University of Alberta

Molecular basis of ubiquitin chain synthesis and recognition

by

Craig James Markin

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry

©Craig James Markin

Spring 2013
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

... and if we were to make the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that *all things are made of atoms* and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms.

Richard Feynman

The Feynman Lectures on Physics

Abstract

K63-linked polyubiquitin is synthesized by the E2 Ubc13, and plays non-degradative roles in immunity and the DNA damage response, yet a full molecular understanding of its synthesis and subsequent recognition remains incomplete. Although previously considered a slow enzyme, we demonstrate that Ubc13 is able to achieve a significant rate enhancement in synthesis of K63-linked Ub_2, even in a putative off-state. Using a non-steady state kinetic approach to measure K63-linked polyUb formation, we are able to measure a true, rather than apparent, k_{cat} for Ub_2 formation. Accurate knowledge of this value provides mechanistic insight that would otherwise remain obscure. These non-steady-state approaches provide the groundwork for quantitative measurement of the activated E2 in the presence of E3 enzymes, the latter currently thought to function in activating their cognate E2s.

In response to DNA double strand breaks, K63-linked polyUb is synthesized on adjacent histones. RAP80 recognizes these chains using ubiquitin interacting motifs (UIMs) and recruits downstream repair proteins. The presence of tandem UIMs and multiple Ub molecules tethered together in the chains leverages multivalency to increase the affinity. We use a novel combination of NMR methods and thermodynamic binding models to dissect these complex interactions to develop a molecular basis for signal amplification through multivalency.

For interactions characterized by relatively fast kinetics, NMR is a powerful method for determining binding thermodynamics. We developed two novel methodologies which increase both the accuracy and precision of thermodynamic values obtained from NMR chemical shift titrations. These are sampling schemes in which the concentrations of analyte and titrant are varied simultaneously. Simulations are used to demonstrate the

potential for increased accuracy and precision while titrations of Mms2 with ubiquitin demonstrate the experimental feasibility.

Whereas chemical shifts provide thermodynamic information, the full line-shapes conceal kinetic information. We demonstrate that classical line-shape analysis allows for determination of kinetics over a broad range of biologically relevant exchange rates, which can be widened using our sampling methods. This provides the opportunity of accurately and precisely quantifying both thermodynamics and kinetics from a single NMR chemical shift titration.

Acknowledgements

Apart from the task of having to articulate and condense a half decade of intense research, thesis writing comes with conflicting emotions. The perspective afforded by joining together papers over a period of five to six years yields a sense of satisfaction and a touch of pride at contributing to new knowledge. Though there is a certain concomitant melancholy. The relationships between people that develop over this long but nevertheless limited period provide a strong sense of comfort and friendship when working together to solve similar problems. With this familiarity comes the confidence to express opinions, both scientific and at times otherwise, and often provides an outlet for the frustration seemingly so tied to research. With the latter is tied the feeling that one never quite feels *done* a PhD.

In any event, looking back on the last half decade, I can say without hesitation that what we've done was worth doing. Appreciation of the opportunity to come in to work every day to try and solve problems, and push the boundaries of one's own knowledge and understanding too easily gives way to nonchalance. Perspective is a strange thing; it's often easier to get mired in what one hasn't achieved rather than what one has.

I remember starting a rotation in Leo's lab in September 2006. I wasn't quite sure where I was going with my graduate studies, but I was leaning towards something involving NMR. Its application to proteins had captured my attention in an undergraduate biophysics course, and it was novel, exciting, and seemed to be a growing field. Admittedly, I didn't know much about ubiquitination at that time, other than its canonical role in controlled protein degradation, which to an undergraduate student seemed a bit *blasé*. But Leo and I met and he shared his interest in lysine-63 linked polyubiquitin signalling, and I decided to spend a month in his laboratory. Following this, I spent similar periods

in the laboratories of Dr. Brian Sykes and Dr. Joe Casey, who graciously allowed me to explore their work while still indecisive about where to study. I'm grateful to them for the opportunity, and also to the Department of Biochemistry for encouraging new graduate students to more fully explore the wide range of available research options.

Although this first brief period with Leo was spent mostly fumbling about the laboratory, I ended up working on an early form of the K63-linked Ub_2 formation assays, which, though they were interrupted by other projects described herein, would become the basis of the work described in chapter 4. At that time the lab was relatively unpopulated, Leo's previous two students having graduated a short while earlier, and I received immeasurable support and instruction from Linda Saltibus. The warmness and patience which she showed me during this time was no doubt a significant contributing factor in my choosing to stay. I'm especially grateful to Linda for her guidance in protein expression and purification, which has given me skills useful to almost every protein biochemistry laboratory in the world. I'm also thankful for her friendship. Reflecting on these early times, I owe a debt of gratitude to Dr. Rachel Miler and Dr. Adrienne Wright for everything learned in Biochemistry 401, the undergraduate laboratory course. Among introductions to the most common techniques used in this multifarious discipline, their insistence on proper record-keeping and labelling was no less essential.

In addition to the assistance provided by Linda, Leo himself was always happy and willing to answer questions. When I had decided to continue in his lab after the rotation period, I was enthusiastic and excited, but there was a sense of trepidation as well. My experience with NMR was mainly limited to undergraduate organic chemistry. I had little to no programming experience, and my familiarity with UNIX was limited to a vague notion of its existence. I'm sure I nearly drove Leo to madness with questions during those first months of learning the basics of the spectrometer, and blundering through *Mathematica* code. Leo, I'm so grateful for your help, and the patience and good humour which accompanied it.

Something as significant as a PhD is not achieved by a single person *in vacuo*. Without the support and, truthfully, the often well-needed distraction afforded by friends and family, none of this would have been possible. After being in one for the past six years, I fully believe that people make a laboratory and similarly, a department. I've

mentioned Linda already, but her conversation and outgoing nature lessened the more onerous aspects of research. Melissa Kean, your sense of humour made the lab more fun. Leo, thanks for the camaraderie, the fellowship, the encouragement, and for always pushing me to be the best scientist I can be. It hasn't always been easy (if ever), but I've learned more than I ever expected and I can be satisfied that we've made some great contributions to the field.

Gratitude is not owed merely to colleagues with whom I worked directly. Thanks go to Olivier Julien and Ian Robertson, for the help they provided by sharing insights and tips gleaned from starting a year before me. Thanks also to other members of the Sykes lab past and present, especially Robert Boyko, Dave Corsen, and Dr. Monica Li for their help. And especially for all the golf. During my rotations outside Leo's lab, I had the pleasure becoming friends with Danielle Johnson and Kate Witkowska. Danielle, I'll never forget you trying to teach me how to snowboard, among all the other good times. Thanks also for answering all my questions about Western blots. Kate, what can I say? Great times. You showed me parts of London I never knew existed. Hopefully we have many more similar nights. Similarly, I feel privileged to have had the friendship of colleagues such as Pam Bonar, Brittany Brown, Roshani Payoe, Curtis Hodge, Stephen Campbell, Charles Leung, Daniel Prins, Mike Burkat, Przemek Gorski, Delaine Ceholski, Allison Kraus, Gina Thede, Lucie Bergdoll, Kristen Conn, Tamara Arnold, and Joe Primeau. Joe, thanks for distractions like that one slightly dubious St. Patrick's Day, as well as for being an unwavering training partner.

Furthermore, I'm grateful to the funding agencies which allow us the opportunity to perform this research, and especially to Alberta Innovates - Health Solutions for the graduate studentship. This research was funded by the Canadian Institutes of Health Research (CIHR), the Alberta Heritage Foundation for Medical Research (AHFMR), and the University of Alberta. I'd also like to thank the Canadian National High Field NMR Centre (NANUC) for use of the facilities, as well as for the assistance provided by Ryan McKay and Deryck Webb. Similarly, the technical help provided by Dean Shieve was invaluable. Thanks also to Sue Smith for help with all the administrative tasks and scholarship applications. Thanks to Junjie Chen for providing RAP80 cDNA, and Pascal Mercier for assistance with programming. Many thanks also to Lewis E. Kay, for provid-

ing pulse sequences. Additionally, thanks to J.N. Mark Glover for many helpful discussions and critical reading of manuscripts. Computational studies described herein were enabled by the use of computing resources provided by WestGrid and Compute/Calcul Canada.

The nature of research is that it is not limited to the hours between 9 and 5. Accordingly, there are many people outside of the laboratory environment whose support and encouragement deserve mention, especially Niels Sorensen, Mark Seniuk, Rav Naidoo, Rob Heydari, Jeff Konowalchuk and Susan Ng, Eric Chamney and Bryce Kustra, who in undergrad made university so fun I stayed another six years. A fantastic time, and I'm thrilled we're all still good friends.

It's difficult to find the right words to express my gratitude to Grant Kemp. Knowing each other since we were fifteen years old, I can't begin to encapsulate what your friendship and support has meant to me. Thanks for everything, and I look forward to your graduation soon.

Special thanks as well to Robert Timewell, for his friendship, and for sharing my interest in European culture and history, as well as continental nightlife. Thanks also to my extended family, Dale, Sonja, Trevor and Lisa Markin, and especially the "British side", Simon and Trudy Goodall, Ken and Judith Muir, and Chris and Jennie Timewell, for their support and for broadening my mind - I'm thinking of the Gladstone Pottery "Toilet Museum", Chris! (and for accommodating me on many holidays!). Special thanks also to my grandparents, Jim and Barbara Goodall, who played no small part in encouraging me to intellectual endeavour while making me a more well-rounded person.

Finally, I'm grateful most of all to my parents, Gary and Karel Markin. Without their unwavering love and support, none of this would have been possible.

Contents

List of Tables

List of Figures

Abbreviations

AIC	Akaike's information criteria
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
BARD1	BRCA1-associated RING domain protein 1
BCA	Bicinchoninic acid
β-TrCP	β-transducin repeat containing protein
BRCA1	Breast cancer 1, early onset
CIN	Chromosomal instability phenotype
CPMG	Carr-Purcell-Meiboom-Gill
CRINEPT	Cross-correlated relaxation-enhanced polarization transfer
DDR	DNA damage response
DSB	Double strand break
DSS	2,2-dimethyl-2-silapentane-5-sulfonate
DTT	Dithiothreitol
DUB	Deubiquitinase
E1	Ubiquitin-activating enzyme
E2	Ubiquitin-conjugating enzyme

E3	Ubiquitin ligase
EMF	Electromotive force
FHA	Forkhead associated domain
FID	Free induction decay
FT	Fourier transform
GST	Glutathione S-transferase
HSQC	Heteronuclear single quantum coherence spectroscopy
INEPT	Insensitive nuclei enhanced by polarization transfer
IPTG	Isoproyl thiogalactoside
IPTG	isopropyl β-D-1-thiogalactoside
iRED	Isotropic reorientational eigenmode dynamics
IRIF	Ionizing radiation induced foci
ITC	Isothermal titration calorimetry
K_{D}	Dissociation equilibrium constant
MATH	Merpin and TRAF homology domain
NEMO	NF-κB essential modulator
NF-κB	Nuclear factor *kappa*-light-chain-enhancer of activated B cells
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
ODE	Ordinary differential equation
PAGE	Polyacrylamide gel electrophoresis
PCNA	Proliferating cell nuclear antigen

polyUb	Polyubiquitin chain
RanGAP1	Ran GTPase-activating protein 1
RAP80	Receptor associated protein 80
RAP80-tUIM	Tandem UIM domains of RAP80
RING	Really interesting new gene
SCF	Skp, cullin, F-box containing protein
SDS	Sodium dodecyl sulfate
SUMO	Small ubiquitin-like modifier
TNF	Tumor necrosis factor
TRAF	Tumor necrosis receptor-associated factor
TRAF6-RD	TRAF6 RING domain
TRIS	2-Amino-2-hydroxymethyl-propane-1,3-diol
TROSY	Transverse relaxation optimized spectroscopy
tUIM	Tandem ubiquitin interacting motif
Ub	Ubiquitin
UIM	Ubiquitin interacting motif

Chapter 1

Introduction

The basis of cellular function

The complete human genome having been sequenced, we currently possess a wealth of information as to the composition of the cell. Although this achievement has undeniably increased our understanding of biology, the task of elucidating the roles of the more than 25000 individual proteins encoded by our DNA remains (1). Proteins comprise much of the cellular machinery needed to carry out the myriad of processes required to sustain life. It is of course possible to make inferences based on sequence information alone, but a thorough molecular understanding of protein function typically results from methodical biochemical and structural studies. The function of every protein, in the most general sense, is to interact with something else. Enzymes bind their substrates before catalyzing their subsequent conversion; the components of muscle tissue form organized arrays of thin and thick filaments, achieving contraction by the disruption and subsequent reestablishment of interactions between actin and myosin across these filaments (2); interactions between proteins and 'first messenger' ligands form the basis of signal transduction, that is, they allow the flow of external information past the cell membrane and begin an orchestrated response; transcription factors bind associated proteins and DNA to control gene expression (3), an event which is often the direct consequence of these initial extracellular signals (4). In some cases, the regulated binding of one region of a protein by another is even essential for proper function (5). Interactions

such as these are the foundation of all the processes occurring relentlessly within the cell.

The relationship between its interactions and function naturally presents a mechanism for controlling the behaviour of a protein. It is not sufficient that proteins merely bind to other molecules; the cell must be able to efficiently modulate these interactions in response to changing metabolic and environmental conditions. Along with regulation of translation, transcription, and degradation, processes which nevertheless indirectly influence interactions by modulating the relative amounts of individual proteins, control of specific protein interactions is the crux of proper cellular function.

Regulating protein interactions

Considering the multitude of processes occurring unceasingly within each individual cell, it is unsurprising that numerous mechanisms of regulating the interactions between proteins and between proteins and ligands are employed. The complex network of cellular interactions is further intertwined by the concerted actions of multiple mechanisms, either working in conjunction or antagonistically, to control and tune biochemical pathways.

A common mechanism of control is covalent modification. Perhaps the most established example is phosphorylation. Phosphorylated residues can influence protein interactions directly, converting otherwise inert sequences to docking sites for partner proteins. The addition of a highly negatively charged phosphate group to a target protein can also effect substantial conformational changes. These changes can potentially expose or occlude a binding site for a substrate or partner protein, achieving control through allostery (6).

The mechanism of covalently modifying a protein in order to influence its capacity to bind other molecules is a very general one. Ubiquitination, the covalent addition of the protein ubiquitin to target proteins, was first discovered as the signal controlling the degradation of cellular proteins (7), but numerous non-degradative signalling roles in diverse cellular processes have since been uncovered (8).

The ubiquitin cascade

Ubiquitin is a 76 residue globular protein found in abundance in all eukaryotic cells (9). In a manner somewhat analogous to phosphorylation, ubiquitin is covalently attached to specific residues on target proteins by a unique set of enzymes in an energy dependent manner. Ubiquitin is able to form an isopeptide bond between its C-terminal carboxyl group and the side-chain amine of a substrate lysine. This addition is a multi-step process, requiring the sequential activity of three classes of enzyme. The first step, catalyzed by a ubiquitin-activating enzyme or E1, results in the formation of a thioester bond between the C-terminus of ubiquitin and a cysteine residue of the E1, using the concomitant hydrolysis of ATP to drive the reaction. This ubiquitin is subsequently transferred to a cysteine of an E2 ubiquitin-conjugating enzyme via transthioesterification; in concert with a ubiquitin ligase (E3), the E2 catalyzes the ligation of ubiquitin to a specific substrate protein in the third step (10) (figure 1.1). Possessing seven lysine residues, ubiquitin itself can act as the substrate for further rounds of ubiquitin addition, forming chains (polyubiquitin, or polyUb) in an iterative manner. The variation in linkage type and length of chain that can be synthesized imparts a topological richness to the ubiquitin signalling network (8). Below, differences in these characteristics are shown to be the key determinants of signal specificity.

Ubiquitin activating enzymes

Ubiquitin is initially prepared for its addition to substrate proteins by its E1, Uba1 (and to a lesser extent, Uba6), using three reversible reactions (12). Uba1 is a monomeric protein with six domains, shown in figure 1.2, over which its various functions are distributed. The first reaction occurs at the active adenylation domain (AAD, figure 1.2, cyan), and couples the hydrolysis of ATP to form a high-energy adenylated ubiquitin (AMP-Ub) moiety, which subsequently binds non-covalently with high affinity at this site (13). The increased reactivity of this high energy species is the driving force for the eventual addition of ubiquitin to a substrate protein. Retaining AMP-Ub allows the catalytic cysteine residue from the second catalytic cysteine half-domain (SCCH) of the enzyme to form the covalent linkage, achieved via a putative conformational change that

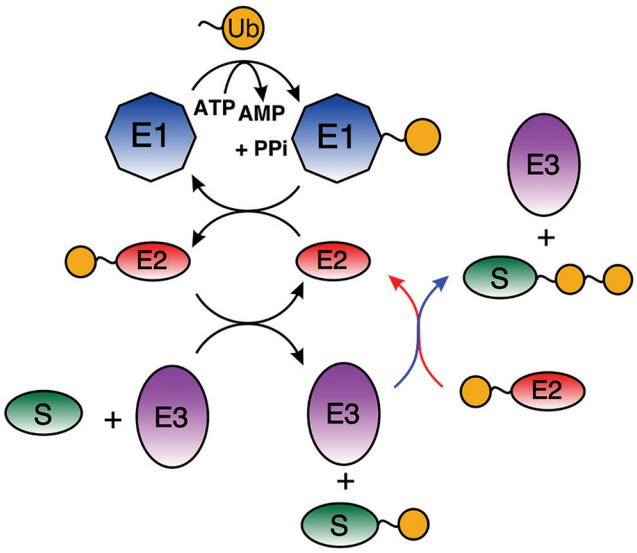

Figure 1.1: The ubiquitin cascade. In the first and only energy requiring step, E1 (blue) forms the high energy adenylated Ub from ATP and Ub (yellow) and subsequently binds the C-terminus of Ub via a thioester bond. In the second step, the bound Ub undergoes a transthioesterification reaction via the catalytic cysteine residue of an E2 (red). In the final step, an E3 ubiquitin ligase interacts with E2 and substrate (S, green) to facilitate the formation of an isopeptide bond between the C-terminus of Ub and a substrate lysine. The reaction of E2 and E3 with ubiquitinated substrate (blue and red arrows, right) illustrates the iterative manner in which polyUb chains can be built. Not illustrated in this schematic is the hierarchical nature of the pathway, with two E1s (Uba1 and to a lesser extent Uba6) responsible for attachment of Ub to the 38 known E2 enzymes. The network of subsequent interactions of these E2s with the more than 600 E3 ligases and their substrates achieves the diverse signalling outcomes of the Ub cascade (11).

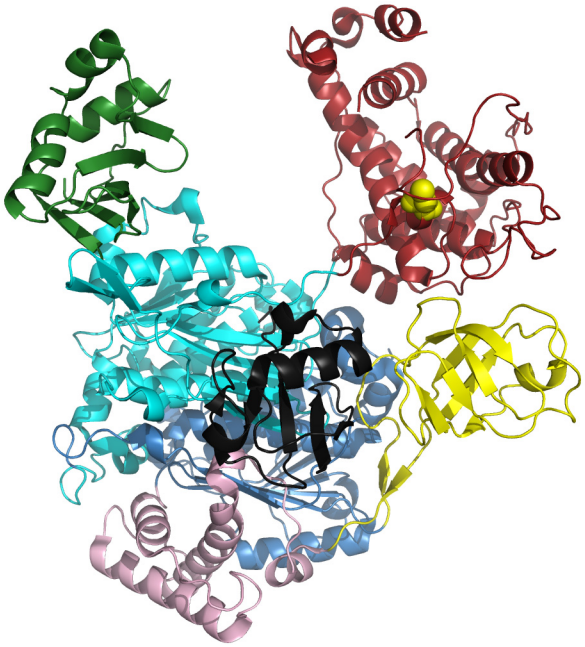

Figure 1.2: Crystal structure of Uba1 non-covalently bound to ubiquitin (PDB ID:3CMM). In blue is the N-terminal inactive adenylation domain (IAD), linked to the first catalytic cysteine half-domain (FCCH) in yellow. This is followed by the four helix bundle (4HB) in pink. In cyan is the active adenylation domain (AAD) with bound Ub (black). Adenylated-Ub remains bound at the AAD so that transfer to the second catalytic cysteine half domain (SCCH, in red, with catalytic cysteine C600 shown in yellow spheres) occurs and a thioester bond is formed between C600 of the E1 and the C-terminus of Ub. The ubiquitin fold domain (UFD, green) recruits an E2 so that transthioesterification occurs and the E2 becomes charged with Ub. Adapted from figure 1 in (14).

brings the reacting groups into closer proximity (14). This frees the ubiquitin-adenylate binding site and a second adenyl-Ub is generated to form a ternary complex in which two Ub molecules are simultaneously bound to the enzyme. The mechanistic purpose of forming this ternary complex is unresolved, but considering the reversibility of these reactions, may assist in driving the overall reaction forward (12). Once ubiquitin is linked to the cysteine via a thioester bond at its C-terminus, the ubiquitin fold domain (UFD) recruits an E2 enzyme. A flexible hinge between the UFD and the AAD allows the catalytic cysteine residues from the E2 and E1 to come into proximity, facilitating the transfer of Ub to the E2 (14).

Ubiquitin conjugating enzymes and ubiquitin ligases

With the exception of the targets of Uba6, Uba1 activates the majority of ubiquitin destined for downstream signalling in the cell (12). In stark contrast to this scarcity of E1s, there are at least 38 known human E2s (11), implicated in diverse signalling pathways. E2s are structurally well characterized, with a $\sim$ 150 residue globular core ubiquitin conjugating (UBC) domain consisting of 4 alpha helices, a 3_{10} helix, and a four stranded antiparallel beta-sheet (10). Certain E2s also have N or C-terminal extensions attached to this core domain, thought to contribute to the functionality of different E2s (15, 16). To facilitate the transthioesterification reaction to maintain a high energy bond for subsequent nucleophilic attack by a lysine side-chain amine, E2s contain a catalytic cysteine residue. Other residues making up the active site are somewhat conserved; however, the mechanisms of action are poorly understood. Non-covalent interactions between the covalently bound Ub and the E2 are sparse, and are limited to the C-terminal tail of Ub (17), though a recent study suggests transient interactions between UbcH5c and covalently bound Ub, and likewise with Ubc13 (18).

The E2 Ubc13 possesses the ability to synthesize both free and substrate bound polyUb chains linked through lysine 63. Ubc13 works in concert with the E2-like, but catalytically inactive, binding partners Mms2 and Uev1A, with which it forms heterodimers (19). Unlike Ubc13, these catalytically inactive Uevs are able to bind ubiquitin and position its K63 side-chain for nucleophilic attack of the thioester bond. In a sense, though it shares the E2-fold, Mms2 plays the role of an E3, but in binding Ub instead of substrate catalyzes the formation of free chains. This binding positions K63 at the active site, accounting for the K63 specificity of this enzyme, but catalytic details remain elusive (20, 21). The mechanism seems to take place via stepwise aminolysis, the C-terminal carbonyl adjacent to the thioester bond undergoing nucleophilic attack by the side-chain of K63 from the non-covalently bound acceptor Ub on Mms2. In a manner somewhat analogous to the stabilization of the tetrahedral intermediate by serine proteases (22), the asparagine (N79 in Ubc13) present in the conserved HPN motif likely stabilizes the oxyanion (23). Recently however, it has been suggested that N79 plays a structural, rather than catalytic, role in the reaction (24), which surprisingly, would

leave the intermediate without an obvious source of stabilization.

At first glance, the proposed mechanism of nucleophilic attack by the side-chain amine of K63 may appear to be at odds with the fact that its pK_a is 10.4. At physiological pH, therefore, the ratio of unreactive protonated -NH_3^+ to the reactive neutral -NH_2 would be 1100. It would therefore seem reasonable that the pK_a is suppressed when the amine comes into proximity of the active site, facilitating deprotonation. For Ubc9, the conjugation enzyme for the ubiquitin-like protein SUMO, comparison of pH rate profiles of the wild type enzyme with those for mutations of residues proposed to suppress the substrate pK_a indeed demonstrates a pK_a downshifting, to which a rate enhancement of $10^{\Delta pK_a}$ can be attributed (25). This work demonstrated a reduced ΔpK_a and decrease in rate when the residues N85, D127 and Y87 were individually mutated, supporting their involvement in this mechanism. It is interesting to note that in regard to the Ubc13-Mms2 reaction with Ub as the substrate, we demonstrated the presence of a similar, albeit weaker, pK_a downshifting of the K63 side-chain amine (-2.6 units compared to -4.1 units in the Ubc9 reaction) (26). In Ubc13 two of these three residues are conserved, N79 and D119, corresponding to N85 and D127 of Ubc9, but the tyrosine is absent (figure 1.3), potentially contributing to the lower ΔpK_a value (figure 1.3). We further demonstrate that pK_a suppression does not account for the overall rate increase of this enzyme, indicating the presence of additional mechanistic factors. These analyses are discussed below in the *Kinetics* section. Comparison of the structures of wild-type Ubc9 and Y87A Ubc9 suggests that the tyrosine residue helps properly orient and restrain the K63 side chain in the active site so that optimal pK_a suppression is achieved (25). In Ubc13 the corresponding residue, D81, has been shown to be important for catalysis, and may play a similar role in positioning the side chain of K63 (27). N123 also seems to play a role in proper orientation by interacting with a backbone NH of the acceptor Ub (21).

There are many more E3s than E2s, so the latter are somewhat promiscuous with regard to E3 binding partners. As illustrated schematically in figure 1.1, E3s facilitate the transfer of ubiquitin from the E2 to a lysine on the substrate, or on a growing polyUb chain. E3 ubiquitin ligases can be subdivided into RING/U-box and HECT domain ligases. There are $\sim$ 30 known HECT domain E3s which, unlike RING/U-box

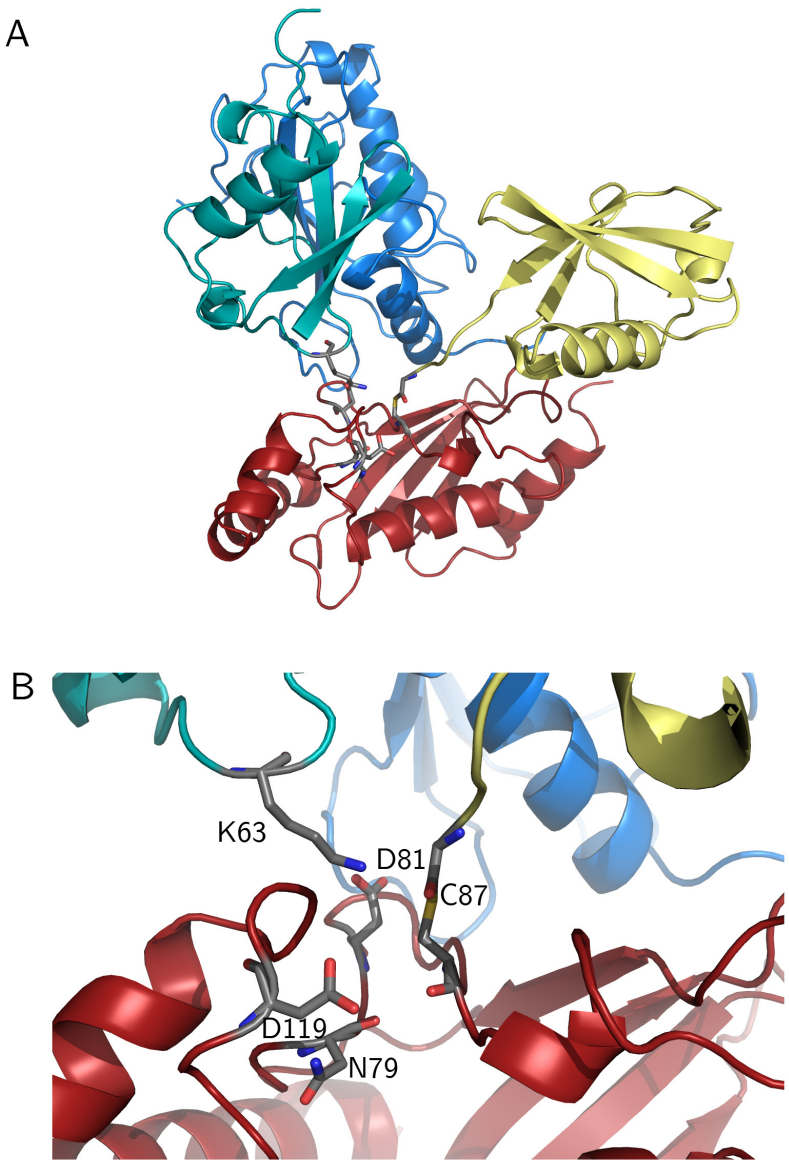

Figure 1.3: A. Model of the tetrameric interaction between charged Ubc13 (red) (Ubc13 covalently modified with donor Ub (yellow), denoted by Ubc13∼Ub), Mms2 (blue), and acceptor Ub (cyan). The thioester bond between the C-terminus of the donor Ub and C87 of Ubc13 is shown in sticks, with K63 of the acceptor Ub positioned in the active site. The mode of binding between Ubc13 and Mms2, and Mms2 and Ub, positions K63 in proximity to the thioester bond, explaining the chain linkage specificity of this enzyme (20, 21). B. Close up of the active site, showing the acceptor lysine in proximity to the thioester bond, and the residues playing a putative role in catalysis, specifically in suppressing the pK_a of the attacking lysine.

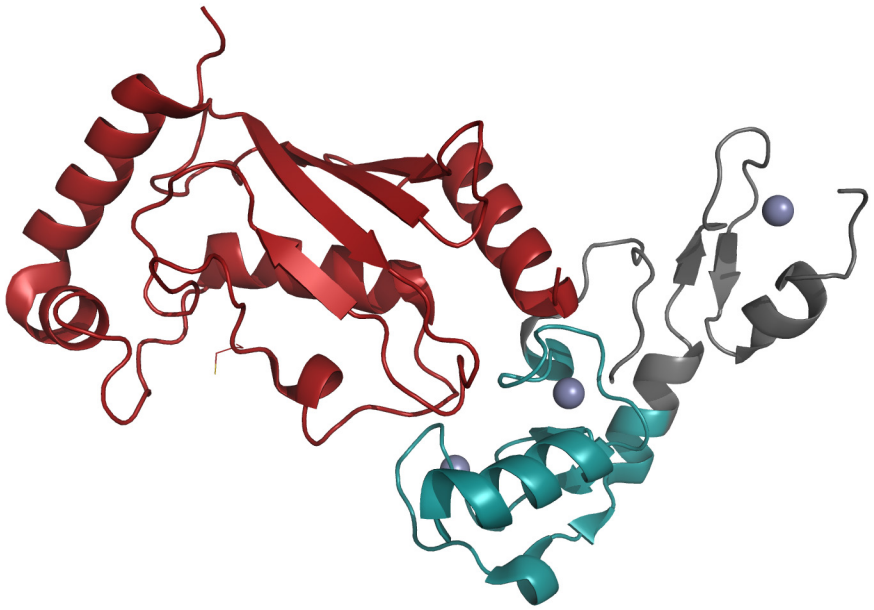

Figure 1.4: Structure of TRAF6 bound to Ubc13 (30). In red is the E2 enzyme, with the catalytic C87 shown in lines as a point of reference. The N-terminal RING domain is coloured in cyan (corresponding to the construct we used in chapter 2), to distinguish it from the first C-terminal Zn^{2+} finger motif. Residues in the Zn^{2+} coordination loops (Zn^{2+} atoms shown as grey spheres) are essential for this interaction, corresponding to the regions typically involved in E2 binding in other RINGs (29).

E3s, form a intermediate thioester bond with the donor Ub before transfer to substrate. E2 binding and the transthioesterification site are located on the HECT domain, whereas substrate binding occurs at diverse N-terminal domains in these proteins (28). The work herein focuses on RING E3 ligases, which make up the majority of E3s. Unlike HECT E3s, RINGs do not form a thioester with donor Ub, but their functions are similarly partitioned, in that the conserved RING domain mediates E2 binding, whereas substrate binding occurs on a variety of separate domains (28, 29).

RING domains are small globular proteins that coordinate two Zn^{2+} ions which are responsible for maintaining the compact fold (32). They are characterized by the conservation of the residues involved in Zn^{2+} coordination, with the canonical consensus sequence being C-X_2-C-$X_{(9-39)}$-C-$X_{(1-3)}$-H-$X_{(2-3)}$-C-X_2-C-$X_{(4-48)}$-C-X_2-C, in which X denotes any residue (29). U-boxes are structurally very similar to RING domains, but a network of hydrogen bonds rather than coordination of Zn^{2+} maintains their stability (33). Remarkably, considering their number, all RING E3s bind E2 enzymes in a

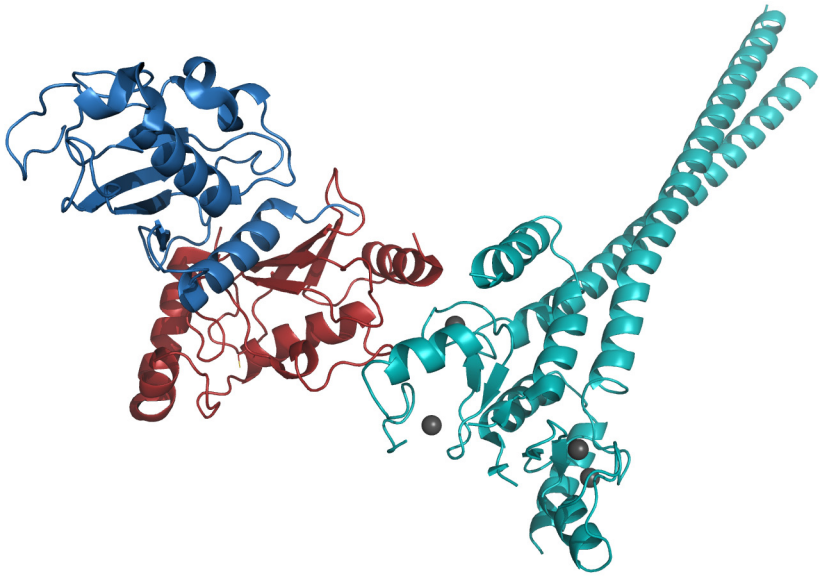

Figure 1.5: Structure of dimeric RNF8 (cyan) bound to the Ubc13-Mms2 heterodimer (Ubc13 in red, and Mms2 in blue). As with the interaction with TRAF6, the Zn^{2+} coordination loops of the RING domain forms interactions with the L1 and L2 loops of Ubc13. Curiously, though RNF8 is a homodimer, only a single Ubc13-Mms2 heterodimer binds to one of the two RING domains (31) (PDB ID: 4EPO).

similar manner, via a shallow cleft on the surface of the RING formed by loops 1 and 2 (figures 1.4 and 1.5). Key residues in these loops interact with residues from helix 1 and loops 1 and 2 of E2 enzymes. While the RING domain is mainly responsible for interacting with the E2, it is typically one of multiple domains or subunits that together facilitate substrate ubiquitination (29).

Although the interactions between many E3s and E2s have been structurally characterized (30, 31, 36, 39, 40), the mechanisms of substrate binding and subsequent ubiquitination remain elusive. For well characterized systems such as SCF complexes the E3 is a multi-subunit protein. The RING fulfills its conventional role in binding the E2, linking the latter to the substrate binding site via the elongated cullin scaffold. The protein responsible for substrate binding typically contains a WD-40 repeat domain or a leucine-rich repeat C-terminal to a F-box domain; the latter links this substrate binding protein to the cullin via an additional adaptor protein, Skp1 in figure 1.6 (41). These substrates usually possess Ser/Thr phosphorylation at specific consensus sites (41); in the case of WD-40 domains, these phosphoserines make specific contacts when bound (37). In these

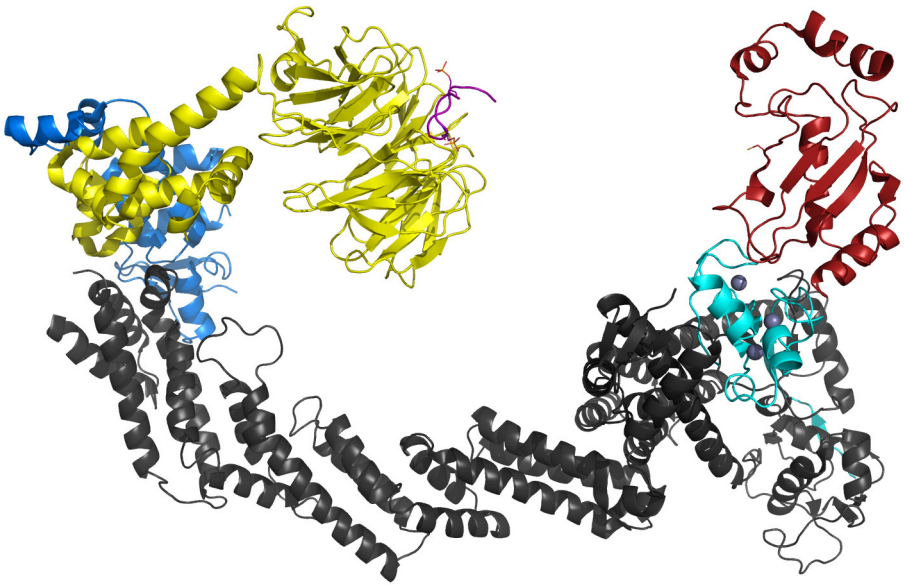

Figure 1.6: Structural model of the $(SCF)^{\beta-TrCP}$ complex. In dark grey is the Cul1 scaffold, interacting with the Rbx-1 RING domain in cyan, with Zn^{2+} atoms shown as grey spheres (PDB ID: 1LDJ/1LDK) (34). The structure of Cdc34 (red) (PDB ID: 3RZ3) (35) was aligned on to the structure of UbcH7 bound to the c-Cbl RING domain (PDB ID: 1FBV) (36); the latter was subsequently aligned with the Rbx-1 RING domain and then removed, in order to model the interaction of Cdc34 with Rbx-1. The Skp1 domain of 1LDK was then aligned with the structure of β-TrCP1 (yellow) bound to Skp1 (blue) and β-catenin substrate (purple) (PDB ID: 1P22) (37). β-TrCP1 contains a WD-40 repeat domain which recognizes the phosphorylated β-catenin substrate, and a F-box domain responsible for interaction with Skp1. All alignments were done using PyMOL (38). To illustrate the ~ 50 Å distance between this substrate and the active site cysteine of Cdc34, the latter side-chain and the substrate phosphoserines are shown in lines.

protein complexes, the E2 binding and substrate binding functionalities are separated in distance, by $\sim$ 50 Å (42) (figure 1.6). The k_{cat} values determined for chain formation and elongation of polyUb on two substrate peptides show distinct dependencies on chain length, suggesting that this distance between the E2 cysteine and substrate binding domains is critical to signalling outcome (43). Furthermore, these distances vary for different substrate binding domains (44). An additional feature of these enzymes is their processivity, whereby the k_{off} for substrate binding is slower than that of E2 binding, so that multiple iterations of chain elongation are achieved before the substrate dissociates from the complex (45). Although the work described herein is mainly concerned with the synthesis and recognition of K63-linked polyUb, chapter 7 details future experiments and methods of analysis designed to understand the mechanisms of these complexes in building polyUb chains.

Regarding other E2-E3 interactions, especially those of Ubc13-Mms2 with TRAF6 (30), RNF8 (31), and RNF168 (31, 46), it remains unclear how transfer of ubiquitin to substrate is achieved. RNF168 possesses MIU domains that interact with ubiquitin (47, 48). RNF8, however, does not seem to bind ubiquitin (31). This would seem to support the latter's role in initiating polyUb chain synthesis at sites of DNA damage, and the putative role of RNF168 in further chain synthesis for signal amplification (49), but actual mechanistic and structural detail of these ubiquitination events is still absent.

There is nevertheless substantial evidence that E3 enzymes act not only as adaptor proteins to facilitate the ubiquitination of substrate proteins, but actually enhance the catalytic rate of ubiquitination. Recently, the RING finger ligase RNF8 was shown qualitatively to greatly enhance polyUb chain synthesis by the heterodimer (31), but as stated above, the mechanism is unclear. Binding studies of the Ub charged E2 UbcH5c with the E3s BRCA1/BARD1 and E4B suggest an allosteric mechanism whereby the E3 reduces the conformational flexibility of UbcH5c Ub, favouring an intramolecular interaction between the hydrophobic patch of the thioester bound Ub with a solvent exposed leucine on the E2 (50). Similar mechanisms have been proposed for the activation of UbcH5a by RNF4 (51) and of UbcH5b by BIRC7 (52). By promoting this intramolecular interaction, the conformation at the active site is presumably altered and made more catalytically viable. Supporting the requirement for active site rearrangement, the crystal structure of

the Ubc13-Mms2 heterodimer with both donor and acceptor Ub molecules bound seems to show an unfavorable orientation for nucleophilic attack, with the carbonyl oxygen rather than carbon oriented towards the lysine (21, 53), so presumably some rearrangement has to occur to allow a more reactive geometry. Furthermore, in the structure of Ubc9 in complex with its substrate RanGAP1 the substrate lysine is ~ 3 Å away from the catalytic cysteine (25, 53); in the crystal structure of charged Ubc13 in complex with Mms2 and acceptor Ub, the substrate lysine nitrogen is 5 Å away (21). The rate of reaction (k_{cat}) in the former is also 330-fold greater, suggesting that this structure reflects a more catalytically potent arrangement. The NMR derived model of the complex also has the N_ϵ further away from C87 than would be expected, at 9 ± 2 Å (20). Additionally, N79 appears to be involved in interactions with the backbone, rather than available for stabilizing an oxyanion (21). Taking all these ambiguities together with the low k_{cat} of K63 polyUb formation (26), these structural models would seem to suggest that E3s might function to allosterically rearrange the active site. This may involve both a rearrangement of N79 and the positioning of substrate amine in closer proximity to the thioester C_α.

Returning to the proposed intramolecular interaction of E2 with its bound Ub, over half of the human E2s with known structure, including Ubc13, possess this conserved leucine (50). RNF8, like E4B and BRCA1/BARD1 with UbcH5c, interacts with Ubc13 via its L1 and L2 loops, so a similar process can be considered a reasonable hypothesis. Indeed, small chemical shift perturbations are observed for UbcH5c and Ubc13 interacting with bound Ub (18). This would seem to contradict the observation that Ubc13 does not bind free Ub (54). However, tethering two molecules with very low affinity together can increase their propensity to bind by increasing the relative concentration of both partners with respect to one another. This is a common theme in biology, especially in ubiquitin signalling, and is discussed in detail below. Nevertheless, if these closed conformations are significantly populated when Ub is covalently attached to Ubc13, one would expect some evidence of weak but nevertheless specific binding between the untethered molecules, especially using NMR chemical shift titrations, which possess high sensitivity to small changes in local chemical environment. Interestingly, the E2-like Mms2 has a glutamine instead of a leucine at this solvent-exposed position and interacts with Ub

in a completely different manner (20), but this cannot be taken as evidence that the leucine facilitates Ub binding in all E2s in which it is present. Considering these recent proposals, the relatively slow k_{cat} measured for Ub_2 synthesis by Ubc13-Mms2 (chapter 4), and the large distance observed between the substrate N_ϵ and thioester bond together with an unfavourable orientration for nucleophilic attack observed in the crystal structure suggest these measurements reflect the catalytic rate of this enzyme in its basal or inactive state. This would explain the characterization of the Ubc13-Mms2 heterodimer as a relatively slow enzyme (55).

Regulatory roles and recognition of K63-linked polyUb signals

So far the discussion has focused mainly on the enzymatic machinery responsible for ubiquinating substrate proteins and assembling polyUb chains. The presence of a polyUb signal on a substrate protein confers recognition by downstream proteins with which it would not otherwise interact. The molecular details of its recognition are therefore of equal importance to those of the synthesis of polyUb. Rather than serving a degradative role, K63-linked polyUb chains are involved in the DNA damage response and in regulating immunity, inflammation, and cell survival through the NF-κB pathway (19).

Ubc13 is an E2 responsible for the synthesis of K63-linked polyUb chains. One determinant of the signalling outcome of K63-linked polyUb is determined by the binding partner of this enzyme. So far, the discussion has focused on Mms2, which shares the E2 UBC fold but lacks the catalytic cysteine. Uev1A, closely related in sequence and similarly inactive from a catalytic standpoint, also forms a heterodimeric complex with Ubc13. It has been shown that Mms2 is the binding partner involved in responding to DNA damage, whereas the Ubc13-Uev1A heterodimer functions to activate the NF-κB pathway (19).

Initiating the NF-κB pathway

The activation of the NF-κB pathway provides an intriguing example of coordination between degradative (K48-linked) and non-degradative (K63-linked) polyUb chain signalling to achieve a desired cellular response. Current knowledge of this complex signalling system is concisely summarized in a review by Liu and Chen (56). The NF-κB protein family consists of five transcription factors which heterodimerize and subsequently translocate to the nucleus to bind DNA and activate target genes involved in regulation of the immune response. In their inactive state, NF-κB proteins are sequestered in the cytosol by the IκB family of proteins. Binding to an extracellular receptor by one of the interleukin-1 (IL-1) family of cytokines or a product of bacterial metabolism indicating the presence of a pathogen results in the intracellular portion of these receptors recruiting adaptor proteins, kinases, and the E3 ligase TRAF6. Recruitment of TRAF6 results in the synthesis of unanchored K63-linked polyUb chains by the Ubc13-Uev1A heterodimer. These chains activate the downstream protein kinase TAK1 by binding to TAB2, an adaptor protein. A cascade is then initiated, activating the kinase IKKβ which subsequently phosphorylates the inhibitory IκB proteins, which become substrates of the $SCF^{\beta-TrCP}$ complex (57), recognized by the WD-40 domain of β-TrCP (figure 1.6) and subsequently modified with a K48-linked polyUb chain. This results in degradation by the 26S proteasome, leaving NF-κB free to translocate to the nucleus. An interesting feature of this signalling pathway is that K63-linked polyUb, though not recognized by the 26S proteasome, nevertheless indirectly results in the controlled degradation of IκB.

Chapter 2 of the present work is concerned with characterizing the dynamics of the RING domain from TRAF6 and explores their biological implications. TRAF6 is a trimer, with a RING domain at the N-terminus, and a peptide binding site at the C-terminus (58, 59). These domains are joined by a coiled-coil domain (58, 59). CPMG relaxation-dispersion experiments (described below) indicate the presence of intramolecular exchange processes on various time-scales. Some are suggestive of slow Zn^{2+} exchange; others, occurring near the E2 binding interface, may play a role in entropic destabilization of the E2-bound complex. These motions may help to maintain weak, transient interactions, perhaps affording a degree of kinetic control to polyUb chain syn-

thesis. Characterization of similar entropically unfavourable exchange phenomena will be a recurring theme in subsequent chapters, especially when discussing binding between K63-linked polyUb and the DNA damage repair mediator RAP80.

Although the full-length TRAF6 protein is trimeric, our data indicate that, unlike many other RING domains, that of TRAF6 is monomeric in solution. The later study of Yin and coworkers (30) came to the conclusion that the RINGs are dimeric, however. In our study of the RING domain our construct did not include the full linker helix, which in the crystal structure contributes to the dimeric interaction. The apparent asymmetry of trimerization at the C-terminus and dimerization at the N-terminus has been suggested as contributing to high-order oligomerization in response to extracellular signals (30); however, as the molecular details of TRAF6-mediated ubiquitination events are not well-understood, it is difficult to ascribe a mechanistic role to higher-order aggregates.

Responding to DNA double-strand breaks

In 2007, a flurry of publications in *Science* demonstrated a role for receptor-associated protein 80, RAP80, as a mediator of the DNA damage response (DDR) (60–62). The DDR functions to either repair DNA damage or induce cellular apoptosis or senescence in order to prevent the damaged cell from replicating (63). DSBs can potentially result in chromosomal defects, which have been associated with the chromosomal instability phenotype (CIN) of many tumor cells. Proper repair or prevention of replication is critical to the prevention of carcinogenesis in such cases, especially during cell division, as loss or gain of large DNA fragments or whole chromosomes can result (64). In much the same way as initiation of the translocation of NF-κB to the nucleus, the general features of the DRR involve the interplay of phosphorylation and ubiquitination to achieve efficient signalling of DSBs. In response to a double strand break event, ATM and ATR kinases phosphorylate S139 of histone H2AX. This serves as a binding platform for the BRCT domain of the mediator MDC1, which, after being itself phosphorylated at specific sites, recruits NBS1 and RNF8 (49, 65). Ubc13 is responsible for K63-linked chain synthesis in this pathway, but interacts with Mms2 and RNF8 rather than Uev1A and TRAF6 to catalyze chain formation (19). In addition to RNF8, a second RING E3 ligase, RNF168, has been identified as directing substrate polyubiquitination, but plays a slightly different

role. RNF8 is responsible for generating initial ubiquitination chains in response to phosphorylated γ-H2AX, whereas RNF168 serves to recognize and amplify the initial ubiquitination events (47). This functionality is consistent with the different domain organization of the two E3s. As alluded to above, in addition to both having RING domains, RNF8 contains a forkhead-associated (FHA) domain (66) whereas RNF168 contains two MIU domains (48), which recognize ubiquitin (47).

The term mediator arises from the fact that these proteins aid the translocation of downstream proteins such as BRCA1/BARD1 to sites of damage. As with MDC1, which seems to play a scaffolding role in bringing damaged DNA and the ubiquitin machinery into proximity (49), RAP80 fulfills a similar role in recruiting repair factors downstream of the ubiquitination signal, recognizing K63-linked polyUb chains through two ubiquitin interacting motifs (UIMs) located in tandem near the N-terminus (60–62). The collective action of these mediators creates DNA damage foci (IRIFs, or ionizing radiation induced foci), microscopically visible cytological structures containing proteins involved in DNA damage signalling and repair, transiently formed in response to genomic damage (67, 68).

Recognizing ubiquitin signals

The UIM (UIM) is a well-conserved class of ubiquitin interacting protein. It is characterized by a single 20-residue α-helix with a consensus sequence of X-Ac-Ac-Ac-Ac-Ψ-X-X-A-X-X-X-S-X-X-Ac-X-X-X-X, where Ψ denotes a large hydrophobic residue, Ac an acidic residue, and X represents residues of relatively low conservation (69). UIMs interact, like the majority of ubiquitin interacting proteins (70), mainly with the L8-I44-V70 hydrophobic patch of ubiquitin (figure 1.7) via the conserved alanine and other hydrophobic residue in the middle of the helix, while forming a hydrogen bond from the conserved serine (69, 71). The role of the N-terminal acidic motif is somewhat uncertain, the side-chain of the second Ac in the above sequence having been suggested as accepting a hydrogen bond from the backbone amides of L71 and R72 in the case of Vps27 (72) or of L73 in the case of RAP80-tUIM (73). In their determination of the solution structure of Vps27 with Ub however, the authors suggested that the purpose of this negatively charged region was to form more general electrostatic interactions with R42 and R72 of Ub (71). These acidic residues may also contribute indirectly to the affinity of binding

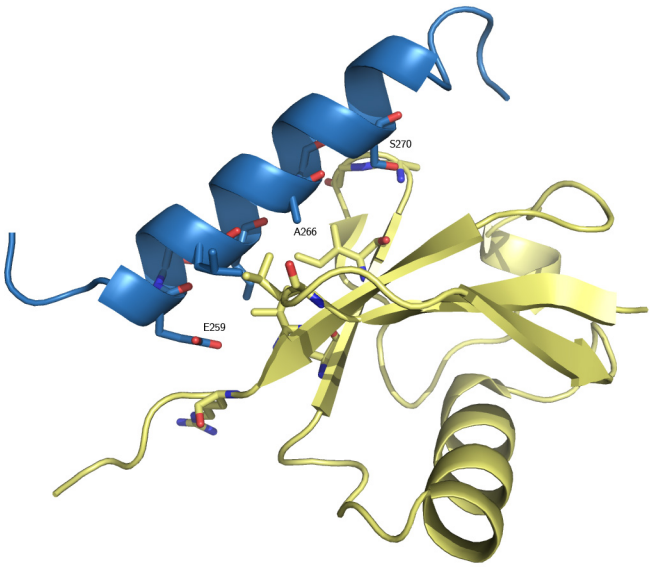

Figure 1.7: NMR structure of UIM-1 from Vps27 bound to Ub. Key conserved interactions between UIM and Ub are shown in sticks. At the centre of the dominant hydrophobic interaction is A266, which interacts with the L8-I44-V70 hydrophobic patch on Ub. The interactions between the negatively charged N-terminus of the UIM and R42 and R72 of Ub, as well as between the conserved S270 of the UIM and the backbone amides of A46 and G47 of Ub are also shown (71).

by stabilizing the helical conformation and thereby increasing the fraction of binding competent UIM. Our work has shown that the RAP80-tUIMs are undergoing rapid and pervading exchange between helical and random coil conformations, both being equally populated at 37 ℃ (74). The role of these acidic residues is a subject of ongoing interest to our group and future experiments designed to address this are outlined in chapter 7.

The strength of UIM-Ub interactions is typically weak, in the high µM range (69, 74). Rather than evolving a tight interaction, UIMs as well as other ubiquitin interacting proteins often leverage multivalency to increase affinity, by being tethered together to favour interaction with polyUb rather than free Ub (75), or by having binding sites on either side of the same UIM helix (76) or UBA domain (77). The prevalence of this tethering in varied cellular signalling systems provides insight into the recognition of polyUb chains and discrimination between the different linkage types. It has been shown that the different chain linkages impart different topological structure to polyUb chains. K48-linked chains tend to be more compact, whereas K63-linked and linear polyUb chains

are more extended (8, 78, 79).

The diversity of ubiquitin signalling immediately underscores the necessity of maintaining signal specificity. An additional complication is the presence of the free pool of Ub, which must be considered in relation to these pathways. True to its name, ubiquitin is found in cells at a concentration of 10-20 µM (80). In spite of this abundance and the varied cellular processes in which it is involved, signalling specificity is nonetheless achieved. Harnessing the multivalency of a polyUb chain to achieve this signal discrimination is a common theme, and we demonstrate its enhancing effect on apparent, or overall, affinity in chapter 3. The presence of multiple ubiquitin interacting sites accounts for the discrimination between polyUb and free Ub, whereas the spatial distribution of these sites seems to form the basis for specificity between chains. Examples from different ubiquitin signalling pathways illustrates this. The UIMs of Vps27 are separated by a 27 residue flexible linker, whereas RAP80 has a much shorter, 10 residue linker. It has been demonstrated that the length of the latter optimizes the UIMs for binding to K63-linked polyUb. Reducing the length and composition of this linker to match the two residues between the UIMs of the K48-binding DUB ataxin 3 has been shown to cause preferential binding of K48-linked chains. Placing two alanines between the RAP80 UIMs result in a preference for K48-linked chains, indicating that spatial organization, not specific interactions between Ub and the linker, is the main determinant (81). That this is a theme not only of UIMs, but of Ub binding in general by UBDs, is suggested by the UBA domain of hHR23a. This is a three-helix bundle capable of binding two Ub molecules simultaneously and which binds K48-linked Ub_2 in a multivalent mode characterized by 1:1 stoichiometry and an increased affinity (77), whereas K63-linked Ub_2 binds more weakly with 2:1 stoichiometry (79). A similar difference is observed in interactions between these two linkage types with the S5a subunit of the proteasome (82). Thus the interplay of topological differences between chains of varied linkage and the relative spacing of multiple UBDs results in contrasting affinities that discriminate between the assorted polyUb chains (70).

Enhancing affinity through multivalency

The presence of multiple UIMs within the same protein and the series of Ub molecules tethered in close proximity in polyUb chains provides the opportunity for multivalent, or avid, binding (83). There are different ways of interpreting this multivalent effect. In the simplest case, it is defined by the simultaneous interaction of multiple binding sites on both partners; that is to say, a substrate with more than one binding site does not necessarily indicate multivalent binding. For that to take place, the cognate interacting partner must also be polyvalent, and the interaction must be such that at least two of these sites on each partner can interact with one another simultaneously. In chapter 3 we apply a thermodynamic analysis to the multivalent binding between the tandem UIMs of RAP80 with polyUb chains. Instead of only being able to interact multivalently (that is, both UIMs simultaneously bound to two adjacent Ub moieties), we show that the overall binding is a combination of this as well as more typical non-multivalent interactions, where one UIM is bound to one Ub, for example, or when both UIMs are simultaneously bound, but to two independent polyUb molecules. Before discussing the affinity gains achieved through these mechanisms, will must first distinguish between multivalent interactions, and what we will term instead polyvalent or multiple monovalent interactions. This latter category encompasses binding of molecules with more than one available binding site but which cannot support a proper multivalent interaction as previously described, in which multiple sites on *both* partners interact simultaneously in an adjacent manner.

In this latter case, in which only one of the partners is polyvalent, it is straightforward to calculate the effect on the apparent binding thermodynamics as observed via an NMR chemical shift titration by solving the system of chemical equilibria describing the individual binding states. As an illustrative example, consider a titration of ^{15}N-labeled ubiquitin with increasing concentrations of the UIM containing protein Vps27. The dissociation constant of the interaction can be measured by following the chemical shift changes ($\Delta\delta$s) observed in ^{1}H-^{15}N HSQC spectra, the methodology of which is described further below. Vps27 is a protein involved in endocytosis, a additional process in which ubiquitination plays a non-degradative role (84). Vps27, having two UIMs, is able

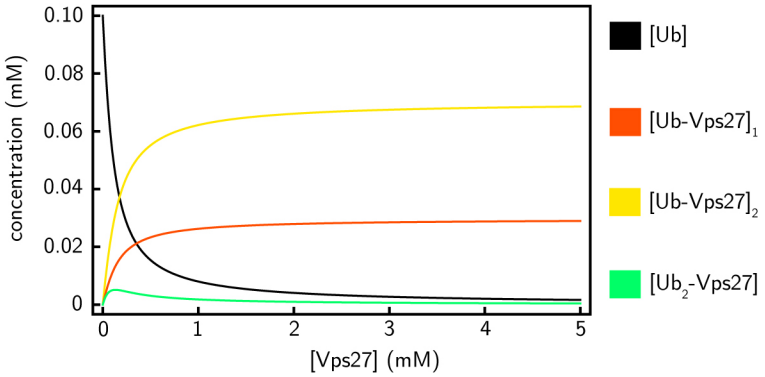

Figure 1.8: Concentrations of the individual binding states of Ub bound to Vps27 (red, yellow, and green curves), as well as that of free Ub (black curve), when titrating 100 µM Ub with increasing amounts of Vps27.

to bind ubiquitin with either or both (71), and a microscopic analysis of these binding modes together requires four equilibria, given in equations 1.1 to 1.4.

$$K_{\mathrm{D},1} = \frac{[\mathrm{Ub}][\mathrm{Vps27}]}{[\mathrm{Ub\text{-}Vps27}]_1} \tag{1.1}$$

$$K_{\mathrm{D},2} = \frac{[\mathrm{Ub}][\mathrm{Vps27}]}{[\mathrm{Ub\text{-}Vps27}]_2} \tag{1.2}$$

$$K_{\mathrm{D},1} = \frac{[\mathrm{Ub}][\mathrm{Ub\text{-}Vps27}]_2}{[\mathrm{Ub_2\text{-}Vps27}]} \tag{1.3}$$

$$K_{\mathrm{D},2} = \frac{[\mathrm{Ub}][\mathrm{Ub\text{-}Vps27}]_1}{[\mathrm{Ub_2\text{-}Vps27}]} \tag{1.4}$$

Solving these equilibria simultaneously with mass-balance expressions for $[\mathrm{Ub}]_0$ and $[\mathrm{Vps27}]_0$, the total ubiquitin and Vps27 concentrations respectively, results in some rather unwieldy equations describing the concentrations of the three bound states as functions of the dissociation constants, $K_{\mathrm{D},1}$ and $K_{\mathrm{D},2}$, and total protein concentrations. Considering their complexity, plotting is more elucidative than simple transciption, although the manner in which these equations are obtained is discussed in more detail in Appendix A. Figure 1.8 shows the concentrations of the three states as a function of total titrant added, when $K_{\mathrm{D},1} = 277$ µM, $K_{\mathrm{D},2} = 117$ µM (the values for the individual UIMs binding to Ub for Vps27, given in (71)) and the concentration of labelled ubiquitin is held at 100 µM. Making the reasonable assumption based on the highly conserved nature of

UIM-Ub binding that the interaction surface on ubiquitin is identical in both complexes (71, 74), the observed chemical shifts will reflect the simultaneous occurrence of all three binding modes, preventing deconvolution of $K_{D,1}$ and $K_{D,2}$; instead, what is measured is an apparent dissociation constant, $K_{D,app}$, described by a 1:1 binding isotherm:

$$[\text{Ub-Vps27}]_{\text{total}} = \frac{1}{2}(K_{D,app} + L_T + P_T) - \sqrt{(-K_{D,app} - L_T - P_T)^2 - 4L_T P_T} \quad (1.5)$$

In equation 1.5, $[\text{Ub-Vps27}]_{\text{total}} = [\text{Ub-Vps27}]_1 + [\text{Ub-Vps27}]_2 + 2[\text{Ub}_2\text{-Vps27}]$, the individual states being indistinguishable. Substituting the equations describing the concentrations of these states (figure 1.8) into eqn. 1.5 and solving with respect to $K_{D,app}$ produces an equation calculating the affinity that would be measured experimentally. As with the equations describing the solved equilibria, it initially appears complex and cumbersome, but can be simplified if the contribution from $[\text{Ub}_2\text{-Vps27}]$ is neglected to the following,

$$K_{D,app} \approx \frac{K_{D,1}K_{D,2}}{K_{D,1} + K_{D,2}} \quad (1.6)$$

which is a reasonable approximation at high concentrations of titrant, as $[\text{Vps27}]_0$ (the titrant) greatly exceeds $[\text{Ub}]_0$ (the analyte), so that $[\text{Ub}_2\text{-Vps27}]$ approaches 0 (figure 1.8, green curve). Regarding this interaction of ubiquitin with Vps27, the $K_{D,1}$ and $K_{D,2}$ values determined from titrations of ubiquitin with constructs of the individual UIMs alone (71) can be substituted into equation 1.6 to calculate a value of 82 µM for $K_{D,app}$. Unsurprisingly, this result confirms the intuitive notion that having multiple binding sites capable of recognizing a partner protein or ligand increases the overall affinity of interaction.

Equation 1.6 is valid in the instance where one component is monovalent and the other divalent. Similar equations can be derived with little additional difficulty for interactions between a monovalent protein or ligand and a polyvalent partner possessing any number of binding sites by solving the system of equations including 1.1 to 1.4 and the necessary additional equilibria.

Returning briefly to Vps27 (71), the authors carried out a titration of isotopically labeled Vps27 UIMs with ubiquitin, but were unable to extract thermodynamic parameters because the former possesses two independent binding sites. In chapter 3 a methodol-

ogy for determining the microscopic K_{D}s for such systems is described and is applied to the interaction of RAP80 with ubiquitin. Although a detailed exposition is beyond the scope of this introduction, it takes advantage of the unique power of the ^{1}H-^{15}N HSQC NMR experiment (85) to observe the effects of a binding event on the individual residues of a labeled protein. This increased 'resolution' allows quantitation of changes induced by ubiquitin addition on residues belonging to both UIMs simultaneously. By constructing and subsequently solving the equilibrium accounting for the possible interactions in much the same manner as done for equations 1.1 to 1.4, the values of $K_{\mathrm{D},1}$ and $K_{\mathrm{D},2}$ that best match the experimental data can be determined by following the chemical shifts of individual residues from both UIMs independently, as a function of total UIM and Ub concentrations. Furthermore, we extend our analyses to characterize the true multivalency, that is the simultaneous interactions between multiple subunits of a polyUb chain and the tandem UIMs of RAP80, that is the source of the greatly enhanced affinity observed in this system (81).

Interpreting multivalency

Unlike the multiple monovalent interactions between a molecule possessing one binding site with a polyvalent binding partner, true multivalency requires that both possess more than one binding site, with binding occurring simultaneously across these sites (83). Compared to the Vps27 example, in which the apparent binding affinity was calculated directly from knowledge of the individual K_{D}s, there is no corollary equation for predicting the strength of a multivalent interaction solely from the component binding equilibria. In his seminal 1981 paper 'on the attribution and additivity of binding energies', Jencks decried the common misconception that "the observed binding energies of two molecules, A and B, are additive in the molecule A-B, so that $\Delta G^{0}_{\mathrm{AB}} = \Delta G^{0}_{\mathrm{A}} + \Delta G^{0}_{\mathrm{B}}$", before explicitly stating that "there is no basis for this assumption" (86). To understand his point, consider ΔG^{0} in terms of its constituent enthalpy and entropy, ΔH^{0} and ΔS^{0}. If the individual interactions constituting a polyvalent system are identical and completely independent, then $\Delta H^{0}_{\mathrm{AB}} = \Delta H^{0}_{\mathrm{A}} + \Delta H^{0}_{\mathrm{B}}$ for two simultaneous interactions, or $\Delta H^{0}_{\mathrm{poly}} = n\Delta H^{0}_{i}$ in the general case, n specifying the number of individual sites and i denoting the enthalpy of the individual interactions. However in practice this is never truly

achieved due to binding at one site imposing a slightly unfavourable conformation or a steric hindrance on the next, reducing the magnitude of ΔH^0 for successive interactions (83).

Multivalent interactions typically achieve affinities greatly exceeding what would be obtained by simply adding the constituent free energies of binding. Much of this gain in affinity is thus contributed by a substantially increased ΔS^0_{poly} (83), as enthalpic gains can only contribute an absolute maximum of $n\Delta H^0_i$, as discussed above. Beginning with a simple case, if one considers the interactions between two completely rigid bivalent binding partners, the formation of the multivalent interaction will be composed of sequential binding events across both sites. The first has a ΔG^0_1 equal to that of an identical interface between two monovalent partners, no interactions between the second site having occurred. Considering what Jencks aptly phrased "the economics of binding" (86), for this intermediate interaction ΔG^0_1 will be composed of an enthalpic gain from the favourable binding and an entropic loss as two proteins free in solution have, as a gross simplification, twice the rotational and translational entropy of a single complex. Considering now the second event, if the two sites are ideally poised so that $\Delta H^0_2 = \Delta H^0_1$, then $\Delta G^0_2 = \Delta H^0_2$, there being no further possible translational or rotational entropic losses (83). Consequently, $\Delta G^0_2 < \Delta G^0_1$, illustrating that the affinity gains observed for multivalent interactions (83) result from decreasing entropic penalties for subsequent binding steps. Entropic effects are thus the main determinant of the affinity gains obtained through multivalent binding (87).

A second way of understanding the multivalent effect is to consider what is often loosely termed 'local concentration'. Quantitive knowledge of this value can be used in combination with the dissociation constants for the individual interactions to calculate the strength of the multivalent interaction. Unlike the above description, this does not require knowledge of the free energy parameters for the stepwise reaction, ΔH^0_i and ΔS^0_i. Furthermore, it is not necessary to consider the interaction as stepwise at all. Instead, we can define a multivalent dissociation constant:

$$K_{\text{D,mv}} = \frac{[\text{P}][\text{L}]}{[\text{PL}_{\text{mv}}]} \tag{1.7}$$

where the subscript *mv* indicates the multivalent bound state. This formation is essentially analogous to the apparent dissociation constant, $K_{\mathrm{D,app}}$, introduced above, as the stepwise binding events are combined in this macroscopic equilibrium. As stated in the previous section, there is no way to calculate $K_{\mathrm{D,mv}}$ with only knowledge of the individual $K_{\mathrm{D},i}$ values. However, by including a term accounting for the increase in local, or effective, concentration of the second site after binding of the first, a theoretical expression for $K_{\mathrm{D,mv}}$ is obtained (88):

$$K_{\mathrm{D,mv}} = \frac{K_{\mathrm{D},1} K_{\mathrm{D},2}}{\rho(\mathrm{r}_0)} \tag{1.8}$$

where $K_{\mathrm{D},1}$ and $K_{\mathrm{D},2}$ are the dissociation constants for the individual interactions, and $\rho(\mathrm{r}_0)$ is the probability density of the end-to-end vector of the linker between the two sites (88, 89). Relating this to the energetics described above, this formulation is equivalent to adding the free energies together with an additional term accounting for the additional decrease in ΔG^0 from entropic effects (87, 88). Interpreting the affinity gains due to multivalency as a higher than in solution concentration of the second binding site is a convenient and intuitive way of understanding this effect. It can be shown that for a general multivalent interaction between a divalent protein and divalent ligand, $\rho(\mathrm{r}_0)$ in equation 1.8 is equal to the effective concentration of the second site by substituting in the concentration terms for $K_{\mathrm{D,mv}}$, $K_{\mathrm{D},1}$ and $K_{\mathrm{D},2}$ and solving for $\rho(\mathrm{r}_0)$:

$$\frac{[\mathrm{P}][\mathrm{L}]_{\mathrm{sol}}}{[\mathrm{PL}_{\mathrm{mv}}]} = \frac{[\mathrm{P}][\mathrm{L}]_{\mathrm{sol}}}{[\mathrm{PL}_1]} \frac{[\mathrm{PL}_1][\mathrm{L}]_{\mathrm{local}}}{[\mathrm{PL}_{\mathrm{mv}}]} \frac{1}{\rho(\mathrm{r}_0)} \tag{1.9}$$

$$1 = \frac{[\mathrm{L}]_{\mathrm{local}}}{\rho(\mathrm{r}_0)} \tag{1.10}$$

$$\rho(\mathrm{r}_0) = [\mathrm{L}]_{\mathrm{local}} \tag{1.11}$$

In the above, $[\mathrm{L}]_{\mathrm{sol}}$ is the concentration of ligand in solution, [P] the protein concentration in solution, $[\mathrm{PL}_1]$ the concentration of the intermediate bound state, and $[\mathrm{L}]_{\mathrm{local}}$ is the local concentration of the the second binding site of the ligand in this intermediate state. Naturally then, $\rho(\mathrm{r}_0)$ has units of concentration.

Having obtained this compact expression describing multivalency in terms of its constituent dissociation constants and a term accounting for the increased concentration of

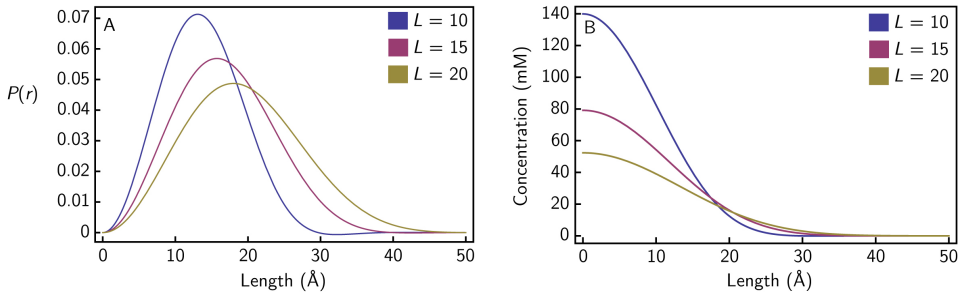

Figure 1.9: A. $P(r)$ calculated using a worm-like chain model and B. Local concentration of the chain end as a function of end-to-end distance, r, for unstructured polypeptide linkers of 10, 15, and 20 residues

ligand, is it not possible to calculate $K_{\mathrm{D,mv}}$ much in the same way as in equation 1.6? Unfortunately it is not as straightforward as the previous case, as $\rho(\mathrm{r}_0)$ is essentially an empirical factor. It depends on the properties of the linker between the two binding sites, such as length, composition, and flexibility (88, 89). From a qualitative standpoint, if the multivalent interaction occurs at r_0, then a short linker will have a large $\rho(\mathrm{r}_0)$, whereas a long and flexible linker will have a smaller $\rho(\mathrm{r}_0)$, due to its capacity for the free end to sample a larger volume relative to the bound end, this latter dispersion resulting in a lower local concentration.

One method of approximating $\rho(\mathrm{r}_0)$ is by modelling protein loops as worm-like chains (89–91). These chains are defined in terms of two parameters, the contour and persistence lengths. The latter is a measure of chain stiffness, whereas the former corresponds to the overall length of the chain. For unstructured polypeptide linkers with total lengths much greater than the characteristic persistence length of the chain ($l_c > 10l_p$ (90, 91)), the probability density of end-to-end distances, $P(r)$, is approximated by (91):

$$\begin{aligned} P(r) =& 4\pi r^2 \left(\frac{3}{4\pi l_p l_c}\right)^{3/2} \exp\left(\frac{-3r^2}{4l_p l_c}\right)\left(1 - \frac{5l_p}{4l_c} + \frac{2r^2}{l_c^2} - \frac{33r^4}{80l_p l_c^3} - \frac{79l_p^2}{160l_c^2} - \right. \\ & \left. \frac{329r^2 l_p}{120l_c^3} + \frac{6799r^4}{1600l_c^4} - \frac{3441r^6}{2800l_p l_c^5} + \frac{1089r^8}{12800l_p^2 l_c^6}\right) \end{aligned} \tag{1.12}$$

In equation 1.12, $l_c = Lb$, where $b = 3.8$ Å, the distance between neighbouring $\mathrm{C}_\alpha - \mathrm{C}_\alpha$ atoms, and $l_p = 3.04$ Å, the value of persistence length taken to best model polypeptide linkers (91). $P(r)$ calculated using this model for chain lengths of 10, 15, and 20 residues is show in figure 1.9A.

Probability however, does not correspond directly to local concentration. For the WLC model, the assumption is that $P(r)$ is dispersed evenly over the surface of the sphere defined by radius r, so that converting to units of concentration requires scaling $P(r)$ by the surface area of the corresponding sphere, $4\pi r^2$, the centre of which is taken to be one end of the chain, so that r also corresponds to the end-to-end distance and the local concentration corresponds to the concentration of the other end. Examination of figure 1.9A demonstrates that very small and very large values of r are unlikely, as chains are rarely fully compact or fully extended, an intuitive result. However, the area that the end of the chain can sample at a particular radius increases proportionally to the square of r so that the effective concentration drops as a function of r. Scaling by the surface area in this manner yields units the somewhat unintutive units of ends per unit volume. As we are only considering a single chain, this is 1 end/unit volume. Dividing by Avogadro's number and converting the length unit of $P(r)$ from Å to dm produces the familiar units of mol L^{-1} (figure 1.9B).

These results have been shown to be consistent with experimental measurements of the binding between linked bispecific antibody fragments to two different sites on the same antigen (92). The WLC model also reasonably reproduces the end-to-end distance distributions obtained via Langevin dynamics simulations of peptides with all attractive forces removed (93).

It should be noted that 1.8 is derived under the assumption that only one of the interacting partners has a flexible linker. It assumes the other is a rigid molecule with two different binding sites distributed on its surface. In many situations this will not be completely true; for instance, the interaction between RAP80 and polyUb chains involves two proteins both having flexibility between binding sites (chapter 3).

Measuring thermodynamics

Having analyzed simple protein-ligand interactions, multisite interactions, and finally multivalent interactions between proteins in terms of microscopic and macroscopic dissociation constants, as well as the associated thermodynamic values of free energy, enthalpy, and entropy, it remains to discuss how these values are actually measured. As alluded

to above in the discussion of simultaneous measurement of K_D values for both UIMs of RAP80, nuclear magnetic resonance (NMR) is a particularly powerful tool for the measurement of interactions between proteins and their cognate ligands or other macromolecules (94, 95).

The use of NMR to measure the thermodynamics associated with a protein of interest followed on the heels of the first reported ^{1}H NMR spectrum of a protein (96, 97). Alongside advances in the assignment and structure determination of proteins using NMR spectroscopy, studies from the 1960s onwards included measurement of residue specific pK_a values (98), characterization of biomolecular interactions such as small drug molecules interacting with proteins (99), calcium binding to proteins such as calsequestrin (100), and the intercalation of small molecules in polynucleotide chains (101).

The most commonly used NMR experiment to obtain quantitative information of binding is the 2D ^{1}H-^{15}N HSQC chemical shift titration. The typical method of conducting this experiment involves adding increasing amounts of unlabelled binding partner to an isotopically ^{15}N labelled protein or peptide, collecting a 2D ^{1}H-^{15}N HSQC spectrum at each titration point. An advantage of NMR chemical shift titrations is that if the resonances are fully assigned the region of the protein responsible for binding can be somewhat elicited, as each amide is an independent probe of the binding event. The labelled protein is typically over-expressed in *E. coli* cells grown in minimal media, containing ^{15}N ammonium sulfate or ammonium chloride as the sole nitrogen source. In this way, ^{15}N is incorporated as the bacteria express the protein under these conditions (102). The protein is then purified and an NMR sample is prepared. Although the appearance of the NMR spectra differ depending on the kinetics of the exchange process (*v.i.*, in *Kinetics by NMR*), for processes characterized by rapid kinetics, the following expression gives the observed chemical shift:

$$\delta_{obs} = f_A\delta_A + f_B\delta_B \tag{1.13}$$

which simply states that the measured shift, δ_{obs}, is the average of the chemical shifts for states A and B, δ_A and δ_B, weighted by the fractional population of each, f_A and f_B. For the opposite case of slow kinetics, rather than an averaged resonance, individual

peaks for both states are present, the areas of which are proportional to their fractional populations. In these cases, the K_D can be determined by integration of the resonances, at various protein and/or ligand concentrations (103).

In the former however, the determination of K_D is less direct. Traditionally, 2D ^{1}H-^{15}N HSQC titrations are conducted with a labelled protein with a concentration of at least 100-200 µM, in order to achieve a sufficient signal to noise ratio allowing the accurate determination of chemical shifts. This requires a large amount of protein relative to other methods for determining binding thermodynamics. The requirement for a high concentration of labelled protein introduces the necessity of adding a high concentration of binding partner. To accurately quantify a dissociation constant from this type of titration requires achieving a high relative degree of saturation, as not only K_D but a second parameter describing the total chemical shift change when the protein is completely bound, $\Delta\delta_{max}$, is necessary. Since this corresponds to the chemical shift at infinite ligand concentration, it is never truly measurable, but as more and more ligand is added the titration curve approaches this value. NMR titrations are consequently always a compromise between having a concentration of analyte in the spectrometer high enough to determine chemical shifts with reasonable accuracy and precision while low enough to achieve a reasonable amount of saturation. In principle, one could add large amounts of ligand to a highly concentrated analyte, but solubility limits often preclude this, especially when the titrant is another protein. Figure 1.10, in which two different approaches to the measurement of a 500 µM K_D and a $\Delta\delta$ of 1.67 ^{15}N PPM are simulated illustrates the effects of this compromise. An arbitrary level of random noise is added to each point in the FID in both cases, and the signal is directly proportional to the total concentration of labelled protein analyte. These simulated data are then Fourier transformed to get the frequency domain spectra shown in figure 1.10, C and D. Apart from providing convenient examples of these titration effects, these line-shape simulations form much of the basis of chapter 6, and are additionally described in Appendix A. Returning to figure 1.10, in the first case, a 1 mM concentration of labelled protein is titrated with 1 mM titrant in 200 µM steps. In the second, a ten-fold dilution of labelled protein is titrated in an identical manner. These simulations were each carried out 100 times (adding random noise each time), and the chemical shifts were determined by fitting to

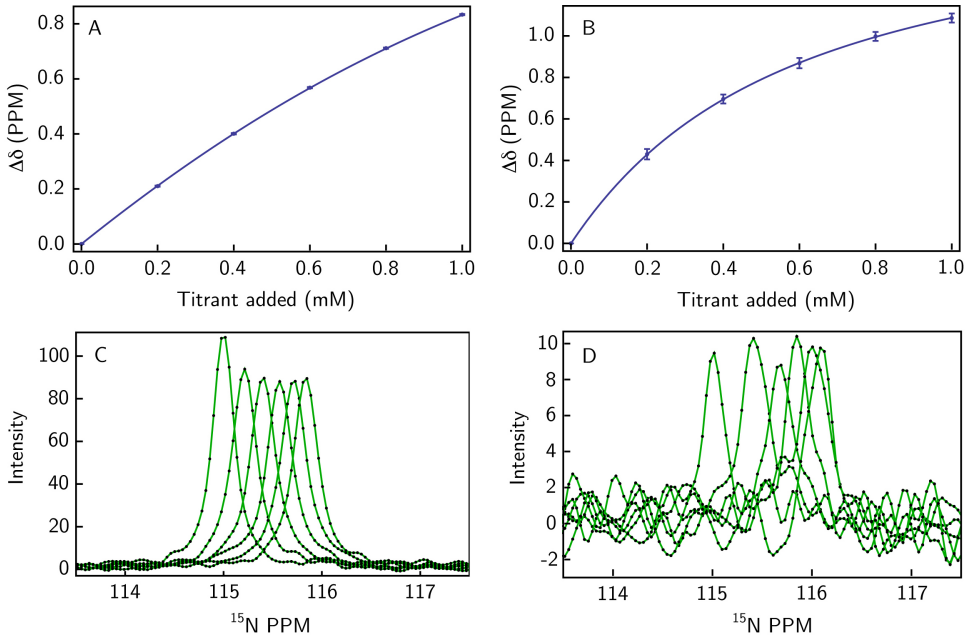

Figure 1.10: Two different approaches to the measurement of a 500 µM K_D

Lorentzian lineshapes. The chemical shift changes were then used to determine the K_D and $\Delta\delta$ by fitting these two parameters to the following equation, which is equation 1.5 recast in terms of the observed chemical shift change $\Delta\delta_{obs}$, the fractional amount of bound protein ($f_b = [PL]/P_T$), and the overall chemical shift change $\Delta\delta_{obs} = f_b\Delta\delta_{max}$:

$$\Delta\delta_{obs} = \frac{(K_{D,app} + L_T + P_T) - \sqrt{(-K_{D,app} - L_T - P_T)^2 - 4L_T P_T}}{2P_T}\Delta\delta_{max} \qquad (1.14)$$

In the first case the increased signal to noise ratio is immediately evident (100 as opposed to 10, the standard deviation of the random noise intensities being approximately 1); however, the overall level of saturation is lower than in the second, in which larger chemical shift changes are observed (*cf.* 0.8 ppm in A and 1.0 ppm in B). The precision of the chemical shift determination is dependent on the overall signal-to-noise of each peak; the standard deviation of the $\Delta\delta_{obs}$ values in the case of 1 mM analyte are ten-fold more precise than those in the 0.1 mM case (0.002 ppm as compared to 0.02 ppm in the latter). One might then expect that the precision of the fitted K_D and $\Delta\delta$ values to be correspondingly more precise in the 1 mM simulation. Surprisingly however, this is not the case (*cf.* Table 1.1). Not only are the determined values of K_D and $\Delta\delta$ more

Table 1.1: Fitted parameters for titrations of 0.1 mM and 1 mM protein analyte

[Analyte] (mM)	K_D (µM)	$\Delta\delta_{max}$ (PPM)
0.1	510 ± 60	1.68 ± 0.08
1.0	544 ± 100	1.7 ± 0.1

precise in the 100 µM example, they are more accurate as well. So from a practical perspective, K_D fitting would seem to be less susceptible to errors caused by less precise shifts than to lack of saturation. It is therefore often better to sacrifice signal-to-noise in the interest of increasing the level of saturation. Closer examination of A and B in figure 1.10 highlights the increased curvature of the latter, which allows for a better estimate of $\Delta\delta_{obs}$. In the former, the function is almost a straight line, so it is difficult to 'pin down' the parameters, even though there is very little error on the chemical shift measurements themselves. It could be argued that one merely needs to add more ligand to the 1 mM case to increase saturation while maintaining excellent signal-to-noise, but often the titrant is another protein burdened with poor expression or limited solubility.

Clearly then, obtaining accurate and precise measurements of K_D proves to be a compromise. In 1983, Joseph Granot examined this in detail and determined that the most precise measurement of K_D for a 1:1 interaction is obtained when the protein concentration is held at $0.5 \times K_D$ and the ligand is varied between $0.4 \times [P_0]$ and at least $11 \times [P_0]$, to achieve sufficient saturation (104). Unfortunately, there are two cases which immediately confound Granot's method. The first, as mentioned above, is that if the full addition of $11 \times [P_0]$ of unlabelled binding partner is thwarted by limited solubility, either the titration will only be taken to the solubility limit or the restriction that $[P_T] = 0.5 \times K_D$ will have to be lifted. In either case, accuracy and precision will be reduced. The second possible issue is one of signal-to-noise. Should the K_D be below the minimum concentration of protein necessary for adequate chemical shift determination, $[P_0]$ will unavoidably be increased and again the precision and accuracy will suffer. Finally, though it cannot strictly be considered an impediment, an additional complication of this method is that it requires knowledge of the K_D before beginning the titration.

While quantifying the strength of binding of the tandem UIM domains of RAP80 with free ubiquitin and polyubiquitin chains, we experienced several challenges that eventually

led us to developing two novel methods for performing NMR chemical shift titrations. Although ubiquitin is a relatively straightforward protein to over-express, purify, and concentrate, polyUb chains can not be expressed and concentrated to the same levels. Further complicating matters, the binding between UIMs and ubiquitin is relatively weak, on the order of a few hundred micromolar. Achieving sufficient saturation therefore requires millimolar concentrations of titrant. This in itself is not overly difficult; these proteins can all be expressed up to at least 1 mM. The difficulty begins when actually conducting the experiment: inevitably, adding high concentrations to an NMR sample will cause volume changes, reducing the concentration of $[P_0]$. If, as is likely, a stock solution of titrant possessing a high enough concentration to allow addition to the NMR tube without causing excessive volume changes is unachievable, there are two options for conducting the titration. The first is to prepare an individual sample for each titration point; the caveat here is that this requires n times the amount of labelled protein, n being the total number of desired titration points. Although it is common to carry out these titrations with a constant $[P_0]$, examination of equation 1.14 highlights that this equation is a function of both $[P_0]$ as well as $[L_0]$. This is illustrated in figure 1.11, where the binding isotherm is shown as a three dimensional surface instead of a two dimensional curve. The traditional way of performing a titration is illustrated with the red and green points. As this does not actually eliminate $[P_0]$ from the equation there is no reason to carry out a titration at constant $[P_0]$, other than the convenience of plotting in only two dimensions. When conducting our titrations of RAP80-tUIM with polyUb chains we added progressive amounts of titrant and allowed the concentration of analyte to drop according to the volume change of the mixture, a more convenient alternative to different samples for every point. At the time, this seemed the most efficient and facile method of obtaining the necessary thermodynamic data in order to develop our multivalent analyses for this system. Following the completion of this analysis, we hypothesized that since the binding isotherm is in fact a function of two variables, then perhaps sampling the effect of changes in both variables simultaneously would not only make titrations more practically feasible, but also result in higher precision in fitted values of K_D and $\Delta\delta$. Such a sampling strategy is illustrated by the blue and magenta points in figure 1.11. In chapter 5, it is demonstrated that this covariation of analyte and titrant concentrations

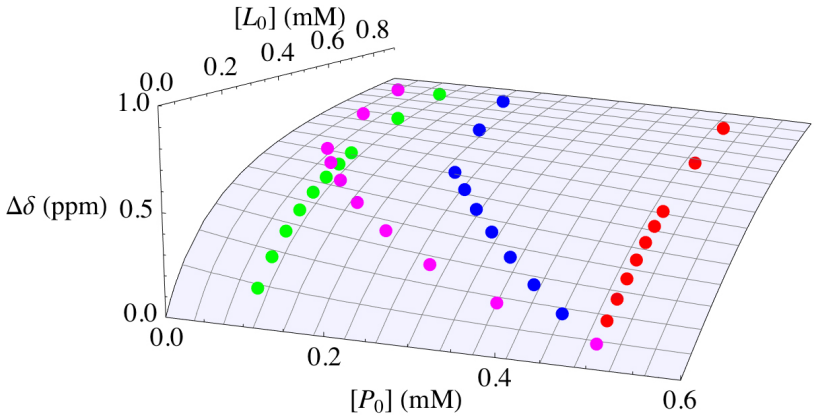

Figure 1.11: Sampling the three dimensional binding isotherm for a 1:1 interaction with $K_{\mathrm{D}} = 200\ \mu M$

simultaneously can achieve precision approaching that obtained via titrations optimized according to Granot's strategy (104). Furthermore, this methodology does not require prior knowledge of the K_{D}; nor is an excessively high concentration of titrant necessary for saturation, due to the decrease in $[\mathrm{P}_0]$.

NMR titration data is often plotted as a function of analyte:titrant ratio. This is a convenient way of presenting the extent of saturation reached in the titration. However, these plots can lead to the erroneous assumption that binding data can actually be fit to the analyte:titrant ratio, instead of the individual concentrations of both components. Although the importance of relative saturation has been stressed, one must be careful not to fall prey to the misconception that the fraction of analyte bound (and by extension, the observed chemical shift change) is directly proportional to this ratio. Figure 1.12 plots the fraction of analyte bound as a function of analyte and titrant concentrations, when the relative ratio is held constant at 1:2. Specifically, the analyte concentration is plotted from 0 to 3.2 mM, whereas the titrant concentration ranges from 0 to 6.4 mM. Therefore, as the total amount of analyte and titrant increases, so does the fraction bound; nevertheless, their ratio remains constant. That this plot is not a straight line immediately highlights the fallacy of plotting titration data solely in terms of analyte:titrant ratio. The only case in which this is acceptable, is if the concentration of one of the components remains fixed; even this, however, is not recommended as the data are only reproducible if all subsequent experiments are carried out not only at the same

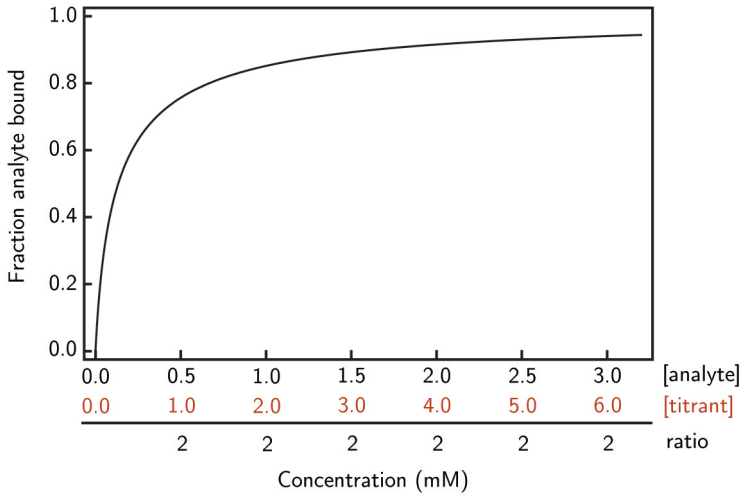

Figure 1.12: Fraction total analyte bound for different total concentrations of analyte and titrant in a 1:2 ratio, respectively

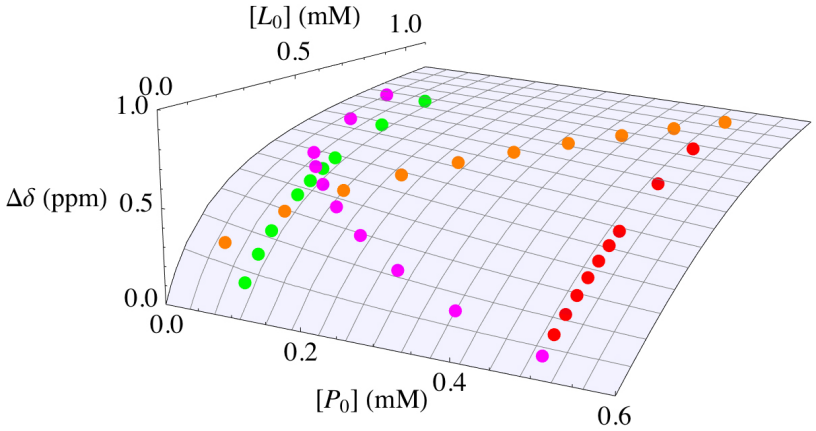

Figure 1.13: A second sampling strategy for a 1:1 interaction with $K_{\mathrm{D}} = 200\ \mu M$

ratios, but using the same absolute analyte and titrant concentrations at every point.

It turns out that the dependence of fraction bound on the total concentrations at constant ratio illustrated in figure 1.12 can be harnessed to obtain K_{D} values having precision comparable to Granot's optimized titrations. This methodology is shown in orange in figure 1.13. In this particular example, a mixture of labelled analyte and unlabelled titrant in a 1:2 concentration ratio is prepared, then serially diluted with buffer for each subsequent titration point. This is analogous to our earlier method, in which both analyte and titrant concentrations are varied throughout the titration, except that the sampling direction is perpendicular to the first (*cf.* the orange points demonstrating

dilution with buffer and the magenta points reflecting dilution with titrant). Chapter 5 presents simulations that indicate both increased precision and accuracy from use of these sampling schemes, and experimental NMR chemical shift titrations of Mms2 with Ub to demonstrate their practical feasibility.

Kinetics

Thermodynamics yields understanding of how polyvalency and multivalency are able to enhance binding affinities, but does not give quantitative information of the rates at which these processes take place. It may be suggested, however, that it offers some insight; measuring a dissociation constant in the mM range, for example, one might be inclined to make the assumption that the rates of binding and dissociation are relatively fast, based on the relationship $K_D = k_{off}/k_{on}$. However, on-rates for protein association can range from 10^3 to 10^9 M^{-1} s^{-1} (105), so for a 1 mM K_D the corresponding off-rate can consequently be as slow as 1 s^{-1} or as fast as 10^6 s^{-1}. Similarly, if two molecules bind with high affinity, it might be concluded that k_{on} is fast and k_{off} is slow. In the vast majority of cases this would likely be the case, but not necessarily so. There are many examples, for instance, of enzyme inhibitors which bind tightly yet associate relatively slowly (106, 107). This slow, tight binding is important for studying their cognate enzymes, as the phenomenon results in different observed reaction curves depending on whether the reaction is started by adding substrate to enzyme pre-incubated with inhibitor or by adding enzyme to a mixture of inhibitor and substrate (*cf.* Figure 1 in (106)). So an understanding of the kinetics of a reaction is important to understanding empirical data, but it is just as necessary for understanding reactions in a biological context.

It could be argued that a study of kinetics is more useful than thermodynamics, as the latter can be calculated from the former. Unsurprisingly, given their higher information content, kinetic measurements tend to be more challenging. Unlike most thermodynamic measurements, which can be probed by equilibrium methods, the nature of measuring a rate often necessitates perturbing the equilibrium in some manner and measuring the resultant recovery as a function of time. NMR, however, can be a powerful tool for

measuring kinetics in many cases, even while maintaining chemical equilibrium. To summarize the techniques used in this work to quantify kinetics of K63-linked polyUb chain synthesis and recognition, it will be convenient to somewhat arbitrarily divide into two groups the types of biological systems for which kinetics are typically of interest.

In the first, which can be considered more commonly amenable to study using NMR techniques, are the association/dissociation rates for protein-protein and protein-ligand interactions, which have already been discussed from a thermodynamic perspective. Also in this category fall intramolecular kinetics, for instance, folding and unfolding events, domain motions, and other structural reorganizations. The important feature of these processes is that they are highly reversible; in other words, they can be characterized over time. This is especially important for multidimensional NMR, where the necessary experiments typically require hours and even days in some cases to complete. Making up the second category are reactions which involve chemical conversion of a reactant to a product, in an irreversible manner. In terms of biological systems, this is typically the binding of a substrate to an enzyme followed by its subsequent conversion. These reactions are typically more difficult to characterize using NMR methods. Considering enzymes exist to enhance reaction rates, and preparing an NMR sample requires a finite amount of time for adding, mixing, and optimizing the spectrometer for that particular sample, often the reaction is complete by the time the acquisition is started. Measuring kinetics for these systems more often than not requires bespoke biochemical assays, in which either a build-up of product or a decrease of substrate is monitored.

Kinetics by NMR

Fundamental to all NMR kinetics measurements is the overall rate of exchange, k_{ex}. For a two-state intramolecular process, $k_{\text{ex}} = k_{\text{on}} + k_{\text{off}}$, whereas for an intermolecular binding event, $k_{\text{ex}} = k_{\text{on}}[\text{L}] + k_{\text{off}}$, [L] being the concentration of unlabelled binding partner (108). The magnitude of k_{ex} determines the timescale of the exchange process, which in turn dictates the NMR experiments best suited to measurement of the kinetics (figure 1.16). A qualitative explanation of exchange necessitates a brief summary of the physics of an NMR experiment (see reference (109), for an excellent detailed discussion of NMR physics). Due to the presence of a strong external magnetic field in the spectrometer, the

vector sum of the individual nuclear spins aligns with this field, along the z axis. Although nuclear spin is a quantum mechanical property, for the purposes of describing relaxation and chemical exchange, a classical treatment of this net magnetization vector will suffice. By applying a radiofrequency (RF) field along the x or y axes, the magnetization vector experiences a torque which rotates it into the xy (transverse) plane. The detection of the NMR signal takes place along x and y, so that once rotated from z measurement can begin. The magnetization now experiences a similar torque from the static magnetic field and begins precessing about z. The frequency of this rotation is related to the measured chemical shift. The precession frequency or chemical shift experienced by the nuclei is not solely dependent on the static magnetic field strength, however; rather it is sensitive to the the net field present in its unique chemical microenvironment. A proton in an electron rich environment will experience a slower precession frequency than it would in isolation, being somewhat shielded from the full effect of the external field. The impressive sensitivity of the spectrometer to these small local field differences contributes greatly to the power and value of this technique. Without these effects, all the nuclei in a protein would experience the same chemical shift and manifest as a single overlapped peak, rather than the wide dispersion of peaks actually observed.

When nuclei are interconverting between states, however, the signals measured by the NMR spectrometer become a combination of the two states. The ability of NMR to measure kinetics is rooted in this transfer of the observed nuclei from one magnetic environment to another. This transfer reflects the physical kinetics of exchange and allows their quantification. However, due to the presence of the external field, the magnetization will not remain in the detection plane, but will decrease in a time-dependent manner. The mechanisms responsible for this realignment are together referred to as relaxation. It will be necessary to briefly describe these processes in order to understand the effect of chemical exchange, as it is itself a mechanism of relaxation.

In the absence of chemical exchange, there are two mechanisms of relaxation, referred to as spin-lattice and spin-spin relaxation, or more commonly T_1 and T_2, respectively. T_1 relaxation refers to the realignment of the magnetization vector with the z-axis. We will be more interested in the latter mechanism for the purposes of the present discussion, as exchange manifests as an additional contribution to T_2, which for proteins is much shorter

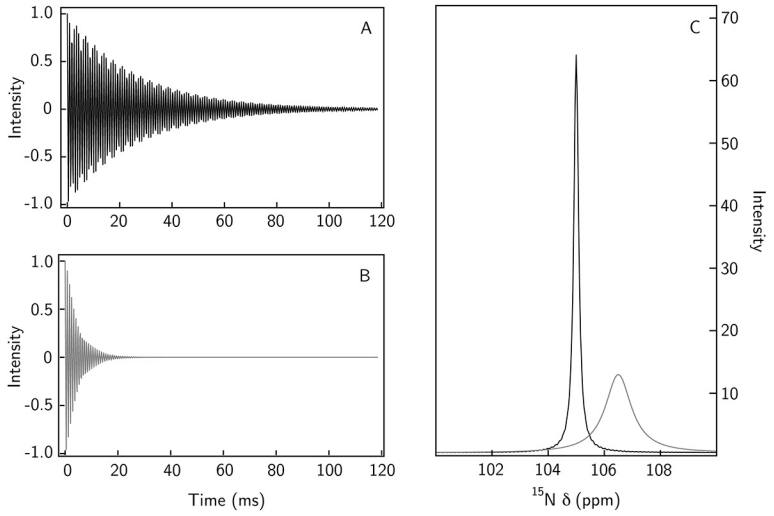

Figure 1.14: The effect of chemical exchange on the NMR signal in the time and frequency domains. (A) A simulated FID for a ^{15}N signal in the absence of chemical exchange; the signal decay is a function of the intrinsic $R_2^0 = 12.0$ Hz. (B) A simulated FID for the same signal in the presence of an intermolecular exchange process (*e.g.* ligand binding) with a $k_{ex} = 2000$ s^{-1}, $\Delta\delta = 180$ Hz, and p_A and p_B both 0.5, and $R_2^0 = 12$ s^{-1} for both the free and bound state; the signal decay is a function of the sum of the intrinsic $R_2^0 = 12$ Hz and the contribution from the exchange process, $R_2^{ex} = 50.9$ Hz, where $R_2^{ex} = p_A p_B \Delta\omega^2 / k_{ex}$, since $k_{ex} \gg \Delta\omega$ (108). (C) Fourier transforms of the simulated FIDs in (A) and (B). FIDs and line-shapes were calculated as described in Appendix 1

than T_1. From a classical perspective, transverse relaxation results from dephasing of the individual spins in the xy-plane as they experience slight differences in the magnetic field; by extension, transfer between sites also contributes to this dephasing, as the average precession frequency (or, from a slightly different perspective, the total angle precessed over the measurement period) for each spin will be slightly different, due to each spending different amounts of time in each state. The net effect of this is that the overall signal decays more quickly than the intrinsic R_2^0 rate due to the magnitude of the vector sum of spins decreasing over time as this dephasing continues. Visually, this manifests as broader line-shapes for exchanging residues (figure 1.14).

Figure 1.14 illustrates the following convenient benefit of using NMR spectroscopy. Often an initial qualitative assessment of the kinetics of any ongoing exchange process can be observed in the shape of the line-shapes themselves. Depending on whether or not a given nucleus is experiencing exchange and the lifetimes in each state, the observed signals will manifest differently. For intramolecular processes occurring on the µs-ms

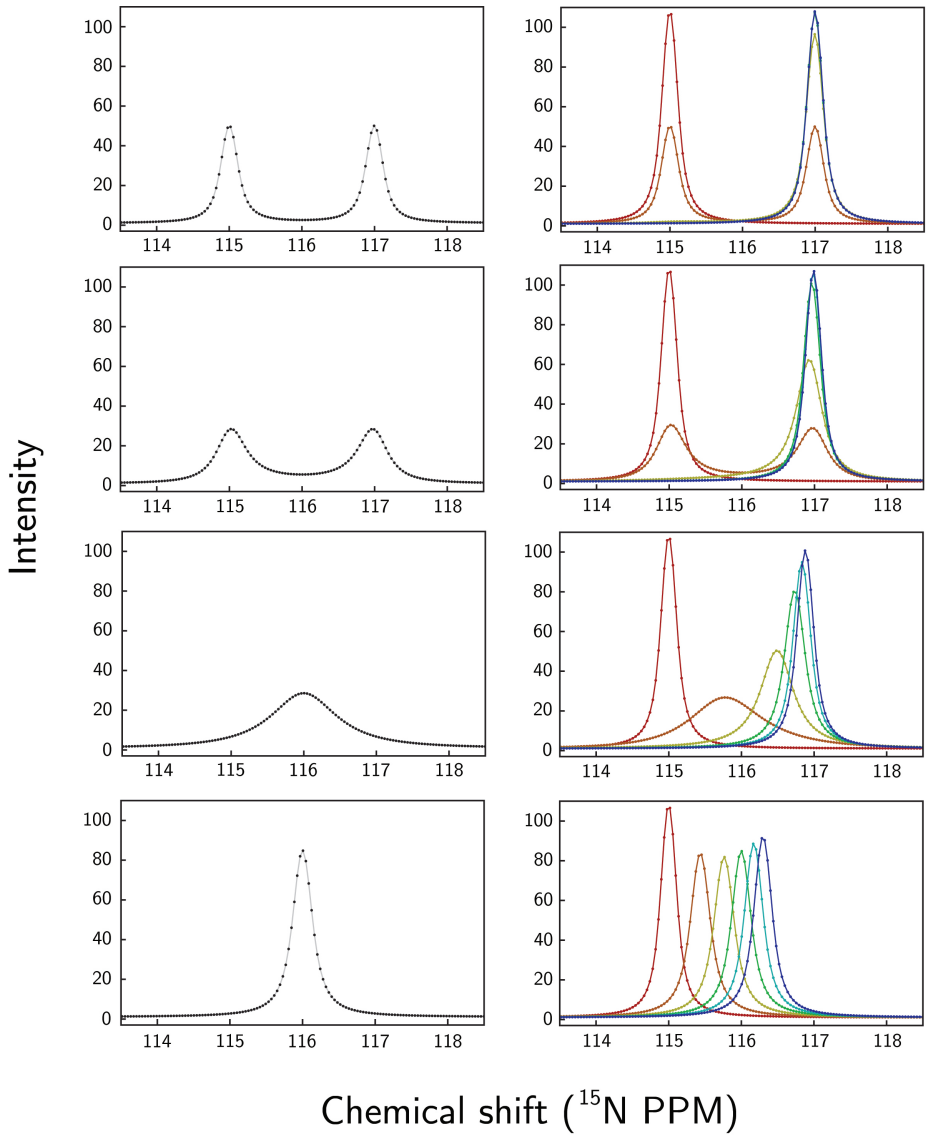

Figure 1.15: Lineshapes for various values of k_{ex}. The left panels illustrate the line-shapes observed for an intramolecular process with k_{off} values (top to bottom) of 5, 50, 500, and 5000 s^{-1}. As the exchange rate increases relative to $\Delta\omega$ the two distinct peaks merge in to a single, population-averaged peak. The right panels illustrate NMR spectra expected from an NMR titration, where increasing amounts of binding partner are added to an isotopically labelled protein. The red peaks simulate 0.5 mM labelled protein in the absence of added ligand, and the remaining peaks reflect the additional presence of 0.25, 0.5, 0.75, 1.0, 1.25, and 1.5 mM ligand (orange to blue curves). As with the intramolecular processes on the left, k_{off} values used were 5, 50, 500, and 5000 s^{-1} (top to bottom). k_{off} is used instead of k_{ex} for purposes of clarity due to $k_{on}^{'} = k_{on}[L]$ for intermolecular exchange; thus $k_{on}^{'}$, and consequently k_{ex}, varies with the amount of added ligand. Simulation of these line-shapes is described in detail in Appendix A. This figure is adapted from figure 1 in (108).

timescale (typically $k_{ex} \leq 5000\ s^{-1}$) the peaks from residues undergoing exchange become broader and relatively less intense than peaks not undergoing exchange. Instead of peak broadening being observed for kinetics in the low ms to s timescales, these residues possess individual peaks for the different states, with the areas proportional to the populations of either state. So for instance, for a two-state exchange process occurring with slow kinetics, two distinct peaks will be observed (Figure 1.15). In practice, one or both of these peaks can often not be actually observed, due to the large degree of broadening lowering the peak intensities below the spectral noise (*cf.* peaks and noise in figure 1.10, C and D). For k_{ex} values greater than a few thousand s^{-1}, a single peak is observed, but there is little to no additional broadening from the presence of an exchange process. Elucidation of kinetics in this regime requires indirect NMR methods. Figure 1.15 also illustrates the differences between spectra of intramolecular processes (left panels) and intermolecular processes (right panels). In the case of intermolecular interactions, by varying one of both of the components and acquiring additional spectra, the peaks either shift progressively or, if the timescale of exchange is slow, then the new populations of each state are reflected in the relative areas of two distinct peaks.

Relaxation experiments

As described above, the kinetics of the exchange process determines the manifestation of the observed NMR signals. The measurement of kinetics for slow and fast exchange require different NMR experiments, each optimized for a particular range of k_{ex} values. In this work described herein, we have had occasion to probe kinetics over most of these timescales (figure 1.16). To determine the k_{on} and k_{off} for Ubc13 binding to Mms2, we used ZZ-exchange experiments, which measure slow exchange processes, for which individual and isolated peaks reflecting each state are observed (111); line-shape analysis was also used as an independent measure of this interaction. Line-shape analysis was used to estimate the kinetics of RAP80-tUIM binding to polyUb as well, this technique being able to probe a wide range of exchange timescales. Chapter 6 specifically deals with the applicability of line-shape analysis to quantify a wide range of exchange processes. For the TRAF6 RING domain, CPMG relaxation-dispersion was used to characterize both zinc and DSS binding by this protein, as well as some putative intramolecular motions

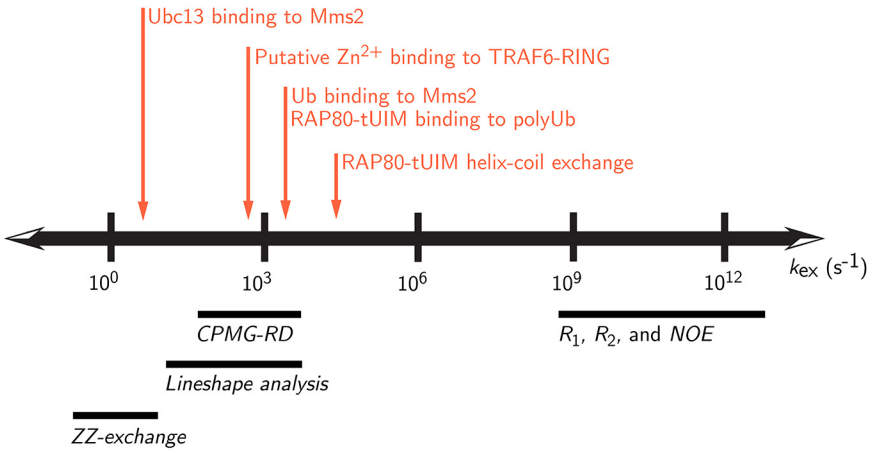

Figure 1.16: Exchange rates (k_{ex}) amenable to NMR relaxation experiments. Shown below the graphic are the approximate ranges of k_{ex} applicable to the NMR experiments used in this work. The exchange rates measured for protein-protein interactions and conformational exchange phenomena described herein are placed accordingly, for illustrative purposes. The helix-coil exchange of the UIMs of RAP80 is placed very approximately, quantification not being possible due to a rapid k_{ex} (> 10000 s^{-1}). Figure adapted from figure 1 in (110).

(figure 1.16).

ZZ-exchange spectroscopy

The principle behind ZZ-exchange is simple and elegant (112). If, for a two-state exchange process, magnetization transfer is slow with respect to the measurement of the chemical shift of the indirectly detected ^{15}N dimension of a 2D ^{1}H-^{15}N HSQC, then two distinct peaks will be observed, reflecting the two states (figure 1.17, top left spectrum). If the HSQC experiment is modified slightly to incorporate a delay between the indirect chemical shift measurement and the ^{1}H detection, some of the nuclei will exchange from the free to the bound state, and vice-versa during this delay. Since this transfer occurs after the ^{15}N chemical shift measurement, these spins will have one chemical shift measured from the first state, and then the directly measured chemical shift from the other. This results in the cross-peaks observed in the other panels of figure 1.17. By varying the mixing time the resultant peak intensities of all four peaks can be fit simultaneously to the Bloch equations modified for a two-state exchange process, which describe the magnetization transfer as a function of the kinetic parameters, k_{on} and k_{off} (figure 1.18)

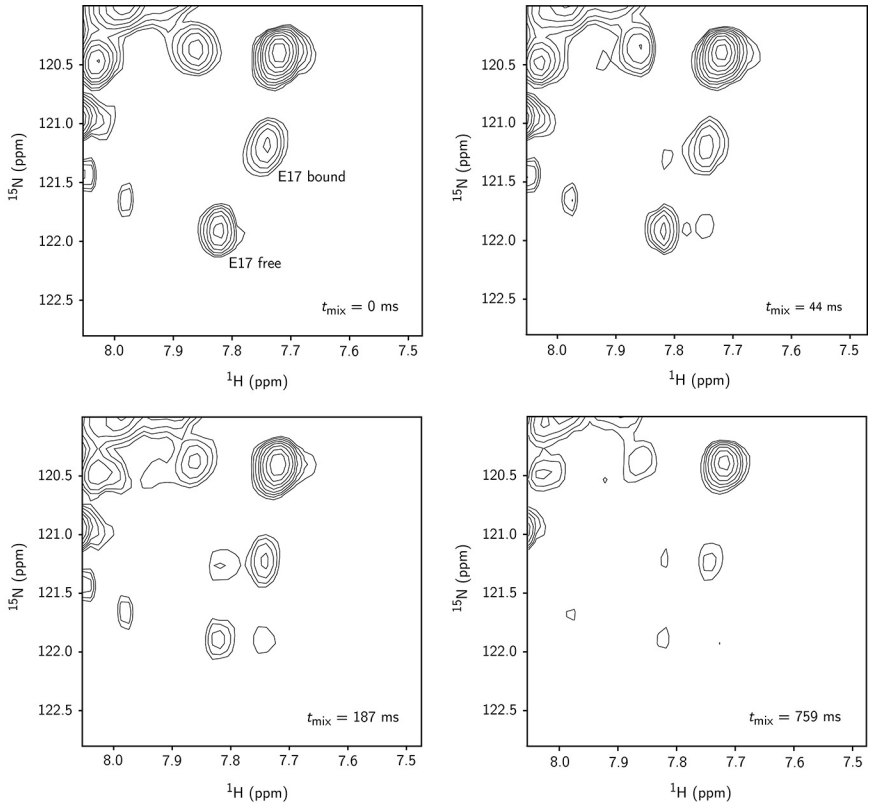

Figure 1.17: Magnetization transfer between the free and bound states of E17 of ^{15}N-labelled Mms2 in a 2:1 ratio with unlabelled Ubc13 as a function of the mixing time in the ZZ-exchange experiment (detailed in chapter 4 (26)). At $t = 0$, only the free and bound peaks are visible; when $t > 0$, the magnetization transfer manifests as cross peaks.

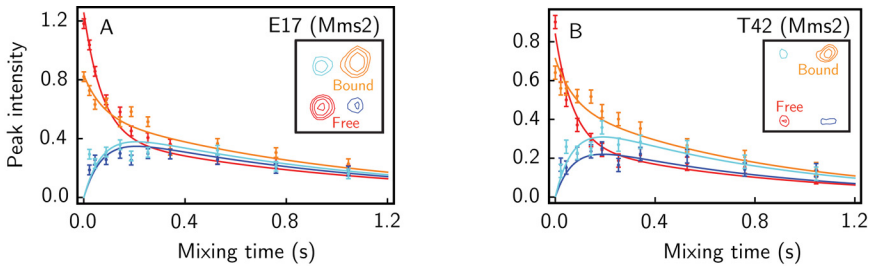

Figure 1.18: Fits of experimental auto and cross-peak intensities as a function of mixing time for E17 and T42 from ^{15}N-labelled Mms2 in a 2:1 mixture with unlabelled Ubc13. By varying t and fitting the time dependence of the intensities of all four peaks simultaneously the kinetics of transfer (k_{on} and k_{off}) can be quantified (chapter 4 (26)).

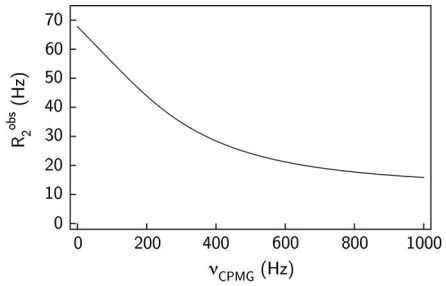

Figure 1.19: Simulated dispersion curve for $k_{ex} = 2000\ s^{-1}$, $\Delta\delta = 180$ Hz, and p_A, p_B both 0.5, and $R_2^0 = 12\ s^{-1}$ for both the free and bound state. As ν_{CPMG} approaches ∞ then $R_{2,obs}$ approaches R_2^0 and the effects of exchange are effectively suppressed. Fitting of experimental dispersion curves to the Carver-Richards equation (108, 114–116) allows determination of k_{ex} and other parameters.

(111).

CPMG relaxation dispersion

CPMG relaxation dispersion experiments differ from ZZ-exchange in that although there is a similar delay in which the nuclei are allowed to exchange between states, the R_2 relaxation rate constant is measured as a function of an increasing frequency of 180° ^{15}N pulses applied during this delay (113). The magnetization is allowed to relax in the transverse xy plane during this pulse train; consequently the measurement is of R_2. The underlying principle of relaxation dispersion is that, for nuclei undergoing exchange on the µs to ms timescale this overall R_2 (hereafter denoted as $R_{2,obs}$, the R_2 actually observed, for clarity) measured is the sum of the intrinsic R_2^0 values of both states scaled by their fractional populations and a contribution from the exchange process itself R_2^{ex}.

By applying the 180° pulses, the dephasing of the individual spins is retarded. To illustrate this, we consider the following simplified situation, depicted in figure 1.20, where the relaxation delay begins with two nuclei in state A, precessing at the same frequency (ω_A). After a period τ, one of the nuclei exchanges in to state B and begins to precess at ω_B. The difference $\Delta\omega$ between the precession frequencies of the two nuclei now results in their dephasing. As τ increases, the more dephasing will result. If, however, a 180° pulse is applied at this point on the x or y axes, the magnetization will flip around this axis and the $\Delta\omega$ will result in rephased spins after a further τ (117). In reality, the situation is clearly more complicated as there are more than two nuclei occupying states at various

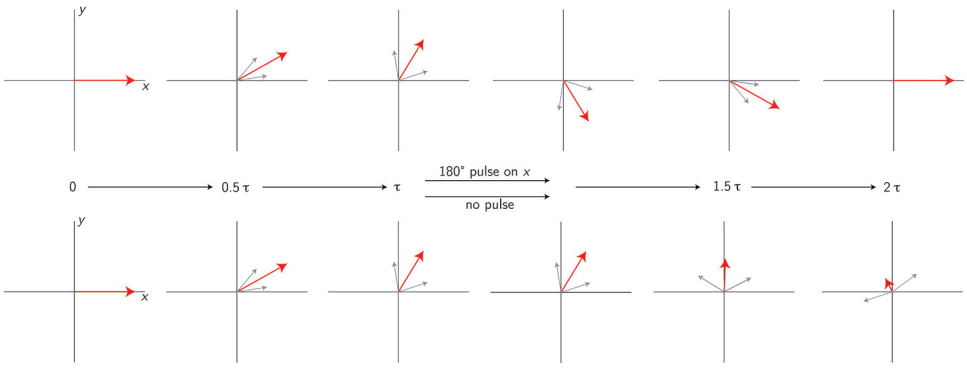

Figure 1.20: Effect of 180° pulses on the net magnetization vector in the presence of two-state chemical exchange (adapted from figure 3 in ref. (117)). At $t = 0$, the two spins are aligned on the x axis in the xy plane, both in state A. At this point, one spin undergoes an exchange into state B, the precession frequency of which is 90 Hz higher than that of A. After a period $t = \tau/2 = 2.5$ ms, both spins have precessed differently, and the magnitude of the net magnetization vector has decreased accordingly. After $t = \tau$, it has decreased still further. Following this delay the upper and lower sequences differ. In the upper sequence, a 180° pulse is applied at $t = \tau$ on x, whereas in the lower no pulse is applied. After $t = 1.5\tau$, the upper magnetization vector has increased as the precession difference between the two spins now rephases them; in the lower sequence, the overall magnetization continues to drop. After the full $t = 2\tau$, the net magnetization vector has fully recovered, whereas in the lower it has decreased still further in the absence of the pulse.

life-times. Essentially however, as the pulsing frequency increases, the overall $R_{2,\mathrm{obs}}$ decreases, and the effect of the exchange process is increasingly suppressed. Although there is an experimental upper limit to this refocusing frequency ($\nu_{\mathrm{CPMG}} \approx 1000$ Hz) (113), as ν_{CPMG} approaches ∞ then $R_{2,\mathrm{obs}}$ approaches R_2^0. Figure 1.19 illustrates the dispersion curve that would be observed for the nuclei undergoing the exchange process simulated in figure 1.14.

By measuring the decrease in $R_{2,\mathrm{obs}}$ as a function of ν_{CPMG}, kinetic parameters can be obtained by fitting to the Carver-Richards equation (108, 114–116):

$$R_2(1/\tau_{\mathrm{cp}}) = \frac{1}{2}(R_{2\mathrm{A}}^0 + R_{2\mathrm{B}}^0 + k_{\mathrm{ex}} - \frac{1}{2\tau_{\mathrm{cp}}}\cosh^{-1}[D_+\cosh(\eta_+) - D_-\cosh(\eta_+)]) \quad (1.15)$$

where $2\tau_{\mathrm{cp}}$ is the delay between refocusing pulses (so that $\nu_{\mathrm{CPMG}} = 1/4\tau_{\mathrm{cp}}$) (115), and:

$$D_\pm = \frac{1}{2}[\pm 1 + \frac{\Psi + 2\Delta\omega^2}{(\Psi^2 + \zeta^2)^{1/2}}] \quad (1.16)$$

$$\eta_{\pm} = \sqrt{2}\tau_{\mathrm{cp}}[\pm\Psi + (\Psi^2 + \zeta^2)^{1/2}]^{1/2} \tag{1.17}$$

$$\Psi = (R^0_{2\mathrm{A}} - R^0_{2\mathrm{B}} - p_{\mathrm{A}}k_{\mathrm{ex}} + p_{\mathrm{B}}k_{\mathrm{ex}})^2 - \Delta\omega^2 + 4p_{\mathrm{A}}p_{\mathrm{B}}k^2_{\mathrm{ex}} \tag{1.18}$$

$$\zeta = 2\Delta\omega(R^0_{2\mathrm{A}} - R^0_{2\mathrm{B}} - p_{\mathrm{A}}k_{\mathrm{ex}} + p_{\mathrm{B}}k_{\mathrm{ex}}) \tag{1.19}$$

where $\Delta\omega$ is the chemical shift difference between sites, $R^0_{2\mathrm{A}}$ and $R^0_{2\mathrm{B}}$ are the intrinsic transverse relaxation rates experienced by the nuclei in sites A and B, respectively, and p_{A} and p_{B} are the fractional populations ($p_{\mathrm{A}} + p_{\mathrm{B}} = 1$) of states A and B. k_{ex} is the overall exchange rate introduced at the beginning of this section. It should be noted that the coefficients multiplied by τ_{cp} differ, as well as the relationship between ν_{CPMG} and τ_{cp}, depending on how this variable is defined in the pulse sequence, whether equal to the time between successive 180° pulses (108), or to half this time (115).

Plotting of the above equation is somewhat more useful for illustrating the dependence of $R_{2,\mathrm{obs}}$ on the parameters $\Delta\delta$, $R^0_{2\mathrm{A}}$, $R^0_{2\mathrm{B}}$, p_{A}, p_{B}, and k_{ex} on the measured $R_{2,\mathrm{obs}}$ values. Figure 1.21 gives an illustration of their effects on the observed value of $R_{2,\mathrm{obs}}$ and their effect on the NMR line-shapes. These curves are also useful for obtaining a qualitative understanding of the effect of these parameters by comparing the values of $R_{2,\mathrm{obs}}$ at $\nu_{\mathrm{CPMG}} = 0$. From a practical perspective, when considering the typical struggles with finite protein solubility, and limited signal-to-noise, it is often desirable to choose an approximate value of $R_{2,\mathrm{obs}}$ that produces a dispersion curve that is not too flat (especially important when $k_{\mathrm{ex}} \gg \Delta\omega$), but low enough that the signal is not too broad relative to the baseline noise. As we typically possess knowledge of the K_{D} when performing these experiments, this is done by choosing values of p_{a} and p_{b} and using the relationship $R^{\mathrm{ex}}_2 = p_{\mathrm{A}}p_{\mathrm{B}}\Delta\omega^2/k_{\mathrm{ex}}$ (108).

Line-shape analysis

Perhaps the most intuitive method of quantifying chemical exchange kinetics by NMR spectroscopy is that of line-shape analysis. In the introductory paragraphs above, it was suggested that one of the advantages of using NMR to uncover exchange processes was that often an initial very rough assessment of the time scale of exchange can be made by careful visual inspection of the spectra. Line-shape analysis can be thought of

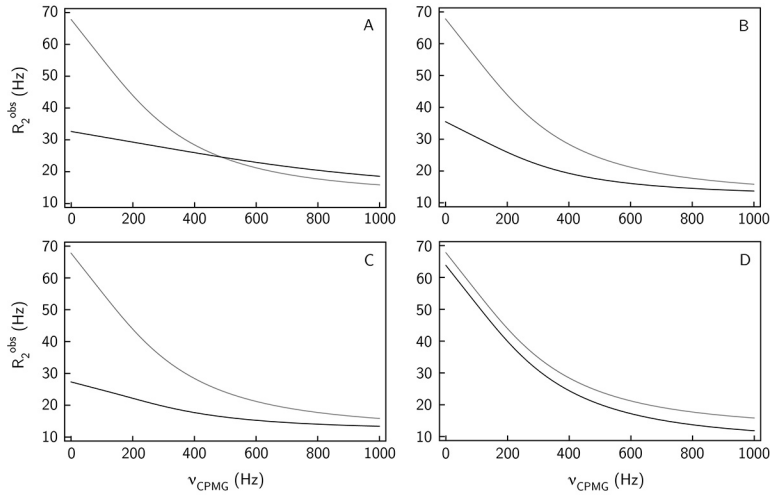

Figure 1.21: The parameter dependence of the Carver-Richards equation. In all panels, the grey curve is that of figure 1.19. (A) Changing k_{ex} from 2000 s^{-1} (grey) to 5000 s^{-1} (black). (B) Changing $\Delta\omega$ from 180 Hz (gray) to 120 Hz (black). (C) Changing p_A from 0.5 (grey) to 0.9 (black). (D) Changing R_2^0 from 12 Hz (grey) to 8 Hz (black)

as the natural development of this property: in essence, it attempts to move from the qualitative to the quantitative. This is no criticism of the preceding methods; CPMG relaxation dispersion, for example, is currently enjoying a renaissance after its origins in the 1950s (117, 118), and has been shown to be especially useful for characterizing "invisible states" of proteins (119, 120). But these methods require specialized NMR experiments. Line-shape analysis can be performed on a series of simple ^{1}H-^{15}N HSQC or even one-dimensional proton spectra, offering the appealing potential for obtaining quantitive measurements of thermodynamics and kinetics from a single chemical shift titration.

As discussed above, there are two manifestations of an exchange process on the observed line-shape. One is the chemical shift: at relatively slow exchange rates $k_{ex} \ll \Delta\omega$, two distinct peaks with chemical shifts will be observed, each corresponding to the nuclei in either state. These converge to a single averaged peak when $\Delta\omega \approx k_{ex}$ (figure 1.15). The second, illustrated in figure 1.14, is an increased line-width for these peaks, due to the exchange contribution to R_2.

Line-shape analysis is typically performed by constructing analytical functions for the line-shapes in the frequency domain, starting from the Bloch-McConnell equations

describing the effect of a two-state exchange process on the NMR signal in the time domain (108):

$$M_{\mathrm{A}}(t) = M_{\mathrm{A}}(0)a_{11}(t) + M_{\mathrm{B}}(0)a_{12}(t) \tag{1.20}$$

$$M_{\mathrm{B}}(t) = M_{\mathrm{B}}(0)a_{22}(t) + M_{\mathrm{A}}(0)a_{21}(t) \tag{1.21}$$

for spins A and B, with coefficients:

$$\begin{aligned} a_{11}(t) = \frac{1}{2}\Bigg[&\left(1 - \frac{-i\Delta\omega + R^0_{2A} - R^0_{2B} + k_{\mathrm{ex}}(p_{\mathrm{B}} - p_{\mathrm{A}})}{\lambda_+ - \lambda_-}\right)\exp\left(-\lambda_- t\right) \\ &+ \left(1 + \frac{-i\Delta\omega + R^0_{2A} - R^0_{2B} + k_{\mathrm{ex}}(p_{\mathrm{B}} - p_{\mathrm{A}})}{\lambda_+ - \lambda_-}\right)\exp\left(-\lambda_- t\right)\Bigg] \end{aligned} \tag{1.22}$$

$$\begin{aligned} a_{22}(t) = \frac{1}{2}\Bigg[&\left(1 + \frac{-i\Delta\omega + R^0_{2A} - R^0_{2B} + k_{\mathrm{ex}}(p_{\mathrm{B}} - p_{\mathrm{A}})}{\lambda_+ - \lambda_-}\right)\exp\left(-\lambda_- t\right) \\ &+ \left(1 - \frac{-i\Delta\omega + R^0_{2A} - R^0_{2B} + k_{\mathrm{ex}}(p_{\mathrm{B}} - p_{\mathrm{A}})}{\lambda_+ - \lambda_-}\right)\exp\left(-\lambda_- t\right)\Bigg] \end{aligned} \tag{1.23}$$

$$a_{12}(t) = \frac{k_{\mathrm{ex}}p_{\mathrm{A}}}{\lambda_+ - \lambda_-}[\exp\left(-\lambda_- t\right) - \exp\left(-\lambda_+ t\right)] \tag{1.24}$$

$$a_{21}(t) = \frac{k_{\mathrm{ex}}p_{\mathrm{B}}}{\lambda_+ - \lambda_-}[\exp\left(-\lambda_- t\right) - \exp\left(-\lambda_+ t\right)] \tag{1.25}$$

where:

$$\begin{aligned} \lambda_\pm = \frac{1}{2}\Bigg\{&-i\Omega_A - i\Omega_B + R^0_{2A} + R^0_{2B} + k_{\mathrm{ex}} \\ &\pm \sqrt{(-i\Delta\omega + R^0_{2A} - R^0_{2B} + k_{\mathrm{ex}}(p_B - p_A))^2 + 4p_a p_b k^2_{\mathrm{ex}}}\Bigg\} \end{aligned} \tag{1.26}$$

These relate the FID to the parameters $R^0_{2\mathrm{A}}$ and $R^0_{2\mathrm{B}}$, which are the intrinsic transverse relaxation constants for states A and B (*e.g.* in the absence of exchange), p_{A} and p_{B}, the fractional populations of states A and B, Ω_{A} and Ω_{B}, the chemical shifts of the two states, as well as the difference between them, $\Delta\omega$, and the kinetics of the exchange process itself, k_{ex}.

To simulate the NMR line-shape, these equations have to be Fourier transformed, so that the effect of the above parameters, especially $\Delta\omega$ and k_{ex} can be ascertained directly. Complicating matters is the usual post-processing that takes place when analyzing NMR

data, such as multiplication by a window function. As this has a direct effect on peak line-width it must be taken into account, otherwise erroneous kinetic parameters may be obtained. To be of any use in quantifying exchange processes, it is necessary that the parameters listed above are variables so that different values can be tested against the experimental data, to determine those that best reproduce it. This requires that the Fourier transform be done symbolically, rather than numerically. Much like the simulations of multivalent binding of Ub to Vps27 discussed in the preceding sections, this calculation results in a ponderous, inelegant expression which, while not ideal from an intuitive standpoint, can nevertheless be used by a computer program to reproduce and fit line-shapes to experimental data.

Our use of line-shape analysis has focused on obtaining kinetic measurements for protein-protein interactions on the intermediate to fast timescales ($k_{ex} > \Delta\omega$) using 2D ^{1}H-^{15}N HSQC titrations where the concentrations of one or both components are changed in stepwise fashion, as described above. The first advantage garnered by this methodology is that K_D is known from the titration itself, by fitting the chemical shift perturbations to equation 1.14. Therefore, typically we reformulate the k_{ex} parameter in eqns 1.20 to 1.26 in terms of k_{off}, using the relationships $k_{ex} = k_{on}[L] + k_{off}$ and $K_D = k_{off}/k_{on}$. This is advantageous as it eliminates the need to fit k_{on}. The protein concentrations at every point being known, [L] can be calculated using:

$$[L] = \frac{1}{2}\left(\sqrt{(-K_D - [L]_0 - [P]_0)^2 - 4[L]_0[P]_0} - K_D + [L]_0 - [P]_0\right) \tag{1.27}$$

the derivation of which is described in Appendix A. Similarly, p_A and p_B are known from the K_D. For protein-protein and protein-ligand equilibria, R^0_{2A} can be obtained from measurements of the labelled component without binding partner using relaxation experiments (121). Regarding R^0_{2B}, for small ligand binding, $R^0_{2B} = R^0_{2A}$, as the intrinsic R^0_2 is largely dependent upon molecular weight; for protein-protein interactions, where the mass of the complex is likely much larger than that of the free proteins, R^0_{2B} can be estimated by multiplying R^0_{2B} by a scaling factor, or, if the saturated complex ($p_B \approx 1$) can be achieved under NMR sample conditions, can be measured as well. Regarding intramolecular processes, sample conditions cannot be as easily tuned so that $p_A =$

1 nor $p_B = 1$ can be achieved, but specialized experiments designed to eliminate the contribution from chemical exchange can be used (122). Additionally, if these direct methods prove unsuccessful, an estimate of the intrinsic R_2^0 can often be obtained from the line-widths of resonances not undergoing exchange. If no independent measurement is available, R_{2B}^0 can also be left as a free parameter; however, it becomes strongly coupled with $\Delta\omega$ so that the two are difficult to fit simultaneously. To alleviate this problem somewhat, simultaneous fitting of multiple line-shapes to different $\Delta\omega$ values and a global ΔR_2 can be used (123). The final parameters needed are those describing the chemical shifts, Ω_A and Ω_B. As with the intrinsic relaxation rates for bimolecular interactions, Ω_A is obtained from a spectrum in the absence of binding partner. Ω_B and $\Delta\omega$ are then fit alongside k_{off}.

It should also be noted that additional parameters are often introduced in order to compensate for differential relaxation losses during the NMR experiment itself, especially during the INEPT transfer periods, which have to be considered when the molecular weight of the bound complex is much higher than that of the free, because as p_B increases so will these losses (124). Although these losses affect neither the line-widths nor the chemical shifts, their effect is to reduce the overall M_0, so that global fitting of this parameter over all the line-shapes from a given titration is no longer appropriate.

In chapter 6, we demonstrate the effective range of k_{ex} over which line-shape analysis is both accurate and precise, for 1:1 protein ligand interactions in the intermediate to fast exchange regime. The ease of use and applicability of the two methods for obtaining more precise K_D values described above, and in chapter 5, led us to the idea that it might be possible to obtain similarly well-determined kinetic measurements from these same titrations. To assess this, we used a combination of line-shape simulations and experimental data. Simulations are a powerful method for assessing the applicability of methods such as these, as they allow one to set all pertinent parameters and then assess how well these parameters can be reproduced using the line-shape analysis. To facilitate this, we performed quantum-mechanical simulations using the program GAMMA (125).

As described above, depending on the relative values of k_{ex} and $\Delta\omega$ the line-shapes can be very different. As k_{ex} increases $\gg \Delta\omega$ the amount of additional broadening, R_{ex}, decreases. There is therefore an upper limit on the k_{ex} that can be determined using line-

shape analysis (figure 1.16). Simulations indicate that reasonable accuracy and precision can be obtained to $k_{ex} = 15000\ s^{-1}$.

Similarly, at intermediate exchange values ($k_{ex} \approx \Delta\omega$), line-shapes can deviate from the Bloch-McConnell equations. This not only somewhat compromises accurate determination of kinetics but also thermodynamic measurements using chemical shift titrations as outlined above. Deviations in peak shape in this exchange regime can introduce errors into the chemical shift measurement and, therefore, in the K_D measured from these shifts. As we use the predetermined K_D to eliminate the need to fit both k_{on} and k_{off} (*v.s.*), any systematic error in K_D will effect the accuracy of the fitted k_{off}. The titration method whereby a mixture of protein and ligand is serially diluted (Method 2 above, figure 1.13) is shown to offer a distinct advantage when obtaining K_D values. This can be rationalized by the observation that this method, unlike a traditional titration, avoids the broadest peaks. Figure 1.22 (here reproduced from chapter 6) demonstrates this, comparing the accuracy and precision of fitted k_{off} for Methods 1, 2, and a traditional titration. It is immediately clear that, on average, the peaks are sharper for Method 2, especially the second and third titration points. This somewhat ameliorates the systematic error inherent in the intermediate exchange regime. We also present a simple method of potential use in determining more accurate K_D values in the intermediate exchange regime, by taking advantage of the fact that many residues typically shift in chemical shift titrations; by comparing the fitted K_D of each of these residues with their $\Delta\omega$ values, any systematic error K_D due to increasing $\Delta\omega$ (and increasing line-broadening) can be identified.

T_1, T_2, and NOE measurements

Unlike the above measurements, which are designed to quantify processes on the µs to ms timescales, T_1, T_2, and NOE measurements reflect molecular motions on the ps to ns timescale. Although beyond the scope of this section, methods of analysis such as model-free (126, 127) and iRED in combination with molecular dynamics simulations (128) are available to extract information on the relative flexibility of proteins on a per-residue basis. With respect to RAP80-tUIM, we utilized these experiments to demonstrate that the two helices are flexible with respect to one another and to show that the linker region

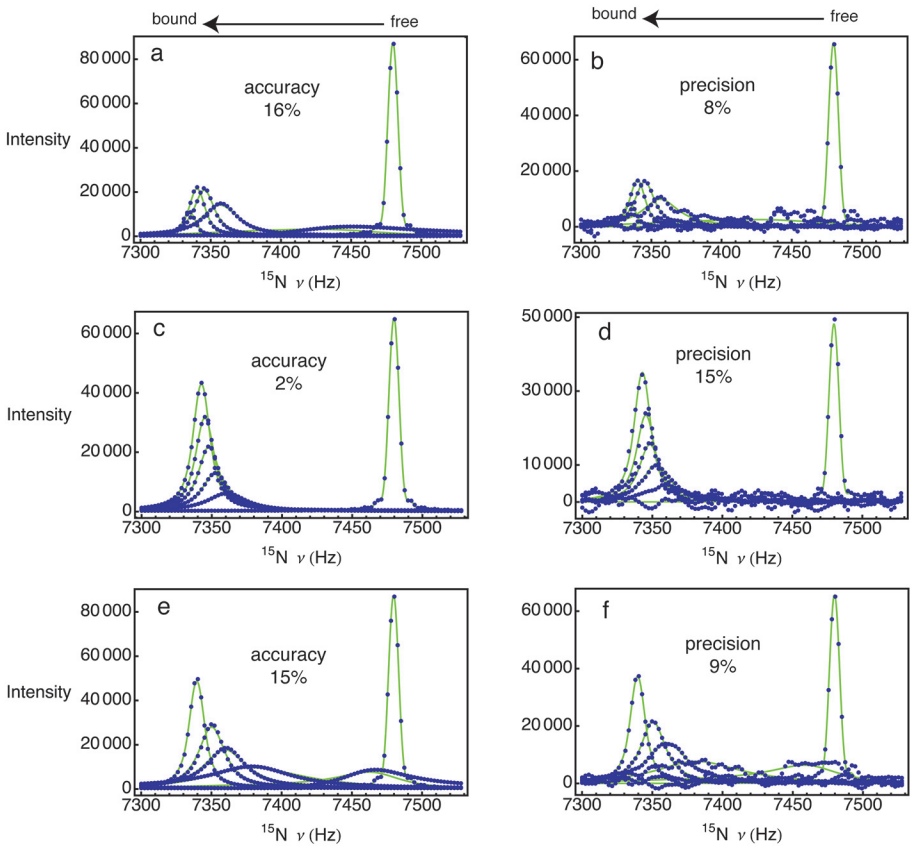

Figure 1.22: Accuracy and precision of k_{off} obtained from line-shape analysis using Methods 1, 2, and a traditional titration with $[P_0]$ fixed at 0.5 mM. In these cases the ^{15}N dimension is simulated, with $K_{\text{D}} = 60$ µM and $k_{\text{off}} = 500$ s^{-1}. Panels on the left are the GAMMA simulations in the absence of noise, to allow determination of the accuracy of the line-shape analysis. The right panels are the spectra in the presence of noise to assess its effect on the precision with which k_{off} can be determined under typical experimental conditions. Line-shapes for method 1 (panels A and B) and the traditional titration (panels E and F) suffer from increased broadening, which worsens the accuracy of k_{off} determined using these methods, in this exchange regime. By avoiding much of this broadening, method 2 is demonstrated to be more amenable to line-shape analysis in exchange regimes with a great deal of broadening.

between the UIMs possesses residual flexibility in the bound state.

Kinetics of ubiquitination

To understand the formation of the polyubiquitin chains so crucial for intracellular signalling, we have used NMR relaxation methods and fluorescence based biochemical enzyme assays. The combination of these two techniques, used together with the power of numerical computational methods, has provided us with a methodological strategy for elucidating the mechanism of the complex enzyme systems responsible for the ubiquitin cascade.

But why the adjective *complex*? What differentiates E2 enzymes from other enzymes? The canonical model for understanding enzymatic reactions is Michaelis-Menten kinetics (129, 130), described by:

$$\mathrm{E} + \mathrm{S} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_{\mathrm{cat}}} \mathrm{P} \tag{1.28}$$

where E and S denote the enzyme and its substrate, coming together non-covalently to form an enzyme-substrate, *ES*, or Michaelis complex; subsequently, at a rate defined by $d[\mathrm{P}]/dt = k_{\mathrm{cat}}[\mathrm{ES}]$, where k_{cat} is the catalytic rate constant, the enzyme catalyzes the formation of product.

Ubiquitin-conjugating enzymes, however, are not as amenable to Michaelis-Menten analysis as others. The primary reason is that, although still catalytic, they act as substrates in their own reactions. Formation of polyUb chains occurs when the side-chain amine of a lysine residue from the growing chain cleaves the thioester bond between the C-terminus of the Ub molecule to be added and the catalytic cysteine of the E2. The polyUb chain then dissociates, but the enzyme is left inert. To undergo a subsequent round, a new Ub must be transferred from an E1 to the catalytic cysteine of the E2. Returning to equation 1.28, the following addition is necessary, now substituting the components of the polyUb reaction under consideration for clarity:

$$\mathrm{Ubc}_{13} \xrightarrow{\mathrm{E1}} \mathrm{Ubc}_{13} \sim \mathrm{Ub} + \mathrm{Ub}_{\mathrm{n}} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{Ubc}_{13} \sim \mathrm{Ub}{-}\mathrm{Ub}_{\mathrm{n}} \xrightarrow{k_{\mathrm{cat}}} \mathrm{Ub}_{\mathrm{n+1}} + \mathrm{Ubc}_{13} \tag{1.29}$$

Further complicating the application of Michaelis-Menten kinetics to the ubiquitination

cascade is that even once charged, E2 enzymes do not typically work in isolation. To synthesize K63-linked polyUb, Ubc13 functions in concert with the E2-like, though inert on account of lacking the catalytic cysteine residue, Mms2. In essence, Mms2 fulfils the substrate binding role for Ubc13 inherently possessed by a more typical enzyme. So not only does an analysis of this enzyme require the presence of an additional reaction, even when in the charged (catalytically active) E $\sim$ Ub state, it does not obey equation 1.28. The reaction schematic is now as follows, with the heterodimeric interaction between Ubc13 and Mms2 a necessary preliminary:

$$\begin{aligned} &\mathrm{Ubc_{13}} \xrightarrow{\mathrm{E1}} \mathrm{Ubc_{13} \sim Ub} + \mathrm{Mms_2} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{Ubc_{13} \sim Ub{-}Mms_2} \\ &\mathrm{Ubc_{13} \sim Ub{-}Mms_2} + \mathrm{Ub_n} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{Ubc_{13} \sim Ub{-}Mms_2{-}Ub_n} \xrightarrow{k_{\mathrm{cat}}} \mathrm{Ub_{n+1}} + \mathrm{Ubc_{13}} \end{aligned} \tag{1.30}$$

In reality, this reaction does not proceed in the stepwise manner implied by equation 1.30, but rather E1 and Ubc13, Ubc13 and Mms2, and Mms2 and Ub are undergoing constant binding and dissociation. Also, as Ubc13 is used up, the resultant uncharged form is able to sequester Mms2. A more illustrative schematic of these numerous processes is given in figure 1.23. From the perspective of Michaelis-Menten kinetics, charged Ubc13 pre-bound to Mms2 binding to substrate to form the ternary Ubc13$\sim$Ub-Mms2-$\mathrm{Ub_n}$ complex is analogous to equation 1.28, but the four additional contributing equilibria are ignored.

In order to elucidate the mechanism of Ubc13, we need to be able to accurately measure the rate of the catalytic step, or $d[\mathrm{Ub_2}]/dt$, the object being to determine the first-order rate constant, k_{cat}. The challenge is that the ancillary equilibria can potentially affect this measurement. Returning to the comparison with Michaelis-Menten kinetics, in the case of $\mathrm{Ub_2}$ formation, rather than the usual complex of enzyme and substrate, the state *ES* is a ternary composition of Ubc13 (the enzyme), Ub (the substrate), as well as Mms2 (equation 1.30). To understand the potential complication, it is useful to consider two extreme cases; first, if the affinity of Ubc13 and Mms2 is so strong that the heterodimer effectively behaves as a monomeric protein (a K_{D} in the nM range), and second, if the affinity is relatively weak (in the µM range). Figure 1.24A illustrates the time-dependent build-up of $\mathrm{Ub_2}$ that would be observed experimentally in these cases, maintaining an identical k_{cat} constant at 0.007 $\mathrm{s^{-1}}$ in both simulations, and

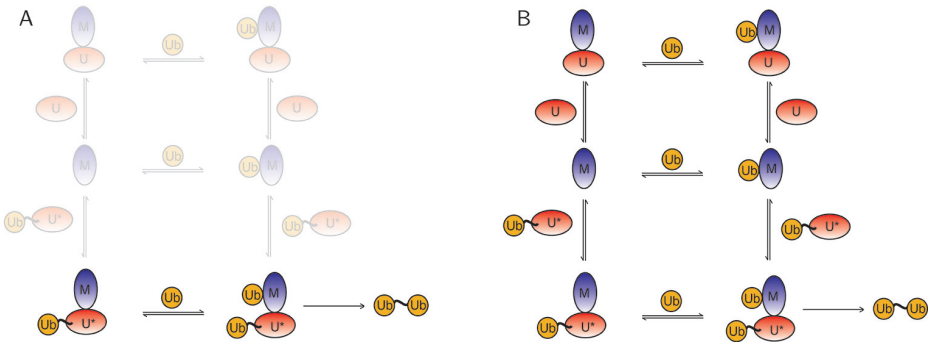

Figure 1.23: Considering the formation of Ub_2 via the charged (thioester bound Ub) Ubc13-Mms2 heterodimer in terms of A) Michaelis-Menten kinetics B) the full equilibrium between charged and uncharged Ubc13 (U* and U), Mms2 (M), and Ub.

a free ubiquitin concentration (substrate) of 100 µM. For these simulations, as for our experimental methodology, Ubc13 is pre-incubated with E1, ubiquitin, and ATP to provide activated enzyme before the reaction is started. Before starting Ub_2 formation, the E1 activity is stopped. Returning to figure 1.24, it is apparent that the overall buildup of Ub_2 is quite different between the two examples. Similar changes in the build-up of Ub_2 are observed for different affinities between Ub and the heterodimer, though variations in this equilibrium are taken in to account when using Michaelis-Menten kinetics, via the equilibrium between $E + S$ and ES complex. These time-dependent curves being what are measured experimentally, without knowledge of the underlying kinetics it is therefore evident that the value of these experiments for determination of k_{cat} is wholly compromised, as discussed in further detail below.

This juncture provides a convenient opportunity for describing the methodology underlying measurement of k_{cat} values, so we will apply it to the simulated build-up curves given in figure 1.24. To begin, we consider only the subset to which Michaelis-Menten can be applied (figure 1.23A). Traditionally, measurement using this model does not involve acquisition of time-course data, instead a plot of initial rates as a function of substrate concentration is obtained. This can be explained as follows. Rate-equations are typically straightforward to construct, even for kinetic schemes as complicated as those describing the synthesis of polyUb chains (*cf.* chapter 4, eqs. 4.22-4.31); solving them to yield expressions for the time-dependence of substrate decrease or product generation is typically much more difficult. During the early to mid twentieth century, the

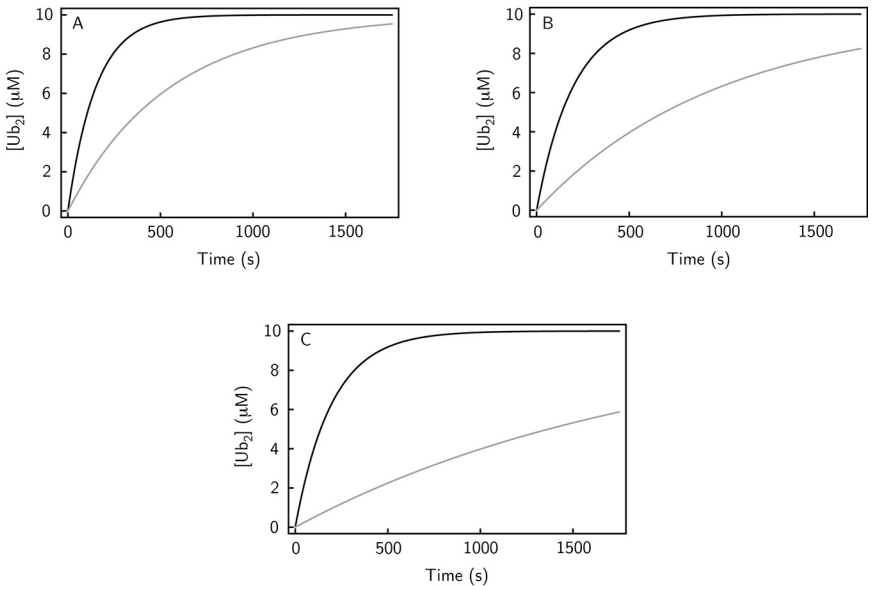

Figure 1.24: Measurement of k_{cat} necessitates knowledge of the underlying kinetics. A. Effect of differences in the binding between Ub and the Ubc13-Mms2 heterodimer. The off-rates used to simulate the Ub_2 buildup curves were 5 s^{-1} (black curve) to 5000 s^{-1} (grey curve). B. Effect of differences in the binding between Ubc13 and Mms2 on the rate of Ub_2 formation. The off-rates for the heterodimer interaction were the experimentally measured value of 4.4 s^{-1} (26) to 4000 s^{-1}. C. Effect of changes in k_{cat} on the rate of Ub_2 formation. The k_{cat} values used were 0.007 s^{-1} and 0.0007 s^{-1}.

classical period of enzymology, it was necessary to make certain simplifying assumptions to make this a more tractable task. For the Michaelis-Menten scheme this is most often achieved by making the steady-state approximation (131), in which the rate of change of the *ES* complex, $d[\mathrm{ES}]/dt = k_{\mathrm{on}}[\mathrm{E}][\mathrm{S}] - k_{\mathrm{off}}[\mathrm{ES}] - k_{\mathrm{cat}}[\mathrm{ES}]$, is taken to be 0. Subsequently solving this equation for the instantaneous concentration of [ES], and substituting the total enzyme concentration $[\mathrm{E}]_0 = [\mathrm{E}] + [\mathrm{ES}]$, gives:

$$[\mathrm{ES}] = \frac{[\mathrm{E}]_0 k_{\mathrm{on}}[\mathrm{S}]}{k_{\mathrm{cat}} + k_{\mathrm{off}} + k_{\mathrm{on}}[\mathrm{S}]} \tag{1.31}$$

Eliminating k_{on} from the numerator and denominator combines the remaining kinetic parameters of the denominator into the well-known Michaelis constant, $K_{\mathrm{m}} = (k_{\mathrm{cat}} + k_{\mathrm{off}})/k_{\mathrm{on}}$. Substitution of this expression into the rate of product build-up yields the familiar expression, relating initial velocity, ν, to initial substrate concentration, [S]:

$$\nu = \frac{V_{\mathrm{max}}[\mathrm{S}]}{K_{\mathrm{m}} + [\mathrm{S}]} \tag{1.32}$$

where V_{max}, the maximum initial rate (at infinite substrate concentration), is equal to the total enzyme concentration multiplied by k_{cat}. It should be noted also that this assumes that $[\mathrm{S}] \gg [\mathrm{E}]_0$, so that [S] does not decrease appreciably upon formation of [ES].

The modern ability to solve differential equations numerically makes simplifying assumptions such as these unnecessary, allowing determination of kinetic parameters directly from the time-course data. This is achieved by numerically integrating the ordinary differential equations describing the Michaelis-Menten kinetic model, rather than the steady-state equation 1.32, above. Furthemore, these assumptions are not only dispensable, but undesirable as well, being a source of error (132, 133). Returning to the curves in figure 1.24, simulated with $k_{\mathrm{cat}} = 0.007\ \mathrm{s}^{-1}$, if we consider them from a Michaelis-Menten perspective it is immediately apparent there is a problem, as the curves saturate at a Ub_2 concentration equal to the starting concentration of charged Ubc13, rather than the starting amount of substrate Ub. This immediately invalidates the integrated full Michaelis-Menten equation (even though there is no longer the requirement

Table 1.2: Backfits of simulated Ub_2 build-up curves to the integrated Michaelis-Menten equation, with the binding kinetics of Ub to the heterodimer known. Curves are shown in figure 1.24, and were sampled at time intervals of 175 s, from 0 to 1750 s in order to fit apparent k_{cat} values using the integrated Michaelis-Menten equation.

$k_{off,Ub\text{-}M}$ (s^{-1})	$k_{off,Ub\text{-}M:U}$ (s^{-1})	$k_{off,M\text{-}U}$ (s^{-1})	$k_{cat,real}$ (s^{-1}) [a]	$k_{cat,fit}$ (s^{-1}) [b]
560	2250	4.4	0.007	0.0065
560	2250	4000	0.007	0.0013
5	2250	4.4	0.007	0.0064
5000	2250	4.4	0.007	0.0066
560	2250	4.4	0.0007	0.00066

a k_{cat} used to simulate data using the full kinetic model for Ub_2 formation
b k_{cat} determined by fitting to the integrated Michaelis-Menten equation

that $d[ES]/dt \sim 0$), as it will try to satisfy the condition that $\lim_{t\to\infty}[P](t) = [S]_0$, P being the product in equation 1.28, $[S]_0$ being the total starting concentration of substrate. This assumes a completely catalytic process where the enzyme is regenerated as part of the reaction mechanism. In a sense then, there is no true enzyme in this system, but a somewhat counterintuitive approximation can be made in which ubiquitin is considered the enzyme for the purposes of a Michaelis-Menten analysis, since its concentration greatly exceeds that of Ubc13 in this example (100 µM versus 10 µM). From this perspective, the curves in figure 1.24A can be fit to the numerically integrated Michaelis-Menten equation, if the proper k_{on} and k_{off} values for binding between Ub and the heterodimer are used. However, in the case of figure 1.24B, where the interaction between Ubc13 and Mms2 is an order of magnitude weaker (in the µM range) even *a priori* knowledge of the Ub binding kinetics is insufficient for accurate determination of k_{cat}. Further illustrating the convolution of all these kinetic parameters is figure 1.24C, in which k_{cat} is increased, to compensate for the drop in heterodimer affinity. The true and apparent k_{cat} rates for these various curves are summarized in Table 1.2. It might be argued that Michaelis-Menten kinetics could be applied if the reactions were allowed to proceed with active E1, so that Ubc13 could be recharged with some of the substrate ubiquitin, but this has the effect of both introducing another k_{cat} into the reaction, as well as using up Ub substrate through this concurrent pathway. Even if the kinetics associated with E2 recharging were fully elucidated, the necessity of knowing the on and off-rates for Mms2 binding to Ubc13 would still be necessary, again precluding use of the Michaelis-Menten model. Comparison of the fitted k_{cat} value of 0.0065 s^{-1} with the true

value of 0.007 s^{-1} (table 1.2) would suggest that since the binding between Ubc13 and Mms2 is tight, it is perhaps not absolutely necessary to account for this equilibrium in our kinetic model. However, when considering the full complexity of the ubiquitin cascade, E2-E3 interactions are often much weaker (μM range) (11), so that full non-steady-state kinetic models will be even more necessary to analyze the effect of E3 binding on the catalytic rate. In chapter 4, the errors associated with the use of steady-state approaches to quantify E3-mediated substrate ubiquitination are discussed. In brief, as in the examples in table 1.2, weak binding between the E2 and E3 leads to apparent k_{cat} values lower than the true values. Furthermore, these errors are exacerbated with increasing k_{cat}, which current evidence suggests occurs in the presence of E3 enzymes.

Returning to figure 1.24 and table 1.2, it was assumed for the purposes of illustration that the kinetics of Ub binding to Ubc13-Mms2 were already known. More typically, analysis using Michaelis-Menten kinetics involves simultaneously elucidating both the kinetics of binding and of catalysis. This cannot, however, be achieved using a single set of time-course data, due to the interplay of the k_{on}, k_{off}, and k_{cat} parameters, much in the same way as the convolution of the kinetics of Ub binding to Mms2, Ub binding to the heterodimer, and binding between Ubc13 and Mms2 themselves complicate our analyses of Ub_2 formation.

The dangers of applying the Michaelis-Menten kinetic scheme to more complex reactions has been demonstrated, but not a more legitimate method of analysis. To accurately determine k_{cat} for Ub_2 formation by the Mms2:Ubc13 heterodimer, the underlying protein-protein interaction kinetics were measured independently using the NMR techniques already described. ZZ-exchange experiments and lineshape analysis were used to quantify the kinetics of Mms2 binding to Ubc13, lineshape analysis of NMR titrations of Mms2 with Ub having been previously performed by our laboratory (134). The on and off-rates thus determined, we constructed the system of ordinary differential equations alluded to above containing all the equilibria shown in figure 1.23 (26). The numerical differential equation solving function implemented in *Mathematica* (135) was used to solve this system to yield the time-course curve of the build-up of Ub_2 as well as of the concomitant decrease in Ubc13-thioester as a function of these kinetics as well as the k_{cat}. Since the protein-protein association kinetics were determined independently, the

experimental data was fit to this theoretical function by varying the value of k_{cat}, as detailed in chapter 4.

Quantifying the ubiquitination cascade

To introduce this section on kinetics, the convenient distinction between those amenable to measurement using NMR relaxation experiments and those for which it is necessary to develop bespoke biochemical assays was made. Falling into this latter category is the actual formation of Ub_2, characterized by the first-order catalytic rate constant, k_{cat}. To measure this, we needed some way of quantifying either the build-up of Ub_2 directly, or the decrease of charged Ubc13. Since the reaction takes place in solution with five proteins (E1, Ubc13, Mms2, Ub, and Ub_2), separation of Ub_2 is necessary. Running the reaction before isolating the components using SDS-PAGE (136) seemed ideally suited for this task, as the proteins in these gels can be quantified by use of an appropriate stain. Unfortunately, the molecular weights of Ub_2, Ubc13, and Mms2 are almost equivalent, precluding resolution between these proteins in the gel. To address this, a fluorophore was attached to the substrate ubiquitin to serve as a filter allowing quantification of labelled Ub alone by imaging the gel specifically for this fluorescence. SDS-PAGE was still necessary of course, as the substrate Ub concentration was in excess of heterodimer, so that there is always fluorescent-labelled substrate present, even after completion, but the molecular weight difference between Ub and Ub_2 is readily resolvable. The fluorophore used was AlexaFluor 488 (AF488) which is attached to a thiol reactive maleimide moiety. To covalently modify the substrate Ub, a N-terminal cysteine residue was introduced. By labelling Ub in this manner, we were also able to follow the drop in charged Ubc13 bound Ub, yielding two probes of the same process.

It would have been possible to run the reactions with active E1 throughout, but we pre-charged Ubc13 with labelled Ub, simplifying the subsequent kinetic analysis. The substrate K63 residue of AF488-Ub was mutated to an arginine, conserving charge but rendering it incompetent for nucleophilic attack on the thioester, ensuring that only Ub_2 can be synthesized, not higher order polyUb chains. To start the reaction, the E1 was stopped and Ub_2 formation was initiated via addition of Mms2 and unlabelled substrate Ub. As with the K63R mutation, stopping the E1 activity before beginning the reaction

ensures that this unlabelled substrate Ub does not recharge Ubc13 and begin subsequent rounds of chain addition.

Having deconvoluted the effects of protein-protein interaction kinetics from the catalytic rate, we can begin to gain some mechanistic insight into the reaction catalyzed by the heterodimer. As discussed above, recent evidence (31, 50) suggests that E3 ligases play a pivotal role in enhancing the catalytic rate of E2 enzymes, though the mechanism of active site rearrangement remains unclear. It therefore seems likely that our measurements of Ub_2 formation reflect the basal rate of this enzyme, in other words, in the off-state. Since Ubc13-Mms2 has the unusual ability to synthesize free polyUb chains, it is likely that the enzyme exists predominantly in this off-state, to prevent undesirable downstream signalling effects and to maintain the free pool of Ub. Nevertheless, this intrinsic basal activity is still useful for elucidating mechanistic insight into this enzyme, as its catalytic proficiency still greatly exceeds that of model substrates in water (chapter 4).

By comparing the rate of reaction to that previously determined for ammonia to model thioesters in water (137), we were able to elucidate that this enhancement is the result of a combination of distinct mechanisms. We established that Ubc13-Mms2, like other E2 enzymes such as Ubc9 (25), though not quite to the same extent, suppresses the pK_a of the donor K63 side-chain amide to increase its nucleophilicity. Using our methodology of determining k_{cat}, we acquired a pH dependent rate profile for the enzyme (26), measuring a ΔpK_a of -2.3. This does not fully account for the overall rate enhancement (k_{cat}/k_{noncat}) of 10^6 to 10^9, achieved even in this putative off-state.

Other sources of rate enhancement must therefore be present. There is undoubtedly an entropic benefit obtained from Mms2 binding Ub in such a way that the K63 lysine side-chain is positioned at the active site. As with the enhancement achieved through multivalency in the case of the RAP80 UIMs, this effect can be partially rationalized in terms of local concentration, whereby the concentration of substrate lysine is increased in this small area relative to its value in solution (138). This may be the mechanistic element responsible for the specific synthesis of K63-linked chains, though it is unlikely the main source of rate enhancement because in the available structures this residue does not appear particularly well-positioned to carry out a nucleophilic attack (20, 21),

suggesting that it still possesses a large degree of conformational freedom. Any entropic contributions are consequently minimized (138). Current thinking ascribes the source of enzymatic rate enhancement to electrostatic stabilization of the transition state (139, 140). To apply Jencks' metaphor in a slightly different context, this type of accounting (86) demonstrates that the catalytic power of E2 enzymes results from the synergistic use of various mechanisms, the exact interplay of which has yet to be fully dissected.

When discussing the ubiquitination cascade above, the putative role of RING and U-box E3 ligases in further activating E2 enzymes to obtain rate enhancements of polyUb chain synthesis was introduced (31, 50). The straightforward manner in which ODEs describing the equilibria between Ubc13, Mms2, and Ub can be used to determine k_{cat} can be applied to the interactions between Ubc13:Mms2 and E3 enzymes, facilitating quantitative measurement of these rate enhancements. As above, this will require measurement of the kinetics of E2-E3 binding, but these should be amenable to the NMR relaxation methods already used.

Summary

K63-linked polyUb chains play essential roles in modulating NF-κB activation and as mediators in the DDR. In chapter 2, the dynamics of the RING domain from TRAF6 are characterized using NMR, and their biological implications are discussed. The following chapter focuses on the recognition of polyUb chains by the tandem UIMs of the DDR protein RAP80, which have been shown as essential for proper IRIF formation in response to DNA double-strand breaks (60–62). By blending a method of analysis synthesizing thermodynamic models describing the interactions between these UIMs and polyUb chains of increasing length with NMR chemical shift titration data, we present a molecular mechanism fully accounting for the enhanced affinity achieved by harnessing multivalency in this signalling pathway. The binding kinetics as well as the intramolecular kinetics of the UIMs themselves are characterized, and their relevance to transient DNA damage signalling is hypothesized. As described above, chapter 4 seeks to understand the molecular mechanisms underlying the synthesis of K63-linked polyUb by the E2 Ubc13.

Reflecting on the practical necessity of diluting both analyte and titrant when performing the NMR ^{1}H-^{15}N HSQC chemical shift titrations of RAP80-tUIM and polyUb led to the supposition that, in addition to being vastly more convenient and efficient, these titrations may be more accurate and precise. Chapter 5 demonstrates the validity and the practicality of this methodology, through simulations of chemical shift titrations and experimental titrations of ^{15}N-labelled Mms2 with its binding partner, Ub. Sampling schemes similar to those used in chapter 3 are here considered, where isotopically labelled analyte is progressively diluted with each addition of unlabelled titrant. Furthermore, the effectiveness of an analogous scheme, the progressive dilution of a 2:1 mixture of analyte and titrant is also studied. Chapter 6 applies these methods to the determination of kinetic parameters for protein-protein and protein-ligand interactions in the intermediate to fast exchange regimes using line-shape analysis. Using the Bloch-McConnell equations to determine kinetic parameters used to generate quantum mechanical line-shape simulations demonstrates the range of timescales amenable to classical line-shape analysis.

In the final chapter, the preceding studies are reflected upon, and future experiments designed to further dissect and unravel the molecular mechanisms underlying the synthesis and recognition of polyUb chains are proposed.

References

1. Venter, J.C. *et al.* (2001) The sequence of the human genome. *Science* 291, 1304–1351.

2. Gordon, A.M., Homsher, E., and Regnier, M. (2000) Regulation of contraction in striated muscle. *Physiological Reviews* 80, 853–924.

3. Wolberger, C. (1999) Multiprotein-DNA complexes in transcriptional regulation. *Annual Review of Biophysics and Biomolecular Structure* 28, 29–56.

4. Karin, M. and Hunter, T. (1995) Transcriptional control by protein phosphorylation: signal transmission from the cell surface to the nucleus. *Current Biology* 5, 747–757.

5. Pufall, M.A. and Graves, B.J. (2002) Autoinhibitory domains: modular effectors of cellular regulation. *Annual Review of Cell and Developmental Biology* 18, 421–462.

6. Huse, M. and Kuriyan, J. (2002) The conformational plasticity of protein kinases. *Cell* 109, 275–282.

7. Ciechanover, A. (1994) The ubiquitin-proteasome proteolytic pathway. *Cell* 79, 13–21.

8. Komander, D. and Rape, M. (2012) The ubiquitin code. *Annual Review of Biochemistry* 81, 203–229.

9. Finley, D. and Varshavsky, A. (1985) The ubiquitin system: functions and mechanisms. *Trends in Biochemical Sciences* 10, 343–347.

10. Pickart, C.M. (2001) Mechanisms underlying ubiquitination. *Annual Review of Biochemistry* 70, 503–533.

11. Ye, Y. and Rape, M. (2009) Building ubiquitin chains: E2 enzymes at work. *Nature Reviews Molecular Cell Biology* 10, 755–764.

12. Schulman, B.A. and Harper, J.W. (2009) Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nature Reviews Molecular Cell Biology* 10, 319–331.

13. Haas, A.L. and Rose, I.A. (1982) The mechanism of ubiquitin activating enzyme. A kinetic and equilibrium analysis. *Journal of Biological Chemistry* 257, 10329–10337.

14. Lee, I. and Schindelin, H. (2008) Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes. *Cell* 134, 268–278.

15. Tong, H., Hateboer, G., Perrakis, A., Bernards, R., and Sixma, T.K. (1997) Crystal structure of murine/human Ubc9 provides insight into the variability of the ubiquitin-conjugating system. *Journal of Biological Chemistry* 272, 21381–21387.

16. Haldeman, M.T., Xia, G., Kasperek, E.M., and Pickart, C.M. (1997) Structure and function of ubiquitin conjugating enzyme E2-25K: the tail is a core-dependent activity element. *Biochemistry* 36, 10526–10537.

17. Brzovic, P.S., Lissounov, A., Christensen, D.E., Hoyt, D.W., and Klevit, R.E. (2006) A UbcH5/ubiquitin noncovalent complex is required for processive BRCA1-directed ubiquitination. *Molecular Cell* 21, 873–880.

18. Pruneda, J.N., Stoll, K.E., Bolton, L.J., Brzovic, P.S., and Klevit, R.E. (2011) Ubiquitin in motion: Structural studies of the E2∼Ub conjugate. *Biochemistry* 50, 1624–1633.

19. Andersen, P.L., Zhou, H., Pastushok, L., Moraes, T., McKenna, S., Ziola, B., Ellison, M.J., Dixit, V.M., and Xiao, W. (2005) Distinct regulation of Ubc13 functions by the two ubiquitin-conjugating enzyme variants Mms2 and Uev1A. *The Journal of Cell Biology* 170, 745–755.

20. Lewis, M.J., Saltibus, L.F., Hau, D.D., Xiao, W., and Spyracopoulos, L. (2006) Structural basis for non-covalent interaction between ubiquitin and the ubiquitin conjugating enzyme variant human MMS2. *Journal of Biomolecular NMR* 34, 89–100.

21. Eddins, M.J., Carlile, C.M., Gomez, K.M., Pickart, C.M., and Wolberger, C. (2006) Mms2–Ubc13 covalently bound to ubiquitin reveals the structural basis of linkage-specific polyubiquitin chain formation. *Nature Structural & Molecular Biology* 13, 915–920.

22. Kraut, J. (1977) Serine proteases: structure and mechanism of catalysis. *Annual Review of Biochemistry* 46, 331–358.

23. Wu, P.Y., Hanlon, M., Eddins, M., Tsui, C., Rogers, R.S., Jensen, J.P., Matunis, M.J., Weissman, A.M., Wolberger, C.P., and Pickart, C.M. (2003) A conserved catalytic residue in the ubiquitin-conjugating enzyme family. *The EMBO Journal* 22, 5241–5250.

24. Berndsen, C.E., Wiener, R., I.W., Yu, Ringel, A.E., and Wolberger, C. (2013) A conserved asparagine has a structural role in ubiquitin-conjugating enzymes. *Nature Chemical Biology* 9, 154–156.

25. Yunus, A.A. and Lima, C.D. (2006) Lysine activation and functional analysis of E2-mediated conjugation in the SUMO pathway. *Nature Structural & Molecular Biology* 13, 491–499.

26. Markin, C.J., Saltibus, L.F., Kean, M.J., McKay, R.T., Xiao, W., and Spyracopoulos, L. (2010) Catalytic proficiency of ubiquitin conjugation enzymes: Balancing pK_{a} suppression, entropy, and electrostatics. *Journal of the American Chemical Society* 132, 17775–17786.

27. VanDemark, A.P., Hofmann, R.M., Tsui, C., Pickart, C.M., and Wolberger, C. (2001) Molecular insights into polyubiquitin chain assembly: crystal structure of the Mms2/Ubc13 heterodimer. *Cell* 105, 711–720.

28. Metzger, M.B., Hristova, V.A., and Weissman, A.M. (2012) HECT and RING finger families of E3 ubiquitin ligases at a glance. *Journal of Cell Science* 125, 531–537.

29. Deshaies, R.J. and Joazeiro, C.A.P. (2009) RING domain E3 ubiquitin ligases. *Annual Review of Biochemistry* 78, 399–434.

30. Yin, Q., Lin, S.C., Lamothe, B., Lu, M., Lo, Y.C., Hura, G., Zheng, L., Rich, R.L., Campos, A.D., Myszka, D.G., *et al.* (2009) E2 interaction and dimerization in the crystal structure of TRAF6. *Nature Structural & Molecular Biology* 16, 658–666.

31. Campbell, S.J., Edwards, R.A., Leung, C.C.Y., Neculai, D., Hodge, C.D., Dhe-Paganon, S., and Glover, J.N.M. (2012) Molecular insights into the function of RING finger (RNF)-containing proteins hRNF8 and hRNF168 in Ubc13/Mms2-dependent ubiquitylation. *Journal of Biological Chemistry* 287, 23900–23910.

32. Borden, K.L., Boddy, M.N., Lally, J., O'Reilly, N.J., Martin, S., Howe, K., Solomon, E., and Freemont, P.S. (1995) The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML. *The EMBO Journal* 14, 1532.

33. Ohi, M.D., Vander Kooi, C.W., Rosenberg, J.A., Chazin, W.J., and Gould, K.L. (2003) Structural insights into the U-box, a domain associated with multi-ubiquitination. *Nature Structural Biology* 10, 250–255.

34. Zheng, N., Schulman, B.A., Song, L., Miller, J.J., Jeffrey, P.D., Wang, P., Chu, C., Koepp, D.M., Elledge, S.J., Pagano, M., Conaway, R.C., Conaway, J.W., Harper, J.W., and Pavletich, N.P. (2002) Structure of the Cul1–Rbx1–Skp1–Fbox^{Skp2} SCF ubiquitin ligase complex. *Nature* 416, 703–709.

35. Ceccarelli, D.F., Tang, X., Pelletier, B., Orlicky, S., Xie, W., Plantevin, V., Neculai, D., Chou, Y.C., Ogunjimi, A., Al-Hakim, A., Varelas, X., Koszela, J., G.A., Wasney, Vedadi, M., Dhe-Paganon, S., Cox, S., Xu, S., Lopez-Girona, A., Mercurio, F., Wrana, J., Durocher, D., Meloche, S., Webb, D.R., Tyers, M., and Sicheri, F. (2011) An allosteric inhibitor of the human Cdc34 ubiquitin-conjugating enzyme. *Cell* 145, 1075–1087.

36. Zheng, N., Wang, P., Jeffrey, P.D., and Pavletich, N.P. (2000) Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* 102, 533–539.

37. Wu, G., Xu, G., Schulman, B.A., Jeffrey, P.D., Harper, J.W., and Pavletich, N.P. (2003) Structure of a β-TrCP1-Skp1-β-catenin complex: destruction motif binding and lysine specificity of the $SCF^{\beta-TrCP1}$ ubiquitin ligase. *Molecular Cell* 11, 1445–1456.

38. Schrödinger, LLC. (2006) The PyMOL molecular graphics system.

39. Nordquist, K.A., Dimitrova, Y.N., Brzovic, P.S., Ridenour, W.B., Munro, K.A., Soss, S.E., Caprioli, R.M., Klevit, R.E., and Chazin, W.J. (2009) Structural and functional characterization of the monomeric U-box domain from E4B. *Biochemistry* 49, 347–355.

40. Brzovic, P.S., Keeffe, J.R., Nishikawa, H., Miyamoto, K., Fox III, D., Fukuda, M., Ohta, T., and Klevit, R. (2003) Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex. *Proceedings of the National Academy of Sciences* 100, 5646–5651.

41. Cardozo, T. and Pagano, M. (2004) The SCF ubiquitin ligase: insights into a molecular machine. *Nature Reviews Molecular Cell Biology* 5, 739–751.

42. Petroski, M.D. and Deshaies, R.J. (2005) Function and regulation of cullin-RING ubiquitin ligases. *Nature Reviews Molecular Cell Biology* 6, 9–20.

43. Pierce, N.W., Kleiger, G., Shan, S., and Deshaies, R.J. (2009) Detection of sequential polyubiquitylation on a millisecond timescale. *Nature* 462, 615–619.

44. Liu, J. and Nussinov, R. (2010) Molecular dynamics reveal the essential role of linker motions in the function of Cullin–RING E3 ligases. *Journal of Molecular Biology* 396, 1508–1523.

45. Kleiger, G., Saha, A., Lewis, S., Kuhlman, B., and Deshaies, R.J. (2009) Rapid E2-E3 assembly and disassembly enable processive ubiquitylation of cullin-RING ubiquitin ligase substrates. *Cell* 139, 957–968.

46. Zhang, X., Chen, J., Wu, M., Wu, H., Arokiaraj, A.W., Wang, C., Zhang, W., Tao, Y., Huen, M.S.Y., and Zang, J. (2013) Structural basis for role of ring finger protein RNF168 RING domain. *Cell Cycle* 12, 312–321.

47. Doil, C., Mailand, N., Bekker-Jensen, S., Menard, P., Larsen, D.H., Pepperkok, R., Ellenberg, J., Panier, S., Durocher, D., Bartek, J., Lukas, J., and Lukas, C. (2009) RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins. *Cell* 136, 435–446.

48. Penengo, L., Mapelli, M., Murachelli, A.G., Confalonieri, S., Magri, L., Musacchio, A., Di Fiore, P.P., Polo, S., and Schneider, T.R. (2006) Crystal structure of the ubiquitin binding domains of rabex-5 reveals two modes of interaction with ubiquitin. *Cell* 124, 1183–1195.

49. Huen, M.S.Y. and Chen, J. (2010) Assembly of checkpoint and repair machineries at DNA damage sites. *Trends in Biochemical Sciences* 35, 101–108.

50. Pruneda, J.N., Littlefield, P.J., Soss, S.E., Nordquist, K.A., Chazin, W.J., Brzovic, P.S., and Klevit, R.E. (2012) Structure of an E3:E2∼Ub complex reveals an allosteric mechanism shared among RING/U-box ligases. *Molecular Cell* 47, 933–942.

51. Plechanovová, A., Jaffray, E.G., Tatham, M.H., Naismith, J.H., and Hay, R.T. (2012) Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* 489, 115–120.

52. Dou, H., Buetow, L., Sibbet, G.J., Cameron, K., and Huang, D.T. (2012) BIRC7-E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nature Structural & Molecular Biology* 19, 876–883.

53. Capili, A.D. and Lima, C.D. (2007) Taking it step by step: mechanistic insights from structural studies of ubiquitin/ubiquitin-like protein modification pathways. *Current Opinion in Structural Biology* 17, 726–735.

54. McKenna, S., Spyracopoulos, L., Moraes, T., Pastushok, L., Ptak, C., Xiao, W., and Ellison, M.J. (2001) Noncovalent interaction between ubiquitin and the human dna

repair protein Mms2 is required for Ubc13-mediated polyubiquitination. *Journal of Biological Chemistry* 276, 40120–40126.

55. Hofmann, R.M. and Pickart, C.M. (2001) In vitro assembly and recognition of Lys-63 polyubiquitin chains. *Journal of Biological Chemistry* 276, 27936–27943.

56. Liu, S. and Chen, Z.J. (2011) Expanding role of ubiquitination in NF-κB signaling. *Cell Research* 21, 6–21.

57. Maniatis, T. (1999) A ubiquitin ligase complex essential for the NF-κB, Wnt/Wingless, and Hedgehog signaling pathways. *Genes & Development* 13, 505–510.

58. Park, Y.C., Burkitt, V., Villa, A.R., Tong, L., and Wu, H. (1999) Structural basis for self-association and receptor recognition of human TRAF2. *Nature* 398, 533–538.

59. Ye, H., Arron, J.R., Lamothe, B., Cirilli, M., Kobayashi, T., Shevde, N.K., Segal, D., Dzivenu, O.K., Vologodskaia, M., Yim, M., *et al.* (2002) Distinct molecular mechanism for initiating TRAF6 signalling. *Nature* 418, 443–447.

60. Kim, H., Chen, J., and Yu, X. (2007) Ubiquitin-binding protein RAP80 mediates BRCA1-dependent DNA damage response. *Science* 316, 1202–1205.

61. Sobhian, B., Shao, G., Lilli, D.R., Culhane, A.C., Moreau, L.A., Xia, B., Livingston, D.M., and Greenberg, R.A. (2007) RAP80 targets BRCA1 to specific ubiquitin structures at DNA damage sites. *Science* 316, 1198–1202.

62. Wang, B., Matsuoka, S., Ballif, B.A., Zhang, D., Smogorzewska, A., Gygi, S.P., and Elledge, S.J. (2007) Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. *Science* 316, 1194–1198.

63. Yan, J. and Jetten, A.M. (2008) RAP80 and RNF8, key players in the recruitment of repair proteins to DNA damage sites. *Cancer Letters* 271, 179–190.

64. van Gent, D.C., Hoeijmakers, J.H.J., and Kanaar, R. (2001) Chromosomal stability and the DNA double-stranded break connection. *Nature Reviews Genetics* 2, 196–206.

65. Harper, J.W. and Elledge, S.J. (2007) The DNA damage response: ten years after. *Molecular Cell* 28, 739–745.

66. Kolas, N.K., Chapman, J.R., Nakada, S., Ylanko, J., Chahwan, R., Sweeney, F.D., Panier, S., Mendez, M., Wildenhain, J., Thomson, T.M., Pelletier, L., Jckson, S.P., and Durocher, D. (2007) Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase. *Science* 318, 1637–1640.

67. Bekker-Jensen, S., Lukas, C., Kitagawa, R., Melander, F., Kastan, M.B., Bartek, J., and Lukas, J. (2006) Spatial organization of the mammalian genome surveillance machinery in response to DNA strand breaks. *The Journal of Cell Biology* 173, 195–206.

68. Polo, S.E. and Jackson, S.P. (2011) Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. *Genes & Development* 25, 409–433.

69. Fisher, R.D., Wang, B., Alam, S.L., Higginson, D.S., Robinson, H., Sundquist, W.I., and Hill, C.P. (2003) Structure and ubiquitin binding of the ubiquitin-interacting motif. *Journal of Biological Chemistry* 278, 28976–28984.

70. Dikic, I., Wakatsuki, S., and Walters, K.J. (2009) Ubiquitin-binding domains-from structures to functions. *Nature Reviews Molecular Cell Biology* 10, 659–671.

71. Swanson, K.A., Kang, R.S., Stamenova, S.D., Hicke, L., and Radhakrishnan, I. (2003) Solution structure of Vps27 UIM–ubiquitin complex important for endosomal sorting and receptor downregulation. *The EMBO Journal* 22, 4597–4606.

72. Nikkilä, J., Coleman, K.A., Morrissey, D., Pylkäs, K., Erkko, H., Messick, T.E., Karppinen, S.M., Amelina, A., Winqvist, R., and Greenberg, R.A. (2009) Familial breast cancer screening reveals an alteration in the RAP80 UIM domain that impairs DNA damage response function. *Oncogene* 28, 1843–1852.

73. Sato, Y., Yoshikawa, A., Mimura, H., Yamashita, M., Yamagata, A., and Fukai, S. (2009) Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by tandem UIMs of RAP80. *The EMBO Journal* 28, 2461–2468.

74. Markin, C.J., Xiao, W., and Spyracopoulos, L. (2010) Mechanism for recognition of polyubiquitin chains: balancing affinity through interplay between multivalent binding and dynamics. *Journal of the American Chemical Society* 132, 11247–11258.

75. Liu, F. and Walters, K.J. (2010) Multitasking with ubiquitin through multivalent interactions. *Trends in Biochemical Sciences* 35, 352–360.

76. Hirano, S., Kawasaki, M., Ura, H., Kato, R., Raiborg, C., Stenmark, H., and Wakatsuki, S. (2006) Double-sided ubiquitin binding of Hrs-UIM in endosomal protein sorting. *Nature Structural & Molecular Biology* 13, 272–277.

77. Varadan, R., Assfalg, M., Raasi, S., Pickart, C., and Fushman, D. (2005) Structural determinants for selective recognition of a lys48-linked polyubiquitin chain by a UBA domain. *Molecular Cell* 18, 687–698.

78. Varadan, R., Walker, O., Pickart, C., and Fushman, D. (2002) Structural properties of polyubiquitin chains in solution. *Journal of Molecular Biology* 324, 637–647.

79. Varadan, R., Assfalg, M., Haririnia, A., Raasi, S., Pickart, C., and Fushman, D. (2004) Solution conformation of Lys63-linked di-ubiquitin chain provides clues to functional diversity of polyubiquitin signaling. *Journal of Biological Chemistry* 279, 7055–7063.

80. Haas, A.L. (1988) Immunochemical probes of ubiquitin pool dynamics. In *Ubiquitin*, ed. Rechsteiner, M. (Plenum Press, New York), pp. 173–206.

81. Sims, J.J. and Cohen, R.E. (2009) Linkage-specific avidity defines the lysine 63-linked polyubiquitin-binding preference of Rap80. *Molecular Cell* 33, 775–783.

82. Haririnia, A., DOnofrio, M., and Fushman, D. (2007) Mapping the interactions between Lys48 and Lys63-linked di-ubiquitins and a ubiquitin-interacting motif of S5a. *Journal of Molecular Biology* 368, 753–766.

83. Mammen, M., Choi, S.K., and Whitesides, G.M. (1998) Polyvalent interactions in biological systems: implications for design and use of multivalent ligands and inhibitors. *Angewandte Chemie International Edition* 37, 2754–2794.

84. Raiborg, C., Rusten, T.E., and Stenmark, H. (2003) Protein sorting into multivesicular endosomes. *Current Opinion in Cell Biology* 15, 446–455.

85. Kay, L., Keifer, P., and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *Journal of the American Chemical Society* 114, 10663–10665.

86. Jencks, W.P. (1981) On the attribution and additivity of binding energies. *Proceedings of the National Academy of Sciences* 78, 4046–4050.

87. Zaman, M.H., Berry, R.S., and Sosnick, T.R. (2002) Entropic benefit of a cross-link in protein association. *Proteins: Structure, Function, and Bioinformatics* 48, 341–351.

88. Zhou, H-X. and Gilson, M.K. (2009) Theory of free energy and entropy in noncovalent binding. *Chemical Reviews* 109, 4092–4107.

89. Zhou, H-X. (2001) The affinity-enhancing roles of flexible linkers in two-domain DNA-binding proteins. *Biochemistry* 40, 15069–15073.

90. Wilhelm, J. and Frey, E. (1996) Radial distribution function of semiflexible polymers. *Physical Review Letters* 77, 2581–2584.

91. Zhou, H-X. (2001) Loops in proteins can be modeled as worm-like chains. *The Journal of Physical Chemistry B* 105, 6763–6766.

92. Zhou, H-X. (2003) Quantitative account of the enhanced affinity of two linked scFvs specific for different epitopes on the same antigen. *Journal of Molecular Biology* 329, 1–8.

93. Lapidus, L.J., Steinbach, P.J., Eaton, W.A., Szabo, A., and Hofrichter, J. (2002) Effects of chain stiffness on the dynamics of loop formation in polypeptides. Appendix: testing a 1-dimensional diffusion model for peptide dynamics. *The Journal of Physical Chemistry B* 106, 11628–11640.

94. Bonvin, A.M.J.J., Boelens, R., and Kaptein, R. (2005) NMR analysis of protein interactions. *Current Opinion in Chemical Biology* 9, 501–508.

95. Takeuchi, K. and Wagner, G. (2006) NMR studies of protein interactions. *Current Opinion in Structural Biology* 16, 109–117.

96. Saunders, M., Wishnia, A., and Kirkwood, J.G. (1957) The nuclear magnetic resonance spectrum of ribonuclease. *Journal of the American Chemical Society* 79, 3289–3290.

97. Markley, J.L. and Ulrich, E.L. (1984) Detailed analysis of protein structure and function by NMR spectroscopy: survey of resonance assignments. *Annual Review of Biophysics and Bioengineering* 13, 493–521.

98. Markley, J.L. (1975) Observation of histidine residues in proteins by nuclear magnetic resonance spectroscopy. *Accounts of Chemical Research* 8, 70–80.

99. Fischer, J.J. and Jardetzky, O. (1965) Nuclear magnetic relaxation study of intermolecular complexes. The mechanism of penicillin binding to serum albumin. *Journal of the American Chemical Society* 87, 3237–3244.

100. Aaron, B.M., Oikawa, K., Reithmeier, R.A., and Sykes, B.D. (1984) Characterization of skeletal muscle calsequestrin by ^{1}H NMR spectroscopy. *Journal of Biological Chemistry* 259, 11876–11881.

101. Krugh, T.R. and Nuss, M.E. (1979) Nuclear magnetic resonance studies of drug-nucleic acid complexes. In *Biological applications of magnetic resonance*, ed. Shulman, R.G. (Academic Press, New York), pp. 113–176.

102. Marley, J., Lu, M., and Bracken, C. (2001) A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR* 20, 71–75.

103. Zuiderweg, E.R.P. (2002) Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* 41, 1–7.

104. Granot, J. (1983) Determination of dissociation constants of 1:1 complexes from NMR data. Optimization of the experimental setup by statistical analysis of simulated experiments. *Journal of Magnetic Resonance (1969)* 55, 216–224.

105. Schreiber, G., Haran, G., and Zhou, H-X. (2009) Fundamental aspects of protein-protein association kinetics. *Chemical Reviews* 109, 839–860.

106. Morrison, J.F. (1982) The slow-binding and slow, tight-binding inhibition of enzyme-catalysed reactions. *Trends in Biochemical Sciences* 7, 102–105.

107. Sculley, M.J., Morrison, J.F., and Cleland, W.W. (1996) Slow-binding inhibition: the general case. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* 1298, 78–86.

108. Palmer III, A.G., Kroenke, C.D., and Loria, J.P. (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods in Enzymology* 339, 204–238.

109. Levitt, M.H. (2001) *Spin dynamics.* (John Wiley and Sons, Ltd., Chichester, UK).

110. Wang, C. and Palmer III, A.G. (2003) Solution nmr methods for quantitative identification of chemical exchange in ^{15}N-labeled proteins. *Magnetic Resonance in Chemistry* 41, 866–876.

111. Farrow, N.A., Zhang, O., Forman-Kay, J.D., and Kay, L.E. (1994) A heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *Journal of Biomolecular NMR* 4, 727–734.

112. Montelione, G.T. and Wagner, G. (1989) 2d chemical exchange NMR spectroscopy by proton-detected heteronuclear correlation. *Journal of the American Chemical Society* 111, 3096–3098.

113. Loria, J.P., Rance, M., and Palmer III, A.G. (1999) A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. *Journal of the American Chemical Society* 121, 2331–2332.

114. Carver, J.P. and Richards, R.E. (1972) A general two-site solution for the chemical exchange produced dependence of t_2 upon the Carr-Purcell pulse separation. *Journal of Magnetic Resonance (1969)* 6, 89–105.

115. Tollinger, M., Skrynnikov, N.R., Mulder, F.A.A., Forman-Kay, J.D., and Kay, L.E. (2001) Slow dynamics in folded and unfolded states of an SH3 domain. *Journal of the American Chemical Society* 123, 11341–11352.

116. Davis, D.G., Perlman, M.E., and London, R.E. (1994) Direct measurements of the dissociation-rate constant for inhibitor-enzyme complexes via the $t_{1\rho}$ and t_2 (CPMG) methods. *Journal of Magnetic Resonance, Series B* 104, 266–275.

117. Carr, H.Y. and Purcell, E.M. (1954) Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review* 94, 630–638.

118. Meiboom, S. and Gill, D. (1958) Modified spin-echo method for measuring nuclear relaxation times. *Review of Scientific Instruments* 29, 688–691.

119. Vallurupalli, P., Hansen, D.F., and Kay, L.E. (2008) Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proceedings of the National Academy of Sciences* 105, 11766–11771.

120. Baldwin, A.J. and Kay, L.E. (2009) NMR spectroscopy brings invisible protein states into focus. *Nature Chemical Biology* 5, 808–814.

121. Kay, L.E., Torchia, D.A., and Bax, A. (1989) Backbone dynamics of proteins as studied by ^{1}5N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* 28, 8972–8979.

122. Hansen, D.F., Yang, D., Feng, H., Zhou, Z., Wiesner, S., Bai, Y., and Kay, L.E. (2007) An exchange-free measure of ^{15}N transverse relaxation: an NMR spectroscopy application to the study of a folding intermediate with pervasive chemical exchange. *Journal of the American Chemical Society* 129, 11468–11479.

123. Korchuganov, D.S., Nolde, S.B., Reibarkh, M.Y., Orekhov, V.Y., Schulga, A.A., Ermolyuk, Y.S., Kirpichnikov, M.P., and Arseniev, A.S. (2001) NMR study of monomer-dimer equilibrium of barstar in solution. *Journal of the American Chemical Society* 123, 2068–2069.

124. Tugarinov, V. and Kay, L.E. (2003) Quantitative NMR studies of high molecular weight proteins: application to domain orientation and ligand binding in the 723 residue enzyme malate synthase G. *Journal of molecular biology* 327, 1121–1133.

125. Smith, S.A., Levante, T.O., Meier, B.H., and Ernst, R.R. (1994) Computer simulations in magnetic resonance. An object-oriented programming approach. *Journal of Magnetic Resonance, Series A* 106, 75–105.

126. Lipari, G. and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *Journal of the American Chemical Society* 104, 4546–4559.

127. Lipari, G. and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *Journal of the American Chemical Society* 104, 4559–4570.

128. Prompers, J.J. and Brüschweiler, R. (2002) General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. *Journal of the American Chemical Society* 124, 4522–4534.

129. Michaelis, L. and Menten, M.L. (1913) Die kinetik der invertinwirkung. *Biochem. z* 49, 333–369.

130. Johnson, K.A. and Goody, R.S. (2011) The original Michaelis constant: translation of the 1913 Michaelis–Menten paper. *Biochemistry* 50, 8264–8269.

131. Briggs, G.E. and Haldane, J.B.S. (1925) A note on the kinetics of enzyme action. *Biochemical Journal* 19, 338–339.

132. Farrow, L.A. and Edelson, D. (1974) The steady-state approximation: Fact or fiction? *International Journal of Chemical Kinetics* 6, 787–800.

133. Edelson, D. (1973) On the solution of differential equations arising in chemical kinetics. *Journal of Computational Physics* 11, 455.

134. McKenna, S., Hu, J., Moraes, T., Xiao, W., Ellison, M.J., and Spyracopoulos, L. (2003) Energetics and specificity of interactions within UbUevUbc13 human ubiquitin conjugation complexes. *Biochemistry* 42, 7922–7930.

135. Wolfram, S. (1999) *The Mathematica book*. (Cambridge University Press).

136. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage t4. *Nature* 227, 680–685.

137. Connors, K.A. and Bender, M.L. (1961) The kinetics of alkaline hydrolysis and *n*-butylaminolysis of ethyl *p*-nitrobenzoate and ethyl *p*-nitrothiolbenzoate. *The Journal of Organic Chemistry* 26, 2498–2504.

138. Page, M.I. and Jencks, W.P. (1971) Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proceedings of the National Academy of Sciences* 68, 1678–1683.

139. Villa, J., Štrajbl, M., Glennon, TM, Sham, YY, Chu, ZT, and Warshel, A. (2000) How important are entropic contributions to enzyme catalysis? *Proceedings of the National Academy of Sciences* 97, 11899–11904.

140. Villa, J. and Warshel, A. (2001) Energetics and dynamics of enzymatic reactions. *The Journal of Physical Chemistry B* 105, 7887–7907.

Chapter 2

Dynamics of the RING domain from human TRAF6 by ^{15}N NMR spectroscopy: implications for biological function*

Introduction

The TNF1 receptor superfamily and the interleukin-1/Toll-like receptors utilize TRAF proteins as signaling adaptors (1, 2). These receptors recruit TRAF proteins and are involved in the activation of cells, cell differentiation, immunity, and signaling for survival (3, 4). The adaptor proteins TRAF2 and TRAF6 have been studied intensively due to

*This chapter has been published. Reproduced with permission from: C.J. Markin, L.F. Saltibus, and L. Spyracopoulos. Dynamics of the RING domain from human TRAF6 by ^{15}N NMR spectroscopy: implications for biological function. *Biochemistry*, 47(38):10010-10017, 2008. Copyright 2008 American Chemical Society.

C.J.M. and L.S. designed experiments. L.S. and C.J.M. wrote the paper. C.J.M. and L.S. carried out the NMR experiments. L.F.S. contributed help with the protein purifications.

their role in the activation of NF-κB (5, 6), a transcription factor that triggers genes involved in the cell cycle, differentiation, apoptosis, and the immune response (7).

The C-terminal regions of TRAF2 and TRAF6 are composed of a MATH domain that is responsible for interactions with the cytoplasmic domains from TNF receptors (8). The N-terminal region of TRAF6 is critical for signaling and contains a RING domain (9, 10). The RING domain from TRAF6 is believed to function as an E3 ubiquitin ligase and plays a role in autoubiquitination by catalyzing covalent attachment of Lys63 poly-Ub chains, through interaction with the E2 Ubc13 (11, 12). The first step of the autoubiquitination process involves attachment of a single Ub to TRAF6 and occurs with slow kinetics (13), consistent with the weak binding observed between the RING domain and Ubc13 (10). In addition, the RING domain from TRAF6 has been found to be important in autodegradation and the induction of apoptosis (14). These studies underscore the fact that the processes underlying substrate ubiquitination and subsequent chain elongation are currently not clearly understood (15).

RING domains from E3 ubiquitin ligases carry out their biological functions by binding their cognate E2 ubiquitin conjugation enzymes, as typified by the structure of c-Cbl bound to UbcH7 (16). It has been noted (17, 18) that many E3 ligases homo- or heterodimerize directly through their RING or RING-like domains to achieve their biological function (19, 20). For example, it is believed that the RING-RING heterodimer is essential for substrate recruitment and efficient ubiquitin transfer for the Bmi-1-RING1B polycomb group ubiquitin ligase complex (20).

In this study, the E3 ubiquitin ligase RING domain from TRAF6 (residues 67-124) was determined to be monomeric in solution, through analysis of ^{1}H-^{15}N two-dimensional (2D) NMR spectra with changes in protein concentration. Furthermore, chemical exchange phenomena on the micro- to millisecond time scale were assessed qualitatively using model-free analysis of ^{15}N R_1, ^{15}N R_2, and $\{^{1}H^{N}\}$-^{15}N NOE and measurement of η_{xy} and η_{z} cross-correlated relaxation rates and analyzed quantitatively using CPMG relaxation dispersion measurements (21–23). The relaxation dispersion experiments were interpreted in a simple fashion by assuming chemical exchange between two sites. For residues near the C-terminal end of the main α-helix within TRAF6-RD, chemical exchange phenomena were rationalized by predicting ^{15}N chemical shifts for the ensemble

of 50 solution structures for TRAF6-RD (2JMD) (10) with SHIFTS (24–26).

Materials and methods

Purification of [*U*-^{15}N]-TRAF6-RD

Overexpression and purification of 15N-labeled TRAF6-RD were carried out as previously described (10), with the exception that LB and M9 minimal media were supplemented with 100 µM $ZnCl_2$ and 50 mM TRIS or BIS-TRIS, 150 mM NaCl, pH 7.0 buffer was used for cell lysis and subsequent protein purification.

^{15}N chemical shift mapping

NMR samples contained 600 µL of a 95:5 H_2O/D_2O mixture, containing 50 mM TRIS, 150 mM NaCl (pH 7.0), and 0.14 mM DSS as a chemical shift reference with protein concentrations of 0.304, 0.149, 0.110, 0.069, and 0.030 mM TRAF6-RD in 5 mm standard NMR tubes. In addition, as a control, a 10-fold dilution of 0.5 mM TRAF6-RD in the absence of DSS was conducted using similar conditions. For chemical shift mapping and ^{15}N R_2 measurements in the presence of DSS, NMR spectra were recorded at 25 ℃ using a Varian Unity INOVA 600 MHz NMR spectrometer and the sensitivity-enhanced, 2D ^{1}H-^{15}N HSQC experiment for measurement of ^{15}N R_2 (21). Spectra were collected with relaxation delays of 10 and 90 ms at each protein concentration. Chemical shift mapping was conducted using spectra acquired with the 10 ms delay. ^{15}N R_2 values were calculated from the natural logarithm of the ratio of cross-peak intensities at 10 and 90 ms divided by 80 ms. Protein concentrations were determined by quantitative amino acid analysis (27). The error in protein concentration was estimated to be 5%, on the basis of previous concentration determinations for the protein troponin C, conducted in triplicate (28). The number of transients collected was 16, 32, 32, 40, and 60 for protein concentrations of 0.304, 0.149, 0.110, 0.069, and 0.030 mM, respectively.

Spectra were processed using NMRPipe (29), and chemical shift assignments for cross-peaks within the 2D ^{1}H-^{15}N HSQC NMR spectra were made with Sparky (30), and previously published chemical shifts for TRAF6-RD (BRMB accession number 15014)

(10). Although the chemical shifts for residues K104 and F118 were unambiguous for the chemical shift mapping experiments, they became overlapped slowly over the course of a few days and could not be unambiguously assigned for the cross-correlated relaxation experiments and model-free analysis described in subsequent sections.

^{15}N chemical shift mapping was accomplished by following changes in 2D ^{1}H-^{15}N HSQC spectra upon dilution. These changes upon dilution occur only in the presence of DSS and therefore reflect binding of DSS to TRAF6. Thus, per residue plots of ^{15}N chemical shift (δ_{obs}) as a function of the protein:DSS ratio were fit to a 1:1 protein-ligand binding equilibrium to extract K_{D}, δ_{free}, and δ_{bound}. The precision of the chemical shift measurement (0.001 and 0.002 ppm for the $^{1}\mathrm{H}^{\mathrm{N}}$ and ^{15}N dimensions, respectively) was determined by taking the mean standard deviation over all assigned residues from five separate samples of [*U*-^{15}N]ubiquitin at concentrations ranging from 0.2 to 1.5 mM, using acquisition parameters similar to those used for TRAF6-RD.

The average ^{15}N R_2 values and their average errors as a function of increasing protein concentration are 5.9 ± 0.7, 5.7 ± 0.4, 6.1 ± 0.4, 6.3 ± 0.3, and 7.2 ± 0.4 s^{-1}. The increase in the average ^{15}N R_2 values as a function of protein concentration indicates that the underlying equilibrium may be more complicated than 1:1 protein-DSS binding.

^{15}N R_2 constant-time, relaxation-compensated CPMG dispersion measurements

^{15}N constant-time, relaxation-compensated CPMG dispersion experiments (23, 31) were carried out at 600 and 800 MHz with a 350 µL sample of 0.43 mM TRAF6-RD in a 95:5 H_2O/D_2O mixture, containing 50 mM BIS-TRIS, 150 mM NaCl, 5 mM DTT, 100 µM zinc sulfate (pH 7.0), and 0.14 mM DSS as an internal chemical shift reference, in a 5 mm Shigemi microcell NMR tube. The gNcpmgex_NH sequence from the Varian Biopack suite of pulse sequences was employed at 600 and 800 MHz. For the ^{15}N CPMG pulse trains, the ^{15}N 90° pulse widths were 44 and 54 µs at 600 and 800 MHz, respectively, and the constant relaxation time was 40 ms. The 800 MHz spectrometer was equipped with a cryogenically cooled probe, and heating compensation (as implemented in the gNcpmgex_NH sequence) was required to maintain constant heating for all CPMG pulse

repetition rates. Per residue $R_{2,\text{eff}}$ values were calculated according to eq 1 in (32). $R_{2,\text{eff}}$ dispersion profiles at both magnetic field strengths were globally fit (33) to eq 25 (Carver-Richards equation) and eq 28 (fast exchange only) in (34) for individual residues using Mathematica, and AIC was used to discriminate between the two functions. Errors in the fitted parameters were obtained using Monte Carlo analyses for global fits. To obtain a qualitative structural interpretation of the dispersion measurements, main chain amide chemical shifts were predicted using the ensemble of 50 NMR solution structures for TRAF6-RD and SHIFTS (version 4.2) (24–26). Predicted ^{15}N $\Delta\delta$ values were taken to be the width of the per residue calculated chemical shift distributions at 1σ, except for F89, for which the distribution was trimmed by removing two predicted shifts with values 122 ppm from the main ensemble of predicted shifts whose values ranged from 113 to 117 ppm.

^{15}N R_1, ^{15}N R_2, {^{1}H}-^{15}N NOE, η_{xy}, and η_z cross-correlated relaxation measurements and model-free analyses

Main chain amide ^{15}N η_{xy} and η_{z} rates were measured using sensitivity-enhanced 2D ^{1}H-^{15}N HSQC NMR experiments (22), at 600 MHz with pulse sequences coded in-house. Experiments were carried out with a 350 µL sample of 0.4 mM TRAF6-RD in a 95:5 H_2O/D_2O mixture containing 50 mM TRIS, 150 mM NaCl (pH 7.0), and 0.14 mM DSS as an internal chemical shift reference, in a 5 mm Shigemi microcell NMR tube. The per residue η_{xy} and η_{z} values were subsequently used with measurements of ^{15}N R_1, ^{15}N R_2, and {^{1}H}-^{15}N NOE (*vide infra*) to calculate per residue $R_{2,0}$ values; these were used to estimate an R_{ex} value (22). Errors were propagated accordingly from the spectral noise. R_{ex} values exceeding the mean by one standard deviation were considered significant. ^{15}N R_1, ^{15}N R_2, and {^{1}H}-^{15}N NOE experiments were conducted at 600 and 800 MHz using pulse sequences described in (21), and with the same protein sample that was used for measurements of η_{xy} and η_{z} cross-correlated relaxation rates. Per residue ^{15}N R_1, ^{15}N R_2, and {^{1}H}-^{15}N NOE values at 600 and 800 MHz were simultaneously subjected to a model-free analysis as previously described (35) using Mathematica (36), with model selection using AIC (37). For model-free analyses, a disadvantage of AIC compared

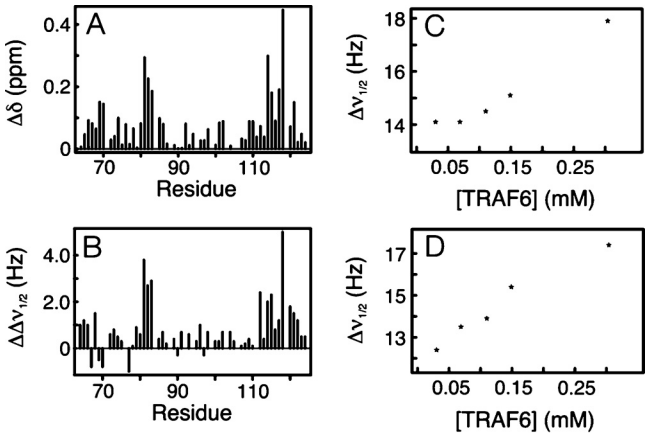

Figure 2.1: (A) Main chain amide ^{15}N chemical shift changes [$\Delta\delta = |\delta$(protein:DSS = 0.2) $- \delta$(protein:DSS = 2.2)|] for TRAF6-RD upon DSS binding. (B) Main chain amide ^{15}N line width changes [$\Delta\delta\nu_{1/2} = \Delta\delta\nu_{1/2}$(protein:DSS = 2.2) $- \Delta\delta\nu_{1/2}$(protein:DSS = 0.2)] for TRAF6-RD upon dilution. (C) Concentration dependence of line widths for TRAF6-RD V81 in the presence of 0.14 mM DSS. (D) Concentration dependence of line widths for TRAF6-RD F118 in the presence of 0.14 mM DSS

to hypothesis testing using α values is that AIC always discriminates between models, whereas hypothesis testing can reveal whether different models are indistinguishable (38). For the model-free analysis, protein tumbling was assumed to be isotropic, with an overall correlation time of 4.17 ns, determined from the R_2/R_1 ratio.

Results

^{15}N and ^{1}H^{N} NMR chemical shift and line width perturbations upon dilution

Line shape perturbations for TRAF6-RD upon dilution from 304 to 30 µM occur only in the presence of DSS and are summarized in figure 2.1. The largest main chain amide ^{15}N chemical shift changes ($\Delta\delta$) exceed the mean by 1.5σ and are observed for residues V81, Q82, E114, and F118. The largest changes in line width ($\Delta\nu_{1/2}$), exceeding the mean line width change by 1.5σ, occur for residues V81 (figure 2.1C), Q82, T83, and F118 (figure 2.1D).

Chemical shift changes in 2D ^{1}H-^{15}N NMR spectra upon dilution for residues V81 and F118 from TRAF6-RD are shown in figure 2.2. It should be noted that these changes

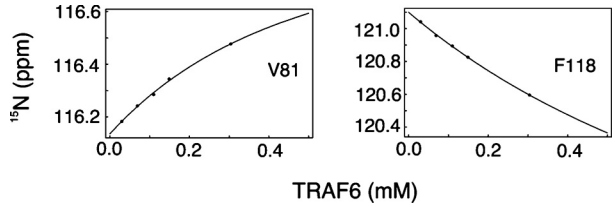

Figure 2.2: Main chain amide ^{15}N chemical shift changes for TRAF6-RD upon dilution, in the presence of 0.14 mM DSS

occur only in the presence of DSS and, in the simplest interpretation, reflect 1:1 protein-DSS binding. 15N $\Delta\delta$ values for V81 and F118 were fit to a 1:1 protein-DSS binding equilibrium to yield K_{D} values of 0.4 ± 0.2 and 1.0 ± 0.4 mM, respectively. In addition, changes in the ^{1}H chemical shift from the methyl groups of DSS upon protein dilution were also fit to a 1:1 protein-DSS binding equilibrium to yield a similar K_{D} of 0.2 ± 0.1 mM. Thus, for an average K_{D} of 0.5 mM, the fraction of free protein is 0.9 at a DSS concentration of 0.14 mM and a protein concentration of 0.4 mM.

^{15}N NMR R_2 relaxation dispersion

^{15}N relaxation dispersion experiments were used to quantify chemical exchange processes on the micro- to millisecond time scale (figures 2.3-2.5 and tables 2.1 and 2.2). Residues V81, N115, L117, and F118 show line shape changes upon dilution (and concomitant increases in protein:DSS ratio), display dispersion curves that are characteristic of the fast exchange limit, and fit with a global k_{ex} of 3514 ± 620 s^{-1} (figure 2.3 and table 2.1). While residue T83 also displays line shape changes upon dilution, dispersion data for this residue were statistically better fit with the Carver-Richards equation (39) and had fitted parameters for chemical exchange similar to those for residues G86 and H87; these residues participate with T83 in an extensive hydrogen bonding network. Thus, residues T83, G86, and H87 were globally fit with the Carver-Richards equation with global p_{a} and k_{ex} values of 0.989 ± 0.004 and 532 ± 165 s^{-1}, respectively (figure 2.4 and table 2.2). Residues I98, R99, and D100 at the C-terminal end of the short α-helix in TRAF6-RD also display exchange with k_{ex} values of 1121 ± 123, 4690 ± 680, and 3569 ± 834 s^{-1}, respectively (figure 2.5 and table 2.1). The relaxation dispersion profile for residue K104 was fit with the Carver-Richards equation with a $\Delta\omega_{\mathrm{N}}$ of 1213 ± 138 rad s^{-1} at 800

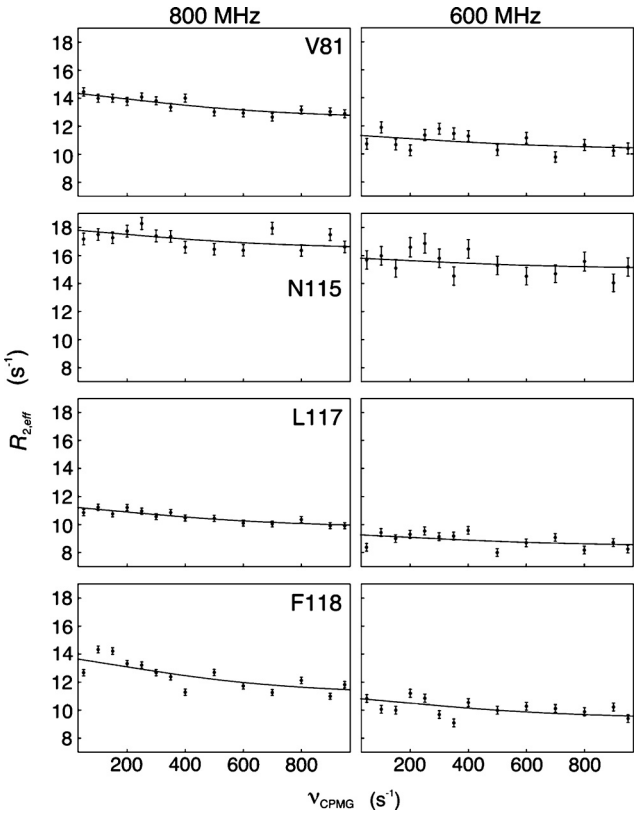

Figure 2.3: ^{15}N R_2 relaxation dispersion for TRAF6-RD, with multiple field data for each residue globally fit to chemical exchange between two sites, and k_{ex} fit globally for all residues

MHz (2 ppm) and with p_a and k_{ex} values of 0.98 ± 0.01 and 255 ± 264 s^{-1}, respectively (figure 2.5). Finally, residues E110 and L112 display relaxation dispersion with k_{ex} values of 2312 ± 629 and 6137 ± 2239 s^{-1}, respectively (table 2.1).

^{15}N NMR cross-correlated relaxation and model-free analysis

Chemical/conformational exchange phenomena were further verified using ^{15}N cross-correlated relaxation measurements (table 2.3). Residues 81-83, 86, 98, and 104/118 display R_{ex} terms that exceed the mean R_{ex} by 1σ. However, the calculated R_{ex} values for residues Q82 and I98 do not exceed 1σ within error. A model-free analysis was also performed using ^{15}N R_1, ^{15}N R_2, and $\{^1H^N\}$-^{15}N NOE data collected at 600 and 800 MHz, with the chief purpose of identifying residues undergoing chemical exchange. Residues 81-83, 86, 87, 98, 99, 104/118, 110, and 114 require R_{ex}^{600} terms ranging from

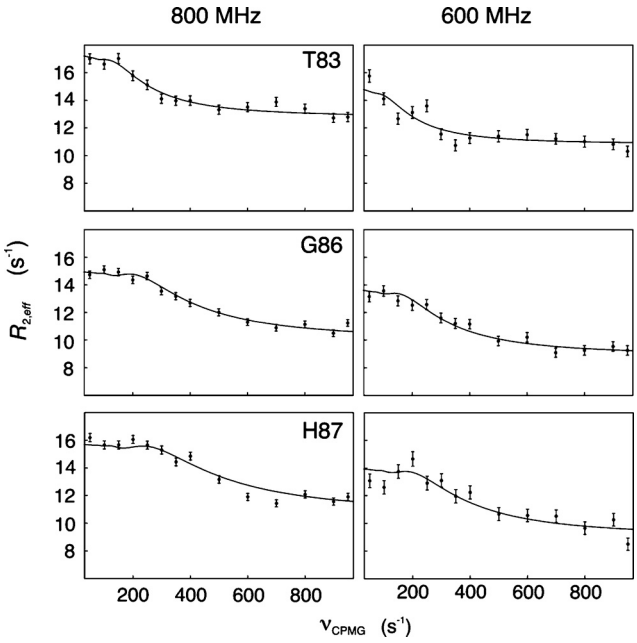

Figure 2.4: ^{15}N R_2 relaxation dispersion for TRAF6-RD, with multiple field data for each residue globally fit to chemical exchange between two sites, and parameters k_{ex} and p_A fit globally for all residues

Table 2.1: Chemical exchange parameters for ^{15}N CPMG R_2 dispersion of TRAF6-RD

Residue[a]	$p_a p_b \Delta\omega^2$ (rad^2 s^{-2})[b]	$R_{2,0}$ (s^{-1})[b]	$R_{2,0}$ (s^{-1})[c]
V81	7190 ± 2154	12.4 ± 0.3	10.2 ± 0.2
I98	4434 ± 414	12.3 ± 0.1	10.9 ± 0.1
R99	18325 ± 4715	9.9 ± 0.4	8.1 ± 0.2
D100	6367 ± 2090	9.8 ± 0.2	7.8 ± 0.1
E110	10984 ± 3466	10.3 ± 0.4	9.4 ± 0.2
L112	39361 ± 30030	9 ± 2	9 ± 1
N115	5396 ± 2647	16.3 ± 0.4	15.0 ± 0.3
L117	5761 ± 1845	9.6 ± 0.2	8.4 ± 0.1
F118	10143 ± 2526	10.9 ± 0.3	9.2 ± 0.2

[a] Residues 81, 115, 117, and 118 were fit simultaneously ($k_{ex} = 3514 \pm 620$ s^{-1}), and residues 98-100, 110, and 112 were fit individually (for k_{ex} values, see Results section of this chapter)
[b] Data collected at 800 MHz
[c] Data collected at 600 MHz

Table 2.2: Chemical exchange parameters for ^{15}N CPMG R_2 dispersion of TRAF6-RD

Residue[a]	$\Delta\omega$ (rad s^{-1})[b]	$R_{2,0}$ (s^{-1})[b]	$R_{2,0}$ (s^{-1})[c]
T83	1378 ± 167	12.8 ± 0.4	10.8 ± 0.2
G86	2375 ± 481	10.0 ± 0.9	8.9 ± 0.5
H87	2941 ± 420	10.7 ± 0.8	9.0 ± 0.4

[a] Simultaneous fits of the data yield a k_{ex} of 523 ± 165 s^{-1} and a p_a of 0.990 ± 0.004.
[b] Data collected at 800 MHz
[c] Data collected at 600 MHz

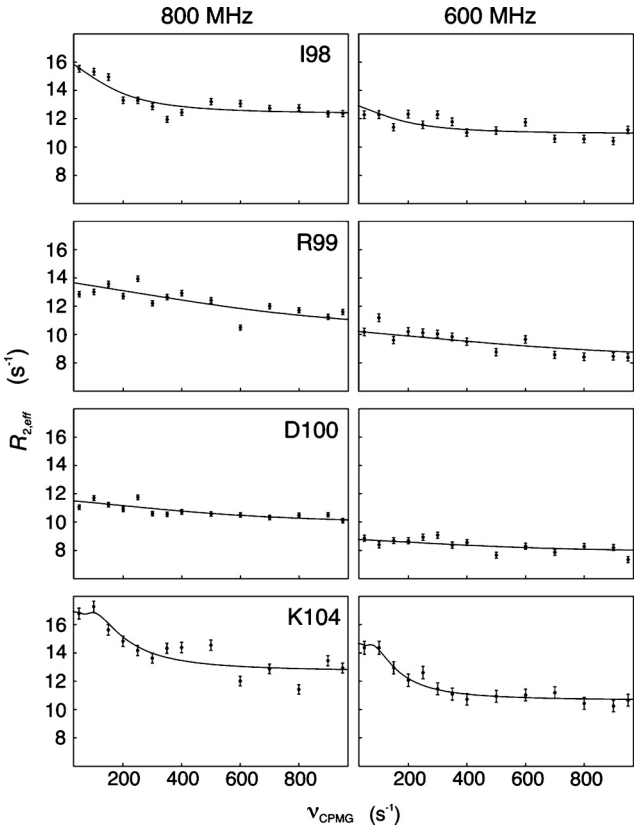

Figure 2.5: ^{15}N R_2 relaxation dispersion for TRAF6-RD, with multiple field data for each residue globally fit to chemical exchange between two sites, and individual residues analyzed separately

Table 2.3: Chemical exchange parameters from ^{15}N cross-correlated relaxation rates

Residue	R_{ex}^{600} (s^{-1})
V81	3.0 ± 0.3
Q82	2.0 ± 0.3
T83	3.1 ± 0.4
G86	2.5 ± 0.6
I98	1.9 ± 0.8
K104/F118[a]	2.9 ± 0.2

[a] K104 and F118 were overlapped for these experiments

Table 2.4: Chemical exchange parameters from ^{15}N model-free analysis

Residue	R_{ex}^{600} (s^{-1})[a]
V81	1.1 ± 0.1
Q82	0.7 ± 0.2
T83	1.6 ± 0.1
G86	2.1 ± 0.2
H87	3.1 ± 0.1
I98	1.6 ± 0.1
R99	0.8 ± 0.1
K104/F118[b]	1.6 ± 0.1
E114	0.7 ± 0.2

[a] Only R_{ex} values of > 0.6 s^{-1} are reported
[b] K104 and F118 were overlapped for these experiments

1 to 3 s^{-1} to fit their relaxation data (table 2.4). Although model 4 (S^2, τ_{ex}, R_{ex}) was chosen by AIC for residue 98, the S^2 value was unrealistic (0.07 ± 0.06), considering it is flanked by two residues with S^2 values of 0.83 and is found in a region of relatively well-defined structure. Thus, the model with the next lowest AIC value was chosen [model 3 (S^2, R_{ex})].

Discussion

Characterization of TRAF6-RD self-association and DSS binding using chemical shift mapping

In this study, we utilized an array of ^{15}N NMR spectroscopic approaches to assess intra- and intermolecular chemical exchange processes for the main chain of the TRAF6 RING domain. For line shape changes upon protein dilution in the presence of DSS, large ^{15}N

$\Delta\delta$ and $\Delta\Delta\nu_{1/2}$ values are indicative of structural changes accompanying self-association. However, chemical shift changes upon 10-fold dilution of 0.5 mM TRAF6-RD are not observed in the absence of DSS. This result indicates that TRAF6-RD is monomeric in solution but associates with DSS, a common chemical shift reference standard (40) that is an anionic detergent. Interestingly, line width increases for residues V81 and F118 in the ^{15}N dimension are observed with increasing protein concentration and the concomitant increase in protein:ligand ratio (figure 2.1C,D). For fast chemical exchange, line widths are expected to decrease for a 1:1 protein-ligand interaction with an increase in protein:ligand ratio, whereas the opposite effect is observed for TRAF6-RD-DSS binding. Therefore, it is likely that the underlying equilibrium is more complex than 1:1 protein-DSS binding.

Characterization of TRAF6-RD chemical exchange processes using relaxation dispersion, model-free analysis, and cross-correlated relaxation measurements

A key issue in the interpretation of relaxation dispersion measurements is the correct attribution of chemical exchange to intra- or intermolecular processes. For example, intramolecular instability for a ubiquitin-associated domain was characterized in terms of a folding-unfolding equilibrium (41). Dispersion experiments conducted at multiple ligand concentrations have been used successfully to separate intra- and intermolecular exchange processes for titration of the regulatory subunit of protein kinase A with cAMP (42), and titration of the phosphorylated kinase inducible activation domain of CREB with the KIX domain from CREB binding protein (43). The kinetics of the monomer-dimer equilibrium for the dimeric four-helix bundle α_2D have been quantified using ^{13}C relaxation dispersion spectroscopy (44), and monomer-dimer exchange for the protein GB1A34F was characterized recently using ^{15}N relaxation dispersion in combination with analysis of 2D ^{1}H-^{15}N NMR spectral changes upon dilution (45).

For the relaxation dispersion of TRAF6-RD, residues V81, N115, L117, and F118 were observed to undergo fast chemical exchange (figure 2.3 and table 2.1). Although only the product $p_{\mathrm{a}}p_{\mathrm{b}}\Delta\omega_N^2$ can be determined from curve fitting, if we assume $p_{\mathrm{a}} = 0.9$

and $p_b = 0.1$ (corresponding to the fractions of free and DSS bound protein with a K_D of 0.5 mM and TRAF6 and DSS concentrations of 0.4 and 0.14 mM, respectively), the fitted values of $p_a p_b \Delta\omega_N^2$ at 800 MHz give rise to maximum $\Delta\delta_N$ values ranging from 180 to 250 rad s^{-1} (from 0.5 to 0.7 ppm). In the presence of DSS, these residues undergo ^{15}N chemical shift changes upon dilution, raising the possibility that the kinetics associated with DSS-protein binding may be the source of chemical exchange. However, these maximum $\Delta\delta N$ values are 10-fold too small if structural changes in the protein accompany DSS binding. In addition, the increase in protein line widths with an increase in protein:DSS ratio (line width decreases are expected) is not consistent with 1:1 protein-DSS binding. These observations suggest that the interaction between protein and DSS is more complicated than 1:1 binding. Given that TRAF6 is a cytoplasmic signaling adaptor protein, DSS binding is not likely to be biologically significant, and a more detailed characterization of DSS binding kinetics is not warranted.

Residues I98, R99, and D100 are found near the C-terminus of the main α-helix in TRAF6-RD. These residues do not exhibit significant chemical shift changes upon dilution. Therefore, it is likely that the observed relaxation dispersion is due to intramolecular chemical exchange. We hypothesize that these residues are undergoing constrained helical fraying, potentially due to the fact that main chain hydrogen bonds near helical termini are not fulfilled. Motions for α-helices that are not constrained at their N- or C-termini typically occur on the nano- to picosecond time scale (46). For TRAF6-RD, the N- and C-termini of the main α-helix are constrained, in the sense that the polypeptide chain participates in additional tertiary structural interactions beyond the termini, and these may serve to slow the typical pico- to nanosecond time scale of motion for unconstrained termini of α-helices to the micro- to millisecond time scale. If we assume that the distribution of structures observed in the NMR ensemble (2JMD) at or near the ends of the main α-helix is a reasonable representation of conformational states between which transitions can occur, the ^{15}N $\Delta\delta$ values predicted by SHIFTS are 1.48, 1.03, and 1.95 ppm for residues 98-100, respectively; these residues are found at the C-terminal end of the α-helix (relaxation dispersion was not detected at the N-terminal end of the helix). In comparison, residues I94 and I95 are located in the middle of the helix and have buried side chains, and the predicted widths of the chemical shift distributions are

0.63 and 0.64 ppm, respectively. Additionally, these residues do not exhibit relaxation dispersion. Residues K96 and S97 are also near the middle of the helix and do not show relaxation dispersion but display larger widths for their predicted chemical shift distributions of 1.01 and 2.38 ppm, respectively. However, the larger widths for these residues arise mainly from contributions to the ^{15}N amide chemical shift due to the conformation of exposed side chains and are likely to be averaged to smaller values by fast time scale side chain fluctuations. In addition, the width of the predicted ^{15}N chemical shift distribution for residue F89 has a smaller value of 0.71 ppm. This residue packs within the hydrophobic core against residue I94 and does not display relaxation dispersion.

It should be noted that on the basis of the SHIFTS predictions, constrained helical fraying can potentially give rise to relaxation dispersion for residues at the N-terminal end of the α-helix, but this was not observed. A more fundamental analysis of the time scale and nature of chemical exchange processes potentially associated with constrained helical fraying can be obtained using molecular dynamics simulations, combined with quantum mechanical or empirical chemical shift calculations (26). However, these simulations are beyond the scope of this study, and the time scale necessary to observe micro- to millisecond time scale fluctuations is computationally intractable.

Finally, we note that for residues 98-100, the widths of their chemical shift distributions exceed the maximum $\Delta\omega$ values from the relaxation dispersion experiments by 4-fold (0.3-0.5 ppm). This observation suggests that a two-state transition is not appropriate if the underlying chemical exchange process is helix-fraying. Additionally, it is possible that the exchange processes observed at the C-terminal end of the α-helix reflect conformational transitions of His103, as observed in relaxation dispersion studies of the protein plastocyanin (47). The main chain amide resonance for this residue is not observed in 2D ^{1}H-^{15}N NMR spectra, presumably due to line broadening, and the adjacent residue K104 undergoes conformational exchange (figure 2.5).

For residues T83, G86, and H87, our straightforward structure-based analysis does not provide meaningful insights into the underlying chemical exchange process. These residues participate in an extensive hydrogen bond network involving H-bond donors and acceptors from both the side chain and main chain. The relaxation dispersion may be the result of chemical exchange processes resulting from zinc binding, given that the side

chain from H87 is believed to coordinate Zn^{2+}. However, Zn^{2+} binding to the BRCA1 RING domain occurs with high affinity (48), and metal exchange for the RING domain from CNOT4 occurs with slow kinetics (49) consistent with the relaxation dispersion experiments. The kinetics of Zn^{2+} binding for TRAF6-RD would need to be quantified to identify the origin of chemical exchange for residues T83, G86, and H87. Regardless of the source of chemical exchange, intra- or intermolecular modulation of the H-bond network in which residues T83, G86, and H87 participate is a likely source for their observed relaxation dispersion.

The remaining residues that display relaxation dispersion (K104, E110, and L112) are found proximal to loop L2 (residues 105-108) that is involved in coordinating zinc. Residue K104 is found in an isolated β-bridge, and residues E110 and L112 are not involved in regular secondary structure, as determined using DSSP (50). These residues do not display chemical shift changes upon dilution, and it is not obvious how their motions may be coupled to zinc binding; thus, the relaxation dispersion results may simply reflect intramolecular main chain dynamics.

In this study, ^{15}N R_1, ^{15}N R_2, $\{^1H^N\}$-^{15}N NOE, and η_{xy} NMR relaxation measurements were used to qualitatively identify residues undergoing chemical exchange. All residues (except Q82) that display R_{ex} terms, as established from cross-correlated relaxation rate measurements, also display ^{15}N relaxation dispersion profiles indicative of chemical exchange. Residues V81, Q82, T83, and F118 display significant line shape perturbations upon dilution and concomitant changes in the protein:DSS ratio, and the R_{ex} values for these residues may reflect DSS binding, as discussed below in further detail for the model-free analysis. The observation of significant R_{ex} values for residues T83, G86, and H87 is consistent with the results from the relaxation dispersion experiments.

We previously conducted a model-free analysis for TRAF6-RD using data collected at 600 MHz (10). In this study, our goal was to identify residues requiring R_{ex} terms to fit their relaxation data, and we have the advantage that data collected at 600 and 800 MHz were simultaneously analyzed, thereby improving model selection, and taking the field dependence of R_{ex} into account. Herein, we have assumed quadratic field dependence for the model-free analysis, whereas for slow and intermediate exchange, this field dependence is not necessarily quadratic (51). From the relaxation dispersion analy-

ses, all residues except T83, G86, H87, and K104 were fit with the relaxation dispersion equation for fast exchange only. Residues T83, G86, H87, and K104 were fit with the Carver-Richards equation, and the values of $k_{ex}/\Delta\omega$ and α range from 0.2 to 0.5 and from 0.1 to 0.4, respectively, indicating that these residues are in the slow exchange regime. For the model-free analysis, and analysis of cross-correlated relaxation rates, chemical exchange due to DSS binding can potentially contribute to the observed R_{ex} terms for residues 81-83 and 118 (table 2.4). For example, we estimate the magnitude of observable R_{ex} terms to be approximately 1 and 2 Hz at 600 and 800 MHz, respectively, given the experimental conditions used to acquire ^{15}N R_2, and assuming a rate of exchange of 2000 s^{-1}, with a p_a of 0.9 and a $\Delta\omega$ of 0.5 ppm. Thus, with the exception of residue Q82, these qualitative results agree with those from the relaxation dispersion experiments. However, we must consider that rotational diffusion anisotropy can potentially contribute to R_{ex} terms determined from cross-correlated relaxation measurements and model-free analysis (22, 52). In this study, anisotropic rotational diffusion does not substantially contribute to R_2 for residues 81-83 and 118. For example, for a protein with an isotropic overall correlation time of 4 ns, an axially symmetric rotational diffusion anisotropy of > 2 is necessary for a maximum contribution of $\sim 2\ s^{-1}$ to R_2, whereas monomeric TRAF6-RD is essentially spherical in shape, and monomeric using the solution conditions employed for relaxation measurements.

Biological implications for lack of TRAF6-RD self-association

Currently, it is not clear if E3 RING homo- or heteroassociation serves a general biological purpose (18). It has been observed that RING self-association creates supramolecular structures in vitro and these serve as polyvalent binding surfaces to provide a platform for multiple partner proteins (53). However, it is not certain if supramolecular structures are necessary for the activity of TRAF proteins, whose main function is to serve as cytoplasmic signaling adaptor proteins. For example, the C-terminal region of TRAF6, which does not contain the RING domain, is known to self-associate as a trimer (54), and this property is necessary for receptor binding (55). That is, the TNF receptor family is believed to be an example for the general theme of receptor activation through oligomerization, as the TNF receptor must associate as a trimer for activity (56). How-

ever, several RING domains are known to associate as homodimers or heterodimers (17, 19, 57, 58), and this property may be critical for biological function. For example, it has been suggested that molecular architectures of E3 ligases are tuned to orient and align E2 enzymes and substrates for efficient ubiquitin transfer, but it is not clear what role RING domain dimerization plays in this function (17). Although it is reasonable to expect that the regions surrounding RING domains are critical for dimerization, the isolated RING domain from Hdm2 was shown to form a homodimer at low concentrations and interact tightly with a RING domain from the homologue HdmX (58). It is interesting to note that the TRAF6 RING domain and surrounding regions are most similar in sequence to the corresponding region in the dimerization domain of RAG1, which contains a RING domain and a C2H2 zinc finger (12, 59). In the RAG1 dimer structure, the RING domains are not in physical contact, separated by 15 Å.

The lack of self-association of the RING domain from TRAF6 and the fact that the intact protein is a trimer suggest that the RING domains may not be closely associated in the intact protein. With respect to functional significance, others have noted that a possible role for RING domains is to serve as protein-protein interaction modules involved in binding RING domains from different proteins (58). Therefore, maintaining the RING domains physically separated in intact TRAF6 may help fulfill this purpose.

Biological implications of TRAF6-RD slow time scale helical motions

The structure of the complex between the RING domain from c-Cbl and UbcH7 is an archetypal E2-E3 interaction and was used to develop a rough model of TRAF6-RD associated with its cognate E2 Ubc13 (figure 2.6). Residues V81, T83, N115, L117, and F118 are involved in DSS binding, and the kinetics of this interaction are manifest as micro- to millisecond time scale chemical exchange and line shape perturbations upon changing the protein:DSS ratio. Interestingly, while the isolated RING domain from TRAF6 does not self-associate, these residues occur at the typical RING dimer interface, and this surface is different from that involved in the TRAF6-Ubc13 interaction. Of the residues involved in typical E2-E3 interactions, residues I98, R99, and D100 display micro- to millisecond

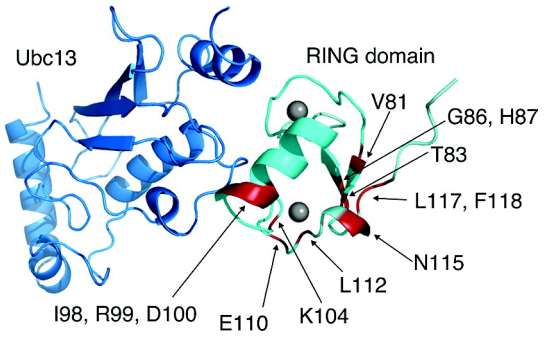

Figure 2.6: Superposition of TRAF6-RD and the E2 enzyme Ubc13 on the c-Cbl RING domain-UbcH7 complex (PDB entry 1FBV). Main chain atoms shown in the cartoon representation and residues displaying dispersion for main chain amide ^{15}N R_2 relaxation rates are colored red

time scale motions. While there are many enthalpic and entropic factors that determine the affinity of protein-protein and protein-ligand interactions, modulation of slow time scale motions observed for residues 98-100 (increases or decreases in flexibility) may be a general mechanism allowing precise adjustment of the affinities of E2-E3 interactions in achieving optimal substrate polyubiquitination. Specifically, quenching of these motions upon E2 binding may contribute unfavorably to the entropy of binding, and this may be an important mechanism for maintaining weak E2-E3 interactions, with fast kinetics, to allow for rapid transfer of ubiquitin to protein substrates bound to the E3 ligase. From a computational perspective, a recent 50 ns molecular dynamics study for receptor-ligand interactions led to the generalization that receptor flexibility contributes a large proportion of the compensating entropy/enthalpy terms of binding free energy (60). Quenching of chemical exchange has been observed experimentally through relaxation dispersion measurements for pheromone binding to major urinary protein (61). Ultimately, further experimental verification and longer time scale simulations will be necessary to test whether this generalization is applicable for slower time scale motions in E2-E3 ubiquitination systems.

References

1. Bradley, J. R. and Pober, J. S. (2001) Tumor necrosis factor receptor-associated factors (TRAFs). *Oncogene* 20, 6482–6491.

2. Bodmer, J.L., Schneider, P., and Tschopp, J. (2002) The molecular architecture of the TNF superfamily. *Trends in Biochemical Sciences* 27, 19–26.

3. Liu, Y.C. (2004) Ubiquitin ligases and the immune response. *Annual Review of Immunology* 22, 81–127.

4. Watts, T.H. (2005) TNF/TNFR family members in costimulation of T cell responses. *Annual Review of Immunology* 23, 23–68.

5. Kobayashi, T., Walsh, M.C., and Choi, Y. (2004) The role of TRAF6 in signal transduction and the immune response. *Microbes and Infection* 6, 1333–1338.

6. Chen, Z.J. (2005) Ubiquitin signalling in the NF-κB pathway. *Nature Cell Biology* 7, 758–765.

7. Ghosh, S., May, M.J., and Kopp, E.B. (1998) NF-κB and Rel proteins: evolutionarily conserved mediators of immune responses. *Annual Review of Immunology* 16, 225–260.

8. Ye, H., Arron, J.R., Lamothe, B., Cirilli, M., Kobayashi, T., Shevde, N.K., Segal, D., Dzivenu, O.K., Vologodskaia, M., Yim, M., Du, K., Singh, S., Pike, J.W., Darnay, B.G., Choi, Y., and Hao, W. (2002) Distinct molecular mechanism for initiating TRAF6 signalling. *Nature* 418, 443–447.

9. Régnier, C.H., Tomasetto, C., Moog-Lutz, C., Chenard, M.P., Wendling, C., Basset, P., and Rio, M.C. (1995) Presence of a new conserved domain in CART1, a novel member of the tumor necrosis factor receptor-associated protein family, which is expressed in breast carcinoma. *Journal of Biological Chemistry* 270, 25715–25721.

10. Mercier, P., Lewis, M.J., Hau, D.D., Saltibus, L.F., Xiao, W., and Spyracopoulos, L. (2007) Structure, interactions, and dynamics of the RING domain from human TRAF6. *Protein Science* 16, 602–614.

11. Deng, L., Wang, C., Spencer, E., Yang, L., Braun, A., You, J., Slaughter, C., Pickart, C., and Chen, Z.J. (2000) Activation of the IκB kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain. *Cell* 103, 351–361.

12. Lamothe, B., Besse, A., Campos, A.D., Webster, W.K., Wu, H., and Darnay, B.G. (2007) Site-specific lys-63-linked tumor necrosis factor receptor-associated factor 6 auto-ubiquitination is a critical determinant of IκB kinase activation. *Journal of Biological Chemistry* 282, 4102–4112.

13. Petroski, M.D., Zhou, X., Dong, G.Q., Daniel-Issakani, S., Payan, D.G., and Huang, J. (2007) Substrate modification with lysine 63-linked ubiquitin chains through the UBC13-UEV1A ubiquitin-conjugating enzyme. *Journal of Biological Chemistry* 282, 29936–29945.

14. He, L., Wu, X., Siegel, R., and Lipsky, P.E. (2006) TRAF6 regulates cell fate decisions by inducing caspase 8-dependent apoptosis and the activation of NF-κB. *Journal of Biological Chemistry* 281, 11235–11249.

15. Windheim, M., Peggie, M., and Cohen, P. (2010) Two different classes of E2 ubiquitin conjugating enzymes are required for the mono-ubiquitination of proteins and elongation by polyubiquitin chains with a specific topology. *Biochemical Journal* 409, 723–729.

16. Zheng, N., Wang, P., Jeffrey, P.D., and Pavletich, N.P. (2000) Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* 102, 533–539.

17. Vander Kooi, C.W., Ohi, M.D., Rosenberg, J.A., Oldham, M.L., Newcomer, M.E., Gould, K.L., and Chazin, W.J. (2006) The Prp19 U-box crystal structure suggests a common dimeric architecture for a class of oligomeric E3 ubiquitin ligases. *Biochemistry* 45, 121–130.

18. Knipscheer, P. and Sixma, T.K. (2007) Protein-protein interactions regulate Ubl conjugation. *Current opinion in structural biology* 17, 665–673.

19. Brzovic, P.S., Rajagopal, P., Hoyt, D.W., King, M.C., and Klevit, R.E. (2001) Structure of a BRCA1-BARD1 heterodimeric RING-RING complex. *Nature Structural and Molecular Biology* 8, 833–837.

20. Li, Z., Cao, R., Wang, M., Myers, M.P., Zhang, Y., and Xu, R.M. (2006) Structure of a Bmi-1-Ring1B polycomb group ubiquitin ligase complex. *Journal of Biological Chemistry* 281, 20643–20649.

21. Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. (1994) Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* 33, 5984–6003.

22. Kroenke, C.D., Loria, J.P., Lee, L.K., Rance, M., and Palmer III, A.G. (1998) Longitudinal and transverse ^{1}H-^{15}N dipolar/^{15}N chemical shift anisotropy relaxation interference: Unambiguous determination of rotational diffusion tensors and chemical exchange effects in biological macromolecules. *Journal of the American Chemical Society* 120, 7905–7915.

23. Tollinger, M., Skrynnikov, N.R., Mulder, F.A.A., Forman-Kay, J.D., and Kay, L.E. (2001) Slow dynamics in folded and unfolded states of an SH3 domain. *Journal of the American Chemical Society* 123, 11341–11352.

24. Xu, X.P. and Case, D.A. (2001) Automated prediction of ^{15}N, $^{13}C_{\alpha}$, $^{13}C_{\beta}$ and ^{13}C` chemical shifts in proteins using a density functional database. *Journal of Biomolecular NMR* 21, 321–333.

25. Xu, X.P. and Case, D.A. (2002) Probing multiple effects on ^{15}N, $^{13}C_{\alpha}$, $^{13}C_{\beta}$ and ^{13}C` chemical shifts in peptides using a density functional theory. *Biopolymers* 65, 408–423.

26. Moon, S. and Case, D.A. (2007) A new model for chemical shifts of amide hydrogens in proteins. *Journal of Biomolecular NMR* 38, 139–150.

27. Smillie, L. and Nattriss, M. (1991) *High-performance liquid chromatography of peptides and proteins: Separation, analysis, and conformation*. (CRC Press).

28. Li, M.X., Spyracopoulos, L., and Sykes, B.D. (1999) Binding of cardiac troponin-$I_{147-163}$ induces a structural opening in human cardiac troponin-C. *Biochemistry* 38, 8289–8298.

29. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, AD. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* 6, 277–293.

30. Goddard, T.D. and Kneller, D.G. (2006). *Sparky* University of California, San Francisco.

31. Loria, J.P., Rance, M., and Palmer III, A.G. (1999) A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by nmr spectroscopy. *Journal of the American Chemical Society* 121, 2331–2332.

32. Mulder, F.A.A., Skrynnikov, N.R., Hon, B., Dahlquist, F.W., and Kay, L.E. (2001) Measurement of slow (μs-ms) time scale dynamics in protein side chains by ^{15}N relaxation dispersion NMR spectroscopy: Application to Asn and Gln residues in a cavity mutant of T4 lysozyme. *Journal of the American Chemical Society* 123, 967–975.

33. Kovrigin, E.L., Kempf, J.G., Grey, M.J., and Loria, J.P. (2006) Faithful estimation of dynamics parameters from CPMG relaxation dispersion measurements. *Journal of Magnetic Resonance* 180, 93–104.

34. Palmer, A.G., Kroenke, C.D., and Loria, J.P. (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods in Enzymology* 339, 204–238.

35. Mandel, A.M., Akke, M., and Palmer, A.G. (1995) Backbone dynamics of *Escherichia coli* ribonuclease HI: Correlations with structure and function in an active enzyme. *Journal of Molecular Biology* 246, 144–163.

36. Spyracopoulos, L. (2006) A suite of Mathematica notebooks for the analysis of protein main chain ^{15}N NMR relaxation data. *Journal of Biomolecular NMR* 36, 215–224.

37. d'Auvergne, E.J. and Gooley, P.R. (2003) The use of model selection in the model-free analysis of protein dynamics. *Journal of Biomolecular NMR* 25, 25–39.

38. Horn, R. (1987) Statistical methods for model discrimination. Applications to gating kinetics and permeation of the acetylcholine receptor channel. *Biophysical Journal* 51, 255–263.

39. Carver, JP and Richards, RE. (1972) A general two-site solution for the chemical exchange produced dependence of T_2 upon the Carr-Purcell pulse separation. *Journal of Magnetic Resonance (1969)* 6, 89–105.

40. Markley, J.L., Bax, A., Arata, Y., Hilbers, CW, Kaptein, R., Sykes, B.D., Wright, P.E., and Wüthrich, K. (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. *Journal of Molecular Biology* 280, 933–952.

41. Murphy, J.M., Korzhnev, D.M., Ceccarelli, D.F., Briant, D.J., Zarrine-Afsar, A., Sicheri, F., Kay, L.E., and Pawson, T. (2007) Conformational instability of the MARK3 UBA domain compromises ubiquitin recognition and promotes interaction with the adjacent kinase domain. *Proceedings of the National Academy of Sciences* 104, 14336–14341.

42. Das, R., Abu-Abed, M., and Melacini, G. (2006) Mapping allostery through equilibrium perturbation NMR spectroscopy. *Journal of the American Chemical Society* 128, 8406–8407.

43. Sugase, K., Dyson, H.J., and Wright, P.E. (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447, 1021–1025.

44. Hill, R.B., Bracken, C., DeGrado, W.F., and Palmer III, A.G. (2000) Molecular motions and protein folding: characterization of the backbone dynamics and folding equilibrium of α2D using ^{13}C NMR spin relaxation. *Journal of the American Chemical Society* 122, 11610–11619.

45. Jee, J.G., Ishima, R., and Gronenborn, A.M. (2008) Characterization of specific protein association by ^{15}N CPMG relaxation dispersion NMR: the GB1^{A34F} monomer-dimer equilibrium. *The Journal of Physical Chemistry B* 112, 6008–6012.

46. Spyracopoulos, L., Lavigne, P., Crump, M.P., Gagné, S.M., Kay, C.M., and Sykes, B.D. (2001) Temperature dependence of dynamics and thermodynamics of the regulatory domain of human cardiac troponin C. *Biochemistry* 40, 12541–12551.

47. Hass, M.A.S., Hansen, D.F., Christensen, H.E.M., Led, J.J., and Kay, L.E. (2008) Characterization of conformational exchange of a histidine side chain: Protonation, rotamerization, and tautomerization of His61 in plastocyanin from *Anabaena variabilis*. *Journal of the American Chemical Society* 130, 8460–8470.

48. Roehm, P.C. and Berg, J.M. (1997) Sequential metal binding by the RING finger domain of BRCA1. *Biochemistry* 36, 10240–10245.

49. Houben, K., Wasielewski, E., Dominguez, C., Kellenberger, E., Atkinson, R.A., Timmers, H., Kieffer, B., and Boelens, R. (2005) Dynamics and metal exchange properties of C4C4 RING domains from CNOT4 and the p44 subunit of TFIIH. *Journal of Molecular Biology* 349, 621–637.

50. Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637.

51. Millet, O., Loria, J.P., Kroenke, C.D., Pons, M., and Palmer III, A.G. (2000) The static magnetic field dependence of chemical exchange linebroadening defines the

NMR chemical shift time scale. *Journal of the American Chemical Society* 122, 2867–2877.

52. Tjandra, N., Feller, S.E., Pastor, R.W., and Bax, A. (1995) Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation. *Journal of the American Chemical Society* 117, 12562–12566.

53. Kentsis, A., Gordon, R.E., and Borden, K.L.B. (2002) Control of biochemical reactions through supramolecular RING domain self-assembly. *Proceedings of the National Academy of Sciences* 99, 15404–15409.

54. Pullen, S.S., Labadia, M.E., Ingraham, R.H., McWhirter, S.M., Everdeen, D.S., Alber, T., Crute, J.J., and Kehry, M.R. (1999) High-affinity interactions of tumor necrosis factor receptor-associated factors (TRAFs) and CD40 require TRAF trimerization and CD40 multimerization. *Biochemistry* 38, 10168–10177.

55. Wu, H. and Arron, J.R. (2003) TRAF6, a molecular bridge spanning adaptive immunity, innate immunity and osteoimmunology. *Bioessays* 25, 1096–1105.

56. Heldin, C.H. (1995) Dimerization of cell surface receptors in signal transduction. *Cell* 80, 213–223.

57. Xu, Z., Devlin, K.I., Ford, M.G., Nix, J.C., Qin, J., and Misra, S. (2006) Structure and interactions of the helical and U-box domains of CHIP, the C terminus of HSP70 interacting protein. *Biochemistry* 45, 4749–4759.

58. Kostic, M., Matt, T., Martinez-Yamout, M.A., Dyson, H.J., and Wright, P.E. (2006) Solution structure of the Hdm2 C2H2C4 RING, a domain critical for ubiquitination of p53. *Journal of Molecular Biology* 363, 433–450.

59. Bellon, S.F., Rodgers, K.K., Schatz, D.G., Coleman, J.E., and Steitz, T.A. (1997) Crystal structure of the RAG1 dimerization domain reveals multiple zinc-binding motifs including a novel zinc binuclear cluster. *Nature Structural & Molecular Biology* 4, 586–591.

60. Baron, R. and McCammon, J.A. (2008) (Thermo)dynamic role of receptor flexibility, entropy, and motional correlation in protein-ligand binding. *ChemPhysChem* 9, 983–988.

61. Perazzolo, C., Verde, M., Homans, S.W., and Bodenhausen, G. (2007) Evidence of chemical exchange in recombinant major urinary protein and quenching thereof upon pheromone binding. *Journal of Biomolecular NMR* 38, 3–9.

Chapter 3

Mechanism for recognition of polyubiquitin chains: balancing affinity through interplay between multivalent binding and dynamics*

Introduction

The covalent attachment of ubiquitin (Ub) to target proteins is crucial to the regulation of life processes such as protein degradation, the cell cycle, DNA repair, the DNA damage response, and trafficking (1). Ultimately, manifestation of the Ub signal on target

*This chapter has been published. Reproduced with permission from: C.J. Markin, W. Xiao, and L. Spyracopoulos. Mechanism for recognition of polyubiquitin chains: balancing affinity through interplay between multivalent binding and dynamics. *Journal of the American Chemical Society*, 132(32):11247-11258, 2010. Copyright 2010 American Chemical Society.

Contributions: W.X. designed and provided the clones of the tandem Ub_n chains. C.J.M. and L.S. designed experiments. C.J.M. made the construct of RAP80-tUIM, performed the protein purifications and NMR titrations. C.J.M. developed the multivalent binding models and analyzed the titration data together with L.S. L.S. carried out the molecular dynamics simulations and carried out the iRED analysis. L.S. and C.J.M. wrote the paper.

proteins may be attachment of a single or multiple Ub molecules to different sites or polyubiquitination through a single type or varied intra-Ub peptide linkages. These various forms of the Ub signal provide a topological richness that can be exploited to achieve specificity in a wide variety of signaling cascades, as typified by the diversity in Ub recognition by domains such as UIM, UEV, CUE, and LUBAC (2). Some of these domains consist of well-structured cores, whereas others, such as the ubiquitin-interacting motif (UIM), consist of single or tandem α-helices. Invariably, recognition of the mono-Ub signal by a single Ub binding domain occurs with weak affinity, having a typical dissociation constant exceeding 100 µM (3). While there are numerous examples that the basic monoUb interaction is intensified by themes such as tandem Ub recognition motifs to achieve biological relevance, it remains a difficult and intricate problem to develop a detailed molecular mechanism to describe the process that leads to amplification of the Ub signal.

Recently, K63-linked polyUb chain recognition has been implicated in the DNA damage response (DDR) through its involvement in the repair of double-stranded DNA breaks (DSB) (4–6). The DDR involves a cascade of altered proteins and protein-protein interactions in response to damaged DNA (7); the process has been likened to signal transduction networks (8). The protein RAP80 has been shown to play a key role in the DDR by recruiting proteins critical for repair to DNA damage foci via recognition of polyUb chains through tandem ubiquitin-interacting motifs (tUIMs) at its N-terminus (4–6). A breakdown in protein targeting to DNA damage foci can lead to tumorigenesis (6). For example, BRCA1 (breast cancer 1, early onset protein) is a tumor suppressor that is mutated in about half of all hereditary early-onset breast and ovarian cancer patients (5), and its recruitment to sites of DNA damage is critical for repair.

RAP80 is a $\sim$ 80 kDa multidomain protein consisting of two putative zinc fingers and two tandem ubiquitin interacting motifs (UIMs). Canonical UIMs consist of a single short α-helix that interacts with the I44 hydrophobic patch on Ub with low affinity (K_{D} 0.1-2 mM). The canonical α-helical UIM binds Ub in a specific fashion, such that the N-terminus of the UIM is adjacent to the C-terminus of Ub and the C-terminus of the UIM is adjacent to the N-terminus (or K63) of Ub (9). Recently, the interaction of RAP80-tUIM with K63-linked Ub_2 and Ub_4 chains was characterized by isothermal titration

calorimetry and fluorescence anisotropy measurements (10), and the structure of the complex was determined by crystallography (11). It was shown that binding of Ub_2 to the tandem RAP80 UIMs is enhanced compared to binding of monoUb, and similarly, binding of Ub_4 is enhanced in comparison to Ub_2. In addition, the length of the linker between the tandem UIMs of RAP80 was shown to influence the affinity of the Ub_2-RAP80-tUIM interaction. Herein, we have used solution-state NMR spectroscopy in combination with molecular dynamics simulations to demonstrate that recognition of extended, tandemly linked polyUb molecules by the tandem UIM domains of RAP80 occurs through interplay between multivalent binding of sequential UIMs to sequential Ub moieties and dynamics between the individual domains. Tandem, or linear, polyUb chains are composed of Ub domains linked sequentially through their respective N- and C-termini. These polyUb chains serve biological roles and, given their close structural and dynamic similarity to K63-linked chains, can also serve as models for understanding recognition processes for K63-linked chains. Importantly, we have developed a basic multivalent binding model for the interaction of RAP80-tUIM with tandem polyUb chains that accounts for the Ub chain length dependence of the affinity of the interaction. Furthermore, we observe fast kinetics for the interaction of tandem Ub_2 with RAP80-tUIM that may be physiologically relevant for maintaining transient protein complexes at DNA damage foci. We also show that conformational selection plays an important role in UIM-polyUb recognition through analysis of the temperature dependence of main-chain $^{13}C_\alpha$ and $^{1}H_\alpha$ chemical shifts and ^{15}N NMR relaxation measurements of the main-chain dynamics for RAP80-tUIM.

Materials and methods

Protein expression and purification

All proteins were cloned and expressed as pGEX6P-1 glutathione S-transferase (GST) fusion constructs in Escherichia coli BL21DE3 cells. Protein sequences are given in the Supporting Information. For expression of the isotopically labeled RAP80-tUIM and polyUb proteins, cells were grown as previously described (12). For unlabeled proteins, the methodology was as follows: Electrocompetent cells were transformed and plated on agar plates containing ampicillin and chloramphenicol overnight. A single colony

was used to inoculate starter cultures of 50 mL of Luria broth (LB) with antibiotics added and allowed to grow overnight. Four flasks of 500 mL of LB were then inoculated with 5 mL of starter culture, grown to OD_{600} of 0.6-0.8, and induced with 1.6 mL of 125 mM isopropyl thiogalactoside (IPTG). Expression was allowed to proceed overnight at 18 ℃ for RAP80-tUIM or at 25 ℃ for tandem polyUb chains. Purification methods for the various proteins used in this study were similar except where noted. After harvesting and lysing of cells, the lysate was passed over a GST affinity column. The ubiquitin constructs were cleaved by use of PreScission protease enzyme in 50 mM tris(hydroxymethyl)aminomethane (Tris) and 200 mM NaCl buffer, pH 7.0. GST-RAP80-tUIM was cleaved in 50 mM Tris and 150 mM NaCl buffer, pH 7.5. Following cleavage, protein solutions were passed over a GST affinity column, followed by size-exclusion chromatography to purify RAP80-tUIM and Ub (Superdex 30) and Ub_2 and Ub_4 (Superdex 75). For Ub_3, ion-exchange purification via a Q-Sepharose column was employed, followed by size-exclusion chromatography on a Superdex 75 column.

NMR spectroscopy

The majority of NMR spectra were collected at 25 ℃ and 600 MHz, with the exception of the titration of [U-^{15}N, U-^{2}H]-RAP80-tUIM with unlabeled Ub_4, which was carried out at 800 MHz. NMR titrations of unlabeled monoUb, tandem Ub_2, tandem Ub_3, and tandem Ub_4 into [U-^{15}N]-RAP80-tUIM (Ub_3) or [U-^{15}N, U-^{2}H]-RAP80-tUIM (Ub, Ub_2, Ub_4) were conducted by following chemical shift changes for backbone amide $^{1}H^{N}$ and ^{15}N resonances in 2D ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) or transverse relaxation-optimized spectroscopy (TROSY) HSQC NMR spectra. In order to ensure consistency of the results, corollary titrations of [U-^{15}N]-Ub_2, [U-^{15}N]-Ub_3, and [U-^{15}N, U-^{2}H]-Ub_4 with unlabeled RAP80-tUIM were conducted in a similar manner. For the 2D ^{1}H-^{15}N HSQC NMR-based titrations, samples initially contained labeled protein in a 95:5 H2O/D2O mixture, containing 50 mM Tris, 150 mM NaCl, and 1 mM dithiothreitol (DTT) (pH 7.3). To this mixture, a stock solution of unlabeled binding partner in the same buffer conditions was added to follow chemical shift changes. For relaxation experiments, 350 µL samples of [U-^{15}N]/[U-^{13}C,U-^{15}N]-labeled RAP80-tUIM were placed in 5 mm Shigemi tubes in either 95:5 H_2O/D_2O or 90:10 H_2O/D_2O containing 50 mM Tris,

150 mM NaCl, 1 mM DTT, and 0.5 µL of 10% NaN_3. [U-^{13}C, U-^{15}N]-labeled samples used for chemical shift assignment were prepared similarly, with the inclusion of 5 µL of 10 mM DSS as a chemical shift reference. Chemical shift assignments were accomplished by standard ^{1}H/^{13}C/^{15}N protein NMR spectroscopy (13).

1:1 Binding isotherms for RAP80-tUIM interactions with tandem Ub_2, Ub_3, and Ub_4

As a first step in analyzing the binding of RAP80-tUIM to polyUb, macroscopic K_D values were determined for each individual titration of unlabeled tandem Ub_2, Ub_3, and Ub_4 into [U-^{15}N]- or [U-^{15}N, U-^{2}H]-RAP80-tUIM by globally fitting titration data for all residues observed to shift significantly with a basic binding isotherm describing 1:1 protein-ligand binding.

Binding isotherm for the interaction of monoUb with RAP80-tUIM

The dissocation constant for the RAP80-tUIM-monoUb interaction ($K_{D,mono}$) was determined experimentally as follows: $K_{D,mono}$ values for both the N- and C-terminal UIMs were extracted from a global fit of all [U-^{15}N, U-^{2}H]-RAP80-tUIM residues with significant chemical shift changes upon addition of monoUb, and that could be followed unambiguously throughout the titration. Resonances from residues belonging to either UIM can be monitored separately by 2D ^{1}H-^{15}N HSQC NMR spectra, allowing for determination of individual $K_{D,mono}$ values for both UIMs. This was accomplished by simultaneously fitting all resonances from the N-terminal UIM with significant chemical shift changes to one $K_{D,mono}$ value, and the C-terminal resonances to a different $K_{D,mono}$. The equation representing the fractional chemical shift changes for a given UIM in RAP80 was modified to account for the binding of Ub to the other UIM, allowing for simultaneous monitoring of binding to both UIMs.

Thermodynamic models for the interaction between RAP80-tUIM and polyUb chains

In order to analyze NMR-based titrations of polyUb chains with RAP80-tUIM, we designed models to take into account both the basic multivalent interaction between tandem UIM units and tandem diUb units, and realistic interactions between individual UIM and individual Ub moieties within chains, as well as states resulting from mixtures of both (Supporting Information). Equilibria for bound states are expressed in terms of a multivalent $K_{\mathrm{D,mv}}$ (the interaction between RAP80-tUIM and two adjacent Ub molecules) and/or the $K_{\mathrm{D,mono}}$ for monoUb binding. Expressions for the various equilibria for monoUb, as well as tandem Ub_2, Ub_3, and Ub_4 binding to RAP80-tUIM, are given in the Supporting Information. The basic multivalent binding model involves the interaction between the two UIM domains from RAP80, and the two Ub moieties from Ub_2. A schematic diagram for this model, and associated binding curves, are shown in figure 3.1.

$K_{\mathrm{D,mv}}$ values for RAP80-tUIM binding to Ub_2, Ub_3, and Ub_4 were determined by globally fitting backbone ^{15}N and ^{1}H chemical shift changes for all residues of RAP80-tUIM that could be followed unambiguously throughout the 2D ^{1}H-^{15}N HSQC or TROSY HSQC titrations by use of an in-house simulated annealing algorithm. The bound fraction is related to the NMR chemical shift changes through $\Delta\delta_{\mathrm{obs}} = f_{\mathrm{b}}\Delta\delta_{\mathrm{sat}}$, where f_{b} is the theoretical bound fraction of [U-^{15}N, U-^{2}H]- or [U-^{15}N]-labeled protein (RAP80-tUIM or polyUb) at given RAP80-tUIM and polyUb concentrations, $\Delta\delta_{\mathrm{sat}}$ is the maximum chemical shift change (corresponding to the saturated state), and $\Delta\delta_{\mathrm{obs}}$ is the chemical shift change at given values for $K_{\mathrm{D,mv}}$ and RAP80-tUIM and polyUb concentrations. $K_{\mathrm{D,mv}}$ was optimized as follows: Theoretical $\Delta\delta_{\mathrm{obs}}$ values for [U-^{15}N, U-^{2}H]- or [U-^{15}N]-labeled protein (RAP80-tUIM or polyUb) at concentrations used in the titrations were generated by initially setting the value of $K_{\mathrm{D,mv}}$, and subsequently numerically solving the coupled equilibrium expressions given in the Supporting Information. The procedure was repeated by parametrically varying the value of $K_{\mathrm{D,mv}}$ with a simulated annealing algorithm to minimize the sum of the squared differences between the experimental and theoretical chemical shift changes.

In addition to $K_{\mathrm{D,mv}}$, the bound chemical shifts were treated as adjustable parameters

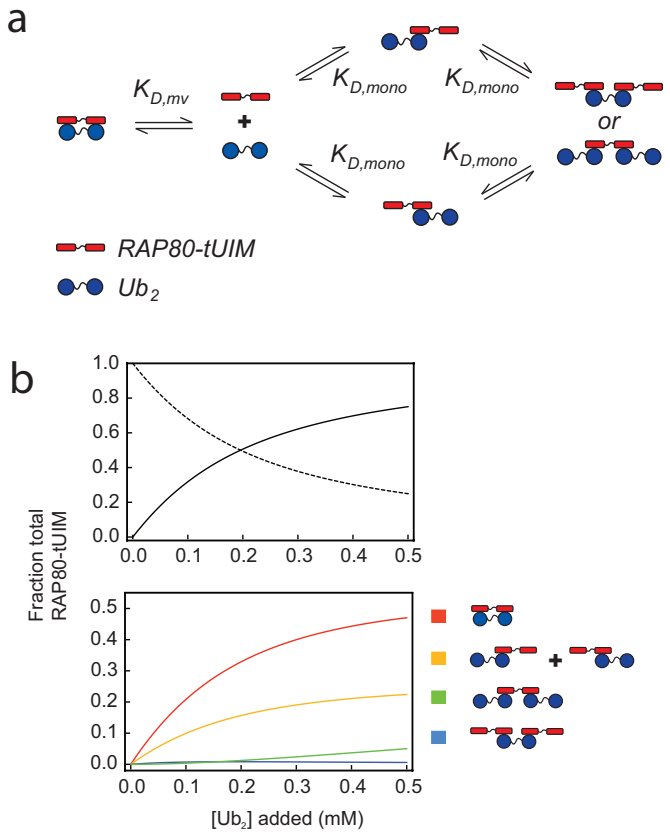

Figure 3.1: a) Equilibria for the interaction between RAP80-tUIM and Ub_2. $K_{\mathrm{D,mv}}$ (223 µM) denotes the dissociation constant for the multivalent interaction between both UIMs and both Ub moieties of Ub_2, whereas $K_{\mathrm{D,mono}}$ (936 µM) is the dissociation constant for the interaction between a single UIM of RAP80 and a single Ub from Ub_2. (b) Simulated curves illustrating the buildup of various UIM-Ub_2 species as Ub_2 is added to 0.1 mM RAP80-tUIM. Curves show the fraction of total UIM in a particular state as a function of [Ub_2]. For the upper plot, the solid curve shows the sum of all RAP80-tUIM-Ub_2 states and the dashed curve shows the fraction of free RAP80-tUIM.

during the fits, given that the combination of low-affinity binding and finite protein concentrations made saturation difficult to achieve. In the case of the [U-^{15}N, U-^{2}H]- or [U-^{15}N]-labeled RAP80-tUIM titrations with polyUb, starting values and constraints for these parameters could be selected on the basis of the final point (highest concentration of polyUb), given that they appear close to saturation. However, for the corollary titrations of the various [U-^{15}N, U-^{2}H]- or [U-^{15}N]-labeled polyUb chains, starting values and constraints were obtained from the final point of [U-^{15}N] Ub_2 titrated with RAP80-tUIM, the titration closest to saturation (72:1 UIM/Ub_2), which consequently likely represents the most accurate estimates of the bound Ub_2 chemical shifts. This approach has the implicit assumption that the bound chemical shifts of the various polyUb chains are the same upon saturation with RAP80-tUIM. This is reasonable, given that the direction and magnitude of chemical shift changes observed during the titrations are similar.

Errors in the optimized parameters were estimated by Monte Carlo trials of the global fits, based on the error in protein concentrations, determined by the bicinchoninic acid (BCA) assay, estimated to be 10%. The accuracy of the BCA assay with respect to determining RAP80-tUIM concentration was confirmed by amino acid analysis, the difference in concentration between the two measurement techniques being less than 2%.

Line-shape analysis for RAP80-tUIM interaction with Ub_2

An analytical expression for the frequency domain NMR spectrum in the presence of two-site exchange was derived from the Bloch-McConnell equations (14), modified to account for the application of a cosine-squared window function. NMR spectra were corrected for concentration differences and number of acquired transients and were numerically fit to the expression for the frequency domain spectrum by optimizing the values of k_{off}, the bound chemical shift, and a global intensity parameter.

Temperature dependence of $^{13}C_\alpha$ and $^{1}H_\alpha$ chemical shifts of RAP80-tUIM

The temperature dependence of the main-chain $^{13}C_\alpha$ and $^{1}H_\alpha$ chemical shifts from 5 to 50 ℃ were used to characterize the helix-coil transition for RAP80-tUIM (15). Briefly,

the observed chemical shift changes as a function of temperature were fit to a two-state cooperative helix-coil transition (eq 9 in (16)) using as fitted parameters the $^{13}C_{\alpha}$ or $^{1}H_{\alpha}$ chemical shifts for the α-helical and random coil states, as well as the midpoint and width of transition.

Main-chain ^{15}N relaxation measurements for RAP80-tUIM

^{15}N-R_1, R_2, and ^{1}H-^{15}N nuclear Overhauser effect (NOE) relaxation data were acquired at 5 and 25 ℃ and two field strengths (600 and 800 MHz) as previously described (17). Additionally, changes in ^{15}N-R_2 (600 MHz and 25 ℃) for free RAP80-tUIM upon binding Ub_2 were also determined.

Contributions from chemical exchange (R_{ex}) to ^{15}N-R_2 were estimated by use of ^{15}N-R_1, R_2, and ^{1}H-^{15}N NOE relaxation data combined with η_{xy} and η_{z} cross-correlated relaxation rates (18). Additionally, a series of main-chain amide ^{1}H-^{15}N single and multiple quantum coherence and two-spin order relaxation rates were used to estimate the rate of chemical exchange for the main-chain amides (19, 20).

Molecular dynamics simulation and iRED analysis for RAP80-tUIM

Calculations were conducted with the AMBER 10 simulation package (21), with the ff99SB force field and the TIP3P water model. Covalent bonds to hydrogen were restrained by use of SHAKE, temperature was regulated by Langevin dynamics with a collision frequency of 1 ps-1, and the cutoff was 8 Å for pairwise nonbonded and electrostatic interactions. A PME approach with default parameters was used for calculating long-range electrostatics. The initial structural model was solvated in a truncated octahedral water box with a minimum of 24 Å between a given protein atom and a given atom of an image in an adjacent unit cell. Six Na^{+} ions were included to ensure the system was neutral. The system was heated gradually over 50 ps to 298 K with 2 kcal/mol restraints on solute atoms and equilibrated to 1 atm pressure for an additional 50 ps. Production dynamics were then conducted for 8 ns. The isotropic reorientational eigenmode (iRED) approach was used to extract correlation functions for the main chain amide ^{1}H-$^{15}N^{H}$

vectors from the molecular dynamics simulation (22). The correlation functions were fit to two-parameter monoexponential decays to extract correlation times; these were subsequently used to calculate ^{15}N-R_1, R_2, and ^{1}H-^{15}N NOE values at 600 and 800 MHz.

Molecular dynamics simulation for tandem Ub_3

Molecular dynamics were carried out with the AMBER 10 biomolecular simulation software suite. The ff99SB force field was employed with the "GBn" generalized Born solvation model with a salt concentration of 150 mM, covalent bonds to hydrogen were restrained by use of SHAKE, and there were no cutoffs for nonbonded interactions and a 25 Å cutoff for pairwise summation of effective Born radii. Temperature was regulated by Langevin dynamics with a collision frequency of 50 ps^{-1}. The system was heated from 0 to 300 K over 50 ps with a time step of 2 fs. Production dynamics were run for 80.5 ns. Cross-correlation functions for vectors from the first two Ub moieties in Ub_3, either perpendicular or parallel to the long axis of the extended Ub_3 chain, were calculated and fit to monoexponential decays to determine the correlation times for reorientation of Ub moieties with respect to each other.

Results and discussion

Interaction of RAP80-tUIM with Ub, tandem Ub_2, tandem Ub_3, and tandem Ub_4

The interaction between RAP80-tUIM and various tandem polyUb chains was analyzed via NMR-based titrations for both binding partners. We titrated [U-^{15}N]- or [U-^{15}N, U-^{2}H]-RAP80-tUIM with unlabeled Ub, Ub_2, Ub_3, and Ub_4 chains (figure 3.2) and carried out corollary titrations of [U-^{15}N]-Ub_2, [U-^{15}N]-Ub_3, and [U-^{15}N, U-^{2}H]-Ub_4 with unlabeled RAP80-tUIM to ensure consistency. The maximum chemical shift changes observed in these titrations are shown in figure 3.3. For RAP80-tUIM, these changes occur within the α-helical UIM domains and predominantly at the canonical I44 hydrophobic binding site for polyUb.

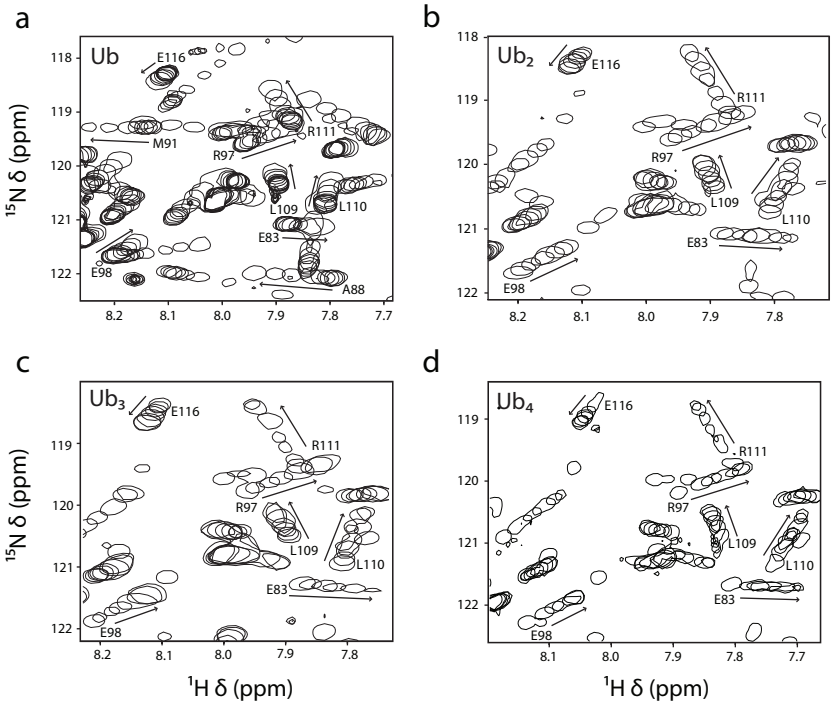

Figure 3.2: Representative regions from 2D ^{1}H-^{15}N HSQC NMR spectra of [*U*-^{15}N, *U*-^{2}H]-RAP80-tUIM or [*U*-^{15}N]-RAP80-tUIM titrated with (a) unlabeled Ub, (b) Ub_2, (c) Ub_3, and (d) Ub_4. Arrows indicate the direction of chemical shift changes upon polyUb addition for labeled RAP80-tUIM resonances.

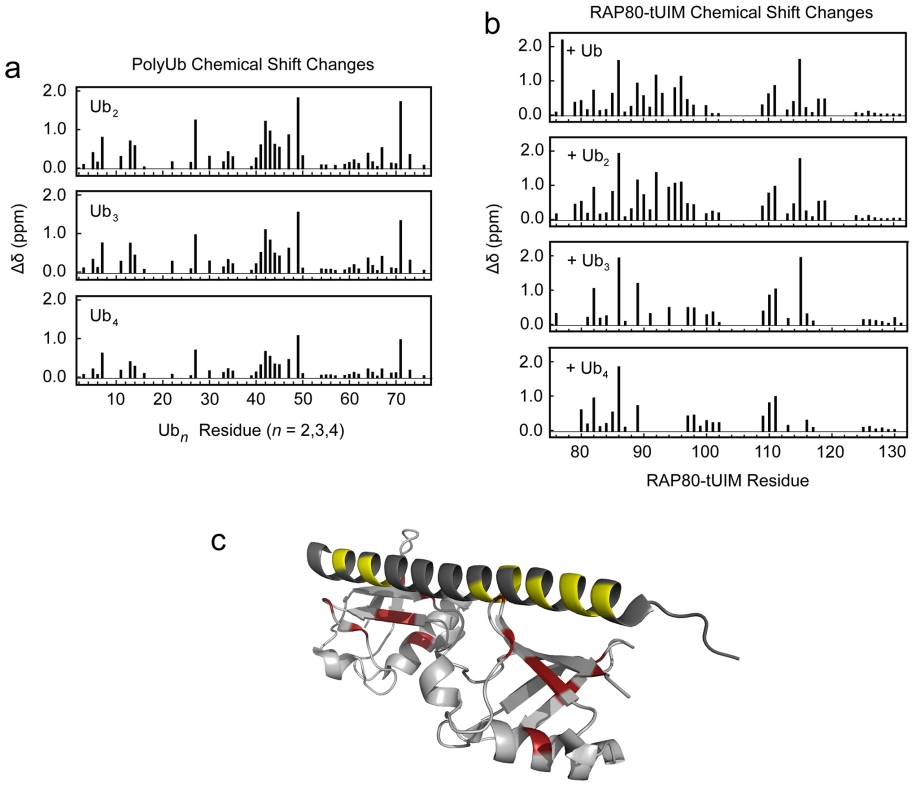

Figure 3.3: Maximum polyUb chemical shift changes, $\Delta\delta = [(\Delta\delta\ ^{15}\mathrm{N})^2 + (\Delta\delta\ ^{1}\mathrm{H})^2]^{1/2}$, observed in titrations of [U-^{15}N, U-^{2}H]- or [U-^{15}N]-labeled polyUb chains with unlabeled RAP80-tUIM (a), and RAP80-tUIM $\Delta\delta$ upon titration of [U-^{15}N, U-^{2}H]- or [U-^{15}N]-labeled RAP80-tUIM with unlabeled polyUb chains (b). Residues 103-108 are a stretch of contiguous Glu residues that are unassigned. Shown in panel c are chemical shift changes exceeding the median by 1 standard deviation ($\Delta\delta + 1\sigma$) for the RAP80-tUIM-Ub_2 titrations, mapped on the RAP80-tUIM-Ub_2 structure, with $\Delta\delta + 1\sigma$ values for RAP80 in yellow and those for Ub_2 in red. $\Delta\delta$ values for several RAP80-tUIM residues were not observed upon titration with Ub_3 and Ub_4 due to line broadening as a result of increased molecular weight of the complexes and chemical exchange. PolyUb and RAP80-tUIM $\Delta\delta$ values are observed to shift in a similar manner in the titrations. Differences in the overall magnitude of $\Delta\delta$ for titrations of longer polyUb chains with unlabeled RAP80-tUIM (a) are likely due to lower levels of saturation resulting from an increased number of binding sites on the longer chains (Supporting Information).

Table 3.1: Dissociation constants for 1:1 binding of RAP80-tUIM to Ub, tandem Ub chains, and K63-linked Ub chains

K_D (μM)[a]						
Ub	tUb_2[b]	K63-Ub_2[c]	tUb_3	K63-Ub_3	tUb_4	K63-Ub_4
960 ± 240[d] 520 ± 130[e]	75 ± 26	22 ± 1[e]	17 ± 7	nd[f]	16 ± 7	4.0 ± 0.2[e]

[a] K_D values were determined with the assumption of 1:1 binding.
[b] tUb_n indicates tandem Ub chains (n = 2, 3, 4) linked sequentially through respective N- and C-termini.
[c] K63-Ub_n indicates tandem Ub chains (n = 2, 3, 4) linked sequentially through respective K63 residues and C-termini.
[d] Average $K_{D,mono}$ for UIM1 (930 ± 162 μM) and UIM2 (942 ± 177 μM), determined in this study by NMR.
[e] Determined by fluorescence anisotrophy (10)
[f] Not determined

The NMR-derived binding curves were initially analyzed with the assumption of 1:1 binding to extract macroscopic dissociation constants, followed by more detailed analyses wherein we assumed that the RAP80-tUIM interaction with polyUb involves two equilibrium constants: $K_{D,mono}$, the dissociation constant for the interaction between a given UIM from RAP80 and monoUb, and $K_{D,mv}$, the multivalent dissociation constant for the simultaneous interaction between both RAP80 tandem UIMs and two neighbouring intramolecular Ub moieties within a polyUb chain (figure 3.1). These detailed analyses of the titrations indicate that the increased macroscopic 1:1 binding affinity observed as a function of increasing chain length can be explained by an increased number of multivalent binding sites in the longer polyUb chains, as well as an increased number of low-affinity interactions between single UIMs from RAP80 with individual Ub moieties in the chain.

Simple analyses of RAP80-tUIM-polyUb binding

NMR-based titrations of unlabeled monoUb into [U-^{15}N, U-^{2}H] RAP80-tUIM indicate that the tandem RAP80 UIMs bind monoUb weakly, with similar dissociation constants ($K_{D,mono}$) for the N- and C-terminal UIMs (figure 3.4 and Table 3.1). The similar $K_{D,mono}$ values for Ub binding to either the N- or C-terminal UIM are reflected in the lack of an initial lag phase in chemical shift titration curves for residues located in either helix.

We obtained an initial view of affinity changes with increasing polyUb chain length by

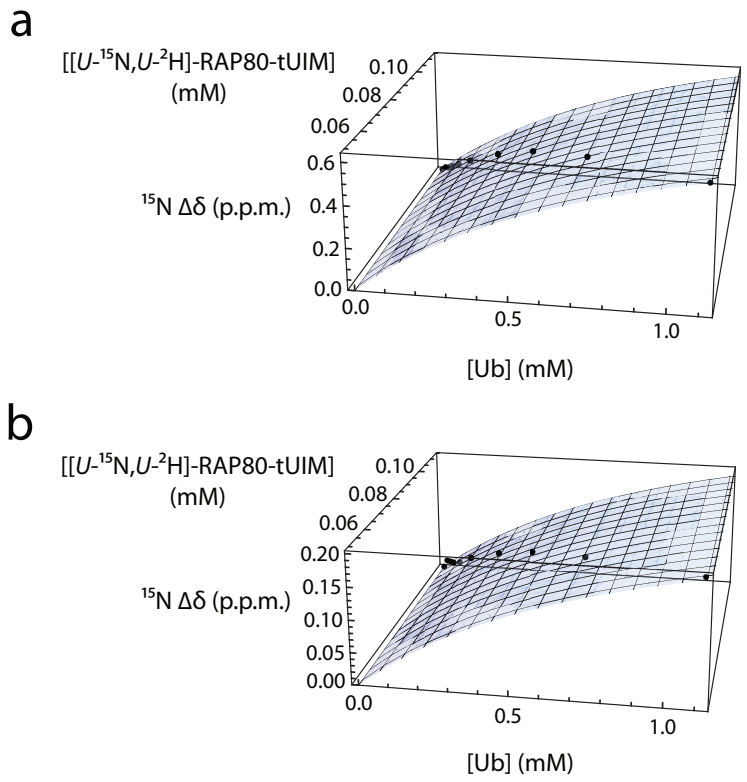

Figure 3.4: Titration of [U-^{15}N, U-^{2}H]-RAP80-tUIM with unlabeled Ub. (a) ^{15}N chemical shift changes for F85 within the N-terminal UIM are indicated on the vertical axis, and concentrations of analyte or titrand (labeled RAP80-tUIM) and unlabeled Ub titrant are indicated on the horizontal axes. (b) ^{15}N chemical shift changes for E116, within the C-terminal UIM. Experimentally determined chemical shift changes are shown as points, and the best fits to associated binding isotherms are shown as surfaces (Materials and Methods and Supporting Information)

analyzing the NMR-based titrations, assuming 1:1 binding. Using this first approach, the RAP80-tUIMs bind tandem Ub_2, Ub_3, and Ub_4 with increasing affinity upon increasing chain length, having macroscopic dissociation constants that decrease 12, 55, and 59-fold, respectively, in comparison to monoUb binding (Table 3.1). These macroscopic dissociation constants indicate similar multivalent effects for Lys63-linked (10) and tandem Ub chains, with a (24 ± 6)-fold increase in affinity for Lys63-linked polyUb chains compared to a (13 ± 5)-fold increase for tandem polyUb chains, for Ub_2 binding in comparison to monoUb binding. Increases in affinity with increasing chain length are the same for the two types of chains within error, with (5.5 ± 0.4)- and (5 ± 2)-fold increases for Ub_4 in comparison to Ub_2 binding, for Lys63-linked and tandem chains, respectively. These results suggest that binding interactions for tandem and K63-linked Ub chains are similar for RAP80-tUIM, as observed for polyUb binding to NEMO (23), and that tandem polyUb chains serve as reasonable models to understand the molecular basis of polyUb chain recognition by tandem UIMs.

The RAP80-tUIM-tandem Ub_2 interaction is specific and multivalent

The binding of RAP80-tUIM to Ub_2 encompasses the fundamental molecular basis, or unit, for recognition of polyUb chains by RAP80-tUIM (figures 3.1 and 3.3c). The titration of [*U*-^{15}N]-tandem Ub_2 with unlabeled RAP80-tUIM indicates that the UIMs of RAP80 bind in a multivalent manner to Ub_2. That is, the individual RAP80 UIMs bind the individual Ub moieties in tandem Ub_2 in a specific orientation. For example, the ^{1}H-^{15}N HSQC NMR spectrum of Ub_2 alone closely resembles that of monoUb. However, as RAP80-tUIM is added, a number of the Ub_2 resonances that respond to UIM binding diverge into two peaks, both of which are in fast exchange with unbound Ub_2 (figure 3.5). This is likely due to sequence differences between the N- and C-terminal UIM domains that give rise to different bound chemical shifts for labeled Ub_2. Importantly, this indicates that the individual Ub moieties of Ub_2 favor one of the tandem UIMs, which can occur only if non-promiscuous, multivalent binding is present. These NMR results are consistent with the specific orientation observed for RAP80-tUIM bound to K63-linked

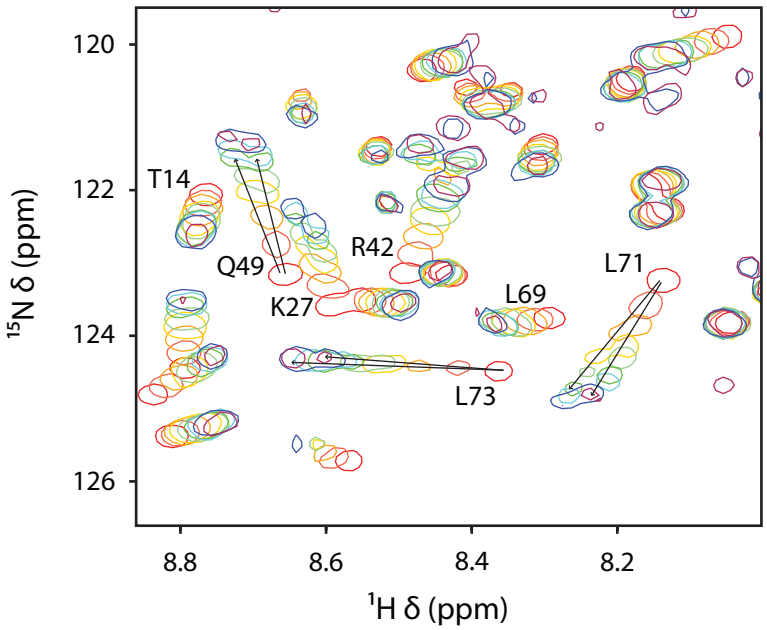

Figure 3.5: Region from the 2D ^{1}H-^{15}N HSQC NMR spectrum of [U-^{15}N]-Ub$_2$ upon titration with unlabeled RAP80-tUIM. Splitting of monoUb resonances into two distinct peaks for various residues upon titration with unlabeled RAP80-tUIM are highlighted by black arrows.

Ub$_2$ (11), which indicates that multivalent binding occurs with the N-terminal Ub bound to the C-terminal UIM and the C-terminal Ub simultaneously bound to the N-terminal UIM. The observed peak splitting also demonstrates that tandem polyUb chains are reasonable mimics for K63-linked chains, with regard to their ability to facilitate multivalent binding of UIMs, a fact that is also borne out by the similar structures for K63-linked and tandem diUb (24).

Multivalent models for analyses of binding between RAP80-tUIM and polyUb

The binding of RAP80-tUIM to polyUb chains was modeled by constructing equilibria for physically reasonable bound states for the various polyUb chains and assuming that the binding process is described by two dissociation constants: binding of a single Ub to a single UIM domain ($K_{\mathrm{D,mono}}$) and multivalent binding of tandem UIMs to sequential diUb units within polyUb chains ($K_{\mathrm{D,mv}}$) (figure 3.1 and Supporting Information). These models assume that binding of both UIMs from a RAP80-tUIM molecule to sequential

diUb units in a polyUb chain does not alter binding of either a single UIM from a RAP80-tUIM molecule to the adjacent free Ub sites or a second RAP80-tUIM unit to adjacent free sequential diUb units in polyUb chains. This assumption is reasonable given the flexibility between UIMs in RAP80-tUIM and between Ub moieties in tandem polyUb, which is discussed in detail in subsequent sections. Additionally, in order for binding of RAP80-tUIM to a sequential diUb unit in a polyUb chain to alter binding of other RAP80-tUIM molecules to immediately adjacent Ub moieties, one would expect a structural interaction between the long C-terminal tail of a bound RAP80-tUIM and the adjacent Ub moieties. However, no significant chemical shift changes for the C-terminal region of RAP80-tUIM are observed upon interaction with longer tandem Ub_3 and Ub_4 chains (figure 3.3b).

$K_{D,mono}$ was experimentally determined, whereas $K_{D,mv}$ was determined by global fitting of NMR-based titrations of RAP80-tUIM with tandem Ub_2, Ub_3, and Ub_4. It should be noted that the common assumption that multivalent association constants are given by the product of individual constants is strictly not legitimate (25, 26). Indeed, affinity increases due to multivalent binding can be larger than expected, given an intrinsically low-affinity individual interaction (27).

The values of $K_{D,mv}$ for multivalent polyUb binding, globally fit by use of NMR titrations of RAP80-tUIM with tandem Ub_2, Ub_3, and Ub_4 and the corollary titrations of RAP80-tUIM into tandem Ub_2, Ub_3, and Ub_4, are shown in figure 3.6 and Table 3.2 and are in good agreement within experimental error. The average $K_{D,mv}$ of 223 $\pm$ 114 µM for the multivalent interaction is 4 times stronger than $K_{D,mono}$. The relatively weak $K_{D,mv}$ for the various polyUb titrations, compared to the increase in affinity with increasing chain length observed for analyses that assume 1:1 binding, underscores the advantage of employing multivalent binding for achieving a large increase in affinity from a fundamentally low-affinity interaction in a biological system. Our detailed analyses of the NMR-based titrations have allowed us to separate $K_{D,mv}$ from the macroscopic K_Ds that were determined with the assumption of 1:1 binding and which represent the sum over all binding modes (Table 3.2). The 4-fold difference between $K_{D,mono}$ and $K_{D,mv}$, combined with increased numbers of binding sites in longer polyUb chains, results in a 59-fold increase in the macroscopic affinity of RAP80-tUIM for tandem Ub_4 compared

Table 3.2: Multivalent dissociation constants for binding of RAP80-tUIM to tandem Ub chains

$K_{D,mv}$ (μM)[a]		
tUb_2[b]	tUb_3	tUb_4
150 ± 60[c]	80 ± 23[c]	200 ± 40[c]
200 ± 40[d]	400 ± 110[d]	320 ± 45[d]

[a] $K_{D,mv}$ (μM) values were determined from bound states and equilibria shown in Supporting Information.
[b] tUb_n indicates tandem Ub chains (n = 2, 3, 4) linked sequentially through respective N- and C-termini.
[c] Titration of tUb_n into [U-^{15}N]- or [U-^{15}N, U-^{2}H]-RAP80-tUIM.
[d] Titration of RAP80-tUIM into [U-^{15}N]- or [U-^{15}N, U-^{2}H]-tUb_n.

to monoUb.

The simplest molecular unit for multivalent binding of RAP80-tUIM to polyUb is characterized by the binding of RAP80-tUIM to Ub_2 (figure 3.3c). From a molecular perspective, the multivalent dissociation constant can be understood as stepwise binding of one UIM, followed by binding of the second UIM, where the second binding event is considered to be intramolecular (25). For this type of molecular model, the multivalent dissociation constant is given by:

$$K_{D,mv} = K_{D,mono1} K_{D,mono2} / \rho_L(r_0) \tag{3.1}$$

where $K_{D,mono1}$ is the dissociation constant for binding of the first UIM, $K_{D,mono2}$ corresponds to binding of the second UIM, and $\rho_L(r_0)$ is the effective concentration, or probability density for the end-to-end length of the linker vector r_0 (25). The effective concentration has typical values in the millimolar range for peptide linkers that are flexible (26). Substituting the values of $K_{D,mono}$ and $K_{D,mv}$ for the RAP80-tUIM-polyUb interaction (Tables 3.1 and 3.2) into eq 3.1 and solving for $\rho_L(r_0)$ gives an effective concentration of 4 mM. For K63-linked chains, the assumption that $K_{D,mv}$ is 22 μM (Table 3.1) gives $\rho_L(r_0)$ 12 mM. The larger $\rho_L(r_0)$ value for K63-linked chains is consistent with the greater affinity of RAP80-tUIM for K63-linked Ub_2. However, this straightforward analysis cannot separate differences in structural and entropic contributions to binding between tandem Ub_2 and K63-linked Ub_2 that are unrelated to linker length; binding differences are simply grouped together as changes in $\rho_L(r_0)$. Furthermore, eq 3.1 as-

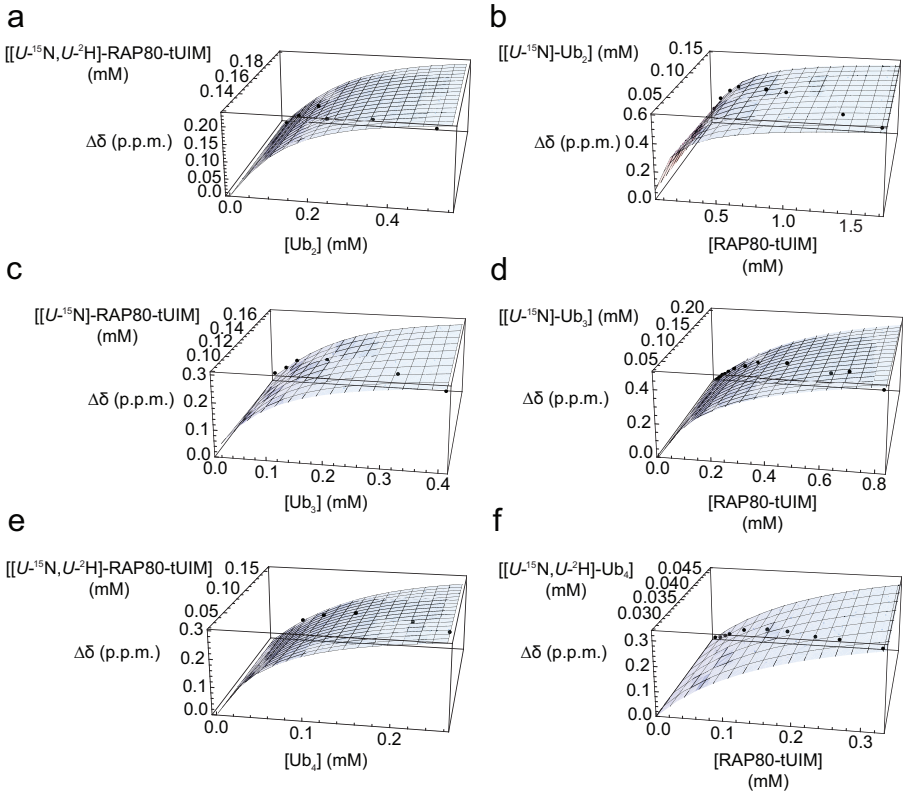

Figure 3.6: Fits of chemical shift perturbation data to multivalent binding models for RAP80-tUIM binding to polyUb. Experimental chemical shift changes are indicated on the vertical axis, and concentrations of analyte or titrand (labeled RAP80-tUIM in left panels, labeled polyUb in right panels) and unlabeled titrant (polyUb in left panels, RAP80-tUIM in right panels) are indicated on the horizontal axes. (a) ^{15}N chemical shift changes for E116 from [U-^{15}N, U-^{2}H]-RAP80-tUIM titrated with unlabeled Ub_2. (b) ^{15}N chemical shift changes for I44 from [U-^{15}N]-Ub_2 titrated with unlabeled RAP80-tUIM. (c) ^{15}N chemical shift changes for E116 from [U-^{15}N]-RAP80-tUIM titrated with unlabeled Ub_3. (d) ^{15}N chemical shift changes for I44 from [U-^{15}N]-Ub_3 titrated with unlabeled RAP80-tUIM. (e) ^{15}N chemical shift changes for E116 from [U-^{15}N, U-^{2}H]-RAP80-tUIM titrated with unlabeled Ub_4. (f) ^{15}N chemical shift changes for I44 from [U-^{15}N, U-^{2}H]-Ub_4 titrated with unlabeled RAP80-tUIM. Experimentally determined chemical shift changes are shown as points, and the best fits to associated binding isotherms are shown as surfaces (Materials and methods and Supporting information)

sumes that the Ub_2 "receptor" is a rigid entity, whereas for both tandem and K63-Ub_2, the Ub moieties are linked by dynamic tethers and are flexible with respect to each other. The consequence of this effect in using equation 3.1 is that it will lead to lower-than-expected effective concentrations. Additionally, the assumption that $K_{D,mv}$ 22 µM for the RAP80-tUIM-Ub_2 interaction should be taken as a lower limit for $K_{D,mv}$, as the actual multivalent dissociation constant is most likely larger, given that this value was determined with the assumption of 1:1 binding, whereas the underlying interactions have more binding modes, which typically leads to a weaker $K_{D,mv}$ (figure 3.1).

The effective concentration values for the interaction of RAP80-tUIM with tandem and K63-Ub_2 can be compared to multivalent binding of the tandem UIMs from S5a (S5a-tUIM), a ubiquitin receptor from the proteasome, to K48-linked Ub_2 (28). Given values of $K_{D,mono1}$ and $K_{D,mono2}$ of 350 and 73 µM, respectively, and if it is assumed that $K_{D,mv}$ is similar to the dissociation constant for the K48-Ub_2-S5a-tUIM interaction (9 µM), then $\rho_L(r_0)$ 3 mM. The decrease in $\rho_L(r_0)$ for the K48-Ub_2-S5a-tUIM interaction compared to the RAP80-tUIM interaction with tandem or K63-linked Ub_2 is consistent with the significantly longer inter-UIM linker (60 residues) compared to that of RAP80-tUIM (10 residues). However, the linker between UIMs in S5a has substantial stretches of α-helical structure (30 residues), whereas equation 3.1 is formulated under the assumption that the linker region is entirely flexible.

In addition to stepwise binding and effective concentration as a mechanism for multivalent binding in RAP80-tUIM, it is also possible that cooperative effects play a role. For example, binding of the first UIM may induce helical structure in the linker region, thereby stabilizing helical structure in the second UIM. Furthermore, conformational selection can also have an impact on the basic tandem UIM-Ub_2 interaction; these effects are discussed in more detail below.

We complemented our NMR-based equilibrium binding studies with measurements of the kinetics of binding and the dynamics of the individual interacting partners. These results indicate that gains in affinity from multivalent binding are balanced by dynamic effects to provide a combined driving force for recognition of extended polyUb chains; this balance between enthalpy and entropy may have physiological relevance with respect to initiating transient protein-protein complexes at sites of DNA damage.

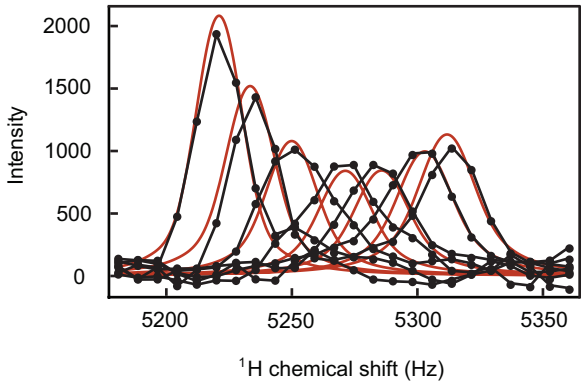

Figure 3.7: Line-shape analysis for [U-^{15}N]-RAP80-tUIM E81 $^1H^N$ upon titration with tandem Ub_2. Experimental data are shown in black, and the fit is shown in red

Binding kinetics for the RAP80-tUIM-Ub_2 interaction

The kinetics of the interaction between RAP80-tUIM and tandem Ub_2 were determined with the assumption of 1:1 binding and by use of line-shape analysis for ^{1}H-^{15}N HSQC NMR spectra acquired during the titration of unlabeled Ub_2 into [U-^{15}N, U-^{2}H]-RAP80-tUIM, to yield estimates of 1.4×10^8 M^{-1} s^{-1} and 10000 s^{-1} for k_{on} and k_{off}, respectively (figure 3.7). This fast off-rate has important implications for signal amplification through multivalent binding utilizing numerous weak interactions. The likely role of the polyUb signal in the DNA damage response is to act as a transient initiator of dynamic protein-protein complexes at sites of damage. From a physiological standpoint, multivalent binding that is inherently weak with fast kinetics ensures that temporal and spatial fluctuations of protein-protein signaling cascades, such as those initiated by recognition of polyUb chains, occur on rapid time scales.

Conformational selection and UIM-Ub recognition: impact of the UIM helix-coil transition

The helix-coil transition for the RAP80-tUIM α-helices may have a significant impact on the binding of Ub and polyUb chains. The transition for RAP80-tUIM was studied by following the temperature dependence of main-chain $^1H_\alpha$ and $^{13}C_\alpha$ chemical shifts (figure 3.8) and was assumed to obey a two-state, cooperative unfolding process, yielding a melting temperature of 34 ℃, a thermal transition width of 28 ℃, and per-residue

chemical shift differences between helical and unfolded states. The width of the thermal transition is defined as the temperature range that covers the middle 80% of the transition (eq 1 in (15)). Thus, at 25 ℃, the α-helices of RAP80-tUIM are 70% helical. These results are consistent with similar studies using α-helical peptides (16). The kinetics of the helix-coil transition was characterized by measuring cross-correlated main-chain amide ^{1}H-^{15}N single quantum coherence and two-spin order relaxation rates (18), and a series of main-chain amide ^{1}H-^{15}N single and multiple quantum coherence and two-spin order relaxation rates (19, 20). The resulting average rate of chemical exchange (R_{ex}) for the α-helices at 5 ℃, presumably arising from the helix-coil transition, is 1.0 ± 0.8 s^{-1}. This R_{ex} value can be used to determine an approximate estimate for the rate of exchange (k_{ex}) for the helix-coil transition (19):

$$R_{ex} = \frac{p_H p_C \Delta\omega^2 k_{ex}}{\Omega_C^2 + \omega_1^2 + k_{ex}^2} \tag{3.2}$$

where ω_1 is the spin-lock field strength (radians per second); $\Delta\omega = \Omega_H - \Omega_C$ (radians per second), where Ω_H and Ω_C are the offset frequencies for the helix and coil states from the carrier frequency (radians per second), respectively; $k_{ex} = k_f + k_u$, where k_f and k_u are the folding and unfolding rate constants (per second), respectively; and p_H and p_C are the fractional populations of the helix and coil states. For $p_H = 0.9$ at 5 ℃, determined from the temperature dependence of $^{13}C_\alpha$ and $^{1}H_\alpha$ chemical shifts, an assumed ^{15}N chemical shift difference between states of 5 ppm, and an ^{15}N spin-lock field strength of 1.8 kHz, we calculate an R_{ex} value of 1 s^{-1}, using a value for k_{ex} of 3×10^5 s^{-1}, or $\tau_{ex} = 1/k_{ex} = 3$ µs. This estimate for the rate of exchange due to the helix-coil transition is consistent with previous NMR (16) and temperature jump (29) measurements, which indicate that τ_{ex} for the helix-coil transition is 0.1-1µs.

The helix-coil transition for the α-helices within RAP80-tUIM has the potential to alter the meaning of the macroscopic binding affinity. Assuming that Ub binds only the α-helical state of the UIM, that is, binding is governed by conformational selection (30), we estimate that the macroscopic dissociation constants determined by NMR may be too large by about 4-fold for the basic multivalent RAP80-tUIM-Ub$_2$ interaction. For example, consider the following expression for the concentration of bound protein in a

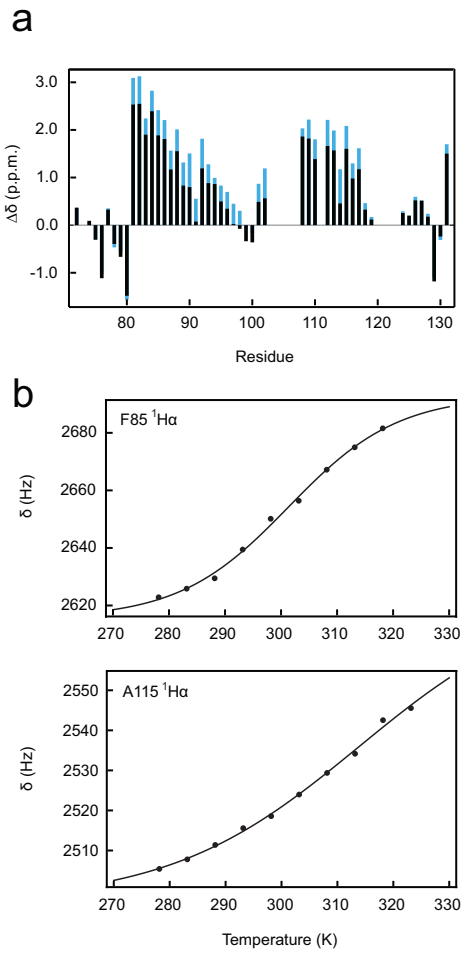

Figure 3.8: (a) Deviation from random coil $^{13}C_{\alpha}$ chemical shifts for [*U*-^{13}C, *U*-^{15}N]-RAP80-tUIM, at 5 ℃ (blue) and 25 ℃ (dark grey). Residues 104-107 are a stretch of contiguous Glu residues that are unassigned. (b) Temperature dependence of $^{1}H_{\alpha}$ chemical shifts for F85 and A115 from [*U*-^{13}C, *U*-^{15}N]-RAP80-tUIM

1:1 protein-ligand interaction:

$$[\mathrm{PL}] = \left\{K_{\mathrm{D1}} + L_0 + P_0 + [K_{\mathrm{D1}} + (L_0 - P_0)^2 + 2K_{\mathrm{D1}}(L_0 + P_0)]^{1/2}\right\}/2 \tag{3.3}$$

where L_0 is the total concentration of ligand, given by the sum of free ligand [L] and ligand bound to protein ([PL]), P_0 is the total protein concentration, given by the sum of free protein [P] and protein bound to ligand ([PL]), and K_{D1} is the dissociation constant for the protein-ligand interaction. Including an equilibrium between α-helical and unfolded states for the ligand in a 1:1 protein-ligand interaction, and assuming that the protein only binds ligand that is in the α-helical conformation (conformational selection), gives the following expression:

$$\begin{aligned}[\mathrm{PL_H}] = \big\{&K_{\mathrm{D2}} + K_{\mathrm{D2}}K_{\mathrm{H}} + K_{\mathrm{H}}L_0 + K_{\mathrm{H}}P_0 - [4K_{\mathrm{D2}}K_{\mathrm{H}}(1 + K_{\mathrm{H}})L_0 \\ &+ (K_{\mathrm{D2}} + K_{\mathrm{D2}}K_{\mathrm{H}} - K_{\mathrm{H}}L_0 + K_{\mathrm{H}}P_0)^2]^{1/2}/2K_{\mathrm{H}}\end{aligned} \tag{3.4}$$

P_0 is the total protein concentration, given by the sum of free protein [P] and protein bound to ligand in the α-helical state ($[\mathrm{PL_H}]$); L_0 is the total concentration of ligand, given by the sum of free ligand in the coil state ($[\mathrm{L_C}]$), ligand in the α-helical state ($[\mathrm{L_H}]$), and ligand in the α-helical state bound to protein ($[\mathrm{PL_H}]$). K_{D2} is the dissocation constant for the interaction between protein and α-helical ligand, and $K_{\mathrm{H}} = (1 - p_{\mathrm{H}})/p_{\mathrm{H}}$, where p_{H} is the fraction of ligand in the α-helical conformation. For $p_{\mathrm{H}} = 0.7$ (the fraction of a given RAP80 UIM in the α-helical conformation at 25 ℃), $K_{\mathrm{D1}} = 900$ μM (similar to $K_{\mathrm{D,mono}}$), $P_0 = 1$ mM, and $L_0 = 2$ mM, equations 3.3 and 3.4 predict that, for $K_{\mathrm{D1}} = K_{\mathrm{D2}}$, conformational selection causes the macroscopic binding isotherm to appear as though binding is weaker, that is, $[\mathrm{PL}]/[\mathrm{PL_H}] = 1.2$ for $P_0 = 1$ mM, and $L_0 = 2$ mM. If K_{D2} is decreased by a factor of 1.6, giving an apparent increase in affinity, then the binding isotherms have a similar appearance with $[\mathrm{PL}]/[\mathrm{PL_H}] = 1$.

The impact of conformational selection in the multivalent binding models can be assessed in the simplest sense by inspection of equation 3.1. If both the UIM domains from RAP80 are in equilibrium between the α-helical and unfolded states, and polyUb binds only the α-helical conformation, both $K_{\mathrm{D,mono1}}$ and $K_{\mathrm{D,mono2}}$ are underestimated by a factor of 2. Therefore, exclusion of conformational selection in the multivalent

binding models gives a 4-fold overestimate of $K_{\mathrm{D,mv}}$; that is, the apparent affinity of the basic multivalent interaction is reduced by conformational selection.

Thus, gains in affinity through multivalent binding are at least partly offset by conformational selection effects arising from the helix-coil transition for the UIM. Balancing affinity gains due to multivalent binding may be important in maintaining transient, dynamic protein?protein complexes at sites of DNA damage undergoing repair processes. Finally, we cannot rule out that an induced-fit mechanism plays a role in the basic, multivalent RAP80-tUIM-Ub_2 interaction. Specifically, binding of one UIM from RAP80 may induce the cooperative formation of α-helical structure in the second UIM α-helix through the linker. Indeed, it has been shown for 1:1 enzyme-ligand binding, wherein the enzyme is in equilbrium between tight and weak ligand binding forms and the rate constants for the underlying equilibria are known, that the flux through both mechanisms can vary, one mechanism can be favoured, or they can occur simultaneously, depending on the enzyme and ligand concentration ratios (31).

Main-Chain dynamics of RAP80-tUIM: impact on multivalent polyUb binding and conformational selection

Fast, pico- to nanosecond time scale fluctuations, or tier-1 and tier-2 dynamics (32), for the protein main chain are often characterized by ^{15}N-R_1, R_2, and NOE NMR relaxation measurements (33–35). These motions play key roles in molecular processes that determine the affinities of proteins for ligands and other proteins, protein stability, and the catalytic activity of enzymes (32). Pico- to nanosecond time scale fluctutations determined by NMR relaxation methods are frequently simulated by molecular dynamics and have the potential to provide important insights into the fundamental determinants of molecular motions (22, 36). In this study, the isotropic reorientational eigenmode (iRED) molecular dynamics approach (22) was used to simulate ^{15}N-R_1, R_2, and NOE NMR relaxation parameters for the main chain of RAP80-tUIM at 25 ℃. The iRED approach has allowed us to investigate the role of tier-1 and tier-2 dynamics in the structure and stability of the UIM domains from RAP80 and the role that flexibility between these domains plays with respect to multivalent binding.

The experimental relaxation data at 25 ℃ and magnetic field strengths of 600 and 800 MHz are in good agreement with those calculated from the molecular dynamics simulation using the iRED general framework (figures 3.9 and 3.10). However, it should be noted that rapid exchange ($k_{ex} > 3 \times 10^5$ s^{-1}) of the α-helical regions between the helical and unfolded states can be expected to give rise to experimental ^{15}N-R_1, R_2, and NOE values that are a weighted average between these states. In addition, the R_2 values may have small contributions (<1 s^{-1} at 5 ℃) arising from chemical exchange due to the helix-coil transition. Thus, experimental R_1 values can be expected to be smaller than those for a rigid helix, experimental R_2 values larger than expected (in the absence of chemical exchange contributions), and NOE values smaller than expected. Additionally, various AMBER force fields are known to overestimate the rigidity of isolated α-helices; indeed, the N- and C-terminal α-helices are 95% helical over the time course of the molecular dynamics simulation, as calculated by the DSSP algorithm. Consistent with this and the expected averaging of the relaxation rates due to fast exchange between helical and unfolded conformations, the experimental NOE values are smaller in magnitude than the simulated values, with rmsd values of 0.3 and 0.2 at 600 and 800 MHz. The simulated R_2 values have rmsd values at both fields of 1 s^{-1} and appear to be overestimated at 600 but not at 800 MHz. The R_1 values are generally in good agreement between experiment and the simulation, with rmsd values of 0.2 s^{-1} at 600 and 800 MHz. With respect to expected R_1 decreases due to rapid averaging, we note that, for a relatively rigid protein, the R_1 value is less sensitive to motions with correlation times in the 1-4 ns range compared to the R_2 and NOE.

Further analysis of simulated pico to nanosecond time scale main-chain dynamics of RAP80-tUIM at 25 ℃ using the iRED approach (22) indicates that the N- and C-terminal helices are flexible with respect to each other (figure 3.9 and 3.10), given that the inter-helical region is flexible. In addition, we have not detected long-range contacts (^{1}H-^{1}H NOEs) between residues in the different UIMs. These results immediately suggest that multivalent effects are not maximized for the UIM-polyUb recognition motif, as this would require the binding partners to be rigid, minimizing the entropic costs of binding. However, the temperature dependence of the $^{13}C_\alpha$ chemical shifts and main-chain ^{15}N relaxation data suggest that the linker is not completely flexible or completely

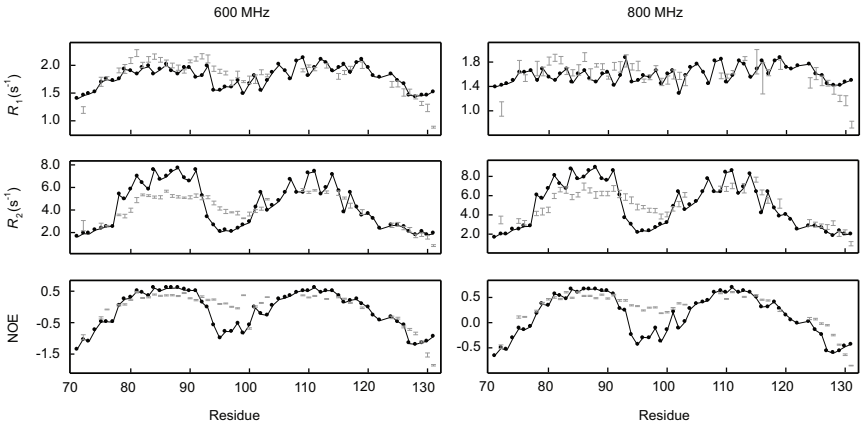

Figure 3.9: Main-chain amide ^{15}N-R_1, R_2, and NOE data at 25 ℃, and 600 and 800 MHz for [U-^{15}N]-RAP80-tUIM. Experimental data are shown as grey bars, and relaxation parameters calculated via iRED analysis are shown as black circles

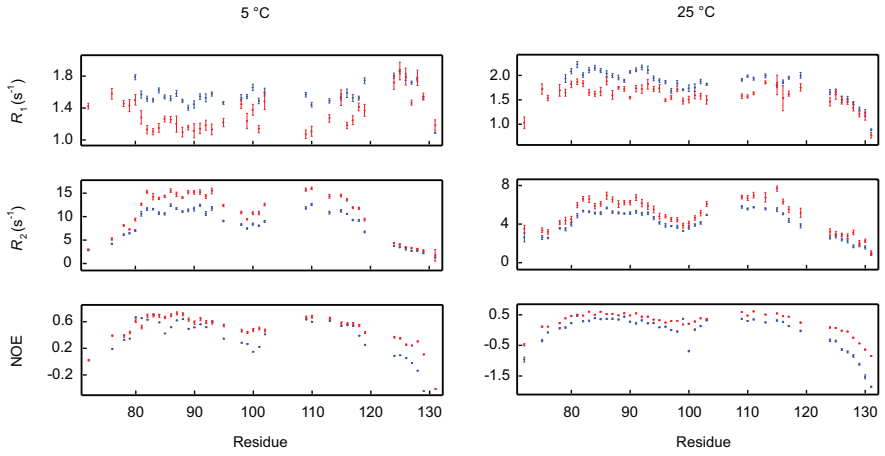

Figure 3.10: Main-chain amide ^{15}N-R_1, R_2, and NOE data at 5 and 25 ℃, and 600 (blue) and 800 MHz (red) for [U-^{15}N]-RAP80-tUIM.

random coil (figures 3.8 and 3.9). Thus, while RAP80-tUIM does not maximize affinity by maintaining a binding-competent orientation for the UIM domains in the unbound state, which would compensate the entropic penalty of binding, the maximum entropic cost is not incurred either, as the linker region appears partly structured, presumably α-helical, as observed in the bound state (figure 3.3c). For example, the temperature dependence of the random coil $^{13}C_\alpha$ chemical shifts in the linker region of RAP80-tUIM (figure 3.8a) is indicative of structural transitions within this region, most likely between random coil and α-helix. This is consistent with the ^{15}N-R_1, R_2, and NOE NMR relaxation parameters for RAP80-tUIM (figure 3.9), which indicate that the linker region is not completely flexible, or at least not as flexible as the N- and C-termini, having values between the extreme terminal regions and the partly (70% at 25 ℃) α-helical UIM domains. As discussed above, the fact that only a fraction of the UIMs from RAP80 are in the α-helical conformation is an important factor underlying the strength of the UIM-Ub interaction. The entropic cost of binding can be expected to be lower than that for Ub binding to an unstructured peptide that subsequently becomes α-helical. However, this affinity gain is offset by conformational selection, or the fact that the fraction of helix is 50% at physiological temperature (37 ℃).

^{15}N main-chain dynamics measurements for RAP80-tUIM indicate that the linker region between UIMs is partly flexible in free and tandem Ub_2-bound RAP80-tUIM (figure 3.11). The flexibility of the linker allows the UIMs to interact with rigid monoUb moieties that are linked through flexible tethers in polyUb. From a molecular dynamics simulation for tandem Ub_3, we estimate that the correlation time for reorientation of Ub moieties with respect to each other in Ub_3 is 15 ns (figure 3.12). Considering the close structural similarity between tandem and Lys63-linked polyUb chains (24), the inter-Ub flexibility observed in the simulation is consistent with previous NMR and small-angle X-ray scattering (SAXS) studies, which indicate that Lys63-linked chains are extended and lack interUb contacts (37, 38). In addition, NMR relaxation measurements indicate that individual Ub moieties in Lys-48-linked Ub_2 reorient on a 10 ns time scale (39). For RAP80, iRED analysis of the ^{15}N relaxation data indicates that N- and C-terminal UIMs reorient with respect to each other on a time scale < 10 ns. With respect to a molecular mechanism for binding, specific selection of polyUb conformations by tandem UIMs can

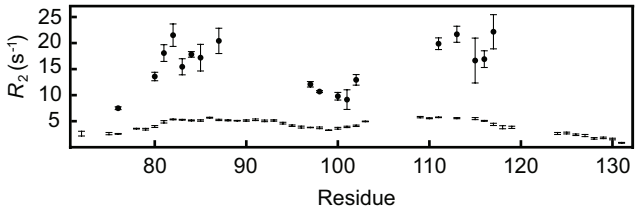

Figure 3.11: ^{15}N-R_2 values for RAP80-tUIM free (bars) and bound to tandem Ub_2 (circles). The baseline ^{15}N-R_2 value is indicative of the overall tumbling of the main chain (UIM helix); decreases below baseline values indicate regions of increased flexibility. Due to the low concentration and the concomitant low signal-to-noise ratio of the RAP80-tUIM sample (120 µM), it was not possible to obtain accurate R2 values for approximately two-thirds of the residues within the bound state; these data were excluded

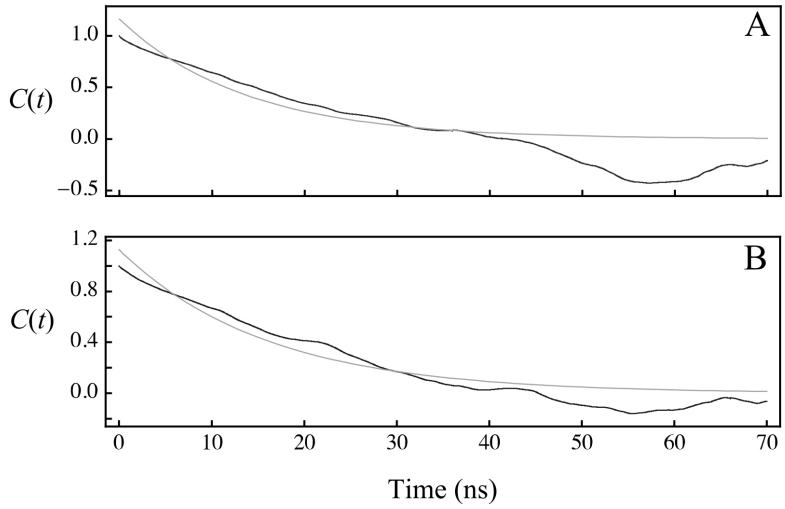

Figure 3.12: Time correlation functions for vectors from the first two interUb subunits perpendicular (A) and parallel (B) to the long axis of a tandem Ub_3 chain. The correlation functions calculated from the MD trajectory are shown in black, fits to a two parameter monoexponential decay are shown in grey

be expected to have a smaller impact on binding than conformational selection arising from the UIM helix-coil transition. For example, the initial binding of one UIM to one Ub moiety can be followed by rapid, nanosecond time scale reorientation and subsequent binding of the neighboring UIM to the neighboring Ub.

There are important consequences regarding nanosecond time scale inter-helical flexibility for the UIMs and inter-Ub flexibility for polyUb chains. To reiterate the previous discussion, it is possible to maximize multivalent affinity in biological systems by employing a rigid scaffold such that tandem UIM and Ub moieties are properly positioned to minimize the entropic cost of binding. However, the flexibility between domains in both

RAP80-tUIM and polyUb indicates that this plays a partial role in the molecular basis for recognition. On the other hand, the entropic penalty incurred through immobilizing adjacent UIM/Ub domains upon binding is not fully realized, as the linker between UIM domains in RAP80 appears partly structured in the unbound state (figures 3.8a and 3.9). This has important implications for an induced-fit mechanism, where binding and stabilization of the α-helical structure of one UIM to one Ub moiety in tandem Ub_2 can lead to a cooperative structuring of the second UIM α-helix through the linker, which subsequently induces binding to the adjacent Ub moiety.

The structure of RAP80-tUIM bound to K63-linked Ub_2 (figure 3.3c, PDB ID 3A1Q) (11) shows that the linker is in a helical conformation. However, the average C_α B-factor for the α-helices from the two molecules of RAP80-tUIM in the asymmetric unit is 22 $Å^2$, and this value increases to 30 $Å^2$ for the linker region. These results indicate that there is disorder and/or flexibility within the linker region of RAP80-tUIM in the fully bound conformation, consistent with the flexibility in the bound state as demonstrated herein by NMR relaxation measurements (figure 3.11). However, given that the underlying contributions to the *B*-factor are multifaceted, there is not a quantitative relationship between *B*-factors and flexibility as measured by NMR relaxation. The NMR results presented herein and the *B*-factors from the crystallographically determined structure suggest that the linker undergoes fluctuations between unstructured and α-helical states when bound to Ub_2. This is consistent with the relatively small chemical shift changes observed for the UIM linker in RAP80-tUIM upon binding tandem Ub_2, Ub_3, and Ub_4 (figure 3.3b), which indicate that while a structural change, or a shift toward a more structured state, occurs within the linker region, it does not become a rigid α-helix. These results suggest that the entropic cost of structuring the linker in RAP80-tUIM is not maximized upon binding Ub_2. Furthermore, the apparent propensity for α-helical structure within the linker region of RAP80-tUIM in the unbound state suggests that an induced-fit mechanism may be possible upon binding Ub_2.

Conclusions

The fundamental question with respect to physiology is why do Ub signaling networks utilize numerous weak ligand-Ub interactions, as opposed to tighter recognition of the Ub receptor? The answer is perhaps 2-fold (40): first, weak binding is likely necessary for desensitization and integration of Ub signaling, because Ub must be recycled back into the free monoUb pool. If specificity and affinity were achieved by strong binding interactions, bound Ub would require removal from the cell, with subsequent up-regulation of Ub expression to replenish the free pool. Second, it is generally believed that weak binding is necessary to maintain protein interaction networks. With respect to K63 polyUb signaling in the DDR pathway, high-resolution imaging of DNA double-strand break foci indicates a remarkable degree of plasticity within these structures; not only do they rapidly form after the initial genotoxic stress, but they also reorganize as evidenced by changes in size and number of foci following damage (41). Our results show that the recognition of polyUb by RAP80 is dynamic, providing a flexible mechanism for the transient initiation of DNA damage foci that ultimately regulate DNA repair.

Supporting information

Bound states and equilibria for the RAP80-tUIM-monoUb interaction

$$[RAP80]_{total} = [RAP80] + [RAP80 - Ub]_1 + [RAP80 - Ub]_2 + [RAP80 - 2Ub] \quad (3.5)$$

$$[Ub]_{total} = [Ub] + [RAP80 - Ub]_1 + [RAP80 - Ub]_2 + 2[RAP80 - 2Ub] \quad (3.6)$$

$$[RAP80 - Ub]_1 = \frac{[RAP80][Ub]}{K_{\mathrm{D,mono}}} \quad (3.7)$$

$$[RAP80 - Ub]_2 = \frac{[RAP80][Ub]}{K_{\mathrm{D,mono}}} \quad (3.8)$$

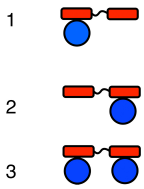

Figure 3.13: Bound states and equilibria for the RAP80-tUIM-monoUb interaction

$$[RAP80-2Ub] = \frac{[RAP80-Ub]_1[Ub]}{2K_{\mathrm{D,mono}}} + \frac{[RAP80-Ub]_2[Ub]}{2K_{\mathrm{D,mono}}} \tag{3.9}$$

Expressions for the fraction of [U-^{15}N]-RAP80-tUIM bound to Ub, for the N-terminal and C-terminal UIMs:

$$f_{\mathrm{bound}}^{[U\text{-}^{15}\mathrm{N}]\text{-RAP80-tUIM (N-UIM)}} = \frac{1[RAP80-Ub]_1 + 1[RAP80-2Ub]}{[RAP80]_{total}} \tag{3.10}$$

$$f_{\mathrm{bound}}^{[U\text{-}^{15}\mathrm{N}]\text{-RAP80-tUIM (C-UIM)}} = \frac{1[RAP80-Ub]_1 + 1[RAP80-2Ub]}{[RAP80]_{total}} \tag{3.11}$$

Bound states and equilibria for the RAP80-tUIM-Ub$_2$ interaction

$$\begin{aligned}[RAP80]_{total} = [RAP80] + [RAP80-Ub_2]_1 + [RAP80-Ub_2]_2 \\ + [RAP80-Ub_2]_3 + 2[2RAP80-Ub_2]_4 + [RAP80-2Ub_2]_5\end{aligned} \tag{3.12}$$

$$\begin{aligned}[Ub_2]_{total} = [Ub_2] + [RAP80-Ub_2]_1 + [RAP80-Ub_2]_2 + [RAP80-Ub_2]_3 \\ + [2RAP80-Ub_2]_4 + 2[RAP80-2Ub_2]_5\end{aligned} \tag{3.13}$$

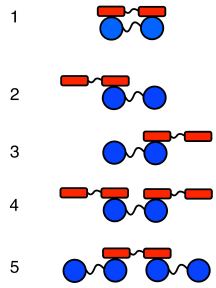

Figure 3.14: Bound states and equilibria for the RAP80-tUIM-Ub_2 interaction

$$[RAP80 - Ub_2]_1 = \frac{[RAP80][Ub_2]}{K_{\mathrm{D,mv}}} \tag{3.14}$$

$$[RAP80 - Ub_2]_2 = \frac{[RAP80][Ub_2]}{K_{\mathrm{D,mono}}} \tag{3.15}$$

$$[RAP80 - Ub_2]_3 = \frac{[RAP80][Ub_2]}{K_{\mathrm{D,mono}}} \tag{3.16}$$

$$[2RAP80 - Ub_2]_4 = \frac{[RAP80 - Ub_2]_2[RAP80]}{2K_{\mathrm{D,mono}}} + \frac{[RAP80 - Ub_2]_3[RAP80]}{2K_{\mathrm{D,mono}}} \tag{3.17}$$

$$[RAP80 - 2Ub_2]_5 = \frac{[RAP80 - Ub_2]_2[Ub_2]}{2K_{\mathrm{D,mono}}} + \frac{[RAP80 - Ub_2]_3[Ub_2]}{2K_{\mathrm{D,mono}}} \tag{3.18}$$

Expression for the fraction of [U-^{15}N]-RAP80-tUIM bound to Ub_2:

$$\begin{aligned} f_{\mathrm{bound}}^{[U\text{-}^{15}\mathrm{N}]\text{-RAP80-tUIM}} = \{&1[RAP80 - Ub_2]_1 + 1[RAP80 - Ub_2] \\ &+ 1[RAP80 - Ub_2]_3 + 1[2RAP80 - Ub_2]_4 \\ &+ 1[RAP80_2Ub_2]_5\}/2[Ub_2]_{total}\}/[RAP80]_{total} \end{aligned} \tag{3.19}$$

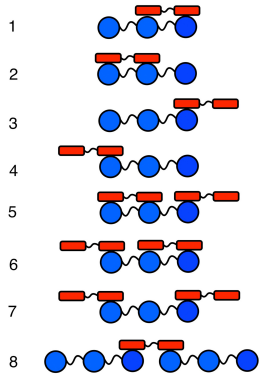

Figure 3.15: Bound states and equilibria for the RAP80-tUIM-Ub_3 interaction

Expression for the fraction of [U-^{15}N]-Ub_2 bound to RAP80-tUIM:

$$\begin{aligned} f_{\text{bound}}^{[U\text{-}^{15}\text{N}]\text{-Ub}_2} = \{&1[RAP80 - Ub_2]_1 + 1[RAP80 - Ub_2] + 0[RAP80 - Ub_2]_3 \\ &+ 1[2RAP80 - Ub_2]_4 + 1[RAP80_2Ub_2]_5\}/2[Ub_2]_{total} \\ &+ \\ &\{1[RAP80 - Ub_2]_1 + 1[RAP80 - Ub_2]_2 + 0[RAP80 - Ub_2]_3 \\ &+ 1[2RAP80 - Ub_2]_4 + 1[RAP80 - 2Ub_2]_5\}/2[Ub_2]_{total} \end{aligned} \tag{3.20}$$

Bound states and equilibria for the RAP80-tUIM-Ub_3 interaction

$$\begin{aligned} [RAP80]_{total} = &[RAP80] + [RAP80 - Ub_3]_1 + [RAP80 - Ub_3]_2 \\ &+ [RAP80 - Ub_3]_3 + [RAP80 - Ub_3]_4 + 2[RAP80 - Ub_3]_5 \\ &+ 2[2RAP80 - Ub_3]_6 + 2[2RAP80 - Ub_3]_7 + [RAP80 - Ub_3]_8 \end{aligned} \tag{3.21}$$

$$[Ub_3]_{total} = [Ub_3] + [RAP80 - Ub_3]_1 + [RAP80 - Ub_3]_2 + [RAP80 - Ub_3]_3 + [RAP80 - Ub_3]_4 + 2[RAP80 - Ub_3]_5 + [2RAP80 - Ub_3]_6 + [2RAP80 - Ub_3]_7 + 2[RAP80 - Ub_3]_8 \quad (3.22)$$

$$[RAP80 - Ub_3]_1 = \frac{[RAP80][Ub_3]}{K_{\mathrm{D,mv}}} \quad (3.23)$$

$$[RAP80 - Ub_3]_2 = \frac{[RAP80][Ub_3]}{K_{\mathrm{D,mv}}} \quad (3.24)$$

$$[RAP80 - Ub_3]_3 = \frac{[RAP80][Ub_3]}{K_{\mathrm{D,mono}}} \quad (3.25)$$

$$[RAP80 - Ub_3]_4 = \frac{[RAP80][Ub_3]}{K_{\mathrm{D,mono}}} \quad (3.26)$$

$$[2RAP80 - Ub_3]_5 = \frac{[RAP80 - Ub_3]_2[RAP80]}{K_{\mathrm{D,mv}} + K_{\mathrm{D,mono}}} + \frac{[RAP80 - Ub_3]_3[RAP80]}{K_{\mathrm{D,mv}} + K_{\mathrm{D,mono}}} \quad (3.27)$$

$$[2RAP80 - Ub_3]_6 = \frac{[RAP80 - Ub_3]_1[RAP80]}{K_{\mathrm{D,mv}} + K_{\mathrm{D,mono}}} + \frac{[RAP80 - Ub_3]_4[RAP80]}{K_{\mathrm{D,mv}} + K_{\mathrm{D,mono}}} \quad (3.28)$$

$$[2RAP80 - Ub_3]_7 = \frac{[RAP80 - Ub_3]_3[RAP80]}{2K_{\mathrm{D,mono}}} + \frac{[RAP80 - Ub_3]_4[RAP80]}{2K_{\mathrm{D,mono}}} \quad (3.29)$$

$$[RAP80 - 2Ub_3]_8 = \frac{[RAP80 - Ub_3]_3[Ub_3]}{2K_{\mathrm{D,mono}}} + \frac{[RAP80 - Ub_3]_4[Ub_3]}{2K_{\mathrm{D,mono}}} \quad (3.30)$$

Expression for the fraction of [U-^{15}N]-RAP80-tUIM bound to Ub_3:

$$\begin{aligned} f_{\text{bound}}^{[U\text{-}^{15}\text{N}]\text{-RAP80-tUIM}} = \{&1[RAP80-Ub_3]_1 + 1[RAP80-Ub_3]_2 \\ &+ 1[RAP80-Ub_3]_3 + 0[RAP80-Ub_3]_4 \\ &+ 2[2RAP80-Ub_3]_5 + 1[2RAP80-Ub_3]_6 \\ &+ 1[2RAP80-Ub_3]_7 + 1[2RAP80-Ub_3]_8\}/[RAP80]_{total} \end{aligned} \tag{3.31}$$

Expression for the fraction of [U-^{15}N]-Ub_3 bound to RAP80-tUIM:

$$\begin{aligned} f_{\text{bound}}^{[U\text{-}^{15}\text{N}]\text{-Ub}_3} = \{&1[RAP80-Ub_3]_1 + 1[RAP80-Ub_3]_2 + 0[RAP80-Ub_3]_3 \\ &+ 1[RAP80-Ub_3]_4 + 1[2RAP80-Ub_3]_5 + 1[2RAP80-Ub_3]_6 \\ &+ 1[2RAP80-Ub_3]_7 + 1[2RAP80-Ub_3]_8\}/3[Ub_3]_{total} \\ &+ \\ &\{1[RAP80-Ub_3]_1 + 1[RAP80-Ub_3]_2 + 0[RAP80-Ub_3]_3 \\ &+ 0[RAP80-Ub_3]_4 + 1[2RAP80-Ub_3]_5 + 1[2RAP80-Ub_3]_6 \\ &+ 0[2RAP80-Ub_3]_7 + 0[2RAP80-Ub_3]_8\}/3[Ub_3]_{total} \\ &+ \\ &\{1[RAP80-Ub_3]_1 + 0[RAP80-Ub_3]_2 + 1[RAP80-Ub_3]_3 \\ &+ 0[RAP80-Ub_3]_4 + 1[2RAP80-Ub_3]_5 + 1[2RAP80-Ub_3]_6 \\ &+ 1[2RAP80-Ub_3]_7 + 1[2RAP80-Ub_3]_8\}/3[Ub_3]_{total} \end{aligned} \tag{3.32}$$

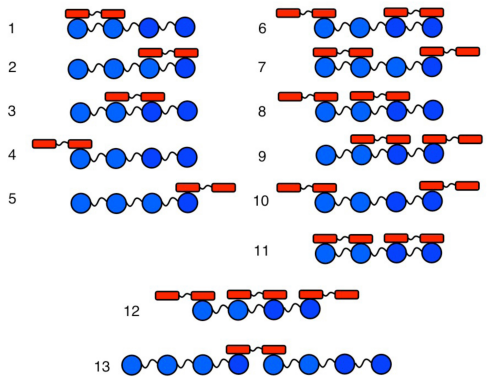

Figure 3.16: Bound states and equilibria for the RAP80-tUIM-Ub_4 interaction

Bound states and equilibria for the RAP80-tUIM-Ub_4 interaction

$$\begin{aligned}[RAP80]_{total} = {} & [RAP80] + [RAP80 - Ub_4]_1 + [RAP80 - Ub_4]_2 \\ & + [RAP80 - Ub_4]_3 + [RAP80 - Ub_4]_4 + [RAP80 - Ub_4]_5 \\ & + 2[2RAP80 - Ub_4]_6 + 2[2RAP80 - Ub_4]_7 + 2[2RAP80 - Ub_4]_8 \\ & + 2[2RAP80 - Ub_4]_9 + 2[2RAP80 - Ub_4]_{10} + 2[2RAP80 - Ub_4]_{11} \\ & + 3[3RAP80 - 2Ub_4]_{12} + [RAP80 - Ub_4]_{13}\end{aligned} \tag{3.33}$$

$$\begin{aligned}[Ub_4]_{total} = {} & [Ub] + [RAP80 - Ub_4]_1 + [RAP80 - Ub_4]_2 + [RAP80 - Ub_4]_3 \\ & + [RAP80 - Ub_4]_4 + [RAP80 - Ub_4]_5 + [2RAP80 - Ub_4]_6 \\ & + [2RAP80 - Ub_4]_7 + [2RAP80 - Ub_4]_8 + [2RAP80 - Ub_4]_9 \\ & + [2RAP80 - Ub_4]_{10} + [2RAP80 - Ub_4]_{11} + [3RAP80 - Ub_4]_{12} \\ & + 2[RAP80 - 2Ub_4]_{13}\end{aligned} \tag{3.34}$$

$$[RAP80 - Ub_4]_1 = \frac{[RAP80][Ub_4]}{K_{\mathrm{D,mv}}} \tag{3.35}$$

$$[RAP80 - Ub_4]_2 = \frac{[RAP80][Ub_4]}{K_{\mathrm{D,mv}}} \tag{3.36}$$

$$[RAP80 - Ub_4]_3 = \frac{[RAP80][Ub_4]}{K_{\mathrm{D,mv}}} \tag{3.37}$$

$$[RAP80 - Ub_4]_4 = \frac{[RAP80][Ub_4]}{K_{\mathrm{D,mono}}} \tag{3.38}$$

$$[RAP80 - Ub_4]_5 = \frac{[RAP80][Ub_4]}{K_{\mathrm{D,mono}}} \tag{3.39}$$

$$[2RAP80 - Ub_4]_6 = \frac{[RAP80 - Ub_4]_2[RAP80]}{K_{\mathrm{D,mv}} + K_{\mathrm{D,mono}}} + \frac{[RAP80 - Ub_4]_4[RAP80]}{K_{\mathrm{D,mv}} + K_{\mathrm{D,mono}}} \tag{3.40}$$

$$[2RAP80 - Ub_4]_7 = \frac{[RAP80 - Ub_4]_1[RAP80]}{K_{\mathrm{D,mv}} + K_{\mathrm{D,mono}}} + \frac{[RAP80 - Ub_4]_5[RAP80]}{K_{\mathrm{D,mv}} + K_{\mathrm{D,mono}}} \tag{3.41}$$

$$[2RAP80 - Ub_4]_8 = \frac{[RAP80 - Ub_4]_3[RAP80]}{K_{\mathrm{D,mv}} + K_{\mathrm{D,mono}}} + \frac{[RAP80 - Ub_4]_4[RAP80]}{K_{\mathrm{D,mv}} + K_{\mathrm{D,mono}}} \tag{3.42}$$

$$[2RAP80 - Ub_4]_9 = \frac{[RAP80 - Ub_4]_3[RAP80]}{K_{\mathrm{D,mv}} + K_{\mathrm{D,mono}}} + \frac{[RAP80 - Ub_4]_5[RAP80]}{K_{\mathrm{D,mv}} + K_{\mathrm{D,mono}}} \tag{3.43}$$

$$[2RAP80 - Ub_4]_{10} = \frac{[RAP80 - Ub_4]_4[RAP80]}{2K_{\mathrm{D,mono}}} + \frac{[RAP80 - Ub_4]_5[RAP80]}{2K_{\mathrm{D,mono}}} \tag{3.44}$$

$$[2RAP80 - Ub_4]_{11} = \frac{[RAP80 - Ub_4]_1[RAP80]}{2K_{\mathrm{D,mv}}} + \frac{[RAP80 - Ub_4]_2[RAP80]}{2K_{\mathrm{D,mv}}} \tag{3.45}$$

$$[3RAP80-Ub_4]_{12} = \frac{[RAP80-Ub_4]_3[RAP80]}{K_{\mathrm{D,mv}}+2K_{\mathrm{D,mono}}} + \frac{[RAP80-Ub_4]_4[RAP80]}{K_{\mathrm{D,mv}}+2K_{\mathrm{D,mono}}} + \frac{[RAP80-Ub_4]_5[RAP80]}{K_{\mathrm{D,mv}}+2K_{\mathrm{D,mono}}} \tag{3.46}$$

$$[RAP80-2Ub_4]_{13} = \frac{[RAP80-Ub_4]_4[Ub_4]}{2K_{\mathrm{D,mono}}} + \frac{[RAP80-Ub_4]_5[Ub4]}{2K_{\mathrm{D,mono}}} \tag{3.47}$$

Expression for the fraction of [U-^{15}N]-RAP80-tUIM bound to Ub_4:

$$\begin{aligned} f_{\mathrm{bound}}^{[U\text{-}^{15}\mathrm{N}]\text{-RAP80-tUIM}} = \{&1[RAP80-Ub_4]_1 + 1[RAP80-Ub_4]_2 \\ &+ 1[RAP80-Ub_4]_3 + 0[RAP80-Ub_4]_4 \\ &+ 1[RAP80-Ub_4]_5 + 1[2RAP80-Ub_4]_6 \\ &+ 2[2RAP80-Ub_4]_7 + 1[2RAP80-Ub_4]_8 \\ &+ 2[2RAP80-Ub_4]_9 + 1[2RAP80-Ub_4]_{10} \\ &+ 2[2RAP80-Ub_4]_{11} + 2[3RAP80-Ub_4]_{12} \\ &+ 1[RAP80-2Ub_4]_{13}\}/[RAP80]_{total} \end{aligned} \tag{3.48}$$

Expression for the fraction of [U-^{15}N]-Ub_4 bound to RAP80-tUIM:

$$
\begin{aligned}
f_{\text{bound}}^{[U\text{-}^{15}\text{N}]\text{-Ub}_4} = \{&1[RAP80 - Ub_4]_1 + 0[RAP80 - Ub_4]_2 + 0[RAP80 - Ub_4]_3 \\
&+ 1[RAP80 - Ub_4]_4 + 0[RAP80 - Ub_4]_5 + 1[2RAP80 - Ub_4]_6 \\
&+ 1[2RAP80 - Ub_4]_7 + 1[2RAP80 - Ub_4]_8 + 0[2RAP80 - Ub_4]_9 \\
&+ 1[2RAP80 - Ub_4]_{10} + 1[2RAP80 - Ub_4]_{11} + 1[3RAP80 - Ub_4]_{12} \\
&+ 1[RAP80 - 2Ub_4]_{13}\}/4[Ub_4]_{total} \\
&+ \\
&\{1[RAP80 - Ub_4]_1 + 0[RAP80 - Ub_4]_2 + 1[RAP80 - Ub_4]_3 \\
&+ 0[RAP80 - Ub_4]_4 + 0[RAP80 - Ub_4]_5 + 0[2RAP80 - Ub_4]_6 \\
&+ 1[2RAP80 - Ub_4]_7 + 1[2RAP80 - Ub_4]_8 + 0[2RAP80 - Ub_4]_9 \\
&+ 0[2RAP80 - Ub_4]_{10} + 1[2RAP80 - Ub_4]_{11} + 1[3RAP80 - Ub_4]_{12} \\
&+ 0[RAP80 - 2Ub_4]_{13}\}/4[Ub_4]_{total} \\
&+ \\
&\{0[RAP80 - Ub_4]_1 + 1[RAP80 - Ub_4]_2 + 1[RAP80 - Ub_4]_3 \\
&+ 0[RAP80 - Ub_4]_4 + 0[RAP80 - Ub_4]_5 + 1[2RAP80 - Ub_4]_6 \\
&+ 0[2RAP80 - Ub_4]_7 + 1[2RAP80 - Ub_4]_8 + 1[2RAP80 - Ub_4]_9 \\
&+ 0[2RAP80 - Ub_4]_{10} + 1[2RAP80 - Ub_4]_{11} + 1[3RAP80 - Ub_4]_{12} \\
&+ 0[RAP80 - 2Ub_4]_{13}\}/4[Ub_4]_{total} \\
&+ \\
&\{0[RAP80 - Ub_4]_1 + 1[RAP80 - Ub_4]_2 + 0[RAP80 - Ub_4]_3 \\
&+ 0[RAP80 - Ub_4]_4 + 1[RAP80 - Ub_4]_5 + 1[2RAP80 - Ub_4]_6 \\
&+ 1[2RAP80 - Ub_4]_7 + 0[2RAP80 - Ub_4]_8 + 1[2RAP80 - Ub_4]_9 \\
&+ 1[2RAP80 - Ub_4]_{10} + 1[2RAP80 - Ub_4]_{11} + 1[3RAP80 - Ub_4]_{12} \\
&+ 1[RAP80 - 2Ub_4]_{13}\}/4[Ub_4]_{total}
\end{aligned}
\tag{3.49}
$$

Weights for the contribution of a bound state to the total fraction bound

The chemical shift changes of [U-^{15}N]-RAP80-tUIM upon titration with unlabeled polyUb chains are described by the equations given above for fraction bound (a function of $K_{\mathrm{D,mv}}$). These expressions reflect binding to either the N-terminal or C-terminal UIM. However, not all bound states will contribute to chemical shift changes for a given UIM, and therefore must be weighted accordingly. For example, chemical shift changes for a residue within the N-terminal UIM will not reflect states where the C-terminal UIM is bound but the N-terminal UIM is not. The equations describing fraction [U-^{15}N]-RAP80-tUIM bound, are used to fit binding to either the N-terminal or C-terminal UIM, given the nearly identical KD,mono values for each UIM.

Analysis of the corollary titrations, for which [U-^{15}N]-polyUb chains are titrated with unlabeled RAP80-tUIM are complicated with respect to fraction bound due to the degeneracy of Ub chemical shifts for a given Ub moiety within a polyUb chain. In this case, binding of an individual UIM from RAP80 to an individual Ub moiety of the chain will result in a chemical shift change. As in the case of [U-^{15}N]-RAP80-tUIM, there are a number of states in which some Ub moieties are bound, whereas others are not. To fit polyUb chemical shift changes for these titrations in order to extract KD,mv values, weights are given to each individual bound state. To reach the fully bound chemical shift, all Ub moieties within a chain must be bound to a UIM from RAP80. With respect to degeneracy in chemical shifts for [U-^{15}N]-polyUb spectra, one might, in principle, expect slightly different chemical shift changes for individual Ub moieties during titration with RAP80-tUIM. However, in practice, all resonances shift in a similar fashion, likely due to the fact that differences in chemical shifts between free and bound states are small in most cases ($<$ 2 ppm for ^{15}N); we estimate that for a given resonance, the maximum difference in chemical shift difference between individual Ub moieties within a chain during titration is 0.2 ppm (^{15}N), or 10% of the maximum chemical shift change upon binding. Given that ^{15}N line-widths for polyUb resonances are on the order of 0.3 ppm, this precludes resolution of differentially shifting resonances from individual Ub moieties. To account for this degeneracy, we averaged the theoretical expression (corresponding to

the fraction bound) for differentially shifting resonances for the individual Ub moieties within a polyUb chain to analyze the titration data.

References

1. Komander, D. (2009) The emerging complexity of protein ubiquitination. *Biochemical Society Transactions* 37, 937–953.

2. Dikic, I., Wakatsuki, S., and Walters, K.J. (2009) Ubiquitin-binding domains-from structures to functions. *Nature Reviews Molecular Cell Biology* 10, 659–671.

3. Hurley, J.H., Lee, S., and Prag, G. (2006) Ubiquitin-binding domains. *Biochemical Journal* 399, 361–372.

4. Wang, B., Matsuoka, S., Ballif, B.A., Zhang, D., Smogorzewska, A., Gygi, S.P., and Elledge, S.J. (2007) Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. *Science* 316, 1194–1198.

5. Kim, H., Chen, J., and Yu, X. (2007) Ubiquitin-binding protein RAP80 mediates BRCA1-dependent DNA damage response. *Science* 316, 1202–1205.

6. Sobhian, B., Shao, G., Lilli, D.R., Culhane, A.C., Moreau, L.A., Xia, B., Livingston, D.M., and Greenberg, R.A. (2007) RAP80 targets BRCA1 to specific ubiquitin structures at DNA damage sites. *Science* 316, 1198–1202.

7. Matsuoka, S., Ballif, B.A., Smogorzewska, A., McDonald III, E.R., Hurov, K.E., Luo, J., Bakalarski, C.E., Zhao, Z., Solimini, N., Lerenthal, Y., Shiloh, Y., Gygi, S.P., and Elledge, S.J. (2007) ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* 316, 1160–1166.

8. Petrini, J.H. (2007) Cell signaling. a touching response to damage. *Science* 316, 1138–1139.

9. Swanson, K.A., Kang, R.S., Stamenova, S.D., Hicke, L., and Radhakrishnan, I. (2003) Solution structure of Vps27 UIM-ubiquitin complex important for endosomal sorting and receptor downregulation. *The EMBO Journal* 22, 4597–4606.

10. Sims, J.J. and Cohen, R.E. (2009) Linkage-specific avidity defines the lysine 63-linked polyubiquitin-binding preference of Rap80. *Molecular Cell* 33, 775–783.

11. Sato, Y., Yoshikawa, A., Mimura, H., Yamashita, M., Yamagata, A., and Fukai, S. (2009) Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by tandem UIMs of RAP80. *The EMBO Journal* 28, 2461–2468.

12. Marley, J., Lu, M., and Bracken, C. (2001) A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR* 20, 71–75.

13. Sattler, M., Schleucher, J., and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* 34, 93–158.

14. Palmer, A.G., Kroenke, C.D., and Loria, J.P. (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods in Enzymology* 339, 204–238.

15. Shalongo, W., Dugad, L., and Stellwagen, E. (1994) Analysis of the thermal transitions of a model helical peptide using ^{13}C NMR. *Journal of the American Chemical Society* 116, 2500–2507.

16. Nesmelova, I., Krushelnitsky, A., Idiyatullin, D., Blanco, F., Ramirez-Alvarado, M., Daragan, V.A., Serrano, L., and Mayo, K.H. (2001) Conformational exchange on the microsecond time scale in α-helix and β-hairpin peptides measured by ^{13}C NMR transverse relaxation. *Biochemistry* 40, 2844–2853.

17. Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. (1994) Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* 33, 5984–6003.

18. Kroenke, C.D., Loria, J.P., Lee, L.K., Rance, M., and Palmer III, A.G. (1998) Longitudinal and transverse ^{1}H-^{15}N dipolar/^{15}N chemical shift anisotropy relaxation interference: Unambiguous determination of rotational diffusion tensors and chemical exchange effects in biological macromolecules. *Journal of the American Chemical Society* 120, 7905–7915.

19. Hansen, D.F., Yang, D., Feng, H., Zhou, Z., Wiesner, S., Bai, Y., and Kay, L.E. (2007) An exchange-free measure of ^{15}N transverse relaxation: an NMR spectroscopy application to the study of a folding intermediate with pervasive chemical exchange. *Journal of the American Chemical Society* 129, 11468–11479.

20. Hansen, A.L., Nikolova, E.N., Casiano-Negroni, A., and Al-Hashimi, H.M. (2009) Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon $R_1\rho$ NMR spectroscopy. *Journal of the American Chemical Society* 131, 3818–3819.

21. Case, D.A., Darden, T.A., III, T.E. Cheatham, Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Walker, R.C., Zhang, W., Merz, K.M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Goetz, A.W., Kolossvai, I., Wong, K.F., Paesani, F., Vanicek, J., Wolf, R.M., Liu, J., Wu, X., Brozell, S.R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M.-J., Cui, G., Roe, D.R., Mathews, D.H., Seetin, M.G., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P.A. (2008). *AMBER 10* University of California, San Francisco.

22. Prompers, J.J. and Brüschweiler, R. (2002) General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. *Journal of the American Chemical Society* 124, 4522–4534.

23. Lo, Y.C., Lin, S.C., Rospigliosi, C.C., Conze, D.B., Wu, C.J., Ashwell, J.D., Eliezer, D., and Wu, H. (2009) Structural basis for recognition of diubiquitins by NEMO. *Molecular Cell* 33, 602–615.

24. Komander, D., Reyes-Turcu, F., Licchesi, J.D.F., Odenwaelder, P., Wilkinson, K.D., and Barford, D. (2009) Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains. *EMBO Reports* 10, 466–473.

25. Jencks, W.P. (1981) On the attribution and additivity of binding energies. *Proceedings of the National Academy of Sciences* 78, 4046–4050.

26. Zhou, H.X. and Gilson, M.K. (2009) Theory of free energy and entropy in noncovalent binding. *Chemical Reviews* 109, 4092–4107.

27. Lee, R.T. and Lee, Y.C. (2000) Affinity enhancement by multivalent lectin-carbohydrate interaction. *Glycoconjugate Journal* 17, 543–551.

28. Zhang, N., Wang, Q., Ehlinger, A., Randles, L., Lary, J.W., Kang, Y., Haririnia, A., Storaska, A.J., Cole, J.L., Fushman, D., and Walters, K.J. (2009) Structure of the S5a:K48-linked diubiquitin complex and its interactions with Rpn13. *Molecular Cell* 35, 280–290.

29. Eaton, W.A., Muñoz, V., Thompson, P.A., Henry, E.R., and Hofrichter, J. (1998) Kinetics and dynamics of loops, α-helices, β-hairpins, and fast-folding proteins. *Accounts of Chemical Research* 31, 745–753.

30. Kumar, S., Ma, B., Tsai, C.J., Sinha, N., and Nussinov, R. (2000) Folding and binding cascades: dynamic landscapes and population shifts. *Protein Science* 9, 10–19.

31. Hammes, G.G., Chang, Y.C., and Oas, T.G. (2009) Conformational selection or induced fit: a flux description of reaction mechanism. *Proceedings of the National Academy of Sciences* 106, 13737–13741.

32. Henzler-Wildman, K. and Kern, D. (2007) Dynamic personalities of proteins. *Nature* 450, 964–972.

33. Spyracopoulos, L. (2005) Thermodynamic interpretation of protein dynamics from NMR relaxation measurements. *Protein and Peptide Letters* 12, 235–240.

34. Jarymowycz, V.A. and Stone, M.J. (2006) Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chemical Reviews* 106, 1624–1671.

35. Mittermaier, A.K. and Kay, L.E. (2009) Observing biological dynamics at atomic resolution using NMR. *Trends in Biochemical Sciences* 34, 601–611.

36. Tong, Y., Ji, C.G., Mei, Y., and Zhang, J.Z.H. (2009) Simulation of NMR data reveals that proteins' local structures are stabilized by electronic polarization. *Journal of the American Chemical Society* 131, 8636–8641.

37. Tenno, T., Fujiwara, K., Tochio, H., Iwai, K., Morita, E.H., Hayashi, H., Murata, S., Hiroaki, H., Sato, M., Tanaka, K., and Shirakawa, M. (2004) Structural basis for distinct roles of Lys63-and Lys48-linked polyubiquitin chains. *Genes to Cells* 9, 865–875.

38. Varadan, R., Assfalg, M., Haririnia, A., Raasi, S., Pickart, C., and Fushman, D. (2004) Solution conformation of Lys63-linked di-ubiquitin chain provides clues to functional diversity of polyubiquitin signaling. *Journal of Biological Chemistry* 279, 7055–7063.

39. Ryabov, Y. and Fushman, D. (2006) Interdomain mobility in di-ubiquitin revealed by NMR. *Proteins: Structure, Function, and Bioinformatics* 63, 787–796.

40. Hicke, L., Schubert, H.L., and Hill, C.P. (2005) Ubiquitin-binding domains. *Nature Reviews Molecular Cell Biology* 6, 610–621.

41. Aten, J.A., Stap, J., Krawczyk, P.M., van Oven, C.H., Hoebe, R.A., Essers, J., and Kanaar, R. (2004) Dynamics of DNA double-strand breaks revealed by clustering of damaged chromosome domains. *Science* 303, 92–95.

Chapter 4

Catalytic proficiency of ubiquitin conjugation enzymes: balancing pK_a suppression, entropy, and electrostatics*

Introduction

Biological organisms orchestrate exquisite responses to various stimuli through temporal fluctuations in protein-protein interaction networks. The ebb and flow of information within these networks is central to life processes, encompassing diverse roles such as repair of damaged DNA (1) and innate immune responses to bacterial pathogens (2). A key crossroad in such pathways involves the build up and recognition of polyubiquitin (polyUb) chains on adaptor proteins (3). Ubiquitination is realized through the combined

*This chapter has been published. Reproduced with permission from: C.J. Markin, L.F. Saltibus, M.J. Kean, R.T. McKay, W. Xiao, and L. Spyracopoulos. Catalytic proficiency of ubiquitin conjugation enzymes: balancing pK_a suppression, entropy, and electrostatics. *Journal of the American Chemical Society*, 132(50):17775-17786, 2010. Copyright 2010 American Chemical Society.

Contributions: L.F.S., M.J.K., and L.S. contributed to the original design of the AF488-Ub assay. C.J.M. and L.S. further optimized the assay, performed NMR experiments, and developed the system of ODEs and the software necessary for analyzing the Ub_2 build-up curves. C.J.M. performed the assays. W.X. designed and supplied clones. R.T.M. provided assistance with NMR data acquisition. L.S. and C.J.M. wrote the paper.

catalytic activity of an enzyme cascade that is initiated by covalent attachment of the C-terminus of Ub (Gly76) to the active site cysteine of a Ub activating enzyme (E1), followed by transfer as a thioester conjugate to a Ub conjugating enzyme (E2). Ultimately, Ub is attached to target proteins through the action of a ubiquitin ligase (E3) which binds the E2 enzyme and the target to facilitate nucleophilic attack of the amino group from a substrate lysine on the thioester bond between the C-terminus of Ub and the conserved cysteine within the E2 active site (4).

The diversity of the molecular architecture for the Ub signal is abundant, extending well beyond the archetypal K48-linked polyUb chain, which destines proteins for degradation. Numerous topological variants of the Ub signal have been identified, such as attachment of a single, or multiple Ub molecules to different target protein sites, or polyubiquitination through single or variable Ub peptide bonds (3). This wide topological variety for the Ub signal likely imparts organisms with a biological advantage for achieving specificity in diverse signaling cascades.

The first step of the ubiquitination cascade, catalyzed by E1 enzyme, is the only one requiring energy through hydrolysis of ATP. The multiple steps in the kinetics of the reaction catalyzed by E1 enzyme are characterized by a turnover number of $1-2\ s^{-1}$ (5). In contrast, the catalytic rate for conjugation of Ub to substrates or build up of polyUb chains is substantially slower, with apparent kcat values ranging from 0.01 to 0.1 s^{-1} (6, 7). This 10-fold variability in apparent rates combined with differences in substrate binding (K_M) suggest a functional diversity in the E2 mechanism that may be important in achieving specificity for the various biological reactions regulated by ubiquitination (7). Indeed, it has recently been demonstrated that attachment of Lys48-linked polyUb to substrates derived from β-catenin and cyclin E1 proteins by their cognate E2-E3 pairs Cdc34-SCF^{Cdc34} and Cdc34-SCF$^{\beta\text{-TrCP}}$, respectively, occurs through sequential addition of Ub molecules to substrate (8). Importantly, a 10-fold difference in the overall rate at which the first Ub molecule was attached to the different substrates was observed. Within the context of a sequentially processive kinetic model, that is, sequential addition of Ub molecules to a substrate (8), a small rate difference for the attachment of the first Ub molecule leads to a significant difference in the fraction of substrates bearing chains of four Ubs, the minimum required for degradation by the proteasome. As interesting as

this kinetic diversity appears, it remains a problem to develop a molecular mechanism for this key regulatory process in ubiquitination.

The rate for covalent attachment of the first Ub to a substrate depends on the catalytic proficiency of E2 enzymes or $(k_{cat}/K_M)/k_{noncat}$. The magnitude of the catalytic proficiency ranges from 10^8 to 10^{23} M^{-1} and is indicative of the ability of a given enzyme to produce a large rate enhancement (k_{cat}/k_{noncat}). Alternatively stated, it is the degree to which an enzyme reduces the activation barrier for a reaction in comparison to the reaction in water (9, 10). In comparison, the efficiency of an enzyme (k_{cat}/k_M) has a maximum value equal to the diffusion limit (10^9 M^{-1} s^{-1}) and represents the efficiency of catalysis with respect to substrate binding. Thus, variability in either k_{cat} or K_M among E2 enzymes represents a regulatory mechanism for polyubiquitination of substrate proteins and the ensuing biological outcomes. The central problem, therefore, is to understand the chemical mechanism for ubiquitination within the context of the catalytic proficiency of E2 enzymes. Kinetic studies of nucleophilic addition of ammonia to thioesters in water using model compounds (11) can be used to calculate a noncatalyzed rate (k_{noncat}), giving a rate enhancement $k_{cat}/k_{noncat} = 10^6 - 10^9$. Thus, the catalytic proficiency of E2 enzymes is modest, with $(k_{cat}/K_M)/k_{noncat} = 10^{11} - 10^{14}$ M^{-1} (10). In comparison, the most proficient enzyme is OMP decarboxylase, with a catalytic proficiency of 2×10^{23} M^{-1}, and the least proficient enzyme is cyclophilin with a catalytic proficiency of 5×10^8 M^{-1} (9).

Recently, for the biological reaction involving attachment of the Ub-like modifier SUMO to the protein substrate GST-RanGAP1 catalyzed by the E2 Ubc9, the catalytic rate enhancement was attributed to a downshift of the pK_a of the substrate lysine ($\Delta pK_a \approx -4$), arising from desolvation effects within the active site (12). That is, the active site environment favors deprotonation of the substrate lysine to promote nucleophilic attack on the thioester bond. This type of catalytic enhancement was proposed on the basis of the pH-rate profile for the acylation of acetoacetate decarboxylase (13). However, lysine pKa suppression may not be the sole source of the catalytic power of E2 enzymes, given that 3-4 ΔpK_a units correspond to a rate enhancement of 1000 to 10000 fold. In addition, the emerging view from computational studies of enzyme mechanism indicates that the main source of catalytic power for many enzymes is stabilization of transition

states by electrostatic complementarity (14, 15). In this regard, the entropic effects of binding were considered to be a major mechanism by which enzymes achieve large rate enhancements (16). However, computational studies indicate that these entropic effects are much smaller than previously suggested (17).

In this study, we performed enzyme assays to measure non-steady-state kinetics for assembly of Lys63-linked polyUb chains catalyzed by the heterodimeric enzyme Mms2-Ubc13. Lys63-linked polyUb chains are critical for recruiting DNA polymerases to the sliding clamp PCNA, thereby initiating error-free postreplicative DNA repair (18–20). Lys63-linked chains play key roles in the immune response through polyubiquitination of the signaling adapter protein TRAF6 and ultimately result in altered gene expression by activation of the transcription factor NF-κB (21). These chains are also essential for recruitment of DNA repair protein complexes to sites of DNA double-strand breaks (22–24).

The heterodimeric enzyme Mms2-Ubc13 is composed of the E2 Ubc13, which contains a catalytic cysteine, and Mms2, an E2-like protein that is structurally similar to Ubc13 but lacks a catalytic cysteine. For Ubc13, E1 enzyme catalyzes the covalent attachment of the C-terminus of a donor Ub to the catalytic cysteine through a thioester bond. Mms2 is tightly bound to Ubc13 and binds an acceptor Ub such that Lys63 is proximal to the thioester bond of the catalytic cysteine from Ubc13, facilitating nucleophilic attack of the acceptor Ub on the thioester bond of the donor Ub (25, 26). In general, k_{cat} and K_{M} values for E2 enzymes are determined using initial rate analysis under the assumption of steady-state, Michaelis-Menten kinetics. However, E2-catalyzed reactions are strictly non-steady state, given that E2 enzyme with a thioester-conjugated Ub molecule is consumed during the reaction with substrate. Furthemore, the noncovalent interaction between Mms2 and Ubc13 is not accounted for in the basic Michaelis-Menten enzymatic reaction. Thus, we derived rate laws to describe the non-steady-state kinetics for catalysis of polyUb chains by Mms2-Ubc13, which included the kinetics of the various protein-protein interactions involved in the mechanism. The protein-protein interaction kinetics were characterized using NMR line-shape and ZZ-exchange analyses. The results indicate that different E2s have 10- to 100-fold variability in rate enhancement. Combined with a 10-fold variability in K_{M} for substrate binding gives a 1000-fold window in

k_{cat}/K_M, raising the intriguing possibility that the mechanism underlying the specificity of E2-catalyzed reactions can be determined in part by differences in proficiency among this family of enzymes. The results have broad implications for regulation of the assembly of polyUb chains and importantly attachment of Ub or Ub-like modifiers to substrates through mediators such as E3 Ub ligases. Additionally, the results suggest a mechanism for kinetic control of the ubiquitination of diverse substrates for a given E2 enzyme as well as control of cross talk between pathways regulated by both Ub and Ub-like covalent modifications, such as those involved in postreplicative DNA repair (20).

Materials and methods

Kinetics of the Mms2-Ubc13 interaction from NMR spectroscopy: line-shape analysis

Protein expression and purification was conducted as previously described (27). 2D ^{1}H-^{15}N HSQC NMR spectra for the titration of 0.34 mM [U-^{15}N]-Ubc13 with unlabeled Mms2 were collected at 25 ℃ and 600 MHz. Protein concentration ratios (Ubc13:Mms2) for the titration were 8.3, 4.3, 2.9, 2.1, 1.6, 1.3, 1.1, and 0.9. The NMR sample contained 0.34 mM [U-^{15}N]-Ubc13 in 9:1 H_2O/D_2O, containing 150 mM HEPES, 75 mM NaCl, 10 mM DTT, 1 mM EDTA, 1 mM DSS (internal chemical shift reference), pH 7.5, in a 5 mm NMR tube. Protein concentrations were determined using the bicinchonininc acid (BCA) assay. The error in protein concentration using the BCA assay is 10%. Line-shape analysis for the titration was conducted using the Bloch-McConnell equations for two-site exchange (28), adjusted to include the application of a cosine-squared window function. In addition, to account for differential relaxation losses during the INEPT delays, the equations included a 2-fold loss of intensity for the bound state (Mms-Ubc13 complex), given the 2-fold increase in size compared to Mms2 alone. NMR spectra were numerically fit to the frequency domain expression derived from the Bloch-McConnell equations by optimizing the values of k_{off}, the bound chemical shift, intensity parameters for individual spectra, and a global intensity parameter. The association rate was $k_{on} = k_{off}/K_D$ with $K_D = 49$ nM, as previously determined (29). The error in k_{off} was determined by

line-shape analysis of the titration at the upper (1.1[Ubc13]$_{\text{total}}$, 0.9[Mms2]$_n$) and lower (0.9[Ubc13]$_{\text{total}}$, 1.1[Mms2]$_n$) error limits for protein concentrations, where [Mms2]$_n$ indicates the concentration of Mms2 at the n^{th} titration point.

Kinetics of the Mms2-Ubc13 interaction from NMR spectroscopy: ZZ-exchange

The rate of exchange of ^{15}N longitudinal magnetization between the free and Ubc13-bound states of Mms2 was characterized using 2D ZZ-exchange spectroscopy (30, 31). 2D ^{1}H-^{15}N HSQC NMR spectra were collected using mixing times of 0, 0.022, 0.044, 0.088, 0.143, 0.187, 0.253, 0.341, 0.528, 0.759, and 1.045 s at 600 MHz. NMR samples contained 350 µL of 0.65 mM [U-^{15}N]-Mms2 with 0.32 mM Ubc13 (2:1 ratio) in a mixed buffer containing 20 mM sodium phosphate, 20 mM Tris, 150 mM NaCl, 1 mM DTT at pH 7.0 in 9:1 H_2O/D_2O. The time dependence of the auto- and cross-peak intensities were fit to eq 3 in (32), wherein the pseudo-first-order rate constant (k_{on}[Ubc13]) for ligand binding was employed. The different relaxation properties for the free and Ubc13-bound states of Mms2 during the INEPT transfers were accounted for by optimization of parameters to adjust the intensities of the auto- and cross-peaks (32, 33). Errors for the kinetic parameters from ZZ-exchange were determined using 500 Monte Carlo trials of the fits using a 10% error in protein concentration, as determined using the BCA assay. ^{15}N-R_1 values for the main chain amides of free and Ubc13-bound Mms2 were collected as previously described; these values were used when fitting the ZZ-exchange data (34, 35).

Enzyme assays for Lys63-linked Ub_2 synthesis catalyzed by Mms2-Ubc13

All proteins other than E1 enzyme and acceptor Ub (purchased from Boston Biochem, Cambridge, MA) were expressed as GST fusion proteins and purified using GST affinity and size-exclusion chromatography, as previously described (27). The kinetics of polyUb chain formation catalyzed by Mms2-Ubc13 were characterized by analyzing enzyme assays with SDS-PAGE and employing donor Ub site-specifically labeled at the N-terminus

with the AlexaFluor 488 probe using maleimide chemistry. Briefly, given that Ub does not contain Cys residues, molecular cloning techniques were used to introduce an N-terminal Cys plus three residues (Leu-Gly-Ser) from Met1 of Ub. On the basis of a recent crystal structure of Mms2/Ubc13 with donor Ub attached to Ubc13 through an oxoester link (PDB ID 2GMI) (26), this location is not expected to interfere with donor Ub interactions with Mms2-Ubc13.

Single-turnover ubiquitination assays were conducted by using E1 to conjugate AlexaFluor 488 labeled Ub (donor Ub) to Ubc13 first at pH 7.3, then quenching the reaction with EDTA. Ub_2 chain synthesis was subsequently initiated by adjusting the pH to the desired value (typically 8.0 but variable for the pH-rate profile) and adding Mms2 and 49-197 µM Ub (acceptor Ub) after quenching the E1 reaction, allowing the kinetics of polyUb chain formation to be followed directly. It should be noted that Ub derived from pGEX-6P1 GST fusion constructs contain a GPLGS cloning artifact at the N-terminus of Ub. Employing acceptor Ub derived from this GST fusion construct in enzyme assays gives k_{cat} values that are underestimated by 4-fold. Therefore, we used wild-type human Ub purchased from Boston Biochem (Cambridge, MA) for enzyme assays. Reaction times for enzyme assays ranged from 2 to 30 min, and typical protein concentrations were 200 nM E1, 8 µM Ubc13, 10 µM Mms2, 10.7 µM Alexa Fluor 488 labeled Ub-K63R, 4 mM $MgCl_2$, and 4 mM ATP in buffer containing 150 mM NaCl and 37 mM BIS-TRIS propane. Enzyme reactions were quenched by flash freezing in liquid nitrogen. Concentrations of Ubc13 and Mms2 in the reaction mixture varied slightly depending on the amount of acceptor Ub added; these differences were accounted for when analyzing enzyme assays. The E1 reaction was quenched with 51 mM EDTA. The fraction of Ubc13 covalently conjugated to AlexaFluor 488 labeled Ub-K63R after reaction with E1 enzyme for 12 min varied between 15% and 60%. Gels were visualized using fluorescence at 517 nm with a Typhoon 9400 Imager (figure 4.11). Protein concentrations were determined by BCA assays; the quantity of Ubc13 that was conjugated to Ub in each assay was determined by using SYPRO Ruby total protein stain following fluorescent imaging of gels.

Analysis of the kinetics of Lys63-linked Ub_2 synthesis catalyzed by Mms2-Ubc13

To analyze enzyme kinetics, we used single-turnover assays to quantify the rate constant for Ub_2 formation catalyzed by human Mms2-Ubc13. The coupled differential equations describing the non-steady-state rate of Ub_2 formation and loss of Ubc13∼Ub thioester (Results section) were numerically integrated by setting values for k_{cat} and parameters to adjust for total Ub_2 (product) formed and total Ubc13∼Ub consumed, specific to each data set. To determine a global k_{cat}, values for these parameters were optimized by global minimization of the squared difference between four experimental enzyme assays and theoretical Ub_2 and Ubc13˜Ub concentrations using an in-house simulated annealing algorithm. Determination of the pH-rate profile for Mms2-Ubc13 was achieved by conducting enzyme assays at pH values of 7.0, 7.5, 7.75, 8.0, 8.5, and 9.0. The value of k_{cat} was numerically optimized for each individual pH as described above. The pH-dependent k_{cat} data were fit to the function $k_{cat} = (10^{pH} k_{cat,base})/(10^{pH} + 10^{pK_a})$, with $k_{cat,base}$ and pK_a as adjustable parameters (12). $k_{cat,base}$ is the rate constant corresponding to pH values for which the substrate Lys exists entirely as the neutral form (Lys-NH_2), with an associated pK_a. For the numerical integration and optimization, values for k_{on} and k_{off} of $(2.0 \pm 0.6) \times 10^7$ M^{-1} s^{-1} and 600 ± 200 s^{-1}, respectively, were used for the interaction of Ub with the Mms2-Ubc13 heterodimer (29). In addition, we assumed that acceptor and donor Ub molecules do not interact. This latter assumption is reasonable given recent crystal structures and chemical shift data (structures Ub_2, 2JF5, Mms2-Ubc13Ub, 2GMI; chemical shifts, ref 27). Furthermore, the kinetics of the Ubc13-Mms2 interaction are included in the numerical integration; these k_{on} and k_{off} values were determined as described above, with further details in the Results section. The error for the global value of k_{cat} and the values of k_{cat} from the pH-rate profile were determined using 25 Monte Carlo trials of the fits using a 10% error in the concentrations of Ubc13, Mms2, and wild-type acceptor Ub as well as the errors for the various k_{on} and k_{off} values for the interaction between Mms2 and Ub, Mms2-Ubc13 and Ub, and between Mms2 and Ubc13.

Results

E2 enzymes achieve large rate enhancements compared to the reaction in water

Modulation of the catalytic proficiency ($(k_{cat}/K_M)/k_{noncat}$) of E2 enzymes, through modest variations in k_{cat}, K_M, or both, likely represents a mechanism by which different E2 enzymes achieve different biological outcomes. While there have been numerous measurements of k_{cat} and K_M, for E2 enzymes, the catalytic proficiency remains unspecified. To determine the overall rate enhancement achieved by the enzyme, a noncatalyzed rate for thioester aminolysis under physiological conditions needs to be calculated to facilitate comparison to the enzyme-catalyzed rate. The mechanism for thioester aminolysis in water is widely considered to be a nucleophilic acyl substitution reaction (36, 37) (figure 4.1). The rate equation describing this reaction in aqueous solution and relevant to the reaction catalyzed by E2 enzymes is given by (11):

$$\frac{d[\mathrm{TE}]}{dt} = k_1[\mathrm{TE}][\mathrm{RNH_2}] + k_2[\mathrm{TE}][\mathrm{RNH_2}][\mathrm{OH^-}] \tag{4.1}$$

where TE is thioester (R'COSR"), $k_1 = 0.015$ $\mathrm{M^{-1}\ s^{-1}}$, and $k_2 = 13.6$ $\mathrm{M^{-2}\ s^{-1}}$. To facilitate comparison to the enzyme-catalyzed reaction, terms involving OH^--catalyzed hydrolysis of thioester and RNH_2-catalzyed thioester aminolysis are excluded. The biological reaction occurs at an intracellular pH of 7.2 and 37 ℃. For a thioester concentration of 10 µM, an estimate for the maximum intracellular concentration of thioester-charged E2 enzyme (38), and a total amine concentration (RNH_2 and RNH_3^+) of 20 µM, an estimate for the upper limit for the intracellular concentration of Ub (39, 40), eq 4.1 gives a rate of reaction of 7.7×10^{-15} M $\mathrm{s^{-1}}$ (pH 7.5 to facilitate comparison to the enzyme reaction). At pH 7.5, the RNH_3^+:RNH_2 ratio is 1100, for an amine pK_a of 10.54, using the equation:

$$[\mathrm{RNH_2}] = \frac{10^{\mathrm{pH}}[\mathrm{RNH_3^+} + \mathrm{RNH_2}]}{(10^{\mathrm{pH}} + 10^{\mathrm{p}K_a})} \tag{4.2}$$

To approximate the rate enhancement due to enzyme, we use for comparison the

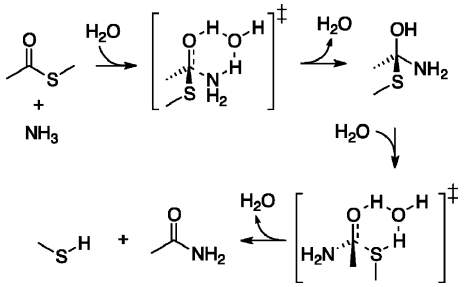

Figure 4.1: Stepwise reaction mechanism for aminolysis of methyl thioacetate

reaction catalyzed by the E2 Ubc9, with the substrate GST-RanGAP1 (an E2 with the largest known k_{cat}) (12). We assume that the mechanism in the enzyme is similar to that in water and given by the scheme in figure 4.2. It is typical to analyze the kinetics of E2-catalyzed reactions using the Michaelis-Menten equation under the assumption of steady-state kinetics (41)

$$v_i = \frac{d[P]}{dt} = \frac{[E_0][\mathrm{S}]k_{\mathrm{cat}}}{K_{\mathrm{M}} + [\mathrm{S}]} \tag{4.3}$$

for the enzymatic reaction

$$\mathrm{E} + \mathrm{S} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_{\mathrm{cat}}} \mathrm{E} + \mathrm{P} \tag{4.4}$$

where v_i is the initial rate, P is the product, E_0 is the total enzyme concentration, S is substrate, and K_{M} is given by

$$K_{\mathrm{M}} = \frac{k_{\mathrm{off}} + k_{\mathrm{cat}}}{k_{\mathrm{on}}} \tag{4.5}$$

K_{M} is approximately equal to K_{D}, the dissociation constant for the enzyme substrate complex, under the conditions $k_{\mathrm{cat}} \ll k_{\mathrm{off}}$. For the E2 Ubc9, with the substrate GST-RanGAP1, using Michaelis-Menten kinetics (eq 4.3) gives values of $k_{\mathrm{cat}} = 0.66\ \mathrm{s}^{-1}$ and $K_{\mathrm{M}} = 2.9\ \mu\mathrm{M}$ (12). Using these values with an enzyme concentration of 10 µM (thioester) and a substrate concentration of 20 µM in eq 4.3 gives a reaction rate of $5.8 \times 10^{-6}\ \mathrm{M\ s}^{-1}$. Thus, the enzymatic rate enhancement for these conditions is 7.5×10^8. Determination of the catalytic proficiency requires an understanding of how the activation barrier for the reaction in water is reduced in the enzyme active site; this is discussed in more detail in the subsequent sections.

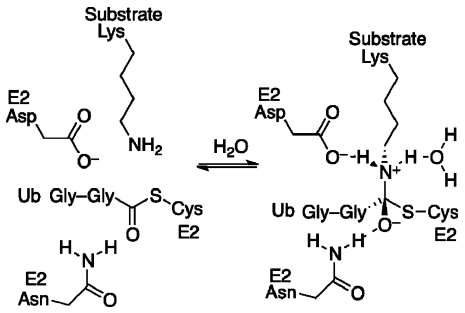

Figure 4.2: First step in a stepwise reaction mechanism for thioester aminolysis catalyzed by E2 enzymes. The transition state is shown on the right. For the serine proteases, the oxyanion is stabilized by two hydrogen bonds, whereas only one such H-bond has been identified for E2 enzymes. Proton transfer from the amine to the sulfur (likely facilitated by water), with cleavage of the C-S bond, completes the reaction. Active site E2 residues include a conserved Asn and Cys and an Asp/Glu proximal to the substrate Lys. The C-terminal Gly from Ub or Ubl modifiers is attached to the active site Cys of the E2 through a thioester bond, and Lys from the substrate performs the nucleophilic attack.

Overall activation barrier for the E2-catalyzed Reaction is smaller compared to the reaction in water

Numerous computational studies for enzyme reactions indicate that complementary electrostatic interactions between the charged transition state and the active site residues leads to a substantial reduction in the activation barrier for a reaction in an enzyme compared to the reaction within a water cage, that is, $\Delta\Delta G^{\ddagger}_{\rm cat} = \Delta G^{\ddagger}_{\rm cat} - \Delta G^{\ddagger}_{\rm cage} < 0$ (14). The first step of the thioester aminolysis reaction involves a proton transfer step from NH_4^+ or lysine NH_3^+ to a water molecule

$$NH_4^+ + H_2O \rightleftharpoons NH_3 + H_3O^+ \tag{4.6}$$

The free energy in solution is given by (42)

$$\Delta G^{S}_{\rm PT} = 2.3RT\{{\rm p}K_{\rm a}(NH_4^+) - {\rm p}K_{\rm a}(H_3O^+)\} \tag{4.7}$$

where R is the gas constant and T is the temperature. Using ${\rm p}K_{\rm a}(NH_4^+) = 10.54$, corresponding to the ${\rm p}K_{\rm a}$ for lysine, and ${\rm p}K_{\rm a}(H_3O^+) = -1$ in eq 4.7 gives $\Delta G^{\rm PT}_{\rm S} = 15.7$ kcal mol^{-1} for the proton transfer step in water.

For the second step of the reaction, the nucleophilic attack of ammonia on the

thioester carbonyl, the rate constant in eq 4.1 (k_2 = 13.6 M^{-2} s^{-1}) can be used to calculate an activation barrier using transition state theory (TST). The rate constant for a chemical reaction is given by (14, 15)

$$k = \kappa k_{\mathrm{TST}} = \kappa \left(\frac{k_{\mathrm{B}} T}{h} \right) \exp \left[-\frac{\Delta G^{\ddagger}}{RT} \right] \tag{4.8}$$

where κ is the transmission coefficient, R, k_{B}, and h are the gas, Boltzmann, and Planck constants, respectively, T is temperature, and $\Delta G^{\ddagger}$ is the free energy functional evaluated at the reaction coordinate value for the transition state at standard state concentration. Assuming that the transmission coefficient is 1, solving for $\Delta G^{\ddagger}$ gives the activation barrier in water (T = 25.6 ℃), $\Delta G^{\ddagger}$ = 15.9 kcal mol^{-1}. The calculation assumes a single transition state, whereas the overall reaction involves two transition states (figure 4.1). However, the largest barrier to the reaction is formation of the first transition state, such that eq 4.8 should provide a reasonable estimate for the rate of reaction in water. For example, the relative energy of the first transition state (figure 4.1) is calculated to be 10.9 kcal in water using electronic structure theory (37). Correcting this value for the entropic cost of bringing the reactants together in a solvent cage (2.4 kcal mol^{-1}) (17) gives a theoretical $\Delta G^{\ddagger}$ = 13.3 kcal mol^{-1}, in reasonable agreement with the experimental value of 15.9 kcal mol^{-1}. It should be noted, however, that the nature of the model compounds differs between the experimental and the theoretical studies.

To calculate the activation barrier for the reaction catalyzed by E2 enzyme ($\Delta G^{\ddagger}$), it is reasonable to assume that the transmission coefficient is similar in water and the enzyme, with magnitude 1 (14). For attachment of the Ub-like modifier SUMO to the substrate GST-RanGAP1 by the E2 Ubc9 (pH 7.5, 37 ℃), k_{cat} = 0.66 s^{-1}; thus, eq 8 gives $\Delta G^{\ddagger}$ = 18.4 kcal mol^{-1} for the overall reaction barrier for Ubc9, an E2 with the largest measured k_{cat} (12). Thus, the reduction in the activation barrier for this enzyme is $\Delta\Delta G^{\ddagger} = \Delta G^{\ddagger}_{\mathrm{cat}} - (\Delta G^{\ddagger}_{\mathrm{w}} + \Delta G^{\ddagger}_{\mathrm{PT}}) = 18.4 - (15.7 + 15.9) = -13.2$ kcal mol^{-1}. As discussed in further detail below, the contributing factors to $\Delta\Delta G^{\ddagger}_{\mathrm{cat}}$ are entropic effects of binding, electrostatic stabilization of the transition state, as well as suppression of the substrate lysine pKa through electrostatic effects.

Errors associated with measurement of $k_{\mathbf{cat}}$ and $K_{\mathbf{M}}$ for E2 enzymes using steady-state Michaelis-Menten kinetics

E2-catalyzed reactions are often studied using steady-state Michaelis-Menten kinetics (7, 12, 40, 43, 44). This approach has often been used to measure relative apparent catalytic rates. However, in a strict sense, the kinetics are non-steady state given that the enzyme, E2 thioester, is consumed during the reaction. To estimate the magnitude of the error associated with the steady-state assumption, we employ the following rate equations for E2-catalyzed reactions, wherein the substrate directly binds E2, to calculate reaction rates

$$\frac{d[\text{E2}\tilde{}\text{Ub}]}{dt} = k_{\text{off,S}}[\text{S}]][\text{E2}\tilde{}\text{Ub}] - k_{\text{on,S}} + k_{\text{off,S}}[\text{Ub}\tilde{}\text{S-E2}\tilde{}\text{Ub}] - k_{\text{on,S}}[\text{Ub}\tilde{}\text{S}][\text{E2}\tilde{}\text{Ub}] \quad (4.9)$$

$$\frac{d[\text{S-E2}\tilde{}\text{Ub}]}{dt} = -k_{\text{off,S}}[\text{S-E2}\tilde{}\text{Ub}] + k_{\text{on,S}}[\text{S}][\text{E2}\tilde{}\text{Ub}] - k_{\text{cat}}[\text{S-E2}\tilde{}\text{Ub}] \quad (4.10)$$

$$\frac{d[\text{Ub}\tilde{}\text{S-E2}\tilde{}\text{Ub}]}{dt} = -k_{\text{off,S}}[\text{Ub}\tilde{}\text{S-E2}\tilde{}\text{Ub}] + k_{\text{on,S}}[\text{Ub}\tilde{}\text{S}][\text{E2}\tilde{}\text{Ub}] \quad (4.11)$$

$$\frac{d[\text{E2}]}{dt} = k_{\text{off,S}}[\text{S}\tilde{}\text{E2}] - k_{\text{on,S}}[\text{S}][\text{E2}] + k_{\text{off,S}}[\text{Ub}\tilde{}\text{S-E2}] - k_{\text{on,S}}[\text{Ub}\tilde{}][\text{E2}] \quad (4.12)$$

$$\frac{d[\text{S-E2}]}{dt} = -k_{\text{off,S}}[\text{S-E2}] + k_{\text{on,S}}[\text{S}][\text{E2}] \quad (4.13)$$

$$\frac{d[\text{Ub}\tilde{}\text{S-E2}]}{dt} = -k_{\text{off,S}}[\text{Ub}\tilde{}\text{S-E2}] + k_{\text{on,S}}[\text{Ub}\tilde{}\text{S}][\text{E2}] + k_{\text{cat}}[\text{S-E2}\tilde{}\text{Ub}] \quad (4.14)$$

$$\frac{[\text{S}]}{dt} = k_{\text{off,S}}[\text{S-E2}\tilde{}\text{Ub}] - k_{\text{on,S}}[\text{S}][\text{E2}\tilde{}\text{Ub}] + k_{\text{off,S}}[\text{S-E2}] - k_{\text{on,S}}[\text{S}][\text{E2}] \quad (4.15)$$

$$\begin{aligned}\frac{d[\text{Ub}\tilde{}\text{S}]}{dt} = {} & k_{\text{off,S}}[\text{Ub}\tilde{}\text{S-E2}\tilde{}\text{Ub}] - k_{\text{on,S}}[\text{Ub}\tilde{}\text{S}][\text{E2}\tilde{}\text{Ub}] + k_{\text{off,S}}[\text{Ub}\tilde{}\text{S-E2}] \\ & - k_{\text{on,S}}[\text{Ub}\tilde{}\text{S}][\text{E2}]\end{aligned} \quad (4.16)$$

where S is substrate protein, $k_{\text{on,S}}$ and $k_{\text{off,S}}$ are the on and off rates of substrate binding to charged (E2∼Ub) and uncharged E2, respectively, k_{cat} is the catalytic rate constant, bullet points indicate a noncovalent interaction, and wavy lines indicate a thioester bond (E2∼Ub) or an amide bond for the final attachment of Ub to substrate. The corresponding equilibria for eqs 4.9 through 4.16 are shown in figure 4.3. These rate equations can be solved for the concentration of product (S∼Ub) under a given set of reaction

conditions, true rate constants, various substrate concentrations, and various times. The product build up for the various substrate concentrations can subsequently be linearly fit to determine simulated initial rates of substrate ubiquitination. These simulated initial rates can then be fit to eq 4.3 to yield fitted k_{cat} and K_M values. Differences between the true k_{cat} and K_M and the fitted values give the error due to the assumption of steady-state, Michaelis-Menten kinetics. For example, given an initial enzyme concentration (E2∼Ub, or E_0) of 21 nM, $K_D = k_{off}/k_{on} = 2.8$ µM ($k_{off} = 56$ s^{-1}, $k_{on} = 2 \times 10^7$ M^{-1} s^{-1}) for binding of substrate to E2, reaction times of 37, 74, 111, and 148 ms, and a true k_{cat} of 1.0 s^{-1}, we calculated initial rates of 2.7, 4.7, 6.3, and 7.7 nM s^{-1} at substrate concentrations of 0.5, 1.0, 1.5, and 2.0 µM, respectively (figure 4.4). Fitting these initial rates to the Michaelis-Menten equation (eq 4.3) gives a k_{cat} of 0.95 s^{-1} and a K_M of 3.1 µM, corresponding to errors of 4% and 17%, respectively, in comparison to the true values (figure 4.4). Changing the reaction times to 1, 2, 3, and 4 s gives a k_{cat} of 0.28 s^{-1} and a K_M of 0.8 µM, corresponding to errors of 72% and 71%, respectively (figure 4.4). This analysis gives a general idea of the accuracy of the steady-state Michaelis-Menten approach to E2-catalyzed reactions where the E2 binds substrate directly. Furthermore, we note that the maximum concentration of substrate must exceed K_M in order to ensure accuracy for this fitted parameter. To further illustrate these points, figure 4.5 shows the error in k_{cat} for various values of maximum reaction time (t_{max}) and k_{cat} for two types of E2 enzyme: one with tight substrate binding and another with weak substrate binding. Accuracy deteriorates with increasing k_{cat} and increasing t_{max}, the time over which initial rates are obtained.

For the rate equations of thioester aminolysis catalyzed by E2 enzymes (eq 9), the rate of OH^--catalyzed hydrolysis of thioester was ignored given that it is slow (11, 45). For example, the second-order rate constant at 26 ℃ is $k_{OH} = 0.53$ M^{-1} s^{-1} (11). Thus, at pH 7.5 and 37 ℃ and with a thioester concentration of 15 µM, the rate of hydrolysis is given by $k_{OH}[TE][OH^-] = 7 \times 10^{-12}$ M s^{-1}. This value compares favorably with the first-order rate of hydrolysis for the Ubc9SUMO thioester, $k = (5 \pm 3) \times 10^{-5}$ s^{-1} (45). This gives a rate of 8×10^{-10} M s^{-1} at pH 7.4, 37 ℃, and with a thioester concentration of 15 µM.

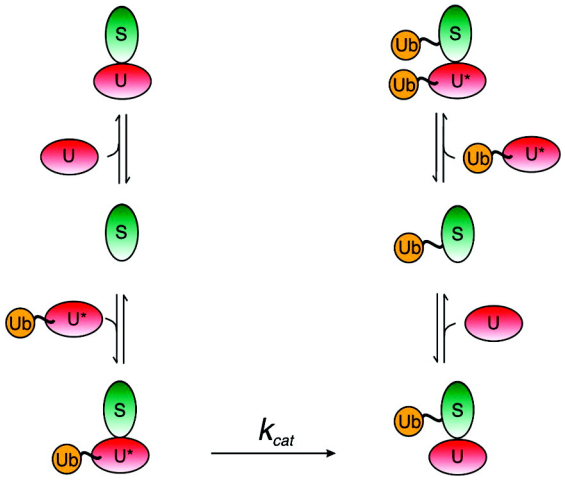

Figure 4.3: Kinetics for the direct ubiquitination of substrates by E2 enzymes. Substrate protein is indicated by an S and shown in green, Ubc13 covalently attached to Ub through a thioester bond (charged Ubc13) is indicated by U* and colored red, uncharged Ubc13 is indicated by a U and shown in red, and Ub is colored yellow.

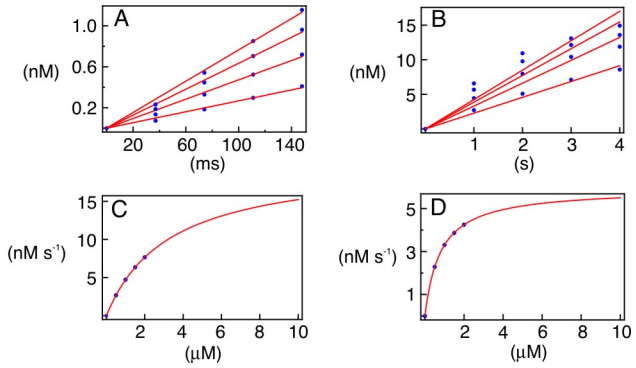

Figure 4.4: Steady-state Michaelis-Menten simulations for direct ubiquitination of substrates by E2 enzymes. Initial rates were determined from linear fits of short time scale (A) and long time scale reactions (B). The initial rates were fit to the Michaelis-Menten equation for the short (C) and long (D) time scale reactions.

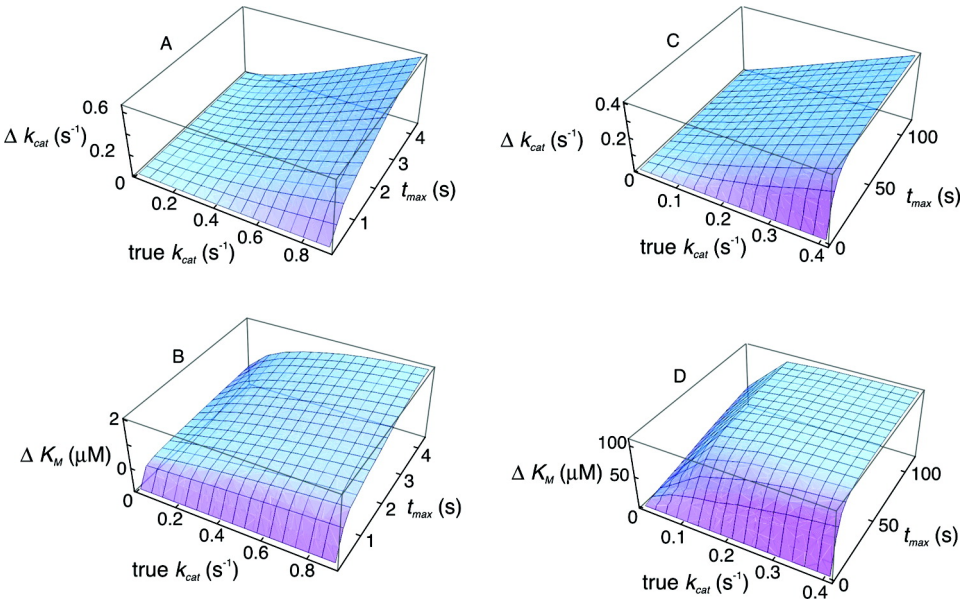

Figure 4.5: Errors associated with the assumption of steady-state Michaelis-Menten kinetics for E2-catalyzed reactions. Δk_{cat} = true k_{cat} − fitted k_{cat}, ΔK_{M} = true K_{D} − fitted K_{M}, and t_{max} is the maximum time over which initial rates were taken in steps of $t_{\mathrm{max}}/4$ (figure 4.4A and 4.4B). The initial enzyme concentration, E_0 or E2∼Ub, was 21 nM, the K_{D} for substrate binding to E2 was 2.8 μM ($k_{\mathrm{off}} = 56\ \mathrm{s}^{-1}$, $k_{\mathrm{on}} = 2 \times 10^7\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$), the substrate concentrations were 0.5, 1.0, 1.5, and 2.0 μM for A and B, whereas in C and D E_0 was 21 nM, the K_{D} for substrate binding to E2 was 110 μM ($k_{\mathrm{off}} = 2200\ \mathrm{s}^{-1}$, $k_{\mathrm{on}} = 2 \times 10^7\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$), and the substrate concentrations were 0-400 μM in steps of 10 μM. Inclusion of product (ubiquitinated substrate) binding to E2 or E2∼Ub in the rate equations has a negligible effect on Δk_{cat} and ΔK_{M} under the specified conditions and with the assumption that K_{D} remains the same as substrate binding.

Non-steady-state kinetics underlying assembly of Lys63-linked polyUb chains

Mms2 and Ubc13 form a noncovalent enzyme complex, and the kinetics of this interaction are not accounted for in the basic Michaelis-Menten enzymatic reaction (eqs 4.3 to 4.5). To derive the rate equations for the chemical reaction catalyzed by Mms2-Ubc13, the starting point is a description of the equilibria for the interaction of Mms2 and Ubc13 with each other and Ub (figure 4.6).

$$\mathrm{Mms_2} + \mathrm{Ub} \underset{k_{\mathrm{off,Ub}}}{\overset{k_{\mathrm{on,Ub}}}{\rightleftharpoons}} \mathrm{Mms_2{-}Ub} \tag{4.17}$$

$$\mathrm{Mms_2{-}Ubc_{13}} + \mathrm{Ub} \underset{k_{\mathrm{off,Ub}}}{\overset{k_{\mathrm{on,Ub}}}{\rightleftharpoons}} \mathrm{Ub{-}Mms_2{-}Ubc_{13}} \tag{4.18}$$

$$\text{Mms2-Ubc13}\tilde{}\text{Ub} \underset{k_{\mathrm{off,Ub}}}{\overset{k_{\mathrm{on,Ub}}}{\rightleftharpoons}} \text{Ub-Mms2-Ubc13}\tilde{}\text{Ub} \tag{4.19}$$

$$\text{Mms2} + \text{Ubc13} \underset{k_{\mathrm{off,MU}}}{\overset{k_{\mathrm{on,MU}}}{\rightleftharpoons}} \text{Mms2-Ubc13} \tag{4.20}$$

$$\text{Mms2} + \text{Ubc13}\tilde{}\text{Ub} \underset{k_{\mathrm{off,MU}}}{\overset{k_{\mathrm{on,MU}}}{\rightleftharpoons}} \text{Mms2-Ubc13}\tilde{}\text{Ub} \tag{4.21}$$

where bullet points indicate a noncovalent interaction and wavy lines indicate a thioester bond. The K_{D} ($k_{\mathrm{off,Ub}}/k_{\mathrm{on,Ub}}$) for binding of substrate or acceptor Ub by Mms2 is 98 and 28 μM for free and Ubc13-bound Mms2, respectively, determined from NMR chemical shift titrations (29). Given the lack of interaction between donor and substrate Ub molecules (26, 27), we assume that these K_{D}s are the same when a donor Ub is covalently attached to the active site cysteine of Ubc13. The K_{D} ($k_{\mathrm{off,MU}}/k_{\mathrm{on,MU}}$) for binding of Mms2 and Ubc13 was previously determined to be 49 nM using ITC (29). We assume that the magnitude of K_{D} for this high-affinity interaction remains the same in the presence of Ub covalently attached to Ubc13 or Ub noncovalently bound to Mms2. In addition to these thermodynamic measurements, the rates of protein-protein association/dissociation for the interaction between Mms2 and Ub were previously measured using NMR line-shape analysis (29). The values of $k_{\mathrm{on,Ub}}$ and $k_{\mathrm{off,Ub}}$ were determined to be $(2.0{\pm}0.5){\times}10^7$ $\mathrm{M^{-1}\ s^{-1}}$ and $2250 \pm 500\ \mathrm{s^{-1}}$, respectively, with an upper limit for k_{off} of $600 \pm 200\ \mathrm{s^{-1}}$ for binding of Ub to the Mms2-Ubc13 complex (29). This upper limit was determined by

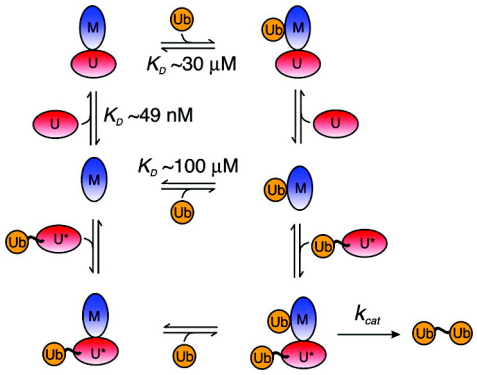

Figure 4.6: Complex kinetics for the Mms2-Ubc13 E2 enzyme system for synthesis of Lys63-linked Ub_2. Mms2 is indicated by an M and shown in blue, Ubc13 covalently attached to Ub through a thioester bond (charged Ubc13) is indicated by U* and colored red, uncharged Ubc13 is indicated by a U and shown in red, and Ub is colored yellow.

assuming that k_{on} for the interaction of Ub with the Mms2-Ubc13 complex is the same as that for the interaction of Ub with Mms2; the different K_D values for Mms2-Ub and Mms2-Ubc13-Ub give a decrease in k_{off} ($k_{off} = k_{on} \times K_D$).

In this study, we determined the kinetics of the interaction between Mms2 and Ubc13 ($k_{on,MU}$ and $k_{off,MU}$) using NMR line-shape analysis and ZZ-exchange (figure 4.7). The previously measured K_D of 49 nM for the interaction of Mms2 with Ubc13 was used in NMR line-shape analysis ($k_{on} = k_{off}/K_D$) to yield an optimized k_{off} of 4.4 s^{-1} for Mms2-Ubc13 binding. Assuming a 10% error on the concentrations of Mms2 and Ubc13 yields upper and lower limits for k_{off} of 15 and 5.8×10^{-10} s^{-1}, respectively. The lower limit appears physically unrealistic; however, assuming that k_{on} is diffusion controlled with a lower limit of 10^5 M^{-1} s^{-1} (46) gives a corresponding lower limit for k_{off} of 0.005 s^{-1}. From the ZZ-exchange measurements, the values for k_{on} and k_{off} are $(1.7 \pm 0.5) \times 10^8$ M^{-1} s^{-1} and 4.4 ± 0.3 s^{-1} respectively, for the Mms2-Ubc13 interaction.

Derivation of the rate equations governing non-steady-state kinetics for assembly of Lys63-linked polyUb chains

Using the reaction scheme for the Mms2-Ubc13 system shown in figure 4.6 and eqns 4.17 to 4.21, the time-dependent changes in the concentrations of the various protein species for catalysis of Lys63-linked polyUb chains by the Mms2-Ubc13 heterodimeric E2 enzyme

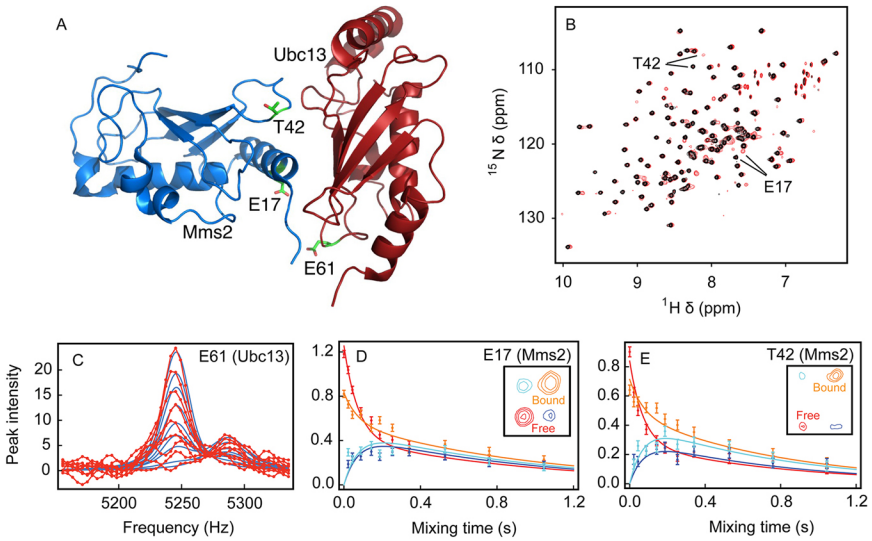

Figure 4.7: NMR measurements for the kinetics of the Mms2-Ubc13 interaction. (A) Structure for the Mms2-Ubc13 complex (PDB ID 1J7D). Residues at the interface undergoing exchange due to the protein-protein interaction are indicated. (B) NMR spectra of free [$U-^{15}$N]-Mms2 (black) and [$U-^{15}$N]-Mms2 bound to Ubc13 (red). Residues located at the binding interface are indicated. (C) NMR line-shape analysis for the interaction between Mms2 and Ubc13. Experimental data for [$U-^{15}$N]-Ubc13 E61 $^{1}H^{N}$ upon titration with Mms2 are shown in red, and simulations are shown in blue. ^{15}N ZZ-exchange profiles for E17 (D) and T42 (E) from [$U-^{15}$N]-Mms2 in a 2:1 ratio with Ubc13. Peak intensity profiles for free and bound peaks are indicated by red and yellow circles, respectively, with the best fits shown as lines. Intensity profiles for cross-peaks between the free and the bound states are shown as dark and light blue circles, with the best fits shown as lines. The insets show the ZZ-spectra taken at mixing times of 143 ms for the auto- and cross-peaks of E17 and T42 from free and Ubc13-bound Mms2.

are given by

$$\begin{aligned}\frac{d[\text{Mms2}]}{dt} &= k_{\text{off,Ub}}[\text{Mms2-Ub}] - k_{\text{on,Ub}}[\text{Mms2}][\text{Ub}] \\ &+ k_{\text{off,MU}}[\text{Mms2-Ubc13˜Ub}] - k_{\text{on,MU}}[\text{Mms2}][\text{Ubc13˜Ub}] \\ &+ k_{\text{off,MU}}[\text{Mms2-Ubc13}] - k_{\text{on,MU}}[\text{Mms2}][\text{Ubc13}]\end{aligned} \tag{4.22}$$

$$\begin{aligned}\frac{d[\text{Mms2-Ub}]}{dt} &= k_{\text{on,Ub}}[\text{Mms2}][\text{Ub}] - k_{\text{off,Ub}}[\text{Mms2-Ub}] \\ &+ k_{\text{off,MU}}[\text{Ub-Mms2-Ubc13˜Ub}] - k_{\text{on,MU}}[\text{Mms2-Ub}][\text{Ubc13˜Ub}] \\ &+ k_{\text{off,MU}}[\text{Ub-Mms2-Ubc13}] - k_{\text{on,MU}}[\text{Mms2-Ub}][\text{Ubc13}]\end{aligned} \tag{4.23}$$

$$\begin{aligned}\frac{d[\text{Ubc13˜Ub}]}{dt} &= k_{\text{off,MU}}[\text{Ub-Mms2-Ubc13˜Ub}] - k_{\text{on,MU}}[\text{Mms2-Ub}][\text{Ubc13˜Ub}] \\ &+ k_{\text{off,MU}}[\text{Mms2-Ubc13˜Ub}] - k_{\text{on,MU}}[\text{Mms2}][\text{Ubc13˜Ub}]\end{aligned} \tag{4.24}$$

$$\begin{aligned}\frac{d[\text{Mms2-Ubc13˜Ub}]}{dt} &= k_{\text{on,MU}}[\text{Mms2}][\text{Ubc13˜Ub}] - k_{\text{off,MU}}[\text{Mms2-Ubc13˜Ub}] \\ &- k_{\text{on,Ub}}[\text{Mms2-Ubc13˜Ub}][\text{Ub}] + k_{\text{off,Ub}}[\text{Ub-Mms2-Ubc13˜Ub}]\end{aligned} \tag{4.25}$$

$$\begin{aligned}\frac{d[\text{Ub-Mms2-Ubc13˜Ub}]}{dt} &= k_{\text{on,MU}}[\text{Mms2-Ub}][\text{Mms2-Ubc13˜Ub}] \\ &- k_{\text{off,MU}}[\text{Ub-Mms2-Ubc13˜Ub}] \\ &+ k_{\text{on,Ub}}[\text{Mms2-Ubc13˜Ub}][\text{Ub}] \\ &- k_{\text{off,Ub}}[\text{Ub-Mms2-Ubc13˜Ub}] \\ &- k_{\text{cat}}[\text{Ub-Mms2-Ubc13˜Ub}]\end{aligned} \tag{4.26}$$

$$\begin{aligned}\frac{d[\text{Mms2-Ubc13}]}{dt} &= k_{\text{off,Ub}}[\text{Mms2}][\text{Ubc13}] - k_{\text{off,MU}}[\text{Mms2-Ubc13}] \\ &- k_{\text{on,Ub}}[\text{Mms2-Ubc13}][\text{Ub}] + k_{\text{off,Ub}}[\text{Ub-Mms2-Ubc13}] \\ &+ k_{\text{cat}}[\text{Ub-Mms2-Ubc13˜Ub}]\end{aligned} \tag{4.27}$$

$$\begin{aligned}\frac{d[\text{Ub-Mms2-Ubc13}]}{dt} &= k_{\text{on,MU}}[\text{Ub-Mms2}][\text{Ubc13}] - k_{\text{off,MU}}[\text{Ub-Mms2-Ubc13}] \\ &+ k_{\text{on,Ub}}[\text{Mms2-Ubc13}][\text{Ub}] - k_{\text{off,Ub}}[\text{Ub-Mms2-Ubc13}]\end{aligned} \tag{4.28}$$

$$\frac{d[\text{Ubc13}]}{dt} = k_{\text{off,MU}}[\text{Ub-Mms2-Ubc13}] - k_{\text{on,MU}}[\text{Mms2-Ub}][\text{Ubc13}] + k_{\text{off,MU}}[\text{Mms2-Ubc13}] - k_{\text{on,MU}}[\text{Mms2}][\text{Ubc13}] \tag{4.29}$$

$$\begin{aligned}\frac{d[\text{Ub}]}{dt} = {} & k_{\text{off,Ub}}[\text{Mms2-Ub}] - k_{\text{on,Ub}}[\text{Mms2}][\text{Ub}] \\ & + k_{\text{off,Ub}}[\text{Ub-Mms2-Ubc13}\tilde{}\text{Ub}] - k_{\text{on,Ub}}[\text{Ub}][\text{Mms2-Ubc13}\tilde{}\text{Ub}] \\ & + k_{\text{off,Ub}}[\text{Ub-Mms2-Ubc13}] - k_{\text{on,Ub}}[\text{Mms2-Ubc13}][\text{Ub}]\end{aligned} \tag{4.30}$$

$$\frac{d[\text{Ub}_2]}{dt} = k_{\text{cat}}[\text{Ub-Mms2-Ubc13}\tilde{}\text{Ub}] \tag{4.31}$$

where the bullet points indicate a noncovalent interaction and the wavy lines indicate a thioester bond for E2∼Ub or an amide bond for Ub_2. In principle, an analytical solution for these coupled differential equations can be used to obtain expressions for the concentrations of the various protein species, which can then be used to fit experimental data. However, an analytical solution was not possible; thus, analysis of the kinetics was accomplished by numerical integration of the rate equations after choosing a value for k_{cat} and including experimentally determined values for $k_{\text{on,MU}}$, $k_{\text{off,MU}}$, $k_{\text{on,Ub}}$, and $k_{\text{off,Ub}}$ to obtain theoretical protein (Ub_2 and total Ubc13∼Ub) concentrations. This procedure was repeated by varying the value of k_{cat} using a simulated annealing algorithm to minimize the sum of the squared differences between theoretical and experimental protein concentrations, giving a rate for the chemical step of 0.007 ± 0.001 and 0.0020 ± 0.0004 s^{-1} at pH 8.0 and 7.5, respectively (Figures 4.8 and 4.9). Additionally, we measured k_{cat} at various pH values for the reaction to determine a ΔpK_a of -2.3 ± 0.1 for the substrate Lys of acceptor Ub (Figure 4.8).

Discussion

Activation barrier for the E2-catalyzed reaction is substantially reduced compared to the reaction in water

E2 enzymes are involved in the regulation of a myriad of cellular processes ranging from trafficking, to protein degradation, the DNA damage response, control of the cell cycle, and DNA repair (3). One of the central questions regarding E2 enzyme activity is

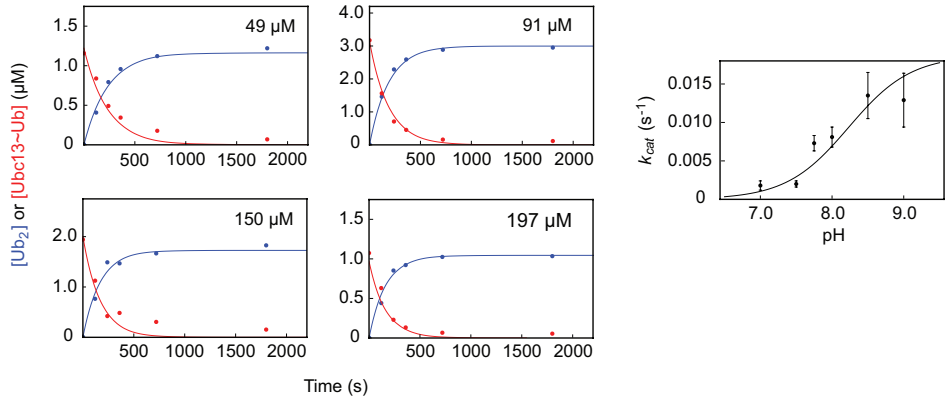

Figure 4.8: (Left) Kinetics of Ub_2 synthesis catalyzed by the E2 heterodimer Mms2-Ubc13. Enzyme assays involved following the decrease in Ubc13∼Ub thioester (red) or the increase in Lys63-linked Ub_2 (blue). Concentrations of Ub substrate are indicated in the panels for the four separate enzyme assays. Solid lines indicate global fits of the data to the integrated rate equations (eqns 4.22 to 4.31). (Right) pH-rate profile for catalysis of the synthesis of Ub_2 by Mms2-Ubc13.

whether biological diversity can be achieved through variability in chemical mechanism. Thus, the problem to be addressed is identification of the fundamental source for the catalytic power of these enzymes. This is established by assuming that the reaction carried out by enzyme follows a mechanism similar to that in water; for bimolecular reactions, the second-order rate constant in water is compared to the enzyme-catalyzed rate (9, 10). However, such comparisons can be complicated as a result of different reaction conditions, the nature of the model compounds with respect to biological substrates, and complex kinetics in enzyme-catalyzed multisubstrate reactions (47). For reactions involving model compounds relevant to the chemistry of ubiquitination, kinetic studies of aminolysis of thioester bonds in water reveal that the reaction depends on pH and amine concentration and is catalyzed by OH^- (11). From the rate constant reported in this early work (k_2 in eq 4.1) and the free energy of proton transfer from ammonia to water (eq 4.7), the total activation barrier is $\Delta G^{\ddagger}_{w} \approx 31.6$ kcal mol^{-1} with a corresponding $k_{noncat} \approx 3.5 \times 10^{-10}$ s^{-1} (eq 4.8). For the SUMO-GST-RanGAP1-Ubc9 reaction, the apparent k_{cat} is 0.66 s^{-1} (12) (one of the largest k_{cat} values for E2 enzymes), giving $k_{cat}/k_{noncat} = 1.9 \times 10^9$, with a catalytic proficiency $(k_{cat}/K_M)/k_{noncat} = 6.5 \times 10^{14}$ M^{-1} at pH 7.5 and 37 ℃. Thus, E2 enzymes provide a substantial enhancement to the rate of thioester aminolysis that occurs in biological systems insofar as the reaction between *p*-nitrothiolbenzoate and

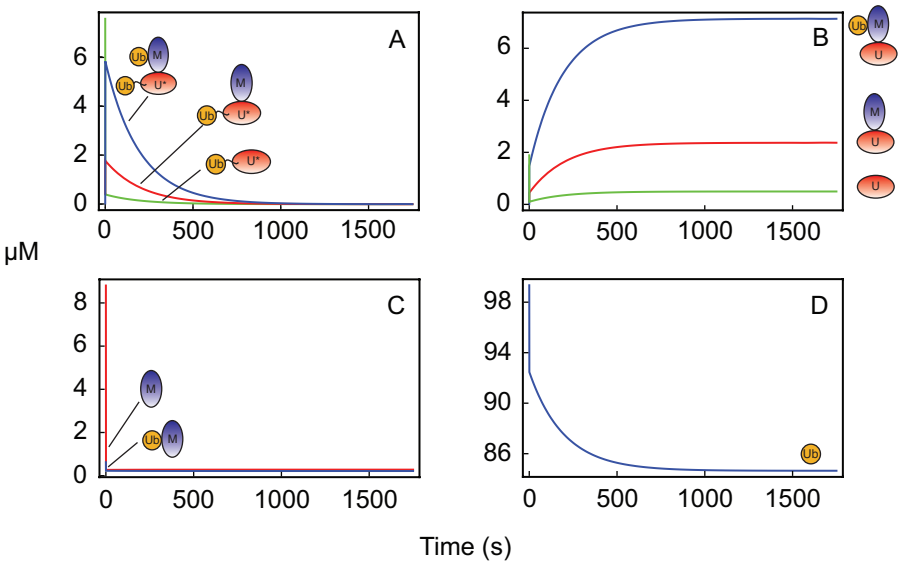

Figure 4.9: Theoretical curves for the kinetics of Ub_2 synthesis catalyzed by Mms2-Ubc13. (A) Concentrations of various Ubc13∼Ub thioester or charged species: Mms2-Ubc13∼Ub (red), Ub-Mms2-Ubc13∼Ub (blue), Ubc13∼Ub (green). (B) Concentrations of various uncharged Ubc13 species: Mms2-Ubc13 (red), Ub-Mms2-Ubc13 (blue), Ubc13 (green). (C) Concentrations of various Mms2 species: free Mms2 (red), Mms2-Ub (blue). (D) Concentration of free Ub. Curves were generated from global fitting of enzyme assays to the integrated rate equations (eqns 4.22 to 4.31) with the following parameters: $k_{cat} = 0.007\ s^{-1}$, $k_{on} = 2 \times 10^7\ M^{-1}\ s^{-1}$ and $k_{off} = 2250\ s^{-1}$ for Ub binding to Mms2, $k_{on} = 2 \times 10^7\ M^{-1}\ s^{-1}$ and $k_{off} = 560\ s^{-1}$ for Ub binding to the Mms2-Ubc13 complex, $k_{on} = 1.7 \times 10^8\ M^{-1}\ s^{-1}$ and $k_{off} = 4.4\ s^{-1}$ for Ubc13 binding to Mms2, the concentration of Ub substrate was 100 μM, the total concentration of Mms2 and Ubc13 was 10 μM each, and the total concentration of thioester charged Ubc13 was 80% of the total Ubc13 concentration.

n-butylamine is representative of the biological reaction. For Ubc9, there is variability in k_{cat} for different substrates. In comparison to the reaction with GST-RanGAP1, k_{cat} is 0.021 and 0.027 s^{-1} (pH 7.5 and 37 ℃) for the substrates p53 (C-terminal tetramerization domain) and GST-PML, respectively, determined using steady-state (p53) and non-steady-state (GST-PML) approaches (12, 48). The rate enhancement for these substrates is $(6-8)\times10^7$, with a catalytic proficiency of $(0.9-15)\times10^{11}$ M^{-1}. Thus, the catalytic proficiency of E2 enzymes is modest, occurring at the lower end of the spectrum in comparison to a number of other enzymes (10). In comparison, k_{cat}/k_{noncat} is 5.7×10^6 for the Mms2-Ubc13 holoenzyme at pH 7.5 and 37 ℃. This enhancement corresponds to a $\Delta\Delta G^{\ddagger}_{cat}$ of -9.6 kcal mol^{-1} and a catalytic proficiency of 1.8×10^{11} M^{-1}. The k_{cat} value for Mms2-Ubc13 at pH 7.5 differs 14-fold compared to the rate of attachment of SUMO to the protein substrate GST-PML, catalyzed by the E2 Ubc9 ($k_{cat}\approx0.027$ s^{-1}, pH 7.5 and 37 ℃) (48). Importantly, the kinetics for the GST-PML/Ubc9 and Ub/Mms2-Ubc13 reactions are directly comparable as they are the only E2-catalyzed reactions analyzed using appropriate non-steady-state approaches.

Suppression of the substrate lysine pK_a provides only part of the reduction to the activation barrier for the enzyme-catalyzed reaction compared to the reaction in water

The simplest mechanism to account for the catalytic function of E2 enzymes involves lowering the pK_a of the substrate lysine to allow deprotonation and subsequent nucleophilic attack of the thioester carbonyl (12, 27). However, this mechanism must be reconciled with generalizations from computational studies that indicate a key source for the catalytic power of enzymes is electrostatic complementarity within the active site (14, 15). Specifically, for attachment of the Ub-like modifier SUMO to the substrate RanGAP1 by the E2 Ubc9, the pH-rate profile for the reaction indicates that the pK_a of the substrate lysine is decreased by -4.1 units (12). This magnitude of ΔpK_a corresponds to a decrease in the free energy of the proton transfer step from the substrate lysine to water by 5.6 kcal mol^{-1} ($1\Delta pK_a = 1.36$ kcal mol^{-1}) or an 8000-fold increase in reaction rate using TST (eq 4.8). Changes in substrate lysine pKa compared to the values in water

for E2-catalyzed reactions are essentially electrostatic in nature and reflect changes in the solvation of the ionizable side chain between water and the polar environment of the enzyme active site (eq 13 in (49), for example).

For Mms2-Ubc13 and Ubc9, ΔpK_a shifts of -2.6 and -4.1 units, respectively, provide maximum contributions to the catalytic rate enhancement of 200- and 8000-fold (or 3.2 and 5.6 kcal mol^{-1} to $\Delta\Delta G^{\ddagger}_{cat}$). Thus, suppression of substrate lysine pK_a represents only part of the total $\Delta\Delta G^{\ddagger}_{cat}$ values of -9.6 and -13.2 kcal mol^{-1} for Mms2-Ubc13 and Ubc9, respectively.

Electrostatic complementarity and entropic effects play roles in reducing the activation barrier for the enzyme-catalyzed reaction in comparison to the reaction in water

In addition to substrate lysine pK_a suppression, entropic effects of substrate binding (16) and electrostatic complementarity (14) provide additional contributions to $\Delta\Delta G^{\ddagger}_{cat}$ for E2 enzymes. Recently, computational studies for a number of enzyme-catalyzed reactions have led to the generalization that enzymes employ electrostatic interactions to stabilize transition states and/or intermediates; the barrier to a reaction is lowered by electrostatic effects from the preorganized, polar environment of the enzyme active site (14, 15). It has been argued that the free energy of binding two reactants can pay for the entropic cost of forming a transition state (16), and this mechanism was considered to be a major catalytic effect in enzymes. More recent views do not consider this mechanism to be as effective as electrostatic complementarity (17, 50). For example, it is estimated that entropic effects may contribute only 2.5 kcal mol^{-1} to catalytic rate enhancement for the enzymatic hydrolysis of a dipeptide using molecular dynamics simulations (17). Furthermore, small molecule catalysts that exploit entropy effects do not reach the efficiency of enzymes (51), whereas small molecule catalysts designed to exploit electrostatic complementarity in reduced polarity solvent produce astounding rate accelerations (52).

With respect to the reaction catalyzed by E2 enzymes, computational studies for small molecules using ab initio electronic structure theory have been employed to model the chemical mechanism (37, 53). The key features of the reaction indicate that catalysis

by water facilitates proton transfer from the nucleophilic amine to the thioester carbonyl oxygen and finally the thioester sulfur, proceeding through two transition states and a tetrahedral intermediate. Indeed, elegant isotope exchange experiments are regarded as definitive proof of the existence of the tetrahedral intermediate and the validity of this mechanism (36, 54). From the computational studies, the largest barrier to the reaction involves formation of the first tetrahedral transition state. For the uncatalyzed reaction in water, the charged transition state is stabilized by a water molecule that bridges opposite charges (figure 4.1). For E2 enzymes, mutational studies have led to the suggestion that stabilization of the oxyanion transition state occurs through a conserved Asn (55). Furthermore, it is likely that an active site Asp residue stabilizes the developing positive charge on the attacking lysine. Stabilization of this transition state in the enzyme active site by hydrogen bonding to the oxyanion and complementing the developing positive charge on the nucleophilic lysine nitrogen may play an important role in the catalytic power of E2 enzymes such as Ubc9 and Ubc13 (figure 4.2). Indeed, mutation of a number of charged residues surrounding the substrate lysine in Ubc9 and Ubc13 have been shown to lead to qualitative reductions in the catalytic rate (12, 26, 56). These observations are consistent with a key role for electrostatic complementarity in the function of E2 enzymes. Combining entropic effects and pK_a suppression gives estimates for $\Delta\Delta G^{\ddagger}_{cat}$ of -5.7 and -8.1 kcal mol^{-1} for Mms2-Ubc13 and Ubc9, respectively; underestimates in comparison to the observed values of -9.6 and -13.2 kcal mol^{-1}, these values indicate that electrostatic stabilization of the transition state for E2 enzymes contributes an additional 4-5 kcal mol^{-1} of stabilization or a 2000-3000-fold rate enhancement. Variations in k_{cat} due to differences in electrostatic effects such as pK_a suppression and/or transition state stabilization represent a reasonable molecular basis for the observation that a 10-fold difference in the rate of attachment for the first Ub to a substrate has a large influence on the subsequent fraction of substrate that develops polyUb chains long enough to signal for degradation by the proteasome (8).

Steady-state approaches for determining k_{cat} and K_M for E3-mediated attachment of Ub to substrates

Determination of apparent k_{cat} and K_M values using the steady-state assumption with Michaelis-Menten kinetics is potentially inaccurate for kinetically complicated reactions such as E3-mediated transfer of Ub to substrates wherein both E2 and substrate bind E3. In particular, the weaker an E3 binds substrate, the greater the inaccuracy, especially for rate measurements at low substrate concentration. This has important consequences for k_{cat} measurements of E2 catalytic activity in the presence of E3s, given that these ubiquitin ligases bind their cognate E2s weakly, with micromolar dissociation constants. The effect of weak binding between Mms2 and Ubc13 or an analogous E3-E2 interaction will lead to apparent k_{cat} values that are lower than the true k_{cat} (Figure 10). For E2 enzymes that can directly bind substrate, these apparent values can potentially allow for relative comparisons between different substrates, such as attachment of the Ub-like modifier SUMO to p53 and GST-RanGAP1 by the E2 Ubc9 (12). However, as shown in figure 4.5, the inaccuracy for k_{cat} and K_M increases as k_{cat} increases. If initial rates are determined by linear fitting of kinetic data and subsequently analyzed using the Michaelis-Menten equation, care must be taken to ensure that reactions with larger k_{cat} values (0.05 s^{-1}) are rapidly sampled on the millisecond time scale. In addition, the rates of direct substrate modification catalyzed by E2 can be enhanced by E3 (E3:E2 Nup358:Ubc9, with substrate RanBP2) (57), and the complex kinetics for this multi-substrate reaction will give rise to inaccurate apparent kinetic parameters under the assumption of simpler kinetic schemes, as discussed above.

Conclusion

E2 enzymes play crucial roles in regulating a wide variety of life processes by catalyzing the covalent addition of Ub or Ub-like modifiers to target proteins, thereby modifying the function of the target or signaling for its degradation. Thus, it is important to determine if the biological outcome of ubiquitination can be regulated by variations in the chemical mechanism for different E2 enzymes, as manifested in differences in k_{cat} and K_M. We determined that the catalytic proficiency of E2 enzymes is modest $10^{11} - 10^{14}$ M^{-1} with

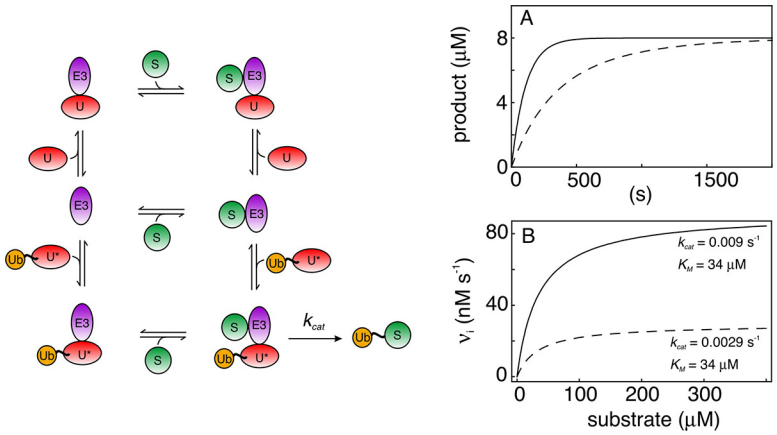

Figure 4.10: E3-mediated substrate ubiquitination by E2 enzymes is shown in the left panel and is essentially the same as the reaction scheme for catalysis of the synthesis of Ub_2 by Mms2-Ubc13 (figure 4.6). In the right panel, the rate equations (eqs 4.22 to 4.31) for catalysis of the synthesis of Ub_2 by Mms2-Ubc13 are applied to the E3-mediated formation of ubiquitinated substrate. (A) The solid curve shows the theoretical dependence of the rate of formation of product (substrateUb) on the strength of the E2-E3 interaction for $k_{\rm cat} = 0.01$ s^{-1} and $K_{\rm D} = 33$ nM for the E3-E2 interaction ($k_{\rm on} = 1.5\times10^8$ M^{-1} s^{-1} and $k_{\rm off} = 5$ s^{-1}), whereas the dashed curve illustrates substrate$\sim$Ub formation upon increasing $K_{\rm D}$ to 17 μM ($k_{\rm on} = 1.5 \times 10^8$ M^{-1} s^{-1} and $k_{\rm off} = 2500$ s^{-1}). The $K_{\rm D}$ for the E3-substrate interaction was 30 μM ($k_{\rm on} = 2.0\times10^7$ M^{-1} s^{-1} and $k_{\rm off} = 300$ s^{-1}). (B) Analysis of product build-up under the assumption of steady-state Michaelis-Menten kinetics gives a 3-fold underestimate for $k_{\rm cat}$ using an E3-E2 $K_{\rm D}$ of 17 μM (dashed curve) compared to a $K_{\rm D}$ of 33 nM (solid curve). Simulated initial rates were calculated in a similar fashion as those for direct ubiquitination of substrates (Results and figure 4.4A), with $t_{\rm max} = 10$ s.

a magnitude for $\Delta\Delta G^{\ddagger}_{\mathrm{cat}}$ from -10 to -13 kcal mol^{-1}. The factors that contribute to the catalytic enhancement include substrate lysine pK_{a} suppression (3-5 kcal mol^{-1}), entropic effects (3 kcal mol^{-1}), and electrostatic complementarity (4-5 kcal mol^{-1}). A recent kinetic model for substrate polyubiquitination by the E2 Cdc34 in combination with the E3 SCF demonstrates that a 10-fold difference in the rate of attachment of the first Ub to a substrate can give rise to a substantial difference in the fraction of substrate-bearing polyUb chains long enough to be degraded by the proteasome (8). Such a modest change in rate can be achieved by modest differences in k_{cat}, K_{M}, or both. Interestingly, slight changes in any of the mechanisms for catalysis by E2 enzymes can give rise to variations in k_{cat} and lead to the 1000-fold difference in proficiency among this family of enzymes. Tuning of the catalytic proficiency of different E2s or E2-E3 combinations likely plays a key role in determining the specificity and biological function of E2-catalyzed reactions.

Supporting information

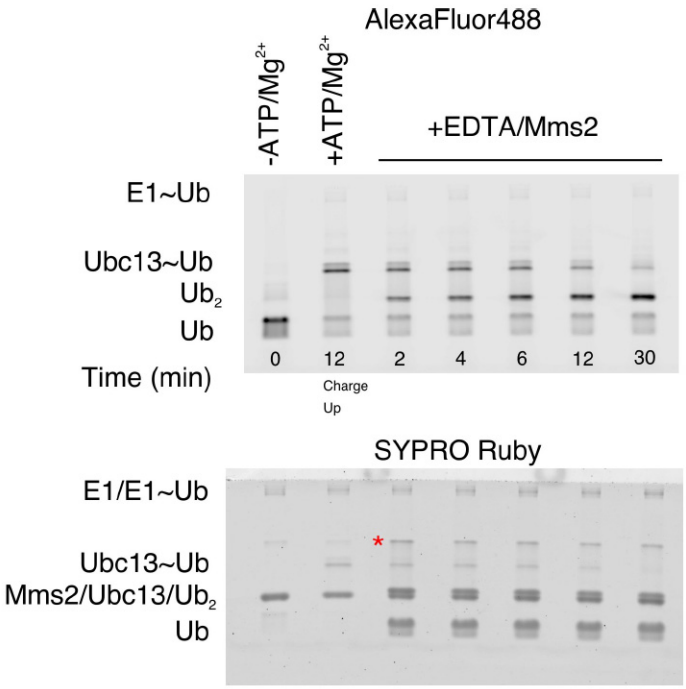

Figure 4.11: SDS-PAGE non-reducing gel for representative Ub_2 assays with the E2 heterodimer Mms2/Ubc13 conjugated with AlexaFluor 488-labeled donor Ub (top panel), and acceptor Ub as the substrate. Lanes labeled + or ATP/Mg^{2+} indicate assays with Ubc13 and E1 in the presence and absence of ATP/Mg^{2+}, respectively, with a reaction time of 12 minutes. The time course indicates the time after which the E1 enzyme was stopped with EDTA, followed by the addition of 50 µM acceptor Ub and Mms2 (to an approximately equimolar ratio of Ubc13). Reactions that included Mms2 were quenched by flash freezing in liquid nitrogen. The bottom panel shows the gel in the top panel, stained with SYPRO Ruby total protein stain. The red asterisk indicates an impurity present in commercially available Ub.

References

1. Petrini, J.H. (2007) Cell signaling. A touching response to damage. *Science* 316, 1138–1139.

2. Covert, M.W., Leung, T.H., Gaston, J.E., and Baltimore, D. (2005) Achieving stability of lipopolysaccharide-induced NF-κB activation. *Science* 309, 1854–1857.

3. Komander, D. (2009) The emerging complexity of protein ubiquitination. *Biochemical Society Transactions* 37, 937–953.

4. Glickman, M.H. and Ciechanover, A. (2002) The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiological Reviews* 82, 373–428.

5. Haas, A.L. and Rose, I.A. (1982) The mechanism of ubiquitin activating enzyme. A kinetic and equilibrium analysis. *Journal of Biological Chemistry* 257, 10329–10337.

6. Pickart, C.M. and Eddins, M.J. (2004) Ubiquitin: structures, functions, mechanisms. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1695, 55–72.

7. Haas, A.L., Bright, P.M., and Jackson, V.E. (1988) Functional diversity among putative E2 isozymes in the mechanism of ubiquitin-histone ligation. *Journal of Biological Chemistry* 263, 13268–13275.

8. Pierce, N.W., Kleiger, G., Shan, S., and Deshaies, R.J. (2009) Detection of sequential polyubiquitylation on a millisecond timescale. *Nature* 462, 615–619.

9. Radzicka, A. and Wolfenden, R. (1995) A proficient enzyme. *Science* 267, 90–93.

10. Wolfenden, R. and Snider, M.J. (2001) The depth of chemical time and the power of enzymes as catalysts. *Accounts of Chemical Research* 34, 938–945.

11. Connors, K.A. and Bender, M.L. (1961) The kinetics of alkaline hydrolysis and *n*-butylaminolysis of ethyl *p*-nitrobenzoate and ethyl *p*-nitrothiolbenzoate. *The Journal of Organic Chemistry* 26, 2498–2504.

12. Yunus, A.A. and Lima, C.D. (2006) Lysine activation and functional analysis of E2-mediated conjugation in the SUMO pathway. *Nature Structural & Molecular Biology* 13, 491–499.

13. Schmidt Jr., D.E. and Westheimer, F.H. (1971) pK of the lysine amino group at the active site of acetoacetate decarboxylase. *Biochemistry* 10, 1249–1253.

14. Villa, J. and Warshel, A. (2001) Energetics and dynamics of enzymatic reactions. *The Journal of Physical Chemistry B* 105, 7887–7907.

15. Garcia-Viloca, M., Gao, J., Karplus, M., and Truhlar, D.G. (2004) How enzymes work: analysis by modern rate theory and computer simulations. *Science* 303, 186–195.

16. Page, M.I. and Jencks, W.P. (1971) Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proceedings of the National Academy of Sciences* 68, 1678–1683.

17. Villa, J., Štrajbl, M., Glennon, TM, Sham, YY, Chu, ZT, and Warshel, A. (2000) How important are entropic contributions to enzyme catalysis? *Proceedings of the National Academy of Sciences* 97, 11899–11904.

18. Xiao, W., Broomfield, S., Chow, B.L., Lin, S.L., and Wei, Y.F. (1998) The products of the yeast $MMS2$ and two human homologs (*hMMS2* and *CROC-1*) define a structurally and functionally conserved Ubc-like protein family. *Nucleic Acids Research* 26, 3908–3914.

19. Hofmann, R.M. and Pickart, C.M. (1999) Noncanonical *MMS2*-encoded ubiquitin-conjugating enzyme functions in assembly of novel polyubiquitin chains for dna repair. *Cell* 96, 645–653.

20. Hoege, C., Pfander, B., Moldovan, G.L., Pyrowolakis, G., Jentsch, S., *et al.* (2002) *RAD6*-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* 419, 135–141.

21. Deng, L., Wang, C., Spencer, E., Yang, L., Braun, A., You, J., Slaughter, C., Pickart, C., and Chen, Z.J. (2000) Activation of the IkappaB kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain. *Cell* 103, 351–361.

22. Wang, B., Matsuoka, S., Ballif, B.A., Zhang, D., Smogorzewska, A., Gygi, S.P., and Elledge, S.J. (2007) Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. *Science* 316, 1194–1198.

23. Kim, H., Chen, J., and Yu, X. (2007) Ubiquitin-binding protein RAP80 mediates BRCA1-dependent DNA damage response. *Science* 316, 1202–1205.

24. Sobhian, B., Shao, G., Lilli, D.R., Culhane, A.C., Moreau, L.A., Xia, B., Livingston, D.M., and Greenberg, R.A. (2007) RAP80 targets BRCA1 to specific ubiquitin structures at DNA damage sites. *Science* 316, 1198–1202.

25. Lewis, M.J., Saltibus, L.F., Hau, D.D., Xiao, W., and Spyracopoulos, L. (2006) Structural basis for non-covalent interaction between ubiquitin and the ubiquitin conjugating enzyme variant human MMS2. *Journal of Biomolecular NMR* 34, 89–100.

26. Eddins, M.J., Carlile, C.M., Gomez, K.M., Pickart, C.M., and Wolberger, C. (2006) Mms2–ubc13 covalently bound to ubiquitin reveals the structural basis of linkage-specific polyubiquitin chain formation. *Nature Structural & Molecular Biology* 13, 915–920.

27. Hau, D.D., Lewis, M.J., Saltibus, L.F., Pastushok, L., Xiao, W., and Spyracopoulos, L. (2006) Structure and interactions of the ubiquitin-conjugating enzyme variant human Uev1a: Implications for enzymatic synthesis of polyubiquitin chains. *Biochemistry* 45, 9866–9877.

28. Palmer, A.G., Kroenke, C.D., and Loria, J.P. (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods in Enzymology* 339, 204–238.

29. McKenna, S., Hu, J., Moraes, T., Xiao, W., Ellison, M.J., and Spyracopoulos, L. (2003) Energetics and specificity of interactions within UbUevUbc13 human ubiquitin conjugation complexes. *Biochemistry* 42, 7922–7930.

30. Montelione, G.T. and Wagner, G. (1989) 2D chemical exchange NMR spectroscopy by proton-detected heteronuclear correlation. *Journal of the American Chemical Society* 111, 3096–3098.

31. Farrow, N.A., Zhang, O., Forman-Kay, J.D., and Kay, L.E. (1994) A heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *Journal of Biomolecular NMR* 4, 727–734.

32. Tollinger, M., Skrynnikov, N.R., Mulder, F.A.A., Forman-Kay, J.D., and Kay, L.E. (2001) Slow dynamics in folded and unfolded states of an SH3 domain. *Journal of the American Chemical Society* 123, 11341–11352.

33. Demers, J.P. and Mittermaier, A. (2009) Binding mechanism of an SH3 domain studied by NMR and ITC. *Journal of the American Chemical Society* 131, 4355–4367.

34. Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. (1994) Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* 33, 5984–6003.

35. Spyracopoulos, L., Lewis, M.J., and Saltibus, L.F. (2005) Main chain and side chain dynamics of the ubiquitin conjugating enzyme variant human Mms2 in the free and ubiquitin-bound states. *Biochemistry* 44, 8770–8781.

36. Bender, M.L. (1951) Oxygen exchange as evidence for the existence of an intermediate in ester hydrolysis. *Journal of the American Chemical Society* 73, 1626–1629.

37. Yang, W. and Drueckhammer, D.G. (2000) Computational studies of the aminolysis of oxoesters and thioesters in aqueous solution. *Organic Letters* 2, 4133–4136.

38. Siepmann, T.J., Bohnsack, R.N., Tokgöz, Z., Baboshina, O.V., and Haas, A.L. (2003) Protein interactions within the N-end rule ubiquitin ligation pathway. *Journal of Biological Chemistry* 278, 9448–9457.

39. Haas, A.L. and Bright, P.M. (1985) The immunochemical detection and quantitation of intracellular ubiquitin-protein conjugates. *Journal of Biological Chemistry* 260, 12464–12473.

40. Hofmann, R.M. and Pickart, C.M. (2001) In vitro assembly and recognition of lys-63 polyubiquitin chains. *Journal of Biological Chemistry* 276, 27936–27943.

41. Fersht, A. (1999) *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding.* (WH Freeman, New York).

42. Warshel, A. (1981) Calculations of enzymic reactions: calculations of pK_{a}, proton transfer reactions, and general acid catalysis reactions in enzymes. *Biochemistry* 20, 3167–3177.

43. Pickart, C.M., Haldeman, MT, Kasperek, EM, and Chen, Z. (1992) Iodination of tyrosine 59 of ubiquitin selectively blocks ubiquitin's acceptor activity in diubiquitin synthesis catalyzed by E2(25K). *Journal of Biological Chemistry* 267, 14418–14423.

44. Haldeman, M.T., Xia, G., Kasperek, E.M., and Pickart, C.M. (1997) Structure and function of ubiquitin conjugating enzyme E2-25K: the tail is a core-dependent activity element. *Biochemistry* 36, 10526–10537.

45. Song, J., Wang, J., Jozwiak, A.A., Hu, W., Swiderski, P.M., and Chen, Y. (2009) Stability of thioester intermediates in ubiquitin-like modifications. *Protein Science* 18, 2492–2499.

46. Schreiber, G., Haran, G., and Zhou, H-X. (2009) Fundamental aspects of protein-protein association kinetics. *Chemical Reviews* 109, 839–860.

47. Frey, P.A. and Hegeman, A.D. (2007) *Enzymatic reaction mechanisms.* (Oxford University Presss).

48. Tatham, M.H., Chen, Y., and Hay, R.T. (2003) Role of two residues proximal to the active site of Ubc9 in substrate recognition by the Ubc9 SUMO-1 thiolester complex. *Biochemistry* 42, 3168–3179.

49. Kamerlin, S.C.L., Haranczyk, M., and Warshel, A. (2008) Progress in ab initio QM/MM free-energy simulations of electrostatic energies in proteins: Accelerated QM/MM studies of pK_a, redox reactions and solvation free energies. *The Journal of Physical Chemistry B* 113, 1253–1272.

50. Bruice, T.C. (2006) Computational approaches: reaction trajectories, structures, and atomic motions. enzyme reactions and proficiency. *Chemical Reviews* 106, 3119–3139.

51. Hammes, G.G. (2008) How do enzymes really work? *Journal of Biological Chemistry* 283, 22337–22346.

52. Alexei, A., Lu, Z.L., Maxwell, C.I., Mohamed, M.F., White, C.J., Tsang, J.S.W., and Brown, R.S. (2006) Combination of a dinuclear Zn^{2+} complex and a medium effect exerts a 10^{12}-fold rate enhancement of cleavage of an RNA and DNA model system. *Journal of the American Chemical Society* 128, 16398–16405.

53. Yang, W. and Drueckhammer, D.G. (2001) Understanding the relative acyl-transfer reactivity of oxoesters and thioesters: computational analysis of transition state delocalization effects. *Journal of the American Chemical Society* 123, 11004–11009.

54. Bender, M.L. and Kemp, K.C. (1957) Oxygen-18 studies of the mechanism of the α-chymotrypsin-catalyzed hydrolysis of esters. *Journal of the American Chemical Society* 79, 111–116.

55. Wu, P.Y., Hanlon, M., Eddins, M., Tsui, C., Rogers, R.S., Jensen, J.P., Matunis, M.J., Weissman, A.M., Wolberger, C.P., and Pickart, C.M. (2003) A conserved catalytic residue in the ubiquitin-conjugating enzyme family. *The EMBO Journal* 22, 5241–5250.

56. VanDemark, A.P., Hofmann, R.M., Tsui, C., Pickart, C.M., and Wolberger, C. (2001) Molecular insights into polyubiquitin chain assembly: crystal structure of the Mms2/Ubc13 heterodimer. *Cell* 105, 711–720.

57. Reverter, D. and Lima, C.D. (2005) Insights into E3 ligase activity revealed by a SUMO–RanGAP1–Ubc9–Nup358 complex. *Nature* 435, 687–692.

Chapter 5

Increased precision for analysis of protein-ligand dissociation constants determined from chemical shift titrations*

Introduction

Biological processes are driven by molecular recognition events involving protein-protein and protein-ligand interactions with dissociation constants (K_{D}) in the µM to mM range and kinetics that span the slow to fast NMR timescales (1–8). For kinetics in the intermediate/fast exchange regimes, the observed NMR resonances for a protein being titrated with cognate binding partner represent weighted averages between the free and bound states of the protein. For 1:1 binding, the observed chemical shift changes follow a hyperbolic dependence on ligand concentration. In general, parameters for the

*This chapter has been published. C.J. Markin, and L. Spyracopoulos. Increased precision for analysis of protein-ligand dissociation constants determined from chemical shift titrations. *Journal of Biomolecular NMR*, 53(2): 125-138, 2012. Reproduced with kind permission from Springer Science and Business Media.

Contributions: C.J.M. and L.S. designed the methods, performed the simulations and analyses, and wrote the paper. C.J.M. prepared and purified proteins, and performed the experimental titrations.

maximum chemical shift change and K_{D} are estimated from nonlinear least squares fits of observed chemical shift changes for the protein as a function of ligand concentration during proteinligand titrations. To obtain accurate values for K_{D} parameters derived from nonlinear least squares fits of 1:1 binding isotherms, statistical analyses of simulated experiments have established that the optimal protein concentration ($[P_0]$) should be held at $0.5 \times K_{\mathrm{D}}$, and the ligand concentration should be varied between $\sim 0.4 \times [P_0]$ and $\sim 11 \times [P_0]$, with 15-20 repetitions of the titration (9). From a practical standpoint, $[P_0]$ is usually between 50 and 500 µM in order to achieve sufficient signal to noise ratio when employing standard ^{1}H-^{15}N or ^{1}H-^{13}C 2D NMR spectroscopic techniques. Thus, to conduct a titration for $[P_0] \sim 500$ µM, and $K_{\mathrm{D}} \sim 200$ µM, the ligand concentration must reach 2 mM during a titration in order to achieve reasonable accuracy, that is, the binding site on the protein must be saturated with ligand. Without prior knowledge of the actual K_{D}, it is difficult to choose the optimal value for $[P_0]$, and larger $[P_0]$ values are more desirable as they yield higher signal to noise ratios in NMR spectra. Another requirement to achieve accuracy is that the binding site should be saturated during the course of the titration; this requires high ligand concentrations, a condition which is often difficult to satisfy for larger K_{D} values. For example, if the ligand is another protein, it may be difficult to prepare in sufficient quantity, or it may not be soluble at higher concentrations. Furthermore, it is difficult to conduct a titration such that the concentration of the fixed protein component is not significantly altered. Ideally, this requires making separate NMR samples for each titration point from single stock solutions of protein and ligand. This is advantageous with respect to controlling buffer conditions, but problematic due to the large volumes of stock solutions required. Whilst there is no simple solution to address the difficulties associated with the experimental design of NMR-monitored titrations, we have devised two straightforward approaches for optimizing the experimental design of protein-ligand titrations such that significant increases in precision can be achieved. Conducting protein ligand titrations whereby the concentration of protein component is allowed to co-vary with the ligand concentration, and employing nonlinear least squares analyses with both ligand and protein concentrations as independent variables and K_{D} and $\Delta\delta_{\mathrm{max}}$ as variable parameters, can allow for a significant increase in the precision of K_{D} compared to the precision obtained at fixed

$[P_0]$, particularly when $[P_0]$ exceeds K_D. We have analyzed two methods by which covariation of protein and ligand concentration can be used to achieve increased precision for fitted K_D values. The first method involves addition of aliquots of ligand solution to a protein solution such that the protein concentration decreases by a constant factor as more ligand is added. The second method involves the serial dilution of a solution containing concentrated protein and concentrated ligand in a 1:2 ratio, respectively. The practicality of both methods was demonstrated by conducting 2D ^{1}H-^{15}N HSQC NMR chemical shift titration experiments for the interaction of human [U-^{15}N]-Mms2 (145 residues) with human ubiquitin (76 residues).

Theory and methods

1:1 binding isotherms

At equilibrium, the relationship between the concentrations of protein, ligand, proteinligand complex, and the equilibrium dissociation constant is given by:

$$K_D = \frac{[P][L]}{[PL]} \tag{5.1}$$

with the total protein and ligand concentrations:

$$P_T = [P] + [PL] \tag{5.2}$$

$$L_T = [L] + [PL] \tag{5.3}$$

Equations 5.1 to 5.3 can be solved to yield the following expression for the bound protein complex:

$$[P_L] = \frac{1}{2}(K_D + L_T + P_T) - \sqrt{(-K_D - L_T - P_T)^2 - 4L_T P_T} \tag{5.4}$$

In the limit of fast exchange on the NMR timescale, the observed chemical shift changes for ^{13}C or ^{15}N labeled protein ($[P_T]$) being titrated with ligand are given by:

$$\Delta\delta_{obs} = f_b \Delta\delta_{max} \tag{5.5}$$

where f_b is the fraction of bound protein, $[P_L]/[P_T]$, and $\Delta\delta_{max}$ is the maximum chemical shift change for a given residue given by the difference $\delta_{free} - \delta_{bound}$, where δ_{free} is the chemical shift for the free state of the protein being observed, and δ_{bound} is the chemical shift for protein being observed in the fully bound state. Thus eq. 5.4 becomes:

$$[PL] = \frac{\Delta\delta_{max}}{2[P_T]}(K_D + L_T + P_T) - \sqrt{(-K_D - L_T - P_T)^2 - 4L_T P_T} \qquad (5.6)$$

Equation 5.6 is applicable, defined by $k_{ex} \gg |\Delta\omega|$, and is roughly valid for exchange rates approaching the intermediate exchange regime, given by (10). The precise limit for the applicability of eq. 5.6 is case-specific, and depends on the magnitude of the line broadening.

Monte Carlo error analysis for simulated titrations

Protein chemical shift changes as a function of ligand concentration can be fit to eq. 5.6 using nonlinear least squares methods with K_D and $\Delta\delta_{max}$ as variable parameters. Typically, the ligand concentration is treated as an independent variable, whereas the initial protein concentration is treated as a constant. However, there is no a priori reason to make this distinction. For example, eq. 5.6 describes a two-dimensional binding curve with $[P_T]$ and $[L_T]$ on the x and y-axes and $\Delta\delta_{obs}$ on the z-axis. If this surface varies as a function of $[P_T]$, then it stands to reason that greater accuracy for a single titration can be achieved by sampling the surface more extensively with respect to $[P_T]$. To that end, we sought to determine the errors associated with fitting chemical shift data to eq. 5.6 with K_D and $\Delta\delta_{max}$ as adjustable parameters and $[P_T]$ and $[L_T]$ as independent variables. We assumed that the two main sources of error in a given titration arise from errors in $\Delta\delta_{obs}$ and the errors in the starting protein and ligand concentrations. Provided that titration points involve addition of a single stock solution of ligand, dilution, or a combination thereof, then the largest concentration error involves only the starting concentrations of protein and ligand, as subsequent errors due to pipetting/dilution are small.

Two specific types of titrations were simulated with the program *Mathematica* 8.0.4 (11) using different concentrations of protein and ligand and seven different values for

Table 5.1: Protein and ligand concentrations (mM) for Method 1 simulations, K_D=2, 20, 60, 200, 600, 1000, and 2000 µM

Case											
1	$[PT]$ ($\Delta_{[P]} = 0$)	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
	$[L_T]$	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.6	0.8
2	$[P_T]$ ($\Delta_{[P]} = 0$)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	$[L_T]$	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.6	0.8
3	$[P_T]$ ($\Delta_{[P]} = 0.1$)	0.5	0.45	0.405	0.364	0.328	0.295	0.266	0.239	0.215	0.194
	$[L_T]$	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.6	0.8
4	$[P_T]$ ($\Delta_{[P]} = 0.25$)	0.5	0.375	0.281	0.211	0.158	0.119	0.089	0.067	0.05	0.037
	$[L_T]$	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.6	0.8

Table 5.2: Protein and ligand concentrations (mM) for Method 2 simulations, K_D=2, 20, 60, 200, 600, 1000, and 2000 µM

Method 2										
$[P_T]$	0.5	0.45	0.4	0.35	0.3	0.25	0.2	0.15	0.1	0.05
$[L_T]$	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1

K_D, ranging from 2 to 2000 µM. The first, Method 1, involves simulation of the standard method of conducting a titration, that is, addition of aliquots of ligand solution to a protein solution, with the exception that the protein concentration is allowed to decrease as more ligand is added (table 5.1). The second type of simulated titration, Method 2, was one in which a solution of concentrated protein and concentrated ligand was serially diluted to produce a binding isotherm (table 5.2). For Method 1, one thousand simulated data sets for titrations were generated for seven different values K_D (2, 20, 60, 200, 600, 1000, and 2000 µM), each with $\Delta\delta_{max}$ = 1 ppm. The initial protein concentrations were randomly chosen from a normal distribution with mean 100 µM (table 5.1, Case 1) or 500 µM (table 5.1, Case 2) and a standard deviation of ±5%; likewise, the initial ligand concentrations were chosen randomly from a normal distribution with mean 0.05 xM and a standard deviation of ±5%, and increased by factors of 2, 3, 4, 5, 6, 7, 8, 12, and 16 to yield the ligand concentration for subsequent titration points (table 5.1). The protein and ligand concentrations, as well as $\Delta\delta_{max}$ were substituted into eq. 6 to yield an ensemble of Monte Carlo $\Delta\delta_{obs}$ values. In addition to the Monte Carlo ensembles corresponding to the seven values of K_D, two further sets of seven Monte Carlo ensembles were generated by choosing two constant factors $\Delta_{[P]} = 0.1$ and 0.25 (table 5.1, Cases 3 and 4, respectively), such that the initial protein concentration $[P_0]$ decreased in the n subsequent titration points according to:

$$[P_n] = [P_{n-1}] - \Delta_{[P]} \times [P_{n-1}] \tag{5.7}$$

for titration points $n > 1$ and $\Delta_{[P]} \leq 1.0$.

For Method 2, the initial protein concentration $[P_0]$ was set to 500 µM, the initial ligand concentration ($[L_0]$) was set to 1 mM, seven values of K_D were chosen (2, 20, 60, 200, 600, 1000, and 2000 µM), with one $\Delta\delta_{max}$ (1 ppm) per K_D (table 5.2). One thousand

simulated data sets were generated by choosing the initial ligand and protein concentrations randomly from normal distributions with means 1 and 0.5 mM respectively, and standard deviations of 5%. The protein and ligand concentrations were allowed to decrease by a constant increment of 50 µM for a total of ten titration points. These protein and ligand concentrations, along with $\Delta\delta_{\max}$, were used to calculate $\Delta\delta_{\text{obs}}$ values for the Monte Carlo ensemble with eq. 5.6.

The Monte Carlo ensembles of titration data were fit to eq. 5.6 using the NonlinearModelFit nonlinear least squares regression package within the program Mathematica 8.04 (11). The NMinimize and DifferentialEvolution options with the default parameters were chosen for the constrained nonlinear optimization algorithm. The resulting ensembles for the K_{D} and $\Delta\delta_{\max}$ parameters were either normal, i.e. Gaussian distributions, or gamma distributions. The normal distributions were fit to the probability density function:

$$P(x) = A_0 \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{2\sigma^2} \tag{5.8}$$

where A_0 is a scaling parameter. The gamma distributions were fit to the following probability density function:

$$P(x) = A_0 \frac{e^{-x/\beta} x^{\alpha-1} \beta^{-\alpha}}{\Gamma(\alpha)} \tag{5.9}$$

where A_0 is a scaling parameter and

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt \tag{5.10}$$

For the normal and gamma distributions, the fitted probability distribution functions were used to calculate the probability of observing a given parameter to within a given error threshold. This approach is useful for gamma probability distribution functions that are typically asymmetric with respect to their median value. The Monte Carlo ensembles for K_{D} and $\Delta\delta_{\max}$ were binned by empirically adjusting the parameters for the Mathematica command BinCounts such that the output closely matched the default settings for the Histogram command given with the option Probability.

^{15}N lineshape analyses for simulated titrations

Free induction decays (FIDs) for the various proteinligand ratios of the simulated titrations were generated from the Bloch-McConnell equations for two-site chemical exchange (12), wherein the time course for transverse magnetization, or the FID, is given by:

$$M_{\mathrm{A}}(t) = M_{\mathrm{A}}(0)a_{11}(t) + M_{\mathrm{B}}(0)a_{12}(t) \tag{5.11}$$

$$M_{\mathrm{B}}(t) = M_{\mathrm{B}}(0)a_{22}(t) + M_{\mathrm{A}}(0)a_{21}(t) \tag{5.12}$$

for two spins (A and B), with the coefficients for the auto-peaks:

$$\begin{aligned} a_{11}(t) = \frac{1}{2}\Bigg[&\left(1 - \frac{-i\Delta\omega + R^0_{2A} - R^0_{2B} + k_{\mathrm{ex}}(p_{\mathrm{B}} - p_{\mathrm{A}})}{\lambda_+ - \lambda_-}\right)\exp\left(-\lambda_- t\right) \\ &+ \left(1 + \frac{-i\Delta\omega + R^0_{2A} - R^0_{2B} + k_{\mathrm{ex}}(p_{\mathrm{B}} - p_{\mathrm{A}})}{\lambda_+ - \lambda_-}\right)\exp\left(-\lambda_- t\right)\Bigg] \end{aligned} \tag{5.13}$$

$$\begin{aligned} a_{22}(t) = \frac{1}{2}\Bigg[&\left(1 + \frac{-i\Delta\omega + R^0_{2A} - R^0_{2B} + k_{\mathrm{ex}}(p_{\mathrm{B}} - p_{\mathrm{A}})}{\lambda_+ - \lambda_-}\right)\exp\left(-\lambda_- t\right) \\ &+ \left(1 - \frac{-i\Delta\omega + R^0_{2A} - R^0_{2B} + k_{\mathrm{ex}}(p_{\mathrm{B}} - p_{\mathrm{A}})}{\lambda_+ - \lambda_-}\right)\exp\left(-\lambda_- t\right)\Bigg] \end{aligned} \tag{5.14}$$

$$a_{12}(t) = \frac{k_{\mathrm{ex}}p_{\mathrm{A}}}{\lambda_+ - \lambda_-}[\exp -\lambda_- t - \exp -\lambda_+ t] \tag{5.15}$$

$$a_{21}(t) = \frac{k_{\mathrm{ex}}p_{\mathrm{B}}}{\lambda_+ - \lambda_-}[\exp -\lambda_- t - \exp -\lambda_+ t] \tag{5.16}$$

where $R^0_{2\mathrm{A}}$ and $R^0_{2\mathrm{B}}$ are the transverse relaxation rates for spins A and B, respectively, in the absence of chemical exchange, $\Delta\omega$ is the difference between the chemical shifts of spin A (Ω_A) and spin B (Ω_B) in rad s^{-1}, k_{ex} is the rate of chemical exchange, $k_{\mathrm{ex}} = k_{\mathrm{on}}[\mathrm{B}] + k_{\mathrm{off}}$, p_A and p_B are the populations of spin A and B, respectively, and

$$\begin{aligned} \lambda_\pm = \frac{1}{2}\Bigg\{&-i\Omega_A - i\Omega_B + R^0_{2A} + R^0_{2B} + k_{\mathrm{ex}} \\ &\pm \sqrt{(-i\Delta\omega + R^0_{2A} - R^0_{2B} + k_{\mathrm{ex}}(p_B - p_A))^2 + 4p_a p_b k^2_{\mathrm{ex}}}\Bigg\} \end{aligned} \tag{5.17}$$

For Method 1, ^{15}N NMR spectra were simulated using the program Mathematica 8.04 (11) for a titration involving a 1:1 protein ligand interaction with $k_{\mathrm{on}} = 1 \times 10^8$ M^1 s^{-1}

and $k_{off} = 2 \times 10^4$ s^{-1} ($K_D = k_{off}/k_{on} = 200$ µM), ^{15}N $\Delta\delta_{max} = 4.64$ ppm ($\Delta\delta_{max} =^{15}$ N $\Delta\delta_{max}/5 = 0.93$ ppm or 278.56 Hz at 60 MHz), $[P_0] = 0.5$ ($K_D = 100$ µM, and ligand concentrations of $[L_0]$, $2[L_0]$, $3[L_0]$, $4[L_0]$, $5[L_0]$, $6[L_0]$, $7[L_0]$, $8[L_0]$, $12[L_0]$, and $16[L_0]$, with $[L_0] = 50$ µM, with the final ligand concentration 16 $[L_0] = 8$ $[P_0]$. These conditions are similar to those in Method 1, Case 1 (table 5.1), but with only one (K_D and a slightly smaller $\Delta\delta_{max}$. The resonance frequencies of the free and bound states were $\nu_A = 7273.76$ and $\nu_B = 7552.32$ Hz, respectively, and the intrinsic ^{15}N linewidths (R^0_{2A}/π and R^0_{2B}/π) of the free and bound states were taken to be 2 Hz, the expected value for an $\sim$ 80 residue protein. FIDs were calculated over an acquisition time of 94.4 ms. FIDs were multiplied by a cosine squared window function and the first point was halved prior to analytical Fourier transformation. ^{15}N NMR spectra for a second and third titration were simulated in a similar fashion with the exception that $[P_0] = 5 \times K_D = 500$ µM with $\Delta_{[P]} = 0$, and $[P_0] = 500$ µM with $\Delta_{[P]} = 0.25$, respectively, similar to Method 1, Cases 2 and 4, respectively (table 5.2). For Method 2, ^{15}N NMR spectra were simulated in a similar fashion as Method 1, but with $k_{on} = 1 \times 10^8$ M^{-1} s^{-1}, $k_{off} = 2 \times 10^2$, 2×10^4, and 2×10^5 s^{-1} (K_Ds of 2 µM, 200 µM, and 2000 µM), with $[L_0] = 1$ mM, $[P_0] = 500$ µM, and a decrease in subsequent concentrations by a constant increment of 0.05 mM, for ten titration points. These conditions are similar to those in Method 2 (table 5.2), but with only three K_Ds and a slightly smaller $\Delta\delta_{max}$.

Sample preparation and experimental NMR-monitored titrations

Human Mms2 (145 residues) was prepared as previously described (13, 14), whereas human ubiquitin (76 residues) was purchased as a lyophilized powder from Boston Biochem (Cambridge, MA). NMR samples for Method 1 employed [U-^{15}N]-Mms2 at a starting concentration $[P_0] \sim 0.33$ mM in 320 µL of 9:1 H_2O/D_2O (pH 7.3), 50 mM TRIS, 200 mM NaCl, 1 mM DTT, 1 mM DSS, 12 µL of 25X stock protease inhibitor cocktail (prepared from Roche cOmplete protease inhibitor tablets, catalog #11697498001), in a Shigemi microcell NMR tube. A stock solution of 4.7 mM human ubiquitin was prepared using the identical buffer as that used to prepare [U-^{15}N]-Mms2. In addition to an initial titration point in the absence of ubiquitin, five titrations points were collected with Mms2 concentrations, $[P_T]$, of 0.28, 0.23, 0.19, 0.15, and 0.06 mM, with corresponding ubiqui-

tin concentrations, $[L_T]$, 0.18, 0.39, 0.56, 0.71, and 1.06 mM (table 5.3). Titrations were conducted by removing the sample from the NMR tube at each titration point using a glass pipette, and subsequently mixing a known amount of the recovered sample with the amount of stock ubiquitin solution and/or stock buffer solution necessary to prepare the next sample. To avoid concentration changes due to small volume losses upon sample recovery, a known amount of recovered sample is removed using a micropipette and buffer is added to this, allowing precise calculation of both concentrations in the new sample for the next titration point. The NMR tube was washed, rinsed, and dried between titration points to minimize unwanted sample dilution.

For Method 2, NMR samples employed [U-^{15}N]-Mms2 at a starting concentration $[P_0] \sim 0.59$ mM and a starting concentration of 1.0 mM ubiquitin in 330 µL of 9:1 H2O/D2O (pH 7.3), 50 mM TRIS, 200 mM NaCl, 1 mM DTT, 1 mM DSS, 13.2 µL of 25 stock protease inhibitor cocktail (prepared from Roche cOmplete protease inhibitor tablets, catalog # 11697498001), in a Shigemi microcell NMR tube. In addition to the initial point, four titration points were collected by diluting the initial sample with stock buffer to achieve Mms2 concentrations, $[P_T]$, of 0.49, 0.39, 0.29, and 0.20 mM, with corresponding Ub concentrations, $[L_T]$, 0.83, 0.67, 0.50, and 0.33 mM (table 5.4). A separate NMR sample containing 0.44 mM [U-^{15}N]-Mms2 in the identical buffer employed in the titration was also prepared to collect free chemical shifts for Mms2. Titrations were conducted by removing the sample from the NMR tube after each titration point using a glass pipette, and subsequently mixing a known amount of the recovered sample with the amount of stock buffer solution necessary to prepare the next sample. Given the small volume loss upon sample recovery, a known amount of recovered sample is removed using a micropipette and buffer is added to this, allowing precise calculation of both concentrations in the new sample for the next titration point. Additionally, the NMR tube was washed, rinsed, and dried between titration points to prevent unwanted sample dilution. For both Methods 1 and 2, the concentrations of stock Mms2 and ubiquitin were determined by amino acid analysis.

Table 5.3: Protein and ligand concentrations for experimental verification of Method 1

Method 1						
$[P_T]$ (mM)	0.33	0.28	0.23	0.19	0.15	0.06
$[L_T]$ (mM)	0.0	0.18	0.39	0.56	0.71	1.06
Transients collected	8	16	16	64	64	128

Table 5.4: Protein and ligand concentrations for experimental verification of Method 2

Method 1					
$[P_T]$ (mM)	0.59	0.49	0.39	0.29	0.20
$[L_T]$ (mM)	1.0	0.83	0.67	0.50	0.33
Transients collected	16	16	16	64	128

NMR spectroscopy

For chemical shift titrations using Method 1 and 2, wherein [U-^{15}N]-Mms2 is monitored, 2D ^{1}H-^{15}N sensitivity enhanced HSQC NMR spectra (15) were acquired at 25 ℃ using a Varian Unity INOVA 600 MHz spectrometer equipped with a room temperature 5-mm triple resonance probe and triple-axis pulsed field gradients. A total of 192 and 977 complex points were collected in the t_1 and t_2 domains, respectively. The number of titration points for Method 1 and Method 2, and the numbers of transients collected per titration point are given in tables 5.3 and 5.4.

NMR data processing and analysis

All spectral processing was accomplished with the program NMRPipe (16). For 2D ^{1}H-^{15}N HSQC spectra, sorting and processing of the superposed orthogonal components for sensitivity enhancement were performed with the ranceY.M macro within the NMRPipe software. Post-acquisition processing of the t_2 interferograms for removal of residual water was employed for 2D ^{1}H-^{15}N HSQC NMR spectra. For 2D ^{1}H-^{15}N and NMR spectra acquired for Mms2, 85°-shifted sine and 90°-shifted sine-squared window functions were applied in t_2 and t_1, respectively. The t_2 and t_1 domains were extended to twice the number of points with zero filling. An automatic polynomial subtraction in the F_2 dimension was used for baseline correction, the region upfield of 6.0 ppm was discarded for 2D ^{1}H-^{15}N NMR spectra. Chemical shifts were assigned using those previously described (17); spectra for the titrations were analyzed with the program Sparky (18).

Nonlinear regression for estimation of $K_{\mathbf{D}}$ and $\Delta\delta$ from NMR-monitored titration data

For Methods 1 and 2, the per residue observed chemical shift changes for various Mms2 residues as a function of Mms2 and ubiquitin concentrations were fit to eq. 5.6 to yield K_{D} and $\Delta\delta_{\mathrm{max}}$ values using the default NMinimize constrained nonlinear least squares fitting algorithm implemented in Mathematica 8.0.4. The precision of K_{D} and $\Delta\delta_{\mathrm{max}}$ were determined for each residue in both the ^{1}H and ^{15}N dimensions independently using Monte Carlo methods within Mathematica. For Methods 1 and 2, the initial protein concentrations by choosing randomly from normal distributions with means of 1 mM, and standard deviations of $\pm 10\%$, and multiplying these values by the experimental protein concentrations; likewise, the initial ligand concentrations were chosen randomly in the same manner. For each Monte Carlo trial, the chemical shifts were also chosen randomly based on the per residue values of the chemical shift precision determined for a given cross-peak (*vide infra*). These data sets were then fit using the same nonlinear least squares fitting algorithm used to determine the experimental K_{D} and $\Delta\delta_{\mathrm{max}}$. The standard deviations of the K_{D} and $\Delta\delta_{\mathrm{max}}$ values from 1000 Monte Carlo trials were taken as the errors in these parameters.

Analyses of chemical shift precision from 2D ^{1}H-^{15}N HSQC NMR spectra

The precision of various per residue ^{1}H and ^{15}N chemical shifts measured from 2D ^{1}H-^{15}N HSQC NMR spectra were determined using Monte Carlo methods within the program Mathematica 8.0.4. For a given 2D cross-peak, the maximum point and the first point on either side of the maximum in either the ^{1}H or ^{15}N dimension was assumed to follow a parabolic dependence on frequency, given by the system of polynomial equations (18):

$$\begin{aligned} a_0 + a_1 x_1 + a_2 x_1^2 &= y_1 \\ a_0 + a_1 x_2 + a_2 x_2^2 &= y_2 \\ a_0 + a_1 x_3 + a_2 x_3^2 &= y_3 \end{aligned} \tag{5.18}$$

where y_2 is the height of the maximum, y_1 and y_3 are the heights of the points on either side, x_1, x_2, and x_3 are the chemical shifts of y_1, y_2, and y_3 respectively, and a_0, a_1, and a_3 are constants. This system of equations can be solved to yield the values of the constants, which are then substituted into the first derivative of the parabolic equation ($a_1 + 2a_2x$), which is set to zero and solved with respect to x to yield an estimate for the chemical shift at the interpolated peak maximum. Ten thousand Monte Carlo trials were conducted by adding random noise, estimated from the baseplane of the NMR spectrum, to every point of the region of the 2D spectrum surrounding the cross-peak of interest, then the y_1, y_2, and y_3 values for these data sets were chosen for either the ^{1}H or ^{15}N resonance for a given cross-peak, and used to interpolate the respective chemical shifts, as described above. The region around a given cross-peak was manually chosen such that the surrounding baseplane was well represented, and lacked other cross-peaks.

Results and discussion

Precision of $K_\mathbf{D}$ and $\Delta\delta_\mathbf{max}$ parameters for simulated titrations conducted with fixed $[P_\mathbf{0}]$

In a typical protein-ligand titration, the initial protein concentration (analyte or titrand), $[P_0]$, is maintained nearly constant during a titration, and the resulting observed chemical shift changes are analyzed using nonlinear regression to estimate the parameters K_D and $\Delta\delta_\mathrm{max}$. In the discussion that follows, we assume throughout that the protein-ligand interaction is described by 1:1 stoichiometry. It has previously been demonstrated that under conditions of nearly fixed $[P_0]$, the most accurate K_D and $\Delta\delta_\mathrm{max}$ parameters will be obtained for $[P_0] \sim 0.5\ K_\mathrm{D}$, a maximum concentration for the titrant (ligand) of $\sim$10 $[P_0]$, and 15 - 20 repetitions of the titration (9). This number of repetitions is typically not practical when employing isotopically labeled proteins and/or ligands that are difficult or costly to prepare. The optimal ligand concentration for achieving accuracy is also not often achievable in practice, particularly if the ligand is a peptide or protein that is difficult to produce, or not highly soluble. Furthermore, in the absence of prior knowledge of K_D, it is difficult to chose $[P_0] \sim 0.5\ K_\mathrm{D}$; moreover, it is often desirable

to have millimolar concentrations of $[P_0]$ to maximize signal intensity in NMR spectra, whereas biologically relevant K_D values are typically in the µM range. Thus, increasing $[P_0]$ to mM concentrations for an initial titration, in the interest of sensitivity gains, will lead to deleterious effects on the accuracy of K_D and $\Delta\delta_{max}$ for biologically relevant K_Ds determined using a traditional titration. In this study, we developed two approaches whereby a titration can be conducted with millimolar starting concentrations for $[P_0]$, while achieving accuracy comparable to, or better than, maintaining $[P_0] \sim 0.5\ K_D$.

Precision of K_D and $\Delta\delta_{max}$ parameters for simulations of Method 1: $[P_0]$ decreases by a constant factor as $[L_0]$ increases

The accuracy of a given titration where $[P_0]$ remains nearly constant can be assessed by considering a titration for a specific 1:1 protein ligand interaction (Method 1, Case 1, figure 5.1A) with K_D = 200 µM, $\Delta\delta_{max}$ = 1 ppm, $[P_0] = 0.5K_D$ = 100 µM, ligand concentrations of $[L_0]$, $2[L_0]$, $3[L_0]$, $4[L_0]$, $5[L_0]$, $6[L_0]$, $7[L_0]$, $8[L_0]$, $12[L_0]$, and $16[L_0]$, with $[L_0]$ = 50 µM, $16[L_0] = 8\ [P_0]$, errors of ± 5% for $[L_0]$ and $[P_0]$, and a chemical shift precision of 0.002 ppm. These conditions represent the traditional approach to conducting an NMR titration and can be considered as Method 1 with $\Delta_{[P]} = 0$. Non-linear least squares regression for an ensemble of 1000 simulated titrations for Method 1, Case 1, yields normally distributed parameters with median values $K_D = 200 \pm 17$ µM (figure 5.1B) and $\Delta\delta_{max} = 1.00 \pm 0.03$ ppm (figure 5.1C). For $[P_0]$ = 0.5 mM, on the other hand (Case 2, figure 5.2A), the parameter ensembles appear as gamma distributions with median values $K_D = 200 \pm 143$ µM (figure 5.2B), $\Delta\delta_{max} = 1.0 \pm 0.2$ ppm (figure 5.2C); values substantially less precise than those obtained at the optimal concentration $[P_0] = 0.5 \times K_D$.

Given the previous titration (Case 1) for $[P_0] = 0.5 \times K_D$, or 100 µM, although good precision is theoretically possible for K_D ($< \pm 10\%$), the main problem is that conducting titrations with protein concentrations of 100 µM or less becomes time consuming as more transients need to be collected to ensure good signal to noise ratio in standard 2D ^{1}H-^{15}N NMR spectra. The precision for Case 2 is poor ($> \pm 50\%$) when $[P_0]$ is 0.5 mM; however, if $[P_0]$ is allowed to decrease by a constant factor in subsequent titration points,

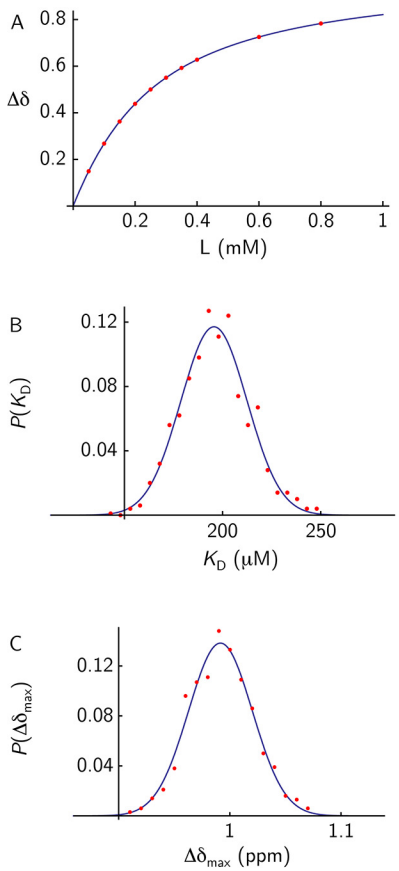

Figure 5.1: (A) Binding isotherm for a 1:1 protein ligand interaction for Method 1, Case 1. (B) Histogram for 1000 K_{D} values determined from Monte Carlo parameter estimation for Case 1 (red dots), and corresponding fit to a normal distribution (blue line, eq. 5.8). (C) Histogram for 1000 $\Delta\delta_{\mathrm{max}}$ values determined from Monte Carlo parameter estimation for Case 1 (red dots), and corresponding fit to a normal distribution (blue line, eq. 5.8).

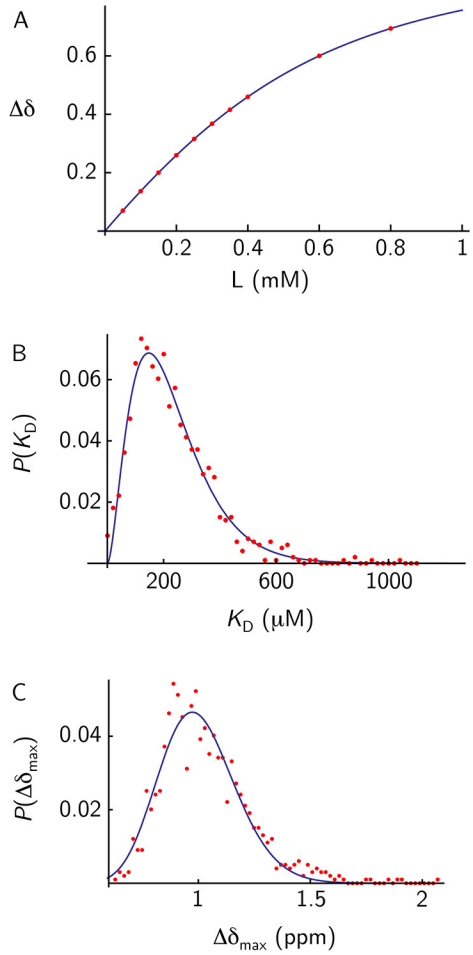

Figure 5.2: (A) Binding isotherm for a 1:1 protein ligand interaction for Method 1, Case 2. (B) Histogram for 1000 K_{D} values determined from Monte Carlo parameter estimation for Case 2 (red dots), and corresponding fit to a gamma distribution (blue line, eqns. 5.9 and 5.10). (C) Histogram for 1000 $\Delta\delta_{\mathrm{max}}$ values determined from Monte Carlo parameter estimation for Case 2 (red dots), and corresponding fit to a gamma distribution (blue line, eqns. 5.9 and 5.10).

precision comparable to Case 1 can be obtained ($[P_0] = 0.5\ K_D$). For example, for $[P_0]$ = 0.5 mM, using $\Delta_{[P]} = 0.25$ in eq. 7 gives $[P_n]$ = 0.5, 0.38, 0.28, 0.21, 0.16, 0.12, 0.09, 0.07, 0.05 and 0.04 mM for n =1, . . . , 10 (Case 3, figure 5.3A). Using these analyte concentrations yields normally distributed parameters with median values for $K_D = 200 \pm 23$ µM (figure 5.3B), and $\Delta\delta_{max} = 1.00 \pm 0.04$ ppm (figure 5.3C). Thus, the precision for both K_D and $\Delta\delta_{max}$ is similar to Case 1 where $[P_0] = 0.5\ K_D$, or 100 µM, for which $\sigma(K_D)$ and $\sigma(\Delta\delta_{max})$ are ±17 µM and ±0.02 ppm, respectively. However, the main advantage for Case 3 is that six of the ten titration points contain protein concentrations exceeding 100 µM, and this is advantageous with respect to signal to noise ratio in NMR spectra in comparison to maintaining $[P_0] = 0.5K_D$. If we assume that NMR spectra are collected using identical parameters for the two different scenarios in Case 1 and Case 3, $[P_0]$ = 100 µM and $[P_0]$ = 0.5 mM with $\Delta_{[P]} = 0.25$, respectively, and linewidth differences between the two cases are small, then an overall increase in the average signal to noise of 1.9-fold can be expected for Case 3 with $[P_0]$ = 0.5 mM and $\Delta_{[P]} = 0.25$ compared to Case 1 where $[P_0]$ = 100 µM, simply on the basis of protein concentration (figure 5.4). Furthermore, in practice, it is not straightforward to conduct an NMR titration at a fixed protein concentration. For Case 1, where $[P_0]$ = 100 µM, if subsequent titration points decreased in concentration by as little as 5%, then the last five titration points would range between a maximum of 80 µM to a minimum of 60 µM. If NMR spectra are collected using identical parameters to Case 3, with minimal linewidth changes between the two cases, then an overall increase in the average signal to noise of 2.4-fold can be expected for Case 3 ($[P_0]$ = 500 µM and $\Delta_{[P]} = 0.25$) compared to Case 1 ($[P_0]$ = 100 µM).

General performance of Method 1 simulations over a range of K_D and with $\Delta_{[P]}$ values of 0.1 and 0.25

In general, the value of K_D for a given protein-ligand interaction is not known prior to planning an NMR titration. Thus, to determine the performance of Method 1 we conducted Monte Carlo simulations using $[P_0]$ = 500 µM, for a series of K_D values (2, 20, 60, 200, 600, 1000, and 2000 µM), and $\Delta_{[P]}$ values (0.1 and 0.25, eq. 7), and ligand

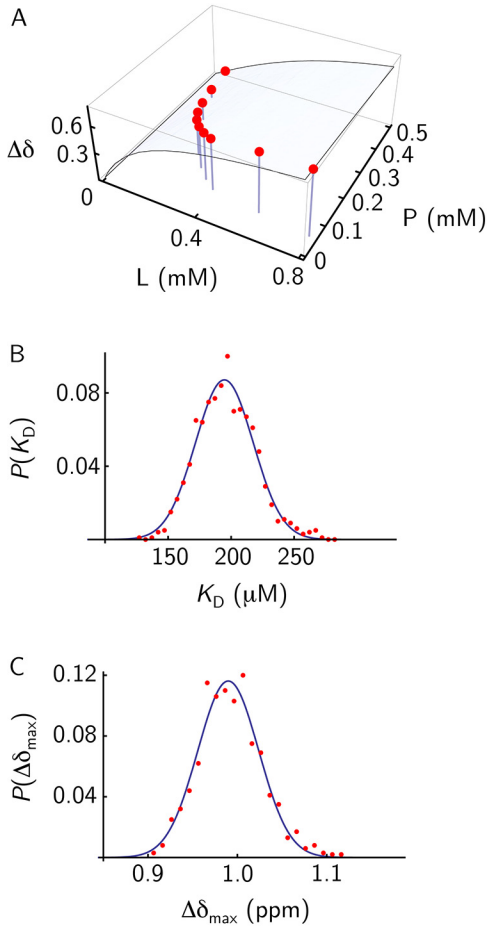

Figure 5.3: (A) Binding isotherm for a 1:1 protein ligand interaction for Method 1, Case 3. (B) Histogram for 1000 K_{D} values determined from Monte Carlo parameter estimation for Case 3 (red dots), and corresponding fit to a normal distribution (blue line, eq. 5.8). (C) Histogram for 1000 $\Delta\delta_{\mathrm{max}}$ values determined from Monte Carlo parameter estimation for Case 3 (red dots), and corresponding fit to a normal distribution (blue line, eq. 5.8).

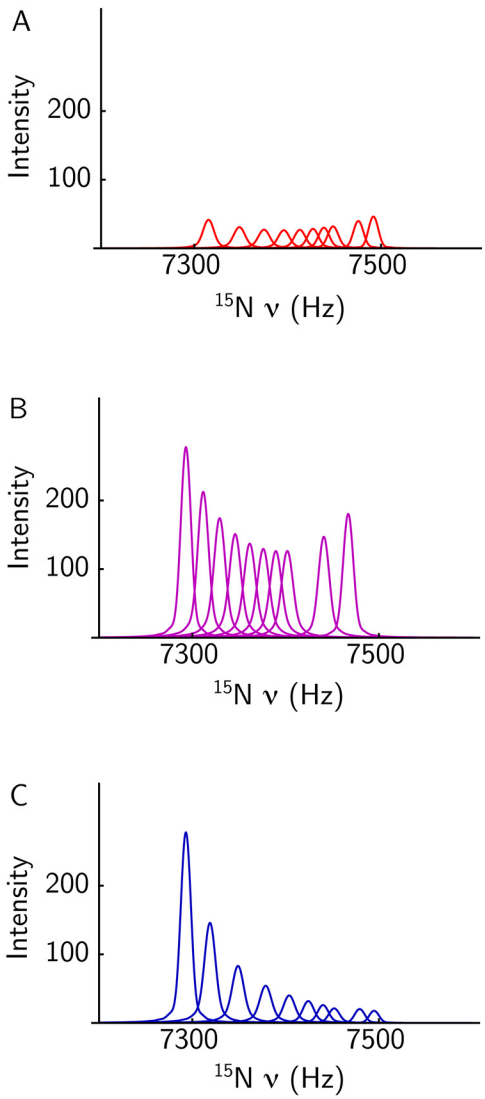

Figure 5.4: (A) Simulated ^{15}N lineshapes for Method 1, Case 1. (B) Simulated ^{15}N lineshapes for Method 1, Case 2. (C) Simulated lineshapes for Method 1, Case 3. For all spectra, the resonance frequencies for the free and bound protein are 7273.76 and 7552.32 Hz, respectively.

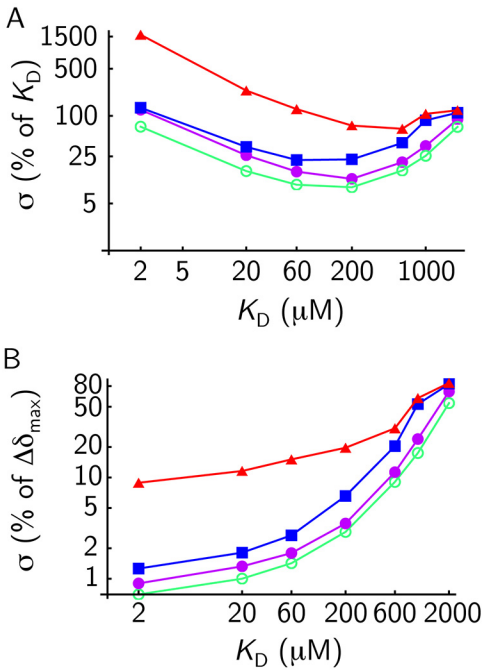

Figure 5.5: (A) Standard deviations of 1000 Monte Carlo trials for various K_D values for Method 1, Case 1 (green open circles), Method 1, Case 2 (red triangles), Method 1, Case 3 (blue squares), Method 1, Case 4 (magenta circles). (B) Standard deviations of 1000 Monte Carlo trials for $\Delta\delta_{max}$ corresponding to various values of K_D for Method 1, Case 1 (green open circles), Method 1, Case 2 (red triangles), Method 1, Case 3 (blue squares), Method 1, Case 4 (magenta circles).

concentrations as outlined in table 5.1. Using these protein and ligand concentrations, a substantial increase in precision is obtained for K_Ds in the range 2 - 1000 μM, as shown in figure 5.5A. However, reasonable accuracy is only obtained within the K_D range 20 - 600 μM, that is, $\sigma \leq\sim 50\%$ of the actual K_D. In comparison to the K_D parameter, reasonable accuracy for $\Delta\delta_{max}$ ($\sigma <\sim 50\%$ of the actual $\Delta\delta_{max}$) is obtained over a K_D range of 2 - 1000 μM (figure 5.5B). Importantly, substantial increases in accuracy for K_D and $\Delta\delta_{max}$ are achieved over the range 20 - 600 μM. These results indicate that in the absence of prior knowledge for the actual value of K_D, Method 1 will produce K_D and $\Delta\delta_{max}$ values with reasonable accuracy for conditions where $[P_0] > (1 - 25K_D)$.

Precision of K_D and $\Delta\delta_{max}$ parameters for simulations of Method 2: $[P_0]$ and $[L_0]$ decrease by a constant increment

For Method 1, co-variation of the ligand and protein concentrations involves successively diluting the initial protein concentration while simultaneously increasing the ligand concentration. Alternatively, for Method 2, a mixture of concentrated ligand and concentrated protein is successively diluted to conduct a titration (figure 5.6A). We performed Monte Carlo simulations using $[L_0]$ = 1 mM and $[P_0]$ = 500 µM, for a series of K_D values: 2, 20, 60, 200, 600, 1000 and 2000 µM, with a decrease in subsequent protein and ligand concentrations by a constant increment of 50 µM, for a total of ten titration points, as outlined in table 5.2. The Monte Carlo ensembles for the simulation with K_D = 200 µM are shown in Figures 6B and C. It is evident that the precision for K_D = 200 µM remains similar to that for Method 1. However, a substantial increase in precision is obtained for other K_D values in comparison to the traditional method of conducting a titration (Method 1, Case 2 with $\Delta_{[P]} = 0$ and $[P_0] = 0.5$ mM), as well as Method 1 with $\Delta_{[P]} = 0.25$ (figure 5.7). For example, for K_D = 2 µM, $\sigma(K_D)$ and $\sigma(\Delta\delta_{max})$ are ±0.4 µM and ±0.002 ppm, respectively, in comparison to Method 1, Case 4, with K_D = 2 µM, a protein concentration of 100 µM, and $\Delta_{[P]} = 0$, for which $\sigma(K_D) = \pm 1.4$ µM and $\sigma(\Delta\delta_{max}) = \pm 0.007$ ppm. For K_D = 2000 M, $\sigma(K_D)$ and $\sigma(\Delta\delta_{max})$ are ±653 µM and ±0.2 ppm, respectively, in comparison to Method 1, with K_D = 2000 µM, a protein concentration of 100 µM, and $\Delta_{[P]} = 0$, for which $\sigma(K_D) = \pm 1370$ µM and $\sigma(\Delta\delta_{max}) = \pm 0.5$ ppm. These results for Method 2 indicate that without knowledge of the actual value of K_D, Method 2 can produce K_D and $\Delta\delta_{max}$ values with excellent accuracy ($\sigma \leq\sim 25\%$) for conditions where $[P_0] > (1 - 25K_D)$.

Simulated NMR spectra for titrations conducted for Method 2 with K_D = 2, 200, and 2000 µM are shown in figure 5.8. These spectra highlight the fact that while Method 2 provides greater precision over a broad range of K_D values in comparison to Method 1, the observed chemical shift changes over the course of a titration are smaller, particularly at extremes of K_D (2 and 2000 µM), and therefore, Method 2 has more stringent requirements for precise chemical shift measurements, and would be most accurate for residues displaying the largest chemical shift changes in a titration ($\Delta\delta_{max} \sim 1$ ppm).

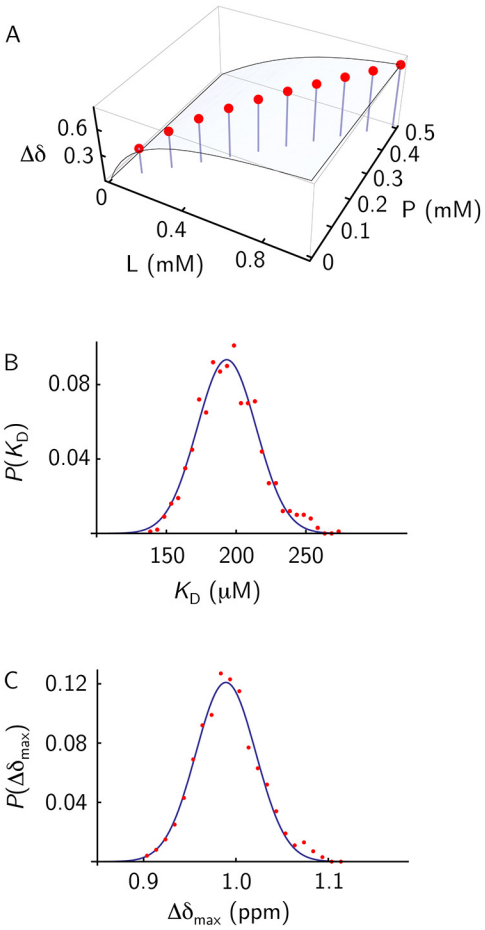

Figure 5.6: (A) Binding isotherm for a 1:1 protein ligand interaction for Method 2, K_{D} = 200 µM. (B) Histogram for 1000 K_{D} values determined from Monte Carlo parameter estimation for Method 2, K_{D} = 200 µM (red dots), and corresponding fit to a normal distribution (blue line, eq. 5.8). (C) Histogram for 1000 $\Delta\delta_{\mathrm{max}}$ values determined from Monte Carlo parameter estimation for Method 2, K_{D} = 200 µM (red dots), and corresponding fit to a normal distribution (blue line, eq. 5.8)

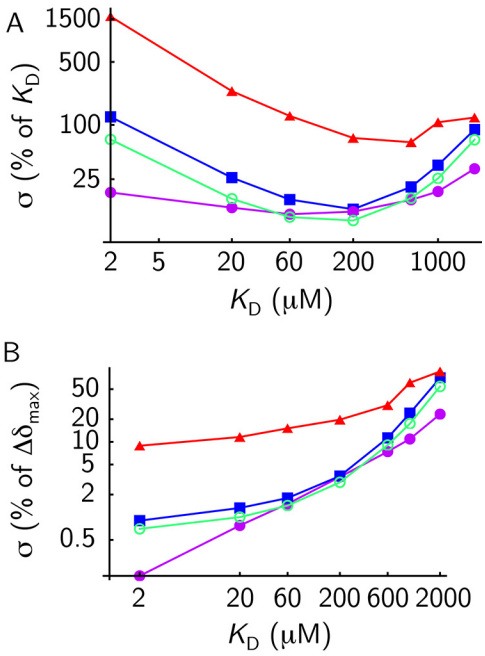

Figure 5.7: (A) Standard deviations of 1000 Monte Carlo trials for various K_D values for Method 1, Case 1 (green open circles), Method 1, Case 2 (red triangles), Method 1, Case 4 (blue squares), Method 2 (magenta circles). (B) Standard deviations of 1000 Monte Carlo trials for $\Delta\delta_{max}$ corresponding to various values of K_D for Method 1, Case 1 (green open circles), Method 1, Case 2 (red triangles), Method 1, Case 4 (blue squares), Method 2 (magenta circles).

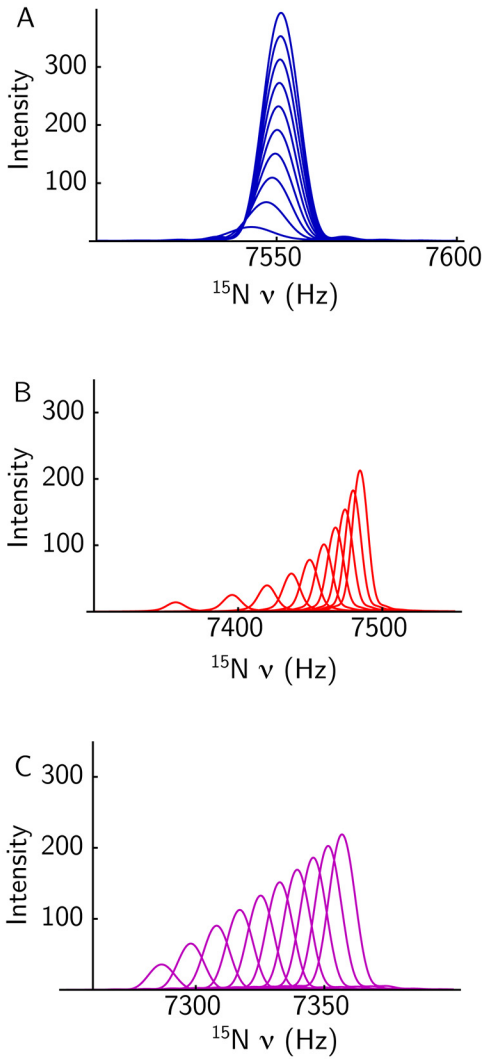

Figure 5.8: (A) Simulated ^{15}N lineshapes for Method 2, $K_{\mathrm{D}} = 2$ µM. (B) Simulated ^{15}N lineshapes for Method 2, $K_{\mathrm{D}} = 200$ µM. (C) Simulated lineshapes for Method 2, $K_{\mathrm{D}} = 2000$ µM. For all spectra, the resonance frequencies for the free and bound protein are 7273.76 and 7552.32 Hz, respectively

In addition to providing increased precision for K_D and $\Delta\delta_{max}$, the average signal to noise ratio for a titration conducted according to Method 2 is greater than that for a traditional NMR titration for which $[P_0]$ is held at a fixed concentration of 0.5 K_D. Assuming that NMR spectra are collected using identical parameters for Method 2 with $[L_0]$ = 1 mM and $[P_0]$ = 500 µM, and a constant decrease in protein and ligand concentrations by 100 µM for a total of ten titration points, and Method 1 with a protein concentration of 100 µM, and $\Delta_{[P]} = 0$, the average signal to noise ratio for Method 2 is 2.75-fold greater. It should be noted that Method 2 is theoretically more accurate than Method 1 because the titration is designed to sample the portion of the two-dimensional binding isotherm that has significant curvature; this occurs at low concentrations of both protein and ligand (figure 5.6A). Thus, care must be taken to ensure sufficient transients are collected for the final points of the titration to ensure adequate signal to noise ratio and therefore, adequate chemical shift precision. In addition, Method 2 requires that δ_{free} for the resonances of the unbound protein are determined from a spectrum separate from the titration. Thus, $\Delta\delta_{max}$ can be expressed as a difference (eq. 5.5), avoiding the use of δ_{free} as a third free parameter in fitting titration data to eq. 5.6.

Experimental validation of Methods 1 and 2 by 2D ^{1}H-^{15}N HSQC NMR-monitored chemical shift titration

A caveat regarding the practical application of Methods 1 and 2 is the necessity of diluting the NMR-observed protein component. This may lead to long acquisition times for the last few NMR spectra to achieve sufficient signal to noise ratio, especially for NMR-monitored titrations for large protein-protein or protein-ligand complexes and/or processes in the intermediate exchange timescale. We conducted NMR-monitored titration experiments using Methods 1 and 2 for the interaction of human [U-^{15}N]-Mms2 (145 residues, 16 kDa) with human ubiquitin (76 residues, 8.5 kDa) (tables 5.3 and 5.4, and figure 5.9). The interaction of Mms2 with ubiquitin has been previously characterized using NMR (14, 19), and the binding occurs with 1:1 stoichiometery (13). A total of five 2D ^{1}H-^{15}N HSQC NMR spectra were collected for both Methods 1 and 2; the associated fits to a 1:1 binding model to extract K_D and $\Delta\delta_{max}$ values for various Mms2 residues are

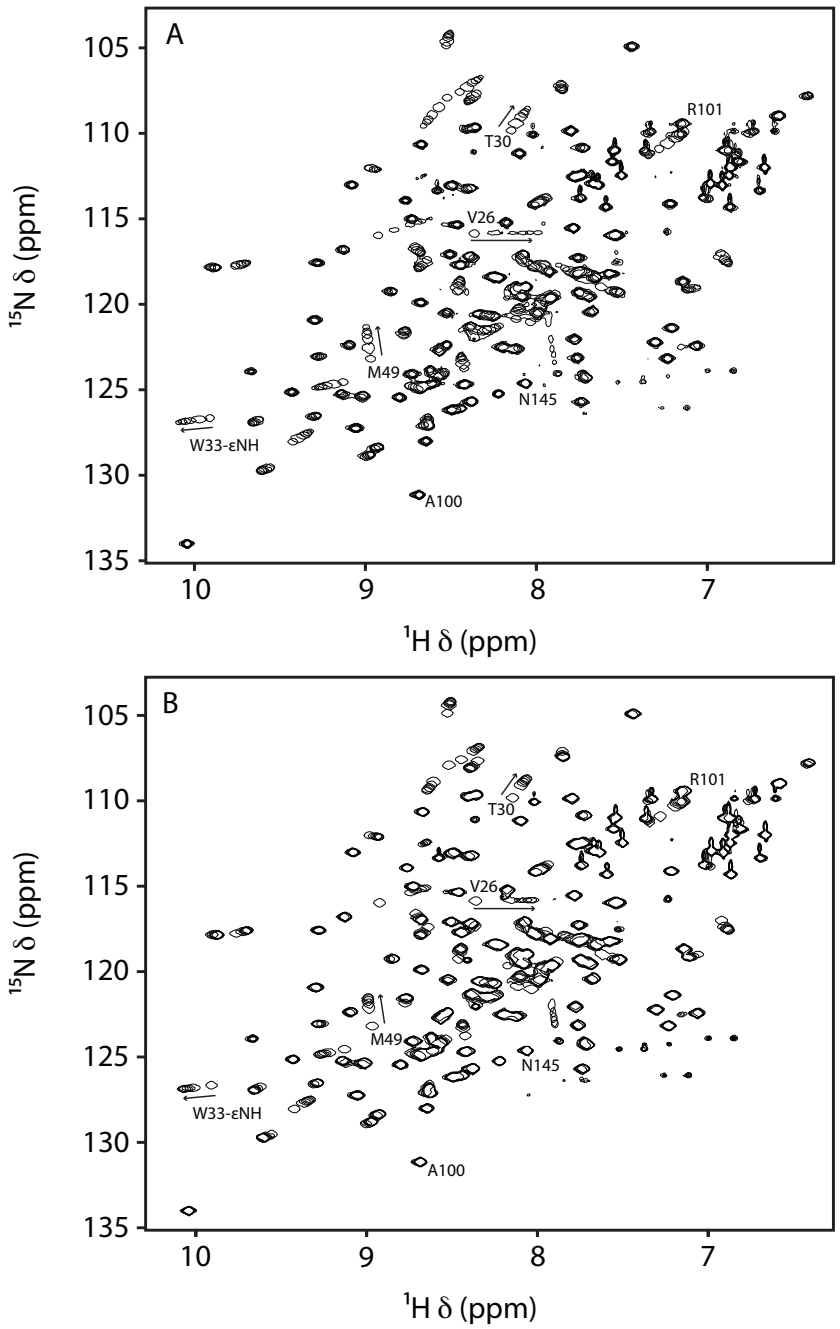

Figure 5.9: 2D ^{1}H-^{15}N HSQC 600 MHz NMR-monitored titrations for the interaction of [U-^{15}N]-Mms2 with ubiquitin for Method 1 (a), and Method 2 (b). A number of residues used in the various analyses are labeled. Arrows indicate the shift from the free to the bound state.

shown in figure 5.10 and tables 5.5 and 5.6. Overall, excellent chemical shift precision and signal to noise can be obtained in a reasonable amount of time using a room temperature triple resonance probe at a magnetic field strength of 600 MHz (14.1 Tesla). For example, for 2D ^{1}H-^{15}N NMR spectra collected at the lowest concentration of the observed protein component, [*U*-^{15}N]-Mms2, the chemical shift precision for the ^{1}H and ^{15}N dimensions is within the ppb range over a wide range of linewidths (figure 5.11). Another indication of the quality of the titration methods are the differences between experimental $\Delta\delta_{\mathrm{obs}}$ values and those determined from the best fits of the data to a 1:1 binding isotherm, given by the expression ($\sqrt{(\Delta\delta_{\mathrm{obs,calc}} - \Delta\delta_{\mathrm{obs,exp}})^2}$). This average difference for M49 ^{15}N and ^{1}H$^{\mathrm{N}}$ chemical shifts for Method 1 is 0.3 and 0.2 ppb, respectively (figure 5.10A). For Method 2, the average difference for M49 ^{15}N and ^{1}H$^{\mathrm{N}}$ chemical shifts is 0.2 and 1.7 ppb, respectively (figure 5.10B). There is excellent agreement between the average K_{D}s for Methods 1 and 2, 0.31 ± 0.02 and 0.35 ± 0.04 mM, respectively (tables 5.5 and 5.6), excluding V26 ^{15}N and M49 ^{1}H$^{\mathrm{N}}$). For the respective experimental conditions, Method 2 is more precise than Method 1, with an average per residue K_{D} error of 0.04 compared to 0.07 mM (tables 5.5 and 5.6). This difference in precision between Methods 1 and 2 is evident in the data in figure 5.10C, which shows that $\pm 1\sigma$ for the K_{D} value of M49 from Method 1 is about twice that of Method 2; consistent with the theoretical prediction that Method 2 should be slightly more precise for $K_{\mathrm{D}} \sim 300$ µM (figure 5.7A). Finally, both Methods 1 and 2 perform reasonably well with respect to the magnitude of $\Delta\delta_{\mathrm{max}}$, the size of the chemical shift change upon binding for a given residue (figure 5.12). Generally, for larger values of $\Delta\delta_{\mathrm{max}}$, between ~ 0.2 and 0.5 ppm, with ^{15}N $\Delta\delta/5$ to facilitate comparison to ^{1}H$^{\mathrm{N}}$, the average error is 18% of the value of K_{D}. In contrast, values of $\Delta\delta_{\mathrm{max}}$ that are smaller than ~ 0.2 ppm have larger errors, $\sim 60\%$ of the value of K_{D}, on average. Thus, as a general rule, only residues with $\Delta\delta_{\mathrm{max}}$ values exceeding ~ 0.2 ppm should be used in determining the K_{D} for a given protein interaction.

Conclusions

For traditional NMR-monitored titrations, $[P_0]$ remains fixed at a constant concentration, and without prior knowledge of the dissociation constant, it is difficult to experimentally

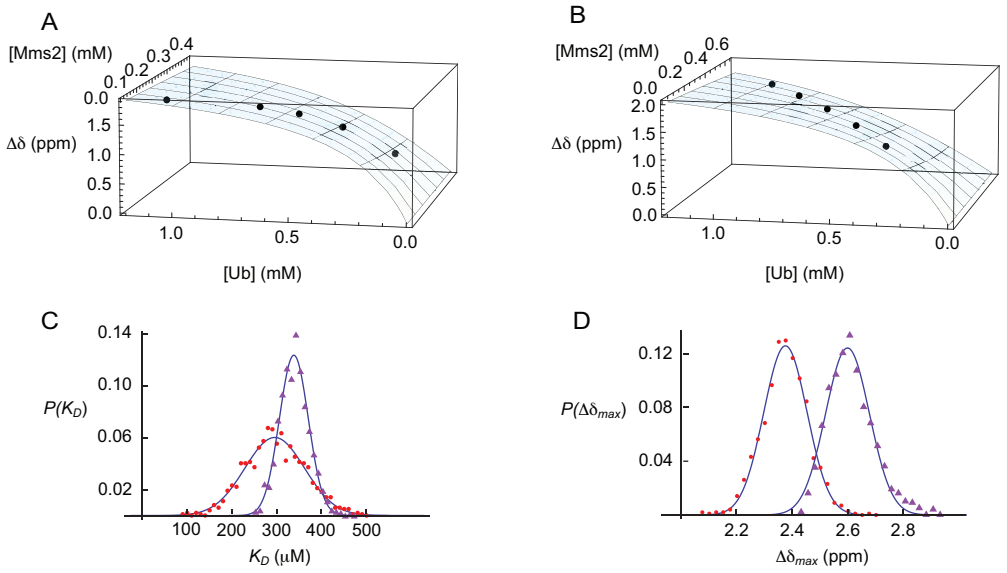

Figure 5.10: (A) Experimental binding isotherm for M49 ^{15}N from NMR-monitored titrations using Method 1 for the interaction of [U-^{15}N]-Mms2 with ubiquitin. (B) Experimental binding isotherm for M49 ^{15}N from NMR-monitored titrations using Method 2 for the interaction of [U-^{15}N]-Mms2 with ubiquitin. (C) Error estimates for the K_{D} parameter from an ensemble of 1000 Monte Carlo simulations for M49 ^{15}N, Method 1 (red circles) and Method 2 (purple triangles). (D) Error estimates for the $\Delta\delta_{\mathrm{max}}$ parameter error for an ensemble of 1000 Monte Carlo simulations for M49 ^{15}N, Method 1 (red circles), and Method 2 (purple triangles). For (C) and (D), the lines through the points represent the best fit to a normal distribution (Eq. 8)

Table 5.5: Fitted parameters from Method 1 for titration of $[U\text{-}^{15}N]$-Mms2 with ubiquitin

Residue	K_{D} (mM)	$\Delta\delta_{\mathrm{max}}$	χ^2 ($\times 10^{-7}$)
V26 ^{15}N	0.01 ± 1	0.01 ± 0.01	4675.38
V26 ^{1}H$^{\mathrm{N}}$	0.33 ± 0.08	0.49 ± 0.03	93.18
M49 ^{15}N	0.30 ± 1	0.48 ± 0.02	4171.11
M49 ^{1}H$^{\mathrm{N}}$	1 ± 3	0.06 ± 0.09	13.30
W33 ^{15}N$_{\epsilon 1}$	0.34 ± 0.1	0.068 ± 0.007	4.25
W33 ^{1}H$_{\epsilon 1}$	0.30 ± 0.07	0.24 ± 0.009	99.15
T30 ^{15}N	0.30 ± 0.06	0.33 ± 0.01	160.05
T30 ^{1}H$^{\mathrm{N}}$	0.27 ± 0.07	0.115 ± 0.006	4.88

Table 5.6: Fitted parameters from Method 2 for titration of [U-^{15}N]-Mms2 with ubiquitin

Residue	K_D (mM)	$\Delta\delta_{max}$	χ^2 ($\times 10^{-7}$)
V26 ^{15}N	10 ± 4	0.10 ± 0.05	1529.85
V26 $^{1}H^{N}$	0.42 ± 0.06	0.55 ± 0.03	801.40
M49 ^{15}N	0.34 ± 0.03	0.52 ± 0.02	2489.66
M49 $^{1}H^{N}$	0.13 ± 0.08	0.0028 ± 0.003	7.31
W33 $^{15}N_{\epsilon 1}$	0.37 ± 0.05	0.071 ± 0.004	270.39
W33 $^{1}H_{\epsilon 1}$	0.32 ± 0.03	0.253 ± 0.008	30.00
T30 ^{15}N	0.32 ± 0.03	0.348 ± 0.007	474.49
T30 $^{1}H^{N}$	0.36 ± 0.04	0.130 ± 0.005	8.71

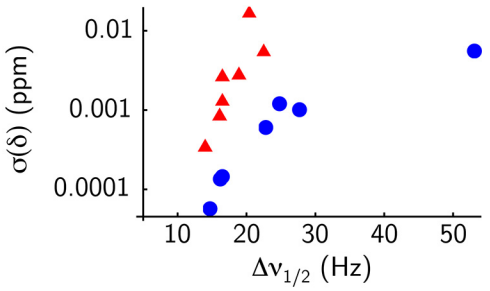

Figure 5.11: Standard deviations of 10000 Monte Carlo trials for chemical shifts determined from parabolic peak interpolation as a function of linewidth. Peak interpolation was conducted on a variety of cross-peaks from 2D $^{1}H^{15}N$ HSQC 600 MHz NMR spectra corresponding to the lowest concentration of [U-^{15}N]-Mms2. Data for $^{1}H^{N}$ are shown as blue circles, whereas data for ^{15}N are shown as red triangles

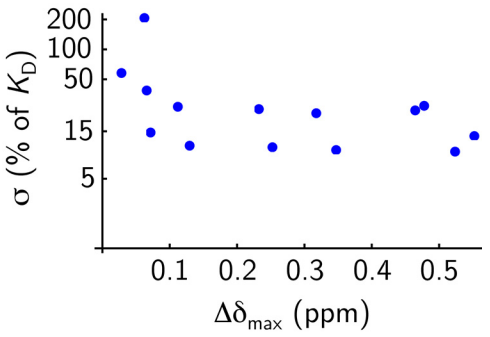

Figure 5.12: Error in K_D as a function of the maximum chemical shift change upon binding ($\Delta\delta_{max}$). Monte Carlo parameter estimation was conducted on a variety of cross-peaks (labeled in fig. 5.9) from NMR-monitored titration data for Methods 1 and 2. Data include both ^{15}N and $^{1}H^{N}$; ^{15}N chemical shifts are scaled by a factor of 1/5

measure an accurate K_D, given the requirement that the optimal value of $[P_0]$ should be $\sim 0.5\ K_D$. We have demonstrated that, in general, co-variation of the ligand and protein concentrations during the course of an NMR-monitored titration leads to increased precision in the parameters K_D and $\Delta\delta_{max}$ for 1:1 protein-protein and protein-ligand interactions. The general utility of co-variation of protein and ligand concentrations is two-fold: accurate K_Ds can be measured using higher starting concentrations for $[P_0]$; secondly, larger $[P_0]$ values translate into greater average signal to noise ratios for the NMR spectra in a given titration. Of the two methods that were analyzed, Method 2, which involves serial dilution of a solution of concentrated protein and ligand, provides fairly robust precision ($\sigma \leq\sim 25\%$) over a broad range of K_D. Some caveats regarding both methods include the fact that observed chemical shift changes for Method 2 are smaller than Method 1, or traditional NMR titrations, and require that chemical shifts be measured with high precision. Secondly, both methods rely on dilution of the NMR-observed protein component, and may require longer acquisition times for the last few titration points to achieve sufficient signal to noise ratio. In this regard, we have demonstrated experimentally that the proposed methods are practical, and straightforward to apply to protein-protein interactions for complexes up to 24 kDa with binding kinetics in the fast exchange regime on the NMR timescale. NMR-monitored titrations involving larger protein-protein or protein-ligand complexes and/or processes in the intermediate exchange timescale become challenging to analyze due to extensive line broadening. However, approaches such as TROSY or CRINEPT NMR spectroscopy (20), coupled with deuteration of the protein components (21), can alleviate problems associated with linebroadening for NMR-monitored titrations (8).

References

1. Zuiderweg, E.R.P. (2002) Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* 41, 1–7.

2. Rintala-Dempsey, A.C., Santamaria-Kisiel, L., Liao, Y., Lajoie, G., and Shaw, G.S. (2006) Insights into S100 target specificity examined by a new interaction between S100A11 and annexin A2. *Biochemistry* 45, 14695–14705.

3. Fielding, L. (2007) NMR methods for the determination of protein-ligand dissociation constants. *Progress in Nuclear Magnetic Resonance Spectroscopy* 51, 219–242.

4. Haririnia, A., D'Onofrio, M., and Fushman, D. (2007) Mapping the interactions between Lys48 and Lys63-linked di-ubiquitins and a ubiquitin-interacting motif of S5a. *Journal of Molecular Biology* 368, 753–766.

5. Marintchev, A., Frueh, D., and Wagner, G. (2007) NMR methods for studying protein–protein interactions involved in translation initiation. *Methods in Enzymology* 430, 283–331.

6. Baryshnikova, O.K., Robertson, I.M., Mercier, P., and Sykes, B.D. (2008) The dilated cardiomyopathy G159D mutation in cardiac troponin C weakens the anchoring interaction with troponin I. *Biochemistry* 47, 10950–10960.

7. Markin, C.J., Saltibus, L.F., Kean, M.J., McKay, R.T., Xiao, W., and Spyracopoulos, L. (2010) Catalytic proficiency of ubiquitin conjugation enzymes: Balancing pK_{a} suppression, entropy, and electrostatics. *Journal of the American Chemical Society* 132, 17775–17786.

8. Markin, C.J., Xiao, W., and Spyracopoulos, L. (2010) Mechanism for recognition of polyubiquitin chains: balancing affinity through interplay between multivalent binding and dynamics. *Journal of the American Chemical Society* 132, 11247–11258.

9. Granot, J. (1983) Determination of dissociation constants of 1:1 complexes from NMR data. Optimization of the experimental setup by statistical analysis of simulated experiments. *Journal of Magnetic Resonance (1969)* 55, 216–224.

10. Cavanagh, J. (2007) *Protein NMR spectroscopy: principles and practice*. (Academic Press).

11. Wolfram, S. (1999) *The Mathematica book*. (Cambridge University Press).

12. Palmer, A.G., Kroenke, C.D., and Loria, J.P. (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods in Enzymology* 339, 204–238.

13. Spyracopoulos, L., Lewis, M.J., and Saltibus, L.F. (2005) Main chain and side chain dynamics of the ubiquitin conjugating enzyme variant human Mms2 in the free and ubiquitin-bound states. *Biochemistry* 44, 8770–8781.

14. Lewis, M.J., Saltibus, L.F., Hau, D.D., Xiao, W., and Spyracopoulos, L. (2006) Structural basis for non-covalent interaction between ubiquitin and the ubiquitin conjugating enzyme variant human MMS2. *Journal of Biomolecular NMR* 34, 89–100.

15. Kay, L.E., Keifer, P., and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *Journal of the American Chemical Society* 114, 10663–10665.

16. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, AD. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* 6, 277–293.

17. McKenna, S., Moraes, T., Pastushok, L., Ptak, C., Xiao, W., Spyracopoulos, L., and Ellison, M.J. (2003) An NMR-based model of the ubiquitin-bound human ubiqui-

tin conjugation complex Mms2Ubc13. *Journal of Biological Chemistry* 278, 13151–13158.

18. Goddard, T.D. and Kneller, D.G. (2006). *Sparky* University of California, San Francisco.

19. McKenna, S., Hu, J., Moraes, T., Xiao, W., Ellison, M.J., and Spyracopoulos, L. (2003) Energetics and specificity of interactions within UbUevUbc13 human ubiquitin conjugation complexes. *Biochemistry* 42, 7922–7930.

20. Riek, R., Pervushin, K., and Wüthrich, K. (2000) TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution. *Trends in Biochemical Sciences* 25, 462–468.

21. Gardner, K.H. and Kay, L.E. (1998) The use of ^{2}H, ^{13}C, an ^{15}N multidimensional NMR to study the structure and dynamics of proteins. *Annual Review of Biophysics and Biomolecular Structure* 27, 357–406.

Chapter 6

Accuracy and precision of protein-ligand interaction kinetics determined from chemical shift titrations*

Introduction

Weak protein-protein interactions with fast kinetics are fundamental for the temporal flow of information in signaling networks, regulating varied life processes such as DNA repair (1), innate immunity (2), and trafficking (3). From its inception, NMR spectroscopy has played a vital role as an enabling technology for the analysis of molecular architecture and dynamics; allowing for pioneering investigations of the thermodynamics and kinetics of chemical exchange in small molecule systems (4, 5), as well as the binding of the

*This chapter has been published. C.J. Markin, and L. Spyracopoulos. Accuracy and precision of protein-ligand interaction kinetics determined from chemical shift titrations. *Journal of Biomolecular NMR*, 54(4): 355-376, 2012. Reproduced with kind permission from Springer Science and Business Media.

Contributions: L.S. performed the GAMMA simulations and line-shape analyses. C.J.M. performed the experimental titrations and together with L.S. performed line-shape analysis on the experimental data. C.J.M. designed and performed the line-shape simulations using the classical Bloch-McConnell equations to assess the accuracy of K_{D} values nearing the intermediate exchange regime. L.S. and C.J.M. wrote the paper.

protein serum albumin to penicillin G (6), through chemical shift titration and line shape analysis.

Traditionally, line shape analysis has involved application of the Bloch equations, derived using classical physics, to describe the motion of nuclear spins, with general modifications to include the effects of chemical exchange (7). Such a phenomenological approach is limited however, in that classical line shape equations are not applicable to coupled spin systems, and may lead to inaccurate kinetics, particularly for high-resolution NMR spectra of small molecules (8, 9). To address this limitation, a unified quantum mechanical theory for the description of NMR line shapes in the presence of chemical exchange has been developed (8). However, in the absence of spin-spin couplings, the modified Bloch equations are equivalent to quantum mechanical formulations (10).

For macromolecular systems, early line shape applications involved one dimensional ^{1}H or heteronuclear NMR spectroscopy with analyses using the Bloch equations or the density matrix approach (11–16). The advent of multidimensional, heteronuclear NMR spectroscopy has allowed for the study of the structures and dynamics of proteins of moderate size ($\sim$ 30 kDa) (17). Subsequent methodological advances have extended the applicability of NMR to proteins in the 100 kDa range (18, 19). Blending multidimensional, heteronuclear NMR methods with chemical shift titrations and line shape analysis has provided detailed insights into the thermodynamics and kinetics of diverse biological processes and molecular recognition events (20–29).

The potential of utilizing multidimensional heteronuclear NMR methods such as the ubiquitous ^{1}H-^{15}N HSQC NMR experiment (30–33) to determine both thermodynamic quantities such as dissociation constants (populations), and kinetics from a single set of spectra corresponding to a protein-protein or protein-ligand titration is alluring. It is noteworthy that line shape analysis is particularly valuable for the study of molecular recognition events in the fast exchange regime ($k_{\rm ex} \gg |\Delta\omega|$), given that values for populations and rates can be determined. Furthermore, line shape analysis facilitates the identification and evaluation of multi-state binding equilibria (34). However, the general consensus is that line shape analysis is mainly a qualitative, or semi-quantitative method, on account of caveats that it is unlikely to be practical for large proteins, and that limited digital resolution in heteronuclear experiments leads to inaccurate kinetics

(28, 35). In addition to these caveats, care must be taken to account for differential intensity losses between the HSQC spectra of a titration due to relaxation and exchange during INEPT polarization transfers, as well as during the collection of two-dimensional spectra (22, 23). Furthermore, the accuracy of the kinetics obtained from line shape analysis is dependent upon the accuracy of the intrinsic transverse relaxation rates (R_2^0), or line widths ($\Delta\nu_{1/2} = R_2^0/\pi$) of the exchanging resonance peaks, and the accuracy of the K_{D} value when it is determined separately and employed in the line shape analysis through the relationship $K_{\mathrm{D}} = k_{\mathrm{off}}/k_{\mathrm{on}}$.

We recently developed 2D ^{1}H-^{15}N HSQC NMR-monitored chemical shift titration methods wherein co-variation of the total protein and ligand concentrations allow for the determination of precise K_{D} values for 1:1 protein-protein interactions with kinetics in the fast exchange regime (36). The methodology was applied to study the interaction of human ubiquitin (76 residues) with the human ubiquitin binding protein [U-^{15}N]-Mms2 (145 residues). In this study, we demonstrate that classical line shape analysis applied to the same set of 2D ^{1}H-^{15}N HSQC NMR spectra from our previous chemical shift titrations (36), yields accurate (< 5 %) and precise (~ 13 %) experimental off-rates (k_{off}) in the range of $\sim$ 3,000 s^{-1} for 1:1 protein-protein ligand interactions.

The accuracy of line shape analysis was determined by conducting quantum mechanical NMR simulations of the chemical shift titration methods with the magnetic resonance toolkit GAMMA (37). For these theoretical simulations, the inclusion of concentration errors for line shape fitting, as well as the addition of noise to FIDs calculated with GAMMA, allowed us to conduct Monte Carlo simulations to evaluate the theoretical precision for k_{off} determined using classical line shape analysis, for the 1:1 interaction between ubiquitin and Mms2. The experimental precision for k_{off} was also determined using Monte Carlo simulations, wherein the main sources of experimental error were assumed to arise from concentration errors and noise in the FID. The consistency between the theoretical and experimental precision and the standard deviation for the per residue k_{off} values indicates that the main sources of random error in the line shape analyses arise from concentration error and the thermal electromotive force (noise) in the NMR probe coil. In addition, the magnitude of the accuracy and precision from Monte Carlo trials of the GAMMA simulations indicate that the typical digital resolution for the di-

rect and indirect dimensions of 2D ^{1}H-^{15}N HSQC NMR spectra is not an impediment to accurate and precise line shape analysis. We also demonstrate that systematic errors in R_2^0 and K_D do not seriously compromise the accuracy of the fitted k_{off} from classical line shape analysis as implemented in this study. We also designed a number of GAMMA simulations to determine that the accuracy and precision of classical line shape analysis over a range of k_{off} from 100 to 15,000 s^{-1} is reasonable. Finally, we show that the applicability of line shape analyses for k_{off} values in the intermediate exchange regime can be facilitated by more accurate K_D values obtained from chemical shift titrations. For these NMR-monitored chemical shift titrations, extrapolation of the theoretical dependence of K_D values on the chemical shift difference between the free and bound states to $\Delta\omega = 0$ leads to more accurate K_D values.

Theory and methods

Chemical exchange for 1:1 protein-ligand interactions

The interaction between a protein and cognate ligand to form a protein-ligand complex is given by the reaction (35, 38):

$$P + L \underset{k_{off}}{\overset{k_{on}[L]}{\rightleftharpoons}} PL \tag{6.1}$$

where P, L, and PL are the concentrations of free protein, free ligand, and protein-ligand complex, respectively, k_{on} is the on-rate constant for association of ligand and protein, and k_{off} is the off-rate constant for the ligand from the complex. Subsequent to initial mixing of protein and ligand, equilibrium is reached. Equation 6.1 describes chemical exchange between two sites, in this case, the free and bound states of a protein. In the absence of scalar coupling, the Bloch equations as modified by McConnell provide a rigorous description of the NMR line shape for a two-site system undergoing chemical exchange; the time-dependence of the transverse magnetization is given by (35):

$$M_A(t) = M_A(0)a_{11}(t) + M_B(0)a_{12}(t) \tag{6.2}$$

$$M_{\mathrm{B}}(t) = M_{\mathrm{B}}(0)a_{22}(t) + M_{\mathrm{A}}(0)a_{21}(t) \tag{6.3}$$

for spins A and B, with coefficients:

$$\begin{aligned} a_{11}(t) = \frac{1}{2}\Bigg[&\left(1 - \frac{-i\Delta\omega + R^0_{2A} - R^0_{2B} + k_{\mathrm{ex}}(p_{\mathrm{B}} - p_{\mathrm{A}})}{\lambda_+ - \lambda_-}\right)\exp\left(-\lambda_- t\right) \\ &+ \left(1 + \frac{-i\Delta\omega + R^0_{2A} - R^0_{2B} + k_{\mathrm{ex}}(p_{\mathrm{B}} - p_{\mathrm{A}})}{\lambda_+ - \lambda_-}\right)\exp\left(-\lambda_- t\right)\Bigg] \end{aligned} \tag{6.4}$$

$$\begin{aligned} a_{22}(t) = \frac{1}{2}\Bigg[&\left(1 + \frac{-i\Delta\omega + R^0_{2A} - R^0_{2B} + k_{\mathrm{ex}}(p_{\mathrm{B}} - p_{\mathrm{A}})}{\lambda_+ - \lambda_-}\right)\exp\left(-\lambda_- t\right) \\ &+ \left(1 - \frac{-i\Delta\omega + R^0_{2A} - R^0_{2B} + k_{\mathrm{ex}}(p_{\mathrm{B}} - p_{\mathrm{A}})}{\lambda_+ - \lambda_-}\right)\exp\left(-\lambda_- t\right)\Bigg] \end{aligned} \tag{6.5}$$

$$a_{12}(t) = \frac{k_{\mathrm{ex}}p_{\mathrm{A}}}{\lambda_+ - \lambda_-}[\exp -\lambda_- t - \exp -\lambda_+ t] \tag{6.6}$$

$$a_{21}(t) = \frac{k_{\mathrm{ex}}p_{\mathrm{B}}}{\lambda_+ - \lambda_-}[\exp -\lambda_- t - \exp -\lambda_+ t] \tag{6.7}$$

In the absence of chemical exchange, R^{02}_{A} and R^{02}_{B} are the intrinsic transverse relaxation rates for spins A and B, respectively, $\Delta\omega$ is the difference between the chemical shifts of spin A (ω_{A}) and spin B (ω_{B}) in rad s^{-1}, the rate of chemical exchange is given by $k_{\mathrm{ex}} = k_{\mathrm{on}}[B] + k_{\mathrm{off}}$, p_{A} and p_{A} are the populations of spin A and B, respectively, and

$$\begin{aligned} \lambda_\pm = \frac{1}{2}\Bigg\{&-i\Omega_A - i\Omega_B + R^0_{2A} + R^0_{2B} + k_{\mathrm{ex}} \\ &\pm \sqrt{(-i\Delta\omega + R^0_{2A} - R^0_{2B} + k_{\mathrm{ex}}(p_B - p_A))^2 + 4p_a p_b k^2_{\mathrm{ex}}}\Bigg\} \end{aligned} \tag{6.8}$$

The NMR line shape for a two-spins-1/2 system without scalar couplings, derived from the Bloch-McConnell equations, is equivalent to that calculated using a quantum mechanical density matrix approach (10). In the absence of FID truncation and post-acquisition processing, NMR spectra from chemical shift titrations can be fit to the Fourier transforms of the sum of the auto and cross-correlation peaks (equations 6.2 and 6.3), or similar frequency domain expressions (39), to estimate the value of k_{ex}.

NMR simulations of chemical shift titrations with GAMMA

We recently developed two chemical shift titration methods by which precise protein-ligand dissociation constants (K_{D}) can be derived through co-variation of the total protein and ligand concentrations (36). The first, Method 1, involves addition of aliquots of concentrated ligand solution to a concentrated protein solution with concomitant decreases in protein concentration by a constant factor. Method 2 involves the sequential dilution of a solution of concentrated ligand and protein, in an initial $\sim$ 2:1 ratio. To assess the accuracy and precision of these methods with respect to the determination of kinetic parameters such as the k_{off} rate constant in Eq. 6.1, we used the magnetic resonance toolkit GAMMA to perform quantum-mechanical simulations (37). Simulations were designed to produce ^{1}H line shapes similar, but not identical, to experimentally observed line shapes for the W33 side chain $^{1}\mathrm{H}\epsilon 1$ in human Mms2 upon titration with human Ub (36). A five spins-1/2 system was employed, and included the scalar coupled indole $^{15}\mathrm{N}_{\epsilon 1}$-$^{1}\mathrm{H}_{\epsilon 1}$ pair (^{1}J was set to 93 Hz, though the actual value is slightly larger at 99 Hz), and three protons within 3 Å of the indole proton from the crystallographically determined structure of Mms2. One-dimensional ^{1}H NMR spectra at a Larmor frequency of 600 MHz (magnetic field strength of 14.1 Tesla) were simulated with the OverBodenhausen pulse sequence (31) (acquisition times $t_1 = 0$, $t_2 = 122$ ms), the ^{1}H spectral width was 8,000 Hz and 977 points were collected. Ideal ^{1}H and ^{15}N pulses were employed, and dipole-dipole relaxation effects between spins were included, using correlation times of 8 ns for free Mms2 (145 residues) and 13 ns for Mms2 bound to Ub (total of 221 residues). The chemical shift of W33 $^{1}\mathrm{H}_{\epsilon 1}$ in the free state was set to 9.909 ppm (5,945.4 Hz), and 10.159 ppm (6,095.4 Hz) in the Ub-bound state. Post-acquisition processing included application of a cosine window function with a cutoff at 99 % of the length of the FID, and zero filling to 2,048 points. To simulate the effects of exchange between the free and Ub-bound states of Mms2, a K_{D} value of 300 µM with a $k_{\mathrm{off}} = 2500$ s^{-1} was used, as well as the protein and ligand concentrations listed in Tables 6.1 and 6.2. The fractional populations of free (p_{A}) and bound (p_{B}) protein (Mms2), for Methods 1

Table 6.1: Protein and ligand concentrations for GAMMA simulations of Method 1, $k_{off} = 2500$ s^{-1}, $K_D = 300$ µM

$[P_T]$[a](mM)	0.33	0.28	0.23	0.19	0.15	0.06
$[L_T]$[b](mM)	0.0	0.18	0.39	0.56	0.71	1.06
p_A[c]	1	0.738	0.518	0.402	0.330	0.228
p_B[d]	0	0.262	0.482	0.598	0.670	0.772
k_{on}[L][e](s^{-1})	0	888.3	2326.2	3719.8	5079.0	8447.5

a Total concentration of protein (Mms2)
b Total concentration of ligand (Ub)
c Population of free Mms2
d Population of Mms2 bound to Ub
e [L] is the free ligand, or Ub concentration, $k_{on} = 8.33 \times 10^6$ M^{-1} s^{-1}

Table 6.2: Protein and ligand concentrations for GAMMA simulations of Method 2, $k_{off} = 2500$ s^{-1}, $K_D = 300$ µM

$[P_T]$[a](mM)	0.44	0.59	0.49	0.39	0.29	0.20
$[L_T]$[b](mM)	0.0	1.0	0.83	0.67	0.50	0.33
p_A[c]	1	0.331	0.366	0.406	0.465	0.555
p_B[d]	0	0.669	0.634	0.594	0.535	0.445
k_{on}[L][e](s^{-1})	0	5045.6	4328.3	3653.7	2874.2	2007.7

a Total concentration of protein (Mms2)
b Total concentration of ligand (Ub)
c Population of free Mms2
d Population of Mms2 bound to Ub
e [L] is the free ligand, or Ub concentration, $k_{on} = 8.33 \times 10^6$ M^{-1} s^{-1}

and 2 were calculated according to:

$$p_A = \frac{-K_D - L_T + P_T + \sqrt{K_D^2 + (L_T - P_T)^2 + 2K_D(L_T + P_T)}}{2P_T} \tag{6.9}$$

$$p_B = 1 - p_A \tag{6.10}$$

where P_T and L_T are the total protein and ligand concentrations, respectively. The free ligand concentration for the pseudo first order rate constant k_{on}[L] (Eq. 6.1) was calculated according to $k_{on} = k_{off}/K_D$ using:

$$[L] = \frac{1}{2}\left(-K_D + L_T - P_T + \sqrt{K_D^2 + (L_T - P_T)^2 + 2K_D(L_T + P_T)}\right) \tag{6.11}$$

To assess the range of validity of line shape analysis, we also carried out GAMMA simulations for Methods 1 and 2 using the protein and ligand concentrations in Tables 6.1 and 6.2, with k_{off} values of 1,500 and 15,000 s^{-1} and respective K_D values of 180 and 1,800 µM. We also assessed the accuracy of a traditional titration (constant P_T) for

k_{off} values of 1,500 and 15,000 s^{-1} and where the protein concentration is held at 0.5 mM, whereas the ligand concentration is varied according to values given in Table 6.1. To assess the impact of unresolved and partially resolved $^{1}H^{N}$-$^{1}H_{\alpha}$ ^{3}J couplings on the accuracy of line shape analysis, three GAMMA simulations were conducted using ligand concentrations as described in Table 6.1 for Method 1 and the protein concentration fixed at 0.5 mM, with the amide proton weakly coupled to a spin-1/2 nucleus with J values of 1.8, 5, and 10 Hz.

GAMMA simulations for ^{15}N were designed to produce similar, not identical, spectra to those experimentally observed for the main chain ^{15}N from M49 in human Mms2 upon titration with human Ub (36). A three spins-1/2 system was used, which included the scalar coupled main chain ^{15}N-$^{1}H^{N}$ pair (^{1}J set to 93 Hz), and the main chain $^{1}H_{\alpha}$. One-dimensional ^{15}N NMR spectra at a magnetic field strength of 14.1 Tesla (^{1}H Larmor frequency of 600 MHz) were simulated with the OverBodenhausen pulse sequence (acquisition time t_1 = 118.6 ms), the ^{15}N spectral width was 2,428.42 Hz and 288 points were collected. As in the case of simulated ^{1}H spectra, ideal ^{1}H and ^{15}N pulses were employed, and dipole-dipole relaxation effects between spins were included, using correlation times of 8 ns for free Mms2 (145 residues) and 13 ns for Mms2 bound to Ub (total of 221 residues). The chemical shift of M49 ^{15}N in the free state was set to 122.53 ppm (7,479.73 Hz), and 120.0 ppm (7,325.29 Hz) in the Ub-bound state. Post-acquisition processing included application of a cosine window function with a cutoff at 99 % of the length of the FID, and zero filling to 1,024 points. The effects of exchange between free and Ub-bound states of Mms2 for Methods 1 and 2 were simulated using a K_D value of 300 µM with k_{off} = 2500 s^{-1}, and the Mms2 and Ub concentrations listed in Tables 6.1 and 6.2.

To test the range of validity of ^{15}N line shape analysis, we also carried out GAMMA simulations using Methods 1 and 2, with the protein and ligand concentrations in Tables 6.1 and 6.2, and k_{off} values of 100, 500, 1,000, and 15,000 s^{-1}, with respective K_D values of 12, 60, 120, and 1,800 µM. We also assessed the accuracy of traditional titrations that is, P_T held at 0.5 mM, for k_{off} values of 5, 500, 1,000 and 15,000 s^{-1} with respective K_D values of 0.6, 60, 120, and 1,800 µM, and where the ligand concentration is varied according to values given in Table 6.1.

Experimental NMR-monitored titrations

In this study, we applied classical line shape analysis on previous 2D ^{1}H-^{15}N HSQC NMR-monitored titrations of human Ub into [U-^{15}N]-Mms2 conducted using our recently developed chemical shift titrations (Methods 1 and 2) (36). Details regarding sample preparation, protein (Mms2) and ligand (Ub) concentrations, NMR data acquisition, and spectral processing are the same as previously reported (36), with the exception that the ^{15}N dimension for the chemical shift titrations was extended by linear prediction to a total of 384 points, multiplied by a 90°-shifted sine window, and zero filled to 1,024 points.

NMR line shape analyses for experimental and simulated NMR-monitored chemical shift titrations

In general, line shape analyses for simulated and experimental 2D ^{1}H-^{15}N HSQC NMR-monitored titrations were conducted using the Bloch-McConnell equations, modified to account for the application of time-domain apodization functions, within the program Mathematica 8.0.4. The detailed protocol for line shape analysis is as follows: analytical expressions for the frequency domain NMR spectra in the presence of exchange between the free and bound states were obtained from the Fourier transforms of the sum of the FIDs given by eqs. 6.2-6.8, after multiplication by sine time-domain window functions:

$$F(\omega) = M_0 \int_0^{t_{\max}} \sin\left(n\pi + \frac{c\pi(1-n)t}{t_{\max}}\right)^m (a_{11}(t) + a_{12}(t) + a_{22}(t) + a_{21}(t)) \exp(-i\omega t) dt \tag{6.12}$$

where n gives the shift in the time axis from a sine function ($n = 0$) to a cosine function ($n = 0.5$), c is the fractional cut-off in the time axis (typically 0.99), $m = 1$ (sine) or 2 (sine-squared), $t_{\max}$ is the acquisition time, and M_0 is a parameter to scale the arbitrary intensity of the frequency domain spectrum. For simulations of $^{1}H^{N}$ spectra using the Bloch-McConnell equations, in the presence of $^{3}J_{HNH\alpha}$ couplings, the spectrum can be assumed to be a weighted superposition of two uncoupled spectra corresponding to the

upfield and downfield multiplet components, to a first approximation (15):

$$F_{\mathrm{J}}(\omega) = \frac{1}{2}\left(F_{+J/2}(\omega) + F_{-J/2}(\omega)\right) \tag{6.13}$$

where $F_{+J/2}(\omega)$ and $F_{-J/2}(\omega)$ are the J-coupled multiplet components for the free and bound protein resonances. Thus, eqs. 6.2-6.8 are modified such that a spectrum with the free and bound resonance frequencies ω_{A}, ω_{B} is represented by the weighted sum of two spectra at free and bound frequencies of $\omega_{\mathrm{A}} + \pi J$, $\omega_{\mathrm{B}} + \pi J$ and $\omega_{\mathrm{A}} - \pi J$, $\omega_{\mathrm{B}} - \pi J$.

One-dimensional traces from the proton dimension for the W33 indole side chain $^{1}\mathrm{H}_{\epsilon 1}$, and from the nitrogen dimension for the main chain amide for M49 were taken from the series of 2D ^{1}H-^{15}N HSQC NMR spectra acquired for the chemical shift titrations conducted using Methods 1 and 2. For each combination of free and bound protein populations (Tables 6.1, 6.2), traces were multiplied by their respective normalized protein concentrations, relative to the titration point with the most concentrated protein. In addition, each trace was multiplied by a factor to correct for differences in the number of transients collected relative to the spectrum with the fewest number of transients (nt = 8 or 16), that is, the respective factors were 8/nt or 16/nt for traces where nt ¿ 8 or 16. In addition to these corrections, Eq. 6.13 is multiplied by a parameter for the overall intensity (M_0), whose value and that of the intrinsic transverse relaxation rate ($R^0_{2\mathrm{A}}$), are derived from fits to Eq. 6.13 with $p_{\mathrm{A}} = 1$, using the trace taken from the spectrum of free protein prior to line shape fitting of the entire titration. Note that $M_{\mathrm{A}}(0)$ and $M_{\mathrm{B}}(0)$ are set to 1 in eq. 6.12. The intrinsic transverse relaxation rate for the fully bound state, $R^0_{2\mathrm{B}}$, was taken to be linearly dependent on the molecular mass increase for the bound state in comparison to the free state, that is, $R^0_{2\mathrm{B}} = 1.5 \times R^0_{2\mathrm{A}}$.

The preliminary steps described above yield traces from the individual NMR spectra of the chemical shift titrations with intensities corrected to match the following: an arbitrary experimental intensity factor, the concentration of the observed protein component, and the number of transients collected per titration point. Subsequently, traces are subjected to global minimization of the following objective function:

$$\chi(\omega_{\mathrm{B}}, k_{\mathrm{off}}, h_{\mathrm{h}>1})^2 = \sum_{k=1}^{N}\sum_{l=1}^{np}\frac{(h_{\mathrm{k}}F^k_{\mathrm{calc}}(\omega_l) - F^k_{\mathrm{obs}}(\omega_l))^2}{\sigma^2_{\mathrm{k}}} \tag{6.14}$$

where k is the kth NMR spectral trace from the kth titration point from the total N calculated (F_{calc}) or observed (F_{obs}) NMR spectra, with corresponding protein and ligand concentrations, l is the lth frequency point from the total np number of points of the calculated (F_{calc}) or observed (F_{obs}) kth NMR trace, and σ_{k} is the baseline noise from the kth NMR spectral trace from the kth titration point. The adjustable parameters in the objective function include the chemical shift for the bound state of the observed protein component (Ω_{B}), the off-rate of the protein-ligand interaction (k_{off}), and parameters ($h_{k>1}$) to optimize the calculated intensity of the kth trace, with the exception of the first trace, for which $h_1 = 1$. This trace corresponds to the free state of the observed protein; the arbitrary intensity of this trace is determined prior to line shape analysis through optimization of M_0, as described above. The remaining intensity parameters ($h_{k>1}$), are individually optimized for the k individual NMR traces. The individual optimization of intensity parameters has been demonstrated to correct for differential intensity losses, in comparison to the free state, between 2D ^{1}H-^{15}N HSQC spectra for different titration points as a result of relaxation and exchange in the dimension other than that being used for line shape analysis, as well as differential relaxation losses during INEPT polarization transfers (22, 23). For the calculated NMR traces (F_{calc}) in the objective function, k_{ex} (eqs. 6.4-6.8, $k_{\mathrm{ex}} = k_{\mathrm{on}}[\mathrm{L}] + k_{\mathrm{off}}$) is given by:

$$k_{\mathrm{ex}} = \frac{k_{\mathrm{off}}}{2K_{\mathrm{D}}}\left(K_{\mathrm{D}} + L_{\mathrm{T}} - P_{\mathrm{T}} + \sqrt{(P_{\mathrm{T}} + K_{\mathrm{D}})^2 + 2(K_{\mathrm{D}} - P_{\mathrm{T}})L_{\mathrm{T}} + L_{\mathrm{T}}^2}\right) \tag{6.15}$$

Thus, K_{D} ,L_{T}, and P_{T} are fixed at values determined from the experimental or theoretical design of the chosen chemical shift titration method, leaving k_{off} as an adjustable parameter. Finally, the chemical shift for the free state of the observed protein is fixed at the experimentally observed, or theoretically chosen value (Ω_{A}). Following optimization of M_0 and $R_{2\mathrm{A}}^0$ from the individual trace corresponding to the free state ($k = 1$), the objective function (Eq. 6.14) was subjected to nonlinear least squares regression in Mathematica 8.0.4 with the default parameters for the NMinimize and Differential Evolution options, with the fitted parameters k_{off}, Ω_{B}, and five intensity parameters (h_k) corresponding to the $k = 2, 3, 4, 5, 6$ (five total) different NMR traces collected in the presence of ligand. The first h_k intensity parameter for $k = 1$ is set to 1 and not optimized, as

described above.

In general, for 1:1 protein-ligand interactions in the fast exchange regime, the K_D value used in line shape analyses is determined separately through non-linear least squares fits of chemical shift changes during the course of a titration, to a 1:1 binding isotherm. Therefore, in order to assess the impact of potential systematic errors in fitted K_Ds on the accuracy of line shape analyses, we determined K_D values from fits of the chemical shift changes observed in the ^{1}H and ^{15}N GAMMA simulations. These theoretical titrations were conducted using Methods 1 and 2 with the protein and ligand concentrations in Tables 6.1 and 6.2, and a traditional titration, that is, Method 1 with $P_T = 0.5$ mM, k_{off} values of 100, 500, 1,000, 2,500, and 15,000 s^{-1} for ^{15}N and k_{off} values of 1,500, 2,500, and 15,000 s^{-1} for ^{1}H. The chemical shifts of the resonances observed in the GAMMA simulations of the various titrations were determined by fits to the natural NMR, or Lorentzian, line shape function:

$$I(\nu) = \frac{2M_0R_2}{\pi\left(4(\nu-\nu_0)^2 + R_2^2\right)} \tag{6.16}$$

where the fitted parameters are M_0, an arbitrary intensity parameter, R_2, the transverse relaxation rate, and ν_0, the chemical shift of the observed resonance. Chemical shifts for the various titrations with respective k_{off} values were fit to 1:1 binding isotherms to extract K_D values, as previously described (36). Errors in the fitted K_D values were determined using Monte Carlo methods by adding noise to the FIDs simulated with GAMMA, and choosing protein and ligand concentrations randomly from normal distributions as described in the following section.

Monte Carlo error estimation for NMR line shape analyses of simulated and experimental titrations

The errors associated with line shape analyses using the Bloch-McConnell equations for simulated and experimental ^{1}H-^{15}N HSQC NMR-monitored titrations were determined using Monte Carlo methods. For experimental titrations, and those simulated with the program GAMMA, we assumed that the two main sources of experimental error are the thermal electromotive force (EMF, or noise) in the probe coil, and the error in the

starting protein and ligand concentrations. To generate an ensemble of 100 Monte Carlo titration data sets, that is, 600 NMR slices with corresponding protein:ligand ratios used in experimental titrations or those for GAMMA NMR simulations of Methods 1 and 2 (Tables 6.1, 6.2), random noise was added to each point of an FID corresponding to a single NMR slice at a given protein:ligand ratio. For the experimental titrations, theoretical FIDs were calculated from the Bloch-McConnell equations, using the fitted parameters determined from line shape analysis, and noise was added to these back-calculated FIDs. The magnitude of the noise for a given FID was empirically adjusted such that the signal to noise ratio after Fourier transformation and post acquisition processing matched that observed experimentally for the NMR slice corresponding to the free state. In addition, the initial protein and ligand concentrations were chosen randomly from a normal distribution with a standard deviation of 5 %, and subsequent concentrations were calculated according to the respective dilution factor, as previously discussed (36). For titrations conducted through addition of a single stock solution of ligand (Method 1), by dilution (Method 2), or both, the maximum concentration error involves only the starting concentrations of protein and ligand, as subsequent errors due to pipetting and or dilution are small (36). The ensemble of 100 Monte Carlo titration data sets were fit to the Bloch-McConnell equations as described in the previous section, and statistics for the resulting χ^2, k_{off}, and bound chemical shift (Ω_{B}) ensembles were calculated.

Results and discussion

Accuracy of kinetic parameters from NMR-monitored chemical shift titrations

Careful experimental design of 2D ^{1}H-^{15}N NMR-monitored titrations for the study of protein-ligand interactions has the potential to produce a wealth of information regarding thermodynamic parameters such as dissociation constants (K_{D}) and concomitant populations for the free and bound states, obtainable through chemical shift analyses, as well as the kinetics of exchange between states, which can be determined through line shape

analyses.

We recently developed precise 2D ^{1}H-^{15}N NMR-monitored chemical shift titration methods for the determination of protein-ligand K_{D} values with kinetics in the fast exchange regime on the NMR timescale (36). The key to greater accuracy and precision with respect to the determination of K_{D} using these methods is the co-variation of the protein and ligand concentrations during the course of an NMR-monitored titration. One method entails the addition of aliquots of concentrated ligand solution to a concentrated protein solution, wherein the protein concentration decreases by a constant factor (Method 1), or alternatively, sequential dilution of a solution of concentrated protein and ligand in an initial 1:2 ratio (Method 2). This latter method yields fairly robust precision over a broad range of K_{D} values. It is of interest therefore, to determine if classical line shape analysis applied to these chemical shift titration methods can give accurate and precise off-rates for a given protein-ligand interaction.

The first issues to address are generalizations that line shape analyses are rarely practical for large proteins, and that limited digital resolution in heteronuclear 2D NMR spectroscopy may lead to inaccurate kinetics (28, 35). Using the GAMMA program (37), we performed quantum mechanical NMR simulations of the chemical shift titration experiments we developed, in order to assess the impact of digital resolution and post-acquisition processing on the accuracy of kinetic parameters derived from classical line shape analyses. The agreement between titrations simulated with the GAMMA program, and the associated fits to the Bloch-McConnell equations is excellent, as shown in fig. 6.1 and Table 6.3. For ^{15}N, with $k_{\mathrm{off}} = 2500\ \mathrm{s}^{-1}$ and $K_{\mathrm{D}} = 300\ \mu\mathrm{M}$, the absolute difference between the actual and fitted k off is 4.4 and 2.7 % for Methods 1 and 2, respectively, and the absolute difference between the actual and fitted $\Delta\omega$ is 1.3 % for both Methods 1 and 2 (fig. 6.1 and Table 6.3). For the ^{1}H dimension, with $k_{\mathrm{off}} = 2500\ \mathrm{s}^{-1}$ and $K_{\mathrm{D}} = 300\ \mu\mathrm{M}$, the absolute difference between the actual and fitted k_{off} is 0.4 and 0.6 % for Methods 1 and 2, respectively, and the absolute difference between the actual and fitted $\Delta\omega$ is 0.15 and 0.02 % for Methods 1 and 2, respectively (fig. 6.1 and Table 6.3).

The excellent accuracy of classical line shape analysis underscores a number of key points: First, the digital resolution of typical 2D ^{1}H-^{15}N HSQC NMR spectra (^{1}H: 8.2 Hz/point, ^{15}N: 12.6 Hz/point) does not appear to seriously impede the accuracy of line

Table 6.3: Line shape analysis for Methods 1 and 2 simulated with GAMMA

	$\chi^2 \times 10^8$	k_{off} (s^{-1})	ν_B (Hz)	$M_0 \times 10^{6a}$	R^0_{2A} (s^{-1})	h_2	h_3	h_4	h_5	h_6
M1;[b] ^{15}N	0.315	2390.27	7325.99	1.815	11.807	1.032	1.033	1.039	1.028	1.026
M2;[c] ^{15}N	3.053	2432.61	7325.91	1.815	11.809	1.024	1.024	1.025	1.026	1.028
M1, ^{1}H	0.069	2509.93	6095.33	2.954	30.135	0.517	0.425	0.387	0.330	0.147
M2, ^{1}H	0.031	2484.94	6095.43	2.949	29.327	0.721	0.689	0.668	0.633	0.598

a M_0 and R_{2A} were optimized separately from the line shape analysis using the NMR trace corresponding to the free state
b M1 indicates Method 1
c M2 indicates Method 2

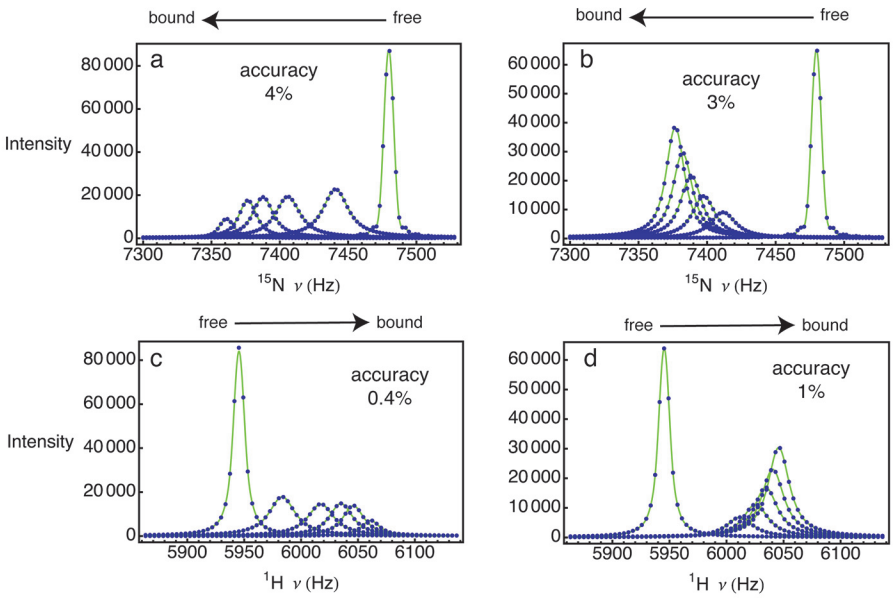

Figure 6.1: GAMMA simulations for (A) M49 ^{15}N Method 1 (blue circles), with corresponding fits to the Bloch-McConnell equations (green lines), (B) M49 ^{15}N for Method 2, (C) W33 $^{1}H_{\epsilon 1}$ for Method 1, and (D) ^{1}H for Method 2. The accuracy of the fitted k_{off} is given in the individual figure panels

shape analysis when typical post-acquisition processing techniques such as multiplication of FIDs by cosine apodization functions, zero-filling (^{1}H and ^{15}N dimensions), and linear prediction (^{15}N dimension) are employed. Secondly, in the absence of spin-spin couplings, the Bloch-McConnell equations, that is, the equations of motion for nuclear spins undergoing chemical exchange derived using classical physics, are essentially identical to those derived using a quantum mechanical approach, as expected (10). Third, optimization of individual intensity parameters for respective ^{1}H and ^{15}N traces taken from individual spectra for various titration points is an effective approach to account for differential relaxation and exchange losses during INEPT transfers and acquisition of two dimensions, as previously pointed out (22, 23). Finally, fitting the line shape of the free state to the Bloch-McConnell equations with $p_{\text{A}} = 1$ (Eq. 6.12) is a reasonable approach to estimate the intrinsic transverse relaxation rate $R^{0}_{2\text{A}}$. For example, using the three spins-1/2 system described in the Theory and Methods section, the calculated M49 ^{15}N dipole-dipole transverse relaxation rate from GAMMA is 11.4 s^{-1}, and the average fitted value is 11.8 s^{-1} for Methods 1 and 2 (Table 6.3). For the five spins-1/2 system

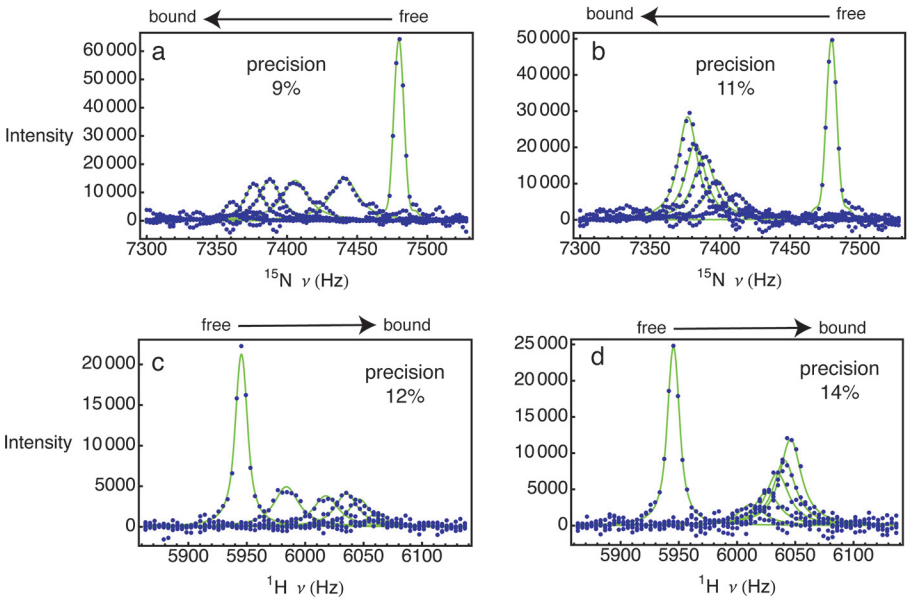

Figure 6.2: (A) One dimensional ^{15}N traces taken through the cross-peaks of M49 from 2D ^{1}H-^{15}N HSQC NMR spectra for Method 1 (blue circles), with corresponding fits to the Bloch-McConnell equations (green lines), and (B) ^{15}N traces taken through the cross-peaks of M49 for Method 2. (C) One dimensional $^{1}H_{\epsilon 1}$ traces taken through the cross-peaks of the W33 from 2D ^{1}H-^{15}N HSQC NMR spectra for Method 1 (blue), with corresponding fits to the Bloch-McConnell equations (green lines), and (D) $^{1}H_{\epsilon 1}$ traces for Method 2. The precision of the fitted k_{off} is given in the individual figure panels

(Theory and Methods), the calculated W33 $^{1}H_{\epsilon 1}$ dipole-dipole transverse relaxation rate from GAMMA is 25.4 s^{-1}, and the average fitted value is 29.7 s^{-1} for Methods 1 and 2 (Table 6.3).

Experimentally, the average k_{off} value determined for the Mms2-Ub interaction using classical line shape analysis for Methods 1 and 2 is 3,347 s^{-1} for M49 ^{15}N and W33 $^{1}H\epsilon 1$ (Table 6.4 and fig. 6.2). The standard deviation in the mean value is 304 s^{-1}, or 9 %. Thus, the excellent agreement for k_{off} between two separate chemical shift titrations (Methods 1 and 2), and different nuclei from two residues, is consistent with the precision of classical line shape analysis as implemented in Methods 1 and 2, as discussed in detail in the subsequent section.

Table 6.4: Experimental line shape analysis for Methods 1 and 2

	$\chi^2 \times 10^8$	k_{off} (s^{-1})	ν_B (Hz)	$M_0 \times 10^{6a}$	R^0_{2A} (s^{-1})	h_2	h_3	h_4	h_5	h_6
M1;[b] ^{15}N	170.684	3004.11	7333.60	1.724	20.515	0.843	0.666	0.652	0.607	0.527
M2;[c] ^{15}N	305.679	3633.51	7321.41	3.706	19.714	0.920	0.911	0.904	0.872	0.854
M1, ^{1}H	33.530	3568.74	6080.97	1.303	48.089	0.634	0.564	0.557	0.558	0.560
M2, ^{1}H	55.548	3179.61	6088.84	2.654	41.313	0.873	0.841	0.735	0.782	0.729

a M_0 and R_{2A} were optimized separately from the line shape analysis using the NMR trace corresponding to the free state
b M1 indicates Method 1
c M2 indicates Method 2

Precision of kinetic parameters from NMR-monitored chemical shift titrations

In the previous section, it was demonstrated that classical line shape analysis is highly accurate (< 5 %) for chemical shift titration methods we recently developed. To estimate the precision of kinetic parameters derived from classical line shape analyses in the presence of experimental noise and random error, we conducted Monte Carlo trials for GAMMA simulations of chemical shift titrations using the experimental designs of Methods 1 and 2. We assumed that the two main sources of experimental error during an NMR-monitored titration arise from thermal EMF in the NMR probe coil due to the Brownian motion of electrons, ultimately manifested as noise in the spectrum (40), and the error in the starting concentrations of protein and ligand (36). Representative spectra for the various chemical shift titration methods in the presence of simulated thermal EMF that is similar in magnitude to that observed experimentally, as well as concentration errors, are shown in fig. 6.3 and can be compared to the analogous spectra in the absence of noise (fig. 6.1).

The average error in k_{off} for residues M49 (^{15}N) and W33 (^{1}H) determined using Monte Carlo parameter estimation for GAMMA NMR simulations of Methods 1 and 2 is 12 % (Table 6.5). Representative Monte Carlo ensembles for line shape analyses for M49 ^{15}N, Methods 1 and 2 are shown in fig. 6.4. The average error in $\Delta\nu$, or $|\Omega_{\text{A}} - \Omega_{\text{B}}|/2\pi$, the maximum chemical shift change, for residues M49 (^{15}N) and W33 (^{1}H) determined using Monte Carlo parameter estimation for GAMMA NMR simulations of Methods 1 and 2 is 0.05 % (Table 6.5). The average parameter bias from Monte Carlo trials of the GAMMA simulations for Methods 1 and 2, that is, the difference between the mean of the ensemble of fitted parameters (k_{off} and ν_{B}) and their actual values is 8 % for k_{off}, and 0.4 % for ν_{B}. The parameter bias is within the error range from the Monte Carlo trials, indicating that the fitting procedure is of good quality, and free of systematic error.

The precision of GAMMA simulations in the presence of noise and concentration errors for Methods 1 and 2 is comparable to the experimentally determined precision. For example, the average error in k off for residues M49 (^{15}N) and W33 (^{1}H) determined using Monte Carlo parameter estimation for Methods 1 and 2 is 10 % (Table 6.6). The

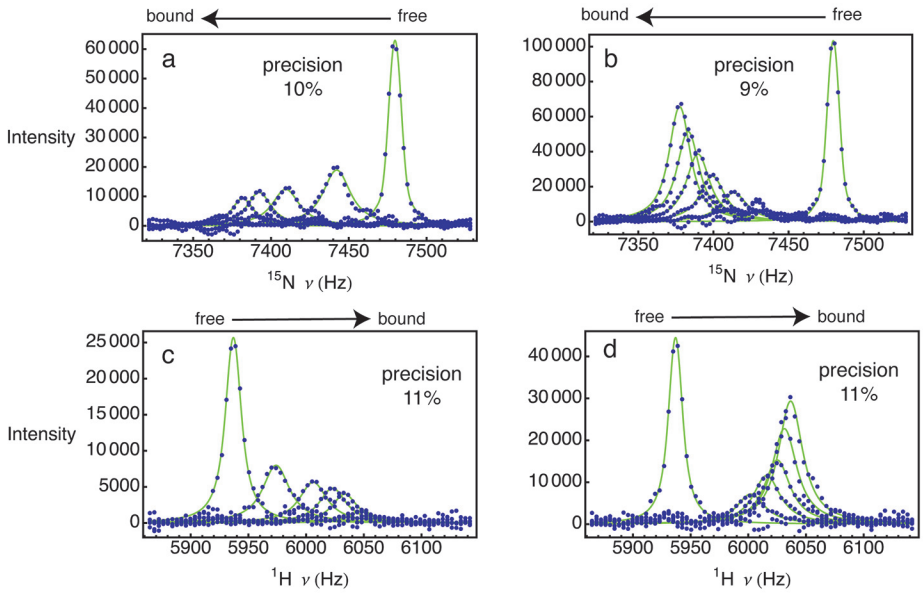

Figure 6.3: GAMMA simulations in the presence of noise and concentration error for (A) M49 ^{15}N Method 1 (blue circles), with corresponding fits to the Bloch-McConnell equations (green lines), (B) M49 ^{15}N for Method 2, (C) W33 $^{1}H_{\epsilon 1}$ for Method 1, and (D) W33 $^{1}H_{\epsilon 1}$ for Method 2. The precision of the fitted k_{off} is given in the individual figure panels

Table 6.5: Errors for line shape analysis for Methods 1 and 2 simulated with GAMMA

	χ^2	k_{off} (s^{-1})	ν_B (Hz)
M1;[a] ^{15}N	295 ± 43	2421 ± 223 (9 %)[b]	7326 ± 3 (0.04 %)[c]
M2;[d] ^{15}N	347 ± 38	2625 ± 286 (11 %)	7326 ± 4 (0.05 %)
M1, ^{1}H	164 ± 19	2780 ± 298 (12 %)	6096 ± 3 (0.05 %)
M2, ^{1}H	850 ± 69	2795 ± 347 (14 %)	6096 ± 4 (0.07 %)

a M1 indicates Method 1
b The standard deviation as a percentage of the actual k_{off} (2,500 s^{-1}) is given in parentheses
c The standard deviation as a percentage of the actual ν_B (7,326 or 6,096 Hz) is given in parentheses
d M2 indicates Method 2

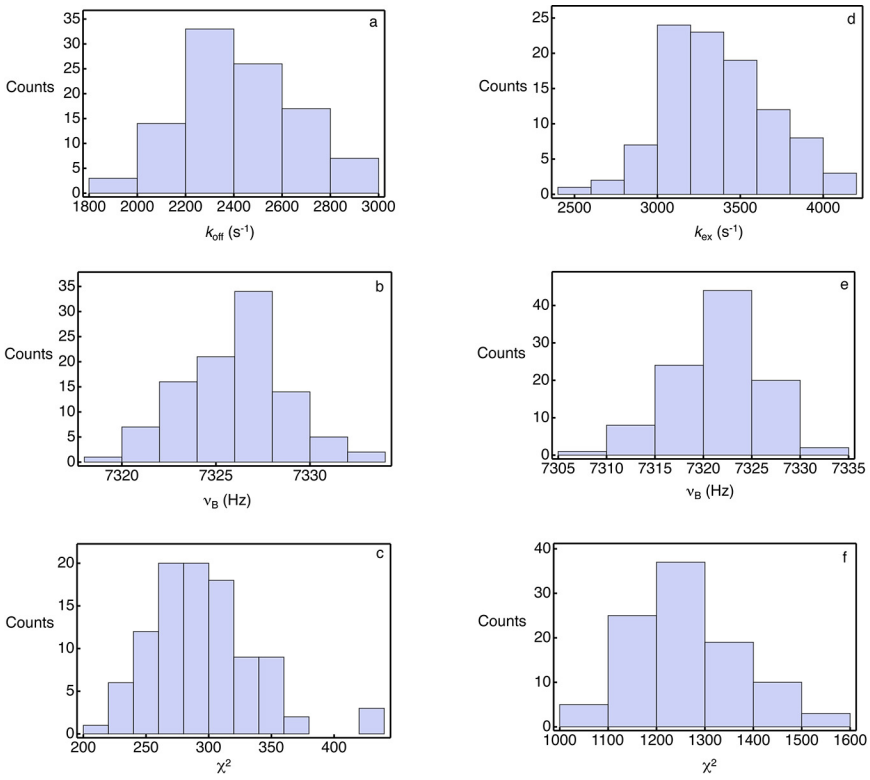

Figure 6.4: Monte Carlo parameter ensembles for line shape analyses of M49 ^{15}N GAMMA simulations, Method 1: (A) k_{off}, (B) the bound chemical shift ν_{B}, and (C) the magnitude of the objective function after optimization, χ^2. Monte Carlo parameter ensembles for line shape analyses of M49 ^{15}N GAMMA simulations, Method 2: (D) k_{off}, (E) the bound chemical shift ν_{B}, and (F) the magnitude of the objective function after optimization, χ^2

Table 6.6: Errors for line shape analysis for Methods 1 and 2

	χ^2	k_{off} (s^{-1})	ν_B (Hz)
M1;[a] ^{15}N	242 ± 36	2788 ± 293 (10 %)[b]	7333 ± 3 (0.04 %)[c]
M2;[d] ^{15}N	1266 ± 107	3364 ± 324 (9 %)	7321 ± 5 (0.07 %)
M1, ^{1}H	216 ± 22	3311 ± 409 (11 %)	6081 ± 3 (0.05 %)
M2, ^{1}H	480 ± 47	3078 ± 345 (11 %)	6089 ± 4 (0.07 %)

a M1 indicates Method 1
b The standard deviation as a percentage of the fitted k_{off} in Table 6.4 is given in parentheses
c The standard deviation as a percentage of the fitted ν_B in Table 6.4 is given in parentheses
d M2 indicates Method 2

average error in the experimentally determined $\Delta\nu$, the maximum chemical shift change, for residues M49 (^{15}N) and W33 (^{1}H) from Monte Carlo parameter estimation for Methods 1 and 2 is 0.06 % (Table 6.6). The parameter biases, the differences between the experimental means and the means of the Monte Carlo ensembles, for k_{off} and $\Delta\nu$ are 6 and 0.3 %, respectively, indicative of good quality fits lacking systematic errors. As previously noted, the mean of the experimental k_{off} has a value for $\pm 1\sigma$ of 9 %, indicating that our estimates for the magnitude of thermal EMF in the probe coil and the protein concentration errors are reasonable, and likely to properly account for the main sources of experimental error.

Theoretical range of validity for line shape analyses of NMR-monitored chemical shift titrations

In addition to GAMMA simulations conducted with a k_{off} value of 2,500 s^{-1}, we also conducted GAMMA simulations with k_{off} values of 500, 1,000 and 15,000 s^{-1} and $K_D = 60, 120$, and 1800 µM, respectively, for M49 ^{15}N using Methods 1 and 2, as well as traditional NMR titrations, that is, Method 1 with P_T held constant at 0.5 mM (figs. 6.5, 6.6, 6.7 and Table 6.7). These k_{off} values were chosen as they represent approximate, qualitative limits of applicability for line shape analysis for this specific protein-protein interaction. The lower k_{off} limit is determined by the amount of line broadening, and the resulting impact on signal to noise ratio (fig. 6.6B, D, E). The upper k_{off} limit was chosen as it represents the approximate K_D value (1.8 mM) beyond which the biological relevance of an interaction becomes questionable.

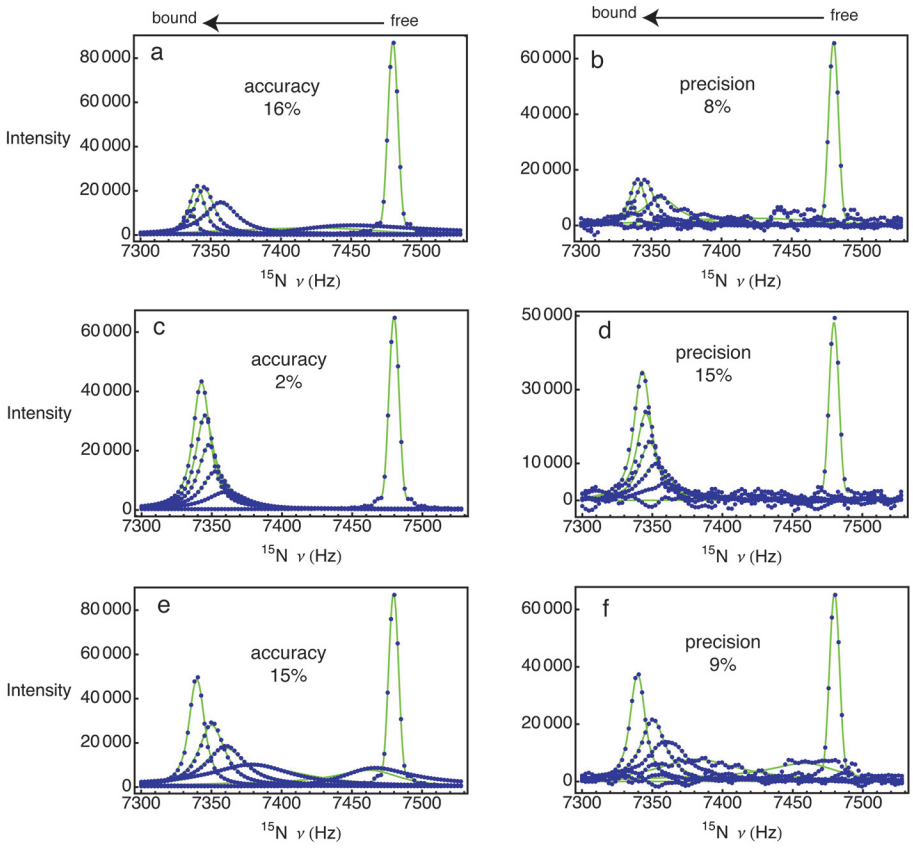

Figure 6.5: GAMMA simulations for K_D = 60 µM and k_{off} = 500 s^{-1} using M49 ^{15}N Method 1 (blue circles), with corresponding fits to the Bloch-McConnell equations (green lines) in absence (A) and the presence (B) of noise and concentration error. ^{15}N Method 2 in the absence (C) and presence (D) of noise and concentration error. ^{15}N Method 1 with P_T fixed at 0.5 mM (traditional titration) in the absence (E) and presence (F) of noise and concentration error. The accuracy or precision of the fitted k_{off} is given in the individual figure panels

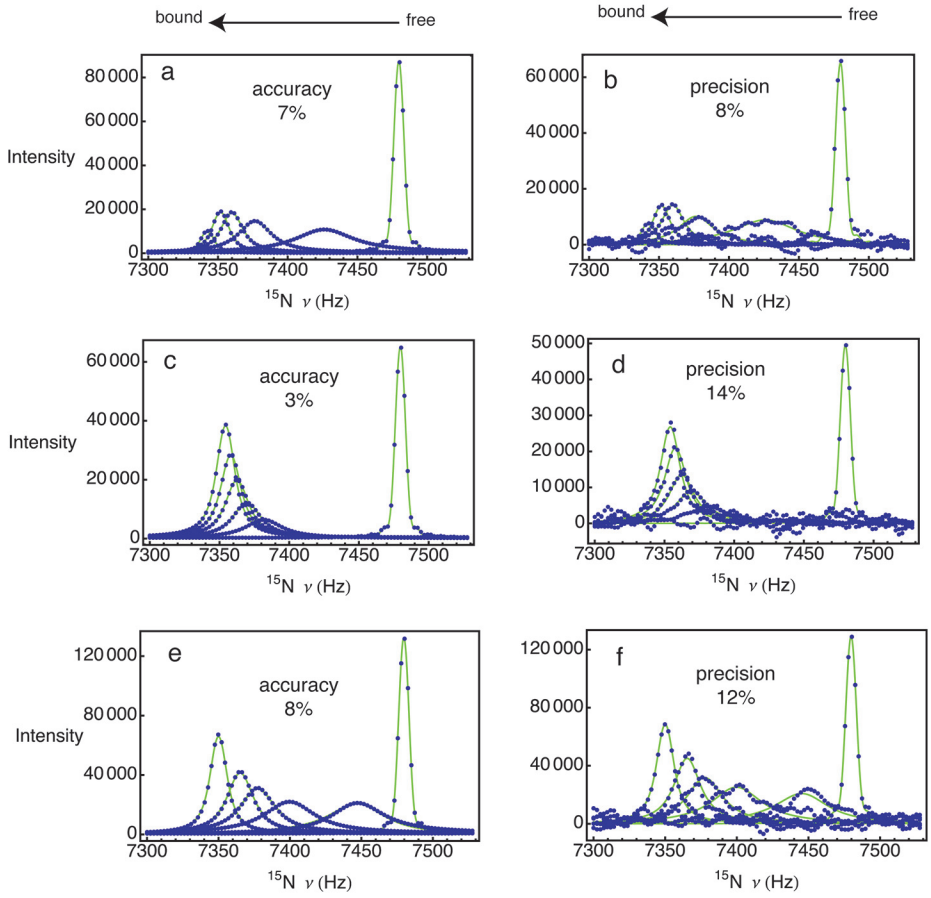

Figure 6.6: GAMMA simulations for $K_D = 120$ µM and $k_{off} = 1000$ s^{-1} using M49 ^{15}N Method 1 (blue circles), with corresponding fits to the Bloch-McConnell equations (green lines) in absence (A) and the presence (B) of noise and concentration error. ^{15}N Method 2 in the absence (C) and presence (D) of noise and concentration error. ^{15}N Method 1 with P_T fixed at 0.5 mM (traditional titration) in the absence (E) and presence (F) of noise and concentration error. The accuracy or precision of the fitted k_{off} is given in the individual figure panels

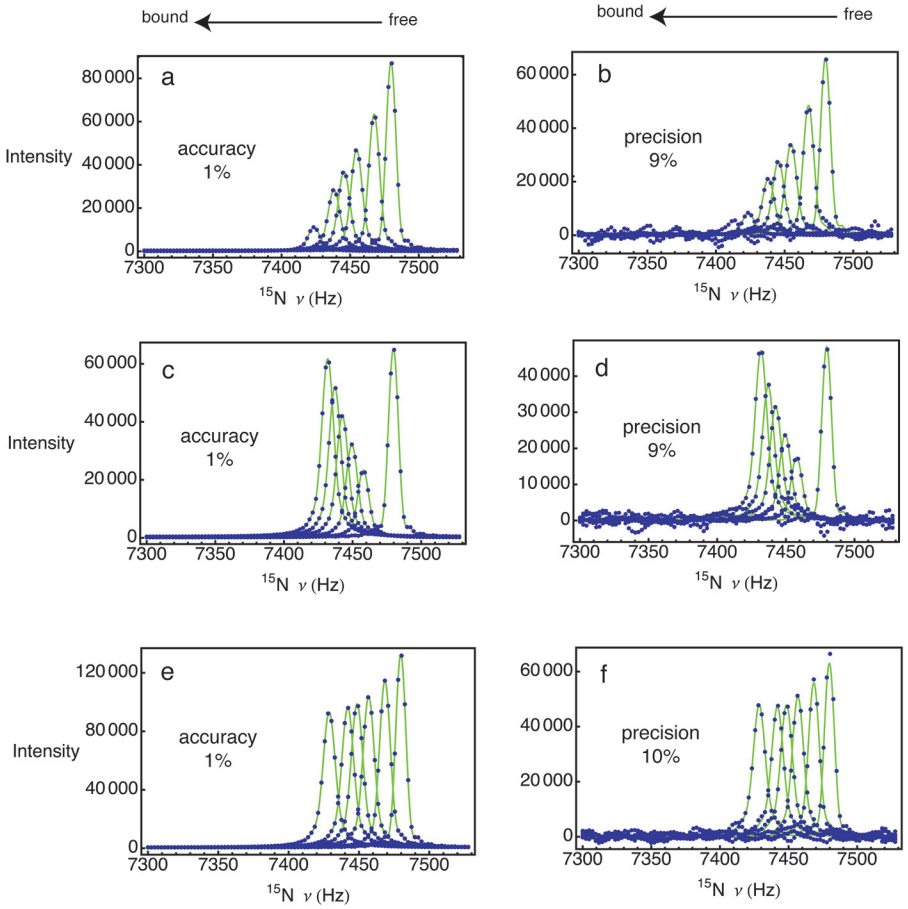

Figure 6.7: GAMMA simulations for $K_D = 1800$ µM and $k_{off} = 15000$ s^{-1} using M49 ^{15}N Method 1 (blue circles), with corresponding fits to the Bloch-McConnell equations (green lines) in absence (A) and the presence (B) of noise and concentration error. ^{15}N Method 2 in the absence (C) and presence (D) of noise and concentration error. ^{15}N Method 1 with P_T fixed at 0.5 mM (traditional titration) in the absence (E) and presence (F) of noise and concentration error. The accuracy or precision of the fitted k_{off} is given in the individual figure panels

Table 6.7: Theoretical accuracy and precision for ^{15}N line shape analysis for Methods 1 and 2 simulated with GAMMA

k_{off}	M1[a]	M2[b]	M1-trad[c]
100 s^{-1}	n.d.[d]	4 (±13)%[e]	n.d.
500 s^{-1}	16 (±8)%	2 (±15)%	15 (±9)%
1000 s^{-1}	7 (±8)%	3 (±14)%	8 (±12)%
15000 s^{-1}	1 (±9)%	1 (±9)%	1 (±10)%

a M1 indicates Method 1 with protein and ligand concentrations in Table 6.1
b M1 indicates Method 2 with protein and ligand concentrations in Table 6.2
c M1-trad indicates Method 1, with P_T fixed at 0.5 mM and ligand concentrations in Table 6.1
d Not determined
e The precision calculated from the Monte Carlo ensemble is given in parentheses

For M49 ^{15}N, Method 2 yields the most accurate values for $k_{off} = 500$ and 1,000 s^{-1}, with absolute differences of 3 and 2 %, respectively, whereas Method 1 is substantially less accurate at 16 and 7 % for $k_{off} = 500$ and 1,000 s^{-1}, respectively (figs. 6.5, 6.6). The traditional method for conducting a titration is also substantially less accurate than Method 2 at 15 and 8 % for $k_{off} = 500$ and 1,000 s^{-1}, respectively (figs. 6.5, 6.6). This decrease in accuracy for Method 1 and traditional titrations is due to the fact that Method 2 has fewer points within the region of protein:ligand ratios that have substantial line broadening (compare figs. 6.5C, 6.6C with figs. 6.5A, 6.6A and 6.5E, 6.6E). In this exchange regime, for the first titration point with $k_{off} = 500$ s^{-1} (fig. 6.5A, E), the free and bound resonances are differentially broadened, have different intensities, and are not averaged to a single resonance. This phenomenon is manifested as a single asymmetric peak in the spectrum. Furthermore, the agreement between spectra calculated using a quantum mechanical approach and the classical Bloch-McConnell equations is somewhat worse than at faster exchange rates with $k_{off} > 500$ s^{-1}, that is, the accuracy in k_{off} drops from 8 % at $k_{off} = 1000$ s^{-1} to 16 % at $k_{off} = 500$ s^{-1}; compare the first titration points in figs. 6.5A, 6.6A and 6.5E, 6.6E. This observation highlights the fact that the lower limit of exchange for which chemical shift titrations can be used to fit K_D values is determined by incomplete averaging of the free and bound peaks to a single resonance. In addition, given that Method 2 does not sample this region to the same extent as Method 1, or traditional titrations, Method 2 provides a means to extend the limit of applicability of chemical shift titrations and line shape analyses to intermediate exchange

Table 6.8: Theoretical accuracy and precision for ^{1}H line shape analysis for Methods 1 and 2 simulated with GAMMA

k_{off}	M1[a]	M2[b]	M1-trad[c]
1500 s^{-1}	1 (±10) %[d]	1 (±17) %[e]	0 (±13) %
15000 s^{-1}	16 (±10) %	1 (±17) %	14 (±13) %

a M1 indicates Method 1 with protein and ligand concentrations in Table 6.1
b M1 indicates Method 2 with protein and ligand concentrations in Table 6.2
c M1-trad indicates Method 1, with P_T fixed at 0.5 mM and ligand concentrations in Table 6.1
d Not determined
e The precision calculated from the Monte Carlo ensemble is given in parentheses

rates; this is discussed in more detail below, and in subsequent sections.

In addition to analyzing accuracy, the precision for k_{off} determined from the various ^{15}N line shape methods was determined from statistical analyses of the ensembles of 100 Monte Carlo distributions. For Methods 1 and 2, with k_{off} values of 500 and 1,000 s^{-1}, the precision is reasonable (8-15 %) given a typical signal to noise ratio for the free peak of ∼50:1, and an error in protein concentration of 5 % (figs. 6.5B, D, F and 6.6B, D, F). For M49 ^{15}N within the very fast exchange regime, k_{off} = 15000 s^{-1} (fig. 6.7), line shape analysis for Methods 1 and 2, as well as a traditional titration with P_T = 0.5 mM, produces accurate values of k_{off}, within 1 % (fig. 6.7A, C, E). In addition, Monte Carlo simulations in the presence of noise and concentration error indicate that the precision is good, 9-10 % for all three methods.

We conducted GAMMA simulations with k_{off} values of 1,500 and 15,000 s^{-1} and K_D = 180 and 1,800 µM, respectively, for W33 ^{1}Hϵ1 using Methods 1 and 2, as well as Method 1 with P_T held at 0.5 mM (figs. 6.8, 6.9, and Table 6.8). For the fast exchange regime k_{off} = 1500 s^{-1}, the accuracy is better than 1 % for all of the titrations (fig. 6.8A, C, E), with the precision ranging from 10 to 17 % (fig. 6.8B, D, F). Within the very fast exchange regime, k_{off} = 15000 s^{-1}, Method 2 reproduces k_{off} with the greatest accuracy (0.5 %, fig. 6.9C), whereas Method 1, as well as a traditional titration with P_T = 0.5 mM, are substantially less accurate at 16 and 14 % (Fig. 9a, e), respectively. Regardless of the differences in accuracy, the precision for the various methods in the very fast exchange regime ranges from 10 to 17 % (fig. 6.9B, D, F).

The theoretical analysis of the accuracy and precision of line shape analysis indicates

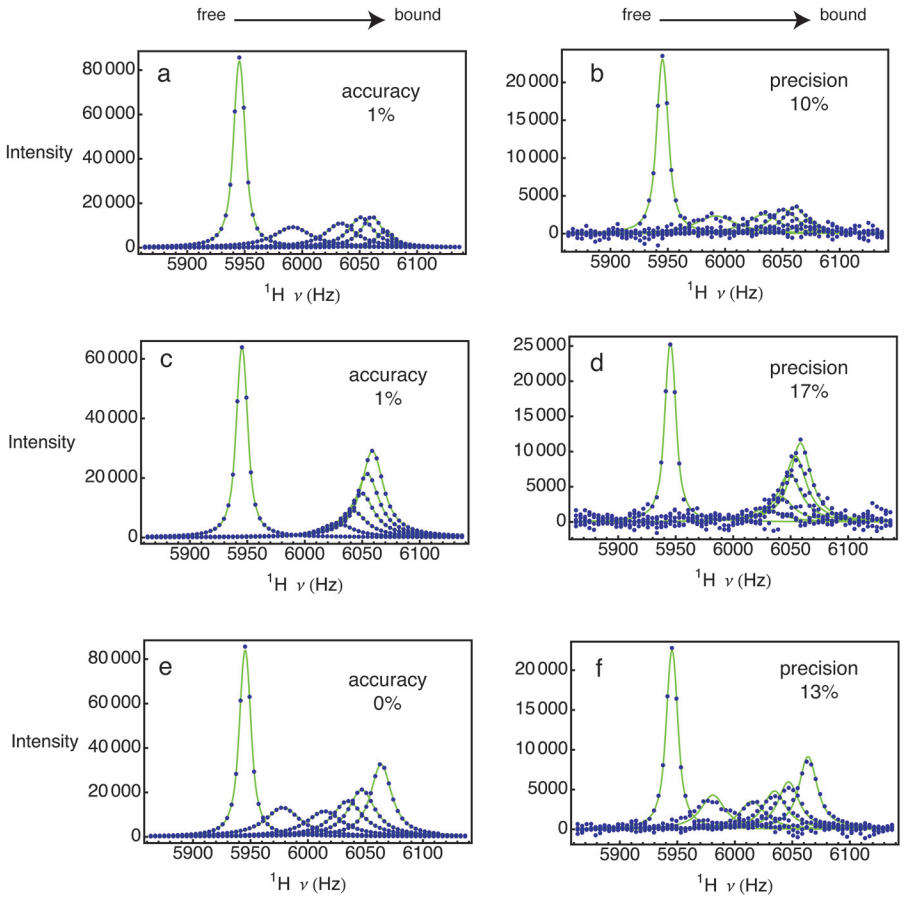

Figure 6.8: GAMMA simulations for $K_D = 180$ µM and $k_{off} = 1500$ s^{-1} using ^{1}H Method 1 (blue circles), with corresponding fits to the Bloch-McConnell equations (green lines) in absence (A) and the presence (B) of noise and concentration error. ^{1}H Method 2 in the absence (C) and presence (D) of noise and concentration error. ^{1}H Method 1 with P_T fixed at 0.5 mM (traditional titration) in the absence (E) and presence (F) of noise and concentration error. The accuracy or precision of the fitted k_{off} is given in the individual figure panels

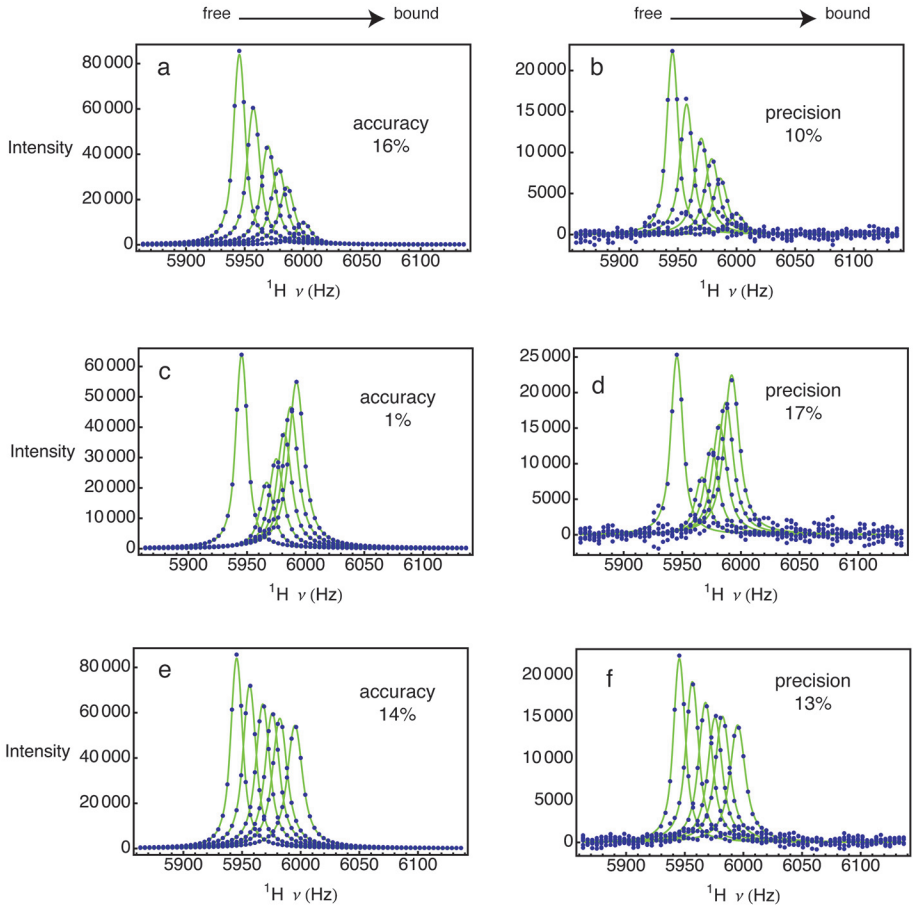

Figure 6.9: GAMMA simulations for K_D = 1800 µM and k_{off} = 15000 s^{-1} using ^{1}H Method 1 (blue circles), with corresponding fits to the Bloch-McConnell equations (green lines) in absence (A) and the presence (B) of noise and concentration error. ^{1}H Method 2 in the absence (C) and presence (D) of noise and concentration error. ^{1}H Method 1 with P_T fixed at 0.5 mM (traditional titration) in the absence (E) and presence (F) of noise and concentration error. The accuracy or precision of the fitted k_{off} is given in the individual figure panels

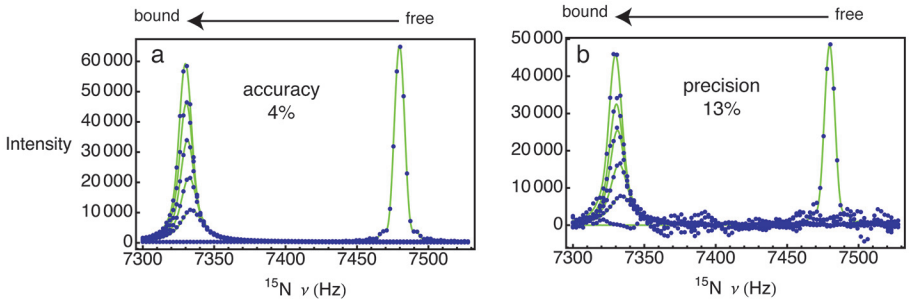

Figure 6.10: GAMMA simulations for $K_D = 12$ µM and $k_{off} = 100$ s^{-1} using ^{15}N Method 2 (blue circles), with corresponding fits to the Bloch-McConnell equations (green lines) in absence (A) and the presence (B) of noise and concentration error. The accuracy or precision of the fitted k_{off} is given in the individual figure panels

that for a protein-protein interaction such as that between Mms2 (145 residues) and ubiquitin (76 residues), with the specific acquisition parameters used in this study (Materials and Methods), line shape analysis using the ^{15}N dimension from 2D ^{1}H-^{15}N HSQC NMR spectra is accurate to better than 8 % and the precision better than 14 % for k_{off} ranging from 1,000 s^{-1} to 15,000 s^{-1} (Table 6.7). To assess an approximate lower limit for k_{off}, it is reasonable to assume that Method 2 is most likely to produce quantitative kinetics for intermediate exchange rates ($k_{off} < 1000$ s^{-1}) where line broadening is extensive, as the experimental design avoids protein:ligand ratios that approach this regime. Therefore, we conducted simulations for ^{15}N with $k_{off} = 100$ s^{-1} (fig. 6.10). From these results, it is evident that Method 2 in the ^{15}N dimension provides good accuracy and precision for k_{off}, better than 4 %, for values of 100 and 500 s^{-1} (Table 6.7 and fig. 6.10). If the K_D value for a protein-protein or protein-ligand interaction is below ∼1 µM, isothermal titration calorimetry can be used to measure K_D, and therefore, can potentially extend the applicability of classical line shape analysis into the slow exchange limit (41, 42). For example, from GAMMA simulations of a traditional titration, or Method 1 with P_T fixed at 0.5 mM, and ligand concentrations in Table 6.1, for ^{15}N with $k_{off} = 5$ s^{-1}, and $K_D = 0.6$ µM, the accuracy for k_{off} is 5 %, however, the precision is poor at 30 % (fig. 6.11).

Line shape analyses using the $^{1}H^{N}$ dimension yields accurate kinetics above ∼ 1,500 s^{-1}; the agreement with the actual k_{off} is better than 1 %, with reasonable precision ranging from 10 to 17 %. Below this k_{off}, line broadening for protein:ligand ratios that

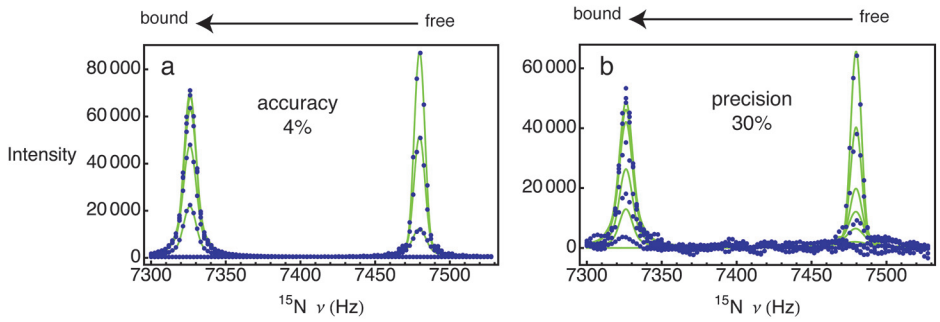

Figure 6.11: GAMMA simulations for $K_D = 0.6$ µM and $k_{off} = 5$ s^{-1} using a traditional titration, that is, Method 1 with $[P_T]$ fixed at 0.5 mM (blue circles), with corresponding fits to the Bloch-McConnell equations (green lines) in absence (A) and the presence (B) of noise and concentration error. The accuracy or precision of the fitted k_{off} is given in the individual figure panels

are ∼1:1 renders line shape analysis using data from the $^1H^N$ dimension difficult. At the biologically relevant upper limit, $k_{off} = 15000$ s^{-1}, the accuracy is impaired (14-16 %) in comparison to the ^{15}N dimension for Method 1, as well as the traditional method of conducting a titration. However, the accuracy of 0.5 % for Method 2 at $k_{off} = 15000$ s^{-1} remains comparable to that obtained in the ^{15}N dimension. Thus, for the protein-protein system and experimental parameters employed in this study, line shape analyses using ^{1}H-^{15}N HSQC 2D NMR monitored titrations have the potential to provide reasonably accurate and precise k_{off} values over a wide range, from 100 to 15,000 s^{-1}. Ultimately, the accuracy of k_{off} is subject to systematic errors arising from $^3J_{HNH\alpha}$ couplings, as well as the accuracy of K_D and R^0_{2A} employed in line shape analyses, as discussed in detail in subsequent sections.

Systematic errors in $^1H^N$ line shape analyses of NMR-monitored chemical shift titrations due to $^3J_{HNH\alpha}$ couplings

For typical 2D ^{1}H-^{15}N HSQC 2D NMR spectra of small to moderately sized proteins (10-30 kDa), per residue cross peaks from the $^1H^N$ dimension display partially resolved or unresolved coupling to the respective $^1H_\alpha$ of a given residue ($^3J_{HNH\alpha}$). The magnitude of the coupling is dependent on the main chain ϕ dihedral angle, according to the Karplus equation, and ranges from ∼4 Hz for α-helical secondary structure, ∼7 Hz for

averaged random coil configurations, and 10 Hz for β-sheet conformations (43). If the phenomenological Bloch equations are employed to analyze kinetic processes using NMR line shape analysis, unresolved $^3J_{\mathrm{HNH}\alpha}$ couplings typically cannot be ignored, as the resulting kinetics may be inaccurate. For example, fig. 6.12 shows the accuracy of $^1\mathrm{H}^\mathrm{N}$ line shape analysis in the presence of unresolved or partially resolved $^3J_{\mathrm{HNH}\alpha}$ couplings. For GAMMA simulations of Method 1 (Table 6.1, $k_{\mathrm{off}} = 2500\ \mathrm{s}^{-1}$) with an unresolved $^3J_{\mathrm{HNH}\alpha}$ of 1.8 Hz for a $^1\mathrm{H}^\mathrm{N}$ resonance peak with an intrinsic R_2^0 of 25 s^{-1}, an acquisition time of 122 ms, a digital resolution of 8.2 Hz/point, and post-acquisition processing of the FID by a cosine window function and zero-filling to 2,048 points, the k_{off} is reproduced with an accuracy of 0.4 % (fig. 6.12A). This indicates that couplings below 2 Hz can be ignored in line shape analyses when employing the Bloch equations at the spectral resolution employed in this work. It should be noted that the $^3J_{\mathrm{H}\epsilon1\mathrm{H}\delta1}$ coupling is 1.8 Hz for the tryptophan side chain, thus, for line shape analyses of the experimental data for W33 $^1\mathrm{H}_{\epsilon1}$ using the Bloch equations (fig. 6.3C), this coupling can be ignored (Table 6.4). The accuracy for an actual value for k_{off} of 2,500 s^{-1} deteriorates to 16 % for the fitted value of 2,887 s^{-1}, in the presence of an unresolved coupling of 5 Hz (fig. 6.12B). For a larger coupling of 10 Hz, employing the Bloch-McConnell equations gives rise to a 72 % error for the fitted value of 4,292 s^{-1} (fig. 6.12D).

The above analysis indicates that if the Bloch equations are to be employed for analysis of $^1\mathrm{H}^\mathrm{N}$ spectral data in general, $^3J_{\mathrm{HH}}$ couplings must be taken into account. There are potentially four different approaches to dealing with $^3J_{\mathrm{HH}}$ couplings in the proton dimension of ^{1}H-^{15}N HSQC 2D NMR spectra. Homonuclear semi-selective shaped pulse decoupling during acquisition can be employed to decouple $^1\mathrm{H}^\mathrm{N}$ and $^1\mathrm{H}_\alpha$, but suffers from substantial signal to noise losses due to intermittent receiver gating, and therefore, may be difficult to apply to protein systems undergoing chemical exchange (44). Second generation pure shift homonuclear decoupling methods (45) offer another option, and have been extended to 2D HSQC NMR experiments (46). However, the pure shift HSQC experiment provides less than half the signal to noise ratio of a regular HSQC. Uniform deuteration of the observed protein component (47, 48) is also a possible approach to minimize the impact of proton couplings on classical line shape analysis, as proton-deuteron coupling constants are reduced by a factor of $\gamma_{\mathrm{H}}/\gamma_{\mathrm{D}} \sim 6.5$ in compari-

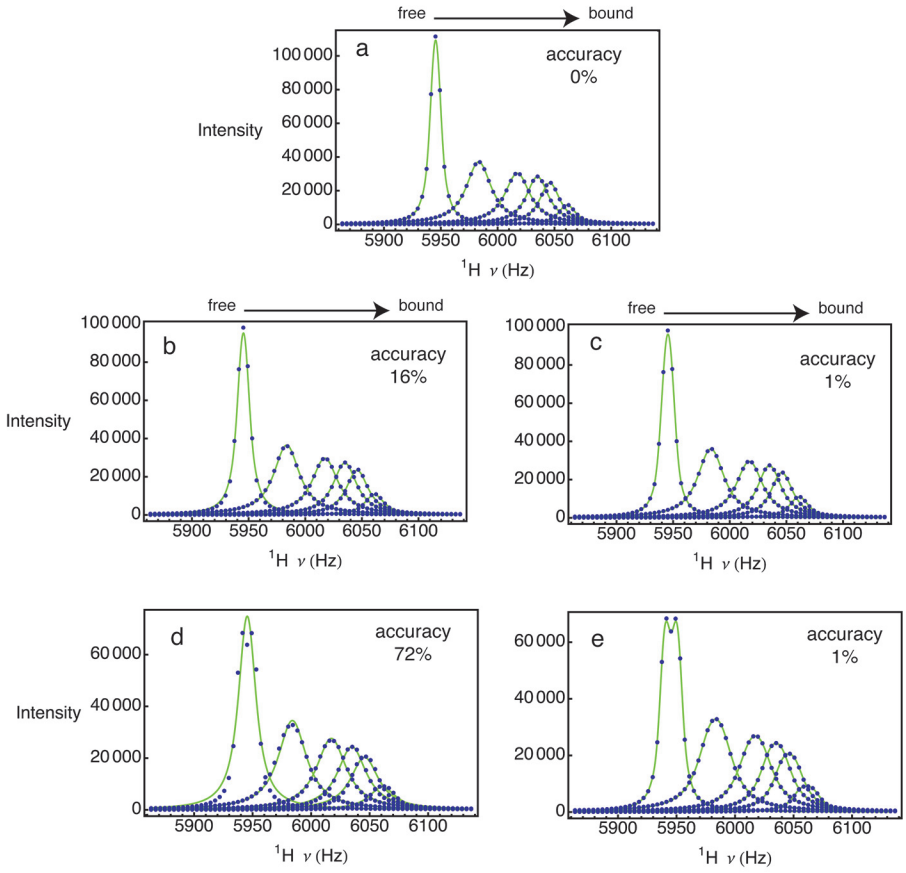

Figure 6.12: GAMMA simulations for K_D = 300 µM and k_{off} = 2500 s^{-1} using ^{1}H Method 1 (blue circles), with corresponding fits to the Bloch-McConnell equations (green lines) in presence of an unresolved 1.8 Hz coupling constant (A), in the presence of an unresolved 5.0 Hz coupling constant (B) in the presence of an unresolved 5.0 Hz coupling constant fit to the Bloch-McConnell equations modified to account for weak coupling according to eq. 6.13 (C), in the presence of a partially resolved and unresolved 10 Hz coupling constant (D) and in the presence of a partially resolved and unresolved 10 Hz coupling constant fit to the Bloch-McConnell equations modified to account for weak coupling according to eq. 6.13 (E). The accuracy of the fitted k_{off} is given in the individual figure panels

son to proton-proton couplings, however, the $^1H^N$ resonance is split into a 1:1:1 triplet. Additionally, interference between dipolar and quadrupolar relaxation mechanisms for proton-deuteron pairs may give rise to asymmetric triplet patterns and line broadening, although these effects are expected to be small for proton-deuteron pairs that are separated by > 2 Å (49). Finally, the doublet splitting arising from $^3J_{HH}$ coupling can be treated as the superposition of two uncoupled NMR resonances to a first approximation (15), but this approach requires measurement of the $^3J_{HH}$ coupling. However, accurate kinetics can be achieved without having to resort to protein perdeuteration, decoupling techniques, or density matrix formulations. For example, using GAMMA simulations for Method 1 (Table 6.1, $k_{off} = 2500$ s^{-1}) with an unresolved $^3J_{HNH\alpha}$ of 5 Hz, the accuracy can be improved from 16 to 1 % (figs. 6.12B and C). For a partially resolved and unresolved 10 Hz coupling, the accuracy is improved from 72 to 1 % (figs. 6.12D and E).

Systematic errors in line shape analyses of NMR-monitored chemical shift titrations

Systematic error adversely impacts accuracy, and is insidious for highly precise line shape parameters, as the precision may be mistaken for high accuracy. However, as we have shown, a thorough Monte Carlo error analysis, including checks on parameter bias, may be helpful to identify potential systematic errors, which may arise from the values of R^0_{2A}, R^0_{2B}, and K_D that are used in line shape analyses. For example, a ±10 % error on the value of R^0_{2A} used in the line shape analysis of the GAMMA simulation for W33 $^1H_{\epsilon 1}$ using Method 1 (Table 6.3), directly translates into a systematic error of ±10 % on the value of the fitted k_{off}. Similarly, ±10 % error on the value of K_D gives a systematic error of ±11 % on the fitted value of k_{off}. Thus, knowledge of the errors on the values of K_D and R^0_{2A} establishes limits on the accuracy and precision of the k_{off}. For our implementation of line shape analysis, Monte Carlo simulations indicate that concentration error and noise in the FID give rise to a ±5 % error on the magnitude of R^0_{2A} when it is determined through fitting the free resonance peak to the Bloch-McConnell equations for $p_A = 1$. Thus, potential systematic errors for R^0_{2A} are about the same as the accuracy for k_{off}, and only half the precision of k_{off}, indicating that the fitted k_{off} will be accurate with

respect to systematic errors in R^0_{2A}.

In general, for weak protein-ligand interactions in the fast exchange regime, dissociation constants determined from fits of chemical shift changes to 1:1 binding isotherms are subsequently used in line shape analyses according to the relationship $k_{on} = k_{off}/K_D$, eliminating the need to fit one of the interaction rate constants. As previously discussed, we recently developed methods to determine precise K_D values from chemical shift titrations (36). For these experimental titrations, the standard deviations for the average per residue K_D value of Methods 1 and 2 are 6 and 11 %, respectively (Table 6.5, (36)). The precision in the K_D value for Method 2 is about the same as the precision in the k_{off} value. Thus, in a worst-case scenario, the fitted k_{off} value would have a systematic error close to the value of $\pm 1\sigma$ of the precision. However, the accuracy remains acceptable, as it would fall within one standard deviation of the experimental precision.

Theoretical accuracy and precision of K_D values from NMR-monitored chemical shift titrations

In addition to conducting line shape analyses, the ^{1}H and ^{15}N spectra (figs. 6.5-6.10) from the GAMMA simulations we designed with different k_{off} values and protein and ligand concentrations given in Tables 6.1 and 6.2, were also used to assess the theoretical range for the accuracy and precision in K_D values determined from chemical shift titrations. We fit ^{1}H and ^{15}N line shapes calculated using GAMMA simulations for Methods 1 and 2, as well as a traditional titration, or Method 1 with P_T fixed at 0.5 mM, and various k_{off} values (Tables 6.7, 6.8) to the Lorentzian line shape (Eq. 6.16) to determine chemical shifts, which were subsequently fit to 1:1 binding isotherms as previously described (36) to extract K_D values (Tables 6.9, 6.10). For $k_{on} = 8.33 \times 10^6$ M^{-1} s^{-1}, and k_{off} values of 100 and 500 s^{-1}, the magnitude of k_{ex} (k_{on}[L] + k_{off}) at the various protein:ligand concentration ratios, ranges from 1,033 s^{-1} to 8,861 s^{-1}. For a chemical shift difference between the free and bound states of 150 Hz, the angular frequency, $\Delta\omega$, is 942 rad s^{-1}. Thus, k_{ex} ranges from the intermediate ($k_{ex} \sim |\Delta\omega|$) to fast ($k_{ex} > |\Delta\omega|$) exchange regimes. As shown in Table 6.9, for ^{15}N, the accuracy in the K_D for all titration methods at k_{off} values of 100-500 s^{-1} is generally poor, no better than 25 %. Therefore, whilst the

accuracy and precision of line shape analyses conducted at these k_{off} values is generally good (Table 6.7), large systematic errors in the fitted K_D values limit the accuracy of line shape analysis.

One solution to this problem is to employ isothermal titration calorimetry to determine K_D values as the intermediate exchange regime is approached for a given protein-ligand interaction (50). Alternatively, the experimental design of chemical shift titrations using Method 2 avoids protein:ligand ratios wherein extensive line broadening is observed. Furthermore, for a given K_D, smaller $\Delta\omega$ values translate into narrower resonances during the titration, raising the possibility that the combination of a chemical shift titration conducted according to Method 2, with an analysis of K_D values determined from residues with different $\Delta\omega$ values can give a more accurate estimate of K_D. Therefore, we conducted simulations of M49 ^{15}N chemical shift titrations with the Bloch-McConnell equations for Methods 1 and 2, and a traditional titration, that is, Method 1 with $P_T = 0.5$ mM, using a k_{off} value of 500 s^{-1}. To test the lower limit of applicability, a titration with a slower k_{off} value of 100 s^{-1} using Method 2 was also simulated as this method avoids much of the broadening that would be observed at protein:ligand ratios used in Method 1 or a traditional titration. The protein and ligand concentrations used with $k_{off} = 100$ and 500 s^{-1} are given in Tables 6.1 and 6.2. The bound chemical shift was varied to allow a range of $\Delta\omega$ to be assessed (188.5, 471.2, 754.0, 1,036.7, 1,319.5, 1,602.2, 1,885.0 rad s^{-1}). Chemical shifts at the various protein:ligand ratios were obtained from the simulated titrations by fitting the spectral data to the Lorentzian line shape function. The mean value and the associated error for the optimized K_D at each $\Delta\omega$ were obtained by generating an ensemble of 100 spectra per titration point through the addition of random noise to each point of FIDs calculated using the Bloch-McConnell equations. Each FID in the ensemble was Fourier transformed, and the resulting spectrum was fit to the Lorentzian line shape in order to obtain the chemical shift for the various titration points. These chemical shifts were subsequently fit to 1:1 binding isotherms (Eq. 6, (36)), to optimize the values of the adjustable parameters K_D and $\Delta\omega$. The results show that the accuracy of the fitted K_D improves with decreasing $\Delta\omega$ for a variety of chemical shift titration methods (fig. 6.13). Importantly, extrapolation of K_D as a function of the free and bound chemical shift difference to $\Delta\omega = 0$, allows for determination of a more

Table 6.9: Theoretical accuracy and precision for K_D values from ^{15}N chemical shift titrations simulated with GAMMA

k_{off}	Actual K_D (µM)	Fitted K_D M1[a](µM)	Fitted K_D M2[b](µM)	Fitted K_D M1-trad[c](µM)
100 s^{-1}	12	n.d.[d]	5 ± 1	n.d.
			56(±8)%[e]	
500 s^{-1}	60	200 ± 52	45 ± 5	34 ± 41
		233(±87)%	25(±8)%	43(±68)%
1000 s^{-1}	120	135 ± 23	113 ± 9	142 ± 56
		13(±9)%	6(±8)%	18(±47)%
15000 s^{-1}	1800	1840 ± 150	1820 ± 90	1840 ± 160
		2(±8)%	1(±5)%	2(±9)%

a M1 indicates Method 1 with protein and ligand concentrations in Table 6.1
b M1 indicates Method 2 with protein and ligand concentrations in Table 6.2
c M1-trad indicates Method 1, with P_T fixed at 0.5 mM and ligand concentrations in Table 6.1
d Not determined
e Accuracy and precision are given as percentages of the absolute difference between the actual and fitted K_Ds, precision is given in parentheses

Table 6.10: Theoretical accuracy and precision for K_D values from ^{1}N chemical shift titrations simulated with GAMMA

k_{off}	Actual K_D (µM)	Fitted K_D M1[a](µM)	Fitted K_D M2[b](µM)	Fitted K_D M1-trad[c](µM)
1500 s^{-1}	180	195 ± 28	182 ± 14	217 ± 61
		8(±16)%[d]	1(±8)%	21(±34)%
15000 s^{-1}	1800	1830 ± 170	1830 ± 110	1860 ± 180
		2(±9)%	2(±6)%	3(±10)%

a M1 indicates Method 1 with protein and ligand concentrations in Table 6.1
b M1 indicates Method 2 with protein and ligand concentrations in Table 6.2
c M1-trad indicates Method 1, with P_T fixed at 0.5 mM and ligand concentrations in Table 6.1
d Not determined

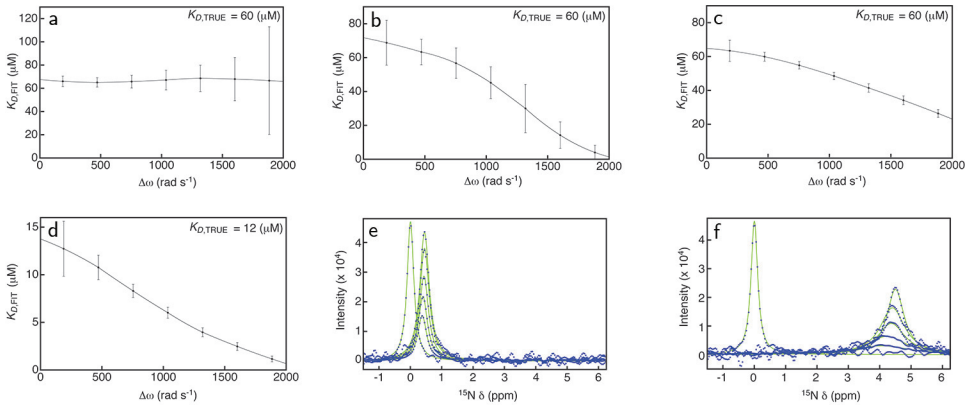

Figure 6.13: Accuracy and precision of K_D values determined from Monte Carlo simulations of ^{15}N chemical shift titrations for (A) Method 1, (B) a traditional titration (Method 1 with P_T = 0.5 mM), and (C, D) Method 2. In A-D, the lines through the points are interpolated with a 2nd order polynomial. Representative spectra from simulated titrations (blue circles) and fits to the Lorentizan line shape (green line) are shown for $\Delta\omega$ values of (E) 188.5 and (F) 1885.0 rad s^{-1}

accurate value for K_D if the titration data include some spectra within the intermediate exchange regime. This regime represents the lower limit of exchange for which chemical shift titrations can be used to fit K_D values, as a result of broadening and incomplete averaging of the free and bound peaks to a single resonance.

In general, fig. 6.13 shows that for chemical shift titrations in which some titration points are in the intermediate exchange regime, the accuracy of the fitted K_D value deteriorates with increasing chemical shift difference between the free and bound states. This is mainly a result of increased line broadening due to increasing $\Delta\omega$ (fig. 6.13B, C, D and a representative spectrum in fig. 6.13F). As $\Delta\omega$ decreases, the line broadening decreases, and the accuracy of K_D increases, as chemical shifts can be more accurately determined for narrower resonances (compare fig. 6.13E, F). Importantly, as shown in fig. 6.13E, for small $\Delta\omega$ values, even in the presence of FID noise, the corresponding small chemical shift changes during a titration can be determined accurately using Method 2. Interpolating K_D values to $\Delta\omega = 0$ is feasible only when resonances from multiple protein residues are observed to shift to varying degrees in the spectra; however, this is often the case for protein:protein or protein:ligand interactions monitored by ^{1}H-^{15}N HSQC 2D NMR spectra, using uniformly ^{15}N-labelled protein as the observed component.

Within the lower range of the fast exchange regime, or $k_{ex} > |\Delta\omega|$, $k_{on} = 8.33 \times 10^6$

M^{-1} s^{-1} and k_{off} values of 1,000 and 1,500 s^{-1}, the accuracy and precision in the fitted K_D values is good for ^1H and ^{15}N using Methods 1 and 2 (19 % or better, Tables 6.9, 6.10). However, for the traditional method of conducting a titration (P_T fixed at 0.5 mM), the precision is poor, not better than 34 %. Thus, within the lower end of the fast exchange regime, the accuracy and precision of Methods 1 and 2 facilitate the determination of both the thermodynamics and kinetics of protein:ligand interactions.

The very fast exchange regime, or $k_{ex} \gg |\Delta\omega|$, with $k_{on} = 8.33 \times 10^6$ M^{-1} s^{-1} and $k_{off} = 15000$ s^{-1}, corresponds to k_{ex} values ranging from $\sim$ 16,000 to 24,000 s^{-1}, depending on the protein:ligand ratio. In this exchange regime, GAMMA simulations indicate that the accuracy in the K_D value determined from the various chemical shift titrations for ^1H and ^{15}N is excellent, 3 % or better, and the precision is 10 % or better, for all titration methods (Tables 6.9, 6.10). These results establish that protein-ligand titrations monitored by ^1H-^{15}N HSQC 2D NMR spectroscopy represent a powerful approach for accurate and precise determination of the thermodynamics, that is K_D values, or populations, as well as kinetics of 1:1 protein-ligand interactions within the fast exchange regime.

Conclusions

We recently developed two chemical shift titration methods for the measurement of precise K_D values for 1:1 protein ligand interactions (36). In this study, we demonstrate that experimental line shape analyses conducted using these methods have an accuracy of < 5 %, and a precision of 13 % in the fitted value of k_{off}. In addition, an analysis of potential systematic errors in the values of R^0_{2A} and K_D indicates that the accuracy of the experimental line shape analysis as implemented in this work may range from half the precision at best, to being about the same as the precision at worst, at a level of $\pm 1\sigma$. The theoretical range of kinetics for which line shape analysis is applicable was established using quantum mechanical simulations with the magnetic resonance toolkit GAMMA. These simulations demonstrate that line shape analysis provides reasonably accurate and precise k_{off} values over a wide range spanning 100-15,000 s^{-1}. Furthermore, theoretical simulations indicate that the applicability of line shape analysis to the lower

range of k_{off} values, in the intermediate exchange regime, may be facilitated by more accurate K_{D} measurements, obtained by NMR-monitored chemical shift titrations where the dependence of K_{D} on the magnitude of chemical shift difference between the free and bound states is extrapolated to $\Delta\omega = 0$. The demonstrated level of accuracy and precision for k_{off} can be expected to be valuable for interpreting biological kinetics in the fast exchange regime for protein-protein interaction networks, where a modest change in the magnitude of a given on or off-rate due to a disease relevant mutation may lead to substantial changes in the populations of downstream protein complexes in a signalling cascade.

References

1. Petrini, J.H. (2007) Cell signaling. a touching response to damage. *Science* 316, 1138–1139.

2. Covert, M.W., Leung, T.H., Gaston, J.E., and Baltimore, D. (2005) Achieving stability of lipopolysaccharide-induced NF-κB activation. *Science* 309, 1854–1857.

3. Haglund, K., Di Fiore, P.P., and Dikic, I. (2003) Distinct monoubiquitin signals in receptor endocytosis. *Trends in Biochemical Sciences* 28, 598–603.

4. Gutowsky, H.S. and Saika, A. (1953) Dissociation, chemical exchange, and the proton magnetic resonance in some aqueous electrolytes. *The Journal of Chemical Physics* 21, 1688–1694.

5. Rogers, M.T. and Woodbrey, J.C. (1962) A proton magnetic resonance study of hindered internal rotation in some substituted N,N-dimethylamides. *The Journal of Physical Chemistry* 66, 540–546.

6. Fischer, J.J. and Jardetzky, O. (1965) Nuclear magnetic relaxation study of intermolecular complexes. The mechanism of penicillin binding to serum albumin. *Journal of the American Chemical Society* 87, 3237–3244.

7. McConnell, H.M. (1958) Reaction rates by nuclear magnetic resonance. *The Journal of Chemical Physics* 28, 430–431.

8. Binsch, G. (1969) Unified theory of exchange effects on nuclear magnetic resonance line shapes. *Journal of the American Chemical Society* 91, 1304–1309.

9. Nageswara Rao, B.D. (1989) Nuclear magnetic resonance line-shape analysis and determination of exchange rates. *Methods in Enzymology* 176, 279–311.

10. Bain, A.D. (2003) Chemical exchange in NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy* 43, 63–104.

11. Baldo, J.H., Halford, S.E., Patt, S.L., and Sykes, B.D. (1975) Stepwise binding of small molecules to proteins. Nuclear magnetic resonance and temperature jump studies of the binding of 4-(N-acetylaminoglucosyl)-N-acetylglucosamine to lysozyme. *Biochemistry* 14, 1893–1899.

12. Vasavada, K.V., Kaplan, J.I., and Nageswara Rao, B.D. (1980) Density matrix theory of ABC⇌AB+C chemical exchange: application to ^{31}P NMR spectra of interconverting enzyme-bound reactants and products of phosphoryl transfer enzymes. *Journal of Magnetic Resonance (1969)* 41, 467–482.

13. Shriver, J.W. and Sykes, B.D. (1981) Energetics and kinetics of the interconversion of two myosin subfragment-1•adenosine 5̀-diphosphate complexes as viewed by phosphorus-31 nuclear magnetic resonance. *Biochemistry* 20, 6357–6362.

14. Kern, D., Kern, G., Scherer, G., Fischer, G., and Drakenberg, T. (1995) Kinetic analysis of cyclophilin-catalyzed prolyl cis/trans isomerization by dynamic NMR spectroscopy. *Biochemistry* 34, 13594–13602.

15. Schmitt, T.H., Zheng, Z., and Jardetzky, O. (1995) Dynamics of tryptophan binding to escherichia coli Trp repressor wild type and AV77 mutant: an NMR study. *Biochemistry* 34, 13183–13189.

16. Johnson, P.E., Creagh, A.L., Brun, E., Joe, K., Tomme, P., Haynes, C.A., and McIntosh, L.P. (1998) Calcium binding by the N-terminal cellulose-binding domain from *Cellulomonas fimi* β-1,4-glucanase CenC. *Biochemistry* 37, 12772–12781.

17. Bax, A.D. and Grzesiek, S. (1993) Methodological advances in protein NMR. *Accounts of Chemical Research* 26, 131–138.

18. Tugarinov, V. and Kay, L.E. (2005) Methyl groups as probes of structure and dynamics in nmr studies of high-molecular-weight proteins. *ChemBioChem* 6, 1567–1577.

19. Gelis, I., Bonvin, A.M.J.J., Keramisanou, D., Koukaki, M., Gouridis, G., Karamanou, S., Economou, A., and Kalodimos, C.G. (2007) Structural basis for signal sequence recognition by the 204-kDa translocase motor SecA determined by NMR. *Cell* 131, 756.

20. McKay, R.T., Tripet, B.P., Pearlstone, J.R., Smillie, L.B., and Sykes, B.D. (1999) Defining the region of troponin-I that binds to troponin-C. *Biochemistry* 38, 5478–5489.

21. Korchuganov, D.S., Nolde, S.B., Reibarkh, M.Y., Orekhov, V.Y., Schulga, A.A., Ermolyuk, Y.S., Kirpichnikov, M.P., and Arseniev, A.S. (2001) NMR study of monomer-dimer equilibrium of barstar in solution. *Journal of the American Chemical Society* 123, 2068–2069.

22. Günther, U.L. and Schaffhausen, B. (2002) NMRKIN: simulating line shapes from two-dimensional spectra of proteins upon ligand binding. *Journal of Biomolecular NMR* 22, 201–209.

23. Tugarinov, V. and Kay, L.E. (2003) Quantitative NMR studies of high molecular weight proteins: application to domain orientation and ligand binding in the 723 residue enzyme malate synthase G. *Journal of Molecular Biology* 327, 1121–1133.

24. Rintala-Dempsey, A.C., Santamaria-Kisiel, L., Liao, Y., Lajoie, G., and Shaw, G.S. (2006) Insights into S100 target specificity examined by a new interaction between S100A11 and annexin A2. *Biochemistry* 45, 14695–14705.

25. Marintchev, A., Frueh, D., and Wagner, G. (2007) NMR methods for studying protein–protein interactions involved in translation initiation. *Methods in Enzymology* 430, 283–331.

26. Markin, C.J., Xiao, W., and Spyracopoulos, L. (2010) Mechanism for recognition of polyubiquitin chains: balancing affinity through interplay between multivalent binding and dynamics. *Journal of the American Chemical Society* 132, 11247–11258.

27. Greenwood, A.I., Rogals, M.J., De, S., Lu, K.P., Kovrigin, E.L., and Nicholson, L.K.

(2011) Complete determination of the Pin1 catalytic domain thermodynamic cycle by NMR lineshape analysis. *Journal of Biomolecular NMR* 51, 21–34.

28. Lian, L.Y. and Roberts, G.C.K. (2011) *Protein NMR Spectroscopy: Practical Techniques and Applications*. (John Wiley & Sons, Ltd, Chichester).

29. Arai, M., Ferreon, J.C., and Wright, P.E. (2012) Quantitative analysis of multisite protein–ligand interactions by NMR: binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP. *Journal of the American Chemical Society* 134, 3792–3803.

30. Morris, G.A. and Freeman, R. (1979) Enhancement of nuclear magnetic resonance signals by polarization transfer. *Journal of the American Chemical Society* 101, 760–762.

31. Bodenhausen, G. and Ruben, D.J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters* 1, 185–189.

32. Palmer, A.G., Cavanagh, J., Wright, P.E., and Rance, M. (1991) Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation nmr spectroscopy. *Journal of Magnetic Resonance (1969)* 93, 151–170.

33. Kay, L.E., Keifer, P., and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *Journal of the American Chemical Society* 114, 10663–10665.

34. Kovrigin, E.L. (2012) NMR line shapes and multi-state binding equilibria. *Journal of Biomolecular NMR* 53, 257–270.

35. Palmer 3rd, AG, Kroenke, C.D., and Loria, J.P. (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods in Enzymology* 339, 204–238.

36. Markin, C.J. and Spyracopoulos, L. (2012) Increased precision for analysis of protein–ligand dissociation constants determined from chemical shift titrations. *Journal of Biomolecular NMR* 53, 125–138.

37. Smith, S.A., Levante, T.O., Meier, B.H., and Ernst, R.R. (1994) Computer simulations in magnetic resonance. An object-oriented programming approach. *Journal of Magnetic Resonance, Series A* 106, 75–105.

38. Hammes, G.G. (2000) *Thermodynamics and kinetics for the biological sciences*. (Wiley-Interscience New York).

39. Sutherland, I.O. (1971) The investigation of the kinetics of conformational changes by nuclear magnetic resonance spectroscopy. *Annual Reports on NMR Spectroscopy* 4, 71–235.

40. Hoult, D.I. (1978) The NMR receiver: a description and analysis of design. *Progress in Nuclear Magnetic Resonance Spectroscopy* 12, 41–77.

41. McKenna, S., Hu, J., Moraes, T., Xiao, W., Ellison, M.J., and Spyracopoulos, L. (2003) Energetics and specificity of interactions within UbUevUbc13 human ubiquitin conjugation complexes. *Biochemistry* 42, 7922–7930.

42. Markin, C.J., Saltibus, L.F., Kean, M.J., McKay, R.T., Xiao, W., and Spyracopoulos, L. (2010) Catalytic proficiency of ubiquitin conjugation enzymes: Balancing pk_a suppression, entropy, and electrostatics. *Journal of the American Chemical Society* 132, 17775–17786.

43. Cavanagh, J. (2007) *Protein NMR spectroscopy: principles and practice*. (Academic Press, Boston).

44. Hammarström, A. and Otting, G. (1994) Improved spectral resolution in ^{1}H NMR spectroscopy by homonuclear semiselective shaped pulse decoupling during acquisition. *Journal of the American Chemical Society* 116, 8847–8848.

45. Aguilar, J.A., Faulkner, S., Nilsson, M., and Morris, G.A. (2010) Pure shift ^{1}H NMR: A resolution of the resolution problem? *Angewandte Chemie International Edition* 49, 3901–3903.

46. Sakhaii, P., Haase, B., and Bermel, W. (2009) Experimental access to HSQC spectra decoupled in all frequency dimensions. *Journal of Magnetic Resonance* 199, 192–198.

47. LeMaster, D.M. (1994) Isotope labeling in solution protein assignment and structural analysis. *Progress in Nuclear Magnetic Resonance Spectroscopy* 26, 371–419.

48. Gardner, K.H. and Kay, L.E. (1998) The use of ^{2}H, ^{13}C, and ^{15}N multidimensional NMR to study the structure and dynamics of proteins. *Annual Review of Biophysics and Biomolecular Structure* 27, 357–406.

49. Grzesiek, S. and Bax, A. (1994) Interference between dipolar and quadrupolar interactions in the slow tumbling limit: a source of line shift and relaxation in ^{2}H-labeled compounds. *Journal of the American Chemical Society* 116, 10196–10201.

50. Ghai, R., Falconer, R.J., and Collins, B.M. (2012) Applications of isothermal titration calorimetry in pure and applied research-survey of the literature from 2010. *Journal of Molecular Recognition* 25, 32–52.

Chapter 7

Conclusions

Although many conclusions can be drawn regarding the synthesis and recognition of polyUb chains, a better description of this chapter is perhaps *future directions*. The work described herein naturally lends itself to further application, which could perhaps be taken as a reflection of its impact, and hopefully its general relevance. As the preceding chapters have all been published, the major conclusions of the work have been articulated at the end of each. That being said, it will nevertheless be useful to take a brief, broader perspective, reaffirming where we are with regard to understanding these systems and to chart a general course for future studies.

The methodological chapters, 5 and 6, seek to improve the measurements of thermodynamics and kinetics using NMR spectroscopy. Although the emphasis in this work is on intermolecular interactions, the potential still remains to see if intramolecular interactions are amenable to similar or analogous methods. The simultaneous sampling of protein and ligand concentrations described in Chapter 5 is by definition limited to intermolecular interactions. NMR, however, is not the only technique used for measuring dissociation constants for protein-protein interactions: it is of interest if methods such as isothermal titration calorimetry (ITC) would also benefit from this approach, especially regarding the dilution of a mixture of protein and ligand. In Chapter 6, the applicability of line-shape analysis over a wide-range of biologically relevant k_{ex} values was demonstrated. Although the simulations were carried out for an intermolecular exchange process, addition of ligand to a binding partner is merely a specific method of

shifting an established equilibrium. Quantifying changes as a function of temperature (*cf.* (1), where CPMG relaxation-dispersion was carried out at various temperatures), or of addition of progressive amounts of denaturant (2), are two examples of methods suited to intramolecular processes.

The work in chapters 3 and 4 provides new understanding of the synthesis and subsequent recognition of polyUb chains. Chapter 3 provided quantitative insight into the mechanism of enhanced affinity using multivalent interactions. The rapid kinetics observed for binding between RAP80-tUIM and polyUb chains suggests a mechanistic basis underlying the transient interactions observed in the DNA damage response. Multivalency is a common mechanism used to increase intrinsic affinity in diverse biological systems (3), many of which should be amenable to these methods. In regard to recognizing and responding to DNA damage, new research has uncovered a potential familial mutation in the UIM motifs of RAP80 which may compromise the ability of the DDR to detect and respond to DNA damage (4). Our methodologies may be of use in determining the molecular basis of this deficiency. The increasing polyvalency acquired by polyUb chains may also be a mechanism of self-regulation of chain synthesis, by resulting in non-productive binding to Mms2. These topics are discussed in the subsections below.

Our research into the catalytic mechanisms underlying polyUb chain synthesis by the E2 Ubc13 and its binding partner Mms2 showed that this enzyme uses a combination of pK_{a} suppression of the acceptor lysine, entropic effects of binding and positioning the substrate Ub, and electrostatic complementarity in the active site. There are many unanswered questions, however, regarding the role of E3 proteins in increasing the catalytic rate of these enzymes. Additionally, given the diverse signalling roles played by Ub, it remains to establish in more detail the unanswered questions regarding the synthesis of not only K63-linked polyUb, but chains of other linkage type as well.

Rate enhancement of polyUb chain synthesis by E3 enzymes

Recent work suggests that RING E3 enzymes, rather than acting only as binding scaffolds to bring together the E2 and substrate, play an additional role in enhancing the catalytic efficiency of E2 enzymes (5, 6). Although the allosteric mechanism proposed is for UbcH5c, the structural similarity and presence of the critical leucine in helix 2 of Ubc13 presents the possibility of a similar activation (6). RNF8 having been shown qualitatively to enhance chain synthesis (5) indicates that our measurements of the k_{cat} of Ub_2 synthesis reflect the basal activity of the enzyme.

Although the binding of E4BU to UbcH5c suggests that the interaction of the donor Ub with UbcH5c is favoured, the details of how this change in conformational mobility of the Ub affects the active site and consequently, the catalytic rate, remain obscure. With the kinetic analyses we developed for polyUb chain formation by the Ubc13-Mms2, we will be able to introduce RNF8 and directly measure the effect of adding this E3 on the k_{cat}. Mutations at or near the active site may also provide insight into the mechanism, if the E3 bound state can be mimicked in this manner.

One possible issue is that addition of RNF8 may increase the rate of reaction to a point where it would be difficult to accurately measure without the use of specialized quench-flow instrumentation. A possible solution to this is to add only a small amount of E3 relative to Ubc13, keeping the fraction of activated E3-bound Ubc13 high enough to result in a measurable increase in rate while low enough so that the increased rate does not preclude accurate quantification. This is a further advantage of the ODE method, whereby knowing the concentrations and affinities of all reaction components makes it possible to deconvolute their effects from changes solely to k_{cat}. Depending on the magnitude of k_{ex}, the kinetics of heterodimer-RNF8 binding should be amenable to one or more of the NMR relaxation experiments outlined in Chapter 1, whether line-shape analysis, CPMG relaxation-dispersion, or ZZ-exchange spectroscopy.

Kinetic control of polyUb chain synthesis

Local concentration and multivalency

As discussed in the first chapter, increasing affinities through multivalency and tethering is a common theme in biological interactions (3). We have developed NMR methods to characterize these effects in terms of the microscopic binding constants for the interaction of divalent RAP80-tUIM with K63-linked polyUb chains. Their role in the ubiquitination cascade, however, does not seem to be limited only to downstream signal recognition; these effects are also crucial to kinetic control of the synthesis of polyUb chains. Consider the phenomenon of processivity introduced in Chapter 1 with regard to the synthesis of K48-linked polyUb chains on substrates targeted for degradation (7). The E3s most commonly responsible for synthesizing K48-linked polyUb are the SCF and SCF-like complexes, binding both E2 and substrate to facilitate ubiquitination of the latter. They consist of a RING subunit, Rbx1, which binds the E2, and a variable F-box protein to bind a specific substrate (8). These proteins are held approximately 50-60 Å apart (9) by binding to the Cul1 scaffold (10) (figure 1.6). The substrate binds on one side with higher relative affinity than E2 binding to Rbx1 on the opposite, facilitating multiple rounds of processive ubiquitination by allowing the E2 to dissociate and recharge between chain lengthening iterations (7, 11). Interestingly, this ubiquitination machinery prefers K48 of Ub acting as the acceptor, with the addition of the first Ub to substrate relatively slow and, therefore, the rate-limiting step (11). In other words, these enzymes seem to be optimized for chain elongation rather than initiation. The elegant work of Deshaies and colleagues has shown that the overall length of chain is related to the k_{cat} for addition of this first Ub to a substrate lysine (12). By extension, the off-rate of substrate binding to E3 takes on an importance no less than that of the overall k_{cat} in determining the final length of chain, the probability of it reacting seemingly being proportional to its bound lifetime. In Pierce *et al.* (12), for CYCE and β-catenin, two substrates with a $\sim$ 10-fold varying affinity between them, the percentage of chains that are long enough to be proteasome substrates differs by a factor of 3. Although the relative interplay of the rate of adding the first Ub to substrate (hereafter referred to as $k^{1}_{\mathrm{cat,Ub}}$ for clarity) and the k_{off} of E3-substrate binding accounts for much of this difference, the rates of subsequent

Ub addition to the growing chain are also different between the two substrates (12).

At first glance, this seems an unusual result. Studying enzymes, we are used to thinking of their reactions taking place in well-defined active sites, and any changes in rate being the result of slightly different interactions or geometry at these sites. This certainly explains the k_{cat} difference between adding the first Ub to a lysine of the substrate itself, as opposed to K48 of Ub on the growing chain, as there are differences in the local environments around these lysine side-chain amines. But, for example, the K48 side-chain of Ub_1 is chemically identical to that of Ub_2, so how is it that $k^2_{cat,Ub}$ is 4 s^{-1} while $k^3_{cat,Ub}$ is merely 1 s^{-1} (12)?

The answer would seem to lie in the conformational flexibility of polyUb chains (13). Due to the flexibility of the tether, as the length increases so does the volume that one end can sample relative to the other. Combining this flexibility with the processivity inherent in K48-linked polyUb chain synthesis by the SCF complex suggests that the rate is dependent on the local concentration of the reactive K48 at the Cdc34 active site. This would indicate that the kinetic rate constants given in reference (12) are apparent, rather than true k_{cat}s, where $k_{cat,true}[K48]_{local} = k^n_{cat,app}[K48]_{solution}$. Interestingly, when discussing the decreased rate for addition of the first Ub to substrate compared to that for subsequent Ub addition, Petrowski and Deshaies suggested that the increased local concentration of substrate lysine facilitated by its binding to the E3 was essential to prevent preferential reaction with the 10-20 µM free Ub in the cell (14). If accurate calculations or measurements of $[K48]_{local}$ were obtained for chains of increasing length, it could be demonstrated that this length dependence is the major determinant of the observed rates of chain extension.

Extended chain synthesis by the Ubc13-Mms2 heterodimer

Similar questions pertain to the length dependence of K63-linked polyUb synthesis by the Ubc13-Mms2 heterodimer. Unlike substrate binding to E3 above, characterized by a k_{off} on the order of 0.4 s^{-1} (12), Ub binds to Mms2 with a k_{off} of between $\sim$ 500 s^{-1} and $\sim$ 2000 s^{-1} (15). This would likely rule out any significant effects of processivity in this system. In a sense, although it shares the E2 fold and binds at a different site on Ubc13, Mms2 acts as an E3 would in binding a substrate to position it for ubiquitination

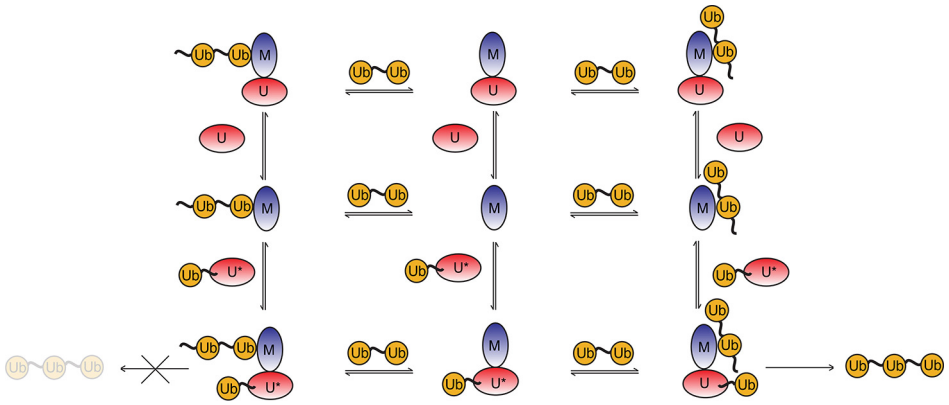

Figure 7.1: Ub_3 formation by the Ubc13-Mms2 heterodimer, accounting for both productive (right) and non-productive binding modes (left) of Ub_2 to Mms2. The K63 sidechain of the acceptor Ub on Ub_2 is indicated by the unjoined wavy line.

(16, 17). In our Ub_2 assays, Mms2 bound mono-Ub, K63 of which is the acceptor lysine. To synthesize longer polyUb chains of length n, where $n > 2$, the $n-1$ Ub of the growing chain must bind Mms2 and be positioned for attack. As the chain grows however, the probability of the proper Ub binding Mms2 should decrease (figure 7.1). This would seem to be an example of polyvalency reducing the apparent catalytic rate. The possibility exists that the E3 plays a role in positioning the proper Ub for attack, but RNF8 has not been shown to bind Ub in GST pulldown assays (5). This result does not preclude the existence of weak, transient interactions between Ub and the coiled-coil domain of RNF8 however. In any event, it remains unclear how the Ubc13-Mms2-RNF8 interaction would specifically recognize and properly position the terminal acceptor Ub on the growing chain. Although RNF168 has Ub binding MIU domains (18, 19), the mechanism by which selection of the proper Ub on growing chains would occur is similarly unclear.

To analyze these kinetics, an approach using the methods described in both Chapters 3 and 4 will be developed. It is a straightforward, though somewhat onerous, task to extend the system of ODEs constructed for Ub_2 synthesis to account for chains of greater length. In order to account for the polyvalency (this term to distinguish a system using multiple binding sites rather than a true multivalent introduction, *cf.* Chapter 1) of the growing chain, models analogous to those given in the supporting information of Chapter 3 will be used. Instead of being formulated in terms of a system of equilibria however, these models will have to be ODEs so that the kinetics, rather than the thermodynamics,

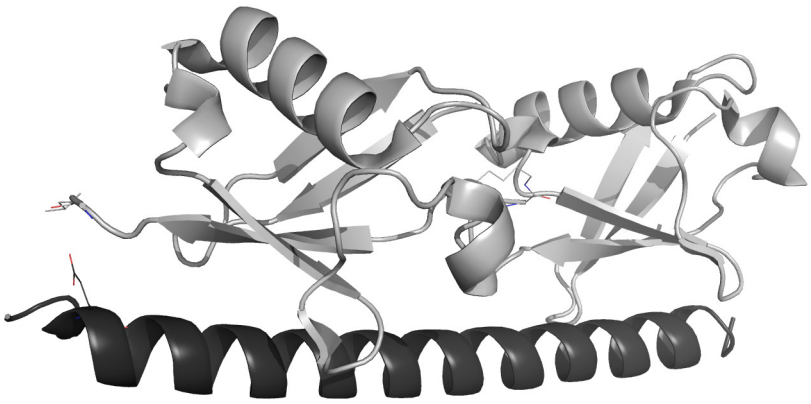

Figure 7.2: Crystal structure of RAP80-tUIM (dark grey) bound to K63-linked Ub_2 (light grey) (21). The backbone and side-chain atoms are shown for E81 of RAP80-tUIM and L73 of the proximal Ub molecule, illustrating the H-bond. The K63-G77 linkage is also shown in sticks, to more easily denote the proximal and distal Ub moieties.

of polyUb binding to Mms2-Ubc13 are the pertinent parameters.

Mutational analyses of RAP80-tUIM

Mechanism of decreased polyUb binding by ΔE81-RAP80-tUIM

While developing our multivalent analysis of RAP80-tUIM binding to polyUb chains, we became aware of research suggesting that a single residue deletion in the UIM region may confer susceptibility to familial breast cancer (4). The deleted residue, E81, is located at the N-terminus of the first UIM; its absence was shown to qualitatively disrupt polyUb binding and reduce the localization of RAP80 to DNA-damage induced foci (4). The authors hypothesized that the ΔE81 mutation results in the inability of the first UIM to form an intermolecular hydrogen bond with the bound Ub, on the basis of the structure of Vps27 bound to Ub (20). A similar H-bond is shown between E81 and L73 of Ub in the crystal structure of K63-linked Ub_2 bound to RAP80-tUIM (21).

Having shown that at 37 ℃ the helices of both RAP80 are only 50 % helical (22), we hypothesize that the loss of E81 may contribute to a decrease in binding affinity due to a further destabilization of the helix. Glutamate is a residue that has been shown to occur frequently at the N-termini of α-helices (23) and to impart a stabilizing

effect (24), its side-chain being able to accept hydrogen bonds from the NH donors of the first four residues of the helix, while the negative charge complements the overall helix dipole (25). Both the N and C-terminal UIMs of RAP80 have multiple glutamate residues at their respective N-termini. Since only one is required for the hydrogen bond with Ub, the presence of additional glutamates would suggest that they may also assist in shifting the helix-coil equilibrium (22) towards the helical state. In reference (20), the authors attribute the presence of multiple glutamines not to a specific H-bond, but rather suggest electrostatic interactions with R42 and R72 on Ub. When discussing their structure however, the authors attribute the main interaction between the two proteins to the L8-I44-V70 hydrophobic patch on Ub with the hydrophobic residues in the middle of the UIM (20). This is affirmed by mutation of I44A of Ub showing a qualitatively large decrease in binding (20). Unfortunately, R42 and R72 were not similarly mutated to assess if loss of this positive patch on Ub would cause a similar reduction in binding affinity. An idea of the relative contributions of H-bonding to the affinity of UIM-Ub interactions may be taken from the A46D mutation however, which forms a H-bond with the conserved S270 (S92 in RAP80-tUIM), and does not appear to have nearly as deleterious an effect on binding (20). On the other hand, deletion of the conserved S270 was shown in an earlier study to reduce Ub binding to Vps27 sixfold (26).

Although its molecular basis remains unresolved, the effect of the ΔE81 mutation on reducing polyUb chain binding is apparent. Preliminary titrations of ΔE81-RAP80-tUIM with Ub and Ub_2 indicates a drastic loss of binding by the N-terminal UIM, though small chemical shift changes are nevertheless present. The decrease in chemical shift perturbation of UIM-1 residues in the latter titration especially indicates an extreme effect on binding, as the local concentration of the proximal Ub upon binding of the distal Ub to the C-terminal UIM is known to be ~ 4 mM (22). The question is therefore whether this loss of affinity is solely the result of being unable to form specific electrostatic or H-bond interactions with Ub, or if an additional reduction in helical stability reduces the fraction of UIM available for productive binding. Assignment of the ΔE81 mutant C_α and H_α resonances and comparison with those previously determined for wild-type RAP80-tUIM (22) may indicate a loss of helicity. NMR temperature titrations and differential scanning calorimetry (DSC) may also be useful in assessing the relative stability of the

wild-type and mutant RAP80-tUIM proteins. Since the H-bond is formed between the amide backbone of L73, mutation of this residue to another would not be expected to disrupt this interaction. On the other hand, if this and the subsequent 3 C-terminal residues of Ub were simply deleted, then this may give an indication of the necessity of this interaction for productive binding.

Linker helicity

Our relaxation data (22) is not in agreement with the crystal structure of RAP80-tUIM bound to K63-linked Ub_2 (21). In the latter, the linker between UIM-1 and UIM-2 is helical, so that the entire RAP80-tUIM molecule forms an unbroken α-helix 7.2. NMR relaxation data of RAP80-tUIM bound with tandem Ub_2 however, indicates that the linker, though less flexible than in the free state, nevertheless is less rigid than the UIMs themselves (*cf.* figure 3.11). A similar conclusion can be drawn from the chemical shift data (*cf.* figure 3.8a), where the linker is shown to be intrinsically less helical than the UIMs in the free state.

This discrepancy could be the result of our use of tandem Ub_2 chains rather than K63-linked Ub_2; however the presence of multivalent binding for tandem chains suggests that the interaction is similar. To further address this unresolved question, binding analyses will be carried out with two linker mutations, N100G and N100P, using our previously developed models and titration methodologies. The presence of a glycine or proline in the linker should disrupt any helical conformation in this region and reduce Ub_2 multivalent binding if a helical linker is necessary for the interaction. To ascertain whether the linker is helical for interaction with K63-linked polyUb only, titrations with these chains will be carried out as well.

References

1. Mulder, F.A.A., Mittermaier, A., Hon, B., Dahlquist, F.W., and Kay, L.E. (2001) Studying excited states of proteins by NMR spectroscopy. *Nature Structural & Molecular Biology* 8, 932–935.

2. Kuhlman, B., Luisi, D.L., Evans, P.A., and Raleigh, D.P. (1998) Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the N-terminal domain of the protein L9. *Journal of Molecular Biology* 284, 1661–1670.

3. Mammen, M., Choi, S.K., and Whitesides, G.M. (1998) Polyvalent interactions in biological systems: implications for design and use of multivalent ligands and inhibitors. *Angewandte Chemie International Edition* 37, 2754–2794.

4. Nikkilä, J., Coleman, K.A., Morrissey, D., Pylkäs, K., Erkko, H., Messick, T.E., Karppinen, S.M., Amelina, A., Winqvist, R., and Greenberg, R.A. (2009) Familial breast cancer screening reveals an alteration in the RAP80 UIM domain that impairs DNA damage response function. *Oncogene* 28, 1843–1852.

5. Campbell, S.J., Edwards, R.A., Leung, C.C.Y., Neculai, D., Hodge, C.D., Dhe-Paganon, S., and Glover, J.N.M. (2012) Molecular insights into the function of RING finger (RNF)-containing proteins hRNF8 and hRNF168 in Ubc13/Mms2-dependent ubiquitylation. *Journal of Biological Chemistry* 287, 23900–23910.

6. Pruneda, J.N., Littlefield, P.J., Soss, S.E., Nordquist, K.A., Chazin, W.J., Brzovic, P.S., and Klevit, R.E. (2012) Structure of an E3:E2∼Ub complex reveals an allosteric mechanism shared among RING/U-box ligases. *Molecular Cell* 47, 933–942.

7. Kleiger, G., Saha, A., Lewis, S., Kuhlman, B., and Deshaies, R.J. (2009) Rapid E2-E3 assembly and disassembly enable processive ubiquitylation of cullin-RING ubiquitin ligase substrates. *Cell* 139, 957–968.

8. Cardozo, T. and Pagano, M. (2004) The SCF ubiquitin ligase: insights into a molecular machine. *Nature Reviews Molecular Cell Biology* 5, 739–751.

9. Liu, J. and Nussinov, R. (2010) Molecular dynamics reveal the essential role of linker motions in the function of cullin–RING E3 ligases. *Journal of Molecular Biology* 396, 1508–1523.

10. Zheng, N., Schulman, B.A., Song, L., Miller, J.J., Jeffrey, P.D., Wang, P., Chu, C., Koepp, D.M., Elledge, S.J., Pagano, M., Conaway, R.C., Conaway, J.W., Harper, J.W., and Pavletich, N.P. (2002) Structure of the Cul1–Rbx1–Skp1–Fbox^{skp2} SCF ubiquitin ligase complex. *Nature* 416, 703–709.

11. Petroski, M.D. and Deshaies, R.J. (2005) Mechanism of lysine 48-linked ubiquitin-chain synthesis by the cullin-RING ubiquitin-ligase complex SCF-Cdc34. *Cell* 123, 1107–1120.

12. Pierce, N.W., Kleiger, G., Shan, S., and Deshaies, R.J. (2009) Detection of sequential polyubiquitylation on a millisecond timescale. *Nature* 462, 615–619.

13. Ryabov, Y. and Fushman, D. (2006) Interdomain mobility in di-ubiquitin revealed by NMR. *Proteins: Structure, Function, and Bioinformatics* 63, 787–796.

14. Haas, A.L. (1988) Ubiquitin. *Ubiquitin*, ed. Rechsteiner, M. (New York: Plenum Press), pp. 173–206.

15. Markin, C.J., Saltibus, L.F., Kean, M.J., McKay, R.T., Xiao, W., and Spyracopoulos, L. (2010) Catalytic proficiency of ubiquitin conjugation enzymes: Balancing pk_a suppression, entropy, and electrostatics. *Journal of the American Chemical Society* 132, 17775–17786.

16. Lewis, M.J., Saltibus, L.F., Hau, D.D., Xiao, W., and Spyracopoulos, L. (2006) Structural basis for non-covalent interaction between ubiquitin and the ubiquitin

conjugating enzyme variant human MMS2. *Journal of Biomolecular NMR* 34, 89–100.

17. Eddins, M.J., Carlile, C.M., Gomez, K.M., Pickart, C.M., and Wolberger, C. (2006) Mms2–Ubc13 covalently bound to ubiquitin reveals the structural basis of linkage-specific polyubiquitin chain formation. *Nature Structural & Molecular Biology* 13, 915–920.

18. Doil, C., Mailand, N., Bekker-Jensen, S., Menard, P., Larsen, D.H., Pepperkok, R., Ellenberg, J., Panier, S., Durocher, D., Bartek, J., Lukas, J., and Lukas, C. (2009) RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins. *Cell* 136, 435–446.

19. Penengo, L., Mapelli, M., Murachelli, A.G., Confalonieri, S., Magri, L., Musacchio, A., Di Fiore, P.P., Polo, S., and Schneider, T.R. (2006) Crystal structure of the ubiquitin binding domains of rabex-5 reveals two modes of interaction with ubiquitin. *Cell* 124, 1183–1195.

20. Swanson, K.A., Kang, R.S., Stamenova, S.D., Hicke, L., and Radhakrishnan, I. (2003) Solution structure of Vps27 UIM–ubiquitin complex important for endosomal sorting and receptor downregulation. *The EMBO Journal* 22, 4597–4606.

21. Sato, Y., Yoshikawa, A., Mimura, H., Yamashita, M., Yamagata, A., and Fukai, S. (2009) Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by tandem UIMs of RAP80. *The EMBO Journal* 28, 2461–2468.

22. Markin, C.J., Xiao, W., and Spyracopoulos, L. (2010) Mechanism for recognition of polyubiquitin chains: balancing affinity through interplay between multivalent binding and dynamics. *Journal of the American Chemical Society* 132, 11247–11258.

23. Aurora, R. and Rose, G.D. (1998) Helix capping. *Protein Science* 7, 21–38.

24. Cochran, D.A.E., Penel, S., and Doig, A.J. (2008) Effect of the N1 residue on the stability of the α-helix for all 20 amino acids. *Protein Science* 10, 463–470.

25. Presta, L.G. and Rose, G.D. (1988) Helix signals in proteins. *Science* 240, 1632–1641.

26. Shih, S.C., Katzmann, D.J., Schnell, J.D., Sutanto, M., Emr, S.D., and Hicke, L. (2002) Epsins and Vps27p/Hrs contain ubiquitin-binding domains that function in receptor endocytosis. *Nature Cell Biology* 4, 389–393.

Appendix A

Kinetic and thermodynamic simulations

In the discussion of thermodynamic and kinetic measurements by NMR in chapter 1, simulations of NMR line-shapes were included to illustrate the underlying physics and mathematics. To avoid unnecessary obfuscation of the contents of that chapter, the details of those simulations are given here for the sake of completeness. For the same reason of clarity, included in this appendix is an explanation of how to solve for the total concentrations of the bound states of the Vps27-Ub interaction in figure 1.8 using *Mathematica* (1).

In the section of chapter 1 entitled *Kinetics of ubiquitination*, backfits of the catalytic rate constant, k_{cat}, were performed on data simulated using eqns. 4.22-4.31 of this work, using the numerically integrated equations describing the Michaelis-Menten kinetic model. These are given below, along with a brief explanatory discussion.

Line-shape simulations

The line-shape simulations used to demonstrate the concepts introduced in chapter 1 were carried out in much the same way as those used in chaper 6 to illustrate the effect of $\Delta\delta$ on the accuracy and precision of the fitted K_{D}, in the section entitled *theoretical accu-*

racy and precision of K_D values from NMR-monitored chemical shift titrations. A function was written in *Mathematica* to calculate discrete FIDs using the Bloch-McConnell equations (eqns. 1.20 to 1.26) (2), sampled at $\Delta t = t_{acq}/np$ s, where t_{acq} is the total acquisition time, 0.119 s, and np is the number of points, 576 in this case. In chapter 6, as the system studied is the interaction of Mms2 with Ub, R^0_{2A} and R^0_{2B} were different, to reflect the increased molecular tumbling time of the bound complex. For the purposes of the introductory chapter however, this was felt to be an unnecessary complication, so $R^0_{2A} = R^0_{2B}$ in these cases. Regarding the other parameters, the actual values of Ω_A and Ω_B are somewhat arbitrary, as after Fourier transformation the resulting spectra are referenced to the desired ^{15}N chemical shift in ppm. Their difference, $\Delta\omega$, as well as p_A, p_B, and k_{ex} are varied depending on the particular simulation. Once the FID is simulated, it is zero-filled to 2048 points, Fourier transformed, and referenced to the desired ^{15}N chemical shift to give an NMR line-shape in the frequency domain.

For intermolecular exchange, k_{ex} is the sum of the pseudo-first-order rate constant, $k'_{on} = k_{on}[L]$ and k_{off}, where [L] is the free ligand concentration in solution. Calculation of this value necessitates solving the system of equations describing a 1:1 protein-ligand interaction for [P], [L], and [PL]:

$$K_D = \frac{[P][L]}{[PL]} \tag{A.1}$$

$$[P]_0 = [P] + [PL] \tag{A.2}$$

$$[L]_0 = [L] + [PL] \tag{A.3}$$

This yields the following solution:

$$[L] = \frac{1}{2}\left(\sqrt{(-K_D - [L]_0 - [P]_0)^2 - 4[L]_0[P]_0} - K_D + [L]_0 - [P]_0\right) \tag{A.4}$$

For the line-shape simulations shown in figure 1.10, demonstrating the effect of different labelled protein concentrations on the determination of K_D, a random FID noise was value simulated and added at every point. This was done by randomly sampling a normal distribution with a mean of 0 and a standard deviation chosen to give a signal-to-noise ratios of 100 and 10, for figures 1.10C and D, respectively. This value is also somewhat arbitrary, as the overall signal-to-noise is determined both by this parameter, and the

overall $M(0)$ values chosen in equations 1.20 and 1.21. In practice, achieving the desired signal-to-noise was done by inspection. The protein and ligand concentrations used to generate the sample titration spectra using this method were as follows. For C, $[P]_0$ was held constant at 1 mM, as described; $[L]_0$ values were 0, 0.2, 0.4, 0.6, 0.8, and 1.0 mM. For D, $[P]_0$ was 0.1 mM; $[L]_0$ values were as for C, to facilitate comparison of the relative precision of the two methods.

To simulate an experimental NMR titration, the chemical shifts of these peaks were determined using Lorentzian fitting and fit to equation 1.14, with K_D and $\Delta\delta$ as fitted parameters. We have previously used a parabolic interpolation of the most intense point, n, and the neighbouring $n-1$ and $n+1$ points to determine chemical shifts (chapter 5); however, in practice we have observed that Lorenztian fitting is a more accurate method (chapter 6), especially for cases of low signal-to-noise, although as being a fitting procedure it requires more computational time. As equation 1.14 relates $\Delta\delta_{obs}$ to K_D and $\Delta\delta_{max}$, these fitted shifts are converted to $\Delta\delta$ values by taking the absolute value of the difference between the shift at $[L]_0 = 0$ and subsequent shifts where $[L]_0 > 0$.

The Monte Carlo analysis to determine the accuracy and precision in these titrations was performed simply by iterating this entire procedure 100 times, then determining the mean and standard deviations of the fitted K_D and $\Delta\delta_{max}$ values.

For the intermolecular exchange processes in the right panel of figure 1.15, line-shapes were simulated in the same manner, except in the absence of FID noise. $k_o n = 1 \times 10^7$ M^{-1} s^{-1}, k'_{on} was calculated by multiplication by [L] using equation A.4, and k_{off} values were 5, 50, 500, and 5000 s^{-1} from top to bottom. The total protein concentration was held constant at 0.5 mM in all these spectra, while $[L]_0$ was 0, 0.25, 0.5, 0.75, 1.0, 1.25, and 1.5 mM (red to blue curves). The left panel shows line-shapes for an intramolecular exchange process, which are calculated in much the same manner except that k_1 (k_1 is used here instead of k_{on} because the latter is only really meaningful for an intermolecular process) can be used directly. For simplicity, all these cases assume an equilibrium constant of 1, so that $p_A = p_B = 0.5$ and $k_1 = k_{-1}$. Values of the kinetic constants used were 5, 50, 500, and 5000 s^{-1}.

In the discussion of CPMG relaxation-dispersion methods, figure 1.14 illustrates the effect of an intermolecular exchange process on the appearance of both the FIDs and the

NMR line-shapes after Fourier transformation. (A) and (B) were simulated in the same manner as for the line-shape figures described above, except that the Fourier transformation was removed to allow plotting of the FIDs directly. The transformation was then reinstated to facilitate plotting of the frequency domain, in (C). Values of p_{A} and p_{B} were 0.5, $R_2^0 = 12\ \mathrm{s}^{-1}$ for both free and bound states, $\Delta\delta = 180$ HZ, and $k_{\mathrm{ex}} = 2000\ \mathrm{s}^{-1}$.

Functions describing the Vps27 equilibrium

In the discussion *enhancing affinity through multivalency*, the equilibrium between the tandem UIMs of Vps27 and Ub was used as an example (figure 1.8). These curves can be generated by solving equations 1.1 to 1.4 using the *Solve* function implemented in *Mathematica* (1):

$$\begin{aligned}
&\mathrm{Solve}[\{K_{\mathrm{D},1} == [\mathrm{Ub}][\mathrm{Vps27}]/[\text{Ub-Vps27}]_1, K_{\mathrm{D},2} == [\mathrm{Ub}][\mathrm{Vps27}]/[\text{Ub-Vps27}]_2,\\
&K_{\mathrm{D},1} == [\mathrm{Ub}][\text{Ub-Vps27}]_2/[\text{Ub-Vps27}], K_{\mathrm{D},1} == [\mathrm{Ub}][\text{Ub-Vps27}]_2/[\text{Ub-Vps27}],\\
&[\mathrm{Vps27}]_0 == [\mathrm{Vps27}] + [\text{Ub-Vps27}]_1 + [\text{Ub-Vps27}]_2 + [\text{Ub-Vps27}],\\
&[\mathrm{Ub}]_0 == [\mathrm{Ub}] + [\text{Ub-Vps27}]_1 + [\text{Ub-Vps27}]_2 + 2[\text{Ub-Vps27}]\},\\
&\{[\mathrm{Ub}] + [\mathrm{Vps27}] + [\text{Ub-Vps27}]_1 + [\text{Ub-Vps27}]_2 + [\text{Ub-Vps27}]\}]
\end{aligned} \tag{A.5}$$

For more complex equilibria, such as those for the multivalent binding of polyUb chains to RAP80-tUIM in chapter 4, *Solve* often is unable to produce a symbolic result. In these cases, *NSolve* is used. This does not return a purely symbolic function; rather, values of parameters have to be set (K_{D}s, $[\mathrm{P}]_0$ and $[\mathrm{L}]_0$) before using the function.

ODEs describing the Michaelis-Menten kinetic scheme

In order to generate data in a form amenable to fitting of k_{cat}, eqns. 4.22-4.31 describing the system of coupled equilibria underlying the formation of Ub_2 by the Ubc13-Mms2 heterodimer were solved with various values of k_{cat}, and k_{off} for the interactions between Ubc13 and Mms2, Ub and Mms2, and Ub and the heterodimer. These values are summarized in table 1.2. For these values, the concentration of Ub_2 was calculated at intervals

of 175 s, from 0 to 1750 s total reaction time. k_{cat} was then fitted to these data with the assumption of Michaelis-Menten kinetics, using the system of ODEs given below, rather than those describing the full equilibrium (*cf.* figure 1.23):

$$\frac{d[\text{E}]}{dt} = -k_{\text{on}}[\text{E}][\text{S}] + k_{\text{off}}[\text{ES}] + k_{\text{cat}}[\text{ES}] \tag{A.6}$$

$$\frac{d[\text{S}]}{dt} = -k_{\text{on}}[\text{E}][\text{S}] + k_{\text{off}}[\text{ES}] \tag{A.7}$$

$$\frac{d[\text{ES}]}{dt} = k_{\text{on}}[\text{E}][\text{S}] - k_{\text{off}}[\text{ES}] - k_{\text{cat}}[\text{ES}] \tag{A.8}$$

$$\frac{d[\text{P}]}{dt} = k_{\text{cat}}[\text{ES}] \tag{A.9}$$

To determine the best back-fitted k_{cat} values the squared difference between the data simulated using eqns. 4.22-4.31 and the product build-up curve generated from the above eqns. A.6 to A.9 at different values of k_{cat} was minimized using an in-house simulated annealing algorithm. As described in chapter 1, in order to make Michaelis-Menten analysis feasible, Ub must be considered the 'enzyme', *E*, and Mms2-Ubc13∼Ub the substrate, *S*, so that the product, *P*, build-up curves saturate at the concentration of thioester, as its concentration is limiting under these reaction conditions. This is a valid, though somewhat unintuitive, perspective, as by numerically integrating the rate equations it is not necessary to make the steady-state approximation. In the above, as the equilibria between Ubc13 and Mms2, as well as Ub binding to free Mms2 are not considered, k_{on} and k_{off} are the on and off-rates for Ub binding to the heterodimer. For all simulated data sets, [E] = 100 µM, [S] = 10 µM, $k_{\text{on,Ub-M:U}} = 2.0 \times 10^7$ s^{-1} (Ub binding to the heterodimer), $k_{\text{on,Ub-U}} = 2.0 \times 10^7$ s^{-1} (Ub binding to Ubc13), and $k_{\text{on,M-U}} = 1.7 \times 10^8$ s^{-1} (Ubc13 binding to Mms2) (3).

References

1. Wolfram, S. (1999) *The Mathematica book*. (Cambridge University Press).

2. Palmer III, A.G., Kroenke, C.D., and Loria, J.P. (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods in Enzymology* 339, 204–238.

3. Markin, C.J., Saltibus, L.F., Kean, M.J., McKay, R.T., Xiao, W., and Spyracopoulos, L. (2010) Catalytic proficiency of ubiquitin conjugation enzymes: Balancing pK_{a} suppression, entropy, and electrostatics. *Journal of the American Chemical Society* 132, 17775–17786.