# Using High-Resolution NMR to Examine the Components that Contribute to the Activity and Stability of Phosphatase of Regenerating Liver (PRL-1) and Their Dependence on the Redox State of Cysteine Residues

By

© 2009
Andria L. Skinner
B.S. Biology and Chemistry, Central Methodist University, 2005
M.S. Pharmaceutical Chemistry, The University of Kansas, 2008

Submitted to the graduate degree program in Pharmaceutical Chemistry and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Committee:_____
Chairperson/Jennifer Laurence

_____
Roberto DeGuzman

_____
C. Russel Middaugh

_____
Christian Schöneich

_____
Mario Rivera

Date Defended:_____

The Dissertation Committee for Andria L. Skinner certifies that this is the approved version of the following dissertation:

# Using High-Resolution NMR to Examine the Components that Contribute to the Activity and Stability of Phosphatase of Regenerating Liver (PRL-1) and Their Dependence on the Redox State of Cysteine Residues

Committee:_____
Chairperson/Jennifer Laurence

_____
Roberto DeGuzman

_____
C. Russell Middaugh

_____
Christian Schöneich

_____
Mario Rivera

Date Approved:_____

As the first graduate student from the lab,

I dedicate this dissertation to my advisor, Dr. Jennifer Laurence.

# ABSTRACT

Phosphatase of regenerating liver (PRL-1) has gained attention as an important therapeutic target because of its involvement in cancer and metastasis, where it promotes proliferation and cell motility. The protein is found at high concentration in numerous cancer tissues and when over-expressed induces cellular transformation. Elevated levels of PRL-1 have been shown to trigger metastasis and to promote cellular transformation and tumorigenesis. Because its enzyme activity contributes to both tumorigenesis and metastasis, this protein is an attractive target for therapeutic intervention. The problem with ubiquitously inhibiting PRL-1 is that this enzyme is also essential to cell survival and tissue regeneration under normal physiological conditions. Inhibition may result in nonspecific and unwanted cell death of healthy tissues. Targeting only the overactive form of an enzyme would provide a reasonable resolution to this problem, but it is currently unknown how PRL-1 prompts proliferation in some cells and differentiation in others. The long-term goal of this project is to identify the crucial structural and dynamic features that support the enzymatic activity of PRL-1 and characterize the different forms to identify ones appropriate to target for therapeutic intervention. In this work, we have shown that the full-length unmodified form of PRL-1 is primarily inactive based on its determined reduction potential (-364.3 ± 1.5 mV) and that *in vivo,* activation is likely conferred through a posttranslational modification at the C-terminus. Using an assortment of biophysical techniques, we have also shown that a mutant form of PRL-1 (PRL-1-C170S/C171S) is reduced and active and can serve as a model for posttranslationally modified PRL-1 at this site.

A second goal of this dissertation project is to use PRL-1 as a model system for studying protein stability. An understanding of the forces that contribute to protein stability is crucial to the development of new biotechnology products with improved half-lives and lower

immunogenicity. Because proteins are held together by numerous weak interactions, understanding the mechanisms by which stabilization is achieved is important to the design of new biotechnology products that better resist unfolding and aggregation. Mechanistic information describing how specific interactions influence stability is lacking, in part because the techniques typically used to study inherent stability do not provide sufficient detail. We used high-resolution solution NMR to examine the details of one aspect of stability in PRL-1. It is well-established that the oxidation state of cysteine residues in proteins are critical to overall physical stability. The presence of disulfide bonds most often imparts thermodynamic stability, and as such, engineered disulfide bonds have become a means for improving the viability of protein therapeutics. In some cases, however, disulfide bonds can diminish stability. PRL-1 has two discernable stable states, one of which contains a disulfide bond (oxidized, inactive) and is slightly less stable than the one that does not (reduced, active). To understand how the reduced form compensates for the loss of the disulfide bond, a series of mutants were analyzed that disrupted disulfide bond formation in distinct ways. The physical stability of each variant is unique with respect to the others, indicating that the stability of the whole protein depends on a combination of local, synergistic interactions with the Cys side chains. High-resolution solution NMR was used to provide site-specific and mechanistic information about how each interaction influences the protein's physical stability. This analysis reveals several key structural components that contribute to PRL-1's overall stability in the absence this disulfide bond. The results provide insight to explain how the reduced protein compensates for the loss of this structural feature, which has implications for its function *in vivo,* and how local instability may lead to aggregation.

# ACKNOWLEDGEMENTS

This dissertation was made possible by NIH grant number P20 RR-017708 from the National Center for Research Resources and the Kansas University Center for Research. This work was additionally supported by fellowships from Amgen, the Edith and Eleta Ernst Cancer Research Fellowship and the Gretta Jean and Gerry D. Goetsch graduate scholarship and continued financial support from the Department of Pharmaceutical Chemistry. This work made use of the KU Biomolecular NMR Core Facility, which is also supported through the KU COBRE Center in Protein Structure and Function by NIH grant number RR-017708. The mass spectrometers used in this study were purchased with support from KSTAR, Kansas administered NSF EPSCoR, the University of Kansas and KCALSI (www.kclifesciences.org). This study also made use of the National Magnetic Resonance Facility at Madison (NMRFAM), which is supported by NIH grants P41RR02301 (BRTP/ NCRR) and P41GM66326 (NIGMS). Additional equipment was purchased with funds from the University of Wisconsin, the NIH (RR02781, RR08438), the NSF (DMB-8415048, OIA-9977486, BIR-9214394), and the USDA. Finally, this work would not be possible if I were not permitted to use many of the analytical instruments in Dr. Russ Middaugh's lab group and so I thank him as well as his lab members who offered training, technical assistance and maintained those instruments.

I would also like to recognize members of the Laurence lab and other collaborators who have contributed to the completion of this work in a significant way. First, I would like to thank Anthony A. Varita, who generated several of the PRL-1 constructs including C49S, Y53E, Y53F and C170S-C171S. Mr. Vartia was a coauthor on the work describing the reduction potential, as he, in collaboration with Todd Williams, characterized the reduced and oxidized forms of PRL-1 by mass spectrometry. Next, I would like to thank Dr. Todd Williams, who also coauthored the

reduction potential study, for his support in collecting and analyzing the mass spectrometry data. Third, I would like to give a special acknowledgement to Dr. Dave Vander Velde and Dr. Asokan Anabanadam for their support, training and maintenance of the NMR spectrometers here at KU. Additional NMR training was received at the National Magnetic Resonance Facility at Madison (NMRFAM) Protein Structure Determination Workshop. The instruction I received here has played a pivotal role in the success of this project and I thank the staff members of that facility who have worked hard to ensure its utility. Finally, I would like to thank Dr. Lisa Crow for traveling to NMRFAM to collect the C49S NMR assignment data and Dr. Klaas Hallenga for his technical assistance with collecting this data set.

Although, only a few former lab members have made specific scientific contributions to this project, the success of my work would not have been feasible without the continued support of all past and current members of the Laurence lab, who have always provided helpful discussions concerning the experimental design and specific conclusions made. Furthermore, I would like to thank my dissertation committee (Drs. Jennifer Laurence, Russ Middaugh, Mario Rivera, Christian Schöneich and Roberto DeGuzman) for their helpful comments and constructive feedback to improve my research and this document as a whole.

On a more personal level, I need to acknowledge the unwavering support of my friends and family, especially my boyfriend Brad Bolton. Their continued encouragement and confidence in my scientific pursuits have been a critical factor to my success at KU and I appreciate them very much.

Finally, I would like to acknowledge my advisor, Dr. Jennifer Laurence. There are truly no words to describe my appreciation for her. Jen’s enthusiasm for scientific learning is contagious and compels me to become a better mentor to the undergraduate and younger

graduate students in our lab. She is a true inspiration on multiple levels and instills a sense of excitement about research, no matter what the outcome. It is professors/mentors/teachers/colleagues like her that will dictate the success of future generations of scientist. It was an honor to be Jen's first graduate student and to play such a pivotal role in shaping the research of her lab. I will always remember and be grateful for the lessons I have learned from her, both professionally and personally speaking. Although my graduate student career is ending, I am thankful that my relationship with my Ph.D. advisor will only continue to grow and mature, a fact that can be attributed to Jen's unique passion and thoughtfulness for her students. For all of these reasons, I love her very much and appreciate everything she has done to ensure my success as a scientist.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

In Alphabetical Order:

| | |
|---|---|
| 3-FABS | 3 fluorine atoms for biochemical screening |
| Abl | abelson tyrosine kinase |
| AChBP | acetylcholine binding protein |
| ADAP | adhesion and degranulating-promoting adapter protein |
| Ala | alanine |
| Ala | alanine |
| Arg | arginine |
| Asn | asparagine |
| Asn | asparagine |
| Asp | aspartic acid |
| ATF-5 | activating transcriptor factor-5 |
| ATP | adenosine triphosphate |
| BME | reduced b-mercaptoethanol |
| BMRB | biological magnetic resonance database |
| CD | circular dichroism |
| Cdc | cell division cycle |
| CML | chronic myelogenous leukemia |
| CRINEPT | cross relaxation-enhance polarization transfer |
| CRIPT | cross relaxation-induced polarization transfer |
| CSA | chemical shift anisotropy |

| | |
|---|---|
| Cys | cysteine |
| Da | daltons |
| DECODES | diffusion encoded spectroscopy |
| DOSY | diffusion ordered spectroscopy |
| dsPTPase | dual-specific protein tyrosine phosphatase |
| DTT | reduced dithiothreitol |
| ESI-MS | electrospray Ionization Mass Spectrometry |
| ExPASY | expert protein analysis system |
| FID | free induction decay |
| Glu | glutamic acid |
| GRX | glutaredoxin |
| GSH | reduced glutathione |
| GSSG | oxidized glutathione; glutathione disulfide |
| HAS | human serum albumin |
| H-bond | hydrogen bond |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| His | histadine |
| HMGB1 | high-mobility group box 1 |
| HSQC | heteronuclear single quantum coherence |
| IDA | iminodiacetic acid |
| Ile | isoleucine |
| ILOES | interligand nuclear overhauser effect |
| IMAC | immobilized metal ion affinity chromatography |

| | |
|---|---|
| INEPT | insensitive nuclei enhanced by polarization transfer |
| INPHARMA | protein mediated interligand NOEs for pharmacophore mapping |
| Kan30 | kanamycin |
| KAP | kidney androgen-related protein |
| kDa | kilodaltons |
| LAR | leukocyte common antigen-related |
| LB | liquid broth |
| Leu | leucine |
| lmwPTPase | low molecular weight protein tyrosine phosphatase |
| MDDR | MDL drug data report |
| Met | methionine |
| MKP | mitogen activated protein kinase phosphatase |
| MW | molecular weight |
| NLS | nuclear localization signal |
| NMR | nuclear magnetic resonance |
| NMR-DOC | nuclear magnetic resonance docking of compounds |
| NMRFAM | National Magnetic Resonance Facility at Madison |
| NOE | nuclear overhauser effect |
| NOESY | NOE spectroscopy |
| oxBME | oxidized b-mercaptoethanol |
| oxDTT | oxidized DTT |
| PCR | polymerase chain reaction |
| PDB | protein data bank |

| | |
|---|---|
| PDI | protein disulfide isomerase |
| Phe | phenylalaline |
| Phe | phenylalanine |
| P-loop | phosphate binding loop |
| pNP | para-nitrophenol/olate |
| pNPP | para-nitrophenyl phosphate |
| PRL-1 | phosphatase of regenerating liver |
| PTK | protein tyrosine kinase |
| PTP1B | protein tyrosine phosphatase-1B |
| PTPase | protein tyrosine phosphatase |
| pTyr | phospho-Tyrosine |
| rf | radio frequency |
| SAR | structure activity relationship |
| SDS-PAGE | sodium dodecylsulfate-polyacrylamide gel elctrophoresis |
| SEA-TROSY | solvent exposed amide-TROSY |
| Ser | serine |
| SH3 | Src homology domain 3 |
| SHP-1 | SH2-containing phosphatase-1 |
| SLS | static light scattering |
| SOLVE | structurally oriented library valency engineering |
| SOS-NMR | structural information using overhauser effects and selective labeling |
| STD | saturation transfer difference |
| SUMO | small ubiquitin-like modifier |

| | |
|---|---|
| Thr | threonine |
| Tris-HCl | tris(hydroxymethyl)aminomethane hydrochloride |
| TROSY | transverse relaxation optimized spectroscopy |
| Trp | tryptophan |
| TRX | thioredoxin |
| tsPTPase | tyrosine-specific protein tyrosine phosphatase |
| Tyr | tyrosine |
| Val | valine |
| Val | valine |
| VHR | Vaccinia H1-related |
| WaterLOGSY | water-ligand observed via gradient spectroscopy |
| XIAP | X-linked inhibitor of apoptosis protein |

Page Left Intentionally Blank

# CHAPTER 1. INTRODUCTION

## 1.1. PRL-1 AS A NOVEL THERAPEUTIC TARGET

In drug discovery research, a drug "target" is a molecule for which there is a reasonable amount of understanding of how it functions in both normal physiology and a diseased state. Upon identification of a target, a lengthy process ensues to validate the significance of the target with respect to the disorder or disease, which is often one of the most critical steps in the development of new therapeutics.[1] Target validation requires more than identifying its expression at the genetic level because many genes often produce different protein isoforms, which can have subtly different functions, or encode posttranslational modification motifs that also give rise to variation in protein function. The ultimate goal is to identify a potent molecule (usually small in size) that modifies the actions of the target and promotes restoration of normal cellular physiology. High-throughput screening experiments rely on the use of large libraries of small molecule candidate compounds that may bind to and modify the activity of the target, most commonly a protein, thereby assessing the "druggability" of a target.

Not all targets can be deemed druggable. Traditionally, classification of a protein as a druggable target is based on the fulfillment of three requirements: 1) the protein must have the ability to bind a small molecule with the appropriate "rule-of-five"* properties, 2) bind to the molecule with reasonable potency and 3) be disease modifying. Using this definition, it is suggested that between 600 and 3000 genes in the human genome code for potential druggable

*The so-called "rule-of-five" predicts that poor absorption or permeation is more likely when 1) there are more than five hydrogen-bond (H-bond) donors, 2) there are more than ten H-bond acceptors 3) the molecular weight (MW) is greater than 500 Daltons (Da), 4) the calculated Log P is greater than 5 (indicates increased lipophilicity) and 5) compound classes that are substrates for biological transporters are exceptions to the rule.[2]

targets.[3] This definition does not, however, factor in specificity. Just because a small molecule drug with good physicochemical properties can be developed that modulates the activity of a protein target does not mean specificity can be achieved. Proteins of the same genomic family usually have functional similarities and conserved structural features within the binding-site, suggesting that if one member of the family were able to bind a drug, other members would also be able to bind the same compound. Furthermore, a single protein may regulate multiple pathways and inhibition or increased activation of the protein may lead to unwanted side effects. In this context structural characterization is an important facet of target validation experiments because the three-dimensional structure can permit identification of both allosteric and active binding sites, which provides further information about the biological role the target plays and facilitates structure based drug design.

This problem of specificity is exemplified by the protein tyrosine kinase (PTK) family of enzymes, which have been studied and targeted for drug design for many years.[4] PTKs play important roles in virtually all aspects of the cellular physiology and malfunction or deletion of these enzymes can lead to diseased states. Nearly all PTK inhibitors function by competing for the highly conserved adenosine triphosphate (ATP) binding site *in vivo*. Specificity, in part, comes from the fact that only a portion of the drug interacts with the residues that bind to ATP, while the rest of the molecule makes contacts with other residues situated outside the ATP-binding pocket. Although a reasonable degree of specificity among related proteins with as much as 60% sequence identity can be achieved with this approach, it has been shown that even more distantly related proteins with less homology can still interact with the same inhibitor.[4] Furthermore, many kinases will alter the phosphorylation state of many protein substrates, which may lead to other undesirable outcomes. A better approach is to target a modified or unique form

of the protein involved in the promotion of a disease but not in normal cell function. This was successfully demonstrated with Gleevaec, a drug developed using structure-based rational design to solely inhibit the aberrant form of Abl kinase and effectively treat chromic mylogenous leukemia (CML).[5,6]

In this work, a similar approach has been initiated to facilitate the design of a specific inhibitor to phosphatase of regenerating liver (PRL-1). PRL-1 is a unique protein tyrosine phosphatase (PTPases) that, together with the opposing actions of PTKs, plays an important role in maintaining appropriate tyrosine phosphorylation levels in the cell during development and tissue regeneration.[7] PRLs are normally expressed at low levels in most mature cells, while higher expression levels are observed during embryonic development[8] and during cellular proliferation[9-13] or differentiation[9,14-16] depending on the cell type. PRL-1 has gained attention as an important therapeutic target because of its involvement in cancer and metastasis, where it promotes proliferation and cell motility when *active.*[17,18] Because its enzyme activity contributes to both tumorigenesis and metastasis, this protein is an attractive target for therapeutic intervention. The problem with ubiquitously inhibiting PRL-1 is that this enzyme is also essential to cell survival and tissue regeneration under normal physiological conditions. Therefore, its complete inhibition may result in nonspecific, unwanted cell death of healthy tissues. Targeting only the overactive form of the enzyme would provide a reasonable resolution to this problem, but it is currently unknown how PRL-1 prompts proliferation in some cells and differentiation in others. Our working hypothesis is that aberrant modification of PRL-1 results in misregulation of phosphatase activity, thereby promoting proliferation and/or motility of transformed cells. Posttranslational modifications may affect the activity of PRL-1 both directly, by altering the conformation of the protein to better prime the active site, and indirectly by altering its

subcellular location where it may act on different downstream effectors. A cartoon of PRL-1 is shown in Figure 1.1, which summarizes known and/or predicted sites of modification. The primary aim of my doctoral research project was to characterize the unmodified, full-length protein and understand how its C-terminal region contributes to the unique regulation of the catalytic activity of PRL-1. The research results described in this dissertation also provide a basis for subsequent development of potential therapeutics by either high-throughput screening or structure-based design using solution NMR.

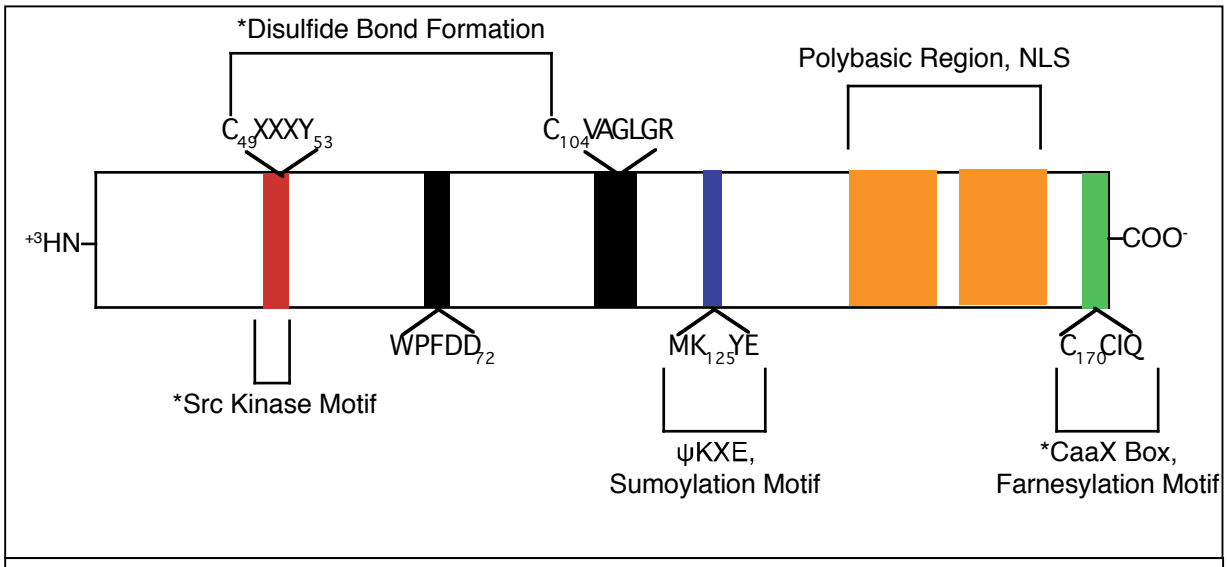



**Figure 1.1. Cartoon of PRL-1**. The potential posttranslational modifications that may regulate PRL-1 are shown with respect to the active site ($C_{104}$VAGLGR) and WPFDD loop (see section 2.4). Those modifications that were investigated in this work are designated with a *.

## 1.2. PRL-1 AS A MODEL PROTEIN TO STUDY PROTEIN STABILITY

Commercial pressure often forces the pharmaceutical industry to direct their efforts towards development of small molecule drugs. Small compounds are far less expensive to make and are often much more stable than protein therapeutics but it can be quite difficult to obtain the desired specificity from a small molecule drug. Proteins offer a very high degree of selectivity in

their mode of action, and as such, these macromolecular molecules are also being pursued as potential drug candidates. Macromolecular drugs are quite diverse and range from peptides and proteins to nucleic acids and vaccines. Protein drugs often include monoclonal antibodies and recombinant proteins like insulin or human growth hormone that target the cell surface or extracellular sites.

Protein molecules do have disadvantages when it comes to their use as therapeutics. Due to the inherent complexity of the molecule, they are often far less stable and susceptible to many pathways of both chemical and physical degradation, often leading to immunogenic responses *in vivo*.[19] It follows that they are far more expensive to make and difficult to formulate than small molecules. Because of these problems, an understanding of the forces that contribute to protein stability is crucial to the development of new biotechnology products having improved half-lives and lower immunogenicity. It is often too expensive and too lengthy a process to conduct such research on viable protein drug candidates because of the need to rapidly identify stabilizing conditions for formulation purposes. In this context, basic research plays a pivotal role in the drug development process. Often times, model proteins (a novel protein target and/or a ubiquitously expressed cell signaling protein) that are well characterized at the molecular level can be used to answer fundamental questions regarding protein stability. Such questions can be conducted much faster with model proteins because the three-dimensional structure has often already been determined and the basic set of assays to produce and characterize the protein are usually known. This approach is used in the present work in which a novel drug target (PRL-1) was used to determine how stabilization or destabilization was achieved in various protein forms with or without a disulfide bond present. PRL-1 serves as an excellent model for this purpose because the information described in the following chapters indicate that this protein strongly

favors the oxidized state *in vitro*, but equilibrium control of the redox state is easily modulated using a redox buffer system to generate the reduced form. Second, a major conformational change occurs upon reduction of the protein, providing well-resolved, site-specific measurable differences between the two states using solution NMR spectroscopy. Additionally, the structures of both reduced and oxidized PRL-1 have been determined previously using X-ray crystallography, which facilitates interpretation of the changes observed in the NMR data with respect to the three-dimensional structure.[20,21]

## 1.3. OUTLINE

The following chapters describe the assays used to characterize full-length PRL-1 in both inactive and active states. Chapter 2 is used to describe the identification of a constitutively reduced and active form of PRL-1 to use as a positive control, followed by Chapter 3, which describes the methods used to obtain sequential backbone NMR assignments of reduced and active PRL-1 and the results therein. Chapter 4 is geared towards quantifying the propensity of PRL-1 to be modulated by redox conditions. The conclusions presented in this section provide the first evidence of how direct modification of PRL-1 can lead to both direct and indirect changes in its activity through structural changes and/or changes in the subcellular location. In Chapter 5, the focus is shifted from understanding the biology of PRL-1 to its stability. In this work, PRL-1 was used as a model to determine how modulation of a disulfide bond in different ways leads to increased or decreased stability and to elucidate the mechanism by which stabilization or destabilization was achieved in the various protein forms. The results of this investigation have provided key insight into how disulfide bonds stabilize proteins in general, which is important information that can be applied to the design of future protein therapeutics

with better physicochemical properties. The final chapter is largely discussion based, as the implications of the findings in this dissertation with respect to the larger goals of this project are considered as well as preliminary data presented that provides a foundation for subsequent studies in the Laurence lab. Two appendices are included in this document to facilitate the understanding of the various methods used. Appendix A focuses on the *in vitro* activity assays used. Appendix B provides a sequence alignment of the PRL family members. Appendix C is a comprehensive summary of NMR methods that can be used to facilitate drug design. Appendix D provides supplementary data for Chapter 5 and Appendix E summarizes a new method for purifying PRL-1 that can potentially be used for preparing samples for *in vitro* posttranslational modification assays.

## 1.4. REFERENCES


1. Szymkowski D. (2003) Drug target validation: Hitting the target. *Nature*, 422, 341-347.
2. Lipinsky C. A., Lombardo F., Dominy B. W. & Feeney P. J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*, 23, 3-25.
3. Hopkins A. L. & Groom C. R. (2002) The druggable genome. *Nat Rev Drug Discov*, 1, 727-730.
4. Cohen P. (1999) The development and therapeutic potential of protein kinase inhibitors. *Curr Opin Cell Biol*, 3, 459-465.
5. Deininger M. W. N. & Druker B. J. (2003) Specific targeted therapy of chronic myelogenous lukemia with Imatinib. *Pharmacol Rev*, 55, 401-423.
6. Druker B. J. & Lydon N. B. (2000) Lessons learned from the development of an Abl tyrosine kinase inhibitor for Chronic Myelogenous Lukemia. *J Clin Invest*, 105, 3-7.
7. Chiarugi P. & Buricchi F. (2007) Protein tyrosine phosphorylation and reversible oxidation: two cross-talking posttranslational modifications. *Antiox Redox Signal*, 9, 1-24.
8. Rundle C. H. & Kappen C. (1999) Developmental expression of the murine Prl-1 protein tyrosine phosphatase gene. *J Exp Zoo*, 283, 612-617.
9. Diamond R. H., Peters C., Jung S. P., Greenbaum L. E., Haber B. A., Silberg D. G., Traber P. G. & Taub R. (1996) Expression of PRL-1 nuclear PTPase is associated with proliferation in liver but with differentiation in intestine. *Am J Physiol Gastrointest Liver Physiol*, 271, G121-G129.
10. Zeng Q., Si X., Horstmann H., Xu Y., Hong W. & Pallen C. J. (2000) Prenylation-dependent association of protein-tyrosine phosphatases PRL-1, -2, and -3 with the plasma membrane and the early endosome. *J Biol Chem*, 275, 21444-21452.
11. Werner S. R., Lee P. A., DeCamp M. W., Crowell D. N., Randall S. K. & Crowell P. L. (2003) Enhanced cell cycle progression and down regulation of p21Cip1/Waf1 by PRL tyrosine phosphatases. *Cancer Lett*, 202, 201-211.
12. Sun J.-P., Luo Y., Yu X., Wang W.-Q., Zhou B., Liang F. & Zhang Z.-Y. (2007) Phosphatase activity, trimerization, and the C-terminal polybasic region are all required for PRL1-mediated cell growth and migration. *J Biol Chem*, 282, 29043-29051.
13. Wang J., Kirby C. E. & Herbst R. (2002) The tyrosine phosphatase PRL-1 localizes to the the endoplasmic reticulum and the mitotic spindle and is required for normal mitosis. *J Biol Chem*, 277, 46659-46668.
14. Kong W., Swain G. P., Li S. & Diamond R. H. (2000) PRL-1 PTPase expression is developmentally regulated with tissue-specific patterns in epithelial tissues. *Am J Physiol Gastrointest Liver Physiol*, 279, 613-621.
15. Yarovinsky T. O., Rickman D. W., Diamond R. H., Taub R., Hageman G. S. & Rickman C. B. (2000) Expression of the protein tyrosine phosphatase, phosphatase of regenerating liver 1, in the outer segments of primate cone photoreceptors. *Mol Brain Res*, 77, 95-103.
16. Takano S., Fukuyama H., Fukumoto M., Kimura J., Xue J.-H., Ohashi H. & Fujita J. (1996) PRL-1, a protein tyrosine phosphatase, is expressed in neurons and oligodendrocytes in the brain and induced in the cerebral cortex following transient forebrain ischemia. *Mol Brain Res*, 40, 105-115.

17. Stephens B. J., Han H., Gokhale V. & Von Hoft D. D. (2005) PRL phosphatases as potential molecular targets in cancer. *Mol Cancer Ther*, 4, 1653-1661.
18. Bessette D. C., Qiu D. & Pallen C. J. (2008) PRL PTPs: mediators and markers of cancer progression. *Cancer Metastasis Rev*, 27, 231-252.
19. Volkin D. B., Mach H. & Middaugh C. R. (1997) Degradative covalent reactions important to protein stability. *Mol Biotechnol*, 8, 105-122.
20. Sun J.-P., Wang W.-Q., Yang H., Liu S., Liang F., Fedorov A. A., Almo S. C. & Zhang Z.-Y. (2005) Structure and biochemical properties of PRL-1, a phosphatase implicated in cell growth, differentiation, and tumor invasion. *Biochemistry*, 44, 12009-12021.
21. Jeong D. G., Kim S. J., Kim J. H., Son J. H., Park M. R., Lim S. M., Yoon T.-S. & Ryu S. E. (2005) Trimeric structure of PRL-1 phosphatase reveals an active enzyme conformation and regulation mechanisms. *J Mol Biol*, 345, 401-413.

Page Left Intentionally Blank

# CHAPTER 2.

# IDENTIFYING A STABLE, REDUCED AND ACTIVE FORM OF PRL-1

## 2.1. INTRODUCTION

### 2.1.1. PRL-1 Belongs to the PTPase Family of Enzymes

Tyrosine phosphorylation (pTyr) is a major protein posttranslational modification that cells utilize in signal transduction to regulate numerous cellular functions including (but not limited to) cell growth, differentiation, cell migration and apoptosis. The opposing activities of protein tyrosine kinases (PTKs), which promote phosphorylation, and protein tyrosine phosphatases (PTPases), which promote dephosphorylation, govern cellular pTyr levels in a reversible and dynamic manner to maintain homeostasis, and imbalances can lead to various diseased states.[2,3] PRL-1 belongs to the PTPase family of enzymes and plays a critical role in normal proliferation during development and regeneration of damaged tissues. Although the biochemical pathways by which PRL-1 functions remain to be determined, mutagenesis studies have shown that the phosphatase activity of PRL-1 is required for normal mitotic progression, cellular proliferation, enhanced cellular motility and metastatic cancer formation, indicating that the active form of PRL-1 is the relevant target for development of anticancer therapeutics.[4-11]

Although members of the PTPase family share an identical catalytic mechanism, the family is subclassified into three groups based on their structure or sequence homology and substrate preference: tyrosine-specific (ts), dual-specific (ds) and low molecular weight (lmw) PTPases. tsPTPases (such as PTP1B) strictly target pTyr substrates, while the dsPTPases can catalyze the hydrolysis of pSer and pThr as well as pTyr. The lmw and dsPTPases display little amino acid sequence identity with classic PTPases beyond the conserved residues in the active

site.[12,13] Although the three subfamilies have little sequence similarity and display differences in substrate specificity, the three-dimensional structures of PTPase catalytic domains have the same general topology with a parallel β-sheet in the central portion of the protein bordered by α-helices on both sides (Figure 2.1). The active site is located within a crevice on the protein surface, which is much deeper for the tsPTPases than for the dsPTPases. The deeper active site pocket in tsPTPases permits selectivity towards pTyr containing substrates.[13,14]

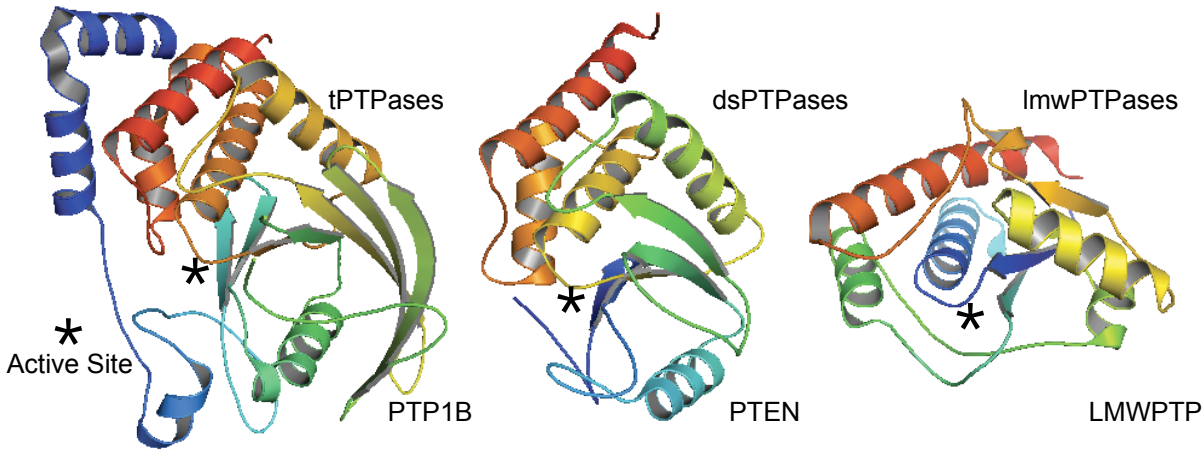
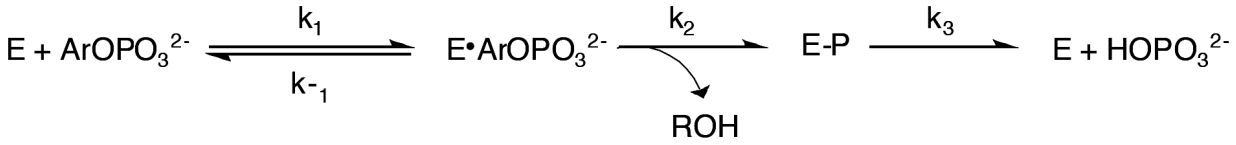



**Figure 2.1. Subclasses of the PTPase Family Members.** A representative member of each PTPase subfamily is presented. The PDB files 2B4S, 1D5R and 1XWW were used to generate the illustrations of PTP1B, PTEN and LMWPTP, respectively. The active site is indicated with a *.[1-3]

The unique feature that is characteristic of all PTPases is the $CX_5R$ (single-letter code for amino acids) signature motif in the active site, where the X represents any amino acid. All PTPases dephosphorylate their pTyr substrates using a common two-step mechanism. Unlike alkaline phosphatases and Ser/Thr phosphatases, this reaction requires no metal ions and is nucleophilic in nature.[12] First, the conserved Cys initiates the catalytic reaction by attacking the phosphorous on the incoming substrate, while the phosphate binding loop (P-loop) and conserved Arg residue bind and position the phosphate group (Figure 2.2). The thiolate of the conserved Cys is utilized as an attacking nucleophile to form a cysteinyl-phosphate enzyme

**2.2.A**




**2.2.B**

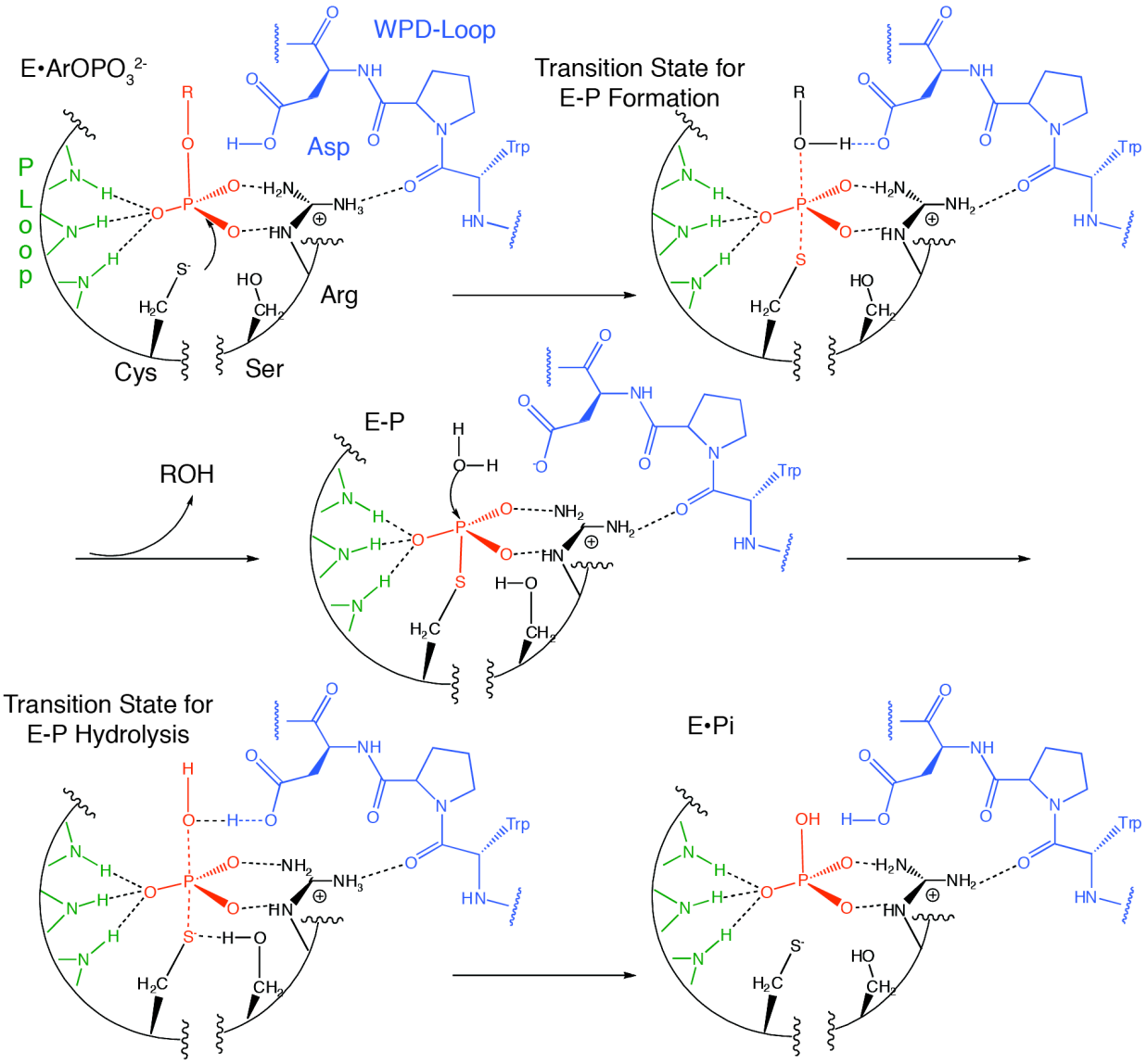



**Figure 2.2. Classic PTPase Catalytic Mechanism. A. Simplified PTPase Reaction Scheme. B. Mechanism of Catalysis.** Although the nature of the chemical reaction transition states is not discussed here, an illustration for each step was included for reference. For a detailed account of how PTPases achieve transition state stabilization, the reader is directed to an excellent review.[12] This figure was created based on Figure 3 from reference 12, with modifications. It should be noted that significant differences between what is illustrated here and the mechanism by which PRL-1 functions exist. The details are discussed in section 2.4.1.

intermediate (E-P), and substitution of the conserved Cys completely abrogates enzyme activity.[12,15] In the second step, enzyme dephosphorylation is accomplished by hydrolysis. A sequentially distal but spatially proximal Asp residue facilitates both steps of catalysis by serving as a general acid during E-P formation, transferring a proton to the leaving tyrosine group, and by serving as a general base during hydrolysis that abstracts a proton from the nucleophilic water molecule (Figure 2.2). Mutation of the Asp residue results in a substrate-trapping enzyme because release of the leaving group is dramatically slowed.[16-19] The conserved Asp residue is located in a flexible loop known as the WPD-loop, which undergoes a major conformational change upon substrate binding.[20,21] In unligated PTPases (closed conformation), the conserved Asp moves as much as 20 Å away from the active site. Previous studies have shown that the invariant Trp in the WPD-loop is an important hinge residue that hydrogen bonds to the conserved Arg to correctly position it in the active site for substrate binding in the open, active conformation. Mutation of the Trp residue completely disables general acid catalysis and diminishes the affinity of the active site for substrate.[18] A conserved Ser or Thr is also often found immediately following the active site motif ($CX_5RS/T$) and mutation of this residue dramatically slows catalysis. The hydroxyl group of this residue facilitates catalysis by stabilizing the negative charge on the leaving thiolate group during E-P hydrolysis.[12,22,23] The PTPase catalytic mechanism has been extensively studied and readers are directed to some excellent reviews for a more detailed description of this information.[12,13]

**2.1.2. Oxidation of PTPases**

Although the PTPase family members display a range of regulatory mechanisms that are specific for each subclass, they do share one common mode of inactivation: reversible oxidation of the catalytic Cys residue, which has been shown to occur *in vivo*.[24,25] Unlike most Cys

residues in proteins that have –SH $pK_a$ values near eight, the side chain of the PTPase catalytic Cys is typically deprotonated ($pK_a \approx 3\text{-}6$) due to the partially positive microenvironment of the active site, making it extremely reactive at physiological pH.[26] The exceptional reactivity of the thiol group is required for enzyme activity but also renders the catalytic Cys highly susceptible to oxidation.[27-29] Members of each PTPase subfamily have previously been shown to be oxidized by treatment with $H_2O_2$ including the tsPTPases, PTP1B and LAR (leukocyte antigen-

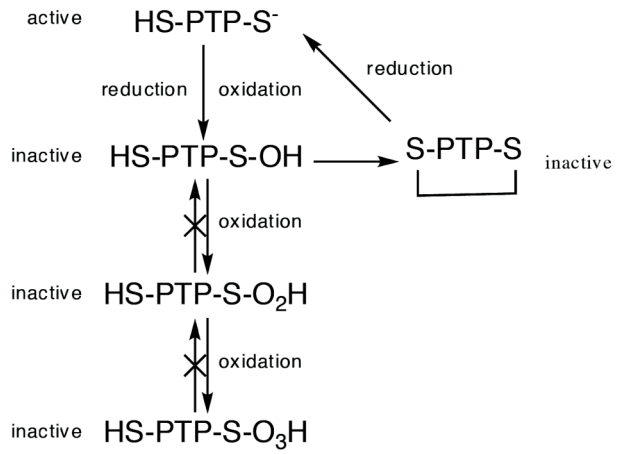



**Figure 2.3. PTPase Oxidation Scheme.** The PTPase catalytic thiolate group can react with reactive oxygen species to form a sulfenic acid intermediate. The addition of one or two more oxygen atoms leads to the formation of sulfinic or sulfonic acid, which are irreversible modifications. On the other hand, the sulfenic acid intermediate can also react with nearby –SH side chains to form a disulfide bond. Disulfide bonds can be re-reduced to obtain the originally active enzyme.

related),[28,30] human LMWPTP,[31] and dsPTPases, VHR (vaccinia H1-related), PTEN and Cdc45.[28,32,33] Oxidation of the catalytic Cys occurs via the formation of sulfenic acid, which inhibits catalysis because the reaction is mediated by the actions of a nucleophilic thiolate side chain.[28] Oxidation of Cys residues to sulfenic acid is reversible upon incubation with thiol compounds. Extreme oxidizing conditions can lead to further oxidation of Cys residues to form sulfinic or sulfonic acid, both of which are irreversible (Figure 2.3).[34] The mechanism common to all PTPases is to prevent further and irreversible oxidation by forming an S-S bridge containing the catalytic Cys. In the case of PTP1B, the sulfenyl-amide intermediate that forms upon oxidation can be glutathionylated to form a mixed disulfide.[35-37] In other cases, intramolecular disulfide bond formation can occur between two Cys residues located within the

active site, as for LMWPTP,[31] or between two distant Cys residues, as for Cdc25,[33] PTEN[32] and SHP-1.[38] This process is important because disulfide bond formation inhibits PTPase activity but protects itself from irreversible oxidation, permitting PTPase reactivation under favorable conditions. Most importantly, the differences in the $pK_a$ values of the catalytic Cys[26] and the structural distance between the two Cys residues will dramatically influence the propensity of an individual PTPase to be rapidly oxidized and regulated by transient changes in intracellular redox conditions.

The previously published crystal structure of wild-type PRL-1 reveals that a disulfide bond forms between the active site Cys (C104) and nearby partner C49, which was subsequently confirmed by Trypsin digestion followed by Mass Spectrometry analysis. Disulfide bond formation is accompanied by substantial changes in the conformation of the catalytically important residues that inhibit the ability of PRL-1 to bind substrate, and the formation of the -S-S- bridge directly prevents formation of the phosphoenzyme intermediate.[39,40] This indicates the *reduced* form of the PRL-1 protein is the relevant target for drug development. The chemically reduced wild-type protein (PRL-1 in the presence of a reducing agent) is not a good candidate for this purpose for several reasons. First, using the chemically reduced wild type may lead to a higher number of false positives or negatives in drug screening assays because 1) the reducing agent transiently interacts with the active-site residues of PRL-1, which may interfere with the binding of molecules of modest affinity, and/or 2) the reducing agent could interact (either chemically or physically) with other chemicals in the screening library. Furthermore, many of the commercially available reducing agents like dithiothreitol (DTT), β-mercaptoethanol (BME) and reduced glutathione (GSH) are readily susceptible to oxidation to form the oxidized counterparts (oxDTT, oxBME and GSSG), the formation of which limits the time during which

the protein remains active. Finally, PRL-1 may require a posttranslational modification *in vivo* for activation, and as such, the chemically reduced wild type would be dramatically different from the biologically relevant form. It follows that in order to conduct biologically relevant *in vitro* drug screening assays, identification of a stable, constitutively active (reduced) form of PRL-1 is needed, which is the subject of the current study.

Several approaches were used to generate a constitutively reduced mimic of PRL-1, including direct modulation of the disulfide bond via mutagenesis. Mutations to C104 cannot be used because a change at this position renders the protein inactive,[5] whereas mutation of C49 would abolish disulfide bond formation but preserve the catalytic thiolate group essential for activity. Additionally, the residues at the predicted Src kinase phosphorylation site (Y53) and residues at the predicted farnesylation site were all tested for their indirect affect on disulfide bond formation (see Figure 1.1). In this approach, protein engineering was used to mutate residues of the respective posttranslational modification sites instead of *in vitro* modification because these reactions can be expensive and practically very difficult to complete. A summary of the findings is described in the following sections.

## 2.2. METHODS

### 2.2.1. Protein Expression and Purification

The PRL-1 gene was cloned into the pET-30 Xa LIC vector as described previously.[41] The PRL-1 wild-type DNA was mutated to C104S, C49S and Y53E using the PCR-based Quickchange method. The primers used to generate these mutants were 5'-ggttgtattgctgttcatTCCgttgcaggccttgg-3', 5'- ggagttaccacaatagtaagagtatGtgaagcaacttatgacac-3' and 5'-gagtatgtgaagcaactGaAgacactactcttgtgg-3', respectively, with capital letters indicating the

**Table 2.1 Purpose of PRL-1 Mutants**

| PRL-1 Variant | Purpose |
|---|---|
| Wild Type | Permits chemical modulation of the disulfide bond with reducing agents |
| C104S | Abolishes disulfide bond formation; inactive |
| C49S | Abolishes disulfide bond formation; catalytic Cys preserved |
| Y53E | Mimics phosphorylation of Y53 |
| C170S-C171S | Mimics modification of the CaaX box at C-terminus |

mutated base. The C170S-C171S double mutant was created using two rounds of mutagenesis with the primers for the C170S and C171S mutations, 5'-ggtcatagaaacaactCttgcat tcaataaggatc-3' and 5'-ggtcatagaaacaa ctgttCcattcaataaggctgtaactc-3', respectively (Integrated DNA Technologies, Coralville, IA). PCR reactions were treated with DpnI (Promega) for 1.5 hours at 37°C, directly transformed into NovaBlue GigaSingles Compentent cells (Novagen) and spread on LB agar plates containing 30 mg/mL kanamycin (Kan30) to select for transformed cells. Individual colonies were cultured overnight in LB Kan30 at 37 °C. The resultant DNA was purified using the Wizard Plus MiniPrep System (Promega). The mutated DNA was confirmed to have the correct sequence by bidirectional sequencing using the T7 promoter and T7 terminator primers (Northwoods DNA, Inc., Bemidjii, MN). The plasmid was transformed into BL21 (DE3) competent cells and selected based on Kan30 resistance. The PRL-1 variants were expressed and purified using the same procedure as described previously for the wild-type protein.[41] All solutions used in the purification and analysis of samples involved in this study were sparged with Ar gas to minimize oxidation derived from $O_2$. Purity of the final protein samples was assessed by SDS-PAGE and protein concentrations were determined by UV absorbance at 280 nm using an extinction coefficient of 19420 $M^{-1}$ $cm^{-1}$. The extinction coefficient was calculated based on the reduced protein using the ProtParam program from ExPASy.[42,43] Final samples were concentrated to 30 mg/mL and stored

at 4 °C in 50 mM HEPES, pH 7.5 until analysis. Table 2.1 summarizes the purpose of studying each mutant.

### 2.2.2. SDS-PAGE

Protein samples were diluted to a final concentration of 1 mg/mL in 50 mM sodium phosphate, pH 7.5, with 100 mM NaCl and were incubated for 1 hour with varying amounts of reduced or oxidized DTT. Oxidized DTT was obtained by incubating $H_2O_2$ with DTT at a 3:1 concentration ratio ($H_2O_2$:DTT) and was confirmed to be complete by monitoring the absorbance at 283 nm.[44,45] After incubation, non-reducing SDS-PAGE loading buffer was added. Samples were then heated for 6 minutes and finally run on a 12% acrylamide gel. Gels were stained using Coomassie (R-250) and the protein bands were used to assess the relative amounts of reduced or oxidized PRL-1. Because $H_2O_2$ also oxidizes Met residues rapidly, oxidized DTT was used to study thiol-specific chemistry changes.

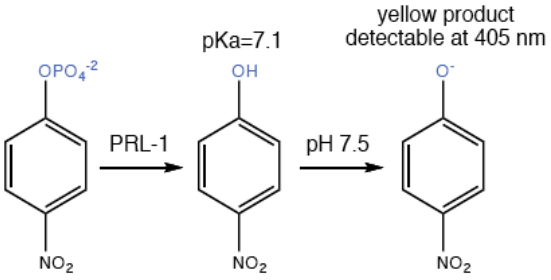



**Figure 2.4. PRL-1/*p*NPP Activity Assay Reaction Scheme.**

### 2.2.3. PTPase Activity Assays

A standard phosphatase activity assay with *p*-nitrophenyl phosphate (*p*NPP) was used to determine the activity of PRL-1 variants according to the scheme illustrated in Figure 2.4.[46] Attempts to fit the activity of PRL-1 to the Michaelis-Menten rate equation were unsuccessful. The details of the work to delineate the reason for this observation are outlined in Appendix A. Instead, the relative activity of PRL-1 at one substrate concentration was evaluated using assay mixtures that contained 50 mM HEPES, pH 7.5, 20 mM freshly prepared *p*NPP and 0-100 mM freshly prepared DTT. Reactions were initiated by the addition of PRL-1 at a final concentration of 1 mg/mL and monitored for

absorption at 405 nm and 37 °C using a Cary 100 UV-Vis spectrophotometer with temperature controller. Reactions were run in triplicate and on a minimum of two independently produced protein samples, totaling at least n=6 replicates. Relative activity was determined at 60 minutes by calculating the amount of *p*-nitrophenolate (*p*NP) produced ($\varepsilon$=1.78x10$^{-4}$ M$^{-1}$·cm$^{-1}$).[47]

## 2.3. RESULTS

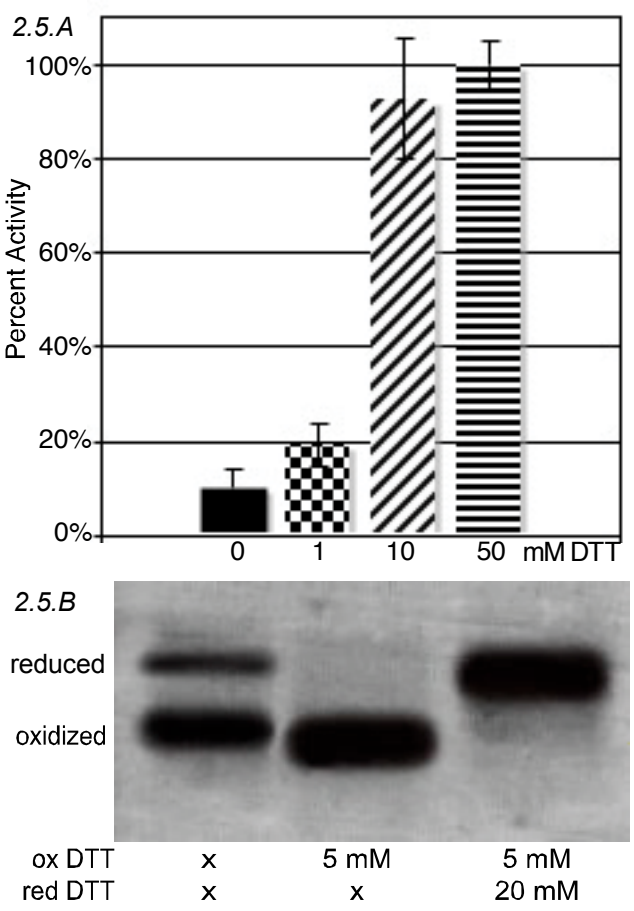



**Figure 2.5. PRL-1 WT Activity A.** Data are displayed as percent maximum activity (100%) based on the average of reduced PRL-1-C170S-C171S samples, following subtraction of the blank involving the inactive PRL-1-C104S mutant. No increase in activity was detected when 100 and 200 mM reduced DTT were added (data not shown). The average of the samples is shown in the figure and error bars represent one standard deviation of the mean. **B. SDS-PAGE of Wild-type PRL-1.** The higher intensity of the lower band indicates that untreated PRL-1-WT is primarily oxidized.

### 2.3.1. Nonreduced Full-length PRL-1 Wild Type is Primarily Inactive

The *p*NPP activity assay performed on non-reduced, purified wild-type protein revealed that 90% of the protein is inactive under non-reducing conditions. As shown in Figure 2.5.A,

PRL-1 requires the addition of a thiol reducing agent for activity. Maximal activity is reached when 10 mM DTT is added to the solution of 1 mg/mL PRL-1. An eight-fold increase in activity is observed when 10 mM DTT is added to the assay mixture. As a negative control, the PRL-1-C104S mutant was constructed. This mutant is completely inactive because the nucleophilic thiol, which is essential for activity, has been removed. Under reducing conditions our activity data is comparable to the previously published kinetic constants, where truncated forms of PRL-1 were found to turn over generic PTPase substrates quite slowly.[21,39]

Because activity was dramatically affected by the addition of reducing agents, we hypothesized that the lack of signal observed for the untreated PRL-1-WT was due to disulfide bond formation at the active site. To test this, we ran untreated PRL-1-WT on SDS-PAGE under nonreducing conditions. Disulfide bond formation can be visualized using nonreducing SDS-PAGE. Oxidized protein, containing a disulfide bond, migrates faster than predicted because the structure is more compact. When left untreated, wild-type PRL-1 runs as two bands on an SDS-PAGE gel, just below the 20-kDa marker (Figure 2.5.B, lane 1). For wild-type PRL-1, the lower band is much more intense, corresponding to a high degree of oxidation. When run under oxidizing conditions, only the lower band is detected (Figure 2.5.B, lane 2). The addition of four-fold excess of reduced DTT to the oxidized sample restores the upper band or reduced species (Figure 2.5.B, lane 3). As a control, the C104S mutant was also examined. At 50 mM oxidized DTT, no band corresponding to the 20 kDa oxidized species was observed (data not shown), indicating the band at 20 kDa reflects formation of a disulfide bond between C104 and C49. This is consistent with the findings of Sun *et al.* (2005), who examined both the C104S and C49S mutant by SDS-PAGE.[39] These studies by Sun *et al.*, however, were conducted on truncated PRL-1 (1-160) and chemical oxidation was required to generate the lower band. As such, direct

confirmation that the presence of the C-terminus does not alter the number or position of disulfide bonds is needed. This is reported in Chapter 4 (4.3.1) but indicates that the bond between C104 and C49 is the only disulfide that forms in PRL-1.

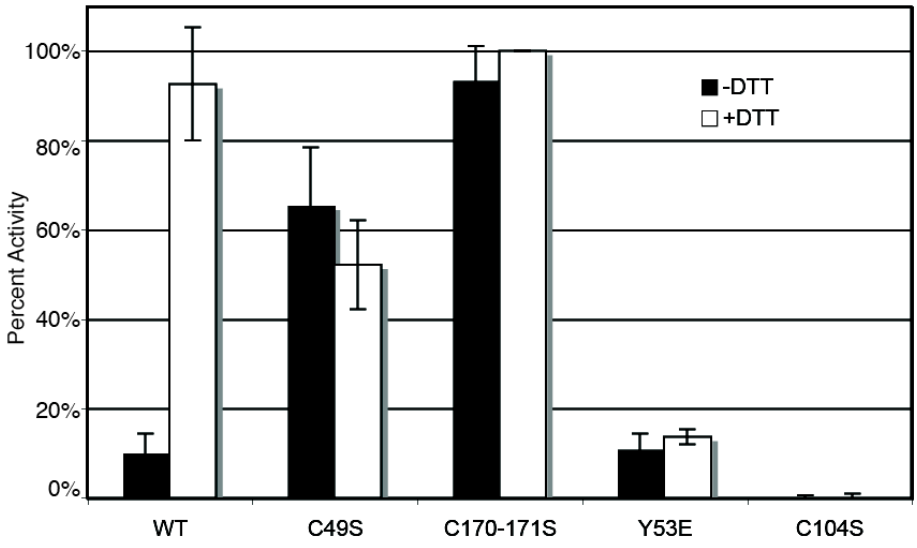



**Figure 2.6. Activity of PRL-1 Mutants**. The data are expressed as the % activity of the C170S-C171S mutant after subtraction of the C104S inactive mutant. C170S-C171S is constitutively active and does not require the addition of reducing agents. Averages of the individual runs are shown and error bars indicate one standard deviation from the mean.

### 2.3.2. PRL-1-C170S-C171S is Reduced and Active

We next began examining the effects of various mutations on the activity of PRL-1. The purpose of each mutant studied is outlined in Table 2.1 As indicated by Figure 2.6, the activity of PRL-1-C49S is slightly greater than the oxidized wild type, but markedly less than the reduced wild type. Only modest activity was observed with the Y53 mutants. In both cases, the activity was approximately nine-fold lower than the reduced wild type but only five-fold higher than the C104S negative control. In stark contrast to untreated wild-type PRL-1, the untreated C170S-C171S mutant exhibits an eight-fold increase in activity. The activity of PRL-1-C170S-C171S is

statistically identical to wild-type PRL-1 under reducing conditions, and the activity of this mutant is not modulated by the addition of reducing agents.

## 2.4. DISCUSSION

The crystal structure of PRL-1 in complex with sulfate is shown in Figure 2.7.A. The secondary structural features and overall topology of PRL-1 are most similar to those of the dsPTPase category. Overall, PRL enzymes share low sequence identity with other dsPTPases (<30%) but have the closest structural homology to VHR, Cdc14, MKP (mitogen-activated protein kinase phosphatase), KAP (kidney androgen-regulated protein) and PTEN.[39,40] Both the catalytic Cys and Asp are present in PRL-1, and the P-loop parallels other dsPTPases in sequence conservation. The so-called WPD-loop of PRL-1 is divergent and the catalytic Asp is located in the +5 rather than +3 position relative to the Trp, which we refer to as the WPFDD-loop. In the closed conformation (PRL-1-C104S; Figure 2.7.A), the active site Arg (R110) is engaged in apolar interactions with the phenyl ring of a nearby Phe (F70) and a polar interaction with the carbonyl oxygen of the same residue (Figure 2.7.B). This important contact positions the most C-terminal Asp residue (D72) to participate in catalysis, which has been confirmed with mutagenesis studies of PRL-3.[21] A sequence alignment of the PRL family members is presented in Appendix B. The most important and unique characteristic of the published PRL-1 structures is the presence of a disulfide bond at the active site (Figure 2.7.C). This has implications for changes in both the structure and activity of the enzyme. Although, Sun *et al.* (2005) showed that disulfide bond formation occurs between C49 and C104 by mass spectrometry, this required the use of $H_2O_2$ and all activity assays were conducted under reducing conditions, which gives no indication of the extent of oxidation in this truncated PRL-1 variant. In Figure 2.7.C, the

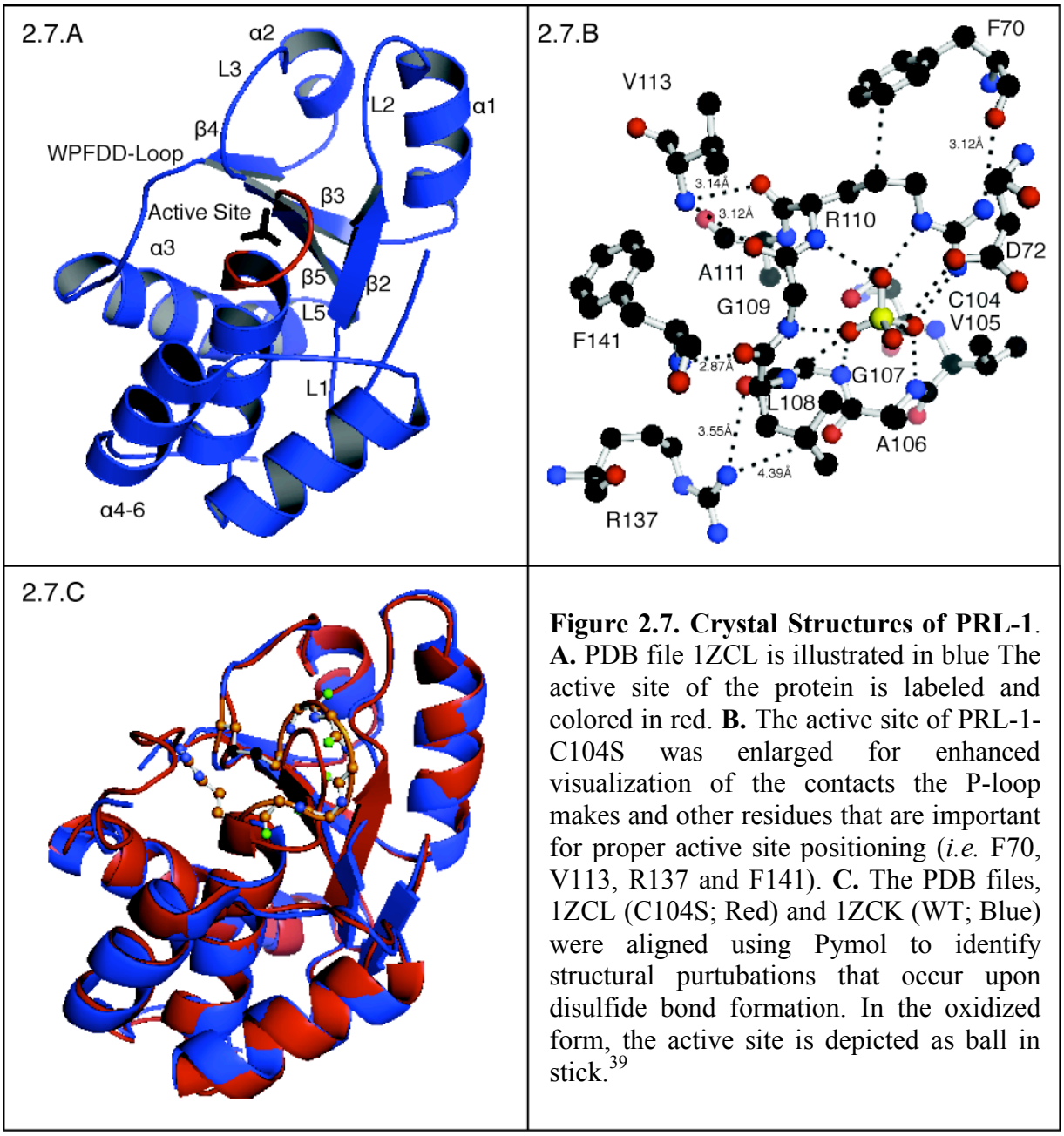



**Figure 2.7. Crystal Structures of PRL-1**. **A.** PDB file 1ZCL is illustrated in blue The active site of the protein is labeled and colored in red. **B.** The active site of PRL-1-C104S was enlarged for enhanced visualization of the contacts the P-loop makes and other residues that are important for proper active site positioning (*i.e.* F70, V113, R137 and F141). **C.** The PDB files, 1ZCL (C104S; Red) and 1ZCK (WT; Blue) were aligned using Pymol to identify structural purtubations that occur upon disulfide bond formation. In the oxidized form, the active site is depicted as ball in stick.[39]

structures of the reduced PRL-1-C104S mutant and oxidized wild type are aligned and oriented to focus on the active site. The crystal structure indicates that structural changes that result from disulfide bond formation are confined to the active site, nearby residues and/or increased flexibility of the WPFDD-loop. For example, the backbone NH groups of P-loop residues lining the active site pocket in PRL-1-C104S are rotated in PRL-1-WT with the carbonyl carbons now

pointing inwards preventing substrate coordination. Additionally, the residues of the WPFDD-loop are absent from the oxidized crystal structure, presumably because of flexibility or disorder. In line with this description, this study shows that wild-type PRL-1 forms a disulfide bond readily *in vitro,* which inhibits its PTPase activity (Figure 2.6).

It appears, however, that upon oxidation, the overall topology of PRL-1 remains the same and that disulfide bond formation is accommodated within the structure without causing considerable conformational changes (Figure 2.7.C). In fact, the rmsd value for alignment of the backbones of the two structures is approximately 0.636 for the two protein forms. Based on this information, it was predicted the C49S enzyme should have comparable activity to the chemically reduced wild type. This was not the case, as indicated by Figure 2.6, which shows that PRL-1-C49S is approximately 1.5-fold less active than the chemically reduced wild type. Closer inspection of the C49S progress curve (data not shown) shows C49S initially turns over substrate at a rate comparable to the reduced wild type. Prior to any significant reduction in the total *p*NPP concentration, however, the rate slows to that of background, indicating substrate turnover has ceased. It appears that C49S can initially rapidly react with *p*NPP to a reasonable extent but has diminished ability to release product. This type of behavior is characteristic of a substrate-trapping mutant, similar to the D72A mutant studied by other groups,[21,39] suggesting that disulfide bond formation may cause a considerably larger change in the protein than indicated by the two crystal structures.

Although maximal activity of wild-type PRL-1 is gained by the addition of 10 mM DTT (Figure 2.6.A), the chemically reduced enzyme still turns over substrate extremely slowly, comparable to the previously published kinetic rate constants.[21,39] Unique structural features of PRL-1's active site may explain this observation. Typical of other dsPTPases, a shallow active

site pocket is present at the surface of the protein, although the β2-α1 loop (loop two; L2) and the β3-α2 loop (loop three; L3) are considerably shorter in PRL-1 and give rise to an unusually wide (almost 9 Å) and shallow catalytic pocket compared to other dsPTPases, which suggests that either PRL-1 may be able to accommodate a unique or broader range of substrates, or its specificity is derived from posttranslational modification and/or binding partners that alter its conformation.

Also unique to the active site of PRL-1 is the substitution of an Ala (A111) for the conserved Ser/Thr in the $CX_5RS/T$ signature motif that facilitates E-P hydrolysis (Figure 2.7.B). Accordingly, it was predicted that this mutation was a major contributing factor to the inexplicably low *in vitro* activity of PRL-1 (and -3) towards generic substrates.[21,39] As shown by Sun *et al.* (2005), substitution of Ala with Ser to create the PRL-1-A111S mutant increases the rate of activity by about 15 fold. Although a 15-fold rate enhancement is significant, it is rather small compared to the differences in reaction rates of other PTPases in which an Ala has been substituted for the Ser/Thr. Others have seen as much as a 100-fold reduction in enzymatic activity with this mutation in other PTPases.[12] There are two additional residues that participate in unique interactions with the active site pocket that also contribute to PRL-1's low *in vitro* activity, albeit to a reasonably small extent (Figure 2.7.B) First, position 113 is predicted (based on structural and sequence homology) to provide two H-bond donors to the carbonyl carbon of G109. The first comes from the backbone NH and the second comes from a conserved Thr side chain. In PRL-1, Val is substituted for the Thr. Furthermore, the NH of this residue is equally close in space to the carbonyl of R110, and it is difficult to predict to which atom it would prefer to donate. Sun *et al.* (2005) found that substitution of Val with Thr increases the enzyme activity by seven fold.[39] In a subsequent study by the same group, it was shown that a double mutant

consisting of both the A111 to Ser and V113 to Thr mutations (PRL-1-A111S/V113T) exhibited a 24-fold increase in PTPase activity compared to the wild type.[7] Second, position 137 is predicted (also based on structural and sequence homology) to provide two H-bond donors to the carbonyl groups of A106 and G107 of the P-loop. In other dsPTPases, this residue is an Arg and its position is stabilized via tertiary contacts with nearby Asp and Asn residues. While the corresponding Arg (R137) residue is present in PRL-1, its interactions with the P-loop are relatively weak because it lacks the appropriate nearby residues to stabilize its position. Instead of Asn or Asp, V10 and presumably A8 are situated in close proximity to R137. Despite missing the appropriate tertiary contacts, Sun *et al.* (2005) found that mutation of R137 to Ala causes a ten-fold reduction in enzyme activity indicating this residue does weakly interact with the active site.[39]

Collectively, this information indicates that structural reorganization of the active site may be needed to confer activity. This could be provided from substrates or binding partners *in vivo*. For example, ATF-5 (activating transcription factor-5; mistakenly called ATF-7 in reference 47) was identified as a PRL-1 binding partner using a yeast two-hybrid system and confirmed using a variety of gel-shift assays *in vitro.*[48] Although the unusually flat surface properties of L2, L3 and the active site is consistent with binding to a coiled-coiled protein such as ATF-5, further confirmation of the PRL-1 and ATF-5 interaction has yet to be described by other groups. [40] This may be due, in part, to the use of inappropriate (incorrectly posttranslationally modified) PRL-1 forms in the binding assays. In agreement with this idea, the majority of the conserved amino acids are located in or surrounding the active site. As such, it is likely that substrate specificity and/or active site rearrangement is determined by other unique

features, distal to the active site where amino acid substitutions among the PRLs are located, which may direct posttranslational modifications of the primary sequence.[39,40]

Y53 is located within a predicted Src kinase motif and is predicted to be phosphorylated *in vivo*.[49] To determine whether modification of this residue affects the activity of PRL-1, the Y53E mutant was constructed. As indicated by Figure 2.6, this mutant has markedly less activity than the reduced wild type and only moderately more activity than the C104S negative control. This finding indicates that phosphorylation (or other modifications of this residue) may negatively regulate PRL-1 activity. On the other hand, modification to residues C170 and C171 dramatically enhanced the activity of PRL-1 and the C170S-C171S double mutant does not require a reducing agent for activity. The PRL enzymes are unique in that each isoform contains a C-terminal prenylation motif or CaaX box, where C stands for cysteine, a represents any aliphatic residue and X corresponds to any amino acid. The prenylation motif directs farnesylation at the conserved CaaX box Cys, followed by proteolytic cleavage of the aaX peptide and methylation of the terminal carboxyl moiety.[50] Farnesylation has been shown to occur *in vivo* for the PRL enzymes.[49,51] In PRL-1, this modification occurs at C170 and often affects its subcellular localization. Because farnesylation occurs on C170 and mutation of this residue increases the *in vitro* activity, we expect farnesylation will also alter the protein's conformation to a more active state. This correlates well with biological studies, which found that when farnesylation is prevented, localization of PRL-1 is altered and its ability to promote proliferation is abrogated. For example, expression of a prenylation-defective mutant of PRL-1 in HeLa cells causes mitotic defects[4] and the loss of PRL-1 prenylation in HEK293 cells inhibits the ability of PRL-1 to promote cell growth and migration.[7]

In summary, the data presented in this chapter show that alterations to two of the predicted posttranslational sites affect the enzyme activity of PRL-1, such that farnesylation at the C-terminus probably promotes catalysis, while phosphorylation at Y53 would inhibit it. These studies also indicate that the C170S-C171S mutant provides the most stable, reduced and active form of PRL-1 of those tested with which inhibition assays and screening for lead compounds for drug design can be pursued.

## 2.5. REFERENCES


1. Zabell A. P., Schroff A. D., Bain B. E., Van Etten R. L., Wiest O. & Stauffacher C. V. (2005) Crystal structure of the human β-form low molecular weight phosphotyrosyl phosphatase at 1.6-A resolution. *J Biol Chem*, 281, 6520-6527.
2. Zhang Z.-Y. (2001) Protein tyrosine phosphatases: prospects for therapeutics. *Curr Opin Chem Biol*, 5, 416-423.
3. Li S., Depetris R. S., Barford D., Chernoff J. & Hubbard S. R. (2005) Crystal structure of a complex between protein tyrosine phosphatase 1B and the insulin receptor tyrosine kinase. *Structure*, 13, 1643-1651.
4. Wang J., Kirby C. E. & Herbst R. (2002) The tyrosine phosphatase PRL-1 localizes to the the endoplasmic reticulum and the mitotic spindle ad is required for normal mitosis. *J Biol Chem*, 277, 46659-46668.
5. Matter W. F., Estridge T., Zhang C., Belgaje R., Stancato L., Dixon J., Johnson B., Bloem L., Pickard T., Donaghue M., Acton S., Jeyaseelan R., Kadambi V. & Vlahos C. J. (2001) Role of PRL-3, a human muscle-specific tyrosine phsophatase, in angiotensin-II signaling. *Biochem Biophys Res Commun*, 283, 1061-1068.
6. Guo K., Li J., Tang J. P., Koh V., Gan B. Q. & Zeng Q. (2004) Catalytic domain of PRL-3 plays an essential role in tumor metastasis. *Cancer Biol Ther*, 3, 945-951.
7. Sun J.-P., Luo Y., Yu X., Wang W.-Q., Zhou B., Liang F. & Zhang Z.-Y. (2007) Phosphatase activity, trimerization, and the C-terminal polybasic region are all required for PRL1-mediated cell growth and migration. *J Biol Chem*, 282, 29043-29051.
8. Fiordalisi J. J., Keller P. J. & Cox A. D. (2006) PRL tyrosine phosphatases regulate Rho family GTPases to promote invasion and motility. *Cancer Res*, 66, 3153-3161.
9. Wu X., Zeng H., Zhang X., Zhao Y., Sha H., Ge X., Zhang M., Gao X. & Xu Q. (2004) Phosphatase of regenerating liver-3 promotes motility and metastasis of mouse melanoma cells. *Am J Pathol*, 164, 2039-2054.
10. Rouleau C., Roy A., St. Martin T. B., Dufault M. R., Boutin P., Liu D., Zhang M., Puorro-Radzwill K., Rulli L., Reczek D., Bagley R., Byrne A., Weber W., Roberts B., Klinger K. W., Brondyk W., Nacht M., Madden S. L., Burrier R., Shankara S. & Teicher B. A. (2006) Protein tyrosine phosphatase PRL-3 in malignant cells and endothelial cells: expression and function. *Mol Cell Biol*, 5, 219-229.
11. Lee J., Yang H., Georgescu M., Di Cristofano A., Maehama T., Shi Y., Dixon J., Pandolfi P. P. & Pavletich N. (1999) Crystal structure of the PTEN tumor suppressor implications for its phosphoinositide phosphatase activity and membrane association. *Cell*, 99, 323-334.
12. Zhang Z.-Y. (2003) Chemical and mechanistic approaches to the study of protein tyrosine phosphatases. *Acc Chem Res*, 36, 385-392.
13. Zhang Z.-Y. (1998) Protein-tyrosine phosphatases: biological function, structural characteristics, and mechanism of catalysis. *Crit Rev Biochem Mol Biol*, 33, 1-52.
14. Jia Z., Barford D., Flint A. J. & Tonks N. K. (1995) Structural basis for phosphotyrosine peptide recognition by protein tyrosine phosphatase 1B. *Science*, 268, 1754-1758.
15. Guan K. & Dixon J. E. (1991) Evidence for protein-tyrosine phosphatase catalysis proceeding via a cysteine-phosphate intermediate. *J Biol Chem*, 266, 17026-17030.

16. Zhang Z.-Y., Maclean D., McNamara D., Sawyer T. K. & Dixon J. E. (1994) Protein tyrosine phosphatase substrate specificity: the minimum size of the peptide and the positioning of the phosphotyrosine. *Biochemistry*, 33, 2285-2290.
17. Hengge A. C., Sowa G., Wu L. & Zhang Z.-Y. (1995) Nature of the transition state of the protein-tyrosine phosphatase-catalyzed reaction. *Biochemistry*, 1995, 13982-13987.
18. Keng Y.-F., Wu L. & Zhang Z.-Y. (1999) Probing the function of the invariant tryptophan in the flexible loop of the *Yersinia* protein-tyrosine phosphatase. *Eur J Biochem*, 259, 809-814.
19. Wu L. & Zhang Z.-Y. (1996) Probing the function of Asp128 in the low molecular weight protein-tyrosine phosphatase catalyzed reaction. A pre-steady-state and steady-state kinetic investigation. *Biochemistry*, 35, 5426-5434.
20. Stuckey J. A., Schubert H. L., Fauman E., Zhang Z.-Y., Dixon J. E. & Saper M. A. (1994) Crystal structure of Yersinia protein tyrosine phosphatase at 2.5 Å and the complex with tungstate. *Nature*, 370, 571-575.
21. Kozlov G., Cheng J., Ziomek E., Banville D., Gehring K. & Ekiel I. (2004) Structural insights into molecular function of the metastasis-associted phosphatase PRL-3. *J Biol Chem*, 279, 1182-1189.
22. Zhang Z.-Y., Palfey B. A., Wu L. & Zhao Y. (1995) Catalytic function of the conserved hydroxyl group in the protein-tyrosine phosphatase signature motif. *Biochemistry*, 34, 16389-16396.
23. Zhao Y. & Zhang Z.-Y. (1996) Reactivity of alcohols toward the phosphoenzyme intermediate in the protein-tyrosine phosphatase catalyzed reaction: probing the transition state of the dephosphorylation step. *Biochemistry*, 35, 11797-11804.
24. Bae Y. S., Kang S. W., Seo M. S., Baines I. C., Tekle E., Chock P. B. & Rhee S. G. (1997) Epidermal growth factor (EGF)-induced generation of hydrogen peroxide. Roe in EGF receptor-mediated tyrosine phosphorylation. *J Biol Chem*, 272 , 217-221.
25. Meng T. C., Fukada T. & Tonks N. K. (2002) Reversible oxidation and inactivation of protein tyrosine phosphatases *in vivo*. *Mol Cell*, 9, 387-399.
26. Peters G. H., Frimurer T. M. & Olsen O. H. (1998) Electrostatic evaluation of the signature motif (H/V)CX5R(S/T) in protein-tyrosine phosphatases. *Biochemistry*, 37, 5383-5393.
27. Chiarugi P. & Buricchi F. (2007) Protein tyrosine phosphorylation and reversible oxidation: two cross-talking posttranslational modifications. *Antiox Redox Signal*, 9, 1-24.
28. Denu J. M. & Tanner K. G. (1998) Specific and reversible inactivation of protein tyrosine phosphatases by hydrogen peroxide: evidence for a sulfenic acid intermediate and implications for redox regulation. *Biochemistry*, 37, 5633-5642.
29. den Hertog J., Groen A. & van der Wijk T. (2005) Redox regulation of protein-tyrosine phosphatases. *Arch Biochem Biophys*, 434, 11-15.
30. Lee S. R., Kwon K. S., Kim S. R. & Rhee S. G. (1998) Reversible inactivation of protein-tyrosine phosphatase 1B in A431 cells stimulated with epidermal growth factor. *J Biol Chem*, 273, 15366-15373.
31. Caselli A., Marzocchini R., Camici G., Manao G., Moneti G., Pieracini G. & Ramponi G. (1998) The inactivation mechanism of low molecular weight phosphotyrosine-protein phosphatase by H2O2. *J Biol Chem*, 273, 32554-32560.

32. Lee S. R., Yang K. S., Kwon J., Lee C., Jeong W. & Rhee S. G. (2002) Reversible inactivation of the tumor suppressor PTEN by H2O2. *J Biol Chem*, 277, 20366-20342.
33. Savitsky P. A. & Finkel T. (2002) Redox regulation of Cdc25C. *J Biol Chem*, 277, 20535-20540.
34. Claiborne A., Miller H., Parsonage D. & Ross R. P. (1993) Protein-sulfenic acid stabilization and function in enzyme catalysis and gene regulation. *FASEB J*, 7, 1483-1490.
35. Salmeen A., Anderson J. N., Myers M. P., Tonks N. K. & Barford D. (2000) Molecular basis for the dephosphorylation of the activation segment of the insulin receptor by protein tyrosine phosphatase 1B. *Mol Cell*, 6, 1401-1412.
36. van Montfort R. L. M., Congreve M., Tisi D., Carr R. & Jhoti H. (2003) Oxidation state of the active-site cysteine in protein tyrosine phosphatase 1B. *Nature*, 423, 773-777.
37. Barrett W. C., DeGnore J. P., Konig S., Fales H. M., Keng Y. F., Zhang Z.-Y., Yim M. B. & Chock P. B. (1999) Regulation of PTP1B via glutathionylation of the active site cysteine 215. *Biochemistry*, 38, 6699-6705.
38. Pregel M. J. & Storer A. C. (1997) Active site Tiration of the tyrosine phosphatases SHP-1 and PTP1B using aromatic disulfides: reaction with the essential cysteine residue in the active site. *J Biol Chem*, 272, 23552-23558.
39. Sun J.-P., Wang W.-Q., Yang H., Liu S., Liang F., Fedorov A. A., Almo S. C. & Zhang Z.-Y. (2005) Structure and biochemical properties of PRL-1, a phosphatase implicated in cell growth, differentiation, and tumor invasion. *Biochem*, 44, 12009-12021.
40. Jeong D. G., Kim S. J., Kim J. H., Son J. H., Park M. R., Lim S. M., Yoon T.-S. & Ryu S. E. (2005) Trimeric structure of PRL-1 phosphatase reveals an active enzyme conformation and regulation mechanisms. *J Mol Biol*, 345, 401-413.
41. Laurence J. S., Hallenga K. & Stauffacher C. V. (2004) $^{1}H$, $^{15}N$, $^{13}C$ resonance assignments of the human protein tyrosine phosphatase PRL-1. *J Biomol NMR*, 29, 417-418.
42. Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R. D. & Bairoch A. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*, 31, 3784-3788.
43. Gill S. C. & von Hippel P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem*, 182, 319-326.
44. Cleland W. W. (1964) Dithiothreitol, a new protective reagent for SH groups. *Biochem*, 3, 480-482.
45. Iyer K. S. & Klee W. A. (1973) Direct spectrophotometric measurement of the rate of reduction of disulfide bonds: the reactivity of the disulfide bonds in bovine alpha-lactalbumin. *J Biol Chem*, 248, 707-710.
46. Denu J. M. & Dixon J. E. (1995) A catalytic mechanism for the dual-specific phosphatases. *Proc Nat Acad Sci*, 92, 5910-5914.
47. Landrieu I., da Costa M., Veylder L. D., Dewitte F., Vandepoele K., Hassan S., Wieruszeski J.-M., Corellou F., Faure J.-D., Montagu M. V., Inze D. & Lippens G. (2004) A small CDC25 dual specificity tyrosine-phosphatase isoform in Arabidopsis thaliana. *Proc Nat Acad Sci*, 101, 13380-13385.
48. Peters C. S., Liang X., Shuixing L., Subburaj K., Peng Y., Taub R. & Diamond R. H. (2001) ATF-7, a novel bZIP protein, interacts with the PRL-1 protein-tyrosine phosphatase. *J Biol Chem*, 276, 13718-13726.

49. Zeng Q., Hong W. & Tan Y. H. (1998) Mouse PRL-2 and PRL-3, two potentially prenylated protein tyrosine phosphatases homologous to PRL-1. *Biochem Biophys Res Commun*, 244, 421-427.
50. Moores S. L., Schaber M. D., Mosser S. D., Rands E., O'Hara M. B., Garsky V. M., Marshall M. S., Pompliano D. L. & Gibbs J. B. (1991) Sequence dependence of protein isoprenylation. *J Biol Chem*, 266, 14603-14610.
51. Zeng Q., Si X., Horstmann H., Xu Y., Hong W. & Pallen C. J. (2000) Prenylation-dependent association of protein-tyrosine phosphatases PRL-1, -2, and -3 with the plasma membrane and the early endosome. *J Biol Chem*, 275, 21444-21452.

Page Left Intentionally Blank

# CHAPTER 3.

# $^{1}H$, $^{15}N$, $^{13}C$ RESONANCE ASSIGNMENTS OF REDUCED AND ACTIVE PRL-1

## 3.1. INTRODUCTION

One of the most important uses of high-field solution nuclear magnetic resonance (NMR) in the pharmaceutical industry relies on its unique ability to examine protein-drug interactions at high resolution. NMR can be used to evaluate the structural, thermodynamic and kinetic aspects of even a weak binding reaction. The basis of NMR screening experiments is that binding causes a perturbation in the physical properties of both molecules. Unique properties of small and macromolecules allow selective detection of either the protein target or ligand, even in a mixture of compounds. This topic has been extensively reviewed in Appendix C, which outlines current methodologies used for assessing protein-ligand interactions from the perspectives of the protein target and ligand and delineates the fundamental principles for understanding NMR approaches in drug research.

In order to conduct protein detected NMR drug screening assays and facilitate structure based drug design, sequence-specific backbone assignments of the protein target are needed. Previously, the $^{1}H$, $^{15}N$, and $^{13}C$ resonance assignments were reported for PRL-1 under nonreducing conditions.[1] Unfortunately, the protein is inactive under these conditions, as indicated by the results presented in Chapter 2. Consequently, a mutant form of PRL-1 (PRL-1-C170S-C171S) has been identified that constitutively exists in the reduced and active state. Although the crystal structures of wild type and reduced but inactive C104S reveal the overall conformation of these two forms is approximately the same, our studies with the C49S mutant

indicate that, in solution, more substantial changes between the two states may be present. Furthermore, the $^{1}H$, $^{15}N$ and $^{13}C$ resonance assignments were reported for PRL-2 and PRL-3 under reducing conditions.[2,3] Comparison of these very closely related homologues (>76% identity) reveals large differences in conformation between the reduced and oxidized forms. See Appendix B for the sequence alignment of the PRL family members. Consequently, the majority of chemical shift positions of resonances from reduced PRL-1 are substantially altered compared to those of the previously reported oxidized form, precluding our ability to deduce assignments based on small chemical shift perturbations. We present here the $^{1}H$, $^{15}N$ and $^{13}C$ resonance assignments of a full-length, reduced and active form of PRL-1 that will facilitate NMR-based drug screening.

## 3.2. METHODS

### 3.2.1. Protein Expression and Purification

PRL-1-WT and C170S-C171S were expressed and purified as previously described in section 2.2.1. All solutions used in the purification and analysis of samples involved in this study were sparged with Ar gas to minimize oxidation derived from $O_2$. Purity of the final protein samples was assessed by SDS-PAGE and protein concentrations were determined by UV Absorbance at 280 nm using an extinction coefficient of 19420 $M^{-1}$ $cm^{-1}$. The extinction coefficient was calculated based on the reduced protein using the ProtParam program from ExPASy.[4,5] Purified protein samples were concentrated to approximately 2.0 mM in 50 mM sodium phosphate at pH 6.5 with 100 mM NaCl.

### 3.2.2. NMR Experiments

For making assignments, the NMR sample was at 2 mM concentration, contained 5% $D_2O$, 50 mM sodium phosphate at pH 6.5 with 100 mM NaCl and was analyzed in a Shigemi tube. 2D $^{1}H$-$^{15}N$ HSQCs and 3D versions of the CBCA(CO)NH, HNCACB, C(CO)NH and HNCO experiments[6] were carried out on a Bruker Avance 800 MHz NMR spectrometer using a cryogenic, triple-resonance probe equipped with pulse field gradients. Water suppression was accomplished using flip-back pulses. All spectra were obtained at 37 °C and referenced relative to DSS.[7] Spectra were processed using nmrPipe [8] and peak picked using Sparky.[9] Peak lists were submitted to the PINE Server through NMRFAM, University of Wisconsin, Madison.[10] All automated assignments were manually verified using Sparky based on the strategies presented by Sattler *et al.* (1999).[6]

For structural comparisons between PRL-1-WT and PRL-1-C170S-C171S, $^{1}H$-$^{15}N$ HSQC spectra were acquired on the same instrument listed above. All spectra were obtained at 37 °C and acquired in 16 scans with 2048 points in the $^{1}H$ dimension and 256 increments for the

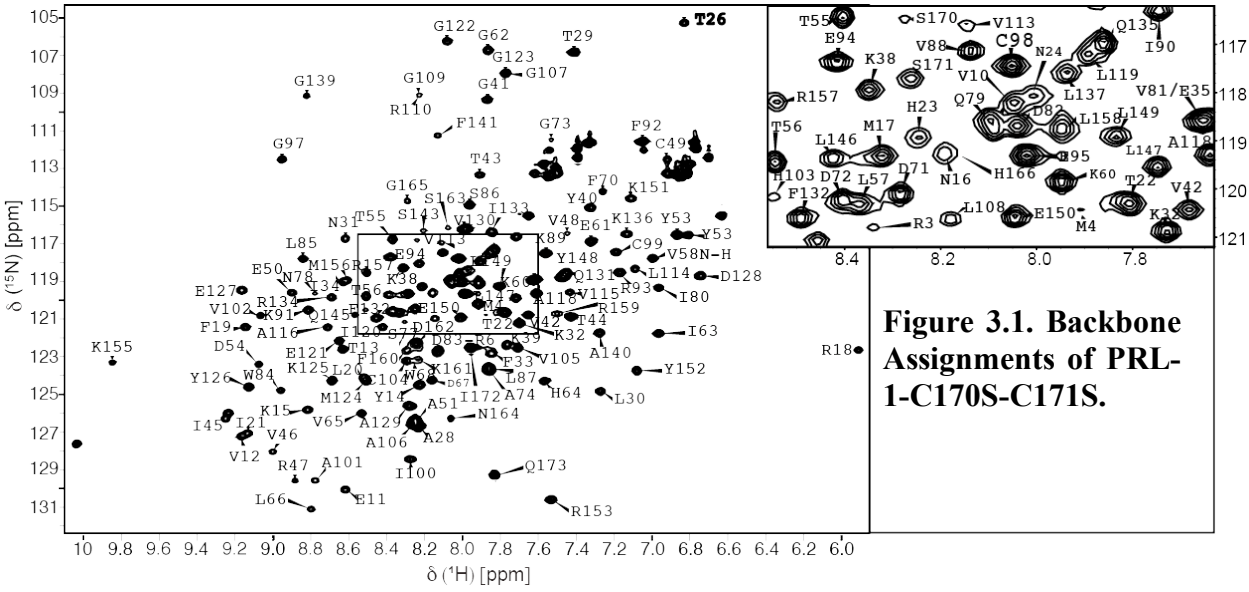



**Figure 3.1. Backbone Assignments of PRL-1-C170S-C171S.**

$^{15}$N dimension. The sample concentrations were 1.0 mM and made up in the same buffer conditions described above. Sample reduction of PRL-1-WT was accomplished by the addition of 10 mM DTT at least 72 hours prior to spectral acquisition. All spectra were referenced as described above, processed using nmPipe and peak picked using Sparky for chemical shift analysis.

## 3.3. RESULTS

### 3.3.1. Completion of Assignments

Greater than 90% of the NH, CO, Cα and Cβ resonances were assigned for PRL-1-C170S-C171S. Complete sequential backbone assignments were achieved for residues V10-Q173 with the exception of residues N27, E36, T52, E59, A111, R137 and R138 and two small regions including residues N142-K144 and R167-N169. Assignments for E36, T52, E59 and R137 are available for wild-type PRL-1, although no $^{13}$C chemical shift information is available for R137. N27, A111 and the C-terminal regions missing in PRL-1-C170S-C171S were also not assigned in the oxidized wild type and are likely undergoing conformational exchange. More importantly, the residues comprising the PTPase active site were fully assigned for PRL-1-C170S-C171S. This information was missing in wild-type PRL-1 and should prove to be useful for NMR-based drug screening assays. More than 80% of the aliphatic side chains were assigned using the CCCONH experiment. Figure 3.1 shows the $^{1}$H-$^{15}$N HSQC spectrum of PRL-1-C170S-C171S with peaks labeled.

### 3.3.2. Chemical Shift Comparisons

The assignments for PRL-1-C170S-C171S correlate well with the assignments previously published for PRL-2 and PRL-3 under reducing conditions, indicating that the C170S-C171S mutant is, in fact, a reduced species. Comparison of PRL-1-C170S-C171S assignments to the

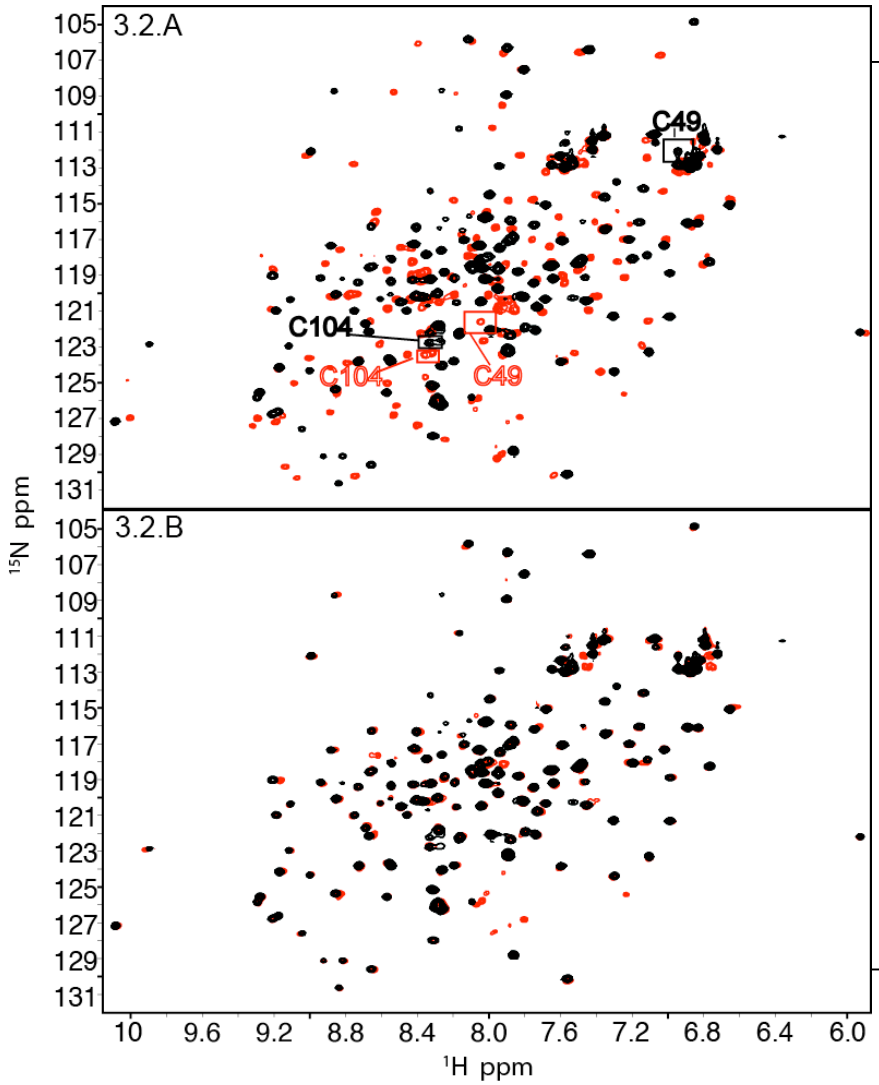



**Figure.3.2. PRL-1 Overlays**.
**A. Oxidized PRL-1-WT (red) and C170S-C171S (black)**. The C49 and C104 residues are highlighted by a black and red boxes, respectively.
**B. Reduced PRL-1-WT (red) and C170S-C171S (black)**. Reduction of the wild type was accomplished by the addition of 10 mM DTT. The small number of unique peaks in the spectra are probably a result of dynamic differences that arise from chemical modulation of the disulfide bond.

non-reduced wild type reveals a difference in conformation between the reduced and oxidized species, which is illustrated in Figure 3.2.A. Overall, more than 90% of the backbone amide resonances shift upon reduction, which is illustrated graphically in Figure 3.3. Exceptions include G123 and Y126, located at the opposite end of the central helix from the active site, and several peaks in the middle of the spectra. The backbone NH resonance of C49, which forms a disulfide bond with the active site residue C104, shifts significantly upfield by more than 9 ppm in nitrogen and 1 ppm in proton (Figure 3.3) with the C170S-C171S mutation, which is also observed in wild type upon chemical reduction. The NH chemical shift of the C49 residue changes to a much greater extent than C104; nonetheless, this shift is significant with the peak

moving upfield just under 1 ppm in nitrogen and 0.1 ppm in proton (Figure 3.3). Cβ chemical shifts can indicate the redox state of cysteine residues [11], and accordingly, the Cβ chemical shifts of the C104 and C49 residues differ between the reduced and oxidized states. The Cβ shifts for C49 and C104 in non-reduced wild type are 35.21 and 36.36 ppm, respectively, while for C170S-C171S, the Cβ shift for C49 is 34.01 and the Cβ peak for C104 is 29.82 ppm. The upfield shift indicates the cysteine residues are in the reduced state in PRL-1-C170S-C171S, although trypsin digestion with mass spectrometry analysis is needed to confirm this and is presented in the next chapter. Lastly, the $^{1}H$-$^{15}N$ HSQC spectra of PRL-1-C170S-C171S overlays exceptionally well with chemically reduced wild type (Figure 3.2.B) and it does not change with the addition of thiol modulating agents (data not shown). This parallels the finding from the previous chapter that PRL-1-C170S-C171S activity also does not change with the

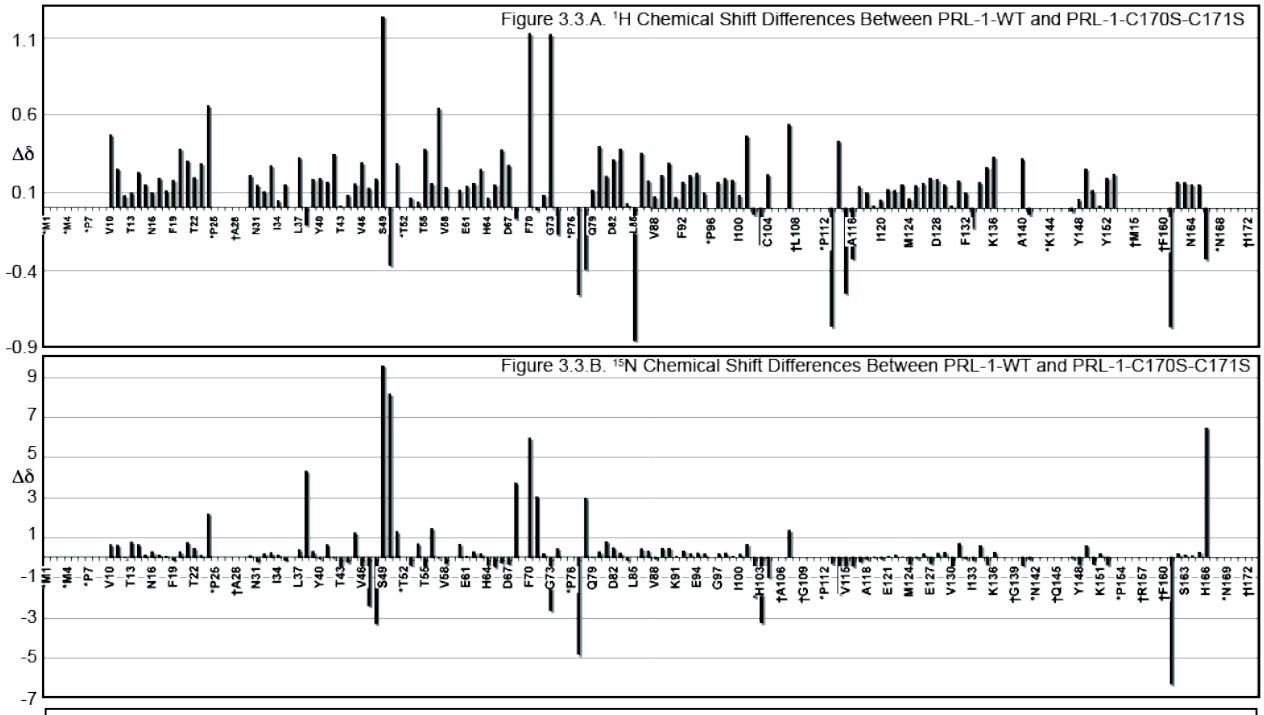



**Figure 3.3. Redox State Based Chemical Shift Changes. A. Δδ ($^{1}H$). B Δδ ($^{15}N$).** Δδ was calculated by subtracting δ (ppm) for PRL-1-C170S-C171S from δ (ppm) for oxidized PRL-1-WT. Residues denoted with ∗, ∘ and †, were either not assigned for both PRL-1 variants, assigned for wild type but not the mutant or assigned for the mutant and not wild type, respectively. Consequently, a Δδ value for these residues was not calculated and appears to be zero in the figure.

addition of DTT (data not shown). Together, the data indicate that PRL-1-C170S-C171S represents the active form of this phosphatase, which can be used to facilitate NMR-based drug screening to identify molecules that inhibit its activity.

### 3.3.3. Data Deposition

The assignments have been deposited in the Biological Magnetic Resonance Databank (BMRB) under the accession number 15949 (http://www.bmrb.wisc.edu).

## 3.4. DISCUSSION

Although the solution NMR data reveal that substantial conformational changes occur in solution between the reduced and oxidized states, such large differences are not apparent in the crystal structures. The majority of the differences are located near the active site when comparing the two crystallized proteins, and these differences are obvious in the NMR data as well. The number and magnitude of changes, however, is much more substantial that would be predicted based on the crystal structures. One explanation for this discrepancy may be the absence of the C-terminus. Crystallized PRL-1 was truncated at residues 156 and 160 for the reduced and oxidized forms, respectively, but we have clearly shown that these residues are important for the structure and activity of the protein. Furthermore, the most dramatic change in the spectrum that occurs upon reduction of PRL-1 corresponds to the chemical shift changes of C49, which is expected. Such a large change can be advantageous for the purpose of studying the reduction potential of the disulfide bond because it provides a large, measurable difference between the two end states. This is the subject of the following chapter.

## 3.5. REFERENCES


1. Laurence J. S., Hallenga K. & Stauggacher C. V. (2004) Letter to the Editor: $^{1}H$, $^{15}N$, $^{13}C$ resonance assignments of the human protein tyrosine phosphatase PRL-1. *J Biomol NMR*, 29, 417-418.
2. Kozlov G., Cheng J., Lievre C., Banville D., Gehring K. & Ekiel I. (2002) Letter to the Editor: $^{1}H$, $^{13}C$ and $^{15}N$ resonance assignments of the human phosphatase PRL-3. *J Biomol NMR*, 24, 169-170.
3. Zhou H., Gallina M., Mao H., Nietlispach D., Betz S. F., Fetrow J. S. & Domaille P. J. (2003) Letter to the Editor: $^{1}H$, $^{13}C$, and $^{15}N$ resonance assignments and secondary structure of the human protein tyrosine phosphatase PRL-2. *J Biomol NMR*, 27, 397-398.
4. Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R. D. & Bairoch A. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*, 31, 3784-3788.
5. Gill S. C. & von Hippel P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem*, 182, 319-326.
6. Sattler M., Schleucher J. & Griesinger C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulse field gradients. *Prog Nucl Magn Reson Spectrosc*, 34, 93-158.
7. Wishart D. S., Bigam C. G., J. Y., Abildgaard F., Dyson H. J., Oldfield E., Markley J. L. & Sykes B. (1995) 1H, 13C, and 15N chemical shift referencing in biomolecular NMR. *J Biomol NMR*, 6, 135-140.
8. Delaglio F., Grzesiek S., Vuister G. W., Zhu G., Pfeifer J. & Bax A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*, 6, 277-293.
9. Goddard T. D. & Kneller D. G. *SPARKY 3*. ed.: University of California, San Francisco.
10. Eghbalnia H. R., Bahrami A., Wang L., Assadi A. & Markley J. L. (2005) Probabilistic identification of spin systems and their assignments including coil-helix interference as output (PISTACHIO). *J Biomol NMR*, 32, 219-233.
11. Sharma D. & Rajarathnam K. (2000) $^{13}C$ NMR chemical shifts can predict disulfide bond formation. *J Biomol NMR*, 18, 165-171.

# CHAPTER 4.
# ENZYME ACTIVITY OF PRL-1 IS CONTROLLED BY REDOX ENVIRONMENT AND ITS C-TERMINAL RESIDUES

## 4.1. INTRODUCTION

PTPases are subject to redox regulation, such that oxidation inhibits their activity and, as such, members of each PTPase subfamily have been shown to be oxidized both *in vitro* and *in vivo*.[1] The mechanism common to all PTPases is to prevent further and irreversible oxidation by forming an S-S bridge containing the catalytic Cys. This process is important because disulfide bond formation inhibits PTPase activity but protects the enzyme from irreversible oxidation, permitting PTPase reactivation under favorable conditions. Intracellular reducing agents such as thioredoxin (TRX), glutaredoxin (GRX) and reduced glutathione (GSH) presumably play essential roles in the recovery of PTPase activity *in vivo*.[2] The propensity for disulfide bond formation will be influenced not only by the $pK_a$ value of the catalytic Cys, but also the structural distance between the two partner Cys residues.[3] Unfortunately, i*n vitro* activity assays and structure analyses are often performed in the presence of high concentrations of reducing agents to prevent inactivation. Despite this precaution, crystallization of PRL-1 produced a structure of the inactive, oxidized form. No other PTPase structure has been reported in the inactive state and many remain active in the absence of reducing agents, suggesting that PRL-1 may be more prone to oxidation than other family members.[4] This is further supported by the data in Chapter 2 indicating that full-length PRL-1 WT is primarily inactive *in vitro* and forms a disulfide bond readily under mild oxidizing conditions.

Disulfide bond formation *in vivo* depends on the presence of reactive oxidants as well as the redox environment of the cell, which is primarily determined by the amount of glutathione and the ratio of its reduced to oxidized forms [GSH:GSSG]. The reduction potential and capacity

of a cell vary with tissue type, age, health and most importantly, cellular compartment.[2] The reported subcellular locations of PRL-1 are quite varied among the different research groups and depend on the presence of posttranslational modifications, especially farnesylation. The specific roles that PRL-1 plays at each of these locations has yet to be elucidated, but because PRL-1 is subject to redox modulation, it is likely that posttranslational modifications may affect the activity of PRL-1 both directly, by altering the conformation of the protein to better prime the active site (as indicated by the C170S-C171S mutant), and indirectly by altering its subcellular location where it may be more or less prone to inactivation by disulfide bond formation. The combination of these regulatory mechanisms provides a novel way for cells to differentially regulate enzymes that share a common mechanism of, but have a different propensity for, oxidative inactivation. In the present study, high-resolution NMR was used to determine the reduction potential of the active site disulfide bond in full-length PRL-1. Determining the reduction potential of the disulfide bond involving the catalytic Cys provides a basis for understanding how subcellular localization may regulate the phosphatase activity of PRL-1 in different types of cellular environments.

## 4.2. METHODS

### 4.2.1. Protein Expression and Purification

All PRL-1 variants were purified as described in section 2.2.1. The primer used to generate the C98A mutant was 5'-cgtgaagaacctggtGCTtgtattgctgttcattgc-3'. Capital letters indicate the mutated bases. All sample buffers for analysis in this study were sparged with Ar gas to minimize oxidation derived from $O_2$. Purity of the final protein samples was assessed by SDS-PAGE and protein concentrations were determined by UV Absorbance at 280 nm using an

extinction coefficient of 19420 $M^{-1}$ $cm^{-1}$. The extinction coefficient was calculated based on the reduced protein using the ProtParam program from ExPASy.[5,6] Purified protein samples were concentrated to approximately 1.5 mM and stored in 50 mM sodium phosphate at pH 7.5 with 100 mM NaCl.

**4.2.2. Electrospray Ionization Mass Spectrometry***

ESI spectra were acquired on a Q-Tof-2 (Micromass Ltd, Manchester UK) Hybrid Mass Spectrometer operated in MS mode employing the time of flight analyzer de-tuned to 8000 resolution (FWHH) for sensitivity. For whole protein ESI spectra, purified protein samples were diluted to 1 $\mu$g/$\mu$L in untreated ("oxidizing") 20 mM ammonium bicarbonate, pH 8.2, and allowed to incubate at room temperature. Time points were taken on the first, fourth and seventh day post-purification and frozen at -80°C until ready for use. Samples were desalted on a short (3 cm x 1 mm I.D.) reverse-phase (RP) HPLC column (Hamilton PRP1, Reno, NV). Samples were loaded onto the column from a 1% formic acid solution with protein (5 $\mu$g), washed in same solution, and eluted with 90% MeOH/0.5% formic acid directly into the ESI source. The cone voltage was 60 eV, and the voltage on the collision cell was 20 V. Spectra were acquired over the mass range 800 to 3000 u, accumulating data for 5 seconds per cycle.

For tryptic digests, PRL-1-WT and PRL-1-C170S-C171S were diluted to 1 mg/mL in ammonium bicarbonate, pH 8.2, with or without 20 mM β-mercaptoethanol (BME), and trypsin (Promega) was added in a 1:50 ratio protease:protein (w/w) and incubated at 37 °C for 16 hours. To confirm complete digestion, samples at various time points were analyzed by SDS-PAGE and visualized by silver staining. Coupled to MS analysis, capillary HPLC separations were performed using a Zorbax SBC18 RP column (5 cm x 0.32 mm I.D., 3.5 μm bead size, 300 Å

* This data was collected in collaboration with Anthony A. Vartia and Todd Williams.

pore size) packed by Micro-Tech Scientific (Sunnyvale, CA), with a chromatograph (Waters capLC XL, Milford, MA) that develops gradients at 10 $\mu$L/min. A linear gradient of 20 to 80% B was applied over 120 minutes. The solvents were: A) 99% $H_2O$, 1% MeOH, B) 99% MeOH, 1% $H_2O$, and both contained 0.08% formic acid. Argon was admitted to the collision cell at a pressure that attenuates the beam to about 20%. This corresponds to 16 psi on the supply regulator or 5.3 x $10^{-5}$ mBar on a penning gauge near the collision cell. The collision cell was operated at 8 V for maximum transmission, and spectra were acquired over the range 250 to 2000 u, accumulating data for 8 seconds per cycle.

**4.2.3. SDS-PAGE**

Protein samples were diluted to 1 mg/mL concentration in 50 mM sodium phosphate, pH 7.5, with 100 mM NaCl and were incubated for 1 hour with varying amounts of reduced or oxidized dithiothreitol (DTT). Oxidized DTT was obtained by incubating $H_2O_2$ with DTT at a 3:1 concentration ratio ($H_2O_2$:DTT); oxidation was confirmed to be complete by monitoring the absorbance at 283 nm.[7,8] After incubation, non-reducing SDS-PAGE loading buffer was added. Samples were then heated for 6 minutes and finally run on a 12% acrylamide gel. Gels were stained using Coomassie (R-250) and the protein bands were used to assess the relative amounts of reduced or oxidized PRL-1. To determine if PRL-1 could be reduced by GSH, a similar study was conducted using GSH as the reducing agent. The protein was incubated under the same conditions with GSH concentrations ranging from 0.1 mM to 200 mM. After incubation, non-reducing SDS-PAGE loading buffer was added. Samples were then heated for 6 minutes and finally run on a 12% acrylamide gel. Gels were stained using Coomassie (R-250) and the protein bands were used to assess the relative amounts of reduced or oxidized PRL-1.

### 4.2.4. Phosphatase Activity Assays

A standard phosphatase activity assay with *p*-nitrophenyl phosphate (*p*NPP) was used to determine the relative activity of PRL-1 variants.[9] All samples were exchanged at least $10^5$ fold by dialysis or using Amicon Ultra 10 kDa centrifugal filters (Millapore) into 50 mM HEPES, pH 7.5 before being tested for phosphatase activity. Assay mixtures contained 50 mM HEPES, pH 7.5, 20 mM *p*-nitrophenyl phosphate (*p*NPP-freshly prepared) and 0-10 mM reduced DTT. Reactions were initiated by the addition of PRL-1 at a final concentration of 1 mg/mL and monitored for absorption at 405 nm and 37 °C using a Cary 100 UV-Vis spectrophotometer with temperature controller. Reactions were run in triplicate and on a minimum of two independently produced protein samples, totaling at least n=6 replicates. Relative activity was determined at 60 minutes by calculating the amount of *p*-nitrophenol produced ($\varepsilon$=1.78x$10^{-4}$ $M^{-1}\cdot cm^{-1}$).[10] Averages and standard deviations of the of the individual runs are shown.

### 4.2.5. NMR Experiments

For structural comparisons of PRL-1 variants, samples were prepared in 50 mM sodium phosphate buffer, 100 mM NaCl, pH 6.5, containing 5% $D_2O$. $^1H$-$^{15}N$ HSQC spectra were acquired on a Bruker Avance 800 MHz spectrometer using a cryogenic, triple resonance probe equipped with pulse field gradients. Water suppression was accomplished using flip-back pulses. The concentrations of PRL-1-WT and PRL-1-C98A were 1 mM and the spectra were collected with 2048 points in $^1H$ and 256 increments in $^{15}N$. Sample reduction was accomplished by the addition of 10 mM DTT at least 72 hours prior to spectral acquisition. All spectra were obtained at 37 °C.

For the redox titration, the method previously published by Piotukh *et al*. (2007) was followed.[11] Briefly, PRL-1-C98A was prepared in 50 mM sodium phosphate, 100 mM NaCl, pH

7.5, containing 5% $D_2O$ at a final concentration of 0.3 mM. As judged by SDS-PAGE, PRL-1-C98A was fully oxidized by adding 5 mM oxidized DTT (DTT oxidation was verified as described above). The sample was then titrated with increasing amounts of reduced DTT (1-200 mM) and a $^{1}H$-$^{15}N$ HSQC spectrum was taken at each point. Spectra were collected on a Bruker Avance 800 spectrometer using a cryogenic, triple resonance probe equipped with pulse field gradients. All spectra were obtained at 37 °C and acquired in at least 16 scans with 2048 points in $^{1}H$ and 256 increments in $^{15}N$. For both studies, $^{1}H$ chemical shifts were referenced with respect to an external DSS standard in $D_2O$. Indirect referencing relative to $^{1}H$ was determined for $^{15}N$ assuming the ratio $^{15}N/^{1}H = 0.101329118$.[12] nmrPipe and SPARKY were used for data processing and spectral analysis.[13,14]

**4.2.6. Determining the Reduction Potential of PRL-1**

Following the method by Piotukh *et al*. (2007) to determine the reduction potential of PRL-1, the heights of several well-resolved peaks in each NMR spectrum were obtained using SPARKY and plotted versus the corresponding half-cell potential of DTT [11]. The half-cell potential of DTT was calculated according to the Nernst equation [2],

$$E_{hc}(mV) = E^{0'} - \frac{R \times T}{n \times F} \times \ln\left(\frac{redDTT}{oxDTT}\right) \times 10^{3} \quad (1)$$

where R is the universal gas constant (8.3145 J $K^{-1}$ $mol^{-1}$), T is the temperature in Kelvin, n=2 for the two-electron reduction, and F is the Faraday constant ($9.6485 \times 10^{4}$ C $mol^{-1}$). For the reduced and oxidized DTT pair, the standard midpoint potential, $E^{0'}$, was previously determined to be -332 mV at pH 7.0 [7]. This value was adjusted to pH 7.5 according to the equation [2,11]

$$E_{pH}(mV) = E^{0} + (pH - 7.0) \times \frac{\Delta E}{\Delta pH} \quad (2)$$

The $\Delta E/\Delta pH$ value for DTT was taken as -61.5 [2].

PRL-1 undergoes a structural transition upon reduction and sequence assignments for the reduced and oxidized species are published elsewhere.[15,16] This information was used to identify chemical shifts for the same residues in each state and to select reference peaks that have no change in chemical shift upon reduction (G123; 7.9 and 107.6 ppm and Y126; 9.3 and 124.3 ppm). The reference peaks were used to control for dynamics changes between the two spectra. The correction factors were averaged for both reference peaks and then applied to the peaks selected for statistical analysis. 10% of the backbone NH resonances that change intensity significantly upon reduction were selected on the basis that they were well-resolved and assigned in both the reduced and oxidized spectra. These peaks were used to calculate $E^{0'}$ of PRL-1. The peak heights of individual resonances were normalized, plotted against the half-cell potential and fit to the following function using Graphpad Prism Software for non-linear curve fitting (Version 5)[11]

$$y = \frac{A - B}{1 + \left(x / E^{0'}\right)p} + B \qquad (3)$$

where A and B are the values for the initial and final plateaus of the sigmoidal decay function and *p* is the curve increase rate from the plot of each resonance. The resulting individual curves, depicting the relative amounts of the corresponding NH resonance in the reduced and oxidized states, intersect at 50% of the maximal intensity. This point defines the standard potential $E^{'}$. Fifteen peaks were fit for an $E^{'}$ value and then averaged to determine the reduction potential.

## 4.3. RESULTS

### 4.3.1. Purified Full-length PRL-1-WT is Inactive

As indicated by our analysis of the full-length, unmodified wild type in Chapter 2, PRL-1-WT is primarily inactive under nonreducing conditions. Because activity was dramatically

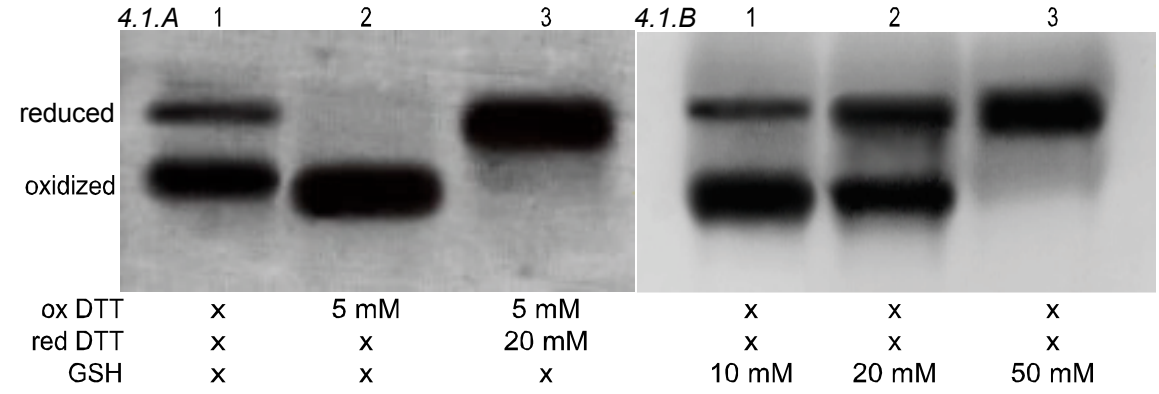



**Figure 4.1. PRL-1-WT. A.** 2 mg/mL PRL-1-WT was incubated with various amounts of oxidized or reduced DTT for approximately one hour. The higher intensity of the lower band indicates that untreated PRL-1-WT is primarily oxidized. **B.** 2 mg/mL PRL-1-WT was incubated with increasing concentrations of reduced glutathione (GSH) and run on SDS-PAGE. Although, only 10-50 mM GSH is shown here, concentrations spanning 0.1-200 mM GSH were tested.

affected by the addition of reducing agents (Figure 2.6.A), we hypothesized that the lack of signal observed for the untreated PRL-1-WT was due to disulfide bond formation at the active site. To test this, we ran untreated PRL-1-WT on SDS-PAGE under non-reducing conditions (Figure 4.1.A, lane 2). When left untreated, wild-type PRL-1 runs as two bands on an SDS-PAGE gel, just below the 20-kDa marker (Figure 1B, lane 1). For wild-type PRL-1, the lower band is much more intense, corresponding to a high degree of oxidation. When run under oxidizing conditions, only the lower band is detected (Figure 4.1.A, lane 2). The addition of a four-fold excess of reduced DTT to the oxidized sample restores the upper band or reduced species. Disulfide bond formation in PRL-1 has been observed by SDS-PAGE previously, but these studies were performed on a truncated version of PRL-1, and hydrogen peroxide was used to accomplish oxidation.[4] Because $H_2O_2$ also oxidizes methionine residues rapidly, we chose to use a thiol-specific chemistry for this study. This approach has the advantage of employing an equilibrium reaction and enables quantitative analysis. Because glutathione (GSH) is the primary

redox buffer *in vivo*, these experiments were also performed using the GSH/GSSG redox couple as a thiol-modulating agent. Although GSSG efficiently oxidized PRL-1 (data not shown), complete reduction by GSH was only accomplished at concentrations near 50 mM (Figure 4.1.B). Because of this, DTT was used in all studies as both an oxidizing and reducing agent.

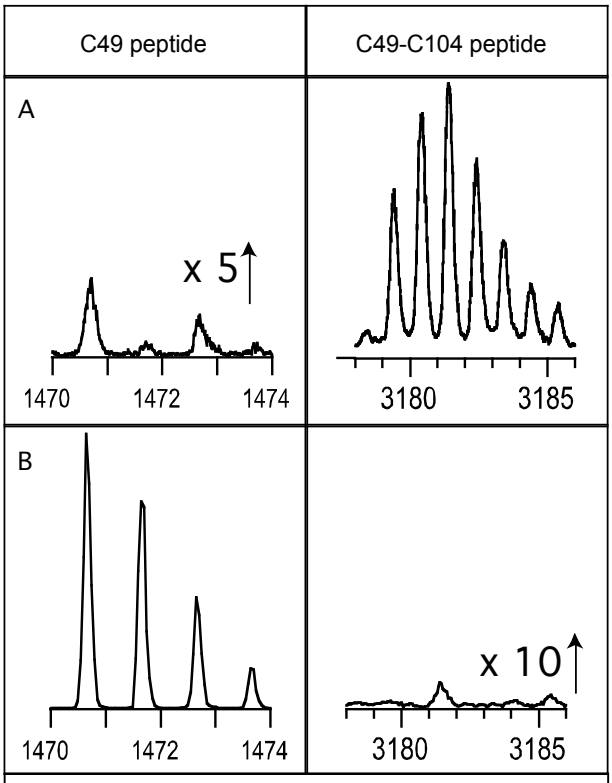



**Figure 4.2. Mass spectra of PRL-1 fragments from a tryptic digest after LC/MS separation.** The column on the left contains data corresponding to the mass series for the C49-containing peptide and the right shows the disulfide bound C49-C104 containing peptides. WT PRL-1 is shown in row **A** and PRL-1-C170S-C171S in row **B**. Some spectra are expanded vertically and the expansion factor is indicated with an arrow in each panel.

To confirm the formation and position of disulfide bonds in the untreated full-length protein, PRL-1 was analyzed using mass spectrometry. Several peaks were observed in the mass spectra of intact protein, which indicate wild-type PRL-1 is oxidized. A mass shift of +2 is observed in the spectra of proteins subject to chemical reduction compared to the oxidized samples, suggesting that one disulfide bond has been broken. To identify the specific sites at which oxidation occurs, PRL-1 was digested with Trypsin and examined using LC-separated ESI-MS under both reducing and oxidizing conditions. Digests of wild-type PRL-1 protein in non-reducing and reducing environments reveal differences in intensity for peaks associated with peptides containing C49 and C104. In the presence of oxygen, PRL-1-WT shows almost exclusively the presence of the C49- and C104-containing fragments joined by a disulfide bond

(Figure 4.2.A). The disulfide-bound fragment is observed as a large *m/z* +1 peak at 3183.3 in the mass spectrum of oxidized PRL-1. The individual C49-containing fragment expected at *m/z* +1 1470.6 in oxidized PRL-1-WT is present only as a trace peak. These two fragments are easily distinguished from other peaks because they are located in unpopulated regions of the spectrum. In contrast, the 1714 *m/z* +1 signal calculated for the C104-containing fragment is buried under other peaks and cannot be easily distinguished among numerous overlapping signals for this sample. Chemical reduction of PRL-1 with reducing agent results in complete loss of signal at 3183.3, indicating the disulfide bond between the two peptides has been broken. Concomitantly, a large peak appears at 1470.6, which corresponds to the individual C49-containing peptide. Additionally, because the nucleophilic cysteine may be oxidized to sulfinic and sulfonic acids, MS data were mined for additional oxidation of the corresponding peptide with two and three oxygen atoms, respectively. Peaks at the predicted masses were not found in any of the samples, indicating that the redox inactivation of PRL-1 occurs predominantly through disulfide bond formation at the catalytic cysteine. Collectively, our data show that purified full-length PRL-1-WT is oxidized and, consequently, requires a reducing agent to restore activity.

### 4.3.2 Complete Oxidation of Full-length PRL-1-WT Leads to Precipitation

NMR can be used to determine the reduction potential of a protein.[11] An accurate calculation of the reduction potential of PRL-1 requires an accurate measurement of the NMR peak heights that correspond to the reduced and oxidized species. This necessitates conversion to either a completely oxidized or reduced species to perform a complete titration and to have an accurate reference baseline. We chose to completely oxidize PRL-1 because it is predominately oxidized following purification without treatment by oxidizing agents. Unexpectedly, incubation of wild-type PRL-1 with oxidizing agents ($H_2O_2$, GSSG and/or oxidized DTT) resulted in

complete precipitation of the protein at one hour. Attempts to re-oxidize PRL-1 following reduction with these components also led to precipitation (data not shown). SDS-PAGE analysis of the insoluble material revealed the precipitate is not mediated by disulfide cross linking (data not shown). Because the NMR experiments necessary to complete a redox titration require the protein to remain stable and soluble for a longer period of time, the wild-type protein was not a suitable candidate for this study.

We have previously shown that the C-terminal mutant, PRL-1-C170S-C171S is primarily reduced and active. Additionally, studies of PRL-1 by other groups indicate that the C-terminus affects solubility.[4,17] As such, we anticipated that precipitation could be avoided by using a C-terminal mutant, to perform the redox titration. As shown in Figure 2.6 (Chapter 2), the untreated C170S-C171S mutant exhibits an eight-fold increase in activity compared to the untreated wild-type protein. Furthermore, the activity of PRL-1-C170S-C171S is statistically identical to wild-type PRL-1 under reducing conditions, and the activity of this mutant is not modulated by the addition of reducing agents. Taken together, these data indicate that the active-site cysteine in the C170S-C171S mutant remains primarily reduced. To confirm this, we performed a trypsin digest and analyzed the peptides by mass spectrometry. The digest of untreated PRL-1-C170S-C171S shows disulfide-linked peptides only at equivalent intensity to background (Figure 4.2.B), suggesting no specific disulfide formation between the C49 and C104-containing peptides occurs. Conversely, the individual C49-containing peptide is exclusively detected. The MS spectrum for PRL-1-C170S-C171S is identical under non-reducing and reducing conditions, as no change in the signal for either mass is observed following addition of reducing agent.

We next investigated the ability to oxidize PRL-1-C170S-C171S. When left untreated, PRL-1-C170S-C171S also runs as two bands on an SDS-PAGE gel; however, the upper band,

which corresponds to reduced PRL-1, is much more intense than the lower band (Figure 4.3.A). We attempted to oxidize PRL-1-C170S-C171S. Based on the intensity of the reduced and oxidized bands on SDS-PAGE, we estimated that this mutant requires at least 20-fold more oxidizing agent than WT (Figure 4.3.B). This finding indicates that PRL-1-C170S-C171S is much less susceptible to oxidation than WT. We did not attempt to quantify the reduction

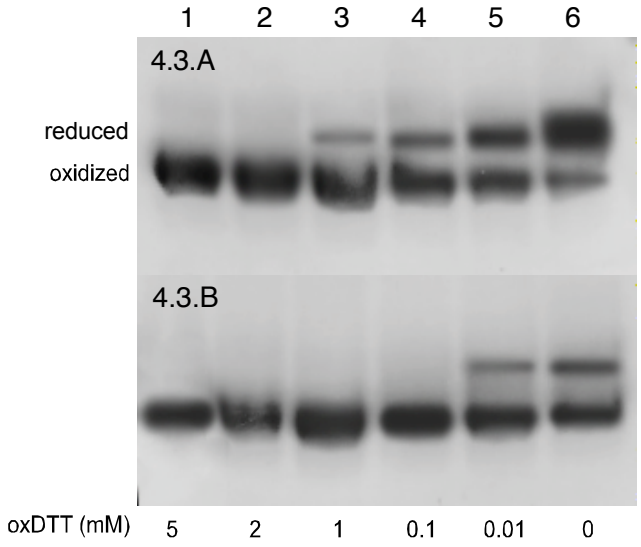



**Figure 4.3. PRL-1-C170S-C171S**. **A.** PRL-1-C170S-C171S was run on SDS-PAGE under various oxidizing conditions. Under non-reducing conditions (lane 6), C170S-C171S runs as two bands. In contrast to wild type, however, C170S-C171S is primarily reduced as indicated by the much greater intensity of the upper reduced band. Complete oxidation of PRL-1-C170-171S occurs near 2 mM oxidized DTT. **B.** PRL-1-WT was run on SDS-PAGE under the same conditions as in B. Complete oxidation of wild type occurs near 0.1 mM oxidized DTT.

potential of this protein because, surprisingly, incubation of PRL-1-C170S-C171S with oxidizing agents also results in complete precipitation of the protein in one hour (data not shown). Instead we pursued identification of a PRL-1 variant that would allow us to determine the active-site reduction potential of full-length PRL-1. We investigated the role of the remaining two Cysteines, C98 and C99, in PRL-1 solubility. The crystal structure of PRL-1 reveals that C98 and C99 are largely buried in the hydrophobic core of the protein, but the side chain of C98 is partially surface-accessible. We hypothesized that oxidation of C98 might lead to formation of a second disulfide bond and cause a structural transition, leading to precipitation. To investigate this hypothesis, we examined the available NMR and MS data more closely.

The Cβ NMR chemical shift of a cysteine residue can be used to deduce the oxidation

state of the thiol moiety. The chemical shift of an oxidized side chain (disulfide bound) should have a downfield shift closer to that of the Cα resonance.[18] The Cβ chemical shift of C98 in untreated wild-type PRL-1 indicates that this residue is reduced (26.449 ppm). For comparison, the Cβ chemical shifts of the disulfide bound C49 and C104 residues are 35.21 ppm and 36.36 ppm, respectively. Reduction of PRL-1 leads to a large change in Cβ chemical shifts for the C49 and C104 residues (34.01 ppm and 29.82 ppm, respectively). Cβ chemical shift information is unavailable for residue C99 in the oxidized state and is absent presumably due to chemical exchange.[15]

To further investigate the propensity of C98 and C99 to be oxidized we examined the MS data of wild type and PRL-1-C170S-C171S. Digested PRL-1-C170S-C171S shows a strong signal for the peptide containing C104 at 1712. The observed value is 2 amu below the predicted mass of 1714. This mass, which is shifted by -2, also is consistently observed in the spectra of the cross-linked species involving the C49- and C104-containing peptides (3181.3) derived from non-reduced wild-type PRL-1. Addition of reducing agent shifts the mass of the C104-containing peptide to 1714 for both wild type and PRL-1-C170S-C171S. Because the C104-containing peptide also includes C98 and C99, it is likely that an intra-peptide disulfide bond between C98 and C99 is also present. MS analysis of intact PRL-1 under non-reducing conditions reveals a peak 4 amu below the calculated mass. This peak shifts to the expected position in the MS spectrum upon addition of reducing agent when no LC separation is performed. As such, it appears that this additional disulfide bond forms after the reducing agent is removed during the LC step.

### 4.3.3. PRL-1-C98A Does Not Precipitate and Is Structurally and Functionally Equivalent to PRL-1-WT

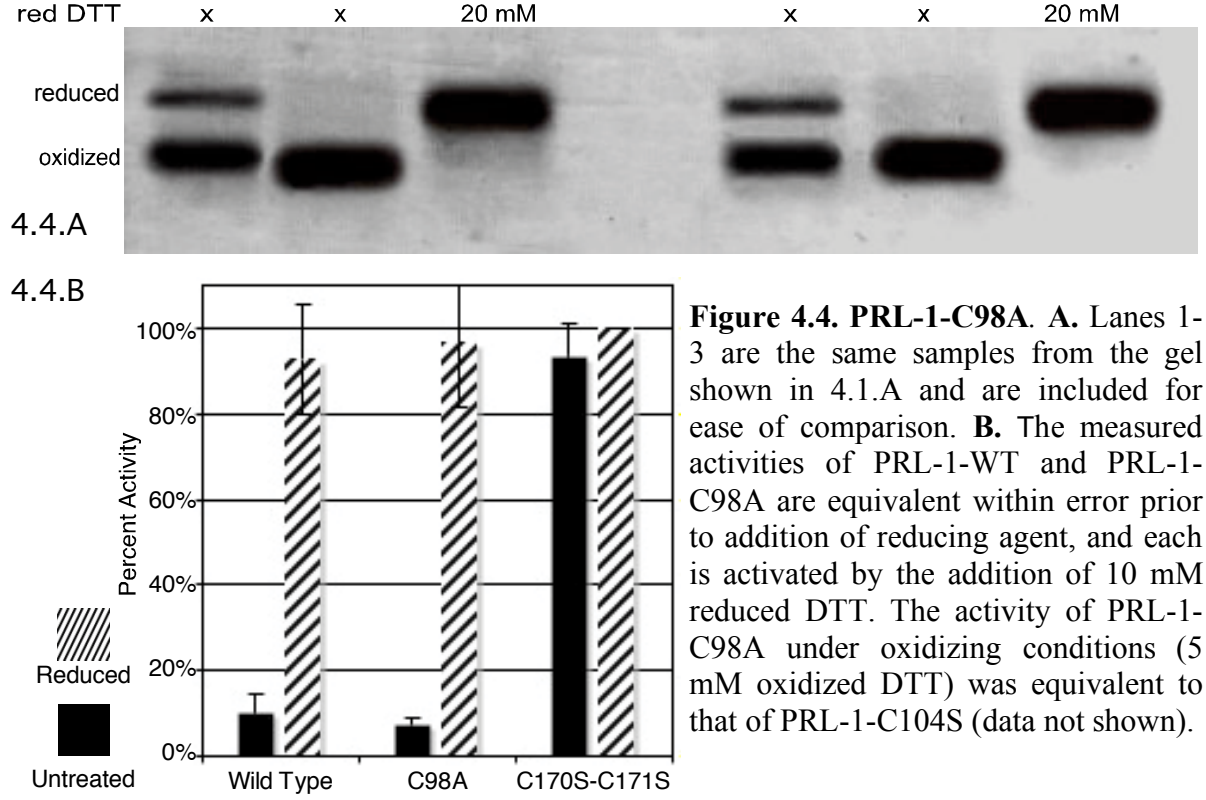



**Figure 4.4. PRL-1-C98A**. **A.** Lanes 1-3 are the same samples from the gel shown in 4.1.A and are included for ease of comparison. **B.** The measured activities of PRL-1-WT and PRL-1-C98A are equivalent within error prior to addition of reducing agent, and each is activated by the addition of 10 mM reduced DTT. The activity of PRL-1-C98A under oxidizing conditions (5 mM oxidized DTT) was equivalent to that of PRL-1-C104S (data not shown).

Because the MS data indicate the possibility of C98-C99 disulfide bond formation upon oxidation, we created the PRL-1-C98A mutant. We hypothesized that chemical oxidation by thiol modulating agents leads to oxidation of the C98 and C99 residues in both PRL-1-WT and PRL-1-C170S-C171S, which results in precipitation. Mutation of C98 mitigates this problem. Untreated PRL-1-C98A runs similarly to PRL-1-WT on SDS-PAGE (Figure 4.4.A, compare lane 1 to 4). Like wild-type PRL-1, complete oxidation of PRL-1-C98A occurs with 5 mM oxidized DTT and is re-reduced by the addition of a four-fold excess of reduced DTT (Figure 4.4.A, compare lane 2 to 5 and 3 to 6). Unlike wild type, incubation of C98A with oxidizing agents for extended periods of time results in little to no precipitation. Consequently, PRL-1-C98A was found to be an appropriate candidate to use for a reversible, equilibrium redox

titration. To ensure that this mutation has not compromised the functional or structural integrity of the protein, we compared the PTPase activity and NMR spectra of PRL-1-WT to PRL-1-C98A.

Wild type and C98A exhibit almost identical substrate turnover when assayed in equivalent redox conditions. PRL-1-C98A shows an equivalent level of activity when left untreated. Oxidation of the protein with 5 mM oxidized DTT completely abolishes the activity of the enzyme (data not shown), while reduction with 10 mM DTT recovers the full activity of this mutant, as shown by comparison to the reduced wild type and C170S-C171S mutant (Figure 4.4.B). The percent activity of PRL-1-C98A under oxidizing conditions was determined to be 0.05 ± 2% and was not included in the figure. Activity of the wild-type protein under oxidizing conditions was not measured because the formation of the precipitate precludes accurate measurement of the assay product.

To verify the C98A mutation does not significantly alter the protein' structure, $^{1}H$-$^{15}N$ HSQC NMR spectra of wild type and PRL-1-C98A were collected. Comparison of the HSQC spectra of wild type and PRL-1-C98A reveals relatively small changes between the two variants (Figure 4.5.A). As such, the overall structure of PRL-1-C98A and PRL-1-WT are virtually identical, as indicated by the large number of peaks in common in these two spectra. Using the previously reported assignments for full-length oxidized PRL-1-WT, the chemical shift changes were mapped to the oxidized crystal structure of truncated PRL-1-WT.[4,15] The significant chemical shift changes observed correspond to residues located near the site of the mutation (Figure 4.5.A, black boxes and Figure 4.5.B). The average chemical shift deviation ($\Delta\delta_{avg}$) for all residues in PRL-1-WT and PRL-1-C98A was calculated by subtracting the chemical shift values of the C98A mutant from those of the wild type and was determined to be 0.07 ppm for $^{15}N$ and

0.01 ppm for $^{1}$H. It should be noted that several new peaks appear in the spectrum of PRL-1-C98A that are not observed for wild type. Peaks for many of the C-terminal residues in PRL-1-WT were absent from the HSQC spectrum, likely because of line broadening due to exchange on the intermediate timescale. We suspect the new peaks in the PRL-1-C98A spectrum are resonances from the C-terminus, and their appearance may reflect a faster rate of exchange on which the motions at these residues occur in the PRL-1-C98A protein. Sequential backbone assignments of the C98A mutant and complete dynamics analysis of both are needed to confirm this, which is beyond the scope of this work.

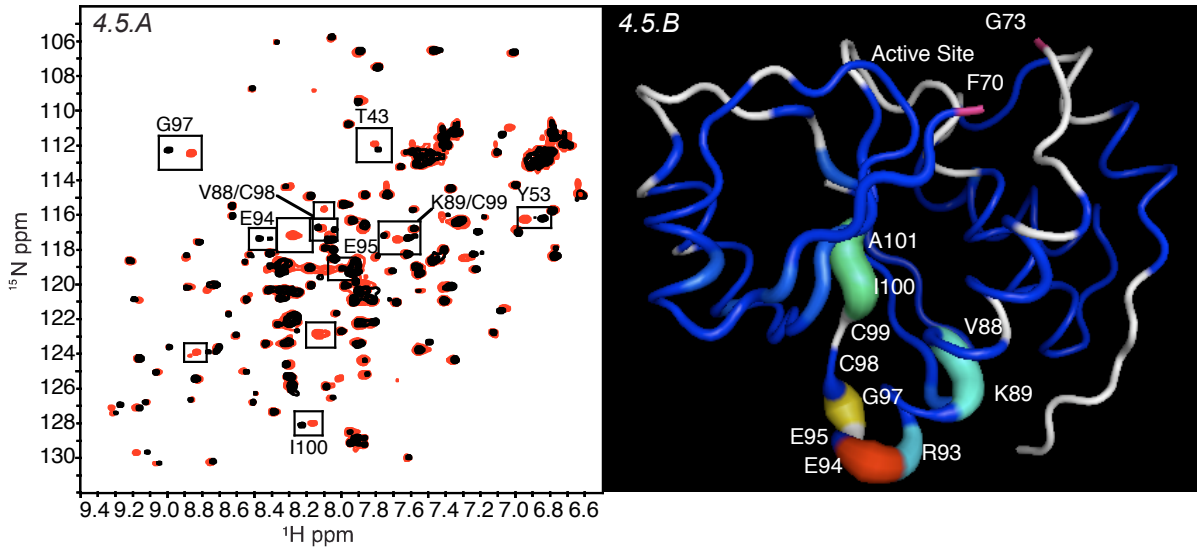



**Figure 4.5. Comparisons of PRL-1-WT and C98A. A.** The spectrum of untreated PRL-1-WT (red) is overlaid with untreated PRL-1-C98A (black). Small differences in structure exist between these two proteins, and the residues that change significantly with mutation are enclosed in boxes. Unlabeled boxes enclose either new peaks in C98A (black) or unassigned peaks in wild type (red). **B.** The chemical shift changes observed between these two proteins were mapped to the available crystal structure of the oxidized wild type. A break in the structure exists between residues F70 and G73 because no information is available for D71 and D72 residues. F70 and G73 are labeled and colored in pink to illustrate the break. For chemical shift mapping, the Δδ value for all assigned residues was used as the b-factor in the pdb file (1ZCK). Residues for which an assignment is missing in one of the two states are colored white. The rainbow color scale is applied with the largest changes being highlighted in red and insignificant changes in blue. Additionally, larger changes in chemical shift correspond to an increase in tube diameter.

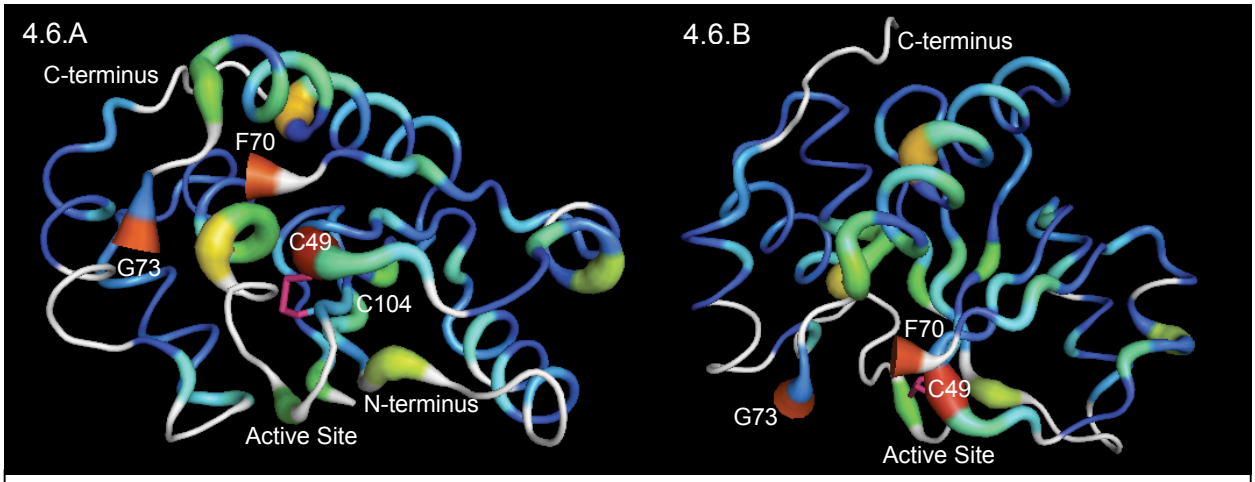



**Figure 4.6. Comparisons of PRL-1-WT and C170S-C171S**. **A.** Differences in NMR chemical shifts between PRL-1-WT and PRL-1-C170-C171S were mapped to the crystal structure (as in figure 4) and displayed with the active site in front. The disulfide bound (pink bond) C104 and C49 are labeled. F70 and G73 are labeled to identify the break in the crystal structure. **B.** As in A, but with the active site rotated downward by approximately 90 degress to reveal the large number of chemical shift deviations in the majority of the protein structure, especially in the central β-strand.

As shown in the previous chapter, reduction of PRL-1 results in a conformational change, causing greater than 90% of the peaks in the wild-type spectrum to move. Backbone chemical shift assignments for both the oxidized and reduced forms were determined, which are reported elsewhere.[15,16] The most dramatic change in the spectrum that occurs upon reduction of PRL-1 corresponds to C49. The position of C49 in the HSQC shifts significantly upfield by more than 9 ppm in nitrogen and 1 ppm in proton compared to the oxidized form. Figure 4.6 maps the observed changes in chemical shifts (Δδ) of assigned resonances to the available crystal structure of oxidized PRL-1. Δδ values were obtained by subtracting the chemical shift of PRL-1-C170S-C171S from those of wild-type PRL-1, which was graphically illustrated in the previous chapter (Figure 3.3). A $\Delta\delta_{avg}$ of 0.91 ppm for $^{15}N$ and 0.24 ppm for $^{1}H$ was determined for these two forms of the protein. Parallel changes in chemical shifts are observed upon reduction of PRL-1-C98A (data not shown). The $\Delta\delta_{avg}$ value was determined to be 0.72 for $^{15}N$ and 0.14 for $^{1}H$ by subtracting the chemical shifts of the reduced protein from those of the oxidized protein. This is

in strong agreement between that observed for wild type and PRL-1-C170S-C171S and confirms that redox modulation at the active site in C98A parallels that of wild type. Collectively, our structural and functional comparisons have shown that PRL-1-C98A is an appropriate analog of the PRL-1 system, from which the reduction potential of the active-site disulfide bond can be determined.

### 4.3.4. Determining the Reduction Potential of the Disulfide Bond in Full-length PRL-1

To determine the reduction potential, oxidized PRL-1 (in the presence of 5 mM oxidized DTT) was gradually reduced by the addition of reduced DTT to a concentration of 200 mM. The peak heights of the following residues were monitored and analyzed as described: V12, R18, L30, T43, R47, C49, V65, L66, I80, R93, I100, A101, L114, E127, and Y152. Peak selection was based on spectral resolution, proximity to the disulfide bond, distance from mutation site and chemical exchange behavior (slow exchanging peaks only). Figure 4.7 displays a representative plot, depicting the results of the titration on residue C49, in which normalized peak height is plotted vs. half-cell potential. Similar results were obtained for all other residues and the chemical shifts of each residue in both states are included in Table 4.1. An average $E^{'}$ value of -364.3 ± 1.5 mV was calculated for the C49-C104 disulfide bond using the data from these fifteen residues.

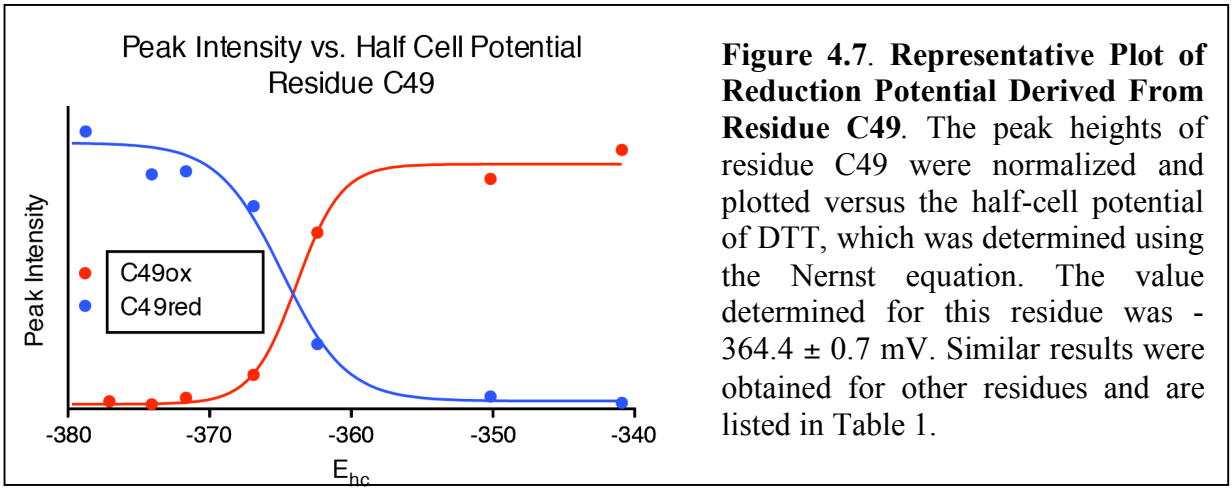



**Figure 4.7. Representative Plot of Reduction Potential Derived From Residue C49.** The peak heights of residue C49 were normalized and plotted versus the half-cell potential of DTT, which was determined using the Nernst equation. The value determined for this residue was -364.4 ± 0.7 mV. Similar results were obtained for other residues and are listed in Table 1.

**Table 4.1. Calculated Reduction Potential for PRL-1 Residues[a]**

| | E' (mV) | ± SD | δ-$^{15}$N ox | δ-$^{1}$H ox | δ-$^{15}$N red | δ-$^{1}$H red | Δδ ($^{15}$N) | Δδ ($^{1}$H) |
|---|---|---|---|---|---|---|---|---|
| V12 | -366.1 | 0.0 | 126.780 | 9.122 | 126.771 | 9.202 | 0.009 | -0.080 |
| R18 | -364.6 | 2.0 | 121.848 | 5.764 | 121.919 | 5.837 | -0.071 | -0.073 |
| L30 | -364.0 | 1.0 | 124.349 | 7.359 | 124.543 | 7.302 | -0.194 | 0.057 |
| T43 | -363.5 | 0.0 | 111.911 | 7.806 | 112.654 | 7.908 | -0.743 | -0.102 |
| R47 | -361.8 | 0.0 | 126.631 | 8.861 | 129.143 | 8.896 | -2.512 | -0.035 |
| C49 | -364.4 | 0.7 | 121.676 | 8.037 | 112.254 | 6.943 | 9.422 | 1.094 |
| V65 | -365.2 | 0.0 | 124.996 | 8.563 | 125.651 | 8.570 | -0.655 | -0.007 |
| L66 | -364.8 | 0.0 | 130.322 | 9.065 | 130.756 | 8.853 | -0.434 | 0.212 |
| I80 | -364.8 | 1.3 | 119.137 | 7.235 | 119.000 | 6.994 | 0.137 | 0.241 |
| R93 | -364.7 | 2.3 | 118.477 | 7.287 | 118.344 | 7.236 | 0.133 | 0.051 |
| I100 | -364.9 | 2.4 | 127.986 | 8.162 | 127.919 | 8.230 | 0.067 | -0.068 |
| A101 | -364.8 | 1.8 | 129.683 | 9.180 | 129.252 | 8.865 | 0.431 | 0.315 |
| L114 | -365.4 | 2.6 | 116.028 | 7.424 | 117.798 | 7.107 | -1.770 | 0.317 |
| E127 | -363.9 | 1.6 | 118.621 | 9.219 | 119.105 | 9.187 | -0.484 | 0.032 |
| Y152 | -362.7 | 1.9 | 122.770 | 7.129 | 123.362 | 7.099 | -0.592 | 0.030 |
| average | -364.3 | 1.5 | | | | | | |

[a]The reported E' value for each residue corresponds to the average of both the reduced and oxidized peak heights fit to the equation listed in the methods section. The standard deviation (SD) is also reported in mV. Δδ values were calculated by subtracting the chemical shift of the reduced species from those of the oxidized species.

## 4.4. DISCUSSION

The reduction potential of the active site disulfide bond in full-length PRL-1 was determined to be approximately -365 mV at pH 7.5. The reported reduction potentials determined for various cellular environments range from -320 (most reducing) to -170 mV (most oxidizing), the details of which have been reviewed extensively.[2,19-22] Based on these values, our results indicate that nascent PRL-1 is rapidly oxidized inside the cell to an inactive state. The reduction potential is considerably lower than those reported for proteins that are known to play an important role in maintaining the redox state of the cell and/or specific proteins. The reduction potential of thioredoxin-1 (Trx-1), found in the cytoplasm and nucleus, has been reported to be -230 mV, while that of Trx-2, found in the mitochondria, is near -330 mV.[19,23] Both enzymes

would have a substantial fraction of protein in the active state in their respective environments, and Trx-1 would be able to act as a reducing agent more effectively in the nucleus. Reduction potentials for cytoplasmic glutaredoxins (Grx) vary depending on the primary sequence and range from -200 to -240 mV.[24] Reduction potentials for proteins that facilitate disulfide bond formation in the endoplasmic reticulum have been determined to be -175 mV (protein disulfide isomerase; PDI) and -275 mV (regulatory disulfide of Ero1α).[25,26] The role of PDI is thought to be to make and break disulfide bonds in target proteins and the reduction potential is appropriate for this enzyme in this compartment. The reduction potential of Ero1α is sufficiently negative to ensure oxidation and facilitate re-oxidation of PDI. To date, analysis of reduction potentials has been focused on a small subset of redox modulating enzymes, but measuring the reduction potentials of other proteins that are subject to redox regulation has recently been recognized as important as well. The absence of general and accurate methods for quantifying these potentials has precluded the ability to identify conditions in which proteins are switched on or off by disulfide bond formation. NMR spectroscopy has recently been shown to be a valuable tool for obtaining this information and was used to determine the reduction potential of Trx-1 with great accuracy.[11] Other proteins have been studied using this method as well. Zimmerman *et al*. (2007) used NMR to determine the reduction potential of the N-terminal hSH3 domain of the immune cell protein ADAP (adhesion and degranulation promoting adapter protein).[27] Sahu *et al*. (2008) used NMR to quantify the reduction potential of the High Mobility Group B1 (HMGB1) protein, which plays an important role in DNA binding as well as chemotaxis.[28] In both examples, disulfide bond formation was shown to lead to a conformational change, which alters the function of the protein. The altered function of the protein also correlates with a change in subcellular location. Likewise, the subcellular location of PRL-1 may affect the activity of the

protein, although no cellular compartment is sufficiently negative to ensure activation. Even in the most reducing environment reported (-320 mV), nascent PRL-1 would exist primarily in an inactive, oxidized state (> 96%). This strongly suggests that *in vivo*, PRL-1 requires a posttranslational modification or binding partner to switch on phosphatase activity.

The data presented earlier demonstrate the importance of the C-terminus to the phosphatase activity of PRL-1. The data presented here show that modification of the C-terminus affects the redox state of the active site, making PRL-1 much less susceptible to inactivation (Figure 4.4). The PRL enzymes each contain a C-terminal prenylation motif or *CaaX* box (*C* stands for cysteine, *a* represents any aliphatic residue and *X* corresponds to any amino acid). The prenylation motif directs farnesylation at the conserved CaaX box cysteine, followed by proteolytic cleavage of the aaX peptide and methylation of the terminal carboxyl moiety.[29] Farnesylation has been shown to occur *in vivo* for the PRL enzymes.[30-32] In PRL-1, this modification occurs at C170. Based on our data, the shift in the reduction potential of the inactivating disulfide bond in PRL-1 to a less negative value appears to be accomplished through a conformational change that depends on the presence of C-terminal residues in the CaaX box. Because farnesylation occurs on C170 and mutation of this residue makes PRL-1 less prone to inactivation, we expect farnesylation will also alter the protein's conformation and reduction potential of the active site disulfide bond. Because farnesylation often leads to membrane association and PRL-1 has been shown to localize to several membranes *in vivo,*[4,17,30,31,33,34] the protein's conformation may be altered more substantially in the context of a lipid bilayer than is apparent from solution studies.

In all work published to date, the phosphatase activity of PRL-1 has been reported to be necessary for the promotion of cellular proliferation and motility by this protein, regardless of

localization.[32,33,35-39] PRL-1 is typically farnesylated in cells [30-32,40] and farnesylated PRL-1 is most often directed to the nucleus.[32,41,42] The fact that PRL-1 is most often localized to the nucleus seems appropriate with respect to cellular redox conditions,[41] because this compartment is often the most reducing organelle at -320 mV.[19] Based on our studies of the C-terminal mutant, it seems that modification at the C-terminus would raise the reduction potential sufficiently to turn on phosphatase activity in this environment. Interestingly, when farnesylation of PRL-1 is prevented in cells, the ability of this protein to promote proliferation is abrogated and its localization is often altered. For example, the lack of prenylation inhibits the ability of PRL-1 to promote invasion and motility of SW480 colon carcinoma cells and to promote HEK293 cell growth and migration.[33,40] In these cases, it was not determined whether localization and/or PTPase activity is responsible for the altered phenotype. In mitotic cells, however, PRL-1 localizes to the centrosomes and spindle apparatus independent of whether farnesylation occurs, and despite proper localization, farnesylation deficient mutants result in mitotic defects.[32] Interestingly, the inactive mutant produced the same result. During mitosis, the nuclear envelope breaks down and the reduction potential of the cell becomes approximately -240 mV.[2,43,44] The reduction potential of unmodified PRL-1 (-365 mV) is much too low to be active in this redox environment. As such, farnesylation may enhance PRL-1 activity during mitosis, and this may explain why mitotic defects were observed with both the catalytically inactive C104 mutant and non-farnesylated proteins.

PRL-1 facilitates the regeneration of damaged tissue by promoting proliferation and migration, notably in the liver, which has a highly reducing environment. Increased activity in this case may be due, at least in part, to elevated expression of the protein.[45] The expression and phosphatase activity of PRL-1 have also been associated with many types of cancer cell

lines.[34,41,46-48] It has been well documented that the interior of cancer cells are generally more reducing than their normal counterparts.[49-51] Because cancer cells have aberrantly regained the ability to proliferate and have lost the ability to maintain redox homeostasis,[2] we hypothesize that the redox environment in some cancer cells may be sufficiently negative to confer activity in PRL-1. This is supported by the fact that abolishing the activity of PRL-1 in cancer cell lines by transfecting them with catalytically inactive PRL mutants decreases cell viability, prevents cells from entering into mitosis and also inhibits anchorage independent growth.[32,33] These and other studies collectively suggest that the active form of PRL-1 may be a good target for therapeutic intervention in the treatment of metastatic cancers.[46]

It is clear from our *in vitro* studies that changes in cellular redox conditions will have a profound affect on the activity of PRL-1 *in vivo* and that modification of residues, especially at the C-terminus, may dramatically alter the propensity of PRL-1 to be oxidized and inactivated inside the cell. Given its diverse localization and multifarious roles in proliferation, differentiation and motility, the data presented here provides the impetus to pursue further studies that investigate the function of PRL-1 under distinct cellular conditions to generate a more accurate account of how PRL-1 is regulated *in vivo* and how its regulation contributes to such diverse cellular outcomes.

## 4.5. REFERENCES


1. Chiarugi P. (2005) PTPs versus PTKs: The redox side of the coin. *Free Rad Res*, 39, 353-364.
2. Schafer F. Q. & Buettner G. R. (2001) Redox environment of the cell as viewed through the redox state of the glutathione disulfide/glutathione couple. *Free Rad Biol Med*, 30, 1191-1212.
3. Peters G. H., Frimurer T. M. & Olsen O. H. (1998) Electrostatic evaluation of the signature motif (H/V)$CX_5$R(S/T) in protein-tyrosine phosphatases. *Biochemistry*, 37, 5383-5393.
4. Sun J.-P., Wang W.-Q., Yang H., Liu S., Liang F., Fedorov A. A., Almo S. C. & Zhang Z.-Y. (2005) Structure and biochemical properties of PRL-1, a phosphatase implicated in cell growth, differentiation, and tumor invasion. *Biochem*, 44, 12009-12021.
5. Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R. D. & Bairoch A. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*, 31, 3784-3788.
6. Gill S. C. & von Hippel P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem*, 182, 319-326.
7. Cleland W. W. (1964) Dithiothreitol, a new protective reagent for SH groups. *Biochem*, 3, 480-482.
8. Iyer K. S. & Klee W. A. (1973) Direct spectrophotometric measurement of the rate of reduction of disulfide bonds: the reactivity of the disulfide bonds in bovine alpha-lactalbumin. *J Biol Chem*, 248, 707-710.
9. Zhang Z.-Y. (2003) Chemical and mechanistic approaches to the study of protein tyrosine phosphatases. *Acc Chem Res*, 36, 385-392.
10. Landrieu I., da Costa M., Veylder L. D., Dewitte F., Vandepoele K., Hassan S., Wieruszeski J.-M., Corellou F., Faure J.-D., Montagu M. V., Inze D. & Lippens G. (2004) A small CDC25 dual specificity tyrosine-phosphatase isoform in Arabidopsis thaliana. *Proc Nat Acad Sci USA*, 101, 13380-13385.
11. Piotukh K., Kosslick D., Zimmermann J., Krause E. & Freund C. (2007) Reversible disulfide bond formation of intracellular proteins probed by NMR spectroscopy. *Free Rad Biol Med*, 43, 1263-1270.
12. Wishart D. S., Bigam C. G., Yoa J., Abildgaard F., Dyson H. H., Oldfield E., Markley J. L. & Sykes B. D. (1995) $^{1}$H, $^{13}$C and $^{15}$N chemical shift referencing in biomolecular NMR. *J Biomol NMR*, 6, 135-140.
13. Goddard T. D. & Kneller D. G. (2004) SPARKY. *University of California, San Francisco*.
14. Delaglio F., Grezesiek S., Vuister G. W., Zhu G., Pfeifer J. & Bax A. (1995) NMRPipe: a multidimensonal spectra processing system based on UNIX pipes. *J Biomol NMR*, 6, 277-293.
15. Laurence J. S., Hallenga K. & Stauffacher C. V. (2004) $^{1}$H, $^{15}$N, $^{13}$C resonance assignments of the human protein tyrosine phosphatase PRL-1. *J Biomol NMR*, 29, 417-418.
16. Skinner A. & Laurence J. S. (2009) $^{1}$H, $^{15}$N, $^{13}$C resonance assignments of the reduced and active form of human protein tyrosine phosphatase, PRL-1. *Biomol NMR Assign*, 3, 61-65.

17. Jeong D. G., Kim S. J., Kim J. H., Son J. H., Park M. R., Lim S. M., Yoon T.-S. & Ryu S. E. (2005) Trimeric structure of PRL-1 phosphatase reveals an active enzyme conformation and regulation mechanisms. *J Mol Biol*, 345, 401-413.
18. Sharma D. & Rajarathnam K. (2000) 13C NMR chemical shits can predict disulfide bonds. *J Biomol NMR*, 18, 165-171.
19. Hansen J. M., Go Y.-M. & Jones D. P. (2006) Nuclear and mitochondrial compartmentaion of oxidative stress and redox signaling. *Annu Rev Pharmacol*, 46, 215-234.
20. Hwang C., Sinskey A. J. & Lodish H. F. (1992) Oxidized redox state of glutathione in the endoplasmic reticulum. *Science*, 257, 1496-1497.
21. Ellgaard L. (2004) Catalysis of disulphide bond formation in the endoplasmic reticulum. *Biochem Soc Trans*, 32, 663-667.
22. Kosower N. S. & Kosower E. M. (1978) The gluathione status of cells. *Int Rev Cytol*, 54, 109-160.
23. Watson W. H., Pohl J., Montfort W. R., Stuchlik O., Reed M. S., Powis G. & Jones D. P. (2003) Redox potential of human thioredoxin 1 and identification of second dithiol/disulfide motif. *J Biol Chem*, 278, 33408-33415.
24. Aslund F., Berndt K. D. & Holmgren A. (1997) Redox potential of glutaredoxins and other thiol-disulfide oxidoreductases of the thioredoxin superfamily determined by direct protein-protein redox equilibria. *J Biol Chem*, 272, 30780-30786.
25. Lundstrom J. & Homgren A. (1993) Determination of the reduction-oxidation potential of the thioredoxin-like domains of protein disulfide-isomerase from the equilibrium with glutathione and thioredoxin. *Biochem*, 32, 6649-6655.
26. Baker K. M., Chakravarthi S., Langton K. P., Sheppard A. m., Lu H. & Bulleid N. J. (2008) Low reduction potential of Ero1a regulatory disulphides ensures tight control of substrate oxidation. *EMBO J*, 27, 2988-2997.
27. Zimmerman J., Kuhne R., Sylvester M. & Freund C. (2007) Redox-regulated conformational changes in an SH3 domain. *Biochem*, 46, 6971-6977.
28. Sahu D., Debnath P., Takayama Y. & Iwahara J. (2008) Redox properties of the A-domain of the HMGB1 protein. *FEBS Lett*, 582, 3973-3978.
29. Moores S. L., Schaber M. D., Mosser S. D., Rands E., O'Hara M. B., Garsky V. M., Marshall M. S., Pompliano D. L. & Gibbs J. B. (1991) Sequence dependence of protein isoprenylation. *J Biol Chem*, 266, 14603-14610.
30. Cates C. A., Michael R. L., Stayrook K. R., Harvey K. A., Burke Y. D., Randall S. K., Crowell P. L. & Crowell D. N. (1996) Prenylation of oncogenic human PTPcaax protein tyrosine phosphatases. *Cancer Lett*, 110, 49-55.
31. Zeng Q., Si X., Horstmann H., Xu Y., Hong W. & Pallen C. J. (2000) Prenylation-dependent association of protein-tyrosine phosphatases PRL-1, -2, and -3 with the plasma membrane and the early endosome. *J Biol Chem*, 275, 21444-21452.
32. Wang J., Kirby C. E. & Herbst R. (2002) The tyrosine phosphatase PRL-1 localizes to the the endoplasmic reticulum and the mitotic spindle ad is required for normal mitosis. *J Biol Chem*, 277, 46659-46668.
33. Sun J.-P., Luo Y., Yu X., Wang W.-Q., Zhou B., Liang F. & Zhang Z.-Y. (2007) Phosphatase activity, trimerization, and the C-terminal polybasic region are all required for PRL1-mediated cell growth and migration. *J Biol Chem*, 282, 29043-29051.

34. Wang Y., Li Z.-F., He J., Li Y.-L., Zhu G.-B., Zhang L.-H. & Li Y.-L. (2007) Expression of human phosphatases of regnerating liver (PRLs) in colonic adenocarcinoma and its correlation with lymph node metastasis. *In J Colorectal Dis*, 22, 1179-1184.
35. Werner S. R., Lee P. A., DeCamp M. W., Crowell D. N., Randall S. K. & Crowell P. L. (2003) Enhanced cell cycle progression and down regulation of p21Cip1/Waf1 by PRL tyrosine phosphatases. *Cancer Lett*, 202, 201-211.
36. Zeng Q., Dong J.-M., Guo K., Li J., Tan H.-X., Koh V., Pallen C. J., Manser E. & Jong W. (2003) PRL-3 and PRL-1 promote cell migration, invasion and metastasis. *Cancer Res*, 63, 2716-2722.
37. Daouti S., Li W.-H., Qian H., Huang K.-S., Holmgren J., Levin W., Reik L., McGady D. L., Olivier A. R., Sergi J. A., Fry D., Danho W., Ritland S., Fotouhi N., Heimbrook D. & Niu H. (2008) A selective phosphatase of regenerating liver phosphatase inhibitor suppresses tumor cell anchorage-independent growth by a novel mechanism involving p130Cas cleavage. *Cancer Res*, 68, 1162-1169.
38. Guo K., Li J., Tang J. P., Koh V., Gan B. Q. & Zeng Q. (2004) Catalytic domain of PRL-3 plays an essential role in tumor metastasis. *Cancer Biol Ther*, 3, 945-951.
39. Wu X., Zeng H., Zhang X., Zhao Y., Sha H., Ge X., Zhang M., Gao X. & Xu Q. (2004) Phosphatase of regenerating liver-3 promotes motility and metastasis of mouse melanoma cells. *Am J Pathol*, 164, 2039-2054.
40. Fiordalisi J. J., Keller P. J. & Cox A. D. (2006) PRL tyrosine phosphatases regulate Rho family GTPases to promote invasion and motility. *Cancer Res*, 66, 3153-3161.
41. Bessette D. C., Qiu D. & Pallen C. J. (2008) PRL PTPs: mediators and markers of cancer progression. *Cancer Metastasis Rev*, 27, 231-252.
42. Gnainsky Y., Spira G., Paizi M., Bruck R., Nagler A., Genina O., Taub R., Halevy O. & Pines M. (2006) Involvement of the tyrosine phosphatase early gene of liver regeneration (PRL-1) in cell cycle and in liver regeneration and fibrosis effect of halofuginone. *Cells Tissue Res*, 324, 385-394.
43. Menon S. G., Sarsour E. H., Spitz D. R., Higashikubo R., Sturm M., Zhang H. & Goswami P. C. (2003) Redox regulation of the G1 to S phase transition in the mouse embryo fibroblast cell cycle. *Cancer Res*, 63, 2109-2117.
44. Harvey A. J., Kind K. L. & Thompson J. G. (2002) Redox regulation of early embryo development. *Reproduction*, 123, 479-486.
45. Diamond R. H., Cressman D. E., Laz T. M., Abrams C. S. & Taub R. (1994) PRL-1, a unique nuclear protein tyrosine phosphatase affects cell growth. *Mol Cell Biol*, 14, 3752-3762.
46. Stephens B. J., Han H., Gokhale V. & Von Hoft D. D. (2005) PRL phosphatases as potential molecular targets in cancer. *Mol Cancer Ther*, 4, 1653-1661.
47. Achiwa H. & Lazo J. S. (2007) PRL-1 tyrosine phosphatase regulates c-Src levels, adherence, and invasion in human lung cancer cells. *Cancer Res*, 67, 643-650.
48. Liu Y.-Q., Li H.-X., Lou X. & Lei J.-Y. (2008) Expression of phosphatase of regnerating liver 1 and 3 mRNA in esophageal squamous cell carcinoma. *Arch Pathol Lab Med*, 132, 1307-1312.
49. Batist G., Behrens B. C., Makuch R., Hamilton T. C., Katki A. G., Louie K. G., Myers C. E. & Ozols R. (1986) Serial determinations of glutathione levels and glutathione-related enzyme activities in human tumor cells *in vitro*. *Biochem Pharmacol*, 35, 2257-2259.

50. McEligot A. J., Yang S. & Meyskens F. L. J. (2005) Redox regulation by intrinsic species and extrinsic nutrients in normal and cancer cells. *Annu Rev Nutr*, 25, 261-295.
51. Wolf C. R., Lewis A. D., Carmichael J., Adams D. J., Allan S. G. & Ansell D. J. (1987) The role of glutathione in determining the response of normal and tumour cells to anticancer drugs. *biochem Soc Trans*, 15, 728-730.

Page Left Intentionally Blank

# CHAPTER 5.
# PROBING RESIDUE-SPECIFIC INTERACTIONS IN THE STABILIZATION OF PROTEINS USING HIGH-RESOLUTION NMR: A STUDY OF DISULFIDE BOND COMPENSATION

## 5.1. INTRODUCTION

An understanding of the forces that contribute to protein stability is crucial to the development of new biotechnology products having improved half-lives, shelf lives and reduced immunogenicity.[1,2] One method commonly used to improve the stability of proteins is engineering cysteine residues into the primary sequence to form a new or additional disulfide bonds.[3-5] This has proved to be important for the stabilization of antibodies.[6-8] Classical theory suggests that disulfide bonds stabilize proteins by reducing the entropy of the unfolded state.[10-12] More recent theories propose disulfide bonds stabilize the folded state enthalpically, presumably through favorable local interactions or by stabilizing the packing of hydrophobic residues.[14,15] Disulfide bonds are not always beneficial. Alterations to a protein's redox status in the form of disulfide exchange can disrupt a protein's native structure and lead to enhanced aggregation by exposing previously buried, often hydrophobic, regions of a protein that can then associate.[3,17] Generally speaking, it is thought that disulfide bonds destabilize the folded state by restricting energetically favorable conformational changes.[17-19] Unfortunately, very little experimental data to support any of the described ideas is available, which makes predicting the circumstances under which disulfide bond formation will stabilize or destabilize a protein fold difficult.[20-23]

Currently, a set of low-resolution experiments is commonly employed to assess structural stability (and more recently dynamics) and this approach has proven beneficial for rapidly identifying stabilizing conditions for individual proteins.[24-27] The methods used provide general information about classes of interactions or structural features in the protein. Despite the obvious

utility of these traditional methodologies, mechanistic understanding cannot be gleaned from these data sets to elucidate why stabilization is affected. Additional information about the contribution of individual residues can be obtained by comparing mutants to the wild-type protein using these methods and calculating changes in the free energy ($\Delta\Delta G$) in an unfolding experiment.[5,28] While this approach can determine whether a specific residue is required for folding, it does not reveal how cooperativity and conformational dynamics contribute to stabilization of the folded state. High-resolution or site-specific detail is required to explain why a protein is inherently more or less stable or is stabilized to a greater or lesser extent by specific solution conditions. Solution NMR studies of protein folding and unfolding are commonly employed to examine folding pathways,[29-31] but this powerful technique has been greatly underutilized to examine the mechanisms of stabilization of the folded state from both the perspective of the packing arrangement in the protein as well as the effects of solution environment on the protein's conformational ensemble. Understanding how structure and dynamics affect the stability of the folded state is important for generating protein therapeutics that better resist unfolding and aggregation.

In the present study, the protein phosphatase of regenerating liver (PRL-1) was used as a model system to investigate the effects of disulfide bond formation on the protein's stability. PRL-1 has two discernable stable states, one of which contains a disulfide bond (oxidized, inactive) and one that does not (reduced, active). Here, the disulfide bond was disrupted in several ways and the protein carefully examined to assess the effects these two cysteines and their oxidation state have on the conformation of the protein. PRL-1 serves as an excellent model for this purpose for several reasons. First, the catalytic Cys (C104) is highly susceptible to inactivation by disulfide bond formation at the active site.[16,32] Our group previously assigned the

NMR resonances for the reduced and oxidized proteins and characterized formation of the disulfide bond between C104 and its partner C49.[33,34] The reduction potential is approximately -364 mV, which indicates that this protein strongly favors the oxidized state *in vitro*, but equilibrium control of the redox state is easily modulated using a redox buffer system to generate the reduced form.[35] Second, a major conformational change occurs upon reduction of the protein, providing well-resolved, site-specific measurable differences between the two states using solution NMR spectroscopy.[35] Additionally, the structures of both reduced and oxidized PRL-1 have been determined previously using X-ray crystallography, which facilitates interpretation of the changes observed in the NMR data with respect to the three-dimensional structure.[13,16]

Standard low-resolution techniques (circular dichroism and static light scattering) were used to show that the reduced form of the protein is slightly more stable than the oxidized state. To understand how specific residues in the reduced form compensate for the loss of the disulfide bond, a series of mutants were analyzed that disrupted disulfide bond formation in distinct ways. The physical stability of each variant is unique with respect to the others, indicating that the global stability of the protein depends on a combination of local, synergistic interactions with the Cys side chains. High-resolution solution NMR provided site-specific and mechanistic information about how each interaction influences the protein's physical stability. This analysis reveals several key structural components that contribute to PRL-1's overall stability in the absence of disulfide bond formation. The results provide insight into how the reduced protein compensates for the loss of this structural feature and how local instability may lead to aggregation.

## 5.2. METHODS

### 5.2.1. Protein Expression and Purification

All PRL-1 variants were expressed and purified as previously described in the preceding chapters. The primer used to generate the Y53F mutant was 5'-gagtatgtgaagcaacttTtgacactactctt-gtgg3'. Purity of the final protein samples was assessed by SDS-PAGE and protein concentrations were determined by UV Absorbance at 280 nm using an extinction coefficient of 19420 $M^{-1}$ $cm^{-1}$. The extinction coefficient was calculated based on the reduced protein using the ProtParam program from ExPASy.[36,37] Final samples were concentrated to 30 mg/mL and stored at 4 °C in 50 mM Tris-Cl, pH 7.4 with 100 mM NaCl until prior to analysis. These conditions are equivalent to purification conditions. All samples for NMR studies were grown on minimal media containing $^{15}N$-labeled ammonium chloride as the sole nitrogen source (Cambridge Isotopes, Andover, MA). Table 5.1 summarizes the reason for studying each mutant. The activity of each mutant was tested using the standard generic substrate *p*-nitrophenyl phosphate (*p*NPP) in a colorimetric assay previously published.[35]

**Table 5.1. Purpose of PRL-1 Mutants**

| PRL-1 Variant | Purpose |
|---|---|
| Wild Type (WT) | Permits chemical modification of the disulfide bond |
| C104S | Disrupts the disulfide bond directly; abolishes enzymatic activity |
| C49S | Disrupts the disulfide bond directly; retains activity |
| Y53E | Breaks hydrogen bond between Y53 and H103; increases solvent exposure of loop three |
| Y53F | Breaks hydrogen bond between Y53 and H103; retains apolar side chain |
| C98A[a] | Control for WT that permits full oxidation (see Chapter 4) |
| C170S-C171S[a] | Indirectly affects disulfide bond formation |

[a]This chapter will not cover the stability results for the C98A and C170S-C171S mutants. The data for these mutants will be presented in Appendix C.

### 5.2.2. Static Light Scattering (SLS)

To assess the aggregation propensity of the various PRL-1 variants, the optical density at 350 nm was monitored with increasing temperature. Protein samples were diluted to a final concentration of 1 mg/mL into 50 mM HEPES or 50 mM sodium phosphate, pH 7.5 (I=0.1) with or without 10 mM dithiothreitol (DTT). All data were collected on a Cary 100 UV-VIS spectrophotometer equipped with 12-cell changer and temperature controller. Data points were collected every 2 °C at a rate of 5 °C per minute from 10-85 °C. Onset temperatures were determined from the plots of optical density versus temperature. The data were analyzed in duplicate and the average of these is shown. The error bars represent one standard deviation of the mean.

### 5.2.3. Circular Dichroism (CD)

CD spectra at 20 and 85 °C were acquired using a Jasco J-810 spectropolarimeter equipped with a 6-position Peltier temperature controller from 260-190 nm using a scanning speed of 50 nm/min and 0.5 nm resolution. Thermal melts were also performed by monitoring the CD signal at 222 nm every 1 °C from 20-85 °C. The temperature was gradually increased at a rate of 2 °C per minute, and the sample was equilibrated for five minutes at each temperature. The cuvette path length was 0.1 cm. As indicated by the SLS data, aggregation of the protein occurs at elevated temperatures, and based on our structural analysis of the protein, this process is probably mediated via association of β-strands (see section 5.3). Because of this, we monitored the loss of α-helical content to avoid complications when fitting the data. The CD signal was converted to molar ellipticity using the Spectra Manager$^{TM}$ software (Jasco Inc., Easton, MD). $T_m$s were determined by fitting the CD curve to a sigmoidal function with Prism 5 software (GraphPad, La Jolla, CA). The $T_m$ was defined by the midpoint of the sigmoidal curve. Protein

samples were diluted to a final concentration of 0.4 mg/mL in either 50 mM HEPES or 50 mM sodium phosphate, pH 7.5 (I=0.1) with or without 1 mM DTT. Ionic strength was maintained by the addition of sodium chloride. Two independently prepared samples were analyzed in duplicate (n=4) to generate error bars, which represent one standard deviation of the mean. The one-sample t-test was used to determine statistical significance using the reduced wild type examined in phosphate buffer as the theoretical mean.

**5.2.4. NMR Experiments**

For structural analysis of the various PRL-1 mutants, samples were concentrated to approximately 1 mM (~30 mg/mL) in 50 mM sodium phosphate buffer, 100 mM NaCl, pH 6.5, containing 5% $D_2O$. $^{1}H$-$^{15}N$ Heteronuclear Single Quantum Coherence (HSQC) spectra were acquired on a Bruker Avance 800 MHz spectrometer using a cryogenic, triple resonance probe equipped with pulse field gradients. The spectra were acquired in 8 scans with 2048 points in $^{1}H$ and 256* increments in $^{15}N$. All spectra were obtained at 37 °C. NMRpipe[38] and SPARKY[39] were used for spectral processing and data analysis. Chemical shift changes (Δδ) were calculated by subtracting the peak positions of the mutant protein (C104S or C49S) from the chemically reduced wild type. Chemical shift changes were expressed as the square root of the sum of the squares of chemical frequency differences in the $^{1}H$ and $^{15}N$ dimensions between two given proteins for each assigned residue as given in Equation 1.[40]

$$\Delta f = \left[ \left( \Delta\delta_{1H} \left( 800.234 \frac{Hz}{ppm} \right) \right)^2 + \left( \Delta\delta_{15N} \left( 81.096 \frac{Hz}{ppm} \right) \right)^2 \right]^{1/2} \quad (1)$$

A change was deemed significant if $\Delta f$ was greater than 33.81. This value represents the average line width of the reduced wild-type protein and was determined by summing the squares of the individual resonant line widths for $^{1}H$ and $^{15}N$ using Equation 1 and averaging. Line widths were

measured in SPARKY.

Because the C49S mutant was significantly different in structure from the reduced wild-type protein, 2D $^{1}H$-$^{15}N$ HSQCs and 3D versions of the CBCA(CO)NH, HNCACB, C(CO)NH and HNCO experiments were collected to obtain sequential backbone assignments. The sample was concentrated to approximately 1.9 mM in 50 mM sodium phosphate, pH 6.5, with 100 mM NaCl and 5% $D_2O$ and analyzed in a Shigemi tube. Experiments were carried out on a Varian Inova 800 MHz NMR spectrometer using a cryogenic, triple-resonance probe equipped with pulse field gradients. All spectra were obtained at 37 °C. Spectra were processed using NMRpipe[38] and peak picked using Sparky.[39] Peak lists were submitted to the PINE Server through NMRFAM, University of Wisconsin, Madison[41] and cross-checked using the assignments for reduced wild type.[34]

Hydrogen-deuterium (H-D) exchange experiments were conducted to assess the solvent accessibility of residues. Post-purification, samples were exchanged at least $10^5$ fold in 50 mM sodium phosphate with 100 mM NaCl at pH 6.5 and concentrated to 25 μL immediately prior to the experiment. The samples were then diluted to 600 μL (approximately 1 mM protein) in 100% $D_2O$ in a standard NMR tube and immediately placed in the NMR spectrometer. $^{1}H$-$^{15}N$ HSQC spectra were recorded at 5, 10, 30, 60 and 120 minutes as well as 12 and 24 hours at 37 °C. Data was collected for the C104S, C49S and reduced C98A mutants. The C98A mutant was used in lieu of the wild-type protein because a wider range of redox conditions can be used with this mutant (see Chapter 4).[35] The spectra were processed with nmrPipe[38] and peak picked using Sparky.[39] The intensity of the relevant peaks were plotted versus time and fit to an exponential curve with Prism 5 software (GraphPad, La Jolla, CA). Many residues could not be fit because they exchanged completely in 5 minutes. These residues were classified as ++. Conversely,

residues that could not be fit because they were fully retained after 24 hours were classified as --.

For all NMR experiments described above, $^{1}H$ chemical shifts were referenced with respect to an external DSS standard in $D_2O$. Indirect referencing relative to $^{1}H$ was determined for $^{15}N$ assuming a ratio of $^{15}N/^{1}H = 0.101329118$.[42]

## 5.3. RESULTS AND DISCUSSION

### 5.3.1. Physical Stability of PRL-1

To determine the overall physical stability of the different forms of PRL-1, CD melts and SLS melts were performed on each variant. For CD, a $T_m$ was calculated by fitting the data to a sigmoidal curve. For SLS, the onset of the transition temperature is reported because saturation of the detector occurred in many cases, an indication of insoluble aggregation formation. Visible observation of the formation of solid material following the temperature titration indicates the process is not entirely reversible. For this reason, thermodynamic comparisons (ΔG measurements) could not be made. The data presented in this study are interpreted as a reflection of relative differences observed under equivalent conditions rather than as a thermodynamic quantity. Table 5.2 summarizes this data. The CD and SLS plots for all PRL-1 variants in phosphate buffer are shown in Figure 5.1.

Several trends in the CD data were observed with the PRL-1 variants. First, the reduced wild-type protein is slightly more stable than the oxidized protein, with melting temperatures at 67.8 ± 0.4 versus 65.5 ± 1.3 °C, respectively (p=0.0351). We anticipated the opposite result because, most often, native disulfide bonds improve the ability of proteins to resist unfolding.[3-8,10-12,14,15] To understand how the reduced protein compensates for the loss of the disulfide bond, the same analyses were also performed on each of the two cysteine mutants (C49S and C104S)

that directly prevent formation of this bond under equivalent conditions (Table 5.1).

**Table 5.2 Physical Stability of PRL-1[a]**

| | Phosphate | | | | HEPES | | | |
|---|---|---|---|---|---|---|---|---|
| | CD | | SLS | | CD | | SLS | |
| | $T_m$ | ± SD | Onset T | ± SD | $T_m$ | ± SD | Onset T | ± SD |
| nrWT | 65.5 | 1.1 | 59.9 | 0.3 | 60.5 | 2.1 | 53.3 | 0.4 |
| redWT[bc] | 67.8 | 0.4 | 63.8 | 0.4 | 62.0 | 1.3 | 57.4 | 0.3 |
| redWT[d] | 79.0[e] | 0.7 | nd[f] | | nd | | nd | |
| C49S | 62.8 | 2.0 | 59.7 | 0.2 | 56.3 | 0.3 | 50.0 | 0.3 |
| redC49S[b] | 63.5 | 0.8 | 59.6 | 0.2 | 58.1 | 2.2 | 49.9 | 0.3 |
| C104S | 76.3 | 1.1 | 72.1 | 0.3 | 64.9 | 2.1 | 59.3 | 0.2 |
| redC104S[b] | 76.7 | 1.8 | 73.0 | 0.4 | 62.2 | 1.1 | 59.3 | 0.3 |
| Y53E | 62.3 | 0.4 | nd | | nd | | nd | |
| Y53F | 62.6 | 0.3 | nd | | nd | | nd | |

[a]All temperatures reported are in °C
[b]For CD, reducing conditions were achieved by the addition of 1 mM DTT. For SLS, reducing conditions were achieved by the addition of 10 mM DTT.
[d]Reduced PRL-1-WT (redWT) was studied in 20 mM phosphate buffer at pH 6.5 with equivalent ionic strength (I=0.1). All other values reported, including those for HEPES buffer, are at pH 7.5.
[e]Melting was incomplete at the maximum temperature attainable. To determine the $T_m$, the data were fit by constraining the final plateau to a constant value typically observed with the other redWT samples ($-15 \times 10^6$).
[f]Not determined (nd)

Interestingly, the C49S and C104S mutants each have a unique melting profile that is different from each other, as well as the reduced wild type. Compared to the chemically reduced wild type, C49S has a significantly lower melting temperature (62.8 ± 2.0 °C; p=0.0491), while the C104S mutant has a significantly higher melting temperature (76.3 ± 1.1 °C; p=0.0005). Such large differences between the C49S, C104S and reduced wild-type proteins were unexpected because, with only a single atom changed, each should mimic the reduced conformation of PRL-1. The fact that these mutants have a 13.5 °C separation in their physical stability strongly indicates that the overall stability of PRL-1 is greatly impacted by local interactions with the side chain moieties of the C49 and C104 residues. Similar trends in the aggregation propensity of PRL-1 variants were also observed by SLS. In phosphate buffer, the

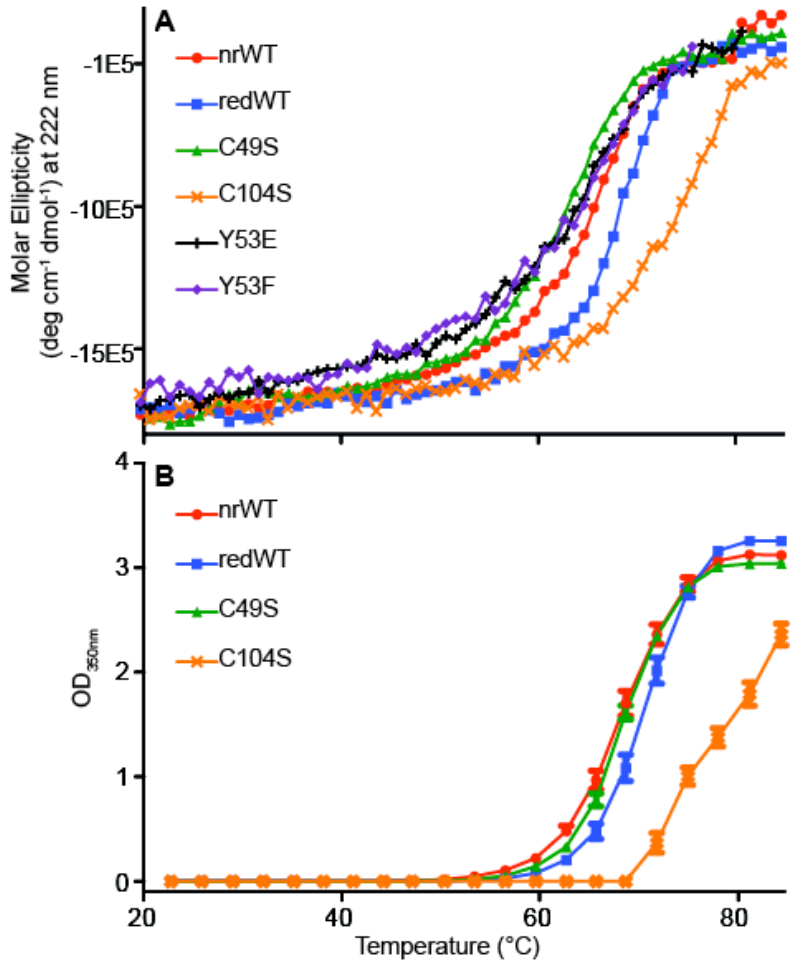



**Figure 5.1. Physical Stability of PRL-1**. The physical stability of PRL-1 was monitored by either **CD (A)** or **SLS (B)**. $T_m$s determined from the data shown here are summarized in Table 2. All data presented in this figure were collected in phosphate buffer at pH 7.5. For CD analysis, data points represent the average of two replicates of two independently prepared samples. For SLS analysis, data points represent the average of two replicates.

reduced wild type was slightly less vulnerable to aggregation than the oxidized wild type ($\Delta T_{onset}$=3.8 °C). The C104S protein is the least aggregation prone ($T_{onset}$=72.1 ± 0.3 °C), while the C49S protein self associates most readily ($T_{onset}$=59.7 ± 0.2 °C). These data clearly show that stability depends on more than just disulfide bond formation at the active site.

Table 5.2 shows that changing the buffer system from phosphate to HEPES reduces the physical stability of PRL-1 for all variants studied, decreasing $T_m$ values by 5-6 °C. Similarly, the aggregation propensity measured by SLS also decreases consistently by approximately 6 °C with the same buffer switch. Because the biological role of PRL-1 is to facilitate phosphate hydrolysis from a phosphotyrosine substrate, this result suggests that the decrease in physical stability can be attributed to whether or not the protein can bind to the buffer species, albeit too

weakly in this case to determine the $K_D$ even by NMR, placing it above the mid mM range (data not shown). The phosphate group is expected to bind at the active site of the protein, much like sulfate occupies the active site in the available crystal structure of C104S (Figure 5.2.A).[9,16] HEPES (sulfonate ion) apparently does not participate in such an interaction, and each variant is destabilized to an equivalent extent by this buffer switch. Because both buffers were prepared to have equivalent ionic strengths, it is unlikely that the effect results from non-specific electrostatic influences, further suggesting that weak phosphate binding imparts stabilization to this protein.

Active site binding also explains the dramatic difference in physical stability between the C104S and reduced wild-type proteins. The catalytic Cys in the reduced wild type exists as a thiolate anion at physiological pH, which is typical for this family of enzymes because of the

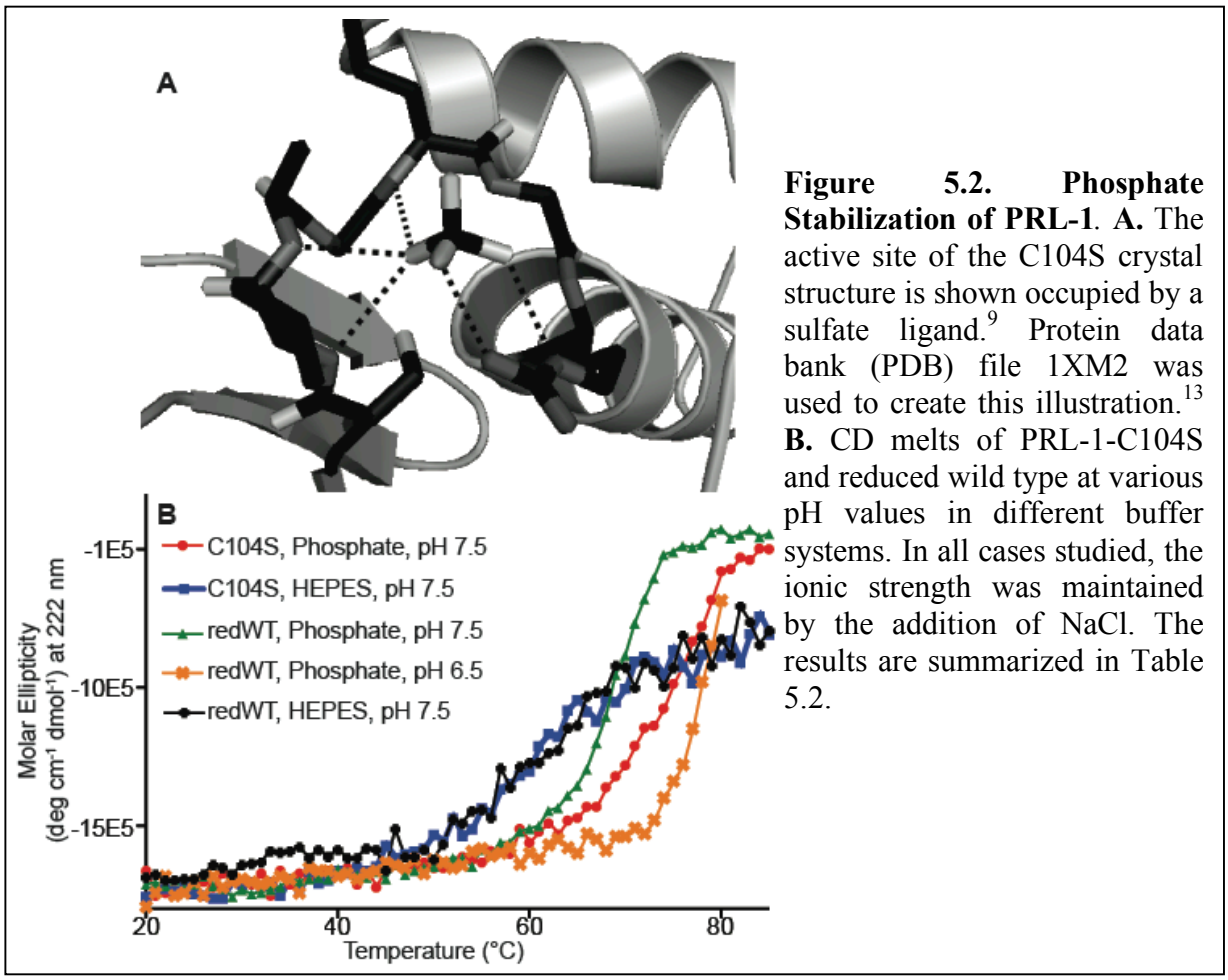



**Figure 5.2. Phosphate Stabilization of PRL-1**. **A.** The active site of the C104S crystal structure is shown occupied by a sulfate ligand.[9] Protein data bank (PDB) file 1XM2 was used to create this illustration.[13] **B.** CD melts of PRL-1-C104S and reduced wild type at various pH values in different buffer systems. In all cases studied, the ionic strength was maintained by the addition of NaCl. The results are summarized in Table 5.2.

surrounding partially positive microenvironment and structural organization of the active site.[16,32] With PRL-1-C104S, the thiolate side chain has been replaced with a hydoxyl group, which remains protonated, eliminating the negative charge at this position. Because less charge repulsion is present between phosphate and the –OH group from Ser than between phosphate and the negatively charged thiolate anion, binding by phosphate would occur to C104S more readily because of the differences in charge state of the active site. This conclusion is further supported by the fact that the melting temperatures measured by CD for the reduced wild type and C104S are statistically identical in HEPES buffer ($\Delta T_m$=2.9, p=0.4207) but differ significantly in phosphate buffer ($\Delta T_m$=8.5 °C, p=0.0005; Table 2). Although no statistical analysis could be conducted on the SLS data because of the small number of replicates, similar trends were observed with the SLS data. The $\Delta T_{onset}$ value for these two proteins in HEPES is considerably smaller than for phosphate buffer (1.9 versus 8.4 °C, respectively; Table 2), further indicating that phosphate binding imparts structural stability. As an alternative way of investigating the effect of charge repulsion on stability, the pH of the phosphate buffer was lowered to 6.5 from 7.5 while maintaining an equivalent ionic strength. This too enhanced the physical stability of the reduced wild type, as measured by CD ($T_m$=79.1 °C). The primary phosphate species at pH 6.5 has a -1 charge, whereas at 7.5, just over 50% of the phosphate has a -2 charge. The smaller degree of charge repulsion between the thiolate and singly charged phosphate ion allows the molecule to more readily occupy the active site and stabilize the protein. A comparison of the CD melts of these mutants in the different buffers is shown in Figure 5.2.B.

Phosphate binding does not fully explain the difference in structural stability of the reduced forms, as the $T_m$ values measured by CD and onset temperatures measured by SLS remain consistently varied for the different forms across each buffer system (Table 5.2). When

compared to the reduced wild-type protein in HEPES buffer, C49S consistently exhibits a lower melting temperature. This suggests that a specific local packing interaction involving C49 is perturbed in this protein that causes destabilization. The side chain of C49 is completely buried in the protein's hydrophobic core in the crystal structure of PRL-1-C104S. One notable feature in the crystal structure of the reduced state is the presence of a completely buried hydrogen bond between residues Y53 and H103 (Figure 5.3).[13,16] In the reduced state, this hydrogen bond forms between the hydroxyl from the tyrosine side chain and the imidazolium Nε of the histidine and has a bond length of 2.85 Å. The C49 side chain is pointed inwards towards this interaction. In the oxidized state, the distance between these atoms lengthens to 3.07 Å, and the angle (107°) further deviates from ideality (120°)[43] compared to the reduced form (113°). The C49 side chain is disulfide bound and points away from these residues. There is a concomitant rearrangement of residues in the region surrounding both the active site and the Y53-H103 hydrogen bond. Based on the small (2 °C) difference in $T_m$ between the reduced and oxidized wild-type states, the optimization of the hydrogen bond and concomitant packing rearrangement around it may compensate for the loss of the covalent disulfide bond. Based on this information, we hypothesized that the C49S mutant is destabilized because the introduction of the polar group favors greater solvent exposure and disrupts favorable packing interactions around the Y53-H103 hydrogen bond, despite being

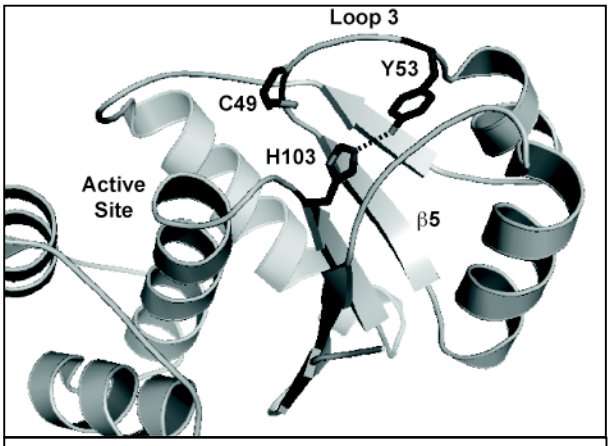



**Figure 5.3. Y53-H103 Hydrogen Bonding in PRL-1**. PDB code 1XM2 was used to illustrate the proximity of the Y53 and H103 side chains.[13] In the figure, the side chain of these residues, as well as C49 are shown as sticks. In the oxidized crystal structure (PDB 1ZCK),[16] the interaction between Y53 and H103 is perturbed.

in the reduced conformation at the active site. To test this hypothesis, we created the Y53E and Y53F mutants and studied their physical stability by CD. As indicated by Figure 5.1 and Table 5.2, Y53E and Y53F have equivalent $T_m$ values to the C49S mutant, suggesting a common mode of destabilization.

### 5.3.2. Structural Analysis of PRL-1

We first collected CD spectra of each PRL-1 variant to determine if substantial changes in the total secondary structure occur as the result of any of the mutations. The results are shown in Figure 5.4. Despite the significant deviations in physical stability between the reduced wild type, C49S and C104S proteins, the total secondary structure content of each mutant remains very similar. Error bars were omitted in Figure 5.4 for presentation purposes. The values below 203 nm are within the standard deviation of each curve, whereas values in the α-helix and β-sheet regions are statistically different. These differences are disproportionately small compared to the differences in stability. Unfortunately, there is no systematic way to extract information from this data set to explain why the $T_m$ values differ so substantially. High-resolution methods are needed to provide site-specific information capable of explaining

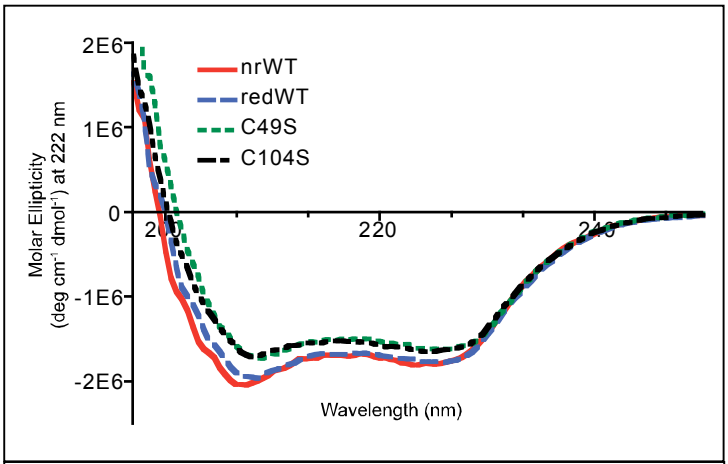



**Figure 5.4. CD Spectra of PRL-1 Variants**. The secondary structure elements of the various mutants were analyzed by monitoring the CD signal from 190-250 nm. The data points represent the average of two replicates of two independently prepared samples (n=4). Error bars were omitted for figure quality purposes. All data points below 203 nm are statistically identical. At 208, 218 and 222 nm, where absorbance for α-helix and β-sheet are commonly observed, the values are statistically different.

the mechanism by which stabilization or destabilization is achieved with the introduction of various mutations or changes in solution conditions. We performed high-resolution solution NMR studies to obtain such information.

The $^{1}H$-$^{15}N$ HSQC NMR experiment was used because this method selectively detects protons that are directly coupled to a nitrogen atom. This is particularly useful for the analysis of protein structure because of the repeating pattern of amides in the polypeptide backbone. A cross peak for each amide in the protein is observed, providing a fingerprint of the protein's unique structure. This experiment serves as the basis for assessing the overall fold and dynamic behavior of a protein. Because chemical shift is so sensitive to an atom's local environment, changes in a protein's tertiary structure, such as those introduced by point mutations, are manifested as changes in the position and/or shape of the peaks in the HSQC spectrum.[44] For this analysis, backbone NMR assignments for reduced PRL-1, which we previously published, were used to assign the respective mutants and provide site-specific information.[35] Because of the large number of changes between the reduced wild-type and C49S proteins (see below), independent backbone assignments for C49S were also made to identify additional peaks that could not be deduced using the reduced PRL-1 spectra.[41]

To determine which residues are perturbed by mutations that affect disulfide bond formation, we first compared the HSQC spectrum of the chemically reduced wild-type protein with that of C104S (Figure 5.5.A). Although many of the peaks overlap, there are a number of significant changes observed between these two proteins, some of which are highlighted by boxes in the figure. To illustrate where these residues are located with respect to the mutation site, we mapped these changes to the available C104S crystal structure (Figure 5.5.B).[13] The largest chemical shift deviations (blue) are observed in or near the active site (V47, W68, V102-

C104, G109, R110, F141), which is expected because the catalytic cysteine has been modified. The active site is illustrated in ball and stick format for reference. Other small changes (red) are observed near the active site and include H23, N24, R47, E50, F70, A74 and V81. Additional smaller changes are present in α4-α6 (V113-V115, I133, G139-A140, L146) because a hydrogen bond between the active site and small loop containing F141 is perturbed. This change propagates, affecting the β1-containing residues V10-Y14 and β2-containing I21. On the opposite side of the active site, there are perturbations in two loops. Because the amides from these residues are already solvent exposed, the change in their chemical shift position is small. Residues Y53 and D54 in the flexible β2-α3 loop (loop three; L3) are slightly perturbed, as well as T26, which is located in loop two (L2) directly adjacent to L3. The perturbation to the Y53 backbone is likely a result of a change in conformational dynamics revolving around the orientation of H103, which results from active site stabilization upon phosphate binding. Overall, when comparing the NMR spectra of reduced wild type to that of C104S, greater than 97% of the changes observed originate from the active site. This strongly suggests that a major component of the stability difference between the chemically reduced wild type and C104S proteins is due to phosphate binding and the diminished charge repulsion at the active site in the hydroxyl-containing mutant. Moreover, previous unpublished work by the Stauffacher lab indicates that increasing the concentration of phosphate in solution sharpens NMR line widths throughout the protein but more so for the resonances associated with the active site. Because broadening of lines likely reflects conformational averaging, this data suggests that phosphate binding also shifts the dynamic equilibrium of the whole protein toward a more rigid state (data not shown).

Despite the inability of the disulfide bound protein to bind phosphate at the active site,

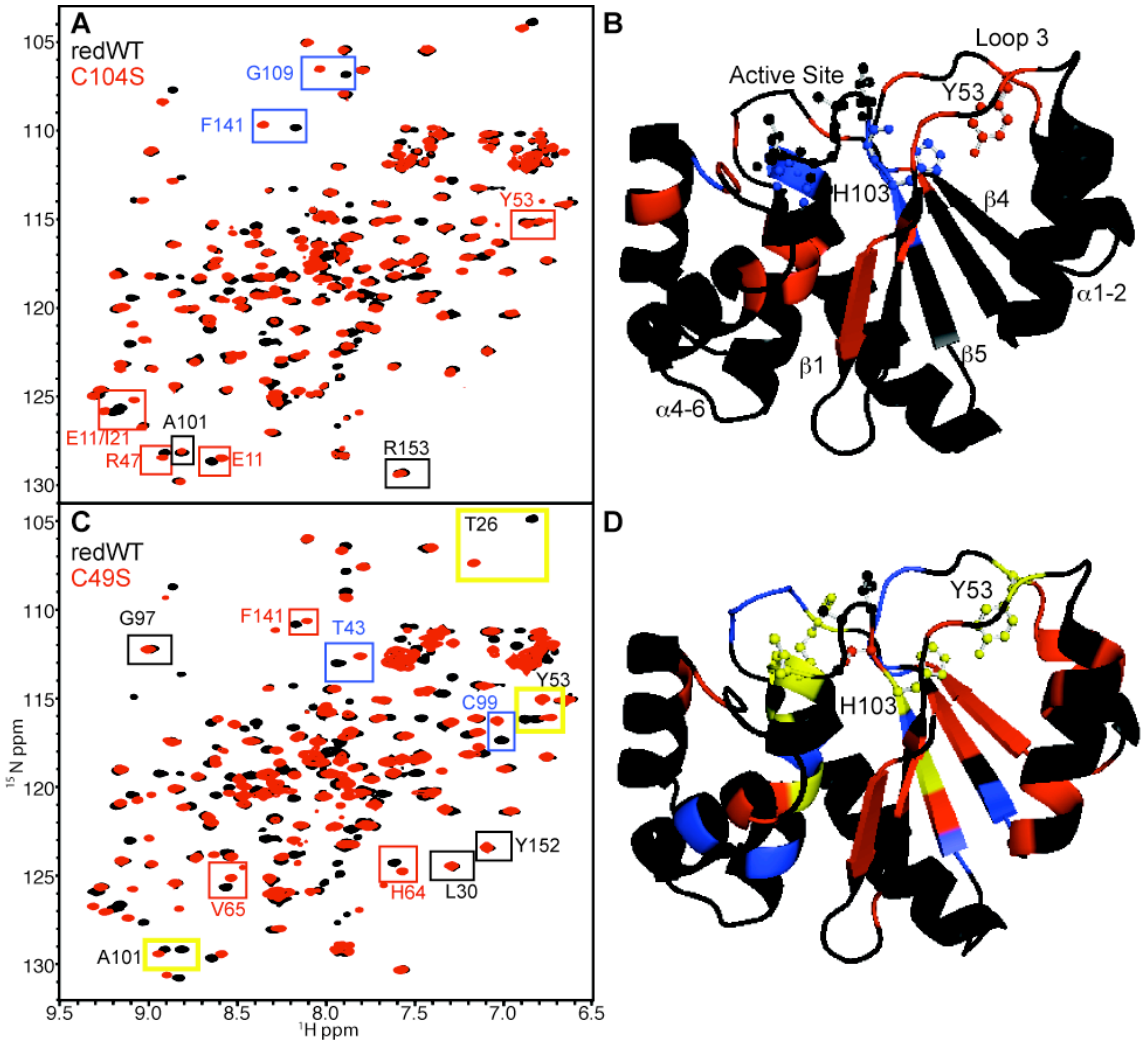



**Figure 5.5. Chemical Shift Mapping of PRL-1. A**. $^{1}$H-$^{15}$N HSQC spectra of reduced of wild type (black) and C104S (red) are shown. Chemical shift changes ($\Delta f$) between the two spectra were analyzed as described in the methods section. A representative example from each category described below is boxed in the spectra. Black boxes highlight residues with $\Delta f$ < 33.8. Red boxes emphasize residues with 33.8 < $\Delta f$ < 100 Hz. Blue boxes specify residues with 300 Hz < $\Delta f$ > 100 Hz. ***B***. Chemical shift changes between reduced wild type and C104S were mapped to the available C104S (PDB 1XM2)[16] crystal structure and color-coded as in A. The active site, Y53 and H103 residues are depicted as ball and sticks. Relevant regions of the protein are also labeled. **C**. $^{1}$H-$^{15}$N HSQC spectra of reduced wild type (black) and C49S (red) with colored boxes highlighting chemical shift change from the corresponding categories. Additionally, residues with $\Delta f$ > 300 Hz are specified by yellow boxes with black lettering. **D**. As in B, except changes between the reduced wild type and C49S are mapped. The same chemical shift change categories used in panel C were used in panel D (no change=black; small change=red; medium change=blue; large change=yellow).

the choice of buffer still affects the stability of the oxidized species such that phosphate is more stabilizing. While studies to explain this phenomenon are beyond the scope of this work, a secondary phosphate interaction site has already been proposed. The C-terminal tail of PRL-1 contains a polybasic region, which has been shown to be important for binding to phospholipid membranes.[45] The stabilization of oxidized wild type by phosphate may be a result of an interaction with this region of the protein. These residues are missing from the crystal structures but large perturbations corresponding to these residues in the NMR spectra occur upon active site oxidation,[35] suggesting they are organized differently in the two states.

We next compared the NMR spectrum of the chemically reduced wild-type protein to that of C49S (Figure 5.5.C). Surprisingly, there are far more changes between these two protein forms than observed between the chemically reduced wild type and C104S. Representative large (yellow), medium (blue) and small (red) changes are highlighted with corresponding colored boxes in the figure. Again for visualization, the changes were mapped to the available C104S crystal structure with each residue colored according to the appropriate chemical shift change category (Figure 5.5.D).[13] Many of the largest chemical shift changes are located in or near the active site (V48-E50, W68, F70-A74, H103, C104, L108-R110). Similar to the C104S-reduced wild type comparison, several additional perturbations to α4-α6 are observed (L114-L117, L119, Q131, Q135, F141 and Q145). These changes are again propagated to β1 and β2 (V10-V12). Because changes in α4-α6, β1 and β2 are confined to the internal face and are relatively small across variants with differing stabilities, we conclude that this region of the protein does not appear to correlate with aggregation propensity. Loop three containing Y53 is also altered. Interestingly, in each region of the protein described above, the number and magnitude of chemical shift changes are larger in C49S, and the specific residues involved in some cases

differ. For example, Y53 and nearby residues, especially T26, are perturbed to a much greater extent with the C49S mutation than with the C104S. The $^{1}$H line width of the Y53 peak increases from 22.7 to 53.9 Hz (137% increase), despite the fact that the average line width for all C49S peaks is approximately the same as that for the reduced wild type (33.05 and 31.8 Hz, respectively). This data supports our hypothesis that burial of the C49 side chain stabilizes the Y53-H103 interaction and incorporation of a polar moiety at this position increases conformational flexibility in this region.

To further probe the role of the hydrogen bond between Y53 and H103, we eliminated the interaction by substituting Phe for Tyr. An NMR spectrum of Y53F was collected, and the broad line widths of the peaks indicate that this protein exists in a molten globule-like state. With Y53F, the Phe side chain is likely exchanging between a buried and exposed conformation of loop three, and consequently, the conformation of this residue and the surrounding residues is not restricted to a specific organization (Figure 5.6). Because the entire set of peaks is broadened to a similar extent in this spectrum, it appears that the Y53-H103 hydrogen bond organizes a

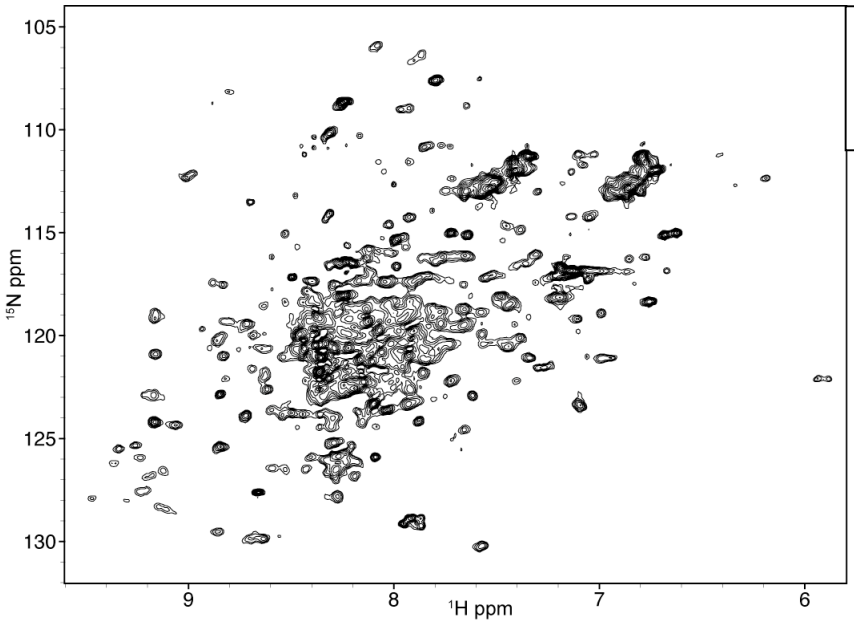



**Figure 5.6. $^{1}$H-$^{15}$N HSQC spectrum of PRL-1-Y53F.**

particular conformation of the protein, which has increased physical stability. Y53E was made because it was expected that mutation to Glu would favor a more exposed state for this region, due to the introduction of a charged moiety. The NMR spectrum of Y53E was collected and compared with that of C49S (Figure 5.7). Overall, the NMR spectrum of Y53E has many more peaks than would be predicted for a single stable conformation, and many of the peaks in the central portion of the spectrum are overlapped and unresolved. Based on the line widths, it appears that a mixture of two or more distinct conformations exists, possibly a set of more open conformers that undergo slow conformational exchange on the NMR timescale. Despite this complexity, critical portions of the Y53E spectrum are remarkably similar to C49S. Of importance is the fact that the peaks for Y53 and E53 are virtually identical. This information supports our hypothesis that the C49S mutation also alters the status of the Y53-H103 hydrogen bond. To eliminate the possibility that mutation of Y53 might alter the redox state of the protein, resulting in the increased number of peaks, reducing agent was added to the sample. The addition

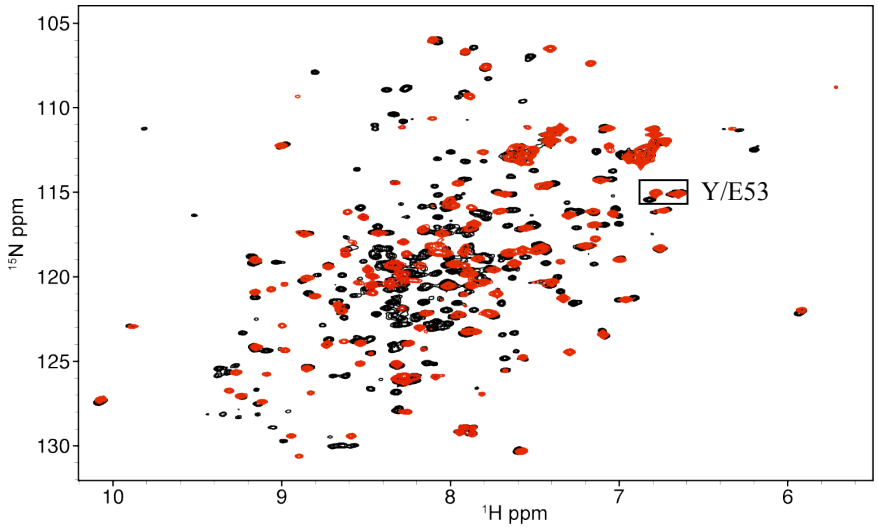



**Figure 5.7. Structural Comparison of C49S and Y53E**. The $^{1}$H-$^{15}$N HSQC spectrum of PRL-1-C49S (red) is overlaid with that of PRL-1-Y53E (black). The number of peaks in the Y53E spectrum indicates that Y53E samples multiple conformations in solution. One of these conformations is similar to the structure of C49S. Although, the chemical shift difference for Y/E53 is significant, the change is rather small (compare to Figure 5C). The small difference is likely a direct result of the side chain substitution.

of a reducing agent does not alter the appearance of the Y53E (or Y53F) spectrum in any way (data not shown), indicating that the protein exists as multiple conformers independent of disulfide bond formation. As such, the formation of the disulfide bond appears to depend on the presence of the Y53-H103 bond, and the local packing organization imparted to the active site and L3 are a result of this interaction.

Based on the above data and the assumption that the charge state of the active site for C49S and chemically reduced wild type are approximately the same, it seems that the active site is coupled to the rest of the protein through H103 via its hydrogen bond to Y53 in L3. Although more extensive structural analysis is needed to fully explain the cooperativity, the β-sheet also packs against the large hydrophobic surface generated by the Y53-H103 hydrogen bond. An unusually large number of substantial differences appear in the central β-strand (β5) located in the hydrophobic core of the protein variants. A101 is the central residue of this structural feature, and the C49S mutation causes a considerably larger change in frequency than does C104S for this residue when compared to the chemically reduced wild type ($\Delta\delta_{1H}$ and $\Delta\delta_{15N}$ for C49S are 0.116 and 2.54 ppm, respectively; $\Delta\delta_{1H}$ and $\Delta\delta_{15N}$ for C104S are 0.000 and 0.073 ppm, respectively). To better illustrate the differences between C104S and C49S, the chemical shift changes observed between the NMR spectra of these two proteins were again mapped to the structure. The results are shown in Figure 5.8. As can be observed in Figure 5.8, the largest changes are observed in the central strand (β5) with residues C99, A101, V102 and H103. The remaining blue change is V42 in β2, and smaller changes are observed for all residues in β4. These changes may be a reflection of the altered packing arrangement of surrounding residues and/or dynamics in L3 where both C49 and Y53, which hydrogen bonds to H103 positioned at the edge of the active site in β5, are located. C49S, which lacks both the Y53-H103 hydrogen

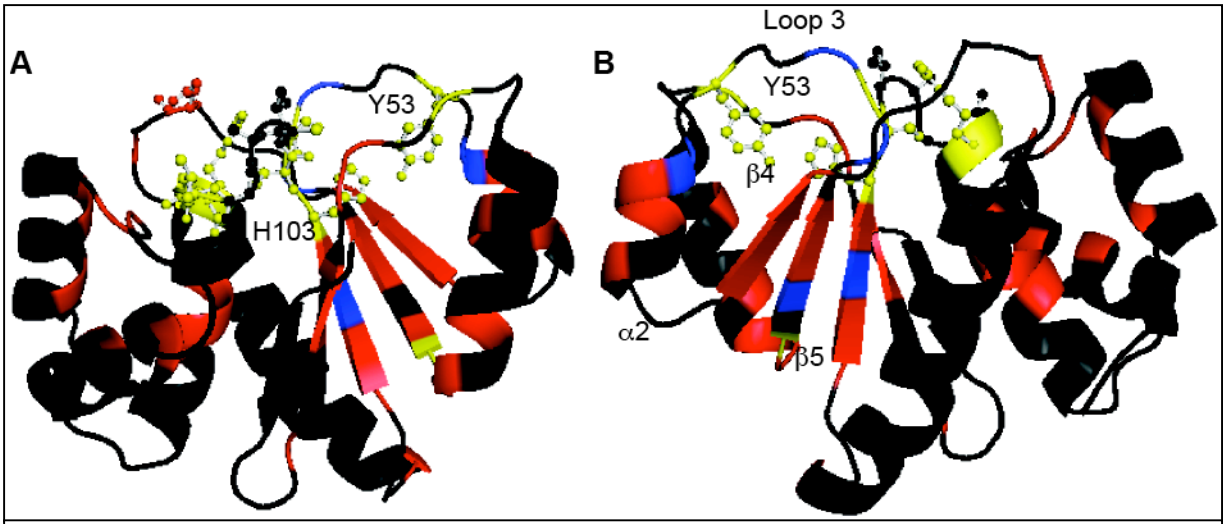



**Figure 5.8. Chemical Shift Mapping of PRL-1**. **A**. Chemical shift changes ($\Delta f$) between the C104S and C49S proteins is shown with the following color categories: no change=black; small change=red; medium change=blue; large change=yellow. Changes were mapped to the available C104S (PDB 1XM2)[16] crystal structure. The NMR spectra are not shown. **B.** As in A, except the molecule is rotated by 180° along the Y axis to better illustrate the differences observed to α1.

bond and C49-C104 disulfide bond, is consistently less stable than all other forms studied, as indicated by its dramatically lower melting temperature in both phosphate and HEPES and increased propensity to aggregate. As such, increased protein aggregation is correlated with increased flexibility in this region of the protein. This is supported by the fact that the oxidized wild type, which contains the disulfide bond but lacks the hydrogen bond, is more stable than C49S (Table 5.3).

In this context,[22] the disulfide bond increases the stability of the protein by restricting the

**Table 5.3. Effects of Specific Bonds on PRL-1 Stability**[a]

| | **Decreasing Stability ⇒⇒⇒** | | | |
|---|---|---|---|---|
| | C104S | Reduced WT | Oxidized WT | C49S |
| Y53-H103 hydrogen bond | Yes | Yes | No | No |
| C49-C104 disulfide bond | No | No | Yes | No |
| Active site $PO_4$ binding | Yes | Partial | No | Partial |
| $T_m$ (pH 7.5) | 76.3 | 67.8 | 65.5 | 62.8 |

[a]All temperatures reported are in °C

conformational flexibility of L3 in a different manner. When comparing the reduced and oxidized wild type proteins,[35] similar changes to the central β-sheet are observed, as well as large changes in α-helix two (α2). Similar changes to α2 are detected when comparing C49S to reduced wild type and/or C104S and can be seen in Figures 5.5.D and 5.8. Because β4 is composed of entirely hydrophobic residues and such sequences are prone to aggregation,[46] we hypothesize that stable packing of L3 and α2 provide protection in the reduced state against edge strand initiated aggregation. Formation of the disulfide bond causes the Y53-H103 bond to become distorted and L3 consequently becomes more flexible, leading to slightly increased aggregation propensity in the oxidized state. Complete disruption of the hydrogen bond, as is the case with C49S, may lead to unraveling of α2, and the increased susceptibility to aggregation may be caused by a larger degree of β-sheet exposure. Many other groups have found that the proportion of β-sheet structure increases in protein aggregates, while α-helical content is diminished.[47,48] Interestingly, full spectral CD analysis reveals that much of the β-sheet structure remains intact at the onset temperature. Although only a representative plot of C49S is shown in Figure 5.9, similar results were seen for the other PRL-1 mutants. This hypothesis is further

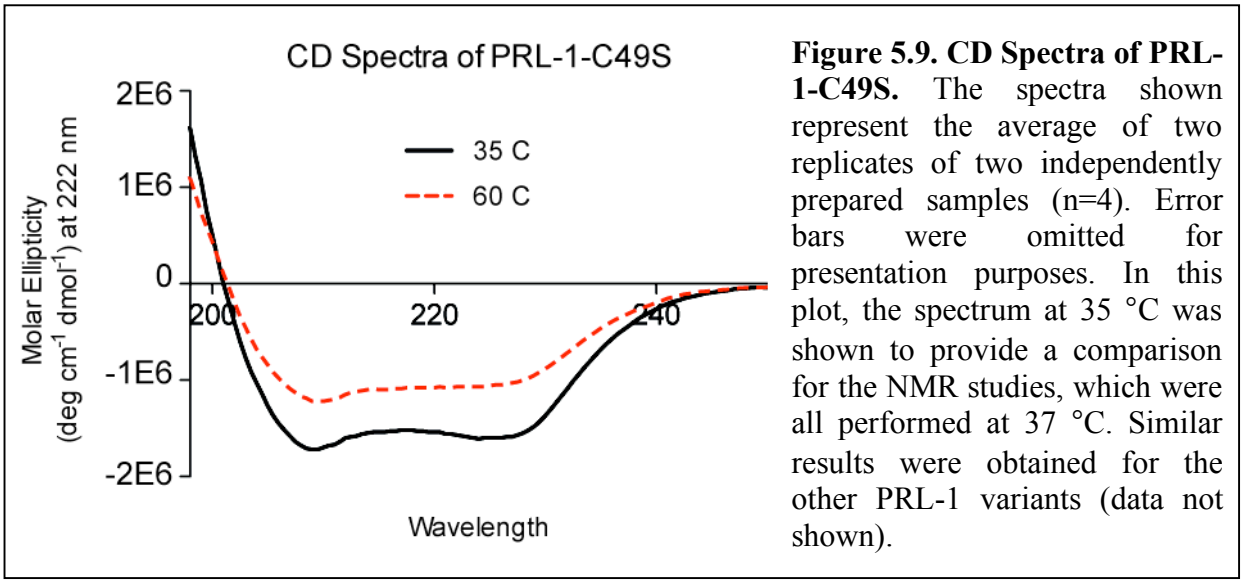



**Figure 5.9. CD Spectra of PRL-1-C49S.** The spectra shown represent the average of two replicates of two independently prepared samples (n=4). Error bars were omitted for presentation purposes. In this plot, the spectrum at 35 °C was shown to provide a comparison for the NMR studies, which were all performed at 37 °C. Similar results were obtained for the other PRL-1 variants (data not shown).

supported by the results of the H-D exchange NMR experiments, in which labile protons in the protein are exchanged with deuterons when the protein is dissolved in $D_2O$. Only protonated groups generate signal in the NMR experiment, and the loss of intensity for each chemical shift can be monitored to determine the rate of exchange at individual positions. These rates will differ depending on the location of the amide in the protein and the interactions in which it participates. For example, based on the crystal structure, residue T56 is located near the edge of α2 and the amide of this residue does not participate in hydrogen bonding. As expected, the exchange rate at this position is very fast for all three protein forms. In contrast, W84 is located in the center of α3 and its amide hydrogen participates in hydrogen bonding with the carbonyl in the i-4 position. The amide for this residue is completely retained for all three variants after 24 hours. As another example, the NH of L66 is pointed towards the interior of the protein and hydrogen bonds with the carbonyl of I45 in β3. The amide should be solvent inaccessible, and as anticipated, the H-D exchange data indicates that the NH of L66 is completely retained after 24 hours as well. Although the vast majority of amides exchanged either too slowly or quickly to obtain quantitative values, F92, which is located in α3 close to the C-terminal end of the helix, exchanged at a measureable rate. The data for this residue was analyzed and is shown because the exchange rate for this amide could be determined for all three proteins. Only very small differences are observed among the three variants indicating that this portion of the structure behaves similarly and experiences no structural or dynamic perturbations with either Cys mutation. This data is summarized in Table 5.4.

In Table 5.4, the amide half-lives determined from the exponential fit of the signal decay from 0-24 hours for a few representative residues are shown. The exchange rate values, although quantitatively determined, are a qualitative way of identifying trends in structural stability among

**Table 5.4. H-D Exchange Data**[a]

| | Protein Region | C49S | redWT | C104S |
|---|---|---|---|---|
| T56 | α2 | ++[b] | ++ | ++ |
| L57 | α2 | ++ | 1.4 | 2.1 |
| V58 | α2 | 26.8 | 34.2 | --[c] |
| H64 | β4 | 26.7 | 42.0 | 54.0 |
| V65 | β4 | ++ | ++ | 2.1 |
| L66 | β4 | -- | -- | -- |
| D67 | β4 | ++ | 2.1 | 3.8 |
| W84 | α3 | -- | -- | -- |
| F92 | α3 | 28.2 | 24.7 | 21.8 |

[a]All data are reported as half-lives of NH peak intensity in minutes.
[b]Could not be fit because the NH exchanges completely within 5 minutes.
[c]Could not be fit because NH is completely retained after 24 hours

the PRL-1 variants at specific sites in the protein. According to the crystal structures, residue V58 is the central residue in α2 and the amide hydrogen should be buried in a hydrogen bond. It therefore should be less solvent accessible and exchange more slowly than random coil if participating in α-helix interactions. As can be seen with Table 5.4, the half-life for this amide in C49S is shorter compared to the reduced wild type, and reduced wild type is shorter than the C104S protein. A similar trend is observed for L57, although these differences are very small. When inspecting residues located in β4, the edge strand, similar trends are observed, especially for residues H64 and, to a lesser extent, D67. This data indicates these residues are more solvent exposed in C49S than in the reduced wild type or C104S protein and parallels the trends observed with our reported $T_{onset}$ and $T_m$ values. Based on these observations, it is likely that unfurling of the structure at L3 and α2 exposes the β-sheet and may permit edge-strand association to initiate aggregation. Because large protein aggregates cannot be observed using solution NMR, the aggregation mechanism will have to be investigated further using other methods.

It is particularly interesting that the hydrogen bond between Y53 and H103 is better able to enhance stabilization of the protein than the disulfide bond. Other incidents of hydrogen bonds contributing substantially to the stability of proteins have been documented.[28,49,50] The methodologies used in these studies, however, could not provide information about cooperative or dynamic influences resulting from disruption of the hydrogen bond and compensation effects

could not be observed. The high-resolution approach used here shows that modification of the hydrogen bond causes local perturbations as well as increased dynamics in coupled regions. For example, because disruption of the Y53-H103 bond in PRL-1 indirectly leads to drastic changes in the central β-sheet, packing of these structural features must be interrelated.

Reduction of the disulfide bond in PRL-1 causes more than 90% of the NMR chemical shifts to change significantly, without drastically altering the secondary structure elements.[35] This structural transition probably reflects changes not only to the conformation of the active site but also the interaction between Y53 and H103 and cooperative hydrophobic packing interactions with the β-sheet in the protein core. Although the solution NMR data reveal that substantial conformational changes occur in solution between the reduced and oxidized states,[35] such large differences are not apparent in the crystal structures. The majority of differences are located near the active site when comparing the two crystallized proteins, and these differences are obvious in the NMR data as well. What is not apparent in a comparison of the crystal structures is the importance of the small rearrangements between the β-sheet, L3 and α2 in the two states. Based on the crystal data, one would predict that the hydrophobic packing interactions in the reduced and oxidized states are largely the same,[13,16] but the NMR data clearly show otherwise. One explanation for this discrepancy may be the absence of several residues from the C-terminus in the crystallized proteins. Crystallized PRL-1 was truncated at residues 156 and 160 for the reduced and oxidized forms, respectively. Although the full-length protein has not been crystallized, studies from our lab previously showed that the missing residues, which compose the polybasic sequence, modulate the structure and activity of PRL-1, such that their removal abrogates disulfide bond formation.[35] This result indicates an additional unique coupling exists in the oxidized protein that involves the tail residues.

### 5.3.3. Concluding Remarks

This study shows that stabilization or destabilization of a protein can be achieved in the folded state through the collaboration of specific, local interactions. The strategies presented here illustrate the utility of NMR in pharmaceutical research, as they can be applied to therapeutically relevant proteins to identify stabilizing forces within the structure. Identification of these forces at atomic resolution will facilitate the rational design of improved protein therapeutics and biotechnology products. Furthermore, this work demonstrates that the effects of the oxidation state of cysteine residues on a protein's stability derive from the overall context in which they reside. Understanding this relationship is valuable because it may guide disulfide bond engineering in protein pharmaceuticals.

## 5.4. REFERENCESB


1. Volkin D. B., Mach H. & Middaugh C. R. (1997) Degradative covalent reactions important to protein stability. *Mol Biotechnol*, 8, 105-122.
2. Manning M. C., Patel K. & Borchardt R. T. (1989) Stability of protein pharmaceuticals. *Pharm Res*, 6, 903-918.
3. Chang S. G., Choi K. D., Jang S. H. & Shin H. C. (2003) Role of disulfide bonds in the structure and activity of human insulin. *Mol Cells*, 16, 323-330.
4. Matsumura M., Becktel W. J., Levitt M. & Matthews B. W. (1989) Stabilization of phage T4 lysozyme by engineered disulfide bonds. *Proc Natl Acad Sci*, 86, 6562-6566.
5. Fontana A. (1991) Analysis and modulation of protein stability. *Curr Opin Biotechnol*, 2, 551-560.
6. Hagihara Y., Mine S. & Uegaki K. (2007) Stabilization of an immunoglobulin fold domain by an engineered disulfide bond at the buried hydrophobic region. *J Biol Chem*, 282, 36489-36495.
7. Saerens D., Conrath K., Govaert J. & Muyldermans S. (2008) Disulfide bond introduction for general stabilization of immunoglobulin heavy-chain variable domains. *J Mol Biol*, 377, 478-488.
8. Gong R., Vu B. K., Feng Y., Prieto D. A., Dyba M. A., Walsh J. D., Prabakaran P., Veenstra T. D., Tarasov S. G., Ishima R. & Dimitrov D. S. (2009) Engineered human antibody constant domains with increased stability. *J Biol Chem*, 284, 14203-14210.
9. Zhang Z.-Y. (2003) Chemical and mechanistic approaches to the study of protein tyrosine phosphatases. *Acc Chem Res*, 36, 385-392.
10. Betz S. F. (1993) Disulfide bonds and the stability of globular proteins. *Protein Sci*, 2, 1551-1558.
11. Pace C. N., Grimsley G. R., Thomson J. A. & Barnett B. J. (1988) Conformational stability and activity of ribonuclease T1 with zero, one, and two intact disulfide bonds. *J Biol Chem*, 263, 11820-11825.
12. Robinson C. R. & Sauer R. T. (2000) Striking stabilization of Arc repressor by an engineered disulfide bond. *Biochem*, 39, 12494-12502.
13. Jeong D. G., Kim S. J., Kim J. H., Son J. H., Park M. R., Lim S. M., Yoon T.-S. & Ryu S. E. (2005) Trimeric structure of PRL-1 phosphatase reveals an active enzyme conformation and regulation mechanisms. *J Mol Biol*, 345, 401-413.
14. Wedemeyer W. J., Welker E., Narayan M. & Scheraga H. A. (2000) Disulfide bonds and protein folding. *Biochem*, 39, 4207-4216.
15. Bhattacharyya R., Pal D. & Chakrabarti P. (2004) Disulfide bonds, their stereospecific environment and conservation in protein structures. *Protein Eng Des Sel*, 17, 795-808.
16. Sun J.-P., Wang W.-Q., Yang H., Liu S., Liang F., Fedorov A. A., Almo S. C. & Zhang Z.-Y. (2005) Structure and biochemical properties of PRL-1, a phosphatase implicated in cell growth, differentiation, and tumor invasion. *Biochem*, 44, 12009-12021.
17. Fu X., Li W., Mao Q. & Chang Z. (2003) Disulfide bonds convert small heat shock protein Hsp16.3 from chaperone to a non-chaperone: implications for the evolution of cystein in molecular chaperones. *Biochem Biophys Res Commun*, 308, 627-635.
18. Hogg P. J. (2003) Disulfide bonds as switches for protein function. *Trends Biochem Sci*, 28, 210-214.

19. Ishikawa H., Kim S., Kwak K., Wakasugi K. & Fayer M. D. (2007) Disulfide bond influence on the protein structural dynamics probed with 2D-IR vibrational echo spectroscopy. *Proc Natl Acad Sci*, 104, 19309-19314.
20. Zavodszky M., Chen C.-W., Huang J.-K., Zolkiewski M., Wen L. & Krishnamoothi R. (2001) Disulfide bond effects on protein stability: designed variants of Cucurbita maxima trypsin inhibitor-V. *Protein Sci*, 10, 149-160.
21. Ogawa K., Sonoyama T., Takeda T., Ichiki S., Nakamura S., Kobayashi Y., Uchiyama S., Nakasone K., Takayama S. J., Mita H., Yamamoto Y. & Y. S. (2007) Roles of a short connecting disulfide bond in the stability and function of pyschrophilic Shewanella violacea cytochrome c5. *Extremophiles*, 11, 797-807.
22. Santiveri C. M., Leon E., Rico M. & Jimenez M. A. (2008) Context-dependence of the contribution of disulfide bonds to β-hairpin stability. *Chemistry*, 14, 488-499.
23. Ciaccio N. A. & Laurence J. S. (2009) Effects of disulfide bond formation and protein helicity on the aggregation of activating transcription factor 5. *Mol Pharm*, Epub.
24. Thirumangalathu R., Krishnan S., Ricci M. S., Brems D. N., Randolph T. W. & Carpenter J. F. (2009) Silicone oil- and agitation-induced aggregation of a monoclonal antibody in aqueous solution. *J Pharm Sci*, 98, 3167-3181.
25. Kamerzell T. J., Ramsey J. D. & Middaugh C. R. (2008) Immunoglobulin dynamics, conformational fluctuations, and nonlinear elasticity and their effects on stability. *J Phys Chem B*, 112, 3240-3250.
26. Esfandiary R., Kickhoefer V. A., Rome L. H., Joshi S. B. & Middaugh C. R. (2009) Structural stability of vault proteins. *J Pharm Sci*, 98, 1376-1386.
27. Fan H. F., Ralston J., Dibiase M., Faulkner E. & Middaugh C. R. (2005) Solution behavior of IFN-β-1a: An empirical phase diagram based approach. *J Pharm Sci*, 94, 1893-1911.
28. Thurlkill R. L., Grimsley G. R., Scholtz J. M. & Pace C. N. (2006) Hydrogen bonding markedly reduces the pK of buried carboxyl groups in proteins. *J Mol Biol*, 362, 594-604.
29. Kuwata K., Hoshino M., Era S., Batt C. A. & Goto Y. (1998) α to β transition of β-lactogobulin as evidenced by heteronuclear NMR. *J Mol Biol*, 283, 731-739.
30. Kuwata K., Shastry R., Cheng H., Hoshino M., Batt C., Goto Y. & Roder H. (2001) Structural and kinetic characterization of early folding events in β-lactoglobulin. *Nature Struct Biol*, 8, 151-155.
31. Sakurai K., Konuma T., Yagi M. & Goto Y. (2009) Structural dynamics and folding of β-lactoglobulin probed by heteronuclear NMR. *Biochim Biophys Acta*, 1790, 527-537.
32. Chiarugi P. & Buricchi F. (2007) Protein tyrosin phosphorylation and reversible oxidation: two cross-talking posttranslation modifications. *Antiox Redox Signal*, 9, 1-24.
33. Laurence J. S., Hallenga K. & Stauffacher C. V. (2004) $^{1}H$, $^{15}N$, $^{13}C$ resonance assignments of the human protein tyrosine phosphatase PRL-1. *J Biomol NMR*, 29, 417-418.
34. Skinner A. & Laurence J. S. (2009) $^{1}H$, $^{15}N$, $^{13}C$ resonance assignments of the reduced and active form of human protein tyrosine phosphatase, PRL-1. *Biomol NMR Assign*, 3, 61-65.
35. Skinner A. L., Vartia A. A., Williams T. D. & Laurence J. S. (2009) Enzyme activity of phosphatase of regenerating liver is controlled by the redox environment and its C-terminal residues. *Biochemistry,* 48, 4262-4272.

36. Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R. D. & Bairoch A. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*, 31, 3784-3788.
37. Gill S. C. & von Hippel P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem*, 182, 319-326.
38. Delaglio F., Grezesiek S., Vuister G. W., Zhu G., Pfeifer J. & Bax A. (1995) NMRPipe: a multidimensonal spectra processing system based on UNIX pipes. *J Biomol NMR*, 6, 277-293.
39. Goddard T. D. & Kneller D. G. (2004) SPARKY. *University of California, San Francisco*.
40. Jaren O. R., Kranz J. K., Sorensen B. R., Wand J. & Shea M. A. (2002) Calcium iduced conformational switching of paramecium calmodulin provides evidence for domain coupling. *Biochem*, 41, 14158-14166.
41. Eghbalnia H. R., Bahrami A., Wang L., Assadi A. & Markley J. L. (2005) Probabilistic identification of spin systems and their assignments including coil-helix interference as output (PISTACHIO). *J Biomol NMR*, 32, 219-233.
42. Wishart D. S., Bigam C. G., Yoa J., Abildgaard F., Dyson H. H., Oldfield E., Markley J. L. & Sykes B. D. (1995) $^{1}H$, $^{13}C$ and $^{15}N$ chemical shift referencing in biomolecular NMR. *J Biomol NMR*, 6, 135-140.
43. Stickle D. F., Presta L. G., Dill K. A. & Rose G. D. (1992) Hydrogen bonding in globular proteins. *J Mol Biol*, 226, 1143-1159.
44. Cavanagh J., Fairbrother W. J., Palmer A. G., III, Rance M. & Skelton N. (2007) *Protein NMR Spectroscopy: Principles and Practice*. 2nd ed., Boston: Academic Press.
45. Sun J.-P., Luo Y., Yu X., Wang W.-Q., Zhou B., Liang F. & Zhang Z.-Y. (2007) Phosphatase activity, trimerization, and the C-terminal polybasic region are all required for PRL1-mediated cell growth and migration. *J Biol Chem*, 282, 29043-29051.
46. Richardson J. S. & Richardson D. C. (2002) Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc Natl Acad Sci*, 99, 2754-2759.
47. Petsko G. A. & Dagmar R. (2004) *Protein Structure and Function*. 1st ed. Stramford: Sinauer Associates. p 160-161.
48. Pan K.-M., Baldwin M., Nguyen J., Gasset M., Serban A., Groth D., Mehlhorn I., Huang Z., Fletterick R. J., Cohen F. E. & Prusiner S. B. (1993) Conversion of α-helices into β sheets features in the formation of the scrapie prion protein. *Proc Natl Acad Sci*, 90, 10962-10966.
49. Shirley B. A., Stanssens P., Hahn U. & Pace C. N. (1992) Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. *Biochem*, 31, 725-732.
50. Pace C. N., Hebert E. J., Shaw K. L., Schell D., Both V., Krajcikova D., Sevcik J., Wilson K. S., Dauter Z., Hartley R. W. & Grimsley G. R. (1998) Conformational stability and thermodynamics of folding of ribonucleases Sa, Sa2 and Sa3. *J Mol Biol*, 279, 271-286.

# CHAPTER 6.

# CONCLUSIONS, FUTURE WORK AND IMPLICATIONS FOR DRUG DESIGN

## 6.1. MAJOR CONCLUSIONS

The major conclusion of this work is that the PTPase activity of PRL-1 is controlled by redox conditions and is dramatically affected by modification of the C-terminus. Modification of the C-terminus confers activity through structural rearrangement of the protein and yields a conformation in which disulfide bond formation is no longer apparent. In the following two sections, the implications these findings have for PRL-1-based drug design and understanding of protein stability are discussed.

### 6.1.1. PRL-1 as a Novel Therapeutic Target

PRL-1 is a unique PTPase that has gained attention over the last decade because it promotes cancer and metastasis, when active.[1,2] Because PRL-1 is a PTPase and is subject to redox regulation, it follows that identification of a stable, constitutively active (reduced) form of PRL-1 is needed to facilitate the design of potential therapeutics by either high-throughput screening or structure-based drug design using solution NMR. We have clearly shown in Chapter 2 that two of the predicted posttranslational sites of PRL-1 dramatically affect the enzyme activity of PRL-1, such that farnesylation at the C-terminus likely promotes catalysis, while phosphorylation at Y53 would inhibit it, indicating that the C170S-C171S mutant provides a stable, reduced and active form of PRL-1 with which inhibition assays and screening for lead compounds for drug design can be pursued. Our data correlates well with previously published biological studies that showed inhibition of farnesylation abolishes PRL-1 mediated cell growth

and motility and causes mitotic defects.[3,4] This finding is also significant because the CaaX box motif is unique to the PRL enzymes and drug molecules that target the farnesylated form may provide a means of specificity without modifying the actions of other essential PTPases.

Direct confirmation that farnesylation activates the enzyme is needed to validate our model protein for drug design. *In vitro* conjugation of the farnesyl group and PRL-1 has been observed previously but has been difficult to reproduce.[5,6] Our lab has previously shown that PRL-1 contains a novel metal coordination center at the site of farnesylation (unpublished results)[*] and we suspect that metal binding interferes with the reaction. Although a number of metals have been found to coordinate to PRL-1, zinc coordination is likely biologically relevant because it is required for the farnesyl transfer reaction.[7] With this in mind, I have developed a new method for purifying recombinant PRL-1 with zinc coordinated at the C-terminus, the details of which are described in Appendix E. Attempts to purify PRL-1 without metal for the purpose of quantifying a metal binding constant and performing *in vitro* farnesyl conjugation were unsuccessful because the protein precipitates rapidly out of solution, indicating that the metal ion is an important structural feature. No other groups have reported PRL-1 metal binding. Others have reported difficulty obtaining crystals of the full-length form, but it is not certain whether metal binding was a factor in these studies or not.[8,9] We have not yet attempted to farnesylate Zn-bound PRL-1 because the reaction is quite complicated and the reagents are expensive. A thorough screening of assay conditions will be required to identify conditions that promote this modification and is well beyond the scope of this dissertation project.

---

[*] This work was completed by Anthony A. Vartia. Site-directed mutagenesis in combination with inductively coupled plasma mass spectrometry and UV-Vis spectroscopy was used to identify the metal coordination site, the species present (nickel, copper, zinc) and the most likely coordination geometry (square planar).

In addition to providing a model protein for drug screening assays, the biophysical and biochemical assays used to characterize the full-length and C-terminally modified proteins have provided, for the first time, a basis for explaining some of the varied findings regarding the role PRL-1 plays in controlled or aberrant growth processes. For example, although PRL-1 most frequently promotes cellular proliferation, it also likely plays a role in differentiation under certain conditions. For example, elevated levels of PRL-1 have been detected in differentiated tissues such as the intestinal villi,[10] confluent CaCo2 cells,[10] the outer segments of photoreceptors,[11] neurons and oligodendrocytes.[12] This is interesting with respect to redox conditions because differentiated cells are generally more oxidizing than proliferating ones. In fact, Yu *et al.* (2007) showed that oxidative stress induced the expression of PRL-1 in the retina and a photoreceptor-derived cell line.[13] Because the PRL-1 promotes several distinct phenotypic response (*i.e.* proliferation and differentiation) it is likely that this enzyme acts in multiple signaling pathways on a series of substrates and the cellular response may be modulated by the level of phosphatase activity or confinement of the protein to distinct subcellular locations. In Chapter 4, the reduction potential for the active site disulfide bond was calculated, which provides a basis for understanding how subcellular localization may regulate the phosphatase activity of PRL-1 in different types of cellular environments.[14] Determining the differences in the subcellular location and/or posttranslational processing of PRL-1 expressed in terminally differentiated versus proliferating tissues will generate a more accurate account of how PRL-1 is regulated *in vivo* and how its regulation contributes to such diverse cellular outcomes. This concept has yet to be explored by any group.

Interestingly, it is not clear whether the PTPase activity of PRL-1 is required for differentiation and it is possible that during differentiation, PRL-1 functions in a completely

unrelated way than it does during proliferation. In Chapters 3 and 4, it was shown that PRL-1 undergoes a major conformational change upon reduction (or modification to the C-terminus) in that greater than 90% of the amide resonances are perturbed to a significant extent.[15] When mapped to the available crystal structure, the perturbations are larger in number and magnitude in the central β-sheet located in the hydrophobic core of the protein (Figure 4.6). Such a drastic change in the protein's hydrophobic core is quite unanticipated. In general, the intrinsic dynamics of a protein predict that changes involving flexible loops or active-site residues occur, and changes of this nature are indeed observed in this study. Rearrangement of a protein's hydrophobic core, however, is rare.[16] Even proteins with large conformational changes share similar hydrophobic packing interactions.[9,17,18] Recently, a structural change that involves repacking of the protein core has been documented. In this example, virtually all of the tertiary contacts of the lyphotactin fold are replaced with new contacts and each of the two states have a unique function.[16] The structural transition in PRL-1 likely reflects changes not only to the conformation of the active site but also the interaction of Y53 and H103 and cooperative hydrophobic packing interactions with the β-sheet in the protein core, which was shown with the work described in Chapter 5. Additional, higher-resolution NMR studies are being pursued to further assess alterations in direct contacts between specific residues in the two states.

### 6.1.2. PRL-1 as a Model Protein to Study Protein Stability

Chapter 5 was dedicated to studying the effects of disulfide bond formation on the stability of PRL-1. The results of this study are significant because they show that that the integrity of the native PRL-1 structure does not depend on the absence or presence of the disulfide bond directly, but rather on the local packing interactions among nearby side chains and coupling between different regions. Each Cys-modulated PRL-1 variant has a unique physical

stability based on retention of specific features in its three-dimensional structure. The data indicate the reduced form of the protein compensates for disulfide bond cleavage by a compilation of adjustments focused around a central hydrogen bond between Y53 and H103. Structural instability and aggregation is correlated with flexibility in L3 (containing Y53), and C49S, which lacks both the Y53-H103 hydrogen bond and disulfide bond, is less stable than the oxidized wild type. In this context,[19] the disulfide bond increases the stability of the protein by restricting the conformational flexibility of L3 in a different manner. The work demonstrates that the effects disulfide bond formation has on a protein's stability derive from the overall context in which they reside, and understanding this relationship is valuable because it may permit a more rational approach to disulfide bond engineering in protein pharmaceuticals.

The other significant part of this work is how the *combination* of low-resolution and high-resolution methods gives a more detailed account of how stabilization is achieved than either method does alone. In this study, low-resolution methods (CD and SLS) were used to rapidly determine whether differences in stability between the various protein forms exist. These experiments are incredibly important because they facilitate rapid identification of stabilizing conditions for protein formulations. In this study, the hypotheses regarding the mechanism of protein stabilization or destabilization derived from low-resolution analysis were tested using high-resolution solution NMR. NMR has long been recognized as a valuable tool for structure determination and studying protein folding. In this work, however, NMR was used to correlate the observed changes in stability with specific perturbations in the structure of the folded protein to identify which residues are important for stabilization of the folded state. The strategies presented here illustrate the utility of NMR in pharmaceutical research, as they can be applied to therapeutically relevant proteins to identify stabilizing forces within the structure. Identification

of these forces at atomic resolution should further facilitate the rational design of improved protein therapeutics and biotechnology products.

## 6.2. REFERENCES


1. Stephens B. J., Han H., Gokhale V. & Von Hoft D. D. (2005) PRL phosphatases as potential molecular targets in cancer. *Mol Cancer Ther*, 4, 1653-1661.
2. Bessette D. C., Qiu D. & Pallen C. J. (2008) PRL PTPs: mediators and markers of cancer progression. *Cancer Metastasis Rev*, 27, 231-252.
3. Wang J., Kirby C. E. & Herbst R. (2002) The tyrosine phosphatase PRL-1 localizes to the the endoplasmic reticulum and the mitotic spindle ad is required for normal mitosis. *J Biol Chem*, 277, 46659-46668.
4. Sun J.-P., Luo Y., Yu X., Wang W.-Q., Zhou B., Liang F. & Zhang Z.-Y. (2007) Phosphatase activity, trimerization, and the C-terminal polybasic region are all required for PRL1-mediated cell growth and migration. *J Biol Chem*, 282, 29043-29051.
5. Cates C. A., Michael R. L., Stayrook K. R., Harvey K. A., Burke Y. D., Randall S. K., Crowell P. L. & Crowell D. N. (1996) Prenylation of oncogenic human PTPcaax protein tyrosine phosphatases. *Cancer Lett*, 110, 49-55.
6. Zeng Q., Si X., Horstmann H., Xu Y., Hong W. & Pallen C. J. (2000) Prenylation-dependent association of protein-tyrosine phosphatases PRL-1, -2, and -3 with the plasma membrane and the early endosome. *J Biol Chem*, 275, 21444-21452.
7. Huang C.-C., Hightower K. E. & Fierke C. A. (2000) Mechanistic studies of rat protein farnesyltransferase indicate an associative transition state. *Biochemistry*, 39, 2593-2602.
8. Jeong D. G., Kim S. J., Kim J. H., Son J. H., Park M. R., Lim S. M., Yoon T.-S. & Ryu S. E. (2005) Trimeric structure of PRL-1 phosphatase reveals an active enzyme conformation and regulation mechanisms. *J Mol Biol*, 345, 401-413.
9. Sun J.-P., Wang W.-Q., Yang H., Liu S., Liang F., Fedorov A. A., Almo S. C. & Zhang Z.-Y. (2005) Structure and biochemical properties of PRL-1, a phosphatase implicated in cell growth, differentiation, and tumor invasion. *Biochem*, 44, 12009-12021.
10. Diamond R. H., Peters C., Jung S. P., Greenbaum L. E., Haber B. A., Silberg D. G., Traber P. G. & Taub R. (1996) Expression of PRL-1 nuclear PTPase is associated with proliferation in liver but with differentiation in intestine. *Am J Physiol Gastrointest Liver Physiol*, 271, G121-G129.
11. Yarovinsky T. O., Rickman D. W., Diamond R. H., Taub R., Hageman G. S. & Rickman C. B. (2000) Expression of the protein tyrosine phosphatase, phosphatase of regenerating liver 1, in the outer segments of primate cone photoreceptors. *Mol Brain Res*, 77, 95-103.
12. Takano S., Fukuyama H., Fukumoto M., Kimura J., Xue J.-H., Ohashi H. & Fujita J. (1996) PRL-1, a protein tyrosine phosphatase, is expressed in neurons and oligodendrocytes in the brain and induced in the cerebral cortex following transient forebrain ischemia. *Mol Brain Res*, 40, 105-115.
13. Yu L., Kelly U., Ebright J. N., Malek G., Saloupis P., Rickman D. W., McKay B. S., Arshavsky V. Y. & Rickman C. B. (2007) Oxidative stress-induced expression and modulation of phosphatase of regenerating liver-1 (PRL-1) in mammalian retina. *Biochim Biophys Acta*, 1773, 1473-1482.
14. Skinner A. L., Vartia A. A., Williams T. D. & Laurence J. S. (2009) Enzyme activity of phosphatase of regenerating liver is controlled by the redox environment and its C-terminal residues. *Biochemistry,* 48, 4262-4272.

15. Skinner A. & Laurence J. S. (2009) $^{1}H$, $^{15}N$, $^{13}C$ resonance assignments of the reduced and active form of human protein tyrosine phosphatase, PRL-1. *Biomol NMR Assign*, online first.
16. Tuinstra R. L., Peterson F. C., Kutlesa S., Elgin E. S., Kron M. A. & Volkman B. F. (2008) Interconversion between two unrelated protein folds in the lyphotactin native state. *Proc Natl Acad Sci*, 105, 5057-5062.
17. Cordes M. H. J., Burton R. E., Walsh N. P., McKnight J. & Sauer R. T. (2000) An evolutionary bridge to a new protein fold. *Nature Struct Biol*, 7, 1129-1132.
18. Luo X., Tang Z., Xia G., Wassmann K., Matsumoto T., Riza J. & Yu H. (2004) The Mad2 spindle checkpoint protein has two distinct natively folded states. *Nature Struct Biol*, 11, 338-345.
19. Santiveri C. M., Leon E., Rico M. & Jimenez M. A. (2008) Context-dependence of the contribution of disulfide bonds to beta-hairpin stability. *Chemistry*, 14, 488-499.

# APPENDIX A. *p*NPP INDUCES OLIGOMERIZATION OF PRL-1

## A.1. *p*NPP INDUCES OLIGOMERIZATION OF PRL-1

In an attempt to determine Michaelis-Menten kinetic constants and compare our findings with those previously published, the *p*NPP substrate concentration was varied from 0.2-60 mM and the reaction was monitored over time. As can be observed in Figure A.1, PRL-1 exhibits an atypical kinetic profile in that the reaction is inhibited with increasing amounts of substrates in a non-linear fashion. This was observed for all mutants examined and a representative plot of the data for PRL-1-C170S-C171S is shown in Figure A.1.A. Because PRL-1 is generating free phosphate during the course of the experiment, which could cause feedback inhibition, attempts were made to precipitate excess phosphate out of solution by the addition of $CaCl_2$ with little success due to interference with the measurement because of light scattering. The progress curve for PRL-1-C49S is also shown in (Figure A.1.B). It is apparent that this mutant behaves even more anomalously than the C-terminal mutant, displaying a profile reminiscent of aggregation behavior. The visible range used for this assay is susceptible to interference from light scattering effects, and as

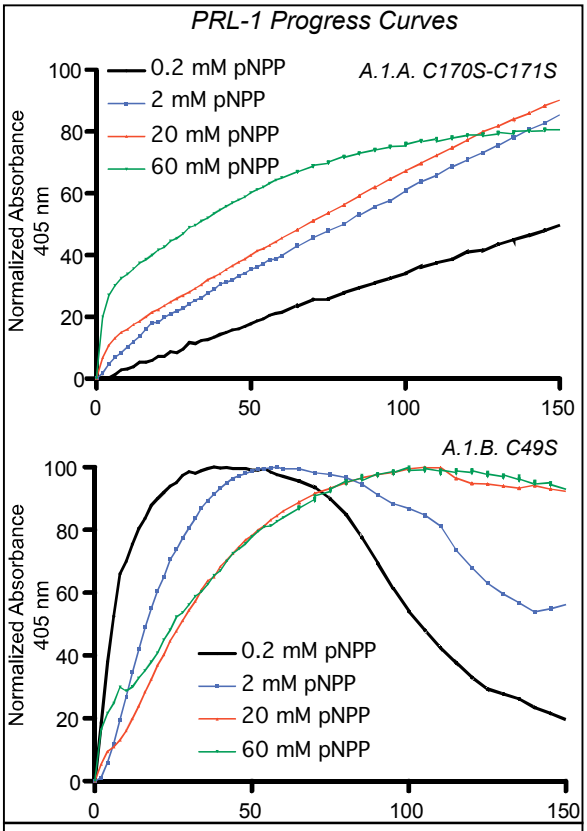



**Figure A.1. PRL-1 Progress Curves. A.** PRL-1-C170S-C171S was allowed to react with increasing concentrations of *p*NPP for 150 minutes and the amount of *p*NP product formed over time was monitored at 405 nm. **B.** As in A, except C49S was monitored.

such, a study to determine if oligomerization or aggregation of the protein was causing this atypical profile was conducted. First, the protein was incubated in the assay conditions without substrate and monitored over the time course of the kinetic reaction. A small amount of background absorbance from the protein itself was detected, but not in any significant quantity to cause such an anomalous progress curve (data not shown). Next the ability of *p*NPP to induce aggregation was evaluated. In this study, the $^{1}$H-$^{15}$N HSQC NMR peaks of PRL-1 were monitored over time, and as indicated in Figure A.2 by the loss of protein signal after the addition of *p*NPP, aggregation of the protein occurs within 60 minutes.

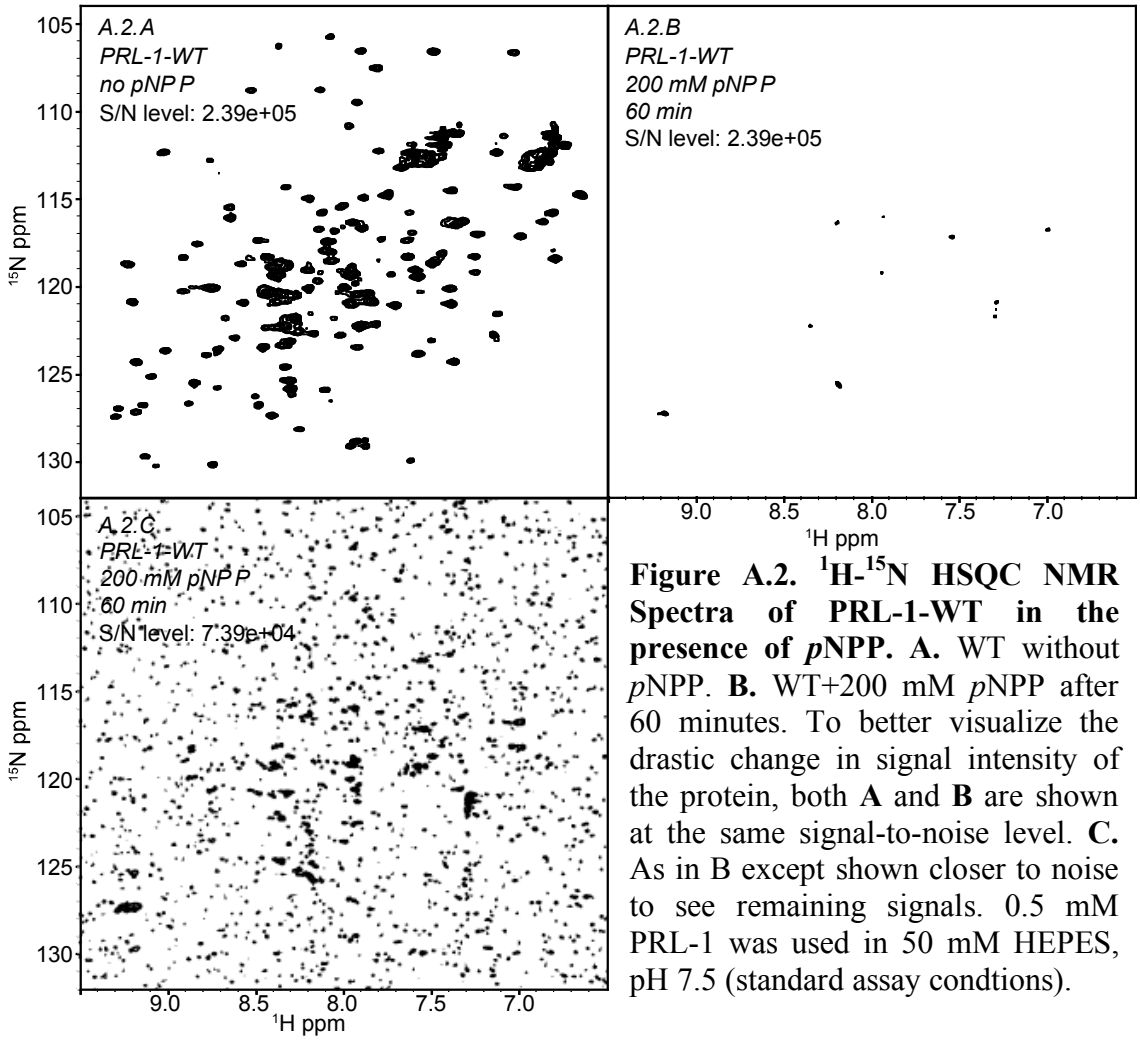



**Figure A.2. $^{1}$H-$^{15}$N HSQC NMR Spectra of PRL-1-WT in the presence of *p*NPP. A.** WT without *p*NPP. **B.** WT+200 mM *p*NPP after 60 minutes. To better visualize the drastic change in signal intensity of the protein, both **A** and **B** are shown at the same signal-to-noise level. **C.** As in B except shown closer to noise to see remaining signals. 0.5 mM PRL-1 was used in 50 mM HEPES, pH 7.5 (standard assay condtions).

The fact that the generic substrate can induce oligomerization is significant because this substrate has been used for all other previously reported kinetic studies of PRL-1 and PRL-3.[1-3] Unfortunately, the actual progress curves are rarely presented because discrete time point assays are often performed, so it is uncertain whether other groups have experienced a similar problem. It is shown here that above concentrations of 20 mM *p*NPP, aggregation can be quite significant, which precludes any kind of quantitative analysis. Unfortunately, Sun *et al*. (2005) did not report the range of substrate concentrations used; only that they spanned from 0.2-5 mM times the $K_M$. Using this information, the concentrations could be extrapolated if the enzyme concentration is known. This is also not reported in their published manuscript.[1] Consequently, only the relative activity of a given mutant at a given concentration is reported for PRL-1 for this project. In these experiments, the reaction is allowed to proceed and an absolute rate value is calculated using the molar absorptivity coefficient of *p*-NP. In this case, 20 mM pNPP was used because there is only a small degree of aggregation at this concentration (see Figure A.3) and because the enzyme is to inherently too slow to use smaller concentrations. Upon examining the progress curves, it is clear that the C170S-C171S mutant has the most activity. This value was, therefore set as the theoretical maximum, while the value of the inactive C104S protein was used as a theoretical minimum. Other groups have used a variety of other substrates to measure the kinetics of other

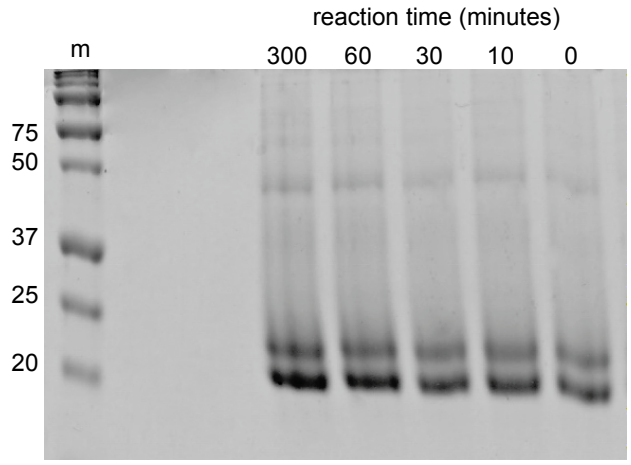



**Figure A.3. SDS-PAGE Analysis of PRL-1-WT in the Presence of 20 mM *p*NPP.** The *p*NPP assay was allowed to proceed and gel samples were taken out at various time points up to 300 minutes. Once collected, non-reducing loading buffer was added and the sample heated to stop the reaction. This reaction was done under non-reducing conditions, which explains the two bands for PRL-1 at 20 kDa.

PTPase family members. It is unclear whether these substrates may also induce oligomerization or if this observation is specific to *p*NPP and/or PRL-1. Such studies are on going in the Laurence lab.

## A.2. REFERENCES


1. Sun J.-P., Wang W.-Q., Yang H., Liu S., Liang F., Fedorov A. A., Almo S. C. & Zhang Z.-Y. (2005) Structure and biochemical properties of PRL-1, a phosphatase implicated in cell growth, differentiation, and tumor invasion. *Biochem*, 44, 12009-12021.
2. Kozlov G., Cheng J., Ziomek E., Banville D., Gehring K. & Ekiel I. (2004) Structural insights into molecular function of the metastasis-associted phosphatase PRL-3. *J Biol Chem*, 279, 1182-1189.
3. Matter W. F., Estridge T., Zhang C., Belgaje R., Stancato L., Dixon J., Johnson B., Bloem L., Pickard T., Donaghue M., Acton S., Jeyaseelan R., Kadambi V. & Vlahos C. J. (2001) Role of PRL-3, a human muscle-specific tyrosine phsophatase, in angiotensin-II signaling. *Biochem Biophys Res Commun*, 283, 1061-1068.

# APPENDIX B. SEQUENCE ALIGNMENT OF PRL FAMILY MEMBERS


The sequence alignment of the human PRL family members is shown above. The sequence alignment shown is based on the results presented in Figure 1 of Sun *et al*. (2005). PRL-1 is 87% identical to PRL-2 and 79% identical to PRL-3. Overall, the three members share 76% sequence identity with each other.

## B.1. REFERENCES

1. Sun J.-P., Wang W.-Q., Yang H., Liu S., Liang F., Fedorov A. A., Almo S. C. & Zhang Z.-Y. (2005) Structure and biochemical properties of PRL-1, a phosphatase implicated in cell growth, differentiation, and tumor invasion. *Biochem*, 44, 12009-12021.

Page Left Intentionally Blank

# APPENDIX C. HIGH-FIELD SOLUTION NMR SPECTROSCOPY AS A TOOL FOR ASSESSING PROTEIN INTERACTIONS WITH SMALL MOLECULE LIGANDS

## C.1. INTRODUCTION

One of the most important uses of high-field solution nuclear magnetic resonance (NMR) in the pharmaceutical industry relies on its unique ability to examine protein-drug interactions at very high resolution. NMR experiments have provided crucial data to assess the ability of a small molecule drug to bind and modify the activity of a disease-modifying protein at a structurally defined binding site.[1] Upon the interaction of a small molecule ligand with a protein target, a perturbation of the NMR spectrum occurs, and detection of the perturbation is the basis of ligand binding analysis by NMR. The perturbation assay is divided into two categories based on whether the signals from the protein target or the ligand are observed. An expanding set of NMR experiments is available to map either the small or macromolecule in a site-specific manner to chemically and physically characterize small molecule ligand interactions with protein targets. A significant advantage of these NMR assays is that they are useful for identifying binding partners from a mixture of compounds (NMR screening). Each method will be discussed in further detail in the appropriate sections, and a summary of the advantages and disadvantages is outlined in Table C.1.

Solution NMR is often segregated into two fields. Low-field NMR is primarily utilized for small molecule analysis, while macromolecular studies require high-field instruments (greater than 500 MHz). Higher field spectrometers offer advantages for the study of both small

and large compounds. First, the probes accompanying higher fields allow selective detection of bonds or atoms to obtain site-specific information about the protein target or small molecule. Additionally, increased field strength improves sensitivity and provides better resolution and narrower line widths, which are necessary to analyze large molecules that generate many signals. Improved sensitivity allows lower concentrations of analytes to be detected. Although small molecule analysis traditionally does not require high-field spectrometers, the higher field instruments are advantageous because the increased resolution allows for improved ability to examine small molecules, specifically in the context of their interactions with macromolecules where they are undergoing exchange processes, which leads to line broadening.[2]

**Table C.1. Summary of Current NMR screening Methods**

| | **Limits and Requirements** | | | **Permits Identification of:** | | |
|---|---|---|---|---|---|---|
| **Protein-Target Observed** | Target MW | $K_D$ | Isotopic Labeling | Binding Site | Binding Epitope | Selectivity |
| [a]Chemical Shift Perturbations | [b]30 kDa | $10^{-9}$-$10^{-3}$ | $^{13}C$ and/or $^{15}N$ | [c]✓ | | |
| $^{19}F$ Relaxation and Chemical Shift Perturbations | unlimited | $10^{-6}$-$10^{-3}$ | $^{19}F$ | [d]✓ | | |
| **Ligand Observed** | | | | | | |
| Relaxation Rates | unlimited | [e]$10^{-6}$-$10^{-3}$ | none | | ✓ | ✓ |
| Diffusion Coefficients | unlimited | $10^{-6}$-$10^{-3}$ | none | | | ✓ |
| Nuclear Overhauser Effects | unlimited | $10^{-6}$-$10^{-3}$ | none | | ✓ | ✓ |
| Magnetization Transfer | unlimited | [e]$10^{-6}$-$10^{-3}$ | none | | ✓ | ✓ |
| $^{19}F$ Relaxation and Chemical Shift Perturbations | unlimited | [e]$10^{-6}$-$10^{-3}$ | [f]none | | | ✓ |

[a]Requires the use of two-dimensional $^{1}H$-$^{15}N$ HSQC or $^{1}H$-$^{13}C$ HSQC experiments.
[b]Requires the use of advanced pulse sequences and/or isotopic labeling strategies.
[c]Partial or complete resonance assignments are required.
[d19]F resonances must be assigned if more than one aromatic residue occurs in the protein.
[e]Competition based screening permits determination of a wider range of binding affinities.
[f]Isotopic labeling is not required if the ligand naturally contains an $^{19}F$ moiety

High-field solution NMR can provide supplementary or additional information about a binding reaction in addition to facilitating determination of the association constant between a drug and binding partner, as can be obtained with traditional functional assays. Most notably, NMR provides detailed structural information about the protein-ligand binding site, as well as data from which the $K_D$ value can be determined, providing a structure-activity relationship (SAR).[3] Moreover, high-field NMR can be used to structurally characterize binding of two or more ligands simultaneously.[4,5] The experimental setup is straightforward and provides a universal process, applicable to virtually any protein, for the detection of intermolecular interactions.[4] Two unique advantages of analyzing molecular interactions by NMR are that: 1) it requires no target-specific assay or knowledge of the target protein's function[6] and 2) NMR experiments can detect species that interact weakly with each other (mM dissociation constants). The ability to detect direct interactions between ligand and target over a broad $K_D$ range illustrates the robustness of this technique.[5]

In brief, high-field solution NMR offers the ability to easily evaluate the structural, thermodynamic and kinetic aspects of a binding reaction.[7] This appendix outlines the current methodologies used to assess protein-ligand interactions from both the perspectives of the protein target and ligand molecule. Appropriate background information is provided to describe the NMR experiments available for assessing ligand binding and the corresponding data interpretation. Finally, advances in NMR technology and experimentation are presented to illustrate the utility of solution NMR in the identification and characterization of potential therapeutics to protein targets.

## C.2. "PROTEIN TARGET" DETECTED METHODS

Most target-observed screening applications rely on chemical shift changes as indicators for intermolecular binding.[6] This type of screening generally is referred to as a perturbation assay, and it is used to evaluate the resonances of the protein target as a function of ligand concentration in a titration experiment.[8] Changes are observed upon binding because chemical shift reports on the local electronic environment of a nucleus. NMR commonly detects spin ½ nuclei ($^{1}H$, $^{13}C$, $^{15}N$, $^{19}F$), and the type of nucleus is a major determinant of the chemical shift value.[9] Additionally, the dielectric properties of the solvent, covalent or hydrogen bonds, van der Waals interactions and bond angles influence the electronic environment of an atom, which impact its chemical shift.[10] Changes in NMR chemical shifts can result from any physical or chemical process that changes the magnetic environment of a nucleus, including dynamic processes such as conformational changes and solvent effects. The chemical shifts of surface exposed nuclei are affected by bulk solution. Proteins are virtually always analyzed in aqueous solution, where the main contribution to solvent effects comes from the pH and ionic species present. Generally, low ionic strength and slightly acidic conditions are more conducive to generating high-quality spectra, but the requirements for each protein are unique.[11] On the interior of the protein, chemical shift values are dictated largely by packing interactions and bond geometries. Chemical shifts corresponding to regions experiencing dynamic fluctuations may be substantially influenced by both structure and solvent contact. When interactions occur between analytes in solution different values of NMR chemical shifts reflect the different nuclear environments undergoing exchange and depend on the nature of the exchange or binding process.[12] A nucleus involved in a binding reaction will experience a different environment in its

free and bound form, generating signal that reflects the combination of both. The appearance of the NMR spectrum also depends on the rate at which the exchange phenomenon occurs (Figure C.1). A slowly exchanging system exists when the exchange occurs more slowly than the difference in chemical shift in frequency units. When the exchange is faster than the difference in chemical shift, the system is classified as a fast-exchange system. NMR experiments detect processes that occur on the microsecond to millisecond time scale. If the nuclear environments change slowly on the NMR timescale, the spectrum will capture two distinct sets of resonances for the free and bound states (Figure C.1.A), whereas rapidly-exchanging nuclei will produce a single set of resonances representing a weighted average of the chemical shifts of the free and bound signals (Figure C.1.C).[9,12]

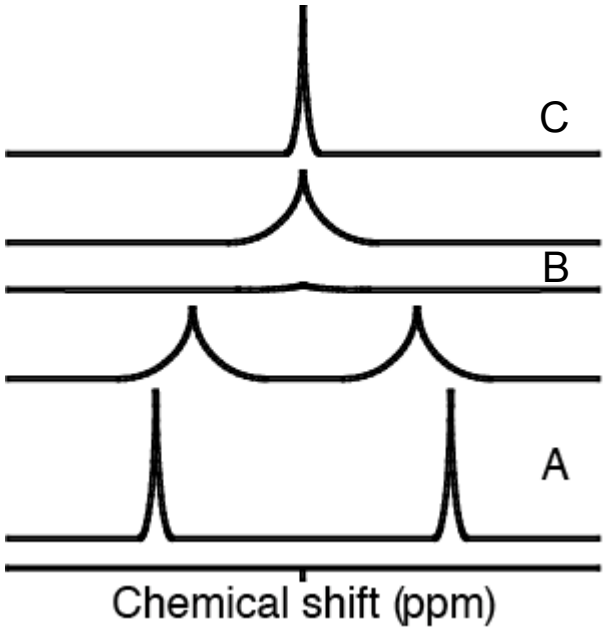



**Figure C.1. Exchange Regimens Observed with NMR. A.** Two resonances appear because the observed nucleus exchanges slower than the NMR scale. **B.** Nuclei that exchange on the intermediate time scale cannot be detected because they are too broad. **C.** One resonance is observed because the detected nucleus exchanges faster than the NMR time scale. The peak represents a weighted average of the two exchanging environments.

The ability to interpret a NMR spectrum becomes increasingly difficult for higher molecular weight species. A small molecule with relatively few atoms will produce well-resolved, sharp NMR signals with large chemical shift dispersion (Figure C.2.A). In contrast, a simple one-dimensional spectrum of a protein contains thousands of signals generated by

numerous similar chemical moieties that appear in the same regions of the spectrum. The presence of secondary and tertiary structure elements provides improved dispersion, distinguishing the chemical shifts from those of the random state as the microenvironment of a particular nucleus dictates its chemical shift in an extremely sensitive manner. Nonetheless, due to the large number of signals, one-dimensional observation typically lacks sufficient resolution to evaluate structure and dynamics and quantify binding (Figure C.2.B).[13] Consequently, this has led to the development of multidimensional correlation experiments. These experiments resolve overlapping signals by expanding the spectrum into two or three dimensions based on different nuclei or differences in connectivity.[12]

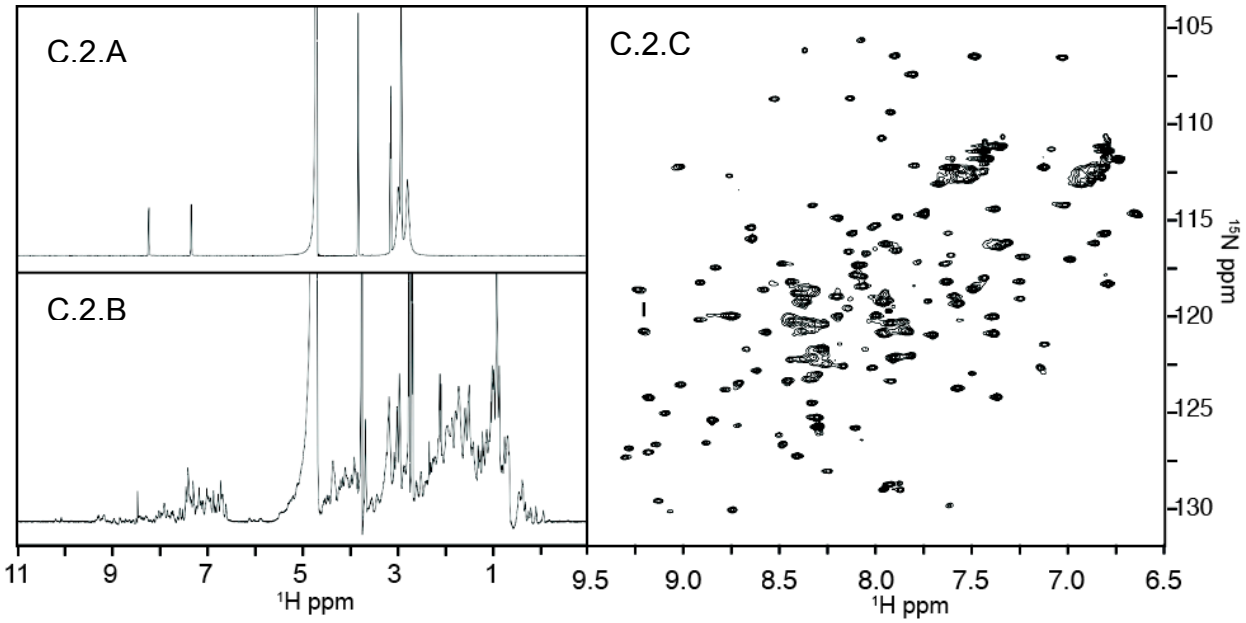



**Figure C.2. Multidimensional NMR. A.** $^{1}$H NMR spectrum of *p*NPP in 50 mM HEPES, pH 7.5. The two resonances at approximately 7 and 8 ppm correspond to the two aromatic protons. The water peak resonates at 4.703 ppm and the remaining peaks below 4 ppm correspond to the HEPES buffer. **B.** $^{1}$H NMR Spectrum of PRL-1. The peaks above 6 ppm are the protein's backbone amides and side chains. The water peak resonates at 4.703 ppm and the remaining signals respresent the protein's aliphatic protons. The signals are too numerous and overlapped to make any specific assignments. **C.** $^{1}$H-$^{15}$N HSQC spectrum of PRL-1. Each peak represents a single NH group in the protein, including backbone and side-chain amides. Using three-dimensional correlation experiments, each peak can be assigned to a given amino acid in the protein sequence to yield structural information that can guide structure-based inhibitor design to the protein target.

Two-dimensional correlation spectroscopy refers to a class of experiments that probe the connectivity between nuclei. Each axis shows the chemical shift of a specific nucleus, and cross peaks result only for those that are correlated to each other.[14] In general, the experiment can be applied to investigate homo or heteronuclear associations. The Heteronuclear Single Quantum Correlation (HSQC) experiment is particularly useful for the analysis of proteins. The $^{1}H$-$^{15}N$ HSQC experiment exploits the repeating nature of the protein's primary sequence and three-dimensional structure. The amide bond connects the amino acids that compose the protein sequence, creating a repeating series of NH groups that become chemically unique in the context of the protein's tertiary structure. The experiment only detects those protons directly coupled to nitrogen, so the resulting spectrum lacks any overlap from proton signals associated with carbon (typically aromatics in the relevant ppm range).[12] The spectrum displays one cross peak for every amide in the protein at a position characterized by its $^{1}H$ and $^{15}N$ chemical shifts.[15] In total, the spectrum contains one signal for each residue except proline, which lacks an HN, two paired peaks generated by the asparagine and glutamine $NH_2$ groups and additional signals from NH containing side chains (Figure C.2.C).[7] The HSQC spectrum provides an overall map or fingerprint of the protein target and serves as the basis for assessing ligand binding and the overall fold and behavior of a protein. This experiment is fast and easily interpreted but requires isotopic labeling of the protein with $^{15}N$ because of its low natural abundance (0.37%).[10] When labeling is not practical, a $^{1}H$-$^{13}C$ version of the HSQC can be performed using unlabeled protein to evaluate the aliphatic protons; however, this experiment requires longer data acquisition. Both experiments reveal information about the protein's structure. The $^{1}H$-$^{15}N$-HSQC has the advantage of probing each amino acid and revealing whether the protein is folded, as the amide proton chemical shift distribution for each amino acid is significantly influenced by its secondary

structure and collapses to a narrow ppm range upon unfolding.[16] Also, the $^{13}C_{\alpha}$ alpha and $^{13}C_{\beta}$ and, to a lesser extent, $^{1}H_{\alpha}$ chemical shifts correlate with α-helical and β-sheet secondary structure elements, but the spectrum is more difficult to interpret with respect to structural changes than the $^{15}N$ version.

The HSQC experiment is a powerful tool for binding studies because of the sensitivity of coupled nuclei to changes in the environment. The $^{1}H$-$^{15}N$ HSQC detects changes in the backbone amide bonds upon addition of ligand, whereas the $^{1}H$-$^{13}C$ HSQC detects changes in aliphatic and aromatic chemical shifts of the side chains. Monitoring the chemical shift changes as a function of ligand concentration can be performed to accurately measure the affinity constant between the ligand and the target (Figure C.3).[7,17-19] The magnitude of the protein's chemical shift change will differ depending on how the small molecule interacts with the protein. For example, hydrogen bonding will have a different effect on the chemical shift than covalent attachment.[20] Knowledge of the protein's resonance assignments reveals the residues involved in binding. In this case, most studies employ the $^{1}H$-$^{15}N$ HSQC because backbone resonances are more easily obtained.[7] Moreover, binding site identification allows differentiation of specific from nonspecific binding, which is characterized by small environmental changes in the NMR spectrum similar to the effects of a change in temperature.[21] NMR is, thus, a unique tool with the ability to immediately identify false positives.[4] It is important to note that even if the protein target resonances have not been assigned, target-observed assays may still be used to identify ligands that bind strongly to the target and reveal whether multiple ligands bind using the same or different binding sites.[22] In addition to detecting tight binding, this method can detect interactions that are as least as weak as mM binding $K_D$s because observation of the protein target does not rely on tight binding to acquire bound-state information.[21,23] A consequence of

weaker binding is that the time scale on which the ligand exchanges leads to line broadening. For systems exchanging on this intermediate timescale (Figure C.1.B), the line width of the protein resonances will sharpen as saturation binding is approached. To confirm that a ligand is binding in a single conformation the concentration of the ligand should be titrated until saturation is reached. If no change in the line width is observed, structural investigations may be complicated, as the ligand likely occupies the binding site in multiple conformations of comparable energy.[24]

Upon the addition of ligand, the signals of those amides whose environments are perturbed by ligand binding change position.[15] Specific binding of a small molecule and protein target can take place using two modes.[6] A lock and key interaction of the ligand to the protein has little or no effect on residues that are not directly involved. Consequently, only those involved experience a different environment and resonate at a different frequency. Some have reported that as few as eight out of 107 resonances change upon compound binding.[25] Induced fit binding leads to a conformational change, often resulting in a global change of the protein as a whole, where the majority of the resonances are altered.[7,26] In this case, a series of experiments that compare chemical shift changes induced by a series of closely related ligands may be used to identify the binding site.[20] After mapping the changes in chemical shifts from the titration of a first ligand, a second, structurally-related ligand differing by a functional group is added to the protein target. Most of the perturbations will parallel the changes of

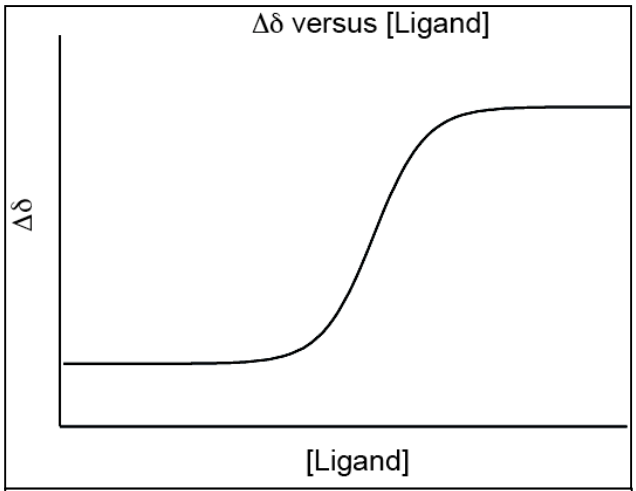



**Figure C.3. Δδ vs [ligand].** Changes in chemical shift are monitored over a series of ligand concentrations and plotted to determine the dissociation constant.

first ligand except for the resonances at the binding site. This occurs because of the proximity of the protein's binding site to the substituted functional group on the ligand and because the magnitude of the chemical shift change depends on the ligand moiety involved with binding.[7,20]

The chemical shift perturbation assay has been used extensively to identify the ligand binding site of many important protein targets and receptors. For example, cyclosporine A bound to human cyclophillin,[27] interleukin-8 complexed with the N-terminal region of the human interleukin-8 receptor,[28] the Fab fragment of IgG interacting with streptococcal protein G[29] and $NADP^+$ bound to the MurB enzyme[30] have all been characterized using the chemical shift perturbation method. Importantly, these studies helped lead to the advent of the SAR (structure-activity relationship) by NMR method developed by S. Fesik at Abbott Laboratories.[25] SAR by NMR detects changes in the chemical shifts of the protein target upon addition of ligands. Weakly interacting compounds that bind at adjacent sites on the protein target are identified and combined with structural information about the orientation of the bound ligands to guide a linked-fragment approach to generate lead drug compounds with greatly increased binding. Figure C.4 outlines the experimental procedure. The first step requires screening a library of ligands using the perturbation method with the assumption that a change in chemical shift greater than 0.1 ppm for at least two peaks in the spectrum constitutes a significant change and indicates binding.[31] After hit identification, monitoring chemical shift changes as a function of ligand concentration in a titration experiment is performed to accurately deduce the binding constant.[32] Following identification of a binding partner to the first site, screening of related compounds is conducted to increase and optimize the binding affinity for that site. The same perturbation method is applied to identify a second low-affinity hit in the presence of the optimized first ligand. Following identification and optimization of the second binding site, structural data

obtained by additional experiments is used to determine the location and orientation of the protein target in complex with its two low affinity hits. Maintaining the spatial orientation of the compounds with respect to each other and the protein target, the ligands are synthetically linked together to produce the high-affinity ligand.[25]

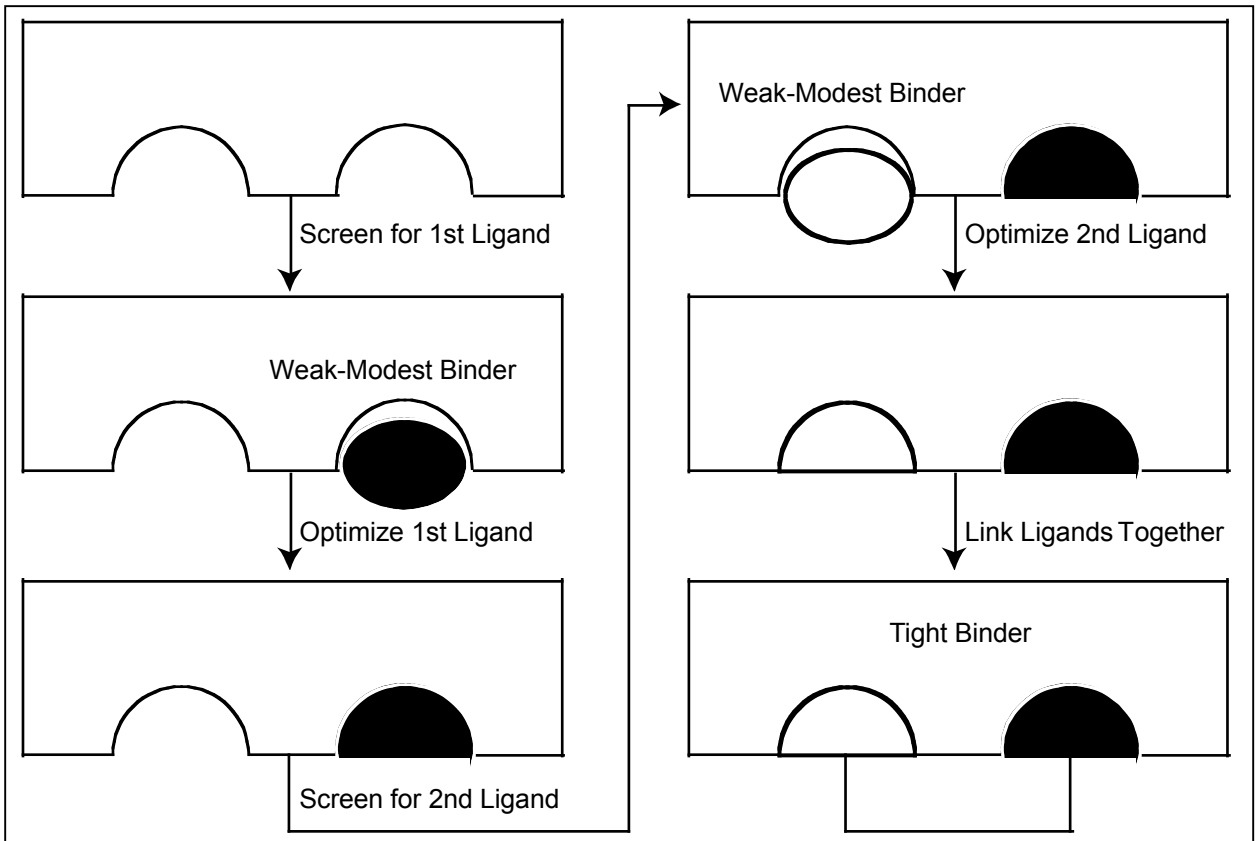



**Figure C.4. SAR by NMR.** Binding of the first ligand is determined by detecting changes in the $^{1}H$-$^{15}N$ HSQC of the protein target upon titration of a mixture of ligands. After a binding ligand is identified and optimized, a mixture is screened for a second low affinity binder to an adjacent site on the protein target. Once optimized, the ligands are synthetically linked together to create one, high-affinity lead drug compound.

The SAR by NMR method has facilitated the development of inhibitors of proteins, for example, stromelysin, to combat its role in arthritis and tumor metastasis.[32] Additionally, the methodology has supported the characterization and development of inhibitors that block DNA binding by the human papillomavirus E2 protein, to treat cervical cancer,[33] and antagonists of Erm Methyltransferases to fight resistant bacteria.[34] More currently, the SAR by NMR method

served as the primary tool for characterizing protein-ligand complexes of the Bcl-2 family, a class of proteins involved in the apoptotic pathways.[35] Additionally, variations of this method have been developed and utilized recently to examine higher molecular weight proteins and analyze selective binding among related proteins, including SAR by NMR with $^{13}$C-labeled methyl groups (see section C.4.3)[36] and RAMPED-UP NMR.[26] RAMPED-UP NMR (Rapid Analysis and Multiplexing of Experimentally Discriminated Uniquely labeled Proteins using NMR) is an important advance, because it provides a way of simultaneously evaluating the specificity of a ligand for a protein among a mixture or class of proteins. Each protein target is labeled with one unique amino acid type and studied simultaneously in a mixture using a ligand titration experiment. The ability to multiplex the protein target not only increases the screening throughput but also reveals whether the ligand of interest will target a single family member in a specific manner or bind to several members from a class of proteins.

Although the perturbation assay and SAR by NMR are powerful techniques for characterizing ligand-protein interactions, the size, solubility and stability of the protein, as well as the ability to isotopically label the protein, may limit the ability to screen for binding using the protein target resonances.[4] The line width of NMR signals increases with increasing molecular weight, resulting in spectral overlap and loss of signal and resolution.[37] To acquire protein spectra with reasonable signal-to-noise ratios in a timely manner, the protein must stay soluble at concentrations around 0.1-1 mM without degradation for the duration of the experiments. The length of the NMR experiments that detect the protein target depend on the type of experiment, which can last anywhere from a few minutes to several days.[38] Furthermore, the $^{1}$H-$^{15}$N HSQC requires isotopic labeling of the protein target.[39] Although, the $^{13}$C nucleus occurs at 1% natural abundance, studies without $^{13}$C enrichment can provide valuable information, but this requires

longer data acquisition. Although isotopically labeled protein production can increase cost, labeling facilitates the observation of specific nuclei or bonds of one species. With protein target detected NMR, for example, because the ligand is left unlabeled, ligand resonances do not complicate the protein target spectrum, eliminating the need for spectral subtraction (difference spectroscopy) (see section C.3.1.c).[40] $^{15}N$ and $^{13}C$ isotopic labeling permits the assignment of backbone amides using a series of three-dimensional NMR experiments.[13,16,38] A description of how assignments are completed is beyond the scope of this paper, and a detailed discussion of the topic can be found elsewhere.[38] Completion of sequential assignments is necessary to obtain structural information about the protein target's binding site and can be used to develop structure activity relationships. As the size of the protein increases, resonance assignments become more difficult to obtain, and as such, different labeling strategies are used to facilitate the evaluation of larger protein targets with molecular weights greater than 30,000 kDa (see section C.4.3).

Target observation also may result in lower throughput as a result of more time-consuming, multidimensional spectral analysis and the need to deconvolute hits from a small molecule compound library.[4] Selecting an optimal mixture size from the library to be screened will have a great impact on the efficiency of the perturbation assay. A comprehensive review of the determinants has been presented by Mercier and Powers.[41] The time needed to deconvolute hits from a mixture depends largely on the number of compounds in a mixture but also on the physicochemical properties of the mixture's components. For example, the main goals to achieve when selecting a compound mixture include minimizing reactivity and interactions of compounds, achieving structural diversity, maximizing solubility, and preserving constant physical properties (pH and ionic strength). Deconvoluting mixtures of ligands also has been addressed by an automated system in which 306 individual compounds were screened for

binding to a $^{15}$N-labeled target. This robotic sample preparation, in conjunction with the automated data acquisition and analysis tool is applicable to many systems. It can be used to evaluate a large number of ligand-protein interactions without the need to deconvolute hits from the larger library.[42,43]

An alternative approach to minimizing the spectral complexity of proteins is to introduce a spin probe. $^{19}$F is a unique spin ½ nucleus that can be incorporated into a protein at Tyr, Phe, His, or Trp residues using selective labeling techniques.[44] Aromatic residues are useful probes of structural integrity and binding because, respectively, they are often buried in the hydrophobic core of the proteins and/or located at the interface of intermolecular interactions.[45] Several useful properties of the $^{19}$F nucleus make it an ideal label for NMR studies, which have been described previously.[46,47] First, $^{19}$F is present at 100% natural abundance and has 83% of the sensitivity of $^{1}$H providing a reasonable alternative to $^{1}$H NMR. Next, $^{19}$F incorporation generally produces no dramatic structural effects when substituted for hydrogen in an amino acid side chain. Although, $^{19}$F is substantial larger than $^{1}$H, the two have similar covalent radii (1.3Å for $^{19}$F and 1.2Å for $^{1}$H).[46] Third, the considerably different dipole moment means that fluorine lone pair electrons primarily control the $^{19}$F chemical shift, resulting in a 100-fold larger chemical shift range than $^{1}$H. This also makes the nucleus especially responsive to local changes in van der Waals environments, electrostatic fields and hydrogen bonding.[47] As such, small chemical shift perturbations are easily detected in one-dimensional $^{19}$F spectra. Finally, $^{19}$F does not occur naturally in proteins so no background signals complicate the spectrum. A protein labeled with $^{19}$F-labeled amino acid will exhibit NMR peaks due only to the label, not to the backbone of the protein permitting the use of simple one-dimensional NMR experiments to detect changes in the protein structure upon addition of ligand. Because one-dimensional experiments are faster, $^{19}$F

NMR studies can be performed with lower protein concentrations, and thus, spectral acquisition is shorter than traditional protein detection techniques. Furthermore, $^{19}$F-NMR can be used to examine structural aspects of much higher molecular weight proteins, which traditional multidimensional NMR experiments fail to resolve. For example, selective $^{19}$F labeling and subsequent $^{19}$F NMR analysis of the 497 residue XIAP (X-linked inhibitor of apoptosis protein) with and without the Smac peptide revealed the interaction of the peptide to protein in a 2:1 ratio and the ability of the Smac peptide to inhibit the interaction of XIAP with its downstream partner.[48,49] Additionally, in an investigation of inhibitors to the anti-apoptotic protein Bcl-xL (181 amino acids), selective $^{19}$F-Phe labeling was used to obtain supplementary structural constraints for defining the orientation of the inhibitor in the Bcl-xL-inhibitor complex.[45]

## C.3. LIGAND DETECTED METHODS

Ligand observation experiments offer a useful alternative to protein target detection for larger proteins or those that oligomerize in solution at higher concentrations[23] or have no suitable expression host that permits isotopic labeling.[21] The use of simple one-dimensional spectra and ability to screen mixtures without deconvolution, results in higher throughput and its use as a well-suited primary screening technique.[4] Ligand examination has the technical advantage of requiring a smaller amount of unlabeled protein[9] and may be used to identify the binding epitope of the ligand.[4] Furthermore, the size of the protein target does not limit this technique, and in many examples, larger proteins provide better sensitivity because binding of a large protein causes a more distinguishable change to the observable NMR signals.[3] This broader approach to NMR screening relies on either detection of an altered hydrodynamic property of the ligand upon

addition of protein target or exchange-mediated transfer of bound state information to the free state.[50]

### C.3.1. Hydrodynamic Property Experiments

The NMR spectral properties of proteins and small molecules differ because of differences in hydrodynamic properties, which affect the translational and rotational mobility of a molecule in solution. Experiments in this category exploit the fact that a bound ligand transiently adopts the NMR properties characteristic of a large protein target, which are manifest in the spectrum of the ligand.[21,40]

#### C.3.1.a. Relaxation Experiments

Comparison of the ligand's relaxation rate in the presence and absence of the protein target represents the most well-established class of NMR binding assays. Relaxation reflects the hydrodynamic radius and rotational tumbling rate of species in solution.[9,21,23] Following delivery of a rf-pulse of energy, the system is perturbed and must re-establish its equilibrium by dissipating the absorbed energy supplied by the pulse. This loss occurs by transferring energy to either its surroundings or to other spins. Relaxation simply implies the movement of bulk nuclear magnetization towards equilibrium and occurs through two distinct mechanisms. Thermal equilibrium governs the recovery of magnetization in line with the magnetic field, meaning that the nuclear spin exchanges energy, in the form of heat, with its surroundings (lattice) to return to its ground state.[39] The rate at which relaxation proceeds via the spin-lattice mechanism is called $R_1$ and effectively occurs at a similar rate for small and large molecules.[10] The second mechanism by which energy is transferred is more complex but fundamentally depends on the size of the molecule, such that relaxation proceeds more rapidly with increasing size.[51] After an rf-pulse, the bulk magnetization becomes perpendicular to the magnetic field (transverse plane).

Although the pulse initially aligns the nuclei in the same plane, spin diffusion ensues and the transfer of energy from one spin to another spin (dipolar coupling) proceeds until the net magnetization dissipates to zero (Figure C.5).[10] The rate at which relaxation occurs via the spin-spin mechanism is called $R_2$ (transverse relaxation) and occurs much faster for large, slowly tumbling molecules.[51] The rate of transfer strongly depends on the chemical environments of the different nuclei and how efficiently the nuclei exchange energy; the larger the molecule, the more efficient the exchange and the faster the rate of relaxation.[37]

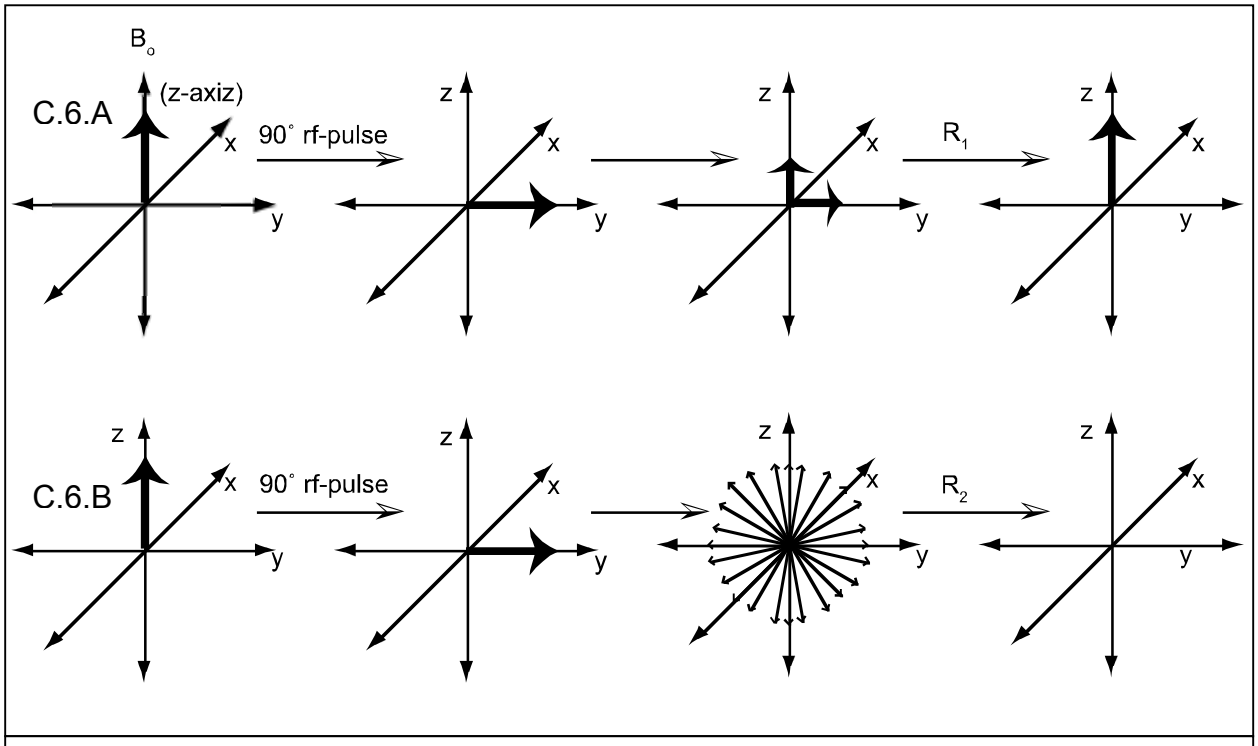



**Figure C.6. Nuclear Relaxation. A. Longitudinal Relaxation** (Spin-Lattice Relaxation). After an rf-pulse of energy is applied to the nuclei in solution, the atoms must relax back to their equilibrium position in line with the z-axis by releasing energy in the form of heat to its surroundings. **B. Transverse Relaxation** (Spin-Spin Relaxation). After an rf-pulse of energy is applied to the nuclei, the signals corresponding to the different nuclei fan out away from the y-axis at different rates due to chemical shift dispersion and diffusion and the transfer of energy from one spin to another spin, which proceeds until the net magnetization becomes zero. For small molecules the rate of transverse relaxation often equals the rate of longitudinal relaxation because there lacks sufficient interaction between the atoms to permit the transfer of excitation energy. For larger molecules, especially proteins, the rate of transverse relaxation is generally faster than longitudinal relaxation because the transfer of energy between spins occurs much more quickly and efficiently than the transfer of energy to the surroundings.

Additionally, the rate of transverse relaxation determines the line width of the observed resonances. Line width is directly proportional to the rate of spin-spin relaxation (line width=$R_2/\pi$), which is a consequence of detecting the NMR signal in the transverse plane, perpendicular to the magnetic field. Transverse relaxation governs the rate at which the signal or Free Induction Decay (FID) falls off, such that a large $R_2$ results in a FID that decays rapidly. Therefore, upon Fourier transformation to the frequency domain, faster relaxing species produce broader, less resolved peaks in the spectral window.[10]

When a small molecule binds to a large molecule, the small molecule transiently possesses similar NMR properties as the large molecule and assumes fast transverse relaxation rates leading to spectral line broadening of the small molecule signal.[9] The extent of ligand binding and size of the protein target directly correlates with line width changes.[52] As such, larger proteins usually produce stronger effects, increasing the sensitivity of the experiment.[3] For ligands with weaker dissociation constants (1-10 mM) to smaller protein targets, it is difficult to obtain reproducible and reliable line-broadening data, while binding to a larger protein (greater than 60 kDa) with the same affinity for the ligand is easily detected.[53] The most basic method to evaluate binding in this manner involves the acquisition of a ligand spectrum (or mixture of compounds) in the absence and presence of a protein target, from which the differences in line shape and relaxation rates are determined. Enhanced transverse relaxation of ligand upon the addition of the protein target indicates the transient formation of a bound complex.[21,23,54] For example, relaxation experiments were used to obtain information regarding the interaction of the FK506/FKBP complex (FK506-binding protein) with its downstream target calcineurin to improve the structure-based design of immunosuppressive agents. By detecting relaxation differences in different regions of the FK506/FKBP complex upon binding to calcineurin, the

region of the complex that interacts with the protein target was identified, which was important because the immunophilin/drug complex (FKBP/FK506 respectively), not the drug or protein alone, inhibits an immune response.[55]

Differential line broadening assays form the basis of the SHAPES screening method. A small but diverse library known as the SHAPES library is screened against any protein target of interest. The library consists of molecular shapes found commonly among commercially available therapeutics.[53] This method is similar in concept to the SAR by NMR technique. Once two shapes have been identified as weak binders by differential line broadening assays, a non-binding shape is used to link the two binding components to yield a higher affinity lead compound. To demonstrate the value of the SHAPES screening method, molecules from the SHAPES library were screened against p38 MAP kinase, an important protein target involved in cellular responses to external stress signals. The p38 MAP kinase served as a model protein target with which the utility of the SHAPES method could be established, because several inhibitors of p38 had been developed previously. SHAPES screening was utilized to discover a unique tri-substituted imidazole lead compound. The hit came from the same family of drugs as other p38 inhibitors that have important anti-inflammatory effects. The effectiveness of SHAPES screening against p38 MAP kinase indicates its potential to be applied to other novel protein targets to develop inhibitors to their function.[53]

Although most ligand screening methods rely on $^{1}$H-NMR, the favorable NMR properties of the fluorine nucleus make $^{19}$F-NMR an extremely useful method for detecting the binding of fluorinated compounds to a target.[3,56] The properties of fluorine that make it useful for protein target detected screening apply to ligand-detected methods as well. Most notably, the extraordinary sensitivity of $^{19}$F to its environment and local shielding effects means that its

resonances are dispersed over a large chemical shift range and it is extremely sensitive to relaxation changes caused by binding, which can provide more resolution than $^{1}$H-NMR relaxation experiments. Furthermore, because $^{19}$F does not occur naturally in proteins, no background signals complicate the one-dimensional spectrum, which eliminates the need for difference spectroscopy (see below).[4] Most NMR solvent and buffer components lack fluorine moieties, which means buffers and solvents also do not interfere with data acquisition.[57] Additionally, the broad chemical shift dispersion allows the use of large compound libraries minimizing overlap of individual ligand resonances.[21] Although protons are found more ubiquitously,[4] approximately 12% of compounds (equating to 200,000 compounds) in the Available Chemical Directory of Screening Compounds[58] and 17% of the MDDR database[45] contain fluorine moieties, making it quite possible to obtain a $^{19}$F-containing lead compound without resorting to special synthetic efforts.[57] For example, $^{19}$F-NMR has been used to characterize the multifunctional calcium-dependent Calmodulin protein bound to the fluorinated antipsychotic inhibitor, trifluoperazine (TFP) to determine that TFP binding can occur without calcium present.[59,60] Additionally, fluorine incorporation often improves the pharmacokinetic properties and potency of the drug.[21,45] Binding interactions between the plasma protein human serum albumin (HSA) and drugs can have detrimental effects on the drug's pharmacokinetic properties. $^{19}$F-NMR has been used to report on specific interactions between HSA and fluorinated dugs prior to lengthy pharmacokinetic studies.[61] This has recently been applied to the interactions between niflumic acid (a potent analgesic anti-inflammatory drug prescribed for rheumatoid arthritis) and HAS, providing insight into its pharmacokinetic and toxicological properties.[62] $^{19}$F NMR also has been applied to detect metabolism of cytochrome P450 substrates.[63] Cytochrome P450 enzymes are largely responsible for drug metabolism, and

analysis of fluorinated drug interactions with cytochrome P450 enzymes by NMR can be used to avoid adverse pharmacodynamic effects resulting from the accumulation of metabolites. Additionally, $^{19}$F-NMR facilitated the determination of the metal to substrate distances and geometries within the active site of P450s, providing valuable structural information for drug optimization.[64] Quantitative information about substrate turnover by enzymes, including P450s, can be obtained using the recently developed 3-FABS (three Fluorine Atoms for Biochemical Screening) technique. The substrate is labeled with a $CF_3$ moiety and detecting by using $^{19}$F NMR. The substrate and product concentrations are monitored as a function of time to determine inhibition constants.[65-67] For a more extensive analysis of ligand-detected $^{19}$F-NMR methods, a comprehensive review recently published is recommended.[68]

Another useful but less common spin probe used for analyzing protein-ligand interactions is $^{31}$P. $^{31}$P can be used to evaluate enzyme activity and inhibition. Screening for inhibitors can be conducted using this unique nucleus when the enzyme binds a phosphorylated cofactor, which would be perturbed by interactions with ligand molecules. A family of small molecule inhibitors of Ras was examined for their ability to stabilize the inactive Ras conformation by monitoring the $^{31}$P signals from the bound GTP.[69]

**C.3.1.b. Diffusion Experiments**

Comparison of the ligand's diffusion coefficient, which reflects the translational mobility of the molecule, in the presence and absence of the protein target, may be used to accurately examine intermolecular interactions between protein targets and ligands. Lin, Shapiro, and Wareing illustrated the use of diffusion-edited NMR (affinity NMR) spectroscopy for screening compound libraries in 1997.[70-72] The spectral differences upon the addition of protein target result from the inverse dependence of diffusion coefficients on the hydrodynamic radius of small

and large molecules. Diffusion experiments follow the same basic principles as relaxation experiments but rely on differences in translational instead of rotational motion. The small molecule transiently adopts the properties of the large molecule and, as such, exhibits a decrease in translational diffusion, which can be readily assessed by pulsed-field gradient NMR.[73] Because the signal intensity of the ligand resonances depends strongly on the strength of the gradient pulse, diffusion-edited NMR experiments distinguish between molecules that interact with the protein target and those that do not by increasing the strength of the gradient pulse, which selects for resonances from bound species (Diffusion Encoded Spectroscopy-DECODES).[22,74] The Diffusion Ordered Spectroscopy (DOSY) experiment extrapolates the one-dimensional chemical shift information into a pseudo-second dimension of diffusion coefficients and eliminates the need for deconvolution of the signals from a mixture of ligands.[75,76] A change in the second dimension indicates binding in a straightforward manner, as the data acquisition does not differ from simple one-dimensional experiments because the second dimension encodes a property of the molecule, not a nucleus.[52] As with relaxation experiments, diffusion experiments can be used to examine compound mixtures or libraries for protein target interactions in a powerful, high-throughput manner. This method has been applied to the identification of peptide compounds that bind to vancomycin to study antibiotic resistance. Vancomycin is a potent antibiotic for streptococcal and staphylococcal bacterial infections because it specifically binds to the carboxy-terminal D-Ala-D-Ala sequence of the bacterial peptidoglycan, inhibiting cell wall synthesis. Antibiotic resistance results because of a substitution to one of the amino acids in the target bacterial cell wall, which inhibits binding of vancomycin. The ability to study binding interactions between vancomycin (or its derivatives) and peptides that mimic the bacterial cell wall will facilitate the development of additional agents

that may overcome bacterial resistance. In this example, the limits of diffusion-edited NMR were challenged because the vancomycin receptor was only three times the size of the peptides tested. The results obtained, however, produced a spectrum with adequate resolution to identify interacting ligands.[77]

**C.3.1.c. Difference Spectroscopy**

Relaxation and diffusion experiments have great utility, but both experiments require the use of spectral subtraction (difference spectroscopy), which can have disadvantages. When examining smaller proteins, the signal intensity from the protein approaches that of the ligand, and the protein resonances may overlap in the region of the ligand signals, such that they obscure the signal from the ligand. Consequently, a large excess of ligand is required to ensure ligand observation. Most studies of this nature require a minimum of ten-fold ligand molar excess.[21] The experimental setup requires a one-dimensional spectrum of the protein alone and subsequent subtraction of the protein resonances from a spectrum of ligand in the presence of protein target. Subtraction of the resonances may lead to artifacts in the resulting spectrum when the protein's chemical shifts change upon binding. Relaxation and diffusion filtering eliminates the signals of the protein target and bound ligands and simplifies the subtraction process, minimizing the number of artifacts observed.[74] Relaxation filtered experiments employ the use of spin-locks, while diffusion filtered experiments utilize pulsed field gradients to attenuate the protein target and bound ligand signals. These edited experiments work because the rotational or translational motion properties of the bound ligand transfer to the detected free ligand. To detect binding, a spectrum of the ligand in the presence of the protein target is subtracted from a spectrum of the ligand in the absence of the protein target. If the ligand interacts with the protein, the resulting difference spectrum should resemble that of the ligand alone with decreased signal intensities.

For ligands that do not bind to the protein target, the resonances will be completely eliminated upon subtraction.

Both relaxation and diffusion experiments require a fast-exchange regimen between the ligand and the protein target. If a ligand briefly exchanges into another environment by binding to a protein and obtains a very different chemical shift during NMR detection, the signals from the ligand will dramatically decrease and broaden. The accumulation of these exchange events with time leads to destructive interference in the FID and loss of signal. Slowly exchanging nuclei cause these errors to accumulate and result in broad NMR signals indistinguishable from the noise. Fast-exchanging nuclei will average out these errors and produce one resonance line.[10,39] This limits the technique to weakly binding ligands in rapid exchange, and consequently, these experiments cannot differentiate the nature of the binding reaction (specific from non-specific binding),[78] leading to more frequent detection of false positives.[3,21] On the other hand, because relaxation and diffusion experiments rely on detection of the ligand resonances, the size, composition and oligomeric status of the protein target are essentially unlimited, which is advantageous over protein target detected methods.[23] Furthermore, the extent of ligand binding and size of the protein target bound directly correlates to line width changes.[52] As such, larger proteins usually produce stronger effects, increasing the sensitivity of relaxation experiments.[3,53] Additionally, there is no need for multidimensional or correlation experiments, simplifying the experimental setup.

**C.3.2. Exchange Transferred NOEs**

To determine the specific conformation of weak to medium affinity ligands when bound to a large protein target, the one- and two-dimensional exchange transferred NOE experiments (etNOE) have been used extensively.[52,79] Like relaxation and diffusion-based experiments, the

use of the transferred nuclear Overhauser effects (NOEs) also depends on the ability of the ligand to acquire the relaxation properties of the protein target.[50] Upon binding of a ligand to a protein target, the NOEs observed for the free ligand change because the molecule retains a "memory" of its bound state properties. The ligand acquires distinct magnetic properties while bound to the protein, which are retained even when the complex dissociates.[80]

Direct magnetic interactions between nuclear dipoles located close to each other in space result in dipolar coupling, leading to time-dependent mutual cross-relaxation. These time-dependent dipolar couplings provide an important source of nuclear relaxation in solution. They also provide incoherent magnetization transfer pathways that result in direct through-space interactions between nuclei called nuclear Overhauser effects (NOEs). Experimental observation of the NOE requires a change in signal intensity from one spin after the perturbation of a nearby spin with radio-frequency excitation. The rate constants governing cross-relaxation depend on the distance between two nuclei and the rotational mobility of the vector connecting the two atoms.[22,54] NOEs are sensitive probes of short-range, through-space, intramolecular and intermolecular interactions,[9] because the intensity of the NOE falls off rapidly with increasing distance ($1/r^6$).[37] As such, the NOE provides relative spatial information for structure determination, and allows the use of solution NMR to examine the interactions between nuclei that are not necessarily covalently bound but are within five angstroms of each other.[22] Differences in Brownian motions of the molecular structure also intimately reflect alterations in NOE patterns. For rapidly tumbling small molecules, two-dimensional $^1H$-$^1H$ nuclear Overhauser spectroscopy (NOESY) experiments result in negative NOE cross peaks with weak intensity. For slowly tumbling large molecules, stronger positive cross peaks result.[6,9,22] The differences in sign

and intensity of the NOE serve as the basis for extracting information about binding from etNOE experiments.

When a ligand binds to the protein target, the ligand protons experience additional dipole-dipole interactions from the protein's protons. In addition, the ligand transiently adopts the tumbling properties of the large protein target. Detection of binding relies on intraligand NOEs that develop in the bound state, where the dipole-dipole interaction caused by the decreased molecular tumbling rate occurs much more efficiently than in the free state. As such, positive NOE cross peaks for small molecules in the presence of the protein target clearly indicate binding. Unbound ligand molecules produce weak, negative NOE cross peaks typical of small molecules. These so-called exchange transferred NOEs have great utility in the examination of the interactions between ligands and protein targets.[3,9,22,54,79] etNOEs were used to examine the interactions between the antibody SM3 and the aberrantly glycosylated MUC-1 protein (which is associated with breast cancer) because specificity of the SM3 antibody to cancerous MUC-1 glycoproteins may result from decreased carbohydrate chain length or a conformational change. etNOES were used to determine the conformation of differentially glycosylated MUC-1 peptides bound to the SM3 antibody and found that the less glycosylated form is bound preferentially because the peptide acquires a more knob-like structure. The ability to determine the conformation of the MUC-1 peptides bound to SM3 has facilitated the development of SM3 for diagnostic and therapeutic purposes.[81]

Once a ligand has been confirmed to bind to a protein by methods such as etNOEs, two recently developed techniques may be used to facilitate identification of the bound ligand's orientation without the need for complete structure determination of the protein-ligand complex under investigation. The INPHARMA (Protein Mediated Interligand NOEs for Pharmacophore

Mapping) method was developed to investigate the binding orientation of a competing ligand when a structure of a similar protein-ligand complex is known.[82] Interligand NOEs are used to gain information about the binding mode of one ligand with respect to the other. Interligand NOEs arise between the known binder and new compound, which are used to determine the orientation of the new complex without the need for complete structure identification. The SAR by ILOEs (interligand nuclear Overhauser effect) approach uses etNOEs to identify small molecule ligands bound to the protein target simultaneously in close proximity to each other. Interligand NOEs then are used to provide information about the orientation of ligands, which subsequently may be linked to yield a high-affinity lead compound in the correct conformation.[83] The SAR by ILOEs approach is advantageous over the SAR by NMR method previously described because it provides information about the relative orientation of the two compounds in the binding pocket without needing to solve the structure of the ternary complex.[80]

Limitations to the etNOE method's utility include the sensitivity of the $^{1}H$-$^{1}H$ NOESY experiment itself. The low-sensitivity of the two-dimensional homonuclear NOE experiment requires that relatively long acquisition times are used to provide spectra of sufficient quality and precludes the high-throughput analysis of large libraries of compounds.[3,7] Furthermore, the spectrum contains strong diagonal peaks that prohibit the observation of cross peaks between ligand resonances having similar chemical shifts. The strong diagonal peaks may also introduce noise or baseline problems that interfere with the observation of weak cross peaks.[52] An additional constraint is that the transferred-NOE experiment most efficiently characterizes ligands with a $K_D$ value in the μM to mM range. To observe changes in NOEs, the dissociation of the ligand must occur quickly enough such that a sufficient percentage of the free ligand will remember the bound state and generate intense, positive NOE cross peaks. If the on rate is too

slow (much slower than diffusion allows), the transfer NOE will be too weak for detection.[6,23] The dissociation constants of slowly exchanging systems can be determined using the etNOE experiment but requires a complex mathematical treatment, which has been reviewed previously and is beyond the scope of this paper.[80,84] This limits the transfer NOE technique to weakly binding ligands in fast-exchange and, as such, requires large ligand concentrations in excess of the protein target (10:1-20:1 ligand:target ratio). This large ratio introduces the possibility of secondary binding, which may hamper unambiguous interpretation of the results. Also, compounds of limited solubility can aggregate in solution and exhibit large positive NOE cross peaks in the absence of the protein target, producing false positives.[9,52,85] False positives are easily identified by examining control samples containing only the small molecules. Nonetheless, the transferred NOE method has great advantages because of its ability to generate information on both the binding affinity and the geometry of the ligand in the bound state from the same experiment. etNOE experiments are often employed as one of the final steps in the SAR by NMR method, when sequential assignments of the labeled protein target are known, to determine the orientation of two adjacently bound ligands to a protein target.[25] Although it is possible to attain even higher resolution information when coupled with sequential assignments, the transferred NOE method always retains the important advantage of being able to provide structural information about ligands interacting with unlabeled protein targets of unlimited molecular weight.

**C.3.3. Exchange-Mediated Saturation Transfer Experiments**

The second category of ligand-detected methods relies on exchange-mediated transfer of another type of bound-state information to the free state.[50] The basis of this type of NMR experiment is similar to transferred NOE experiments in that magnetization transfer can occur

between ligand and target through NOEs.[23,86] The difference from etNOE experiments is the fact that detection of binding does not rely on the bound ligand exhibiting transient relaxation properties of the protein target. Instead, the protein target is saturated with rf-energy, which is transferred to an exchanging ligand to affect the one-dimensional proton signal intensities of the ligand. The Saturation Transfer Difference (STD) experiment utilizes this form of magnetization transfer in conjunction with difference spectroscopy and manifests the binding interaction as an apparent decrease in signal intensity of the ligand. The STD experiment is advantageous because it provides more sensitivity than previously described ligand-detected methods and may be used to quantify binding affinities between ligand and target and identify the binding epitope of the ligand.[87]

STD experiments require the application of a train of selective radio-frequency pulses until the protons in the protein target can no longer absorb any more energy (saturation). Ligands that come in close contact with the target receive part of this saturation via intermolecular $^{1}H$-$^{1}H$ cross relaxation pathways, which results in a decrease in the ligand's line intensity. An apparent decrease in the one-dimensional signal intensity of the ligand-resonances upon subtraction of a reference spectrum (lacking target saturation, and as such, magnetization transfer) accurately indicates binding of the ligand to the protein target.[21] STD studies often employ extremely low, undetectable concentrations of the target, and ideally the resulting difference spectrum contains only ligand signals corresponding to protons making NOE contacts with the protein; however, for studies requiring high concentration of the protein target, relaxation filters may be applied to suppress the signals from the protein.[3,54] Differences in the observed signal intensities of the resonances from atoms in the ligand indicate which moieties are embedded in the protein-ligand interface. Because the magnetization transfer depends on the distance between the ligand and

protein target, proximal interaction leads to more efficient transfer of saturation, which in turn leads to a greater loss of signal, allowing identification of the ligand's binding epitope with the difference spectrum.[3] This method was used to characterize the binding epitope of the cyclic RGD peptides that inhibit fibrinogen binding to integrins to prevent platelet aggregation and to permit optimization of these peptides as potential drug candidates. Conformational changes of the RGD peptide significantly impact selectivity for specific integrins involved in clot formation, and the appropriate conformation was identified to selectively target these peptides to the correct integrin type.[88] When used with isotopic labeling strategies (See Advances in NMR; Isotopic Labeling), STD NMR can identify the ligand binding epitope and the amino acid composition of the ligand-binding site simultaneously. This setup forms the basis of the SOS-NMR experiment (Structural information using Overhauser effects and Selective labeling). SOS utilizes STD NMR to examine a ligand complexed to a series of selectively protonated protein target samples to characterize the amino acid composition of the ligand-binding site, while concomitantly identifying the ligand binding epitope.[89]

The STD experiment offers more sensitivity than relaxation, diffusion, and transfer NOE experiments.[52] The increased sensitivity depends on the ability to selectively pulse the protons of the protein target and the ability of the pulse to saturate all of its spins. Typically, the radio-frequency pulse used to saturate the protein resonances occurs at ppm values that lack ligand signals in the outermost region of the spectrum (typically below -2 or above 10 ppm). Because the pulse only perturbs selective protons of the protein target, the saturation must transfer to the rest of the protein signals via dipolar interactions and, as such, depends on the protein's ability to transfer energy. STD NMR is ideally suited for protein targets of 30,000 molecular weight or greater, because the saturation disperses more quickly to the other protons of the protein target

that were not directly perturbed by the selective radio-frequency pulse. The larger the protein, the more efficient transfer of magnetization is to other nuclei, including the bound ligand, making this technique an extremely useful alternative to protein target detected methods for species of large molecular weight.[21,23,87] Additionally, the better sensitivity of the STD experiment means that these experiments require even smaller concentrations of the protein target; however, optimal setup, use, and interpretation of the STD experiments require some familiarity with the exchange processes at work, and as such, some prior knowledge the protein target's structure or function.

Ligands with $K_D$ values in the range of $10^{-8}$-$10^{-3}$ are ideally suited for STD methods. For extremely weak binders, too many protein target molecules will lack a ligand in the binding site to achieve detectable signal by STD. For tight binders, the exchange occurs so infrequently that the pool of free-ligand is barely influenced by its visit to the saturated protein target and relaxes back to equilibrium before it is released from the protein. The population of saturated ligands will not influence the total ligand magnetization to produce a measurable STD signal.[21] Despite these limitations, a broad range of binding constants can be evaluated. By expressing the signal intensity of the STD spectrum as a fraction of the intensity of an unsaturated reference spectrum $[(I_0-I_{sat})/I_0]$, the binding affinity can be calculated. In the preceding equation, the intensity of the reference peak is $I_0$ and the intensity of the peak after saturation is $I_{sat}$. A value of 0.5 corresponds to 50% saturation of the ligands (equating to 50% of the ligands in exchange with the protein target).[86] This value is then multiplied by the ligand:protein ratio to generate the STD factor. For a series of ligand concentrations, plotting the STD factor versus ligand concentration generates a curve from which the $K_D$ value is determined. The ability to accurately quantify binding constants is a unique advantage of this NMR method.

An important adaptation of the STD method is the waterLOGSY (water-ligand observed via gradient spectroscopy) experiment.[90,91] The aim of both methods is to transfer magnetization to the bound ligand and detect differences in the spectrum of the free ligand. Where the STD method achieves this by directly saturating the protein, the waterLOGSY method accomplishes transfer indirectly through excitation of the bulk water magnetization. The excited water interacts with the protein-ligand complex transferring the magnetization, which is subsequently retained by the ligand after it dissociates. Bound ligand directly interacts with bound water molecules within the binding site via cross-relaxation mechanisms. Alternatively, labile protons (NH and OH groups) in the protein target exchange with bulk water and acquire the saturated magnetization previously applied. Then the magnetization is propagated to the free ligand when the water molecules interact directly with the bound ligand at the protein-ligand interface via the same types of intermolecular cross-relaxation mechanisms.[21] In both of these cases, the magnetization transfer pathway conserves the sign of the magnetization from the bulk water and, as such, non-binding compounds appear with opposite sign compared to the water signal. Nonbinders and binders are easily discriminated from each other in a mixture because they give waterLOGSY signals of opposite sign.[21,91] False positives may be observed if magnetization is transferred directly to the free ligand via chemical exchange of bulk water with exchangeable ligand protons. The exchangeable proton resonances, when visible, may complicate the interpretation of waterLOGSY data. To overcome this problem, a waterLOGSY spectrum for the free ligands without protein target can be collected as a control.[21] Like all ligand-based screening methods, the waterLOGSY experiment is limited to the detection of weakly binding ligands in the low μM range. Ligands with a tighter affinity will stay bound to the protein too long and relax back to their original state before dissociating. An important advantage of the

waterLOGSY method is that it is much more sensitive than other ligand-detected methods. This may be attributed to the large excess of solvent available and the large number of exchangeable protons in a protein-ligand complex. The exceptional sensitivity of this method permits the use of nanomolar concentrations of unlabeled protein target, making it a powerful method for primary screening of compounds that bind to therapeutically relevant protein targets.[91]

## C.4. ADVANCES IN NMR

Many recent advances in solution NMR have served as tools to overcome the limitations encountered with each screening methodology discussed. Advances in instrumentation and innovative experimental protocols (pulse sequences) have been utilized to overcome the problems of fast relaxing signals and lack of resolution with higher molecular weight proteins. Isotopic labeling strategies have facilitated the study of higher molecular weight protein targets (up to 900,000 kDa) by decreasing the number of observable signals and reducing spectral crowding. In addition, computational methods have improved the amount of information gleaned from limited NMR data sets. The development of competition experiments has expanded the range of binding affinities that can be evaluated by ligand-detected screening methods. These advances have contributed to a more extensive use of NMR as a high-throughput screening tool.

### C.4.1. Instrumentation Advances

Instrumentation development has rapidly improved the quality of data available from NMR experiments, especially the development of higher frequency spectrometers. Increasing the magnetic field strength simplifies complex or overlapped spectra by generating more dispersion along the observed chemical shift axis, because the chemical shift dispersion increases linearly with field strength, while the line widths of aliphatic protons and carbons remain unchanged.[12]

Higher field strength not only resolves overlapping signals but also increases the NMR sensitivity because stronger magnets achieve better signal-to-noise, increasing roughly as the 3/2 power of the magnetic field.[39] For example, moving from a 400 MHz spectrometer to an 800 MHz spectrometer increases the sensitivity by a factor of approximately three and resolution by a factor of two. Because of this, field strengths of 14.1 Tesla (600 MHz) and above are routinely utilized for drug screening and development. Additionally, magnets with proton resonance frequencies up to 1 GHz are becoming available and are employed currently for the study of higher molecular weight protein targets.[92] Because of the increase in signal-to-noise, better spectra can be recorded in less time using less material. This has greatly improved the quality, speed, and resolution of data acquired for protein target structure determination and ligand interactions.[16]

Increasing the sensitivity of the NMR signals observed may be accomplished by either an increase in the signals or decrease in noise. Increasing the sample concentration or improving the field strength can enhance the signal-to-noise by directly increasing the intensity of the resonances. Cryogenically cooled probes (probes which have had their electronics cooled to 20-25 K) greatly diminish the level of thermal noise, increasing the signal to noise ratio by a factor of three or more and, consequently, improving the sensitivity of the NMR experiment.[92] This is important because a three-fold enhancement in sensitivity corresponds to approximately a ten-fold shorter experiment time or a three-fold decrease in the protein concentration needed to achieve the same signal-to-noise with a room temperature probe.[3] A cryogenic probe can effectively convert a lower field instrument (500 MHz) to the equivalence of a higher field instrument (800 MHz) at a greatly reduced cost compared to obtaining a higher field spectrometer. It should be noted that the signal-to-noise in a cryogenic probe depends strongly

on the conductivity and dielectric losses of the solution, such that an increase in ionic strength corresponds to an increase in noise. At approximately 100 mM salt concentration in a standard 5 mm tube, the signal-to-noise advantage over a conventional room temperature probe begins to diminish.[93-96] This can be overcome by using a narrow diameter tube, which positions the ions in a region with the lowest electric field; however, the sample concentration must be increased to account for the decrease in volume being analyzed. Using a cryogenic probe with low ionic strength solutions, 2D $^{1}H$-$^{15}N$ HSQC spectra can be obtained in a few hours on protein samples as low as 50 μM. This advance has facilitated high-throughput analysis of binding interactions, because minimizing the protein concentration permits a greater number of compounds to be screened simultaneously: 100 compounds at 50 μM each may be used while keeping the total concentration of added components to 5 mM. This increases the throughput of the NMR screening assay by ten fold, because earlier studies used concentrations of one millimolar protein target and were limited to ten compounds in each mixture, allowing a maximum throughput of 1000 compounds per day. Using this strategy, more than 10,000 compounds can be screened in one day, greatly facilitating the drug discovery process.[16,92,97] This can be an advantage when examining systems with very low hit rates.[41]

**C.4.2. Pulse Sequence Improvements**

The greatest limitation of early protein target observed NMR screening was the inability to observe higher molecular weight protein targets because the relaxation mechanisms of the larger proteins cause lines to broaden and signals to disappear; however, the introduction of Tansverse Relaxation Optimized SpectroscopY (TROSY) has supported the investigation of many protein targets up to a molecular weight of 100 kDa.[98,99] Protein target NMR screening studies require well-resolved, high-quality spectra to determine the involvement of specific

residues in the binding reaction. The slower tumbling of larger proteins results in very efficient spin-spin relaxation between protons and primarily affects transverse relaxation processes. This effect is reflected in the line widths of the observed resonances, making these basic spectral requirements harder to fulfill.[92] The TROSY technique reduces signal loss due to increased relaxation rates and consequent line broadening by exploiting the differential relaxation effects arising from chemical shift anisotropy (CSA) and dipole-dipole couplings, which are quite significant at higher magnetic fields (greater than 500 MHz). The measurement of these two phenomena using an uncoupled $^{1}H$-$^{15}N$ HSQC spectrum results in four differently shaped peaks, only one of which is sharp and narrow. The TROSY protocol detects only the sharp, narrow peak resulting in a better-resolved NMR spectrum for higher molecular weight protein targets.[16,22,98] Dipole-dipole couplings arise in proteins independent of the field strength, while CSA increases at higher fields. Consequently, an optimal field strength exists at which the rate in transverse relaxation in TROSY experiments approaches zero (1 GHz).[9] As such, the increase in sensitivity due to the TROSY pulse sequence is more pronounced at higher field,[3,16] further illustrating the importance of higher field instruments. With TROSY, the molecular size of proteins accessible for detailed NMR investigations has been extended several fold. The approach may be applied to a variety of NMR experiments, including two- and three-dimensional experiments used for sequential assignments to identify important residues of the protein target involved in binding.[13] To fully benefit from the line narrowing effects of TROSY, other relaxation mechanisms should be suppressed. These mechanisms include proton-proton dipolar relaxation between NH and CH protons and intermediate exchange broadening due to segmental motion. Deuteration of the amino acid side chains may be performed to facilitate the suppression of these relaxation mechanisms.

Cross Relaxation-Enhance Polarization Transfer (CRINEPT) and Cross Relaxation-Induced Polarization Transfer (CRIPT) experiments can yield up to a three-fold signal-to-noise enhancement for amide groups and have been applied to a 110 kDa protein[100] and 900 kDa protein complex.[101] In a correlation experiment such as the $^{1}H$-$^{15}N$ HSQC, the pulse sequence includes a step in which transfer of the magnetization of a sensitive nucleus ($^{1}H$) to an insensitive nucleus ($^{15}N$) via spin-spin couplings takes place (Insensitive Nuclei Enhanced by Polarization Transfer; INEPT). This key step has been incorporated into many of the multidimensional NMR experiments used for sequential assignments and structure determination. The efficiency of the INEPT sequences depends on the strength of the magnetization of the sensitive nuclei, and as such, also depends on the transverse relaxation rate (an increase in $R_2$ causes rapid deterioration of the magnetization). For protein targets with molecular weights beyond 100 kDa, transverse relaxation during the transfer time may become a limiting factor because the magnetization of a fast relaxing $^{1}H$ signal is relatively short lived and cannot transfer its magnetization to the bonded $^{15}N$. CRINEPT and CRIPT experiments utilize cross-relaxation pathways (NOEs) to transfer the magnetization to the $^{15}N$ nucleus. An increase in molecular weight increases the efficiency of cross-relaxation pathways, increasing the sensitivity of the experiment for larger protein targets. The CRINEPT and CRIPT experiments permit high-resolution analysis of membrane bound protein targets and protein targets that oligomerize in solution.[7,100]

While TROSY based experiments address the relaxation and resolution difficulties of higher molecular weight proteins, the problem of resonance overlap due to a large number of amino acids and corresponding backbone amides remains a major obstacle for resonance assignment and chemical shift perturbation studies. The Solvent Exposed Amide-TROSY (SEA-TROSY) experiment provides a modification of the $^{1}H$-$^{15}N$ correlation experiments to reduce the

problem of resonance overlap in very large proteins. This type of pulse sequence assumes that only the amides exposed to the solvent contribute to the intermolecular interactions involved in binding, so that those buried in the core of the protein may be eliminated and ignored. After a partially or completely deuterated protein sample (see next section) is dissolved in water, the labile deuterons near the surface of the protein will exchange rapidly with protons from the aqueous bulk solution. The pulse sequence manipulates the rf-energy in such a way as to eliminate any magnetization generated from amide protons, without modifying the water signal. The water magnetization exchanges with exposed amide protons, which can be detected with a TROSY-type correlation experiment. This eliminates any signals that may result from amides in the core of the protein. The resulting spectrum contains fewer resonances with peaks for only the amides near the surface of the protein, which then may be monitored upon the addition of small molecule ligands to examine interactions between ligands and surface residues.[102]

**C.4.3. Isotopic Labeling Strategies**

Many newly emerging isotope-labeling techniques address the limitations of examining large protein targets by solution NMR. Specifically, the overlap of NMR signals from large protein targets precludes full interpretation of the spectral data, and selective isotopic labeling strategies can permit observation of only specific moieties, reducing the number of signals and decreasing spectral crowding.[103] For example, the use of deuteration at non-exchangeable sites (aliphatic sites) in combination with uniformly $^{13}C$- and $^{15}N$-labeling dramatically improves the quality of NMR spectra of larger proteins. Expressing the protein in partially or completely deuterated growth media followed by purification in protonated buffers causes most of the deuterated amides at the surface of the protein to be replaced by protons, permitting the $^{1}H$-$^{15}N$ HSQC to be utilized for screening.[22] Replacing protons in the core of a large protein with

deuterons not only improves spectral overlap by decreasing the number of observable signals but also results in a narrowing of the remaining proton signals,[22] increasing the overall quality of the spectra obtained. This effect occurs because deuteration results in a considerable reduction in the proton spin density, which decreases proton spin-spin interactions and, therefore, the overall relaxation rate of the protein target.[104] Additionally, the inclusion of specifically protonated amino acids or amino acid precursors in $D_2O$ growth media permits selective retention of protons at aliphatic resiudes (Ala, Val, Leu, Ile) or aromatic side chains (Phe, Tyr), which yields important structural information about protein target hydrophobic cores.[105-107] Phe and Tyr side chains are often located at binding interfaces as well,[103] and as such, protonation at these specific residues may be used to detect direct binding of a ligand.

Selective $^{13}$C-labeling of Val, Leu and Ile methyl groups and subsequent $^{1}$H-$^{13}$C HSQC chemical shift perturbation assays results in a three-fold increase in sensitivity compared to the $^{1}$H-$^{15}$N experiment.[36] This labeling strategy, of course, improves the molecular weight limit of the protein target and has been used to study proteins with as many as 723 amino acids.[108] More importantly, $^{13}$C-methyl labeling greatly simplifies the $^{1}$H-$^{13}$C HSQC spectrum of the protein target and makes it amenable to implementation in high-throughput protocols. Additionally, the increase in signal-to-noise of the $^{1}$H-$^{13}$C cross peaks shortens the required acquisition time and permits a lower concentration of protein to be used, facilitating the screening of many ligand mixtures in one day.[7] $^{13}$C-methyl labeling utilizes $^{13}$C-labeled amino acid precursors instead of $^{13}$C-labeled glucose, making the cost of protein preparation comparable to uniform $^{15}$N-labeling.[36]

With prior knowledge of the structure or function of a protein, a variety of $^{15}$N- and $^{13}$C-labeling techniques can be used to simplify the spectrum observed upon addition of ligand to

provide information about a specific region of interest in the protein target. Selective enrichment, applicable to many naturally occurring amino acids, is accomplished using supplemented media with one or more specific $^{15}$N-amino acids.[109] For example, the specific $^{15}$N labeling of active site residues simplifies the $^{1}$H-$^{15}$N HSQC perturbation assay. With knowledge of the protein target's function unique amino acid sequences can be selectively labeled and used to detect binding using a simple approach. $^{13}$C and $^{15}$N labeling of two consecutive amino acids respectively ($^{13}$C-X, $^{15}$N-Y) in the active site of the protein target will yield only one signal or a limited number of signals in a $^{13}$C-$^{15}$N-$^{1}$H three-dimensional correlation experiment (HNCO),[38] provided that only one or a limited number of the chosen amino acid pairs occur in the protein sequence. This particular labeling strategy facilitates the identification of ligand binding to a selective site without the need for sequence specific assignments because only adjoined $^{13}$C-$^{15}$N labeled nuclei are detected with the HNCO experiment (see Figure 3).[110] In another application, prior knowledge of the protein's structure or function can guide segmental $^{13}$C and $^{15}$N-labeling of multidomain protein targets via trans-splicing[111,112] or chemical ligation techniques.[113] This reduces the number of observed signals in the correlation experiments, simplifies the analysis of chemical shifts upon titration of ligand and permits the evaluation of multidomain proteins in their native form. Selective $^{15}$N-labeling of specific amino acids is an additional approach that facilitates evaluation of larger molecular weight proteins as well as multidomain proteins. This type of labeling scheme also permits structural investigations of proteins that undergo allosteric regulation. For example, induced conformational changes of the acetylcholine binding protein (AChBP) pentamer were studied by selectively labeling Cys residues in the protein with $^{15}$N. Conformational changes in the Cys-loop distant from the binding site were observed upon the

addition of ACh, which confirmed the important regulatory role of the Cys-loop to AChBP function.[114]

### C.4.4. Computational Methodologies

When full structure determination by NMR is not practical, computational strategies can be used to generate structural information about the protein-ligand complex to guide drug development using a limited amount of NMR data. NMR-DOC[115] (Nuclear Magnetic Resonance DOcking of Compounds) can be used to develop models of the ligand bound to the protein target once the residues involved in binding have been identified. To determine important residues involved in binding, selective isotopic labeling is performed to observe a small number of key residues in the binding site, eliminating the need for complete sequential assignments during the screening process. Using a previously determined x-ray or solution structure of the protein as a starting point and experimentally measured intermolecular NOEs that arise between the ligand and the selectively labeled protein, models are developed that dock the ligand in the binding site. Selective labeling simplifies interpretation of the NOE spectra because fewer signals are present. This method has great utility because it can be used to examine higher molecular weight proteins with only a few NMR experiments and limited data analysis.

NMR-SOLVE[115] (Structurally Oriented Library Valency Engineering) is a very useful alternative to the SAR by NMR method of drug screening because it is applied to different protein targets that belong to a class of proteins related by binding site properties (pharmacofamily), as opposed to one protein target at a time. Proteins are classified into families based on common structural and functional features. Those from the same family typically share similar attributes, which may be exploited to narrow the search space to more rapidly identify high-affinity lead compounds. With NMR-SOLVE, the development of high-affinity ligands is

tailored to the family of protein targets related by the presence of a common binding site adjacent to a variable binding site.[116] A known binder (often a cofactor) and its interactions with the common binding site are first studied by protein-detected NMR to identify important residues involved in binding. Common ligand mimics that bind with low affinity to a model member of the protein-family are then investigated by the same means. Intermolecular NOEs that develop between the ligand and the common and variable binding sites are used to develop NMR-DOC models based on the previously determined structure of the reference family member. The model of the low affinity ligand bound to the common binding site is then used to design a linker from which various molecular fragments can be added. Because the protein family selected has the same geometric relationship between the common and variable sites, the addition of various molecular fragments to the linker yields a compound library tailored to a given protein family without specific knowledge of their three-dimensional structure.[115] Once the compound library has been constructed, screening of additional family members may be performed using other methods with a higher probability of finding a high-affinity (nanomolar) hit. This approach may prove to be an important drug design strategy in the post-genomics era.[9]

The protein-ligand NOE matching[117] method was developed in order to bypass the requirement of backbone resonance assignments of a protein's binding site in order to determine the location, orientation, and internal conformation (binding pose) of a ligand bound to the protein. The binding pose of the complex can be determined at high-resolution using isotope-filtered NMR experiments[118] but may be difficult to accomplish in a reasonable time frame, as the time-consuming task of assigning binding site residues for each protein-ligand complex studied does not support repetitive cycles of structure-based ligand design. In this method, trial binding poses are scored based on matching an experimental pattern of intermolecular protein-

ligand NOEs to predicted NOE patterns based on theoretical calculations. Trial protein-ligand binding poses are generated by any suitable computational docking method and are based on a previously determined structure (either X-ray or NMR) of a protein target of interest. For each binding pose, the intermolecular NOE pattern is predicted and matched to the experimental patterns. The scoring is based on a fast, deterministic algorithm, making it suitable for scoring a large number of trials in a short amount of time. Only the proton assignments for the bound ligand are required, making this a useful alternative to traditional methods when a structure of the unbound protein is available.[117]

**C.4.5. Competition Experiments**

While it is possible to examine a wide range of binding affinities by NMR, ligand-detected NMR screening has been expanded further to detect strongly binding ligands with slow dissociation rates by the use of competition assays. The inability to observe tightly bound ligands directly results from the need to distinguish the ligand and target signals by exploiting size-dependent effects. Ligand-based screening methods detect only the free ligand signals and require rapid exchange between free and bound forms. To detect high-affinity ligands, competition-based experiments have been developed that rely on $^{1}H$ or $^{19}F$ relaxation and the STD NMR or waterLOGSY experiments because these methods are amenable to high-throughput protocols in which mixtures of compounds can be screened simultaneously.[58,119-120,121]

Competition-based experiments utilize a reporter ligand that is in rapid exchange between free and bound forms with the protein target of interest.[21] Tight binding by another compound results in displacement of the reporter.[4] Displacement is manifested as a shift of the reporter compound's NMR parameters toward those of the free state.[21] Fluorine moieties on the reporter

molecule are especially useful because they facilitate the rapid high-throughput screening of large chemical mixtures against the protein target of interest because of the absence of overlap with signals from the mixtures to be screened.[58] Competition experiments permit the detection of high-affinity molecules, significantly expanding the $K_D$ range amenable to ligand-based NMR screening to include high-affinity ligands.

## C.5. CONCLUSIONS

High-field NMR is a powerful tool for the investigation of transiently forming complexes, and has become increasingly useful as the methodologies available to study these complexes expand. Methodologies are available to assess protein-ligand interactions from either the perspective of the protein target or the ligand molecule. They take advantage of the differences in the fundamental properties of small and large molecules as well as the presence of unique connectivities or atoms within the molecules to selectively detect binding. Advances in NMR technology and experimentation have greatly increased the size of molecules that can be analyzed, as well as increasing the throughput, sensitivity and resolution of the technique. As a result, high-field solution NMR plays a significant role in the identification of binding partners and the physicochemical characterization of protein-ligand interactions and has been applied successfully to advance pharmaceutical research. NMR's unique ability to examine weak interactions and to structurally characterize binding events has substantially improved rational drug design and development and the analysis of potential therapeutics to protein targets. Given the versatility of high-field NMR and the continual innovation of the drug evaluation process, NMR is likely to find further applications in various aspects of drug development programs.

## C.6. REFERENCES


1. Hopkins A. L. & Groom C. R. (2002) The druggable genome. *Nat Rev Drug Discov*, 1, 727-730.
2. Mulder F. A., Hon B., Muhandiram D. R., Dahlquist F. W. & Kay L. E. (2000) Flexibitility and ligand exchange in a buried cavity mutant of T4 lysozyme studied by multinuclear NMR. *Biochem*, 39, 12614-12622.
3. Wyss D. F., McCoy M. A. & Senior M. M. (2002) NMR-Based Approaches for Lead Discovery. *Curr Opin Drug Discov Devel*, 5, 630-647.
4. Coles M., Heller M. & Kessler H. (2003) NMR-based Screening Technologies. *Drug Discov Today*, 8, 803-810.
5. Jahnke W. (2002) Spin labels as a tool to identify and characterize protein-ligand interactions by NMR spectroscopy. *Chembiochem*, 3, 167-173.
6. Diercks T., Coles M. & Kessler H. (2001) Applications of NMR in drug discovery. *Curr Opin Chem Biol*, 5, 285-291.
7. Salvatella X. & Giralt E. (2003) NMR-based methods and strategies for drug discovery. *Chem Soc Rev*, 32, 365-372.
8. Stockman B. J. (1998) NMR Spectroscopy as a tool for structure-based drug design. *Prog Nucl Magn Reson Spectrosc*, 33, 109-151.
9. Pellecchia M., Sem D. S. & Wuthrich K. (2002) NMR in drug discovery. *Nat Rev Drug Discov*, 1, 211-219.
10. Sanders J. K. M. & Hunter B. K. (1987) *Modern NMR Spectroscopy: A Guide for Chemists*. 1st ed., Oxford: Oxford University Press. p 308.
11. Primrose W. U. (1993) Sample Preparaion. In Roberts G. C. K., editor NMR of Macromolecules, ed., New York: Oxford University Press. p 7-34.
12. Evans J. N. S. (1995) *Biomolecular NMR Spectroscopy*. 1st ed., Oxford: Oxford University Press. p 444.
13. Wider G. (2000) Structure determination of biological macromolecules in solution using NMR spectroscopy. *BioTechniques*, 29, 1278-1294.
14. Levy G. C. & Craik D. J. (1981) Recent Developments in Nuclear Magnetic Resonance Spectroscopy. *Science*, 214, 291-299.
15. Roberts G. C. (2000) Applications of NMR in drug discovery. *Drug Discov Today*, 5, 230-240.
16. Wishart D. (2005) NMR spectroscopy and protein structure determination: applications to drug discovery and development. *Curr Pharm Biotechnol*, 6, 105-120.
17. Fielding L. (2000) Determination of association constants ($K_a$) from solution NMR data. *Tetrahedron*, 56, 6151-6170.
18. Fielding L. (2003) NMR methods for the determination of protein-ligand dissociation constants. *Curr Top Med Chem*, 3, 39-53.
19. Fielding L. (2007) NMR methods for the determination of protein-ligand dissociation constants. *Prog Nucl Magn Reson Spectrosc*, 51, 219-242.
20. Medek A., Hajduk P., Mack J. & Fesik S. W. (2000) The use of differential chemical shifts for determination of the binding site location and orientation of the protein-bound ligands. *J Am Chem Soc*, 122, 1241-1242.
21. Lepre C. A., Moore J. M. & Peng J. W. (2004) Theory and applications of NMR-based screening in pharmaceutical research. *Chem Rev*, 104, 3641-3676.

22. Pochapsky S. S. & Pochapsky T. C. (2001) Nuclear magnetic resonance as a tool in drug discovery, metabolism and disposition. *Curr Top Med Chem*, 1, 427-441.
23. Carlomagno T. (2005) Ligand-target interactions: what can we learn from NMR? *Annu Rev Biophys Biomol Struct*, 34, 245-266.
24. Reibarkh M., Malia T. J. & Wagner G. (2006) NMR distinction of single- and multiple-mode binding of small-molecule protein ligands. *J Am Chem Soc*, 128, 2160-2161.
25. Shuker S. B., Hajduk P. J., Meadows R. P. & Fesik S. W. (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science*, 274, 1531-1534.
26. Zartler E. R., Hanson J., Jones B. E., Kline A. D., Martin G., Mo H., Shapiro M. J., Wang R., Wu H. & Yan J. (2003) RAMPED-UP NMR: multiplexed NMR-based screening for drug discovery. *J Am Chem Soc*, 125, 10941-10946.
27. Kallen J., Spitzfaden C., Zurini M. G. M., Wider G., Widmer H., Wuthrich K. & Walkinshaw M. (1991) Structure of human cyclophilin and its binding site for cyclosporin A determined by X-ray crysallography and NMR spectroscopy. *Nature*, 353, 276-279.
28. Clubb R. T., Omichinski J. G., Clore G. M. & Gronenborn A. M. (1994) Mapping the binding surface of interleukin-8 complexed with an N-terminal fragment of type 1 human interleukin-8 receptor. *FEBS Lett*, 338, 93-97.
29. Lian L.-Y., Barsukov I. L., Derrick J. P. & Robers G. C. K. (1994) Mapping the interactions between streptococcal protein G and the Fab fragment of IgG in solution. *Nat Struct Biol*, 1, 355-357.
30. Farmer B. T., Constantine K. L., Goldfarb V., Friedrichs M. S., Wittekind M., Yanchunas J. J., Robertson J. G. & Mueller L. (1996) Localizing the NADP+ binding site on the MurB enzyme by NMR. *Nat Struct Biol*, 3, 995-997.
31. Hajduk P. J., Boyd S., Netesheim D., Nienaber V., Severin J., Smith R., Davidson D., Rockway T. & Fesik S. W. (2000) Identification of novel inhibitors of urokinase via NMR-based screening. *J Med Chem*, 43, 3862-3866.
32. Hajduk P., Sheppard G., Nettesheim D. G., Olejniczak E. T., Shuker S. B., Meadows R. P., Steinman D. H., Carrera G. M. J., Marcotte P. A., Severin J., Walter K., Smith H., Gubbins E., Simmer R., Holzman T. F., Morgan D. W., Davidsen S. K., Summers J. B. & Fesik S. W. (1997) Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. *J Am Chem Soc*, 119, 5818-5827.
33. Hajduk P. J., Dinges J., Miknis G. F., Merlock M., Middleton T., Kempf D. J., Egan D. A., Walter K. A., Robins T. S., Shuker S. B., Holzman T. F. & Fesik S. W. (1997) NMR-based discovery of lead inhibitors that block DNA binding of human papillomavirus E2 protein. *J Med Chem*, 40, 3144-3150.
34. Hajduk P. J., Dinges J., Schkeryantz J. M., Janowick D., Kaminiski M., Tufano M., Augeri D. J., Petros A., Nienaber V., Zhong P., Hammond R., Coen M., Beutel B., Katz L. & Fesik S. W. (1999) Novel inhibitors of Erm methyltransferases from NMR and parallel synthesis. *J Med Chem*, 42, 3852-3859.
35. Lugovskoy A. A., Degterev A. I., Fahmy A. F., Zhou P., Gross J. D., Yuan J. & Wagner G. (2002) A novel approach for characterizing protein ligand complexes: molecular basis for specificity of small-molecule Bcl-2 inhibitors. *J Am Chem Soc*, 124, 1234-1240.
36. Hajduk P. J., Augeri D. J., Mack J., Mendoza R., Yang J., Betz S. F. & Fesik S. W. (2000) NMR-based screening of proteins containing $^{13}$C-labeled methyl groups. *J Am Chem Soc*, 122, 7898-7904.

37. Reid D. J. editor (1997) *Protein NMR Techniques*. 1st ed.: Humana Press. p 419.
38. Sattler M., Schleucher J. & Griesinger C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Magn Reson Spectrosc*, 34, 93-158.
39. Claridge T. D. W. (1999) *High Resolution NMR Techniques in Organic Chemistry*. 1st ed., Kidlington: Elsevier Science. p 382.
40. Jahnke W. & Widmer H. (2004) Protein NMR in Biomedical Research. *Cellular and Molecular Life Sciences*, 61, 580-599.
41. Mercier K. A. & Powers R. (2005) Determining the optimal size of small molecule mixtures for high throughput NMR screening. *J Biomol NMR*, 31, 243-258.
42. Ross A., Schlotterbeck G., Klaus W. & Senn H. (2000) Automation of NMR measurements and data evaluation for sytemically screening interactions of small molecules with target proteins. *J Biomol NMR*, 16, 139-146.
43. Ross A. & Senn H. (2001) Automation of Measurements and Data Evaluation in Biomolecular NMR Screening. *Drug Discov Today*, 6, 583-593.
44. Frieden C., Hoeltzli S. D. & Bann J. G. (2004) The preparation of $^{19}$F-labeled proteins for NMR studies. *Methods Enzymol*, 380, 400-415.
45. Yu L., Hajduk P. J., Mack J. & Olejniczak E. (2006) Structural Studies of Bcl-xL/ligand Complexes using $^{19}$F NMR. *J Biomol NMR*, 34, 221-227.
46. Danielson M. A. & Falke J. J. (1996) Use of $^{19}$F NMR to probe protein structure and conformational changes. *Annu Rev Biophys Biomol Struct*, 25, 163-195.
47. Gerig J. T. (1994) $^{19}$F NMR of proteins. *Prog Nucl Magn Reson Spectrosc*, 26, 293.
48. Leone M., Rodriguez-Mias R. A. & Pellecchia M. (2003) Selective incorporation of $^{19}$F-labeled Trp side chains for NMR-spectroscopy-based ligand-protein interaction studies. *Chembiochem*, 4, 649-650.
49. Huang Y., Rich R. L., Myszka D. G. & Wu H. (2003) Requirement of both the second and third BIR domains for the relief of X-linked inhibitor of apoptosis protein (XIAP)-mediated caspase inhibition by Smac. *J Biol Chem*, 278, 49517-49522.
50. Van Dongen M., Weigelt J., Uppenberg J., Schultz J. & Wikstrom M. (2002) Structure-based screening and design in drug discovery. *Drug Discov Today*, 7, 471-478.
51. Cavanagh J., Fairbrother W. J., Palmer A. G., III, Rance M. & Skelton N. (2007) *Protein NMR Spectroscopy: Principles and Practice*. 2nd ed., Boston: Academic Press.
52. Stockman B. J. & Dalvit C. (2002) NMR Screening Techniques in Drug Discovery and Drug Design. *Prog Nucl Magn Reson Spectrosc*, 41, 187-231.
53. Fejzo J., Lepre C. A., Peng J. W., Bemis G. W., Ajay, Murcko M. A. & Moore J. M. (1999) The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem Biol*, 6, 755-769.
54. Peng J. W., Lepre C. A., Fejzo J., Abdul-Manan N. & Moore J. M. (2001) Nuclear magnetic resonance-based approaches for lead generation in drug discovery. *Methods Enzymol*, 338, 202-230.
55. Fejzo J., Lepre C. A., Peng J. W., Su M. S.-S., Thomson J. A. & Moore J. M. (1996) Dynamic NMR studies of ligand-receptor interactions: Design and analysis of a rapidly exchanging complex of FKBP-12/FK506 with a 24 kDa clacineurin fragment. *Protein Science*, 5, 1917-1921.
56. Peng J. W. (2001) Cross-correlated $^{19}$F relaxation measurements for the study of fluorinated ligand-receptor interactions. *J Magn Reson*, 153, 32-47.

57. Huth J. R., Sun C., Sauer D. R. & Hajduk P. J. (2005) Utilization of NMR-derived fragment leads in drug design. *Methods Enzymol*, 394, 549-571.
58. Dalvit C., Flocco M., Veronesi M. & Stockman B. J. (2002) Fluorine-NMR competition binding experiments for high-throughput screening of large compound mixtures. *Comb Chem High Throughput Screen*, 5, 605-611.
59. Shimizu T. & Hatano M. (1983) Interaction of trifluoperazine with porcine calmodulin. $^{19}$F NMR and induced CD spectral studies. *FEBS Lett*, 160, 182-186.
60. Shimizu T., Hatano M., Muto Y. & Nozawa Y. (1984) Interaction of trifluoperazine with Tetrahymena calmodulin. A $^{19}$F NMR study. *FEBS Lett*, 166, 373-377.
61. Jenkins B. G. & Lauffer R. B. (1990) Detection of site-specific binding and co-binding of ligands to human serum albumin using 19F NMR. *Mol Pharmacol*, 37.
62. Kitamura K., Omran A. A., Takegami S., Tanaka R. & Kitade T. (2007) $^{19}$F NMR spectroscopic characterization of the interaction of niflumic acid with human serum albumin. *Anal Bioanal Chem*, 387, 2843-2848.
63. Crull G. B., Nardo J. V. & Dawson J. H. (1989) Direct observation of substrate binding to ferrous-CO cytochrome P-450-CAM using $^{19}$F NMR. *FEBS Lett*, 254, 39-42.
64. Crull G. B., Kennington J. W., Garber A. R., Ellis P. D. & Dawson J. H. (1989) $^{19}$F Nuclear magnetic resonance as a probe of the spatial relationship between the heme iron of cytochrome P-450 and its substrate. *J Biol Chem*, 264, 2649-2655.
65. Dalvit C., Ardini E., Flocco M., Fogliatto G. P., Mongelli N. & Veronesi M. (2003) A general NMR method for rapid, efficient, and reliable biochemical screening. *J Am Chem Soc*, 125, 14620-14625.
66. Dalvit C., Ardini E., Fogliatto G. P., Mongelli N. & Veronesi M. (2004) Reliable high-throughput functional screening with 3-FABS. *Drug Discov Today*, 9, 595-602.
67. Fernandez C. & Jahnke W. (2004) New approaches for NMR screening in drug discovery. *Drug Discov Today: Technol*, 1, 277-283.
68. Dalvit C. (2007) Ligand- and substrate-based 19F NMR screening: principles and applications to drug discovery. *Prog Nucl Magn Reson Spectrosc*, 51, 243-271.
69. Spoerner M., Graf T., Burkhard K. & Kalbitzer H. R. (2005) A novel mechanism for the modulation of the Ras-effector interaction by small molecules. *Biochem Biophys Res Commun*, 334, 709-713.
70. Lin M. F. & Shapiro M. J. (1996) Mixture Analysis in combinatorial chemistry. Application of diffusion-resolved NMR spectroscopy. *J Org Chem*, 61, 7617-7619.
71. Lin M., Shapiro M. J. & Wareing J. R. (1997) Diffusion-edited NMR affinity for direct observation of molecular interactions. *J Am Chem Soc*, 119, 5249-5250.
72. Lin M. G. & Shapiro M. J. (1997) Mixture analysis by NMR spectroscopy. *Anal Chem*, 69, 4731-4733.
73. Altieri A. S., Hinton D. P. & Byrd R. A. (1995) Association of biomolecular systems via pulsed field gradient NMR self-diffusion measurements. *J Am Chem Soc*, 117, 7566-7567.
74. Hajduk P. J., Olejniczak E. T. & Fesik S. W. (1997) One-dimensional relaxation and diffusion-edited NMR methods for screening compounds that bind to macromolecules. *J Am Chem Soc*, 119, 12257-12261.
75. Morris K. F. & Johnson C. S., Jr. (1992) Diffusion-ordered two-dimensional nuclear magnetic resonance spectroscopy. *J Am Chem Soc*, 114, 3139-3141.

76. Morris K. F. & Johnson C. S., Jr. (1993) Resolution of discrete and continuous molecular size distributions by means of diffusion-ordered 2D NMR spectroscopy. *J Am Chem Soc*, 115, 4291-4299.
77. Bleicher K., Lin M. F., Shapiro M. J. & Wareing J. R. (1998) Diffusion edited NMR: screening compound mixtures by affinity NMR to detect binding ligands to vancomycin. *J Org Chem*, 63, 8486-8490.
78. Zartler E. R., Yan J., Mo H., Kline A. D. & Shapiro M. J. (2003) 1D NMR Methods in Ligand-Receptor Interactions. *Curr Top Med Chem*, 3, 25-37.
79. Post C. B. (2003) Exchange-transferred NOE spectroscopy and bound ligand structure determination. *Curr Opin Struct Biol*, 13, 581-588.
80. Leone M., Freeze H. H., Chan C. S. & Pellecchia M. (2006) The nuclear overhauser effect in the lead identification process. *Curr Drug Discov Technol*, 3, 91-100.
81. Moller H., Serttas N., Paulsen H., Burcell J. M., Taylor-Papadimitriou J. & Meyer B. (2002) NMR-based determination of the binding epitope and conformational analysis of MUC-1 glycopeptides and peptides bound to the breast cancer-selective monoclonal antibody SM3. *Eur J Biochem*, 269, 1444-1455.
82. Sanchez-Pedregal V. M., Reese M., Meiler J., Blommers M. J. J., Griesinger C. & Carlomagno T. (2005) The INPHARMA method: protein-mediated interligand NOEs for pharmacophore mapping. *An*.
83. Becattini B. & Pellecchia M. (2006) SAR by ILOEs: an NMR-based approach to reverse chemical genetics. *Chemistry*, 12, 2658-2662.
84. London R. E., Perlman M. E. & Davis D. G. (1992) Relaxation-matrix analysis of the transferred nuclear Overhauser effect for finite exchange rates. *J Magn Reson*, 97, 79-98.
85. Hajduk P., Meadows R. P. & Fesik S. W. (1999) NMR-based Screening in Drug Discovery. *Q Rev Biophys*, 32, 211-240.
86. Mayer M. & Meyer B. (2001) Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J Am Chem Soc*, 123, 6108-6117.
87. Mayer M. & Meyer B. (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew Chem Int Ed Engl*, 38, 1784-1788.
88. Meinecke R. & Meyer B. (2001) Determination of the binding specificity of an integral membrane protein by saturation transfer difference NMR: RGD peptide ligands binding to integrin alphaIIbBeta3. *J Med Chem*, 44, 3059-3065.
89. Hajduk P. J., Mack J. C., Olejniczak E. T., Park C., Dandliker P. J. & Beutel B. A. (2004) SOS-NMR: a saturation transfer NMR-based method for determining the structures of protein-ligand complexes. *J Am Chem Soc*, 126, 2390-2398.
90. Dalvit C., Pevarello P., Tato M., Veronesis M., Vulpetti A. & Sundstrom M. (2000) Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *J Biomol NMR*, 18, 65-68.
91. Dalvit C., Fogliatto G., Stewart A., Veronesi M. & Stockman B. J. (2001) WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability. *J Biomol NMR*, 21, 349-359.
92. Keifer P. A. (2000) NMR spectroscopy in drug discovery: tools for combinatorial chemistry, natural products, and metabolism research. *Prog Drug Res*, 55, 137-211.
93. Hallenga K., Tonelli M., Westler W. M., Markley J. L., Takeda M., Shigezane M. & Kainosho M. (2006) *Cryogenic probe sensitivity as a function of tube geometry, sample*

*conductivity and dielectric permittivity of the solven*. Experimental Nuclear Magnetic Resonance Conference, ed., Pacific Grove, California.

94. de Swiet T. M. (2005) Optimal electric fields for different sample shapes in high resolution NMR spectroscopy. *J Magn Reson*, 174, 331-334.
95. Flynn P. F., Mattiello D. L., Hill H. D. W. & Wand A. J. (2000) Optimal use of cryogenic probe technology in NMR studies of proteins. *J Am Chem Soc*, 122, 4823-4824.
96. Voehler M. W., Collier G., Young J. K., Stone M. P. & Germann M. W. (2006) Performance of cryogenic probes as function of ionic strength and sample tube geometry. *J Magn Reson*, 183, 102-109.
97. Hajduk P. J., Gerfin T., Boehlen J. M., Haberli M., Marek D. & Fesik S. W. (1999) High-throughput nuclear magnetic resonance-based screening. *J Med Chem*, 42, 2315-2317.
98. Pervushin K., Riek R., Wider G. & Wuthrich K. (1997) Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA*, 94, 12366-12371.
99. Salzmann M., Pervushin K., Wider G., Senn H. & Wuthrich K. (2000) NMR assignments and secondary structure determination of an octameric 110 kDa protein using TROSY in triple resonance experiments. *J Am Chem Soc*, 122, 7543-7548.
100. Riek R., Wider G., Pervushin K. & Wuthrich K. (1999) Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. *Proc Natl Acad Sci USA*, 96, 4918-4923.
101. Fiaux J., Bertelsen E. B., Horwich A. L. & Wuthrich K. (2002) NMR analysis of a 900K GroEL-GroES complex. *Nature*, 418, 207-211.
102. Pellecchia M., Meininger D., Shen A. L., Jack R., Kasper C. B. & Sem D. S. (2001) SEA-TROSY (Solvent Exposed Amides with TROSY): A method to resolve the problem of spectral overlap in very large proteins. *J Am Chem Soc*, 123, 4633-4634.
103. Goto N. K. & Kay L. E. (2000) New developments in isotope labelling strategies for protein solution NMR spectrscopy. *Curr Opin Struct Biol*, 10, 585-592.
104. Sattler M. & Fesik S. W. (1996) Use of deuterium labeling in NMR: overcoming a sizeable problem. *Structure*, 4, 1245-1249.
105. Venters R. A., Huang C. C., Farmer B. T., 2nd, Trolard R., Spicer L. D. & Fierke C. A. (1995) High-level $^{2}H/^{13}C/^{15}N$ labeling of proteins for NMR studies. *J Biomol NMR*, 5, 339-344.
106. Farmer B. T., 2nd & Venters R. A. (1996) Assignment of aliphatic side-chain $^{1}HN/^{15}N$ resonances in perdeuterated proteins. *J Biomol NMR*, 7, 59-71.
107. Venters R. A., Farmer B. T., 2nd, Fierke C. A. & Spicer L. D. (1996) Characterizing the use of perdeuteration in NMR studies of large proteins: $^{13}C$, $^{15}N$ and $^{1}H$ assignments of human carbonic anhydrase II. *J Mol Biol*, 264, 1101-1116.
108. Tugarinov V. & Kay L. E. (2003) Ile, Leu, and Val methyl assignments of the 723 residue malate synthase G using a new labeling strategy and novel NMR methods. *J Am Chem Soc*, 125, 12868-12878.
109. Waugh D. S. (1996) Genetic tools for selective labeling of proteins with $\alpha$-$^{15}N$-amino acids. *J Biomol NMR*, 8, 184-192.
110. Weigelt J., van Dongen M., Uppenberg J., Schultz J & Wikstrom M. (2002) Site-selective screening by NMR spectroscopy with labeled amino acid pairs. *J Am Chem Soc*, 124, 2446-2447.

111. Yamazaki T., Otomo T., Oda N., Kyogoku Y., Uegaki K., Ito N., Ishino Y. & Nakamura H. (1998) Segmental isotope labeling for protein NMR using peptide splicing. *J Am Chem Soc*, 120, 5591-5592.
112. Otomo T., Teruya K., Uegaki K., Yamazaki T. & Kyogoku Y. (1999) Improved segmental isotope labeling of proteins and application to a larger protein. *J Biomol NMR*, 14, 105-114.
113. Xu R., Ayers B., Cowburn D. & Muir T. W. (1999) Chemical ligation of folded recombinant proteins: segmental isotope labeling of domains for NMR studies. *Proc Natl Acad Sci USA*, 96, 388-393.
114. Gao F., Mer G., Tonelli M., Hansen S. B., Burghardt T. P., Taylor P. & Sine S. M. (2006) Solution NMR of acetylcholine binding protein reveals agonish-mediated conformational change of the C-loop. *Mol Pharmacol*, 70, 1230-1235.
115. Pellecchia M., Meininger D., Dong Q., Chang E., Jack R. & Sam D. S. (2002) NMR-based structural characterization of large protein-ligand interactions. *J Biomol NMR*, 22, 165-173.
116. Sem D. S., Yu L., Coutts S. M. & Jack R. (2001) Object-oriented approach to drug design enabled by NMR SOLVE: first real-time structural tool for characterizing protein-ligand interactions. *J Cell Biochem*, 84, 99-105.
117. Constantine K. L., Davis M. E., Metzler W. J., Mueller L. & Claus B. (2006) Protein-ligand NOE matching: a high throughput method for binding pose evaluation that does not require protein NMR resonance assignments. *J Am Chem Soc*, 128, 7252-7263.
118. Breeze A. L. (2000) Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Prog Nucl Magn Reson Spectrosc*, 36, 323-372.
119. Dalvit C., Flocco M., Knapp S., Mostardini M., Perego R., Stockman B. J., Veronesi M. & Varasi M. (2002) High-throughput NMR-based screening with competition binding experiments. *J Am Chem Soc*, 124, 7702-7709.
120. Wang Y. S., Liu D. & Wyss D. F. (2004) Competition STD NMR for the detection of high-affinity ligands and NMR-based screening. *Magn Reson Chem*, 42, 485-489.
121. Dalvit C., Fasolini M., Flocco M., Knapp S., Pevarello P. & Veronesi M. (2002) NMR-Based Screening with Competition Water-Ligand Observed via Gradient Spectroscopy Experiments: Detection of High-Affinity Ligands. *Journal of Medicinal Chemistry*, 45, 2610-2614.

This Page Left Intentionally Blank

# APPENDIX D.

# STABILITY OF PRL-1-C98A AND PRL-1-C170S-C171S

**Table D.1 Physical Stability of PRL-1**[a]

| | Phosphate | | | | HEPES | | | |
|---|---|---|---|---|---|---|---|---|
| | CD | | SLS | | CD | | SLS | |
| | $T_m$ | ± SD | Onset T | ± SD | $T_m$ | ± SD | Onset T | ± SD |
| nrWT | 65.5 | 1.1 | 59.9 | 0.3 | 60.5 | 2.1 | 53.3 | 0.4 |
| redWT[bc] | 67.8 | 0.4 | 63.8 | 0.4 | 62.0 | 1.3 | 57.4 | 0.3 |
| redWT[d] | 79.0[e] | 0.7 | nd[f] | | nd | | nd | |
| C49S | 62.8 | 2.0 | 59.7 | 0.2 | 56.3 | 0.3 | 50.0 | 0.3 |
| C104S | 76.3 | 1.1 | 72.1 | 0.3 | 64.9 | 2.1 | 59.3 | 0.2 |
| Y53E | 62.3 | 0.4 | nd | | nd | | nd | |
| Y53F | 62.6 | 0.3 | nd | | nd | | nd | |
| oxC98A | 64.5 | 0.6 | 59.4 | 0.3 | 64.7 | 1.6 | 59.9 | 0.3 |
| redC98A | 65.1 | 0.6 | 63.5 | 0.3 | 64.2 | 1.8 | 60.9 | 0.3 |
| oxC170S-171S | 70.2 | 0.5 | 68.6 | 0.3 | 60.1 | 0.1 | 56.6 | 0.4 |
| redC170S-171S | 70.4 | 0.5 | 72.6 | 0.3 | 67.4 | 2.4 | 60.0 | 0.3 |

[a]All temperatures reported are in °C

[b]For CD, reducing conditions were achieved by the addition of 1 mM DTT. For SLS, reducing conditions were achieved by the addition of 10 mM DTT.

[c]CD and/or SLS were also used to study the physical stability of PRL-1 mutants under both reducing and nonreducing conditions, although no differences were detected. This data for this is not shown.

[d]Reduced PRL-1-WT (redWT) was studied in 20 mM phosphate buffer at pH 6.5 with equivalent ionic strength (I=100 mM). All other values reported, including those for HEPES buffer, are at pH 7.5.

[e]Melting was incomplete at the maximum temperature attainable and the data was extrapolated to determine the $T_m$.

[f]Not determined

Table D.1 includes the SLS and CD onset and melting temperatures for all PRL-1 variants examined in the stability study (Chapter 5).

In all conditions examined, the reduced and oxidized C98A mutants have virtually no differences in their physical stabilities as measured by CD or SLS. This is not unexpected because the differences between the reduce and oxidized wild types are, although statistically

different from each other, extremely small (2.1 °C) compared to other differences observed (*i.e.* 8.5 °C difference for C104S and reduced wild type). Inconsistent with all other mutants examined, however, phosphate does not stabilize the conformation of C98A to any reasonable extent. The reason for this observation is not clear but may have to do with local changes in the protein structure or dynamics when comparing the oxidized wild type and reduce wild type.

As expected, there is no difference in $T_m$ values between the reduced and untreated C170S-C171S mutant assayed under phosphate buffered conditions. Interestingly though, the untreated mutant was slightly less stable than the reduced mutant as judged by SLS in either buffer or by CD in HEPES buffer. The reason for this is unclear, but it may be due to the transient ability of the active site to form the disulfide bond. This is supported by the fact that C170S-C171S is less stable than the C104S mutant but more stable than the reduced wild type, possibly because C170S-C171S is less capable of forming the disulfide bond than the reduced wild type. Furthermore, the C104S mutant, which is incapable of forming the disulfide bond, is the most stable species. The fact that there was no difference observed by CD in phosphate may be because the buffer can bind the active site and slows down these conformational changes. Although highly speculative, this data collectively suggests that dynamic exchange of the disulfide bond is destabilizing. Construction of the C104S-C170S-C171S triple mutant may be able to help answer this question.

# APPENDIX E.

# PURIFICATION OF Zn-BOUND PRL-1

PRL-1 is normally expressed as a $His^6$-tagged fusion protein to permit purification by immobilized metal ion affinity chromatography (IMAC). The tag is cleaved by Factor Xa and then separated from the protein by passing the sample back over the IMAC column. Although the purification scheme for the zinc version of PRL-1 is the same in concept, a few minor adjustments were made to accommodate changes in protein solubility. First, the cells are lysed using a French press in Buffer A (50 mM Tris-HCl, pH 7.5 with 100 mM NaCl and 10 mM imidazole) and subsequently centrifuged at 21,000 x g to pellet out insoluble debris (Figure E.1, lane 1). Using a batch method for purification, the soluble cell lysates were then incubated with approximately 3 mL of iminodiacetic acid (IDA) chelating IMAC resin charged with 100 mM $ZnSO_4$ for at least 10 minutes but no more than 30

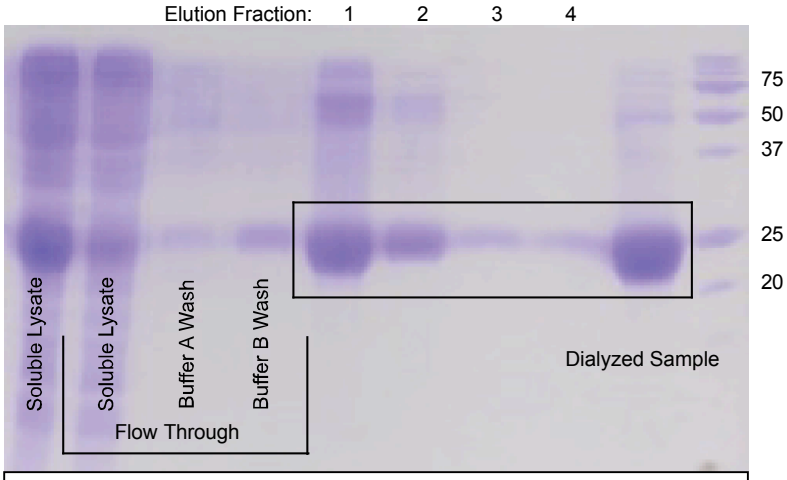



**Figure E.1. Zn-PRL-1 Purification Scheme.** Lane 1: soluble cell lysate. Lane 2: flow through from cell lysate. Lane 3: flow through from Buffer A wash. Lane 4: flow through from Buffer B wash. Lane 5-8: elution fractions 1-4. Lane 9: dialyzed sample that was subsequently treated with Factor Xa to remove the purification tag. Lane 10: Biorad protein standards. $His^6$-tagged PRL-1 runs at approximately 25 kDa. Untagged PRL-1 runs just below 20.

(lane 2). The resin with protein bound was washed with 25 mL of Buffer A five times (lane 3). The resin was then rinsed with 25 mL of Buffer B (50 mM Tris-HCl, pH 7.5 with 200 mM NaCl and 40 mM imidazole) three times to minimize the number of nonspecific protein contaminants (lane 4). Elution was accomplished in 10 mL using Buffer C, repeating four times times (50 mM Tris-HCl, pH 7.5 with 200 mM NaCl, 500 mM imdazole, 100 μM $ZnCl_2$) (lanes 5-8). The fractions were pooled together and a slow dialysis procedure was used to remove the imidazole from solution (lane 9). The 30-40 mL protein sample was dialyzed in 4 L of buffer containing 250 mM imidazole for 1 hr, followed by subsequent dialysis steps with decreasing amounts of imidazole (250, 150, 75, 10 then 0 mM). Dialysis was performed in 50 mM Tris-Cl, pH 7.5 with 100 mM NaCl. The affinity tag was cleaved with factor Xa overnight and subsequently purified by size exclusion on a G75 column in 50 mM sodium phosphate, pH 6.5 with 100 mM NaCl, the next morning to ensure separation of the final cleaved product from any uncleaved sample remaining. Figure E.1 shows the SDS-PAGE gel corresponding to the first-day purification procedure to illustrate the efficiency of the method.

The final samples were concentrated and exchanged $10^5$ fold into 50 mM HEPES, pH 7.5 for *p*NPP activity assays as described in the previous sections. The relative activities of the Zn-bound PRL-1 and Ni-bound PRL-1 under both reducing and nonreducing conditions are essentially identical (data not shown), indicating that the majority of the protein is oxidized even though zinc lacks any specific redox active properties. $^{1}H$-$^{15}N$ HSQC NMR comparisons of the Zn-bound (0.5 mM) and Ni-bound (0.5 mM) PRL-1 forms are shown in Figure E.2. The diagnostic C49 NMR peak corresponding to reduced Zn-PRL-1 is not visible in this spectrum and a peak is observed close to the position of C49 in the oxidized state, further supporting our hypothesis that Zn-bound PRL-1 is primarily oxidized. Although some significant differences are

present between the two spectra, the two protein forms share the same general topology. Reduction of Zn-bound PRL-1 leads to large conformational differences that parallel those observed for the Ni-bound species both in the magnitude and number of changes, but also in the residues involved. Detailed chemical shift analysis cannot be conducted at this time because sequential backbone assignments are needed to clarify some of the larger differences between the two species. Collectively this information indicates that zinc binding in PRL-1 does not grossly affect the activity or structure when compared the nickel counter part.

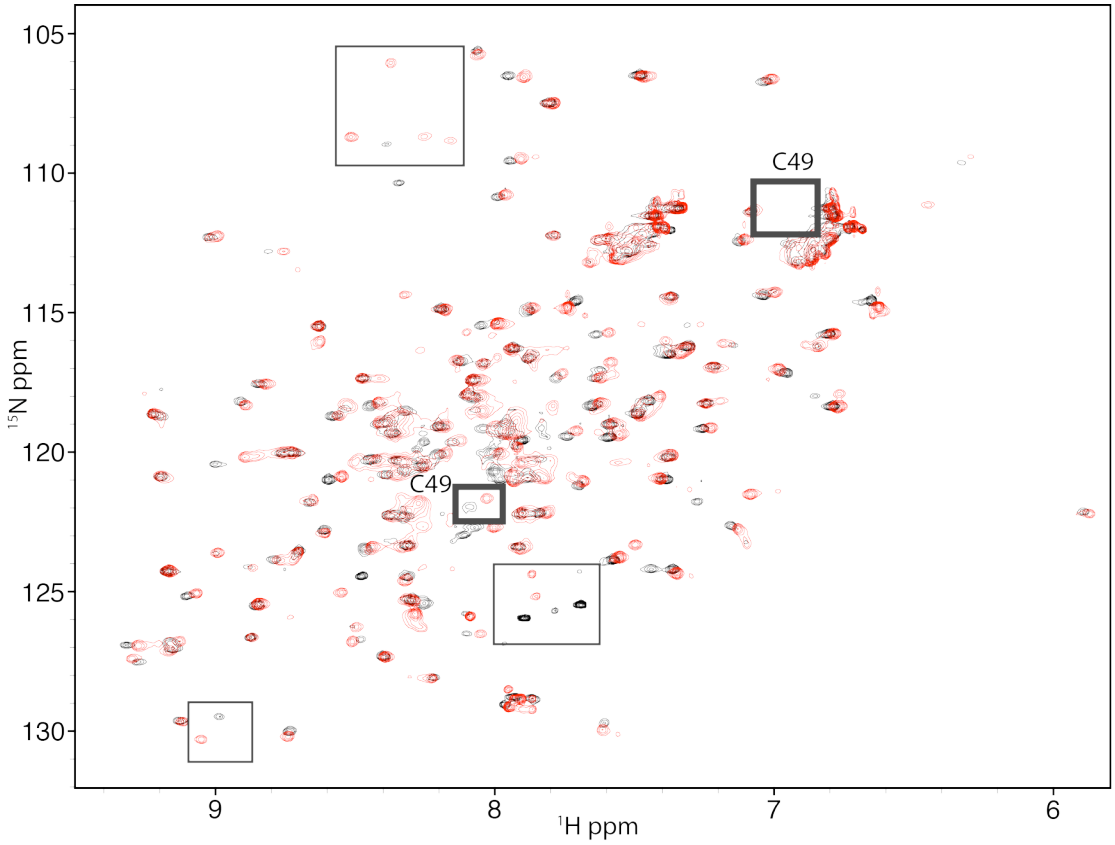



**Figure E.2. Overlay of $^{1}$H-$^{15}$N HSQC Spectra of Zn- and Ni-PRL-1-WT.** Both samples were collected under identical solution conditions at the same concentration. The C49 residues are highlighted by a thick black box and the place where the corresponding C49 peak in the reduced state should be located is demarked as well. Some of the larger changes between the two species are emphasized with the thinner black boxes. Some of these peaks correspond to previously unassignable C-terminal peaks, which are expected to deviate between these two forms because metal binding is accomplished by residues in this region.

Page Left Intentionally Blank