In-cell and in vitro studies of disease related protein-protein interactions using NMR-spectroscopy

Andres Yudiel Maldonado
University at Albany, State University of New York, amaldonado@albany.edu

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Biochemistry Commons, and the Chemistry Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive. Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
In-cell and *In vitro* Studies of Disease Related Protein-Protein Interactions Using NMR-Spectroscopy

by

Andres Y. Maldonado

A Dissertation

Submitted to the University at Albany, State University of New York

in Partial Fulfillment of

the Requirements for the Degree of

Doctor of Philosophy

College of Arts & Sciences

Department of Chemistry

2011
In-cell and *In vitro* Studies of Disease Related Protein-Protein Interactions Using NMR-Spectroscopy

by

Andres Y. Maldonado

COPYRIGHT 2011
Acknowledgements

The projects presented in this dissertation are the culmination of work done dating back to my junior year of college in the Shekhtman Research Lab. Although I am the sole author of this dissertation, this work stands on the shoulders of those who have mentored me along my scientific path.

I would like to thank Professor Alexander Shekhtman for his guidance, patience and support throughout this process. His passion for science has inspired me to always press forward in the face of adversity. I would not have become a scientist if it were not for him. Infinite amounts of this work would have not been possible without Dr. David Burz who has become a mentor and role model from my scientific career. I would also like to thank Dr. Sergey Reverdatto for sharing his knowledge of Biochemistry and Molecular Biology.

I would also like to thank my fellow group members for their cooperation and assistance and to all my friends who have helped nurture my passion for science. Special Thanks to Jacqueline Washington for her support throughout this process which has been invaluable.

Finally, I would like to thank my parents, Dr. Carlos Y. Maldonado, Idalia Burkitt and both my stepparents Donna Rittucci and John Burkitt whose constant support throughout my educational career has culminated to this point. We made it!
Abstract

The receptor for advanced glycation end products (RAGE) is a multiligand cell surface macromolecule that plays a central role in the etiology of diabetes, inflammation, and neurodegeneration. The cytoplasmic domain of RAGE, ctRAGE, is critical for RAGE-dependent signal transduction. As the most membrane proximal event, mDia1 binds to ctRAGE and is essential for RAGE ligand-stimulated phosphorylation of AKT and cell proliferation/migration. We show that ctRAGE contains an unusual α-turn that mediates the mDia1-ctRAGE interaction and is required for RAGE dependent signaling. The results establish a novel mechanism through which an extracellular signal initiated by RAGE ligands regulates RAGE signaling in a manner requiring mDia1.

Interactions between biological macromolecules give rise to and regulate biological activity. This activity is manifest through structural dynamics and changes in the macromolecular structures that comprise these interactions. Until recently, mostly in vitro techniques have been used to study macromolecular interactions that govern biological processes under conditions remote from those existing in the cell. With the advent of in-cell Nuclear Magnetic Resonance (NMR) spectroscopy, these processes can now be studied within a cellular environment. Here we review the latest techniques in order to study proteins under more physiological relevant conditions.

Mycobacterium tuberculosis ability to resist nitric oxide stress, from the immune initiated response, relies on then bacteria’s ability to target misfolded or damaged proteins. Mycobacterium tuberculosis uses (Prokaryotic Ubiquitin like protein) Pup to target proteins for proteasome mediated degradation. Proteins which have been tagged by Pup interact with mycobacterial ATPase (Mpa). Using STINT NMR we characterized the interactions between Pup and Mpa in-cell and show that Pup’s C- terminus is engaged by Mpa. Pups N terminus was also
shown to be engaged only in the presence of the Mpa-proteasome complex. We also demonstrate, through various biochemical assays that Pup which is not deamidated is not a substrate for proteasomal degradation.
List of Abbreviations

3FY 3-fluro-tyrosine
α-SYN α-synuclein
AGEs Advanced Glycated End products
AMC 7-amino-4-methylcoumarin
ATK Agammaglobulinaemia Tyrosine Kinase
BSA Bovine Serum Albumin
BTZ Bortzeomib
CAM-Calmodulin
CEL CarboxyEthylLysine
CI2 Chymotrypsin Inhibitor 2
CK2 Casein Kinase 2
CML CarboxyMethylLysine,
cT-RAGE- Cytoplasmic Tail of RAGE
CP Core Particle
CPP Tat Tat Protein of HIV-1
DAD Diaphanous Activating Domain
DID Diaphanous Inhibitory Domain
DOP Deamindase of Prokaryotic Ubiquitin like Protein
ERK Extracellular signal-Regulated Kinase
FH1 Formin Domain
FKBP FK506 Binding Protein
FRAP Fluorescence recovery after photobleaching
FRB FKBP-Rapamycin-Binding domain of mTOR
FRET Fluorescence Resonance Energy Transfer
GB1 Streptococcal protein G B1 domain
GFP Green Fluorescence Protein
HDH Hisidinol Dehydogenase
HMGB1 High Mobility Group Box-1
HRS Hepatocyte growth factor-regulated tyrosine kinase substrate
HSQC Heteronuclear Single Quantum Coherence
HTS High Throughput Screening
IPTG IsoPropyl 1-Thio-β-D-Galactopyranoside
IS Interaction Surface
iNos Inducible Nitric Oxide Synthase
LB Luria-Bertani medium
MBP Maltose binding domain
mDia1 Mammalian Diaphanous
MHC Major HistoCompatibility
MPA-S Mycobacterium Smegmatis Atpase
MPA-T Mycobacterium Tuberculosis Atpase
mTOR mammalian Target Of Rapamycin
NA's Nucleic Acids
NOE Nuclear Overhauser Enhancement
NOESY Nuclear Overhauser Enhancement Spectroscopy
PAFA Proteasome Accessory Factor A
PBS Phosphate-Buffered Saline
PDB Protein Data Bank
PMCC Pearson product-Moment Correlation Coefficient
PPIase Peptidyl-Prolyl cis–trans Isomerase
PRR Pattern Recognition Receptor
PTM Post Translational Modifications
PUP Prokaryotic Ubiquitin like Protein
PVP polyvinyl pyrrolidone
RAGE Receptor for Advanced Glycation End products
RDC Residual Dipolar Couplings
RP Regulatory Paricle
RPT Atpase Type SubUnit
RPN Non ATPase Type SubUnit
RNI Reactive Nitrogen Intermediates
SDS-PAGE sodium dodecyl sulfate polyacrylamide gel electrophoresis
sRAGE soluble Receptor for Advanced Glycation End products
SMC Smooth Muscle Cells
SMILI-NMR Screening Small molecule Interactor Library using In-cell NMR
STAM2 Signal- Transducing Adapter Molecule 2
STINT-NMR in-cell NMR-based method for mapping the STructural INTeractions
tmfF trifluoromethyl-L-phenylalanine
TGFβ Transforming Growth Factor-β
TOCSY Total Correlation Spectroscopy
TROSY-HSQC Transverse optimized spectroscopy
UBQ Ubiquitin
UCH Ubiquitin Carboxyl-Terminal Hydrolyses
UIM Ubiquitin Interacting Motif
USP Ubiquitin Specific Proteases
List of Figures

Fig. 1.0 ctRAGE contains a folded segment.
Fig. 1.1 Solution structure of the ctRAGE fragment (amino acids 2-15).
Fig. 1.2 ctRAGE interacts with mDia1 FH1.
Fig. 1.3 S100b stimulates AKT activation in vascular smooth muscles cells via ctRAGE-mDia1 FH1 interaction through R5 and Q6.
Fig. 1.4 S100B stimulates migration (A) and proliferation (B) in vascular SMCs mainly via ctRAGE-mDia1 FH1 interaction through R5 and Q6.
Fig. 1.5 Model of RAGE induced activation of mDia1.
Figure 2.0 GB1 in Xenopus oocytes.
Figure 2.1 Ubiquitin in HeLa cells
Figure 2.2 Ubiquitin-ligand complexes in E. coli.
Figure 2.3 Ternary FKBP-FRB-ligand complexes in E. coli
Figure 2.4 Interaction surface maps of Ubiquitin-ligand complexes.
Figure 2.5 In-cell [1H-13C]-CT-TROSY-HSQC spectra of the aromatic regions of 13C/15N-labeled d(GCGAAGC) and r(GGCACUCCGGUGCC) hairpins.
Figure 3.0 The structural arrangement of the subunits α- and β.
Figure 3.1 Overview of the euukaryotic Ubiquitin-proteasome system.
Figure 3.2 Overview of proposed model for Pup-Proteasome pathway in Mycobacterium tuberculosis.
Figure 3.3 Pup does not contain any regions of secondary structure due to macromolecular crowding.
Figure 3.4 In-cell Mpa-S binds to the C-terminal region of Pup.
Figure 3.5 Pup is not degraded in the presence of the Mpa-S - wild type proteasome complex.
Figure 3.6 Order of assembly of the Mpa-Proteasome complex does not affect the binding of Pup.
Figure 3.7 Pup is not degraded in the presence of the Mpa-S- Open gate proteasome complex
Figure 3.8 Pup-GGQ is not degraded in the presence of the Mpa-Proteasome complex in vitro
Figure 3.9 Pup-GGQ is not degraded in the presence of the Mpa-S-Proteasome complex in vitro upon varying the concentrations of complex.
Figure 3.10 The Open gate and wild type proteasome show activity when mixed with 7-amino-4-methylcoumarin (AMC).
Figure 3.11 Proteasome activity was inhibited by addition of Bortzeomib.
Figure 3.11B STINT-NMR reveals an equilibrium of events.
Figure 3.12 Pup gets degraded by the Mpa-proteasome only upon deamidation.
Supplementary Figure 1. ctRAGE does not interact with a model membrane formed by DMP/CHAPSO bicelles.
Supplementary Figure 2. Tertiary structure of ctRAGE is stabilized by electrostatic interactions between R5, E10, E11, and R12.
Supplementary Figure 3. Stereoview of ctRAGE structure.
Supplementary Figure 4. R5AQ6A mutation leads to significant structural changes in the folded region of ctRAGE.
Supplementary Figure 5. Double mutant R5AQ6A ctRAGE does not form a complex with mDia1 FH1.
Supplementary Figure 6. Transfection efficiency of SMCs.
Supplementary Figure 7. Proliferation of SMCs caused by RAGE signaling is attenuated by double mutant RAGE.
Supplementary Figure 8. Pup-GGQ is not degraded in the presence of the Mpa-T-Proteasome complex in vitro.
Supplementary Figure 9. Chemical shift changes in the $^{15}$N-HSQC spectra of in-cell relative to in-vitro Pup.
Supplementary Figure 10. Intensity difference between the $^{15}$N-HSQC spectra of in-cell and $^{15}$N-HSQC of in vitro Pup.
Supplementary Figure 11. Mpa-S binds to two distinct regions on Pup upon titration in-cell.
Supplementary Figure 12. Intensity difference in the $^{15}$N-HSQC spectra of in-cell free Pup with $^{15}$N-HSQC spectra of Pup titrated with Mpa-S in-cell.
Supplementary Figure 13. Mpa-S binds to two distinct regions on Pup upon titration in vivo.
Supplementary Figure 14. Intensity difference in the $^{15}$N-HSQC spectra of free Pup with $^{15}$N-HSQC spectra of Pup over expressed second post Mpa-S over expression in-cell.
Supplementary Figure 15. Mpa-S-Open gate complex interacts with the C-terminus of Pup.
Supplementary Figure 16. Intensity difference in the $^{15}$N-HSQC spectra of free Pup with $^{15}$N-HSQC spectra of Pup expressed first followed by over expression of the Mpa-S-Open gate complex.
Supplementary Figure 17. Mpa-S-Open gate complex interacts with the C-terminus of Pup.
Supplementary Figure 18. Expression of Mpa-S- Open gate complex first followed by Pup induction second is indicative of complete interaction of Pup.
Supplementary Figure 19. Pup’s N-terminus interacts with the Mpa-S-wild type proteasome complex.
Supplementary Figure 20. Expression of Mpa-wild type proteasome complex second indicates only partial interaction with Pup.
Supplementary Figure 21. Pup shows increased surface interaction in the presence of the Mpa-S-wild type proteasome complex when complex is expressed second in cell.
Supplementary Figure 22. Expression of Mpa-S-wild type proteasome complex second is indicative of Pup’s full interaction with the Mpa-S-wild type complex.
Table of Contents

Acknowledgements ............................................................................................................. iii
Abstract .............................................................................................................................. iv
List of Abbreviations .......................................................................................................... vi
List of Figures ...................................................................................................................... viii
Preface ............................................................................................................................... 1

Chapter 1 ............................................................................................................................ 3

Signal Transduction in RAGE: Solution Structure of C-terminal RAGE (ctRAGE) and its Binding to mDia1 ................................................................. 3

Abstract ............................................................................................................................. 4
Introduction ......................................................................................................................... 4

Experimental Procedures .................................................................................................. 6
  Reagents and Chemicals. ................................................................................................. 6
  Plasmid construction. ....................................................................................................... 6
  Labeling, Expression, and Purification of Wild Type and Mutant ctRAGE. ................. 7
  Purification of mDia1 FH1 and its fragments. ................................................................. 8
  Site-Directed Mutagenesis of ctRAGE. ........................................................................ 9
  NMR Experiments ......................................................................................................... 10
  Structure Calculation ..................................................................................................... 11
  Cell Lines and Materials ............................................................................................... 13
  Site-directed mutagenesis of full length RAGE ........................................................... 13
  Transfection ................................................................................................................... 13
  Western blot analysis ..................................................................................................... 13
  Smooth Muscle Cell Migration and Proliferation Assays ............................................. 14
  Data analysis ................................................................................................................... 14

Results ................................................................................................................................ 15
  C-tail of RAGE possesses a region with ordered tertiary structure. ............................ 15
  ctRAGE folds into an alpha-turn. .................................................................................. 18
  In vitro identification of residues critical for the mDia1 FH1-ctRAGE- interaction. ........ 20
  R5AQ6A-RAGE evokes a dominant negative response in RAGE-mDia1 signaling. ......... 23

Discussion ......................................................................................................................... 27

Chapter 2 ............................................................................................................................ 31

In-Cell NMR Spectroscopy ............................................................................................... 31

Introduction ......................................................................................................................... 32
Methodologies ..................................................................................................................... 33
  Target labeling ................................................................................................................. 33
Chapter 3 ............................................................................................................. 75

Fate of Pup during Mtb proteasome proteolysis as studied by in-cell NMR spectroscopy .......................................................... 75

Abstract .................................................................................................................. 76

Background ............................................................................................................ 76

Proteasome: Background and Structure ................................................................ 77

Proteasome Mediated Substrate Degradation .......................................................... 79

Introduction ............................................................................................................ 86

Experimental Procedures ...................................................................................... 88

Plasmid Construction ............................................................................................ 88

Sequential over-expression and labeling .................................................................. 91
SDS PAGE Analysis .........................................................................................................................97
NMR spectroscopy. ..........................................................................................................................97
Expression and purification of proteins for in vitro assays............................................................99
Proteasome Activity Assay .............................................................................................................101
Results ...........................................................................................................................................102
The in-cell spectrum of Pup is disordered .........................................................................................102
Mpa binds to multiple sites on Pup. ..................................................................................................105
Pup is not degraded when bound to Mpa-proteasome complex in vivo ........................................108
Recreation of the Pup proteasome pathway in-vitro reveals that Pup is retained and not degraded. ........................................................................................................................................114
Discussion .....................................................................................................................................119
Appendix.........................................................................................................................................125
Supplementary Information: .........................................................................................................125
Signal Transduction in RAGE: Solution Structure of C-terminal RAGE (ctRAGE) and its Binding to mDia1...........................................................................................................................................125
Fate of Pup during Mtb proteasome proteolysis as studied by in-cell NMR spectroscopy. ..........132
Preface

Interactions between biological macromolecules give rise to and regulate biological activity. This activity is manifested through molecular dynamics and changes in the structures that comprise the interaction. Nuclear Magnetic Resonance (NMR) spectroscopy is a useful tool to monitor and classify macromolecular interactions both in vitro and in vivo. The dissertation presented here is a combination of works in which disease related protein-protein interactions were studied in vitro and in vivo by using NMR spectroscopy.

Chapter one describes a direct application of in vitro NMR spectroscopy for protein structure elucidation. Using a suite of standard triple resonance experiments we assigned the protein amide backbone of the cytoplasmic domain of the Receptor for Advanced Glycation End products (RAGE), ctRAGE, and along with 295 Nuclear Overhauser connectivities (nOe)’s solved the solution structure of ctRAGE. We show that ctRAGE contains an unusual α-turn that mediates the mDia1-ctRAGE interaction and is required for RAGE dependent signaling. We also demonstrate that the FH1 and FH2 domains of mDia1 are required for mDia1 activity. Coupling our in vitro NMR structural data with physiological assays that monitor the rate of AKT phosphorylation along with cellular migration and proliferation, we propose a new model of RAGE induced activation of mDia1.

Chapter two is based on our published review titled “In-cell NMR Spectroscopy” and covers all of the latest methodologies and applications of in-cell NMR. The development of structural biology was historically based on the principle of divide and conquer — individual proteins were purified to homogeneity and their atomic structures were solved in vitro by using either X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy. This approach was tremendously successful, and led to the creation of a protein-structure databank that currently contains more than 50,000 structures. But relating in vitro protein structures to biological...
processes that occur inside the cell is not a trivial task. A traditional approach to solving this problem entails mutating a protein's structure at certain sites based on its in vitro structure and observing the effects of these changes on the cell. This low-resolution validation of high-resolution structures may still lead to situations where the in vitro structure does not fully represent the physiologically active protein structure under the conditions present in a cell. The work presented by various authors in the review reveals new ways to examine biological macromolecules as they exist inside living cells, ushering in a new era of structural biology.

Chapter three describes the direct application of an in-cell NMR-based method for mapping the structural interactions (STINT-NMR) that underlie protein-protein complex formation. The method entails sequentially expressing two (or more) proteins within a single bacterial cell in a time-controlled manner and monitoring their interactions using in-cell NMR spectroscopy. We characterized the interactions between Prokaryotic Ubiquitin-like protein, Pup, and proteasomal ATPase, MPA, and proteasome in-cell. We show that the C-terminus of Pup is engaged by Mpa and the N-terminus is engaged only in the presence of the Mpa-proteasome complex. Through the use of proteasome degradation assays we demonstrate that Pup, which is not deamidated, is not a substrate for the proteasome.
Chapter 1

Signal Transduction in RAGE: Solution Structure of C-terminal RAGE (ctRAGE) and its Binding to mDia1.
Abstract

The receptor for advanced glycation end products (RAGE) is a multiligand cell surface macromolecule that plays a central role in the etiology of diabetes, inflammation, and neurodegeneration. The cytoplasmic domain of RAGE, ctRAGE, is critical for RAGE-dependent signal transduction. As the most membrane proximal event, mDia1 binds to ctRAGE and is essential for RAGE ligand-stimulated phosphorylation of AKT and cell proliferation/migration. We show that ctRAGE contains an unusual α-turn that mediates the mDia1-ctRAGE interaction and is required for RAGE dependent signaling. The results establish a novel mechanism through which an extracellular signal initiated by RAGE ligands regulates RAGE signaling in a manner requiring mDia1.

Introduction

The Receptor for Advanced Glycation End products (RAGE) is a pattern recognition receptor that consists of three extracellular immunoglobulin domains and a short 42 amino acid cytoplasmic domain[1]. RAGE binds diverse ligand families, including advanced glycation end products[1-3], S100/calgranulins[4], High Mobility Group Box-1 (HMGB1)[5], amyloid-β peptide (Aβ), β-sheet fibrils[6], and phosphatidylserine[7]. Studies in vitro and in vivo indicate that RAGE is a signal transduction receptor for these ligand families[4, 8]. The deletion of the cytoplasmic domain of RAGE exerts a “dominant negative” (DN) effect in which the signal transduction response to RAGE ligands is blunted[9-12]. A number of signal transduction cascades are activated upon ligand-RAGE interactions, including mitogen-activated protein kinases, phosphatidylinositol 3-kinase, Jak/STAT (signal transducers and activators of transcription)[4], and the Rho GTPases Rac-1 and Cdc42[13]. RAGE-ligand interactions evoke central changes in cellular properties including stimulation of cellular migration and proliferation.
and leading to such pathological conditions as diabetes, Alzheimer’s disease, inflammation and cancers.

RAGE also plays a pivotal role in the atherosclerotic process[14]. According to the “Response to Injury hypothesis”[15] both migration and proliferation of cells from media to intima are central to the atherosclerotic pathogenesis. RAGE stimulates both processes as a result of the binding of the formin homology (FH1) domain of mammalian Diaphanous-1 (mDia1) to the short cytoplasmic tail of RAGE (ctRAGE)[13]. mDia1 acts as a potent actin and microtubule polymerization factor that regulates a number of processes including cell migration and division (reviewed by Higgs[16] and Wallar et al.[17]). Two formin homology domains of mDia1 are required for mDia1 actin polymerization activity[16]. The precise mechanism by which mDia1 stimulates actin polymerization is not completely understood. However, mDia1 appears to promote polymerization from the barbed end of actin filaments in cooperation with the actin binding protein, profilin [18, 19].

The molecular details of RAGE-mDia1 interactions are required to explore specific pathways leading to RAGE dependent pathologies. Here we present the solution structure of ctRAGE and identify the mDia1 (FH1)-ctRAGE interaction surface. In vitro binding studies reveal that two amino acids, R5 and Q6 of ctRAGE, which are part of the mDia1-ctRAGE interaction surface, are essential for interactions with mDia1. When R5 and Q6 are mutated to alanines, the loss of the binding epitope for mDia1 results in no RAGE ligand-induced downstream phosphorylation of AKT or migration and proliferation of vascular smooth muscle cells(SMCs). These studies suggest a novel signaling paradigm in which extracellular cues stimulated by RAGE ligand binding are transduced to the cytoplasmic domain of the receptor via mDia1 to stimulate a fundamental signaling network.
Experimental Procedures

Reagents and Chemicals.

Restriction enzymes and Phusion polymerase were from NEB. All other chemicals used were reagent grade or better.

Plasmid construction.

Human Dia1 cDNA library clone BC117257 was obtained from Open Biosystems and used as a template for PCR amplifications. DNA coding for the FH1 domain (amino acids 571-835) was PCR-amplified using Phusion polymerase and oligonucleotides 5’-TTTCATATGGCTCAAAACATCACAGCCCGGATTGG and 3’-TTTGTCGACTCATTCTGGCTTCCCAGGAATCTG, which contain 5’-NdeI and 3’-SalI restriction sites. The restriction-digested PCR products were ligated into expression vector H-MBP[20], which confers ampicillin resistance. The resulting plasmid, H-MBP-FH1, expresses a N-terminal His- and MBP-tagged FH1 domain of mDia1. Additionally, the two shorter fragments of FH1 designated as FH1-Pep1 (amino acids 646-660) and FH1-Pep2 (amino acids 683-771) were cloned into expression vector pTM-7[21] between the BamHI and HindIII sites creating pTM-FH1-pep1 and pTM-FH1-pep2 plasmids. The DNA fragment for the shorter peptide Pep1 was assembled from the two complementary oligonucleotides FH1_pept1_F (AGCTTATGTCTGGGGATGCTACCATCCCTCCACCCCTCCTTGGCTGAGGTTGAG) and FH1_pept1_R (GATCCTCAACCTCAGGCAAAGGAGGGGTGAGGAGGATGGTAGCATCCCCAGACAT A), derived from the cDNA sequence of human mDia1. The longer peptide, Pep2, was PCR amplified by using oligonucleotides FH1_pept2_F (TTTAAAGCCTATGGGGAGTGCTAGAATCCCCCCCACCA) and FH1_pept2_R
DNA fragment coding for ctRAGE (amino acids 362-404) was amplified from human cDNA clone ID IOH12890 (Invitrogen) with primers R-CT_Hind III (TTAAAAGCTTATGTGTGGCAAAGGGCGGCAAC) and R-CT_BamH I (ATTGGATCTCTCAAGGGCCCTCCAGTACTACTCTCGC), and cloned into the BamHI and HindIII sites of pTM-7 vector[21], which confers kanamycin resistance. This vector is used to direct small peptides to bacterial inclusion bodies by fusing them with a very hydrophobic protein, trpL[21]. The resulting plasmid, pTM-ctRAGE, expresses an N terminal His-tagged fusion of trpL-ctRAGE domain.

Labeling, Expression, and Purification of Wild Type and Mutant ctRAGE.

To uniformly label ctRAGE, pTM-ctRAGE, pTM-R5A Q6A ctRAGE, pTM-R24A ctRAGE or pTM-E26A ctRAGE were transformed into E. coli strain BL21(DE3) Codon+ (Novagen). For $[^{15}N]$ labeling, cells were grown at 37 °C in minimal medium (M9) containing 35 mg/liter kanamycin and 1 g/liter $[^{15}N]$ammonium chloride as the sole nitrogen source. For $[^{13}C,^{15}N]$ labeling, cells were grown at 37 °C in M9 medium containing 35 mg/liter kanamycin, 1 g/liter $[^{15}N]$ammonium chloride, and 2 g/liter $[^{13}C]$ glucose instead of unlabeled glucose as the sole carbon source. Cells were grown to 0.7 $A_{600}$ at 37 °C, induced with 0.5 mM isopropyl 1-thio-β-d-galactopyranoside (IPTG), and grown overnight. Cells were harvested and resuspended in 20 mM Hepes-Na [pH 7.0] buffer, containing 8 M urea and sonicated. The lysate was centrifuged, and the supernatant was loaded onto a nickel-nitrilotri-acetic acid-agarose (Ni-NTA) column (Qiagen). The column was washed with 20 mM Hepes-Na buffer, pH 7.0, containing 8 M urea and the protein was eluted with 20 mM phosphate buffer, pH 4 containing Urea. Fractions containing the eluted protein were pooled and dialyzed into 10 mM sodium phosphate, pH 6.5,
buffer. 70% of formic acid (v/v) was added to the resultant sample and the N-terminal His tag and trpL were removed by cyanogen bromide cleavage at room temperature for 1 h. The sample was cleared by centrifugation and dialysed into buffer A (10 mM phosphate buffer pH 7.0) before anion exchange chromatography on a Q column (Amersham Biosciences). The protein was eluted with a gradient to buffer B (10 mM phosphate buffer pH 7.0, 1M NaCl). The fractions containing eluted protein were concentrated by using Ultra-Centricones (Millipore). Purity was estimated to be >95% by Coomassie-stained SDS-PAGE.

**Purification of mDia1 FH1 and its fragments.**

To express the His-MBP-FH1 fusion protein, H-MBP-FH1 was transformed into *E. coli* strain BL21(DE3) Codon+ (Novagen). Cells were grown in Luria broth to 0.7 $A_{600}$ at 37 °C, induced with 0.5 mM IPTG, and grown overnight at room temperature. Cells were harvested and resuspended in 20 mM Hepes-Na [pH 7.0], 100 mM NaCl buffer and lysed by sonication. The supernatant was loaded onto a nickel-nitrilotri-acetic acid-agarose (Ni-NTA) column (Qiagen). The column was washed with 20 mM Hepes-Na buffer, pH 7.0 and the protein was eluted with 20 mM phosphate buffer, containing 250 mM Imidazole. The fractions containing the eluted protein were dialyzed into PreScission cleavage buffer (20 mM Tris pH 7.0, 150 mM NaCl, and 1mM DTT). MBP-FH1 was cleaved by PreScission Protease overnight at room temperature. His-MBP fusion was removed by passing the protein sample through a Ni-NTA column equilibrated with PreScission cleavage buffer. The protein was concentrated by using Ultra-Centricones (Millipore).

To express His-TrpL-FH1-pep1 and His-TrpL-FH1-pep2, pTM-FH1-pep1 and pTM-FH1-pep2 were transformed into *E. coli* strain BL21(DE3) Codon+ (Novagen). Cells were grown in Luria broth to 0.7 $A_{600}$ at 37 °C, induced with 0.5 mM IPTG, and grown overnight. Cells were harvested and resuspended in 20 mM Hepes-Na [pH 7.0], 100 mM NaCl buffer containing 8 M
urea and lysed by sonication. The lysate was centrifuged, and the supernatant was loaded onto a Ni-NTA column (Qiagen). The column was washed with 20 mM Hepes-Na buffer, pH 7.0, containing 8 M urea and the protein was eluted with 20 mM phosphate buffer, pH 4 containing 8 M urea. Fractions containing the eluted protein were pooled and dialyzed into 10 mM sodium phosphate, pH 6.5, buffer. 70% of formic acid (v/v) was added to the resultant sample and the N-terminal His tag and trpL were removed by cyanogen bromide cleavage at room temperature for 1 h. Cyanogen bromide also cleaved FH1-pep2 at two internal Met sites creating three additional peptides, FH1-pep2-del1 (amino acids 683-712), FH1-pep2-del2 (714-745), and FH1-pep2-del3 (749-771). The sample was cleared by centrifugation and dialysed into 10% acetic acid. After lyophylization, the sample was resuspended in buffer A (0.01% trifluoroacetic acid (TFA)) before reverse phase chromatography on a C18 column (Waters). The protein was eluted with a gradient to buffer B (90% acetonitrile, 0.015 TFA). The fractions containing the eluted protein were concentrated by using Ultra-Centricones (Millipore). Purity was estimated to be >95% by Coomassie-stained SDS-PAGE.

**Site-Directed Mutagenesis of ctRAGE.**

To doubly mutate the ctRAGE, the QuikChange II XL Site-Directed Mutagenesis Kit (Strategene) was used. Following mutagenic PCR, pTM-ctRAGE was restriction digested with DpnI for 1 hour and transformed into *E. coli* strain DH10B. Mutated plasmid was isolated and purified using Mini-Prep Kit (Qiagen). DNA sequencing identified plasmid pTM- R5A Q6A ctRAGE, which code for the appropriate mutant ctRAGE.
NMR Experiments

Protein samples of the uniformly labeled, \([U^{-13}C,^{15}N]\) and \([U^{-15}N]\), ctRAGE, with concentrations ranging from 60 to 300 \(\mu\)M were dissolved in NMR buffer (10 mM potassium phosphate [pH 6.5], 100 mM NaCl, 0.02% (w/v) NaN\(_3\), in 90%H\(_2\)O/10%D\(_2\)O). To obtain backbone resonance assignments of \([U^{-13}C,^{15}N]\) ctRAGE, standard triple resonance spectra \(^1\)H,\(^1\)N HSQC, HN(CA)CO, HNCO, HN(CO)CA, HNCA, CBCA(CO)NH, and HNCA(CB) were acquired at 298 K using an Avance Bruker spectrometer operating at a \(^1\)H frequency of 700 MHz equipped with a single Z-axis gradient cryoprobe. To obtain the side-chain resonance assignments of ctRAGE, \(^1\)H, \(^{13}\)C HSQC, \(^1\)H, \(^{13}\)C 3D NOESY-HSQC and 3D HCCH-TOCSY experiments [22] were acquired. To obtain steady state \(^{15}\)N nOe values for ctRAGE, standard experiments were used[23]. All spectra were processed using TOPSPIN 2.1 (Bruker, Inc), and assignments were made by using CARA[24].

To perform hydrogen-deuterium exchange experiments, \([U^{-15}N]\) ctRAGE was lyophilized and reconstituted on ice in deuterated NMR buffer (10 mM potassium phosphate [pH 6.5], 100 mM NaCl, 0.02% (w/v) NaN\(_3\), in 100%D\(_2\)O). The final concentration of the NMR sample was 1 mM. Deuterium exchange was monitored by a series of \(^{15}\)N HSQC experiments conducted at 6 °C which were collected every 5 minutes with a duration of 5 minutes.

To study mDia1-ctRAGE interactions, the NMR titration experiments were performed. 0.5 mM of unlabeled mDia1 FH1, in NMR buffer were titrated into 100 \(\mu\)M \([U^{-15}N]\) ctRAGE domain in 4 steps to yield ctRAGE to mDia1 FH1 molar ratios of 5:1, 2:1, 1:1, 1:2 respectively. The results of the titration were monitored by \(^1\)H[25]-HSQC. Over the course of titration, the signal to noise ratio of the peaks that did not show any changes was kept constant by adjusting the number of scans.
Structure Calculation

Structural calculations were carried out with Cyana 2.1[26] using 295 distance restraints derived from $^{13}$C-edited NOESY and $^{15}$N-edited NOESY spectra, 33 pairs of backbone torsion angle restraints derived from TALOS[27], and 2 restraints for hydrogen bonds. nOes were converted to upper limit distances using the CALIBA module in CYANA[26]. The reference volume determined by CALIBA was increased 2 times before conversion in order to loosen the distance restraints. All upper limit distances for intermolecular nOes were set to 6 Å. These experimental restraints are summarized in Table 1. To perform CYANA calculations, a single polypeptide chain was constructed for the ctRAGE molecule. Refinement: The CYANA-generated distance and angle restraints were converted into CNS format in CCPN[28]. A total of 1,000 structures were calculated, and the 200 lowest energy structures were subjected to water refinement and further analysis by PROCHECK_NMR[29]. Eleven of the ctRAGE residues from the structured region 1-15 were in the most favorable regions of the Ramachandran plot and one (E11) was in generously allowed regions. There were no residues in the disallowed regions of the Ramachandran plot. The structural statistics of the 20 best structures are reported in Table 1.
### Table 1  NMR and refinement statistics for ctRAGE structures (residues 2-42)

<table>
<thead>
<tr>
<th>NMR distance and dihedral constraints</th>
<th>ctRAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance constraints</td>
<td></td>
</tr>
<tr>
<td>Total NOE</td>
<td>295</td>
</tr>
<tr>
<td>Intra-residue</td>
<td>82</td>
</tr>
<tr>
<td>Inter-residue</td>
<td>213</td>
</tr>
<tr>
<td>Sequential (</td>
<td>i-j</td>
</tr>
<tr>
<td>Medium-range (</td>
<td>i-j</td>
</tr>
<tr>
<td>Long-range (</td>
<td>i-j</td>
</tr>
<tr>
<td>Hydrogen bonds</td>
<td>2</td>
</tr>
<tr>
<td>Total dihedral angle restraints</td>
<td></td>
</tr>
<tr>
<td>phi</td>
<td>33</td>
</tr>
<tr>
<td>psi</td>
<td>33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structure statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Violations (mean and s.d.)</td>
<td></td>
</tr>
<tr>
<td>Distance constraints (Å)</td>
<td>0.035 +/- 0.0094</td>
</tr>
<tr>
<td>Dihedral angle constraints (°)</td>
<td>2.5 +/- 0.34</td>
</tr>
<tr>
<td>Max. dihedral angle violation (°)</td>
<td>4</td>
</tr>
<tr>
<td>Max. distance constraint violation (Å)</td>
<td>0.2</td>
</tr>
<tr>
<td>Deviations from idealized geometry</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.012</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>0.38</td>
</tr>
<tr>
<td>Impropers (°)</td>
<td>0</td>
</tr>
<tr>
<td>Average pairwise r.m.s.d.* (Å)</td>
<td></td>
</tr>
<tr>
<td>Heavy</td>
<td>1.89 +/- 0.25</td>
</tr>
<tr>
<td>Backbone</td>
<td>0.90 +/- 0.16</td>
</tr>
</tbody>
</table>

**Pairwise r.m.s.d. was calculated for residues 2-15 among 20 refined structures.
Cell Lines and Materials

Wild-type primary murine aortic vascular smooth muscle cells were isolated and employed through passage 5 to 7. Vascular SMCs were grown in 10% FBS containing DMEM medium (Invitrogen).

Site-directed mutagenesis of full length RAGE

The human RAGE full-length cDNA was cloned and inserted into pcDNA3.1 expression vector as described previously[30]. This plasmid DNA was used as the template to create the double mutation at R5 and Q6 within the RAGE cytoplasmic domain using the Quickchange site-directed mutagenesis kit (Stratagene). The mutated plasmid was sequenced to ensure that only the desired substitutions to alanine were present.

Transfection

Wild-type primary aortic vascular smooth muscle cells were transfected with vector or mutant human RAGE cDNA construct using the Nucleofector kit (Lonza).

Western blot analysis

Total cell lysates were immunoblotted and probed with AKT-specific antibody, p-AKT-specific antibody. HRP-conjugated donkey anti-rabbit IgG (Amersham Pharmacia Biotechnology) or HRP-conjugated sheep anti-mouse IgG (Amersham Pharmacia Biotechnology) was used to identify sites of binding of the primary antibody. After probing with the primary antibodies, membranes were stripped and reprobed for relative total AKT protein. Blots were scanned by using an AlfaImager TM 2200 scanner with AlfaEase (AlfaImager) FC 2200 software. Results are reported as relative absorbance of test antigen to relative total protein. In all Western blot studies, at least triplicate cell lysates per group were used; results of representative experiments are shown.
Smooth Muscle Cell Migration and Proliferation Assays.

Wild-type vascular smooth muscle cells were isolated and cultured from the mouse aorta and employed through passage 5 to 7. Migration assays were performed using the QCM Colorimetric Cell Migration Assay (Chemicon). Cells (3x10^5/well) were seeded into the upper chambers fitted with a lower 8 μm porous polycarbonate membrane, and the insert was placed in the lower chamber of a 24-well dish containing Dulbecco’s modified Eagle medium and no stimulant, 10 μg/mL S100B (generously provided by Dr. Guenter Fritz), or 10 ng/mL PDGF (R&D Systems) and incubated at 37 °C for 5 hrs and relative migration was measured according to the manufacturer’s instructions. Proliferation of cultured SMCs was quantified by measuring the incorporation of tritiated thymidine. SMCs were seeded at a density of 2x10^4 cells/well in 24-well tissue culture-treated plates and incubated in serum-free DMEM for 16 hrs. Cells were exposed to serum-free DMEM containing the indicated concentration of S100B or PDGF along with ^3^H-thymidine (1 μCi/well). Cells were harvested 48 hrs after the incubation period, and cellular proliferation was determined based on the incorporation of tritiated thymidine. Cell counting was performed and confirmed that increased trititated thymidine incorporation reflected an increase in cell number.

Data analysis

The mean standard deviation (SD) is reported. Statistical comparisons between groups were determined using one-way analysis of variance (ANOVA); where indicated, individual comparisons were performed using Students’ t-test.
Results

C-tail of RAGE possesses a region with ordered tertiary structure.

To initiate structural analysis of the ctRAGE-mDia1 interaction, we bacterially overexpressed and purified the C-terminal RAGE fragment (ctRAGE) (Fig. 1.0 A). Solution NMR spectroscopy is well suited to characterize small peptides such as ctRAGE, and their interactions. The heteronuclear single quantum coherence NMR experiment (HSQC) provides information about the protein backbone structure and is used to assess the suitability of a protein for structure determination[22]. The $^{15}$N-HSQC spectrum of ctRAGE is well-resolved but has limited chemical shift dispersion (Fig. 1.0B) indicating that the majority of the ctRAGE residues are disordered.
**Fig. 1.0** ctRAGE contains a folded segment. A. Sequence alignment of ctRAGE. Conserved residues are in red. Residues mutated to Ala are in blue. B. $^{15}$N-HSQC NMR spectrum of ctRAGE. The assignments of the backbone amide protons and nitrogens are shown by the residue sequence number. Amide proton peaks show limited chemical shift dispersion, from 7.5 ppm to 8.8 ppm, which is characteristic of predominantly unstructured proteins. C. Steady state $^{15}$N nOe indicates that N-terminus of ctRAGE, residues 2-15, is folded. Residues 16-42 exhibit low or even negative values of steady state $^{15}$N nOe, which indicate an unstructured C-terminal tail. D. Hydrogen-deuterium exchange experiment. The $^{15}$N HSQC spectrum shows residual amide proton and nitrogen peaks from $[U-^{15}$N] ctRAGE remaining after 5 min at 6 °C. The $^{15}$N-HSQC experiment was conducted at 6 °C to slow the hydrogen-deuterium exchange rate. Amide protons, which form hydrogen bonds and also amide protons of charged amino acids exhibit impeded exchange rates.
To identify ordered regions of the ctRAGE primary structure, we assessed the flexibility of the ctRAGE backbone. The nuclear Overhauser effect (nOe), which depends on local motions of the backbone NH vectors, provides a convenient measure to identify residues participating in the folded structure: steady-state $^{15}$N{$^1$H} nOes in proteins are positive and close to one for residues from folded regions and small or negative for residues from disordered segments[23, 31-33]. Steady-state $^{15}$N{$^1$H} nOe NMR experiments identified a region of ctRAGE that has limited flexibility due to structural constraints and is likely to be folded. The results (Fig. 1.0C) show that only the N-terminus of ctRAGE (residues 2-15) possesses large positive $^{15}$N{$^1$H} nOe values and is thus ordered. Amino acids 16-42 exhibit small positive or even negative steady state nOes suggesting unrestricted local flexibility of the backbone indicative of disordered protein fragments[33, 34].

Hydrogen bonds are often critical to hold the structure of small proteins together. To confirm that the N-terminal fragment of ctRAGE may contain intramolecular hydrogen bonds, we performed an amide hydrogen-deuterium exchange experiment in which the extent of proton-deuterium exchange is monitored by NMR spectroscopy[35, 36]. Solvent exposed backbone amide hydrogens are usually efficiently exchanged with deuterium if a protein is dissolved in a deuterated solvent. Protein amide hydrogens involved in intramolecular hydrogen bonding are expected to be exchanged at a slower rate than solvent exposed amide hydrogens. The local electrostatic field of the amide proton can also strongly affect the exchange rate[37]. After 5 minutes of hydrogen deuterium exchange, only a few strong peaks corresponding to R5, G9 and also E16, E26, and N28 can be seen in the $^{15}$N-HSQC spectrum of ctRAGE (Fig. 1.0D) suggesting that hydrogen bonds may indeed be present in the folded part of ctRAGE.
ctRAGE folds into an alpha-turn.

We used standard NMR experiments to solve the solution structure of the ctRAGE ordered region, comprised of residues 2-15. Based on 295 distance constraints and 66 dihedral angle constraints obtained from NMR experiments, 20 structures with the lowest target function values were superimposed (Fig. 1.1A and Table 1). The root mean standard deviation (RMSD) of the backbone atoms of the best 20 structures was 0.9 Å (Fig. 1.1A); well-folded proteins typically exhibit a backbone RMSD below 0.5 Å. The average value of 0.5 for the steady-state $^{15}$N nOes also suggests that the backbone of ctRAGE is flexible[33], which is to be expected for a short non-cyclic peptide[34] These data suggest that the structure is dynamic and may include multiple conformers.

The structure consists of a loop stabilized by electrostatic interactions between R5 and E10, and E11 and R12, as well as two hydrogen bonds between the backbone amide of R5 and carbonyl E11, and the amide of G9 and carbonyl of R5 (Fig. 1.1B and Supplementary (Fig. 1). To validate the solution structure, we made a double mutant in which R5 and Q6 were changed to alanine to create R5AQ6A-ctRAGE. The mutations would eliminate critical electrostatic interactions between R5 and E10 and disrupt the tertiary structure. Indeed, the mutations greatly affect chemical shifts of residues located far away from the mutation site (Supplementary Fig. 2).

The pattern of hydrogen bonds classifies the structure as an alpha-turn[38]. The alpha-turn corresponds to a chain reversal involving five amino acids and that is stabilized by a hydrogen bond between the carbonyl group of the first residue and the amino group of the fifth[38]. This structural element is found at the C-terminal helical part of proteins and peptides, and on the loop side of beta-hairpins[39].
Fig. 1.1. Solution structure of the ctRAGE fragment (amino acids 2-15).  (A) Cluster of 20 solution backbone traces of ctRAGE. The closeness of the traces to each other reflects the overall quality of the solution structure. The RMSD of the ctRAGE cluster is 1.5 Å. (B) Ribbon diagram of ctRAGE. Hydrogen bonds between R5 and E11 and G9 and R5 are indicated by black lines. The hydrogen bond pattern indicates that ctRAGE folds into an alpha-turn.  (C) Electrostatic map of ctRAGE. Positive and negative surfaces are colored blue and red, respectively.
The protein molecular surface is critical for understanding protein-protein interactions. We constructed an electrostatic map of ctRAGE that shows that the ctRAGE molecular surface is highly charged as is expected based on the primary structure (Fig. 1.1C). At the same time, there is a hydrophobic patch formed by the R5 and Q6 side chains that may be important for protein-protein interactions.

In vitro identification of residues critical for the mDia1 FH1-ctRAGE-interaction.

In order to characterize interactions between mDia1 FH1 and ctRAGE, we bacterially overexpressed and purified a maltose binding protein (MBP) fusion construct of FH1. MBP was subsequently removed by using PreScission protease releasing FH1 (Fig. 1.2A). The mDia1 FH1 construct (Fig. 1.2A) is slightly larger than the construct used in the original study[13] that reported the discovery of the mDia1-RAGE interaction. mDia1 FH1 includes poorly conserved polyproline regions as well as a highly conserved region between the FH1 and FH2 domains of mDia1[16].
Fig. 1.2. ctRAGE interacts with mDia1 FH1. (A) Sequence alignment of human mDia1 FH1 construct used in this study (NCBI Accession code NP_005210) and mouse mDia-1 FH1 (NCBI Accession code NP_031884). Conserved residues are in red. The boundaries of mDia1 FH1 small fragments, FH1-pep1, FH1-pep2-del1, FH1-pep2-del2, FH1-pep2-del3, respectively, are indicated by solid lines above the sequence. (B) $^{15}$N-HSQC NMR spectrum of the mDia1 FH1-$[^{15}$N] ctRAGE complex. To form mDia1FH1-ctRAGE complex, 0.5 mM of unlabeled mDia1 FH1, in NMR buffer (10 mM potassium phosphate [pH 6.5], 100 mM NaCl, 0.02% (w/v) NaN₃, in 90%/10% H₂O/D₂O) was added into 100 μM $[^{15}$N] ctRAGE to yield mDia1 FH1 to ctRAGE molar ratios of 1:2. Due to $^{15}$N editing of the experiment, only backbone amide protons and nitrogens of ctRAGE are present in the spectrum. Most peaks do not change their positions reflecting the fact that only a subset of ctRAGE residues interact with FH1. ctRAGE peaks that are substantially or completely broadened are labeled. (C) The mDia1 FH1-ctRAGE interaction map. Residues broadened during the NMR titration experiment are indicated in red.
Because of the exquisite sensitivity of chemical shifts to the chemical environment, solution NMR spectroscopy is used to identify interaction surfaces between reacting molecules[22, 40]. Molecular binding preferentially perturbs the chemical environment of atoms in the immediate vicinity of the binding site leading to changes in the corresponding NMR spectrum. Titrating unlabeled FH1 into uniformly $^{15}$N-labeled ctRAGE, [$U^{-^{15}}$N]-ctRAGE, resulted in specific changes in the HSQC spectrum of ctRAGE (Fig. 1.2B). Due to $^{15}$N editing of the experiment, only backbone amide protons and nitrogens of ctRAGE are present in the spectrum. When the molar ratio between FH1:ctRAGE reached 1:1, peaks corresponding to Q3, R4, R5 and Q6 either disappeared or were substantially broadened when compared to the rest of the peaks. These changes suggest that residues Q3 through Q6 participate in the interaction with FH1. Peak broadening is characteristic of slow exchange and suggests that the dissociation constant of the interaction $K_d$ is $< 10$ mM[3, 41]. This result is in good agreement with the results from FH1-RAGE binding studies reporting a $K_d$ of 30 $\mu$M[13].

Mapping the observed chemical shift changes onto the molecular surface of ctRAGE allowed us to identify the FH1-ctRAGE interaction surface. It consists of a hydrophobic patch formed by R5 and Q6 that is contiguous with a positively charged surface formed by R4 and R5 (Fig. 1.2C). To confirm the FH1-ctRAGE interaction surface we made a ctRAGE double mutant in which residues R5 and Q6 were changed to alanine to create R5AQ6A-ctRAGE (Fig. 1.0A). Titrating unlabeled FH1 into [$U^{-^{15}}$N] R5AQ6A-ctRAGE resulted only in small ($< 0.02$ ppm) changes in the position of one amino acid residue, R10. No other chemical shift changes or substantial broadening of the R5AQ6A-ctRAGE peaks were observed, suggesting that the double mutant interacts with FH1 very weakly, with a $K_d$ well above 1 mM (Supplementary Fig. 3). We conclude that R5 and Q6 are an important part of the FH1-ctRAGE interaction surface.
We attempted to identify the specific region(s) of FH1 that interacts with ctRAGE by creating 4 deletion constructs: FH1-pep1 (amino acids 646-660), FH1-pep2-del1 (amino acids 683-712), FH1-pep2-del2 (714-745), which contains short polyproline fragments, and FH1-pep2-del3 (749-771), which contains a part of the conserved FH1-FH2 linker (Fig. 1.2A). Based on NMR titration experiments, none of these constructs interact with ctRAGE. These results suggest that mDia1 FH1 forms a tertiary structure that is necessary for the mDia1 FH1 interaction with ctRAGE and this structure is absent in the deletion constructs.

*R5AQ6A-RAGE evokes a dominant negative response in RAGE-mDia1 signaling.*

The data suggest that the binding of mDia1 FH1 to ctRAGE depends upon residues R5 and Q6 of ctRAGE. To determine the potential mechanistic implications of these findings, we transfected R5AQ6A-ctRAGE or vector alone into primary murine aortic smooth muscle cells (SMCs). When vector-treated primary murine SMCs were incubated with RAGE ligand S100B, an increase in phosphorylation of AKT was noted after 5, 10 and 20 minutes of incubation. In contrast, in cells transfected with the R5AQ6A mutant, S100B failed to phosphorylate AKT over the same time course (Fig. 1.3)
Fig. 1.3 S100b stimulates AKT activation in vascular smooth muscles cells via ctRAGE-mDia1 FH1 interaction through R5 and Q6. Transiently-transfected vector control, RAGE-overexpressing double mutant primary murine aortic SMCs were stimulated with 10 µg/mL of S100B for the indicated times. Total lysates were Western blotted with antibodies against total AKT or p-AKT. Quantified levels of phosphorylated and total AKT in the wild type-transfected SMCs at different times following stimulation with S100B are shown. Fold changes are relative to control. Error bars represent SD (*P < 0.05).
Fig. 1.4. S100B stimulates migration (A) and proliferation (B) in vascular SMCs mainly via ctRAGE-mDia1 FH1 interaction through R5 and Q6. Vector or mutant human RAGE cDNA construct transfected SMCs were treated with 10 µg/mL of S100B or 10 ng/mL of platelet-derived growth factor (PDGF) for 5 or 48 hours. At the end of that time, migration (A) and proliferation (B), were assessed. Assays were performed in triplicate and the results shown represent three independent experiments. Error bars represent SD (*P < 0.005).
Fig. 1.4 continued. S100B stimulates migration (A) and proliferation (B) in vascular SMCs mainly via ctRAGE-mDia1 FH1 interaction through R5 and Q6. Vector or mutant human RAGE cDNA construct transfected SMCs were treated with 10 µg/mL of S100B or 10 ng/mL of platelet-derived growth factor (PDGF) for 5 or 48 hours. At the end of that time, migration (A) and proliferation (B), were assessed. Assays were performed in triplicate and the results shown represent three independent experiments. Error bars represent SD ( *P < 0.005).

To test the functional implications of these findings, we treated vector- and R5AQ6A-transfected cells with S100B and tested migration and proliferation responses in primary murine aortic SMCs. In vector- treated cells, incubation with S100B resulted in increased SMC migration and proliferation compared to control (Fig. 1.4 A and 1.4 B). However, in R5AQ6A-transfected cells, S100B failed to increase migration or proliferation of SMCs. When vector- and R5AQ6A-transfected SMCs were treated with a non-RAGE ligand, PDGF, migration and proliferation responses of SMCs were not affected (Fig. 1.4 A and 1.4 B).
Discussion

Deducing the interactions between ctRAGE and intracellular proteins is important to understand how RAGE signal transduction is implicated in various disease states. The interaction between ctRAGE and the FH1 domain of mDia1 is clearly involved in migration and cell proliferation: two critical processes in the etiology of atherosclerosis, tumors and immune/inflammatory disorders. Studies in multiple cell types indicate that the short 42 amino acid cytosolic tail of RAGE, ctRAGE, is absolutely required for RAGE signaling[9-12], however, ctRAGE contains no phosphotyrosine, -threonine or -serine motifs that are typically used for intracellular signal propagation (Fig. 1.0A). Here we characterized the intracellular mechanism of RAGE signaling by studying the interaction of mDia1 with ctRAGE.

Our structural analysis revealed that ctRAGE contains an N-terminal segment that folds into an alpha-turn and a long unstructured C-terminal tail. Alpha-turns are characterized by a hydrogen bond between the first and the fifth residue in the turn[38]. In globular proteins, these structures are usually exposed to solvent and protrude outward from the protein surface with a hook-like shape[39]. Alpha-turns function as a protein or ligand interacting module in, for example, the CD4 receptor, cyclooxygenase-1, and alpha-lytic protease[38]. Alpha-turns are also relevant structural domains in small peptides, particularly in cyclopeptides containing 7-9 residues [42, 43]. Alpha-turns contain mainly hydrophilic amino acids[38]. Indeed, the alpha-turn of ctRAGE consists of 4 charged amino acids, R5, R7, R8 and E10.

The FH1-ctRAGE interaction surface lies within the alpha-turn. Since the ctRAGE primary structure is well-conserved beyond the alpha-turn (Fig. 1.0 A), it is likely that mDia1 is not the only intracellular effector of RAGE signaling. The interaction surface is contiguous and contains hydrophilic and polar residues located on one side of the ctRAGE alpha-turn (Fig. 1.2C). Ionic as well as non-ionic interactions may be critical for mDia1 FH1 binding. The R5AQ6A
double mutant abolishes the FH1-ctRAGE interaction and no interaction was detected when short fragments of the FH1 domain were used instead of the intact FH1 domain. These results suggest that the FH1 domain possesses a well-defined structure required for binding to ctRAGE.

Previous cellular and in vivo studies in which novel transgenic mice were generated to express the dominant negative form of RAGE in a cell specific manner, such as in endothelial cells, smooth muscle cells, neurons, and macrophages[9-12], examined the role of the cytoplasmic domain of RAGE in RAGE ligand-stimulated signaling by completely deleting the intracellular domain. The present study establishes that key residues within the cytoplasmic domain are required to bind to the FH1 domain of mDia1, and that such binding is essential for RAGE ligand-induced AKT signaling, migration and proliferation in SMCs. The observation that R5Q6 ctRAGE has no effect on PDGF-stimulated migration and proliferation demonstrates that the double mutant does not non-specifically abolish proliferation and migration responses in SMCs. Although it was postulated that the complete deletion of the cytoplasmic domain exerts its effects by blocking RAGE signaling consequent to ligand binding, it is plausible that the extracellular and transmembrane domains of the receptor lacking the cytoplasmic domain serve as a sink for ligands, thereby reducing the binding to any signaling receptor. The present work supports the concept that RAGE ligands exerts their biological effects by binding to the extracellular domain of the receptor and transducing a signal through the intracellular domain.

The FH1 domain forms an extended structure within mDia1 making the FH1 domain accessible for binding to ctRAGE[44-46]. Indeed, ctRAGE can bind the FH1 domain within full-length mDia1[13]. This result suggests that the alpha-turn of ctRAGE is constitutively present (i.e. even in the absence of an extracellular RAGE ligand), implying that mDia1 may also be constitutively bound. What then triggers RAGE-mDia1 signaling when ligands binds to RAGE?
Many RAGE ligands are multivalent and cause RAGE molecules to move together on the cell surface[3, 41, 47-50]. This clustering will lead to intracellular clustering of ctRAGE, which, in turn, will increase the local concentration of mDia1 bound to ctRAGE (Fig. 1.5). In isolation, mDia1 is autoinhibited by the binding of the C-terminal diaphanous autoinhibitory domain and N-terminal diaphanous inhibitory domain[44, 51-53]. The upstream effector of mDia1, Rho-A, can release this autoinhibition, activating mDia1 for actin polymerization[54]. It is also known that domain-domain contacts within mDia1 are likely to be quite dynamic, suggesting the possibility of intermolecular domain swapping between adjacent mDia1 molecules[44]. We propose that the high local concentration of mDia1, induced by the binding of RAGE ligands, leads to an mDia1 configuration, which is active (Fig. 1.5). It is important to note that this mode of activation is different from that of RhoA. RhoA activation leads to a many-fold increase in phosphorylation of downstream effectors as opposed to the 50% increase observed due to RAGE activation of mDia1 (Fig. 1.3, 1.4 A, and 1.5 B).

In conclusion, our findings show that the FH1-ctRAGE interaction is critical for signaling and functional parameters in primary SMCs, and reveal for the first time the structural mechanism through which the RAGE cytoplasmic domain interacts with the FH1 domain of mDia1. These data identify a novel binding interface as a target for suppression of RAGE ligand-stimulated signal transduction.
**Fig. 1.5. Model of RAGE induced activation of mDia1.** FH1 and FH2 domains of mDia1 are required for mDia1 activity. DID and DAD are the N-terminal Diaphanous Inhibitory Domain and C-terminal Diaphanous Autoregulatory Domain of mDia1, respectively. mDia1 is autoinhibited due to the interaction of the regulatory DID and DAD domains. RAGE can constitutively bind to mDia1 by using the ctRAGE-FH1 interaction. Clustering of mDia1 molecules due to extracellular RAGE (exRAGE) binding to a RAGE ligand results in intermolecular domain swapping leading to partial restoration of mDia1 activity.
Chapter 2

In-Cell NMR Spectroscopy
Introduction

Interactions between biological macromolecules give rise to and regulate biological activity. This activity is manifest through structural dynamics and changes in the macromolecular structures that comprise these interactions [55-57]. Until recently, mostly \textit{in vitro} techniques have been used to study macromolecular interactions that govern biological processes under conditions remote from those existing in the cell [58]. With the advent of in-cell Nuclear Magnetic Resonance (NMR) spectroscopy [58], these processes can now be studied within a cellular environment.

In-cell NMR spectroscopy provides atomic level resolution of molecular structures under physiological conditions. NMR-active nuclei in biological macromolecules are extremely sensitive to changes in the chemical environment resulting from specific and non-specific binding interactions with ions, small effector ligands and macromolecules as well as changes due to biochemical modifications. These interactions alter molecular surfaces and may result in tertiary and quaternary conformational changes, all of which are reflected by changes in the chemical shifts of these nuclei. Thus, by performing NMR spectroscopy on living cells, we can begin to understand the structural underpinning of biological activity.

As the field of in-cell NMR spectroscopy has progressed, the severity of early concerns regarding the validity of in-cell NMR for studying biological macromolecules, has abated. Differences between the resonance peaks of proteins measured in-cell \textit{versus} those measured \textit{in vitro} are small, reflecting the effect of the intracellular environment on the protein structure. Potential pitfalls in the technique regarding low signal intensity requiring the need for abnormally high, non-physiological concentrations of the labeled target, the effects of molecular crowding inherent to the cytosol, the relevance of studying prokaryotic proteins in a eukaryotic intracellular milieu, the viability of cells during data acquisition and the ability to expand in-cell
methodology to eukaryotic cells, have proven to be more tractable than expected [15, 48]. The results have reaffirmed the power of in-cell NMR spectroscopy to measure changes in structure, due to post-translational biochemical modification, interactions with other biological molecules and/or allosteric changes resulting from binding interactions under physiological or near physiological conditions and in determining three-dimensional (3D) structures de novo.

Methodologies

Target labeling

To use NMR spectroscopy to study biological macromolecules in living cells the labeled targets must be easily distinguished from all other species present. Specific labeling schemes are employed to detect and resolve in-cell NMR protein resonance peaks and to yield the lowest background signals.

Backbone group probes

The most commonly used scheme employs $^{15}$NH$_4$Cl as the sole nitrogen source to incorporate NMR-active $^{15}$N nuclei into the peptide backbone of proteins over-expressed in bacterial cells. This uniform [$U$-$^{15}$N] labeling strategy results in diminished background resonance peaks due to the fact that during protein over-expression bacterial growth is significantly reduced. The protein resonance peaks of the resulting $^1$H[25]-HSQC (heteronuclear single quantum coherence) spectra are adequately resolved provided the protein is expressed to a high enough concentration within the cell.

Another backbone labeling scheme uses auxotrophic bacterial strains [59, 60] to incorporate high levels of amino acids that have been specifically enriched with $^{15}$N. Arginine, histidine and lysine are ideal for this type of labeling [59] since these amino acids lie at the end of their respective biosynthetic pathways. Labeled amino acids are chosen so as to comprise a substantial
number of the total residues in the protein being expressed, since only those residues will contribute to the NMR spectrum. As a result of the less extensive labeling, the resulting in-cell spectrum will necessarily exhibit lower resolution than a spectrum obtained for a uniformly labeled target, but will be essentially devoid of background signals. Other amino acids can be used with the caveat that there will be an unavoidably larger background signal due to the presence of multiple metabolic products.

Proteins may also be uniformly labeled in bacteria by using $^{13}$C as the sole carbon source during bacterial over-expression. Uniform [$U^{-13}$C] labeling results in a high background due to metabolic reactions and the natural abundance of $^{13}$C (1.1%), and is generally not used in this capacity. Selective $^{13}$C-isotopic labeling of constituent amino acids offers a better opportunity for enhancing the signal and minimizing the background. This isotope is best suited for labeling amino acid side chain residues, particularly methyl and methylene groups.

*Methyl group probes*

To produce narrow resonance lines for in-cell NMR, the protein of interest must tumble rapidly inside the cell. The rate of tumbling decreases as the size of the molecule increases, limiting in-cell NMR to the study of small molecules. Serber et al. [61] extended in-cell NMR spectroscopy to the study of larger molecules by selectively incorporating isotopically labeled methyl groups into protein over-expressed in bacteria.

Methyl groups contain three protons coupled to carbon and generate an NMR spectrum with three times the sensitivity obtained from a single amide proton. Due to rapid local motions, methyl groups also have longer relaxation times than amide protons, which further increases the sensitivity of the NMR signal. In addition, amide protons chemically exchange with water resulting in a loss of signal due to broadening, whereas, methyl group protons do not. Therefore, methyl groups provide probes that are more sensitive than amide groups for in-cell NMR studies.
Serber et al. [61] tried different labeling schemes to determine which yielded the greatest resolution and lowest background. When $^{13}$C-glucose was used to uniformly label Calmodulin in *E. coli*, the $^1$H($^{13}$C )-HSQC spectrum of these cells showed a high background, which precluded identifying most of the methyl group proton resonances. When $^{13}$C-pyruvate, a precursor for the biosynthesis of amino acids side chains, was used as the sole carbon source [62] to label over-expressed Calmodulin in D$_2$O-minimal medium, the overall number of peaks in the spectrum was reduced relative to those obtained by using full labeling since only the amino acids that required pyruvate as a precursor incorporated the $^{13}$C labeling. Many protein resonance peaks were observed, but assigning specific resonances was hampered by substantial overlap with broad background peaks.

Peak resolution was improved and background reduced by using a single labeled amino acid. ($^{13}$C-methyl)-methionine was used to label Calmodulin overexpressed in minimal medium. The resulting $^1$H($^{13}$C )-HSQC spectrum was dominated by free methionine and well-defined Calmodulin methionine resonances. Free ($^{13}$C-methyl)-methionine was removed from the in-cell sample by washing the cells prior to NMR spectroscopy to yield a further reduction in background peaks. When ($^{13}$C-methyl)-alanine was used to label NmerA, the resulting spectrum displayed a higher background than observed using ($^{13}$C-methyl)-methionine, as well as free alanine and NmerA alanine resonances. By using M9 medium supplemented with amino acids and nucleosides [63], the background was reduced even further. Thus, selective isotopic labeling of methyl groups provides a viable probe to study large molecular complexes by using in-cell NMR spectroscopy.
Fluorine probes

The study of $^{15}$N labeled globular proteins by in-cell NMR is not always possible due to the high viscosity and weak interactions that occur in the cytoplasm. Wang et al. [64] demonstrated an $^{19}$F NMR labeling strategy that is applicable to the in-cell study of both globular and disordered proteins. $^{19}$F labeling is very attractive for in-cell NMR studies. The gyromagnetic ratio and natural abundance of $^{19}$F is very close to that of $^1$H making $^{19}$F as sensitive as protons and the $^{19}$F chemical shift range is at least ten times larger than that of protons, which confers greater resolution. Naturally occurring proteins contain no fluorine so that $^{19}$F labeled amino acids introduced during bacterial over-expression will result in a very low NMR background signal. Since acquisition times to collect $^{19}$F spectra are very short compared to those of 2D-HSQC experiments, proteins may be studied near physiological concentrations.

Wang et al. [64] examined one disordered and five globular proteins labeled with $^{19}$F, ranging in size from 7 to 100 kDa: $\alpha$-synuclein ($\alpha$-SYN), Ubiquitin, Chymotrypsin Inhibitor 2 (CI2), Calmodulin, Green Fluorescence Protein and Hisidinol Dehydrogenase. Two fluorinated amino acid analogs, 3-fluro-tyrosine (3FY), and trifluoromethyl-L-phenylalanine (tfmF) were separately incorporated into each protein. 3FY labeling was accomplished by expressing the protein in minimal media containing 3FY, phenylalanine, tryptophan, and $N$-(phosphonomethyl) glycine. tfmF labeling utilized an orthogonal tRNA synthase system to replace residues with tfmF [65].

$^{19}$F in-cell, supernatant and lysate spectra were acquired for each protein to determine quality and dispersion. All globular proteins examined yielded strong well-dispersed $^{19}$F peaks. Cell leakage was checked by comparing the supernatant spectrum with the in-cell spectrum for $^{19}$F. The results showed that $^{19}$F in-cell NMR is an easy way to monitor protein leakage. Only the globular protein CI2 showed signs of leakage.
The viability of using $^{19}$F NMR for in-cell analyses of disordered proteins was demonstrated by using $\alpha$-SYN as a model disordered protein. Because of the limited chemical shift range exhibited by disordered proteins it was shown that tfmF labeling is preferred due to the fact that any protein amino acid can be replaced with tfmF.

**In-cell NMR samples**

Proper preparation of samples for in-cell NMR spectroscopy is critical for acquiring high quality in-cell spectra. The concentration of labeled nuclei must be high enough to provide well-resolved resonances for unambiguous identification. Several methods have been employed to introduce target proteins into both prokaryotic and eukaryotic cells. These methods have their advantages and disadvantages, but all have proven effective in facilitating the acquisition of reproducible in-cell NMR spectra.

**Over-expression**

The most popular and convenient method for producing labeled targets for in-cell NMR spectroscopy involves over-expressing protein off an inducible plasmid in bacterial cells. The concentration of over-expressed protein can be controlled by using tightly regulated promoters whose level of transcription is proportional to the concentration of inducing molecule (e.g. the arabinose $P_{BAD}$ [66] and rhamnose $P_{RHA}$ [67]). Other methods for controlling the levels of intracellular concentration of over-expressed labeled protein include varying the induction time and using plasmids with greater or lesser copy number. PAGE or Western analyses are typically employed to determine the final intracellular concentration of over-expressed target protein. Eukaryotic cells have been used to produce protein, but uniform high level isotopic labeling in these cells is fraught with difficulties, rendering a homogeneous sample difficult to obtain. Over-expression has the advantage of creating the target protein within the cellular milieu in which it will be studied; no exogenous translocations or extracellular manipulations are involved, making
it the simplest and most straightforward method to generate samples for in-cell NMR spectroscopy.

**Microinjection**

Microinjecting labeled protein into cells has advantages over endogenously over-expressing labeled protein: The concentration of labeled protein can be accurately and reproducibly controlled and the background generated by over-expression is eliminated, allowing for in-cell sample preparations with an effectively zero background. The only significant contribution to background arises from the natural abundance of $^{13}$C (1.1%). The primary disadvantages are that only large, easily manipulated cells, such as *X. laevis* oocytes, are amenable to this procedure, the process is tedious and there is an inherent variability in the oocytes.

Selenko *et al.* [68] used a robotic microinjection device to administer precise quantities of streptococcal protein G B1 domain (GB1) [69] into *X. laevis* oocytes [70]. GB1 was uniformly $[^{15}$N] labeled in and purified from *E. coli*. As a prelude to in-cell NMR spectroscopy, the initial NMR experiments were performed using *X. laevis* egg extracts as a model system to mimic the intracellular milieu. Results showed that intracellular components do not appear to affect the folded state of the protein domain and revealed no intracellular binding partners. The spectrum of $^{15}$N-GB1 obtained in extracts was virtually identical to that obtained *in vitro* using purified protein over the concentration range examined.

Next, *X. laevis* oocytes were injected with purified $^{15}$N-GB1 to intracellular concentrations ranging from 50-500 μM. Intracellular $^{15}$N-GB1 produced high quality 2D spectra, generally matching peaks seen in spectra acquired *in vitro* by using purified protein (Fig. 2). Some peaks displayed a distorted, split profile that was attributed to different intracellular environments. This observation was substantiated by dissecting single oocyte nuclei and demonstrating that GB1 was found in both the cytoplasm and nucleus.
Figure 2.0. GB1 in *Xenopus* oocytes. Typical in-cell NMR sample of 200 oocytes in a Shigemi NMR tube and overlay of the 2D $^1$H($^{15}$N) HSQC spectra of purified GB1 (0.5 mM, black) and of GB1 in *Xenopus* oocytes (50 mM intracellular concentration, red). The graph at the bottom left depicts intensity ratios (buffer over oocytes) for all GB1 residues at the indicated sample concentrations. Reproduced from Selenko *et al*. [68].

There were no discernable differences in the quality of the resulting NMR spectra acquired under identical conditions of temperature and data acquisition time, however, oocytes that had been automatically injected were found to be more viable than those that were manually injected. Control experiments showed no leakage of labeled protein from the cells. This work demonstrated the feasibility of performing in-cell NMR spectroscopy in *X. laevis* oocytes and the use of *Xenopus* egg extracts to optimize conditions for high-resolution NMR spectroscopy in eukaryotic cells.

Bodart *et al*. [71] studied the neuronal protein, Tau, inside *X. laevis* oocytes, an environment in which it is not normally found [72]. Tau interacts with the microtubular network present in oocytes, thus providing an opportunity to perform in-cell NMR spectroscopy on a target protein in the bound state. Tau was uniformly $[^1]U$-$^{15}$N labeled in and purified from bacteria [73, 74].
The protein was microinjected into X. laevis oocytes to a final concentration of ~5 µM, which is close to physiological levels. The resulting $^1$H($^{15}$N)-HSQC spectrum was similar to that of a purified sample obtained in vitro [75], but the cross peaks were broadened and many, attributed to free Tau, were missing. After mechanically lysing the cells and re-acquiring the spectrum, the peak intensity increased but still lacked the peaks associated with free Tau. Instead the spectrum resembled that of Tau bound to Tubulin [75]. Therefore, the peak broadening, which leads to lower resolution spectra, is likely due to Tau interacting with Tubulin and possibly other proteins present in the oocyte.

Bodart et al. [71] examined methods to optimize the reproducibility of spectra acquired from different in-cell preparations and by varying just one acquisition parameter. They concluded that while the overall spectral profile is improved by injecting the same solution into the same injection site on the oocyte and selecting oocytes in a qualitatively reproducible manner, the physiological state of the individual oocytes likely contributes to variations in signal strength, and therefore collecting a series of spectra on different samples will not be straightforward.

The results show that Tau can be studied in cells at an intracellular concentration of ~5 µM, extending the lower limit of concentration that can be studied by using in-cell NMR spectroscopy, but many of the resonances are broadened due to protein-protein interactions, yielding low resolution data.

**Cell penetrating peptides**

Inomata et al. [76] utilized a heretofore untested procedure for introducing isotopically labeled proteins into eukaryotic cells by using cell penetrating peptides [77-79] to transduce human HeLa cells. Uniformly $[^{15}$N] labeled protein was covalently tagged or conjugated with a cell-penetrating peptide derived from the Tat protein of HIV-1 (CPP$_{Tat}$) [80], and incubated with
human HeLa cells and pyrenebutyrate. Pyrenebutyrate mediates the translocation of CPP-tagged proteins into the cytosol [81].

Experiments were performed using a Ubiquitin (Ub-3A) that was mutated at three sites (L8A, I44A, V70A) [82] to preclude binding with ubiquitin-interacting proteins (UIPs). The [$U^-,^{15}N$] labeled Ub-3A contained a C-terminal CPP fusion (Ub-3A-CPP$_{Tat}$). Following transduction, a well-resolved $^1H(^{15}N)$ HSQC spectrum was observed. The in-cell spectrum lacked the cross peak corresponding to the C-terminal CPP residue and showed an intense signal corresponding to the C-terminal glycine (G76) of Ub-3A. Control experiments showed that Ub-3A-CPP$_{Tat}$ was cleaved between G76 and D77, presumably by endogenous ubiquitin-specific C-terminal proteases (DUBs) [83] to yield free Ub-3A (Fig 2.1).

![Figure 2.1 Ubiquitin in HeLa cells.](image)

(a) In-cell $^1H(^{15}N)$ HSQC spectrum of Ub-3A. The C-terminal G76 is indicated. (b) In vitro reference spectrum of [$U^-,^{15}N$] Ub-3A-CPP$_{Tat}$ fusion protein. The C-terminal R88 at the end of CPP$_{Tat}$ is indicated. Reproduced from Inomata et al. [76].
Cleavage of the peptide tag is not only desirable, but necessary. It is well known that CPPs aggregate with many cytosolic components, including the inner plasma membrane [79]. Proteins bound to CPPs also form aggregates, producing broad, overlapping signals in NMR spectra. Experiments performed using an uncleavable CPP demonstrated that cleavage is essential for a well-resolved spectrum of the target protein. In addition, cleaved proteins exhibited a uniform intracellular distribution, whereas, CPP-tagged proteins are heterogeneously dispersed throughout the cytosol. CPP cleavage is therefore essential for uniform protein distribution.

The concentration of transduced Ub-3A-CPP\textsubscript{Tat} was estimated to be 20-30 µM in the cells, about twice the physiological concentration. Cell leakage was negligible; cell viability and membrane integrity testing indicated that no significant toxicity was associated with pyrenebutyrate treatment. Ub-3A-CPP\textsubscript{Tat} was also transduced into monkey COS-7 cells, demonstrating the versatility of this method.

The in-cell spectrum of wild-type Ubiquitin showed extensive peak broadening relative to the in-cell spectrum of Ub-3A and included the G76 cross-peak observed for cleaved Ub-3A. This indicated that at least some of the wild-type Ubiquitin existed in a C-terminally unconjugated state. The difference in peak intensity between the two spectra likely reflect the interaction of wild-type Ubiquitin with endogenous proteins, since the mutated residues prevented binding with UIPs. A similar effect was seen in in-cell NMR experiments performed by using \textit{X. laevis} oocytes [84].

Other methods of CPP-linked transduction were tested. CPPs linked to cargo proteins by using disulfide bonds [85] are cleaved in the cytosol by autonomous reduction. The $^1\text{H}\{^{15}\text{N}\}$ HSQC spectrum of the B1 domain of streptococcal protein G (GB1) conjugated to CPP\textsubscript{Tat} and transduced into HeLa cells was well-resolved and virtually identical to that of a spectrum acquired \textit{in vitro}. Similar results were obtained using Ub-3A.
A final delivery method was used to demonstrate the feasibility of studying drug-protein interactions. A $[U,-^{15}N]$ labeled fusion protein consisting of Ub containing an N-terminal CPP$_{Tat}$ and C-terminal FKBP-12 domain was transduced into HeLa cells. Cleavage was predicted to be mediated by DUBs, which would release free FKBP12 into the cytosol. The resulting in-cell $^1$H$^{15}$N correlation spectrum of FKBP12 was identical to the reference \textit{in vitro} spectrum, whereas the contribution of CPP$_{Tat}$-Ub to the spectrum was minimal. It was concluded that DUB-mediated cleavage released free FKBP12 to yield an analyzable in-cell NMR spectrum while CPP$_{Tat}$-Ub underwent CPP-mediated aggregation within the cell.

The interaction of free FKBP12 with two immunosuppressant drugs, FK506 and rapamycin, were examined. Transduced HeLa cells were incubated with either FK506 or rapamycin. The changes observed in the in-cell $^1$H$^{15}$N correlation spectra obtained for both drugs were distinct from one another but consistent with the reference spectra, acquired \textit{in vitro}, of FKBP12 complexed with each drug. The results showed that exogenously administered drugs entered the cells and formed specific complexes with FKBP12.

This work demonstrated that the technique of using cell penetrating peptides to deliver sufficient concentrations of isotopically labeled proteins into eukaryotic cells to perform high resolution in-cell NMR spectroscopy is effective for studying in-cell protein dynamics, protein-protein and protein-drug interactions.

\textit{Molecular crowding}

Much concern has been generated over the physiological relevance of studying biological macromolecules by performing NMR on living cells. These concerns have focused on the colligative properties of the cellular interior, specifically viscosity and the presence of high concentrations of endogenous and over-expressed macromolecules, and the extent to which these characteristics of the cytosol will affect the resulting in-cell NMR spectra. To examine the effects
of intracellular composition on the structure of in-cell target molecules, two classes of solutes are generally considered: Large macromolecular species and small solvent-like molecules. The first is expected to affect the diffusion and tumbling rates of the target due to frictional interactions, while the second may be responsible for altering the structure of the target through excluded volume effects.

Macromolecular crowding and solvent exclusion result in higher effective concentrations of macromolecules due to the reduction in water concentration. This can result in an increase in both specific and non-specific association reactions, including aggregation, which tend to broaden in-cell NMR signals. It can also promote folding; an inherently disordered protein lacking significant biological activity can be partially or completely folded under crowding conditions to generate a biologically active molecule\[86\]. Indeed, this may be a mechanism of regulating activity in species where, in the absence of crowding, no biological activity is present. Finally, the reduced concentration of water can alter structured water, which, in turn, can alter protein conformation and biological activity \[87\].

\textit{In vitro}, the addition of co-solvents such as glycerol, or large molecular species such as bovine serum albumin (BSA) or polyvinyl pyrrolidone (PVP) are used to mimic the bulk solvent properties of the cytosol. While these species are ideally inert, one must evaluate the ability of these substances to interact with the targets and therefore interfere with the property or process being examined. Another method to mimic cytosolic conditions uses concentrated cell extracts or lysates. The problem here is the lack of cellular regulatory mechanisms that may be required for normal activity.

Selenko \textit{et al.} [68] acquired the in-cell NMR spectrum of streptococcal protein G B1 domain (GB1) and found that the signal intensity was reduced \(~5\)-fold and proton peak widths increased by \(~2.5\)-fold relative to an equimolar \textit{in vitro} spectrum of purified GB1. Analyses of 2D
correlation spectra allowed the global structure to be deduced and the differential line broadening of individual residues to be evaluated. Hydrogen-bonded amide groups in secondary structural elements showed greater peak broadening than those in unstructured regions. This observation implies that the cellular environment affects the relaxation properties of hydrogen-bonded amide groups more strongly than those that are exposed to solvent and is in keeping with the idea that macromolecular crowding slows overall tumbling rates while solvent-exposed groups experience greater degrees of freedom. [88]

To determine whether these observations were due to macromolecular crowding, comparative analyses were performed in vitro by acquiring NMR spectra of GB1 in 150-300 mg/mL BSA. At the lowest BSA concentrations, the in vitro and in-cell spectra were virtually indistinguishable. At higher concentrations, a reduction in signal intensity and corresponding increase in proton peak widths was observed. The results suggested that properties of the intracellular environment can be simulated by using artificially crowded solutions.

To study how molecular crowding affects protein diffusion in response to recombinant protein over-expression, Slade et al. [89] used fluorescence recovery after photobleaching (FRAP) to monitor the diffusion of Green Fluorescent Protein (GFP) in E. coli. In FRAP, a small area/volume of a cell containing fluorescent molecules is bleached using a laser and fluorescence recovery is monitored as unbleached molecules diffuse back into the bleached region. Slade et al. [89] used four different proteins, which covered a range of sizes and shapes: Tau-40 (45 kDa), Maltose binding protein (42 kDa), Calmodulin (17 kDa) and α-Synuclein (14 kDa). In each case GFP was first expressed, followed by separate expression of each of the four proteins. The levels of GFP in each experiment were constant as monitored by SDS-PAGE. It was calculated that each co-expressed protein accounted for approximately 15% of the total intracellular protein. Regardless of the size and shape of each co-expressed protein, the diffusion rate of GFP showed no variation. These findings suggest that molecular crowding due to protein over-expression does
not appreciably affect intracellular protein diffusion in bacteria and that in-cell NMR spectroscopy is an effective method to study proteins in a biologically relevant environment.

**Cell viability**

In-cell NMR yields lower quality spectra than that obtained using highly purified samples typically due to low signal to noise ratios. This shortfall can be compensated for by increasing the intracellular concentration of the target, by increasing the number of scans per experiment and by using a ultrasensitive cryoprobe. Stabilizing the cells by using known cell protectants, such as glycerol or sucrose [90], or alginate encapsulation [91], may also extend the in-cell NMR acquisition time. Finally, by using transverse optimized spectroscopy (TROSY-HSQC) for high molecular weight complexes that have slow tumbling times, the sensitivity of the resulting in-cell spectrum may be increased.

Another important factor that limits the acquisition time of in-cell NMR experiments is cell lysis, which results in leakage of labeled target from the cells [90]. Xie et al. [35] added glycerol (to 10%), to samples as a cryoprotectant for prolonged storage of the cells at -80 °C. Adding glycerol to the NMR buffer maintains the viability of *E. coli* cells at room temperature for more than four hours at densities sufficient to obtain in-cell NMR spectra and to minimize cell lysis. As a result, the NMR signal from the \([U-, ^{15}\text{N}]\)-protein in the supernatant did not exceed the noise level.

To minimize cellular degradation during their study of the Tau protein, Bodart *et al.* [71] suspended *Xenopus* oocytes in a 20% Ficoll solution, which allowed the cells to remain stable overnight [92]. Control experiments showed no Tau in the extracellular medium. This work demonstrated a method for extending the lifetimes of cells in the NMR tube, a critical prerequisite for collecting data at physiological concentrations.
Another method to assess bacterial cell viability is by using a standard colony plating test. In this test, the number of colonies grown on antibiotic selection plates inoculated with the in-cell sample before NMR spectroscopy is compared with the number of colonies grown using cells plated after acquiring experimental data. Colonies are counted by using a molecular imager. The cells are considered to be viable if the number of colonies on the plates are within 10% of each other [93].

In-cell NMR is an effective tool to study proteins in their native environment. However a consequence of this technique is the issue of protein leakage which can lead to sharp signals from the protein molecules in the less viscous extracellular medium, thereby masking the broader signals obtained from protein in the cytosol. To ensure that the NMR signal arose from proteins located inside the cell, Li et al. [91] collected an HSQC spectrum of in-cell sample supernatant immediately after preparing the cells and immediately after acquiring the spectrum. A comparison between the spectra obtained using two proteins, the disordered, 14 kDa α-synuclein (αSN), and the 7 kDa globular protein chymotrypsin inhibitor-2 (CI2) showed that CI2 leaks from the cell, ~20% is lost as confirmed by SDS-PAGE analysis, and αSN remained in-cell with only small metabolites found in the spectrum of the supernatant. To improve the in-cell spectrum of αSN Li et al. [91] used an alginate encapsulation method [94] to stabilize the cells, which, in turn, yielded a clean HSQC for αSN. Encapsulating CI2 resulted in no in-cell NMR signal despite an increase in acquisition time.

*Rotational diffusion in the cell*

The disappearance of the CI2 signal was reasoned to be due to altered dynamics induced by either CI2 binding non-specifically to larger species in the cell or from the increased viscosity of the intracellular medium; both processes; would lead to a decrease in the rotational diffusion of
the protein. The rotational diffusion of proteins in solution affects the longitudinal (R1) and transverse (R2) relaxation rates of NMR active nuclei. Longitudinal relaxation rates determine how fast nuclear magnetization comes to equilibrium, while transverse relaxation rates determine the linewidth of the NMR signal. For the $^1\text{H}^{15}\text{N}$-HSQC experiment, if proton and nitrogen transverse relaxation rates are too fast the resonances tend to be too broad to detect.

The extent of non-specific interactions with cellular components depends on the nature of the protein, the extent of protein over-expression and cell growth conditions, all of which can be modified as necessary to improve the in-cell NMR signal. To test how R1 and R2 vary with viscosity under the conditions of macromolecular crowding, Li et al. [91] used poly(vinyl pyrrolidone (PVP) as a macromolecular crowding agent. R1 changed very little between buffer and PVP for all proteins tested. For the intrinsically disordered αSN protein there was a 1.5 to 6-fold increase in R2 measured in PVP compared to buffer, while the values for CI2 increased 3 to 40-fold. R1 and R2 measurements made using different solvents indicated that the globular protein CI2 is more sensitive to viscosity than αSN. The cause for the differences in relaxation rates may be linked to local and global motions within each protein. CI2 is more sensitive to global motions due to its rigid structure as opposed to αSN, which is unstructured and influenced more by local rotational motions. The findings show in-cell NMR is more well-suited to study disordered proteins than globular ones.

To monitor the effect of PVP on the translational and rotational diffusion of the globular protein CI2 Li et al [95] used NMR spectroscopy. Translational diffusion was monitored by using a heteronuclear stimulated gradient echo sequence [96] and rotational diffusion was assessed from the $^{15}\text{N}$ R2/R1 ratio [97]. According to the classical description of diffusion [98], both translational and rotational diffusion are inversely proportional to the viscosity. This relationship breaks down when PVP is used as a macromolecular crowding agent. Rotational diffusion was affected less than the translational diffusion. This effect was explained by the fact
that large crowding molecules do not make extensive contact with the protein surface due to steric effects and thus are not as effective as small molecules, such as glycerol, in decreasing protein rotational diffusion. Inside the cell, macromolecular crowding is provided by large biological molecules, which will have a similar effect on rotational diffusion, thus improving the sensitivity of in-cell NMR spectroscopy.

Applications

Much of the early work in the field of in-cell NMR utilized E. coli as the host cell. E. coli are easy to handle and grow very rapidly. Proteins are uniformly labeled with NMR-active isotopes, primarily $^{13}$C and $^{15}$N, and over-expressed to high enough intracellular levels to yield high quality HSQC spectra with little or no interfering background. Furthermore, the ability to selectively label proteins and grow cells in D$_2$O-based medium provides the capability to study high molecular weight proteins at physiological conditions. However, while bacterial cells are ideal for simulating an intracellular environment for studying eukaryotic proteins, in that they present the lowest potential for intrinsic binding partners, they lack the inherent ability to affect post-translational modifications and provide compartmentalization for selective activity.

In-cell NMR spectroscopy in eukaryotic cells has been limited to very large and mechanically manipulable cells, such as Xenopus laevis oocytes. Studies employing eukaryotic cells are hampered by the difficulties in over-expressing isotopically labeled proteins inside the cell, the inability to regulate post-translational modification activities and the inefficient delivery of labeled proteins to the cytosol. Initially isotope-labeled molecules were microinjected into the cytosol of these cells. This delivery method was improved by using active transport of isotope-labeled molecules into the cell by linking them to cell penetrating peptides. In last few years
much progress has been made to extend in-cell NMR to many cell types including primary cells and to take advantage of endogenous biochemical activity present in these cells to identify protein-protein and protein-DNA interactions, and sites and mechanisms of post-translational modifications.

**Structure determination**

The ability to determine high resolution 3-dimensional structures of proteins is one of the prime features of NMR spectroscopy. To accomplish this by acquiring spectra from living cells represents the one of the next great steps in the evolution of structural biochemistry.

**De Novo 3D Protein Structure Determination**

The low sensitivity of protein NMR requires very long sampling times: Typical NMR experiments collect data for 1-2 days. Without fresh supplies of nutrients and gases, along with waste removal, *E. coli* cells cannot survive for that long. To minimize the sampling time, Sakakibara *et al.* [99] employed a novel non-linear sampling scheme [18-20] to solve the structure *de novo*, of TTHA1718, a 66 amino acid, putative heavy-metal binding protein from *T. thermophilus*. By combining this scheme with maximum entropy processing [100, 101] the time required to collect sufficient multidimensional data was reduced to several hours.

TTHA1718 was overexpressed in uniform isotopic-labeling medium from *E. coli* to a concentration of 3-4 mM. Control experiments demonstrated that the NMR signal originated from protein within the cells. 63 out of 66 backbone resonances were assigned by using six 3D triple-resonance experiments with samples prepared freshly for each experiment. 86% of Hα, 71% of Hβ and 34% of the aliphatic $^1$H/$^{13}$C side chain resonances were identified. *In vitro* NMR performed on purified protein was used to compare to in-cell assignments. Colony plating tests showed that after 6 hours of data acquisition the viability of the bacteria was 85%.
To improve the signal to noise ratio, each 3D experiment was performed several times and the data sets were combined to enhance the protein signal. The 3D experiments were bracketed by a 2D $^1$H-$^{15}$N-HSQC experiment, which acted as a control to ensure that only data collected from intact cells were included in the combined data set. Each control spectrum following a 3D experiment was compared against the original reference HSQC spectra collected at the beginning of the run. If the control spectrum was significantly different from the reference spectrum, the 3D spectrum that was collected prior to the control spectrum was not added to the composite data.

Side chain methyl groups of Ala, Leu and Val were selectively $^{13}$C-labeled allowing 78% of their side chain resonances to be assigned. In addition, out of a possible total of 148 NOEs involving methyl groups 69 of 89 long range NOEs were assigned. NOE distance restraints, backbone torsion restraints and restraints for hydrogen bonds were incorporated into the calculation of a 3D structure for TTHA1718 using CYANA [102, 103]. The final calculated structure and was similar to the structure determined in vitro for purified TTHA1718, and had a root mean squared deviation (rmsd) of 0.96 Å and a backbone rmsd of 1.16 Å.

This study demonstrates the feasibility of determining high resolution 3D structures of proteins in living bacterial cells. The results were possible because of innovations that allowed rapid data collection and unambiguous identification of long range NOE interactions based on selective labeling of methyl groups. Furthermore, when TTHA1718 was over-expressed to 1.2-1.5 mM and 0.6-0.8 mM, 74% and 61% of the respective NOE cross-peaks used to calculate the structure were observed. This suggests that in-cell structure determination in eukaryotic cells may be possible since labeled proteins can be introduced into Xenopus oocytes at concentrations up to 0.7 mM [68]. Eukaryotic cell stability is still a limiting factor and new methods of either NMR technology and/or sample preparation are required before in-cell NMR protein structure determination in eukaryotic cells becomes possible.
Protein-protein interactions

Protein-protein interactions are the mainstay of biological activity. One of the most difficult challenges in studying protein-protein interactions in cells arises from the fact that proteins bind many types of ligands, from small effectors molecules like protons and ions to large macromolecules such as nucleic acid polymers and other proteins. The range of affinity with which these ligands are bound varies over a broad range of concentrations, offering the possibility of unrestricted protein interactions. To further dissect the differences between in-cell and in vitro studies of protein-protein interactions, we can divide the interactions between expressed proteins and other intracellular molecules into two parts: Specific interactions with affinities lower than 10 uM and non-specific interactions with affinities much greater than 10 uM.

Interactor proteins have to be present in a stoichiometric ratio with the target protein, so that the bulk of the labeled population is bound, to detect specific interactions by using in-cell NMR spectroscopy. Typically, target proteins are over-expressed to concentrations 10 to 100-fold greater than their physiological concentrations, which results in a large population of free protein and no changes in the NMR spectrum. Thus, we can largely ignore the contribution of undesired specific interactions between the target protein and intracellular binding partners to the resulting in-cell NMR spectrum. By using selectively inducible over-expression, specific interactor proteins can be introduced later at concentrations that are comparable to that of the labeled protein, thereby giving rise to specific interactions inside the cell that are detectable by using NMR.

Non-specific interactions between labeled proteins and intracellular molecules are omnipresent and establish the proper physiological environment for the labeled proteins that uniquely distinguishes in-cell NMR from in vitro techniques. This is best evidenced by small differences in the NMR spectra and solution structures of proteins measured in cells and in vitro
The concentrations of molecules inside eukaryotic and prokaryotic cells are very similar to each other; as a result we do not expect that the contribution of non-specific interactions will be dramatically different in these two cases and that the in-cell NMR spectra of cytosolic proteins acquired in either cell type will also be very similar.

**Structural interactions**

Burz *et al.* [104, 105] developed an in-cell NMR-based method for mapping the structural interactions (STINT-NMR) that underlie protein-protein complex formation. The method entails sequentially expressing two (or more) proteins within a single bacterial cell in a time-controlled manner [106] and monitoring their interactions using in-cell NMR spectroscopy [58]. The resulting NMR data provide a complete titration of the interaction and define structural details of the interacting surfaces at atomic resolution. Unlike the case when interacting proteins are simultaneously over-expressed in the labeled medium, in STINT-NMR the spectral complexity is minimized because only the target protein is labeled with NMR-active nuclei, which leaves the interactor protein(s) cryptic.

The target protein, whose NMR structure must be known, is first overexpressed on uniformly labeled ($U^{-15}N$) medium to yield a high-resolution, isotope-edited heteronuclear single quantum coherence ($^{15}N$-HSQC) backbone spectrum of the target protein inside the bacterial cells. The growth medium is changed, and the unlabeled interactor is overexpressed. As the interactor binds to the target, the HSQC spectrum of the target changes to reflect the different chemical environment for residues that have been affected by the binding interaction. The corresponding changes in the peak widths and chemical shifts of the target protein resonances define the interface between the protein and its ligand (interactor), providing atomic resolution information on the interactions.
Changes in chemical shifts and differential broadening of some assigned peaks may be more widespread, however, reflecting rearrangements of secondary structural elements, or a global or allosteric change in the conformation of the target. To assess accurately the changes in the NMR spectrum of a target molecule upon complexation with an interactor molecule, it is imperative that the resonance assignments of the target protein be known beforehand, and that the target is stable and well behaved in the absence of the interactor protein.

The efficacy of this method was demonstrated by over-expressing \([U^{-15}N]\) Ubiquitin followed by over-expression of either one of two Ubiquitin ligands containing the Ubiquitin Interacting Motif (UIM) [107]: a 28-amino acid peptide from Ataxin 3 (AUIM; ~4 kDa) or the Signal-Transducing Adapter Molecule (STAM2 [108]; ~50 kDa) in unlabeled medium. AUIM binds Ubiquitin in vitro with ~230 \(\mu\)M affinity [109, 110] and STAM2, which contains two Ubiquitin interacting surfaces, binds with a higher (~10 \(\mu\)M) overall affinity. These two systems simulated a range of protein-protein interaction affinities and molecular weights.

The \(^1\)H\(\{^{15}N\}\)-HSQC spectrum of Ubiquitin, which was maintained at a single concentration, changed as the concentration of interactor was increased (Fig. 2.2). The chemical shift changes were mapped onto the three-dimensional structure of Ubiquitin. Peaks that underwent substantial (>0.1 ppm) chemical shifts came exclusively from surface-exposed amides. Control experiments using a mutant AUIM, which does not bind to Ubiquitin, demonstrated that the chemical shifts changes result from specific interactions and not merely over-expression of the interacting molecule.
Figure 2.2. Ubiquitin-ligand complexes in *E. coli*. (a) Overlay of $^1$H$^{15}$N HSQC spectra of *E. coli* after 3 hour over-expression of $[U^{15}$N] Ubiquitin and 0 h (black), 2h (red) and 3h (blue) over-expression of AUIM. Individual peaks exhibiting large chemical shifts are labeled with corresponding assignments. (b) Overlay of $^1$H$^{15}$N HSQC spectra of free $[U^{15}$N] Ubiquitin (black) and $[U^{15}$N] Ubiquitin-AUIM complexes at a molar ratio of 1:1 (red) and 1:2 (blue). Reproduced from Burz *et al.* [104, 105].

The method is limited primarily by the concentration level of interacting protein that can be achieved and can be used to study interacting proteins whose structure is unknown, since only one of the interacting species is labeled. It can be used to study proteins that are difficult to purify or are proteolytically labile, since there is no need for purification. Another limiting consideration is the integrity of the interacting proteins. For example, over-expressed proteins may degrade into components that bind non-specifically to each other, thereby presenting multiple and/or incorrect interaction surfaces. For this reason, the target should be stable over the
course of the experiment. Should sample stability become questionable, SDS-PAGE and Western blots can be used to assess the extent of degradation.

**Drug screening**

Xie *et al.* [111] developed an in-cell NMR spectroscopy-based screening procedure, Screening of small Molecule Interactor Library (SMILI-NMR), to rapidly screen for compounds capable of disrupting and enhancing specific interactions between two or more components of a biomolecular complex. SMILI-NMR utilizes STINT-NMR [104, 105] technology to produce biomolecular complexes inside the cell in which one of the constituent proteins is uniformly \([U,^{15}N]\) labeled with NMR-active nuclei. By monitoring the in-cell NMR spectrum of the labeled protein, the formation of high-affinity ternary complexes is observed. STINT-NMR analyses identify changes in structure induced by binding of a small drug-like molecule that disrupts or enhances the stability of the complex and reveal biologically relevant, functional interaction surfaces. In this way, STINT-NMR serves as a direct assay for protein-drug interactions, identifying small drug-like molecules that bind to this surface and facilitating high throughput screening.

A system of two interacting proteins, FKBP and FRB, was used as a model to show the effectiveness of SMILI-NMR to screen small molecules that facilitate heterodimerization. The FKBP-FRB interaction constitutes one of the immunomodulatory systems in mammalian cells [112]. In complex with rapamycin, a macrolide antifungal antibiotic currently undergoing clinical trials for a variety of cancer treatments [113, 114], FKBP binds to FRB. When \([U,^{15}N]\)-FKBP was over-expressed in bacterial cells, the \(^{1}H\{^{15}N\}\)-HSQC spectrum showed no well-resolved peaks, implying that the single species was part of a large complex and therefore invisible to NMR. When unlabeled FRB was then over-expressed in the same cells, the NMR spectrum of FKBP became evident but only at the highest FRB concentrations, indicating the
formation of a complex. Similar results were obtained when [U-,\textsuperscript{15}N]-FRB and unlabeled FKBP were sequentially over-expressed in the same cells. These observations demonstrated that creating a proper protein complex was necessary for high-resolution studies.

Adding rapamycin to the cell suspension resulted in visible changes in the $^1\text{H}\{^{15}\text{N}\}$-HSQC-spectrum of [U-,\textsuperscript{15}N]-FKBP indicating the formation of a high-affinity ternary complex between FKBP-rapamycin and FRB (Fig.2.3). Adding rapamycin to cells overexpressing labeled FKBP in the absence of FRB or labeled FRB in the absence of FKBP did not produce an NMR spectrum. In each case, co-expression of the second protein was required to generate an in-cell NMR spectrum.
Figure 2.3. Ternary FKBPFRB-ligand complexes in E. coli.  (a) Overlay of $^1$H{$^{15}$N} HSQC spectra of E. coli after 4 h of over-expression of [U-, $^{15}$N]-FKBP and 4 h sequential over-expression of FRB in the absence (black) and presence (red) of 150 µM rapamycin.  (b) $^1$H{$^{15}$N} HSQC spectrum of E. coli after 4 h of over-expression of [U-, $^{15}$N]-FKBP and 4 h sequential over-expression of FRB in the presence (red) of 5 mM A-E.  (Bottom) Yeast assay for biological activity of the dipeptide, A-E.  Isogenic haploid yeast strains (S. cerevisiae) that express (FPR1) or lack (frp1) the FKBPFKBP proline isomerase were grown for 3 days on YPD medium.  (c) Control plate.  (d) 100 µM rapamycin.  Expression of FKBPFKBP allows the formation of a toxic FKBPFKBP-rapamycin-FRB bio-complex.  (e) 5 mM A-E.  Results indicate that the dipeptide induces formation of a bio-complex similar to that induced by rapamycin.  The reduced growth in the frp1 strain likely reflects the weaker affinity of A-E for FKBPFKBP. Reproduced from Xie et al. [111].
A dipeptide chemical library [115] composed of 17x17 dipeptides [115, 116] was chosen to provide a collection of compounds that are capable of interacting with the target molecule at a detectable level. These compounds provided suitable starting points for subsequent optimization into credible drug candidates [117] and were considered as potential sources of novel lead structures. Dipeptides have been shown [116, 118] to be an excellent starting point for drug design since 1) they can be prepared at low cost, 2) a library containing only 289 compounds can provide a data set that spans a broad spectrum of physicochemical properties, 3) no deconvolution is required to identify the lead structures [117], and 4) dipeptides can pass through the cell membrane and interact with the target protein directly in vivo [119, 120].

A standard procedure called the matrix method, in which compounds located in one row or one column of a matrix plate were mixed and tested, was used to screen the library. Individual mixtures were examined for their ability to change the in-cell NMR spectrum of the FKBP-FRB biocomplex. Samples exhibiting similar spectral changes, located at the intersection of rows and columns, were used in the second round of screening to deconvolute and validate the initial findings. In this way, a matrix of 289 (17x17) compounds were screened by examining 34 (17+17) samples.

Most of the dipeptide mixtures showed no interaction with the FKBP target. The mixture of A-X (where X is all possible amino acids) elicited a totally different spectrum from that of the rapamycin-induced ternary complex (Fig.2.3). Formation of the dipeptide-induced complex resulted in extreme broadening and the disappearance of some peaks in the NMR spectrum at the highest concentration used. The mixtures of D-X, T-X, L-X, X-E, X-I, X-A, X-T also caused a similar broadening of the spectrum. Thus, the dipeptides located at the intersection of rows, A-X, D-X, T-X, L-X and columns X-E, X-I, X-A, X-T were screened in a second round and titrated into cells individually. Only A-E showed the same interaction with FKBP suggesting that A-E facilitated heterooligomerization of FKBP and FRB. Competition experiments with rapamycin
confirmed that A-E binds specifically to the FKBP-FRB complex. Further confirmation that A-E exhibited biological activity comparable to that of rapamycin was obtained by using a yeast growth assay.

SMILI-NMR provides an important means to bridge the gap between biochemical identification of small ligands capable of interfering with target biocomplexes and the biological activity resulting from the inhibition of cellular processes by these ligands. The method requires minimal sample preparation and eliminates the need for extensive protein purification. Furthermore, SMILI-NMR can be automated by making use of robotic HTS accessories available for modern NMR spectrometers, such as liquid handlers and NMR tube changers.

Metabolic processing

To investigate proteins under intracellular environmental conditions Sakai et al. [84] injected X. laevis oocytes with labeled proteins. The experiments highlight the effects of protein-protein interactions on NMR spectra and the difficulty in obtaining in-cell spectra by using native proteins. They also report the first real-time intracellular observations of enzymatic activity.

Uniformly \([U-,^{15}N]\) labeled Ubiquitin and its derivatives were expressed and purified in bacteria [121]. The purified protein was microinjected into X. laevis oocytes to a maximum final concentration of \(~100 \mu M\). The magnitude of the signal for the \(1H\{^{15}N\}\)-HSQC spectrum of wild-type Ubiquitin injected into the oocytes was very weak. Control experiments verified that the Ubiquitin spectrum originated from an intracellular environment and not from labeled protein that had leaked from the cells into the surrounding media. To counteract this non-desirable interference, cells were washed prior to loading into the NMR tube. A series of \(1H\{^{15}N\}\)-HSQC spectra were obtained and between each measurement the cells were removed from the NMR tube and washed. There was a 13% loss in peak intensities after each wash indicating that the observed NMR signal is mostly derived from intracellular protein.
Experiments performed using a series of mutant Ubiquitins, in which residues implicated in binding ubiquitin-interacting proteins (UIPs), L8, I44 and V70 [82], were changed to alanines. The results showed that the in-cell $^1\text{H}[^{15}\text{N}]$-HSQC spectrum of Ubiquitin could be largely recovered when the UIP interface is perturbed. The perturbation disrupts protein-protein interactions, which decreases peak broadening and resolves the spectrum. Ubiquitin molecules carrying single mutations affecting the UIP binding site (L8A-D77, I44A-D77 & V70A-D77) partially restored the in-cell Ubiquitin $^1\text{H}[^{15}\text{N}]$-HSQC spectrum and the protein containing all three mutations ((L8A, I44A, V70A)-D77) dramatically improved the spectrum. Thus, the inability to resolve an in-cell spectrum for Ubiquitin results from its interacting with endogenous UIPs, preventing a sufficient in-cell concentration of free Ubiquitin for analysis by NMR spectroscopy.

The Ubiquitin used in this study contained a C-terminal D77 residue (Ub-D77), whereas mature Ubiquitin has a G76 in that location [122]. The D77 protein, therefore, mimics a Ubiquitin precursor and can act as a substrate for Ubiquitin C-terminal hydrolase (UCH). The cross peaks attributed to the G76 and D77 amide groups are missing from the in-cell spectrum of Ub-D77, and a single peak corresponding to wild-type G76 is present. This suggests that the G76-D77 bond in this mutant is cleaved in the oocyte.

By pre-injecting oocytes with ubiquitin aldehyde, which specifically inhibits UCH [122], the in-cell spectrum of $^{15}\text{N}$-Ub-D77 showed G76 and D77 cross peaks and a G76 signal that was weaker than control cells pre-injected with water that displayed no G76 and D77 cross peaks. Further experiments showed a dose dependence on residual UCH activity. Thus, by pre-injecting oocytes with ubiquitin aldehyde, UCH activity was inhibited, thereby reducing and curtailing Ubiquitin processing. These observations suggest that in-cell NMR spectroscopic analyses of metabolic processing may be possible under select conditions.
Sakai et al. [56] also microinjected $U,^{15}\text{N}$ labeled Calmodulin into oocytes. The in-cell $^1\text{H}^{^{15}\text{N}}\text{-HSQC}$ spectrum of Calmodulin, acquired without Ca$^{2+}$ in the buffer, resembled that of apo-Calmodulin acquired in vitro, except the peaks were broader, indicating that the majority of the in-cell Calmodulin was Ca$^{2+}$-free. When an excess of Ca$^{2+}$ was co-injected with the protein, the overall spectrum changed and the resulting cross peaks were further broadened, precluding exact assignments of individual resonances. However, more than 10 cross peaks that are consistent with those observed for Ca$^{2+}$-bound Calmodulin were identified. Since Ca$^{2+}$-bound Calmodulin is more likely to interact with downstream effector proteins [123] than apo-Calmodulin, the reduced spectral quality observed in the presence of Ca$^{2+}$ suggests interactions between injected and endogenous proteins.

Post-translational modifications

Post-translational modifications (PTMs) of proteins such as phosphorylation, ubiquitination, sumoylation and glycation are necessary to regulate biological processes. Several studies have taken advantage of in-cell NMR spectroscopy to identify and to examine the ramifications of phosphorylated sites on proteins. The results show that post-translational modifications can be studied in eukaryotic cells where these modifications commonly occur, and in prokaryotic cells, where, by appropriately engineering the transcriptional machinery of the cells, the effects of PTM's on protein structure can be examined without competing reactions, in effect, turning the bacteria into "cellular test tubes".

Identifying phosphorylation sites

Certain patterns of protein phosphorylation contribute to different pathologies [124]. X. laevis oocytes, which contain many kinases, can potentially be used to study protein phosphorylation in vivo. Bodart et al. [61] used in-cell NMR spectroscopy to examine the phosphorylation pattern of the neuronal protein, Tau, in Xenopus oocytes. Tau is involved in
stabilizing microtubules and constitutes a major component of the intracellular tangles of neurons in the cells of Alzheimer diseased patients.

The bulk of the in-cell HSQC NMR spectrum acquired for Tau was consistent with it being bound to Tubulin. The resolution and number of the resonance peaks increased following mechanical homogenization of the sample. Some peaks were attributed to phosphorylated residues of Tau. Comparing the in-cell spectrum to an *in vitro* spectrum of Tau phosphorylated by PKA [125], indicated that the major peak observed in the in-cell spectrum was not generated by this kinase.

Another kinase present in the oocyte, Gsk3β [126], was examined to see if it was responsible for generating the major peak. Gsk3β was overexpressed, purified and used to phosphorylate PKA-phosphorylated Tau *in vitro*. The spectrum obtained from lysed cells that had been injected with PKA/Gsk3β phosphorylated Tau indicated that the major site observed in the in-cell spectrum was not generated by this kinase either. By examining the cell lysate of this sample at a lower threshold, a minor peak corresponding to phosphorylated Ser214 was identified and attributed to in-cell PKA-mediated phosphorylation. Thus, in-cell NMR spectroscopy in *Xenopus* oocytes was used to identify individual phosphorylated residues and the enzyme responsible for a specific post-translational modification. Importantly, this work opens the door for the use of specific probes to study biochemically modified sites.

*Kinetic mechanism for Casein Kinase 2*

Selenko *et al.* [127] used in-cell NMR to observe real time enzymatic reactions in *X. laevis* egg extracts and oocyte cells. The study focused on the phosphorylation of substrates at adjacent sites by Casein kinase 2 (CK2). This naturally abundant kinase is constitutively active in
eukaryotic cells. Extensive mutational analyses were performed to determine the chemical nature and location of amino acids surrounding the modification sites that were critical for the phosphorylation reactions.

Experiments were performed using a model substrate, XT\textsubscript{111,132}GB1, in which the regulatory region of the viral SV40 large T antigen, XT\textsubscript{111,132}, was fused to the B1 domain of streptococcal protein G (GB1). The GB1 domain acted as a C-terminal tag [128] and facilitated the solubility of the fusion protein. The protein was uniformly [$U$, $^{15}$N] labeled and purified from bacterial cells [68]. CK2 phosphorylates serine residues 111 and 112 in this construct, and this modification has been shown to regulate the transportation of full length protein into the nucleus [129, 130].

Initially, $^1$H($^{15}$N)-HSQC spectra, acquired \textit{in vitro} by using labeled substrate and recombinant CK2, revealed that after 60 minutes, S112 was exclusively phosphorylated as indicated by a large change in the chemical shift of the S112 resonance cross peak. The magnitude of the change was consistent with that observed between phosphorylated and unphosphorylated amino acids [131]. This modification proceeded until the peak corresponding to unphosphorylated S112 had disappeared, indicating that all of the S112 sites were phosphorylated, at which time phosphorylation of S111 commenced. After 100 minutes, a population of singly and doubly phosphorylated substrate was observed. S111 phosphorylation continued until all of the substrate had been modified at both sites (~140 minutes).

CK2 phosphorylation of XT\textsubscript{111,132}GB1 was shown to proceed in a two-step process. Since singly modified substrate was observed at intermediate times during the course of the reaction, it was rationalized that CK2 must dissociate from the mono-phosphorylated substrate prior to phosphorylating the second site. This is because the substrate-CK2 complex is too large to be detected, therefore any substrate detected must be in the free state. The ability of the
phosphorylation process to be selective enough to only phosphorylate one site was postulated as necessary to maintain the ultrasensitive nature of the signaling pathway.

The reaction was examined by using *X. laevis* egg extracts [132] and by microinjecting labeled substrate into *X. laevis* oocytes, both of which contain endogenous CK2 [133]. Changes in the cross peak resonances of the modified residues in NMR spectra acquired using egg extracts were indistinguishable from those obtained *in vitro*. In oocytes, the NMR spectral changes were also consistent with stepwise phosphorylation of S112, release of the mono-phosphorylated substrate and the phosphorylation of S111. Thus, CK2 phosphorylated XT_{111-132}GB1 via the same mechanism both *in vitro* and *in vivo*.

The study affirmed that in-cell NMR spectroscopy has advantages over traditional methods that might be used for analyzing protein phosphorylation *in vitro*. The NMR spectra obtained represented the direct detection of phosphorylated substrate residues. The signal does not depend on the location of the modified residue in the substrate or the number of substrates. The NMR process allows for the phosphorylation reaction to be studied over a period of time and the magnitude of the NMR signal provides information on the relative population of the different states of phosphorylation. This work demonstrated that other post-translational modifications that produce distinctive NMR spectra can potentially be studied by using in-cell NMR.

*In-cell biochemistry*

Burz and Shekhtman [134] developed an in-cell methodology to introduce post-translational modifications (PTMs) onto interactor proteins in bacterial cells and identify the changes in the interaction surface of a target protein when bound to the biochemically modified interactors. Modifying the interactor protein causes structural changes that manifest on the interacting surface.
of the target protein and these changes are monitored by using STINT-NMR [104, 105]. The method was used to phosphorylate tyrosine residues on the Signal-Transducing Adapter Molecule 2 (STAM2) and Hepatocyte growth factor-Regulated tyrosine kinase Substrate (Hrs), and identify changes in the interaction surface of Ubiquitin resulting from these post-translational modifications. The work demonstrated that post-translational modification of over-expressed proteins in bacterial cells can be regulated by tight temporal control over protein expression, a process dubbed in-cell biochemistry.

STAM2 and Hrs are components of an endocytic pathway present in eukaryotic cells. Each binds Ubiquitin via a Ubiquitin Interacting Motif (UIM) [135]. In addition, STAM2 has a VHS domain capable of binding Ubiquitin [110]. Evidence suggests that receptor sorting through endocytosis and subsequent degradation is controlled by ubiquitination of both the internalized receptors and components of the endocytic machinery [136].

To study Ubiquitin binding to STAM2 and Hrs, $[^{15}N]$-Ubiquitin over-expression was induced prior to or following 3 or 4 hours of over-expression of STAM2 or Hrs alone, or co-expression of both. STAM2 and Hrs were phosphorylated by inducing over-expression of the constitutively active Src-family tyrosine-kinase, Fyn, for the final 2 hours of STAM2, Hrs, or STAM2-Hrs over-expression. STAM2 and Hrs phosphorylation were confirmed by using Western blots and mass spectroscopic analyses.

The STINT-NMR spectra of Ubiquitin revealed no changes in the interaction surface when bound to non-phosphorylated or phosphorylated Hrs (Fig.2.4). The spectrum of Ubiquitin interacting with STAM2 revealed that a smaller surface is involved in the interaction with phosphorylated STAM2, corresponding to the loss of the interaction surface attributed to the VHS domain (Fig.2.4). Mutational analysis revealed that two STAM2 tyrosines, Y371 and Y374,
located in the conserved ITAM domain, were responsible for these changes. The ITAM domain has been identified as necessary for tyrosine phosphorylation of STAM2 by Jak1 [137].

Figure 2.4. Interaction surface maps of Ubiquitin-ligand complexes. Interaction surface of Ubiquitin mapped onto the three-dimensional structure of Ubiquitin (PDB code 1D3Z). Individual residues exhibiting either a chemical shift change >0.05 ppm or significant differential broadening are indicated in red. All perturbed residues lie on the Ubiquitin surface and, therefore, reflect changes in the interaction surface of the molecule rather than changes in tertiary or quaternary structure. (a) STAM2-Ubq interaction; (b) Hrs-Ubq interaction; (c) STAM2-Hrs-Ubq interaction; (d) phosphorylated STAM2-Ubq interaction (YP-STAM2); (e) phosphorylated Hrs-Ubq interaction (YP-Hrs); (f) phosphorylated STAM2-Hrs-Ubq interaction (YP-STAM2-Hrs). Ubiquitin ligands are indicated in each panel. Reproduced from Burz and Shekhtman [134]
A similar result was obtained for the interaction between Ubiquitin and the STAM2-Hrs heterodimer: Ubiquitin interacted with the phosphorylated ternary complex in much the same way that it interacted with phosphorylated STAM2 and phosphorylated Hrs, involving contact with only the UIMs of both interactor proteins (Fig. 2.4). The commensurate weakening of the binding due to the loss of the second interaction surface is consistent with the idea that phosphorylation mediates the disassembly of Ubiquitin-mediated scaffold complexes during endocytosis.

The introduction of in-cell biochemistry using STINT-NMR facilitates biochemical modification and examination of protein-protein interaction surfaces at the atomic level. The ability to control PTM's in an environment that normally lacks the ability to provide such modifications, *i.e.* bacterial cells, affords an opportunity to examine the effects of PTM's on protein structure without competing reactions. The methodology can be applied to any stable target molecule and may be extended to include other post-translational modifications.

**In-cell NMR of nucleic acids and protein-nucleic acid interactions**

An ongoing part of structural biology is to understand how genetic material is utilized during cellular processes. The extension of in-cell NMR spectroscopy to include nucleic acids and protein-nucleic acid interactions is critical to advancing this area of study. In-cell NMR studies of nucleic acids have been difficult due to the inability to over-express and label small DNA and RNA fragments inside the cell and the intrinsic instability of these fragments in a cellular milieu. Microinjecting labeled RNA and DNA fragments into *X. laevis* oocytes, which are largely devoid of DNAse and RNAse activity, circumvents these problems and permits the study of these structures by using in-cell NMR spectroscopy. To date, the study of protein-nucleic acids interactions inside cells has been limited to non-specific interactions.
Nucleic Acids

Hansel et al. [138] extended in-cell NMR spectroscopy to the study of nucleic acids (NA) by injecting uniformly double labeled ($^{13}$C and $^{15}$N), synthetic d(GCGAAGC) DNA and r(GGCACUUCGGUGCC) RNA hairpins into X. laevis oocytes. The corresponding in-cell [$^{15}$N-$^1$H]-SOFAST-HMQC, HSQC and [$^{13}$C-$^1$H]-CT-TROSY spectra showed line broadening relative to spectra acquired in vitro, effects that were attributed to the increased viscosity and heterogeneity of the oocyte. Over time, the intensity of the DNA aromatic resonance peaks decreased and new peaks emerged (Fig. 2.5). Mass spectroscopic analyses indicated that the injected DNA was degraded to mononucleotides in the oocyte. Most of the resonances of the in-cell RNA spectrum decreased over time as well, the exceptions being U6 and U7 located in the flexible loop [139, 140], but did not give rise to new peaks with chemical shifts characteristic of isolated nucleotides (Fig. 2.5). This suggested that rather than being completely degraded the RNA interacted with large cellular components.
Figure 2.5. In-cell [1H-13C]-CT-TROSY-HSQC spectra of the aromatic regions of 13C/15N-labeled d(GCGAAGC) and r(GGCACUUCGGUGCC) hairpins. (a) and (f) Schematic representations of secondary structures for d(GCGAAGC) and r(GGCACUUCGGUGCC), respectively. (b) and (g) In vitro spectra of d(GCGAAGC) and r(GGCACUUCGGUGCC). (c) and (d) In-cell NMR spectra of d(GCGAAGC) recorded approximately 5 and 19 h after microinjection of the X. laevis oocytes. (h) and (i) In-cell NMR spectra of r(GGCACUUCGGUGCC) recorded approximately 5 and 19 h after microinjection of the X. laevis oocytes. (e) and (j) In vitro spectra of d(GCGAAGC) and r(GGCACUUCGGUGCC) in cleared oocyte lysates. The signals in black (positive) and red (negative) correspond to purines and pyrimidines, respectively. The difference between the signs of the purine and the pyrimidine peaks is due to the constant-time mode of the experiment. Reproduced from Hansel et al [138].
To arrest the degradation of the DNA samples, modified substrates were prepared that replaced the first and second phosphate groups with phosphorothioate groups. The resistance of these modified DNA samples to nuclease cleavage was confirmed by the resulting in-cell NMR spectra. Modifying only the first phosphate group led to a slowdown of the degradation reaction, but did not prevent complete degradation of the samples. The ability to stabilize RNA was examined utilizing two procedures: The first employed phosphorothioate modification of the backbone, similar to the process that was used to decrease the degradation of the DNA samples; the second used substrates that had been modified by methylating the O2'-hydroxyl groups of the RNA. Both of these modifications were successful in extending the stability of the RNA samples.

Injecting synthetically prepared NAs proved to be fatal to the oocytes, likely due to low molecular weight impurities. This problem was overcome by butanol precipitating the NA preparations prior to microinjection. Analyses were performed to determine the extent of leakage from the cells. The potential for cell leakage is greater when using microinjection to introduce labeled target into the cell, due to the incision made during the injection. Results indicated that less than 5% leakage from the oocytes is occurring. This confirms that almost all of the NA signals in the NMR spectra arise from endogenous NAs. Background signals presented little or no problem for unambiguously detecting aromatic, imino and amino resonances associated with the injected NAs. Spectra attributed to aliphatic/sugar regions C2’-H2’/H2” and C5’-H5’/H5” contained significant background signals and the C3’-H3’ region of the $^1$H($^{13}$C)-HSQC spectra was dominated by a water suppression artifact. No oocyte background was observed for signals corresponding to the C1’-H1’ and C4’-H4’ regions.
Hansel *et al.* [138] also examined the feasibility of determining the conformation of injected NAs by studying a short fragment that is a constituent of telomeric DNA. Telomeric DNA forms quadruplex structures at the 3’ ends of chromosomes [141]. Quadruplex DNA is known to inhibit the activity of telomerase [142]. G-quadruplex conformations depend on specific counterions, molecular crowding and sequence composition [143-150]. Unlabeled d(G_3(TTAG_3)T) was injected into oocytes; this sequence was shown to take on a basket-type G-quadruplex [143, 151-160] conformation in solutions containing potassium ions [161]. The resulting in-cell spectra were of low resolution and sufficiently different from those obtained *in vitro* so as to preclude a structural interpretation of the spectrum. The results further demonstrate that there are differences in conformation between *in vitro* and in-cell conditions.

The cost of preparing labeled NAs is high. However, it is possible to recover a large percentage of these NAs after use. The recovery process yield is as high as 70% for the RNA hairpin. Alternatively, it may be possible to bypass the high cost of labeled NAs by working with unlabeled NAs: It was observed that background ⁴H signals in the imino region of oocyte NMR spectra are considerably less than in other regions, resulting in the ability to observe some resonances of unlabeled NAs in the cellular environment.

*Non-specific protein-DNA interactions*

In *E. coli*, expression of the Met regulon is controlled by the transcriptional repressor MetJ. The Met regulon is comprised of 7 sites located throughout the bacterial genome and contains at least twelve genes that code for proteins involved in the biosynthesis and transport of methionine [162, 163]. Each gene carries 2-5 tandem repeats of the MetJ binding site in their promoters called metboxes [164]. MetJ binds to the metboxes to repress [111, 112][165, 166] transcription within 30 minutes of a change in methionine concentration [167, 168]. Augustus *et al.* [169] demonstrated through in-cell NMR experiments that MetJ associates non-specifically with
genomic DNA and postulated that these interactions provide a rapid mechanism for specific binding of MetJ to metbox sequences to regulate methionine biosynthesis.

MetJ was uniformly \([U^-, ^{15}N]\) labeled and overexpressed in bacteria. The resulting in-cell and lysed cell \(^1\text{H}(^{15}\text{N})\)-HSQC spectra contained minimal signals. Control experiments indicated that labeled MetJ, partially purified from the lysates, was present in sufficient concentrations to produce a NMR spectrum comparable to that obtained for purified MetJ \textit{in vitro}. The absence of a NMR spectrum suggested that MetJ was associating with large molecular species inside the cell that restricted its motion, rendering it invisible to NMR spectroscopy.

Since MetJ is a DNA-binding protein and is present at a concentration that is several thousand-fold excess over the concentration of metbox sequences in the cell, non-specific binding to DNA was postulated as the reason that no NMR spectrum was obtained. To test this hypothesis, nonspecific (sonicated salmon sperm) DNA was titrated into a solution of uniformly labeled, purified \([U^-, ^{15}N]\) MetJ \textit{in vitro}. The MetJ peaks originally present in the NMR spectrum gradually disappeared as the concentration of DNA was increased. The progressive decrease, with no concomitant changes in chemical shifts, suggested that MetJ bound nonspecifically to the DNA.

The MetJ NMR spectrum was recovered using either of two methods: The first used DNase and MgSO\(_4\) to digest the DNA and release free MetJ restoring the NMR signal; the second used a short oligonucleotide containing two MetJ binding sequences to compete away the nonspecific DNA to yield an NMR spectrum of MetJ bound to the oligonucleotide. Purified \([U^-, ^{15}N]\) MetJ titrated with unlabeled cell lysate yielded results similar to those observed using labeled MetJ over-expressed in whole cells and cell lysates.
The results indicate that MetJ interacts extensively and non-specifically with genomic DNA inside bacterial cells. This observation is substantiated by the fact that MetJ can locate a target sequence faster in the presence of longer pieces of DNA [170]. The ability of MetJ to interact non-specifically with DNA, rather than relying on free diffusion, provides a mechanism to explain how MetJ locates and binds with high specificity to metbox sequences to rapidly regulate methionine biosynthesis.

Conclusions and future directions.

The application of in-cell NMR spectroscopy to cellular structural biology is still in its infancy. The first steps have been taken to demonstrate the feasibility of using in-cell NMR experiments to study proteins and nucleic acids, and their interactions in prokaryotic and eukaryotic cells. In-cell NMR spectroscopy continues to expand thanks to new procedures for introducing labeled proteins into living cells and advancements in instrumentation. Studies utilizing prokaryotic cells have become quite commonplace, whereas there are still some challenges, such as detecting labeled targets at physiological concentrations and signal broadening due to intracellular interactions, to be overcome before eukaryotic cells can be routinely used for in-cell NMR. Protein in-cell NMR has revealed how the cellular environment influences molecular structure and interactions that are at the core of metabolic and signal transduction pathways. Expanding in-cell NMR to study nucleic acid structure and dynamics will allow us to probe the central dogma of biology under physiological conditions with unprecedented resolution. Ongoing concerted efforts are still needed to realize this potential.

Acknowledgements

This work was supported by NIH R01 GM085006-01A2 grant and American Diabetes Association Career Development Award 1-06-CD-23 to A.S.
Chapter 3

Fate of Pup during *Mtb* proteasome proteolysis as studied by in-cell NMR spectroscopy.
Abstract

*Mycobacterium tuberculosis* (*Mtb*) ability to resist nitric oxide stress, from immune initiated response, relies on then bacteria’s ability to target misfolded or damaged proteins. *Mtb* uses Pup (Prokaryotic Ubiquitin like protein) to target protein for proteasome mediated degradation. Proteins that have been tagged by Pup interact with Mycobacterial ATPase (Mpa). Using STINT NMR we characterized the interactions between Pup and Mpa in-cell and show that Pup’s C-terminus is engaged by Mpa. Pups N-terminus was also shown to be engaged only in the presence of the Mpa-proteasome complex. We also demonstrate through various biochemical assays that Pup that is not deamidated is not a substrate for proteasomal degradation.

Background

*Mtb* persistently infects approximately two billion people worldwide. [171] The proliferation of *Mtb* relies on its ability to resist stress from nitric oxide and other reactive nitrogen intermediates (RNI’s) generated by the immune system’s response. [172] RNI’s are produced by macrophages to slow the growth of invading pathogens and are thought to inhibit microbial growth by damaging nucleic acids, proteins, and lipids. [173, 174] Importantly, the production of NO by nitric oxide synthase (iNOS) present in macrophages is essential to control *Mtb* growth in mice. [175] Although wild-type mice survive much longer than iNOS-deficient mice after *Mtb* infection, the bacteria are rarely eliminated from these immunocompetent animals. The bacterium primarily takes house within the macrophages where the environment has properties of low pH and is also subject to nitrosative and oxidative stress. Oxidative stress results from the reaction of NO with superoxide formed by bacterial metabolism to generate peroxynitrate. [172] It was shown that the immunity of the host via iNOS is non-sterilizing,
resulting in a lifelong infection in humans. Uncovering the underlying mechanism of immunity to host response may shed light on possible new drug targets for the treatment of Mtb.

K. Darwin, discovered two genes associated with a putative proteasome, while screening for hypersusceptibility to acidified nitrate [172] There are two enzymatic activities characteristic of the proteasome gene locus. The first is an ATPase, encoded by mpa that forms a hexameric ring with ATPase activity. The second is a putative protease encoded by prcB in the prcBA operon. [176] The prcBA genes were predicted to be essential for viability and normal growth of Mtb. [176] PrcB and PrcA proteins, which are expressed by Mtb, bind to themselves and each other and assemble into a 28-subunit cylinder with peptidolytic properties and the inhibitor profile of a bona fide proteasome. [176]

Proteasome: Background and Structure

The 20S proteasome was first discovered in 1968 by examining human erythrocyte micrographs. The micrographs revealed a complex with a cylindrical structure. This discovery was followed by the isolation of a 700 kDa protease from bovine pituitary tissue in 1980. [177] The protease was shown to contain three to five types of subunits of mass ranging from 24-28 kDa, and were able to cleave peptide substrates at hydrophobic, basic and acidic sites. Until 1989 it was thought that the 20S proteasomes were constrained to eukaryotic cells. However, in 1991 a large complex with chymotryptic activity was isolated from the archaebacterium, *Thermoplasma acidophilum*, and in 1992 the first bacterial proteasome was purified from the actinomycetale, *Rhodococcus erthropolis*. It now appears that the 20S proteasome is ubiquitous in archa. [177] It is thought that proteasomes are not required in prokaryotic cells due to other redundant proteolytic systems, however the question remains why they are still present. [172]

Proteasomes are multimeric, self-compartmentalizing proteases that appear to be universal and essential in eukaryotes where they rapidly reduce levels of regulatory proteins.
during cellular adaptations such as cell cycling, hypoxic response and inflammation. [178] Proteasomes also degrade irreversibly mis-folded, oxidized and potentially toxic proteins. In eukaryotes the 26S proteasome is composed of two functionally distinct sub-complexes: the 20S core particle (CP) and the 19S regulatory particle (RP) which mediates Ubiquitin (Ub) binding. [176] The main structural feature of the 20S CP, both in prokaryotes and eukaryotes, are the 28 subunits from which they are built, however there are several key differences. Prokaryotic proteasomes consist of 14 copies each of 2 distinct but related polypeptides, α and β, eukaryotic proteasomes are built of 2 copies each of 7 distinct α- and 7 distinct β-type subunits.¹ Despite this difference, the overall architecture of these complexes is conserved: α- and β-type subunits segregate into 7-member homo- [α7β7β7α7 (prokaryotes)] or heterooligomeric [(α1- α7) (β 1-β7)(β1-β7)(α1-α7) (eukaryotes)] rings. Two juxtaposed rings of β-type subunits flanked on top and bottom by a ring of α-type subunits form the barrel-shaped complex (Fig. 3.0). [171]

![Figure 3.0](image.png)

**Figure 3.0 The structural arrangement of the subunits α- and β.** (B) Ribbon structure showing seven fold symmetry for *Mycobacterium tuberculosis* proteasome. (Gin et. al 2004)

The β-subunits of the eukaryotic proteasome are able to cleavage peptides at acidic, basic and hydrophobic sites, allowing the proteasome to cleave most peptide bonds. [171] The main function of the α-rings is to form a gated channel that controls the passage of unfolded substrates
and cleaved particles in and out of the proteolytic chamber. The α-rings also serve as a docking surface for protein complexes such as the 19S RP, which interacts with either entrance to the 20S CP. Proteolytic activity is increased upon deletion of the N-terminal eight amino acids of the α-subunit. [171] This suggests that the α-ring effectively blocks access of even small peptide substrates to the 20S CP. This observation strongly supports the hypothesis that an activating protein or proteins are needed to open the α-ring for substrate degradation.

The 19S RP is composed of multiple ATPase (Rpt) and non-ATPase (Rpn) subunits. [173] The ATPase complex is formed by six distinct ATPases associated with different cellular activities (AAA+) that form a ring and interact with the α-subunits of the 20S CP. [173] The 19S RP also contains numerous proteins that bind and remove Ub from substrates prior to degradation by the 20S CP in eukaryotes. [178] In contrast, no bacterial proteasome has ever been shown to interact with its cognate ATPase in vivo. [177] The mechanism by which substrates are recognized for degradation is unknown because regulatory complexes that associate with proteasome-associated ATPases (PAN) or other prokaryotic PAN have not been identified. [177] It also remains unclear in the case of *Mtb* proteasome how a single type of β-subunit is able to confer such a broad spectrum of protease activity.

### Proteasome Mediated Substrate Degradation

Proteins that are targeted for degradation in eukaryotes are generally tagged with the molecule Ubiquitin (Ub), a small (76 aa) highly conserved regulatory protein. Immature Ub polypeptides are processed by Ubiquitin-specific proteases (USP) and Ubiquitin carboxy-terminal hydrolases (UCH) resulting in mature Ub molecules that end in a Gly-Gly motif. [177-179] The Gly-Gly motif is the feature common to all conjugatable Ubiquitin-like modifiers (UBL’s). The newly exposed C-terminal glycine is subjected to a series of reactions that conjugate Ub to lysines in the target protein. ATP is then used by activating enzymes to
adenylate the C-terminal glycine of Ub, which is then passed to an active site cysteine in the E1 enzyme. Subsequent steps involve the transfer of Ubiquitin to a Ub conjugating enzyme (E2) that delivers Ub to a protein ligase (E3), which catalyzes the formation of an isopeptide bond with a lysine on the target substrate. The E3 ligases can be subdivided into two groups, HECT (homologous to the E6-AP C terminus) and Ring Ligases (really interesting gene). Ring ligases hold both the E2 and the substrate and facilitate the transfer of Ub from E2 to the substrate. In contrast HECT ligases form a thioester bond with Ub before the transfer to a substrate lysine. The various E3 ligases present in eukaryotes give rise to various binding activities that provide specificity to the Ub-proteasome system by determining which substrates are conjugated to Ub. In the eukaryotic Ub proteasome system, only substrates that are polyubiquitinated to lysine 48 will be targeted for degradation (Fig 3.1).[177-179]
Figure 3.1 Overview of the eukaryotic Ubiquitin-proteasome system. (Darwin et. al 2009)
Despite the understanding of the eukaryotic Ub proteasome system, the role of proteasomes in prokaryotic physiology was largely unknown. Ubiquitin or UBL’s had never been successfully identified in bacteria or archaea and it was presumed that they were absent from this domain of life. [180, 181] Recently a small protein modifier, Pup was found to target mycobacterial proteins to the proteasome for degradation. [174, 181] Additional studies on this pathway may yield insight on the physiological implication.

Upon confirmation that prcBA in the Mtb genome encodes a proteasome 20S CP, it was assumed that two separate regions encoding an ATPase and a Proteasome accessory factor, PafA, were involved in the pathway not only due to their close proximity but also due to their presence in bacteria that contain proteasomes. [182] This was supported by reports that mutations in either gene resulted in NO-sensitivity and attenuated virulence, suggesting both participate in the same pathway. [172] The ATPase used in the Mtb proteasome pathway is Mycobacterium proteasome ATPase which is similar to ATPases present in the eukaryotic 19S RP. Studies have shown that Mpa forms hexameric rings similar to AAA+ ATPases. [173, 183]

Although structural analysis predicts that Mpa physically interacts with the Mtb proteasome 20S CP and plays a role in binding, unfolding and translocating substrates into the 20SCP, this has not been experimentally shown. [178] Using Mpa as bait in a bacterial two-hybrid screen of an Mtb genomic library searching for potential binding partners of Mpa, Pearce and colleges discovered the first prokaryotic Ubiquitin-like protein, Pup, a 64 amino acid protein that modifies and targets mycobacterium proteins to the proteasome for degradation. Pup is similar in size to Ub but the two proteins have different sequences and lack structure homology. [179]

Although the end-point for both the Ubiquitin degradation system in eukaryotes and the Pup degradation system in prokaryotes is the proteasome, the two functionally analogous tagging
systems do not share similar methods of activation and conjugation to target proteins. Pup has a di-glycine motif at the penultimate position of the C-terminus, followed by either glutamate (Glu) or glutamine (Gln), depending on the organism. It was initially hypothesized that the C-terminal amino acid (Glu or Gln) of Pup was removed to expose the di-glycine motif for conjugation, which would be analogous to Ubiquitin activation by C-terminal hydrolases. [177] Mass spectrometry on pupylated substrates in Mtb, where the C-terminal residue is Gln, revealed that Gln was not removed; instead it was deamidated to Glu prior to being conjugated to substrate lysines. The deamidation reaction is catalyzed by Dop (deamidase of Pup), which is encoded by a gene upstream of pup and the proteasome core particle genes. [184] Dop shares no homology with Ubiquitin-activating enzymes but bioinformatic analysis suggests structural homology to the carboxylate amine/ammonia ligase super family of glutamine synthetases.[185] This family catalyzes the ligation of amine groups to carboxylates, resulting in an amide linkage. It has been shown that some bacteria that encode Glu at the C-terminus of Pup and therefore do not require deamidation have retained the dop gene. This suggests that dop plays additional roles in the pupylation pathway. [177]

Bioinformatic analysis suggested that PafA (proteasome associated factor A) also has structural homology to the carboxylate amine/ammonia ligase super family. [182] PafA co-localizes with proteasome-associated genes and its gene product catalyzes the conjugation of Pup to the known proteasome substrates FabD and PanB in vitro in the presence of ATP and Dop. [182] ATP is hydrolyzed during the course of the reaction, suggesting that the PafA-catalyzed ligation reaction proceeds through a phosphorylated intermediate, as do other members of the carboxylate amine/ammonia ligase super family of enzymes. It is unknown which C-terminal carboxylate (the backbone carboxylate or the γ-carboxylate on the Glu) is conjugated to substrates. [177, 178] Pup terminating in Glu (Pup-Glu) instead of Gln (Pup-Gln) is a substrate for PafA-catalyzed conjugation in the absence of Dop, suggesting deamidation precedes
conjugation, and that Dop and PafA-mediated reactions are not necessarily coupled in vitro.[177, 178] In contrast to Ubiquitin activation and conjugation, which proceeds through a series of at least four enzymatic steps prior to substrate conjugation, Pup can be activated and conjugated in two steps. In organisms where pup encodes a C-terminal Glu, it is likely that only one enzyme is required for activation and conjugation.[180] In Mtb, a PafA mutation abolishes pupylation, suggesting that it is responsible for most, if not all pupylation in vivo. It is currently not known whether the Pup tag has additional functions besides proteasome targeting. It is possible that additional Pup ligases besides PafA exist in Mtb. These ligases may not always be present and may require additional factors for activity, or may serve functions other than proteasome targeting.[172,177,179,184]
Figure 3.2. Overview of proposed model for Pup-Proteasome pathway in *Mycobacterium tuberculosis*. 
Introduction

Inside a cell, macromolecular machineries have to be assembled with specificity to carry out necessary biological functions in the presence of a crowded cytosol. Usually, multiple ligands or substrates of the macromachineries are also present with a potential of binding the individual subunits and thus changing a pathway of how proper biological macromachinery conformation is achieved. The sequence of events, such a presence of the ligands and substrates prior to expression of macromacineries as opposed to the presence of the macromachines before ligands or substrates come about may play a regulatory role in these processes and it is not clear a priory that the final conformation of macromachineries will differ due to this temporal control.

Traditionally, biochemical and biophysical investigations carried out to study protein–protein interactions involve responses between isolated biopolymers in homogeneous and dilute solutions. A method to examine STructural INTeractions of proteins within their native environment by in-cell NMR, STINT-NMR (Structural Interactions using NMR spectroscopy) can be applied to gain understanding of the assembly of a macromolecular machinery, such as Mtb proteasome, inside the cell.

STINT NMR[104, 105, 134] is an in-cell technique for examining the structural changes in a target protein resulting from protein-protein interactions. Initially protein over expression is induced in a uniformly labeled medium [U-\(^{15}\)N] to produce a target protein containg a NMR-active nuclei. Cells are then transferred to a non-labeling medium to induce overexpression of the interactor protein. \(^{15}\)N- edited heteronuclear single quantum coherence (\(^{1}{H}(^{15}\text{N})\)-HSQC) NMR experiments are then used to monitor chemical shifts of target backbone amide nuclei as the concentration of interactor is increased. Because only the target protein is labeled we monitor the chemical shift changes of solely the target as the interactor remains cryptic. Depending on the
chemical exchange rate between the free and bound states of the target, affected NMR peaks, corresponding to backbone amides, can either shift, broaden their line shape or disappear completely, thereby delineating the intermolecular interaction surface between the target and the interactor(s). We used STINT-NMR to examine changes in the binding surface of the mediator of prokaryotic proteasome, Ubiquitin like protein, Pup, in the presence of the assembly of Mycobacterium ATPase (Mpa) as well as the full length Mpa-Proteasome complex.

Mtb proteasome presents a tractable system to study by STINT-NMR. Pup is a small intrinsically unstructured protein, with NMR relaxation properties favorable for in-cell NMR analysis. It is expressed as a precursor molecule, Pup-GGQ, before its C-terminal Gln is catalytically deaminated to create Pup-GGE during pupylation reaction. We used Pup-GGQ for our studies since it exists as a single molecule inside the cell not attached to its target.

E. coli is a relevant prokaryotic host that provides a proper milieu for the reaction without interfering factors. Indeed, it was used for functional studies of Mtb proteasome. Expression of Pup ligase in E. coli was used to study in-vivo pupylation of the proteins and proved that no additional factors are needed to pupylate proteins. Individual proteins of the proteasome were overexpressed in Ecoli in the active states. Importantly, Mtb proteasome which was expressed in E. coli as a precursor molecule, was maturating into the catalytically active state.

The main objectives of the experiments presented in this work are to gain physiological and structure insight as to how Pup-GGQ interacts with the Mpa-Proteasome in-cell by using STINT-NMR. This was demonstrated by recreation of the final steps in the mycobacterium degradation pathway in-cell. We showed that order of assembly of the Mpa-proteasome complex did not affect how Pup interacted with the Mpa-proteasome complex. In order for Pup’s N and C-terminus to be engaged the proteasome core particle has to be present. We have also shown that Pup-GGQ needs to be deamidated before it becomes a substrate for proteasomal degradation.
**Experimental Procedures**

*Plasmid Construction*

pTM-Pup.

Four separate oligonucleotides containing fragments of Pup, amino acids (1-31), (5-1) - 5’- CCC AAG CTT ATG GCG CAA GAG CAG ACC AAG CGT GGC GGT GGC GGC GAT GAT GAC GAC ATC GCC GGC AGC ACC GCC GCG GGC CAG GAG CGT CGC GAA AAG CTG ACC GAG GAG ACC GAC-3’ and (5-2)- 5’- /PO₄_/ GAT CTG CTC GAC GAA ATC GAC GAC GTC CTC GAG GAG ACC GCC GAG GAC TTC GTC CGC GCA TAC GTC CAA AAG GGC GGA CAG TGA GGA TCC AA-3’ amino acids (32-64), with reverse complements of (3-1) – 5’- / PO₄_/ GGT CTG CTC GAG GAG TTC GTC CGC GCA TAC GTC CAA AAG GGC GGA CAG TGA GGA TCC AA-3’ and (3-2) – 5’ – TTG GAT CCT CAC TGT CCG CCC TTT TGG ACG TAT GCG CGG ACG AAG TCC TCG GCG TTC TCC TCG AGG ACG TCG TCG ATT TCG TCG AGC AGA TCG TC-3’ were Duplexed as follow, (5-1) with (3-1) and (5-2) with (3-2) and ligated together to form the complete gene sequence (amino acids 1-64) flanking containing 5’ BamHI and 3’ HindIII sites restriction sites. The restriction digested product was ligated into pTM vector to yield pTM-Pup. The resulting plasmid, pTM-Pup, expresses Pup from a T7 promoter/ lac operon which is induced by isopropyl-β-D-thiogalactoside (IPTG). This plasmid confer Kanamycin resistance and has a lacI gene and pBR322 origin which encodes for Lac repressor.
pASK3+ -Pup.

DNA coding for full-length Pup was amplified from pTM-Pup using the oligonucleotides 5'- AAA AAA GGT CTC TAA TGG CGC AAG AGC AGA CCA AGC GTG G-3’ and 5'- AAA AAA GGTCTC AGC GCT TCA ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG CTG TCC GCC CTT TTG GAC GTA TGC - 3’. The gene was ligated into pASK-IBA3+ [ IBA] using the Bsall linker sites. The resulting plasmid, pASK3+ -Pup, expresses Pup from a tet promoter/operator, which is induced by tetracycline or anhydrotetracycline. This plasmid confers ampicillin resistance and contains an f1 origin and the tet gene, which codes for Tet repressor.

pASK3+ -Pup-His7x.

DNA coding for full-length Pup was amplified from pASK3+ -Pup using the oligonucleotides 5'- AAA AAA GGT CTC TAA TGG CGC AAG AGC AGA CCA AGC GTG G-3’ and 5'- AAA AAA GGTCTC AGC GCT TCA ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG ATG CTG TCC GCC CTT TTG GAC GTA TGC - 3’. The gene was ligated into pASK-IBA3+ [ IBA] using the Bsall linker sites. The resulting plasmid, pASK3+ -Pup His7x, expresses Pup His7x from a tet promoter/operator, which is induced by tetracycline or anhydrotetracycline. This plasmid confers ampicillin resistance and contains an f1 origin and the tet gene, which codes for Tet repressor.

pRSF-Mpa-S.

DNA coding for full-length Mpa-S (1149 aa) was PCR amplified from isolated genomic DNA extracted from Mycobacterium smegmatis by using Phusion polymerase and oligonucleotide primers 5'- AAA AAA GGT ACC ATG GGT GAG TCA GAG CGT TC- 3’ and 5’AAA AAG CTT TCA CAG GTA CTG GCC GAG GTT GG3’. The gene was ligated into pRSF-1b [Novagen] using the Kpnl and HindIII linker sites. The resulting plasmid, pRSF-Mpa-
S, expresses Mpa from a T7 promoter/lac operator (P_{T7}/lacOp), which is induced by IPTG. These plasmid confers kanamycin resistance and contains an RSF replication origin and the lacI gene, which encodes for Lac repressor.

pRSF-Mpa-T.

DNA coding for full-length Mpa-T (1149 aa) was PCR amplified from, M.tuberculosis H37Rv isolated genomic DNA by using Phusion polymerase and oligonucleotides 5’-AAA AAA GGT ACC ATG GGT GAG TCA GAG CGT TCT GAG G-3’ and 5’-AAA AAG CTT TCA CAG GTA CTG GCC GAG GTT GG3 sites. The gene was ligated into pRSF-1b [Novagen] using the KpnI and HindIII linker sites. The resulting plasmid, pRSF-Mpa-T, expresses Mpa, from a T7 promoter/lac operator (P_{T7}/lacOp), which is induced by IPTG. This plasmid confers kanamycin resistance and contains an RSF replication origin and the lacI gene, which encodes for Lac repressor.

pACYCDuet - PrcAB and pACYCDuet - PrcΔAB.

Plasmids encoding PrcAB in vector pACYDuet-1, a gift from by G.Lin, Weil Medical College of Cornell University, were used to overexpress Tuberculosis proteasome and Mtb proteasome Open gate mutant (PrcΔAB) were. These plasmids expresss Mtb-Proteasome or Open gate proteasome from a T7 promoter/lac operator (P_{T7}/lacOp), which is induced by IPTG. This plasmids confer chloramphenicol resistance and contains an P15A replication origin and the lacI gene, which encodes for Lac repressor.
Sequential over-expression and labeling.

*E. coli* strain BL21(DE3) codon+ [Novagen] was co-transformed with pASK-Pup and pRSF-Mpa-S (Pup-Mpa interaction); or pASK-Pup, pRSF-Mpa and pACYCDuet - PrcAB (Pup-Mpa-wild type proteasome interaction); or pASK-Pup, pRSF-Mpa and pACYCDuet - PrcΔAB (Pup-Mpa-Open gate proteasome interaction) or pASK-Pup, pRSF-Mpa and pACYCDuet - PrcAB (PupHis7x-Mpa-wild type proteasome interaction); or pASK-PupHis7x, pRSF-Mpa and pACYCDuet - PrcΔAB (Pup-His7x-Mpa-Open gate proteasome interaction).

Cells were grown overnight at 37 °C to an OD₆₀₀ of 1.6 in Luria-Bertani medium (LB) supplemented with 150 mg/L of carbenicillin for cultures containing pASK-Pup or pASK-PupHis7x, 35 mg/L of kanamycin for cultures containing pRSF-Mpa-S or pRSF-Mpa-T, and 33 mg/L of chloramphenicol for cultures containing pACYCDuet - PrcAB and pACYCDuet-PrcΔAB. Four protocols were employed: In the first the target Pup was overexpressed and labeled followed by the over expression of Mpa. In the second, Mpa was expressed followed by overexpression and labeling of Pup. In the third, the Pup labeled target was overexpressed and labeled followed by overexpression of the complex of Mpa with (wild-type or Open gate proteasome); in the fourth, the complex of Mpa with (wild type or Open gate proteasome) were overexpressed followed by overexpression and labeling of the Pup target.
Table 2. Plasmid Construction for STINT-NMR.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Gene</th>
<th>Size (Kb)</th>
<th>Ab&lt;sup&gt;R&lt;/sup&gt;</th>
<th>Inducer</th>
<th>Origin</th>
<th>Protein Expressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRSF-1b</td>
<td><em>Mpa</em></td>
<td>5.4</td>
<td>Kn</td>
<td>IPTG</td>
<td>RSF</td>
<td>Mycobacterium smegmatis ATPase</td>
</tr>
<tr>
<td>pACYCDuet</td>
<td><em>prcA-B</em></td>
<td>5.5</td>
<td>Cm</td>
<td>IPTG</td>
<td>P15</td>
<td>Mycobacterium Proteasome wild-type</td>
</tr>
<tr>
<td>pACYCDuet</td>
<td><em>prcΔA-B</em></td>
<td>5.5</td>
<td>Cm</td>
<td>IPTG</td>
<td>P15</td>
<td>Mycobacterium Proteasome Open Gate</td>
</tr>
<tr>
<td>pASK3+</td>
<td><em>pup</em></td>
<td>3.4</td>
<td>Ap/Cb</td>
<td>Tet</td>
<td>ColE1</td>
<td>Prokaryotic Ubiquitin-like Protein</td>
</tr>
</tbody>
</table>

Cells from the overnight culture were washed once with minimal medium (M9) salts and re-suspended to an OD600 of 0.5 in M9 medium containing the appropriate antibiotics, [U-15N] ammonium chloride (0.7 g/L) as the sole nitrogen source and 0.2% glucose as the sole carbon source. For all induced cultures we substituted ampicillin (100 mg/L) for carbenicillin. The cells were incubated at 37 ºC for 10–15 minutes and Pup overexpression was induced by adding 2 mg/mL of anhydrotetracycline in dimethylformamide (DMF) to a final concentration of 0.2 mg/mL. Pup overexpression was allowed to proceed for up to 4 hours. Following the first induction, a 100 mL sample of culture was collected, the cells were centrifuged, washed twice with 50 mL of 10 mM potassium phosphate buffer [pH 6.5], resuspended with 1 mL of 10 mM potassium phosphate buffer [pH 6.5] containing 10% glycerol and stored at -80 ºC for subsequent NMR analysis. The use of a cryoprotectant is critical to eliminate cell lysis or breakage due to repeated freeze-thawing. This control sample was used to assess the extent of overexpression and quality of labeling for a given experiment. Culture samples (50 mL) were taken prior and post induction for SDS-PAGE analysis.

Expression of Mpa-S.

Following Pup overexpression and labeling, the culture was centrifuged and washed once with M9 salts before resuspending a sufficient number of cells to yield an OD600 of 0.5 in LB medium supplemented with the appropriate antibiotics. The culture was incubated at 30 ºC for 10–15 minutes and 1.0 M IPTG was added to a final concentration of 1.0 mM to induce individual over-expression of Mpa-S; induction was allowed to proceed for 16 hours. 100 mL of samples were taken, centrifuged, washed twice with 10 mM potassium phosphate buffer [pH 6.5], resuspended with 1 mL of 10 mM potassium phosphate buffer [pH 6.5] containing 10% glycerol.
and stored at -80 °C for subsequent NMR analysis. Culture samples (50 mL) were taken prior and post induction for SDS-PAGE analysis.

Protocol 2: Expression of Mpa-S.

Cells from the overnight culture were washed once with LB medium and re-suspended to an OD$_{600}$ of 0.5 in LB medium containing the appropriate antibiotics. For all induced cultures we substituted carbenicillin (150 mg/L) for ampicillin. The cells were incubated at 37 °C for 10–15 minutes and 1.0 M IPTG was added to a final concentration of 1.0 mM to induce over-expression of Mpa-S; induction was allowed to proceed for 16 hours. Following the first induction, a 100 mL sample of culture was collected, the cells were centrifuged, washed twice with 50 mL of 10 mM potassium phosphate buffer [pH 6.5], resuspended with 1 mL of 10 mM potassium phosphate buffer [pH 6.5] containing 10% glycerol and stored at -80 °C for subsequent NMR analysis. A control sample was used to assess the extent of overexpression for a given experiment. Culture samples (50 mL) were taken prior and post induction for SDS-PAGE analysis.

Expression of: $[^{15}N]$-Pup.

Following Mpa-S expression the culture was washed once with minimal medium (M9) salts and re-suspended to an OD$_{600}$ of 0.5 in M9 medium containing the appropriate antibiotics, $[^{15}N]$ ammonium chloride (0.7 g/L) as the sole nitrogen source and either 0.2% glucose as the sole carbon source. For all induced cultures we substituted carbenicillin (150 mg/L) for ampicillin. The cells were incubated at 37 °C for 10–15 minutes and Pup overexpression was induced by adding 2 mg/mL anhydrotetracycline in DMF to a final concentration of 0.2 mg/mL. Pup overexpression was allowed to proceed for up to 4 hours. Following the second induction, a 100 mL sample of culture was collected, the cells were centrifuged, washed twice with 50 mL of 10 mM potassium phosphate buffer [pH 6.5], resuspended with 1 mL 10 mM potassium phosphate buffer [pH 6.5] containing 10% glycerol and stored at -80°C for subsequent NMR analysis.
analysis. This control sample was used to assess the extent of overexpression and quality of labeling for a given experiment. Culture samples (50 mL) were taken prior and post induction for SDS-PAGE analysis.

Protocol 3: Expression of $[U^{15}N]$-Pup.

Cells from the overnight culture were washed once with minimal medium (M9) salts and re-suspended to an $OD_{600}$ of 0.5 in M9 medium containing the appropriate antibiotics, $[U^{15}N]$ ammonium chloride (0.7 g/L) as the sole nitrogen source and either 0.2% glucose as the sole carbon source. For all induced cultures we substituted carbenicillin (150 mg/L) for ampicillin. The cells were incubated at 37 ºC for 10–15 minutes and Pup overexpression was induced by adding 2 mg/mL of anhydrotetracycline in DMF to a final concentration of 0.2 mg/mL. Pup overexpression was allowed to proceed for up to 4 hours. Following the first induction, a 100 mL sample of culture was collected, the cells were centrifuged, washed twice with 50 mL of 10 mM potassium phosphate buffer [pH 6.5], resuspended with 1 mL of 10 mM potassium phosphate buffer [pH 6.5] containing 10% glycerol and stored at -80°C for subsequent NMR analysis. A control sample was used to assess the extent of overexpression and quality of labeling for a given experiment. Culture samples (50 mL) were taken prior and post induction for SDS-PAGE analysis.

Expression of Mpa-S and wild- type proteasome.

Following Pup overexpression and labeling, the culture was centrifuged and washed once with M9 salts before resuspending a sufficient number of cells to yield an $OD_{600}$ of 0.5 in LB medium supplemented with the appropriate antibiotics. The culture was incubated at 37ºC for 10–15 minutes and 1.0 M IPTG was added to a final concentration of 1.0 mM to induce overexpression of the Mpa-S wild-type proteasome complex was allowed to proceed for 16 hours. 100 mL samples were taken, centrifuged, washed twice with 10 mM potassium phosphate buffer
[pH 6.5], resuspended with 1 mL 10 mM potassium phosphate buffer [pH 6.5] containing 10% glycerol and stored at -80°C for subsequent NMR analysis. Protocol 3 was also used to express the Open gate proteasome in place of the wild-type via the same procedure. 50 mL Culture samples were taken pre- and post- induction for SDS-PAGE analysis.

Protocol 4: Mpa-S and wild-type proteasome.

Cells from the overnight culture were washed once LB medium and re-suspended to an OD$_{600}$ of 0.5 in LB medium containing the appropriate antibiotics. For all induced cultures we substituted carbenicillin (150 mg/L) for ampicillin. The cells were incubated at 37 °C for 10–15 minutes and Mpa-S and wild type proteasome overexpression was induced by adding 1.0 M IPTG to a final concentration of 1.0 mM induction was allowed to proceed for 16 hours. A 100 mL sample of culture was collected, the cells were centrifuged, washed twice with 50 mL of 10 mM potassium phosphate buffer [pH 6.5], resuspended with 1 mL 10 mM potassium phosphate buffer [pH 6.5] containing 10% glycerol and stored at -80°C for subsequent NMR analysis. 50 mL culture samples were taken pre- and post- induction for SDS-PAGE analysis. Protocol 4 was also used for the co-expression of Mpa-S with Open gate proteasome in place of wild type proteasome. 50 mL culture samples were taken pre- and post- induction for SDS-PAGE analysis.

Expression of [$U^{-15}N$]-Pup.

Following Mpa-S and wild-type proteasome expression the culture was washed once with minimal medium (M9) salts and re-suspended to an OD$_{600}$ of 0.5 in M9 medium containing the appropriate antibiotics, $^{15}$N ammonium chloride (0.7 g/L) as the sole nitrogen source and 0.2% glucose as the sole carbon source. For all induced cultures we substituted ampicillin (100 mg/L) for carbenicillin. The cells were incubated at 37 °C for 10–15 minutes and Pup overexpression was induced by adding 2 mg/mL of anhydrotetracycline in DMF to a final concentration of 0.2 mg/mL. Pup overexpression was allowed to proceed for up to 4 hours. Following the second
induction, a 100 mL sample of culture was collected, the cells were centrifuged, washed twice with 50 mL of 10 mM potassium phosphate buffer [pH 6.5], resuspended with 1 mL of 10 mM potassium phosphate buffer [pH 6.5] containing 10% glycerol and stored at -80 °C for subsequent NMR analysis. A control sample was used to assess the extent of overexpression and quality of labeling for a given experiment.

**SDS PAGE Analysis.**

For protocols 1 and 3, gel analysis was performed by taking 100 mL samples at time points 0 hr, 2 hr, 4 hr, 6 hr post Pup expression followed by 12 hr and 24 hr post Mpa or Mpa-Proteasome expression. For protocols 2 and 4, 100 mL samples were taken at 12 hr and 24 hr post Mpa or Mpa-Proteasome expression followed by 0 hr, 2 hr, 4 hr, 6 hr post Pup expression. Samples were harvested via centrifugation and sonicated. Lysates were analyzed on a 18% acrylamide gel and developed by either Commasie Blue staining or western blotting using SuperSignal West HisProbe Kit [Pierce].

**NMR spectroscopy.**

Cells containing [U-^{15}N] Pup were re-suspended in 0.5 mL of NMR buffer (10 mM potassium phosphate, pH 6.5, 90%/10% H_{2}O/D_{2}O) and transferred to an NMR tube. To rule out the possibility that the NMR spectrum was due to extracellular proteins due to leakage from the cells, we sedimented the cells from the NMR sample and acquired the $^{1}H\{{^{15}N}\}$-HSQC spectrum of the resultant supernatant. No protein NMR signal was visible above the noise level. All NMR experiments were performed using a Bruker Avance 500 MHz NMR spectrometer equipped with a cryoprobe. The cryoprobe affords a four-fold increase in sensitivity allowing data collection within 2 hr for an individual experiment; this is critical to minimize cell leakage. We used a watergate version of the $^{1}H\{{^{15}N}\}$-edited HSQC data were recorded with 32 transients as 512{64}
complex points, apodized with a squared cosine-bell window function and zero-filled to 1k{128} points prior to Fourier transformation. The corresponding sweep widths were 12 and 35 ppm in the $^1$H and $^{15}$N dimensions, respectively. Chemical shifts of $[U-, ^{15}$N]-Pup inside the cell are slightly different from purified Pup. We reassigned the backbone chemical shifts of Pup using the clarified lysate of $[U-, ^{13}$C, $^{15}$N]-Pup and a standard suite of triple resonance experiments. During in-cell titration experiments, we measured the change in the chemical shifts( $\Delta \Omega$) of amide nitrogens and covalently attached amide protons according to the equation: 

$$\Delta \Omega = \sqrt{\delta_H^2 + (0.25)\delta_N^2},$$

where $\delta_{1H(1N)}$ represent hydrogen and nitrogen chemical shifts. Even without quantifying protein concentrations present in the cell, in-cell NMR spectroscopy allows us to make a crude estimate of the protein binding affinities. Depending on the magnitude of the chemical shift change $\Delta \Omega$ and the rate constant, $k_{off}$, between bound and free states, chemical exchange can result in gradual changes of chemical shifts when $\Delta \Omega < k_{off}$ (fast exchange), line broadening when $\Delta \Omega \leq k_{off}$ (intermediate exchange) or the appearance of new peaks when $\Delta \Omega >> k_{off}$ (slow exchange). Assuming that the binding reaction is diffusion limited and the average change of the chemical shift is , 0.01 ppm, the fast exchange regime will occur when the dissociation constant, $K_d$, is larger than 100 $\mu$M and intermediate or slow exchange will occur when the dissociation constant is less than or equal to 10 $\mu$M.
Expression and purification of proteins for in vitro assays.

*E. coli* strain BL21(DE3) codon+ [Novagen] was individually transformed with pASK-Pup His$_{7x}$, pRSF-Mpa-S or pRSF-Mpa-T and pACYCDuet - PrcΔAB, pACYCDuet – PrcAB. Cells expressing Pup His$_{7x}$ were grown in LB medium containing 150 μg/mL of carbenillicin at 37 ºC to an OD$_{600}$ of 0.7 and induced with 2 mg/mL of anhydrotetracycline in DMF to a final concentration of 0.2 mg/mL for 6 hr. Cells expressing Mpa-S were grown in LB broth containing 35 mg/L of kanamycin at 37 ºC to a OD$_{600}$ of 0.7 and induced with 1.0 mM IPTG for 12 hr. Cells expressing Mpa-T were grown in LB broth containing 35 μg/mL of kanamycin at 30 ºC to a OD$_{600}$ of 0.7 and induced with 1.0 mM IPTG for 12 hr. The induced cells were harvested by centrifugation at 4000 g for 10 min at 4 ºC. Cells were suspended in 30 mL of ice-cold lysis buffer (50 mM NaH$_2$PO$_4$ [pH 8.0], 300 mM NaCl, 10mM imidazole), and were lysed by sonication using a Branson instrument at 300W, for 0.3 s intervals followed by 1 s rests for 5 min). The lysate was centrifuged at 20000 g for 20 min at 4 ºC. The supernatant was loaded onto a Ni-NTA resin-filled column (Qiagen). The column was washed with 20 mL of lysis buffer and 30 mL of wash buffer (50 mM NaH$_2$PO$_4$ [pH 8.0], 300 mM NaCl, 20 mM imidazole). Fusion proteins were eluted with elution buffer (50 mM NaH$_2$PO$_4$ [pH 8.0], 300 mM NaCl, 250 mM imidazole). The eluted proteins were buffer exchanged against 100 ml buffer R (50 mM Tris-HCl [pH 7.5], 150 mM 10% (v/v) glycerol, 20 mM MgCl$_2$, 1 mM DTT). $^{15}$N-labelled Pup His$_{7x}$ was prepared in the same way except that LB broth was replaced by M9 medium containing 0.7 g/L of 99% $^{15}$NH$_4$Cl as the sole nitrogen source. SDS-PAGE analysis revealed bands at 15 kDa for Pup, 65 kDa for Mpa-S and 66 kDa for Mpa-T. Protein concentration was calculated by using the Bradford method.

*E. coli* strain C41 DE3+ [Novagen] was individually transformed with pACYCDuet - PrcΔAB and pACYCDuet – PrcAB. Single colonies were used for inoculation into Overnight
Express Autoinduction System 1 medium (Novagen) and grown at 37 °C for 18hr. Cells were centrifuged at 6000 g for 20 min, resuspended in ice-cold lysis buffer and lysed using a Branson sonicator at 300 W, for 0.3 s pulse intervals followed by 1 s rest for 5 min. The remaining purification steps were carried out at 4 °C. Supernatants were centrifuged at 20000 g for 30 min, mixed with Ni²⁺- nitrilotriacetate acid (Ni-NTA) agarose beads and incubated for 2 hr or overnight. The beads were washed four times with wash buffer and proteins eluted in elution buffer. Fractions containing PrcB and PrcA, as judged by Coomassie staining of SDS-PAGE gels, were pooled and purified on Hi-trap Q-Sepharose (Amersham) via FPLC (Amersham, NJ) with a gradient of buffer in 50 mM Tris-HCl, pH 7.5 0–1 M NaCl. The eluted proteins were buffer exchanged against 100 mL reaction buffer R.

In vitro Degradation Assay

In vitro Pup degradation by the Mpa proteasome was performed as described by Striebel et al; (2010) [183]. In summary, the reaction was carried out in buffer R completed with the addition of 25 mM phosphocreatine (Sigma), 1 U/ml creatine phosphokinase (Sigma), 0.3 mM Mpa-S, 0.4 mM open-gate proteasome, 3 mM Pup with 5 mM ATP incubated at 25 °C. Assays in which wild type proteasome was used in place of open gate proteasome or where Mpa-T was used in place of Mpa- S were carried out under identical conditions. Samples were taken at 0, 30, 60, 120, 240 min and the reaction terminated by adding SDS-sample buffer and examined by Coomassie-stained SDS–PAGE and or western analysis using SuperSignal West HisProbe Kit [Pierce].
**Proteasome Activity Assay**

Wild-type/ Open gate proteosomes were purified as described for the *in vitro* assay and buffer exchanged to buffer PA (20 mM Hepes, 0.5 mM EDTA, pH 7.5). Proteasomes (10 ng protein) were incubated at 37°C with 20 μM N-Succinyl-Leu-Leu-Val-Tyr -7-amino-4-methylcoumarin (Suc-LLVK-Amc) using a 96 well black plate (Corning) in a fluorescence spectrophotometer with continuous stirring to monitoring the release of Amc by fluorescence at λ<sub>ex</sub> = 360 nm, λ<sub>em</sub> =440 nm; experiments were performed in triplicate. Inhibition of the proteasome was accomplished by adding (1 nmol ) of Bortzomib to each reaction to ensure that degradation occurred via the proteasome.
Results

The in-cell spectrum of Pup is disordered.

Purified Pup-GGQ is a disordered protein [186]. For in-cell NMR studies, we cloned a wild type *Mtb* Pup-GGQ into pASK3+, which confers ampicillin resistance and permits heterologous protein expression induced by anhydrotetracycline. Curiously, Pup-GGQ stains very weakly by protein stains and Pup-GGQ overexpression in the cells cannot be easily observed on the SDS-PAGE gels. We created Pup-GGQ-6xHis construct, cloned it into pASK3+, and used Western blotting with anti-His to visualize Pup-GGQ expression. In order to determined whether intracellular Pup contains any regions of induced secondary structure due to macromolecular crowding, we overexpressed Pup for three hours in *E. coli* grown on [*U*-\(^{15}\)N] M9 minimal media, washed and re-suspended the cell pellet in the NMR buffer and collected in-cell \(^1\)H[\(^{15}\)N] HSQC NMR spectrum. In-cell NMR spectrum of [*U*-\(^{15}\)N] Pup exhibited at least 5:1 signal to noise ratio and contained well-resolved backbone amide peaks (Fig 3.3). To determine that the in cell NMR spectrum of Pup was due solely to intracellular protein, after collecting in-cell NMR spectrum, the in-cell NMR sample was centrifuged and the supernatant was examined. No NMR signal was observed above the noise level implying that no leakage or cell lysis was occurring during experimental acquisition time.

Following the in-cell NMR spectrum acquisition, cells containing [*U*-\(^{15}\)N] Pup were lysed and the \(^1\)H[\(^{15}\)N]-HSQC spectrum of the lysate was acquired and compared to the in-cell NMR spectrum. All residues peaks from both the in-cell and lysate NMR spectra were found within 0.01 ppm of each other, with uniform peak broadening arising from decreased tumbling of Pup due to increased viscosity within the cytoplasm of the cell. Backbone chemical shifts are exquisitively sensitive to changes in the backbone conformation and expected to exhibit major
 (>0.5 ppm) changes due to acquisition of secondary structure elements. Since only minor changes in chemical shifts and peak intensities have occurred, we concluded that Pup remains an intrinsically disordered protein in vivo. (Supplementary Fig. 9 and 10 )
Figure 3.3. Pup does not contain any regions of secondary structure due to macromolecular crowding. A. In-cell $^{15}$N-HSQC spectrum of $[U-^{15}N]$ Pup. B. In-vitro $^{15}$N-HSQC spectrum of $[U-^{15}N]$ Pup. Broadened peaks as shown in panel A. are due to the high viscosity inside the cell and lower rate of tumbling in the cytosol. Sharpening of peaks in panel B. is due to the lower viscosity upon lysis which allows faster tumbling. Peaks designated by a asterisk are metabolites in cytosol.
Mpa binds to multiple sites on Pup.

Alpha and beta carbon NMR chemical shifts are used to identify the regions of protein helical structure, even though these structures exist in solution only transiently. [187] Even though Pup is disordered in solution, residues 21 through 51 of Pup have propensity to form a transient helical structure. [186] In-vitro NMR studies revealed that the Pup region of transiently helical structure is involved in binding to the proteasomal ATPase, Mpa. Crystal structure of Mpa-Pup complex showed that Pup binds to the coiled coil domain of Mpa and the binding indicates a stable helical conformation in region 21-64 of Pup. The N-terminus of Pup remains unstructured in the Pup-MPA complex. [188]

We used STINT-NMR to analyze the interaction surface of Pup binding to Mpa in vivo. We cloned Msm Mpa into a bacterial plasmid pRSF that places heterologous gene expression under the control of IPTG inducible Lac-promoter. Both pRSF and pASK3+ are compatible plasmids and can be used to sequentially overexpress Mpa and Pup. By overexpressing one protein on [U- $^{15}$N] labeled medium and then switching to the unlabeled medium and inducing overexpression of the second protein, we created complexes of [U-, $^{15}$N]- Pup- Mpa. We varied the order of expression, Pup first then Mpa, or Mpa followed by Pup to assess how the in vivo assembly of Mpa is influenced by the presence of its ligand, Pup. The $^1$H{$^{15}$N}-HSQC spectrum of free Pup was overlayed with $^1$H{$^{15}$N}-HSQC of [U- $^{15}$N] Pup-Mpa complex in order to determine changes in chemical shifts and peak intensities. Upon overexpression of Pup for 3 hours followed by overexpression of Mpa for 6 hours we observed the most pronounced chemical shift changes (>0.06ppm) and 50% of peak intensities broadening in the following residues G9, R28, K31, E34, D37, L39, I43, D44, D45, V46,L47, 50, D53, R56, A57,Y58, and V59 (Fig. 2A). (Supplementary Fig. 11 and 12)
Since binding affinity of Pup for Mpa was estimated to be submicromolar[186] we expected that the selective broadening of the Pup residues, which are in direct contact with Mpa would dominate the spectrum. Indeed, regions of reported Pup helical structure showed the most pronounced changes in chemical shift and intensity. When [$U\cdot^{15}N$] Pup was overexpressed for two hours after overexpressing unlabeled Mpa for 5 hours, the most pronounced chemical shift changes and 85% intensities broadening occurred to residues Q3, G9, A19, A23, L39, I43, D44, D45, V46, L47, D53 R56, V59, Q60, K61, and G63. These results again indicate that the regions of the reported Pup helical structure were broadened out. There was no significant difference between the Pup regions affected by Mpa binding due to the order of sequential expression. We concluded that Mpa assembles into its functional state regardless of the presence or absence of even a high concentration of its physiological ligand. (Supplementary Fig. 13 and 14)
Figure 3.4. In-cell Mpa-S binds to the C-terminal region of Pup. A. Overlay of $^{15}$N-HSQC spectra of Pup titrated with Mpa-S (Protocol 1). B. Overlay $^{15}$N-HSQC of Pup expressed after Mpa-S induction (Protocol 2). Free Pup is red and titrated Pup is black. Panel A is characteristic of an equilibrium between free and bound Pup where residue between S21 and Q64 exhibit the greatest chemical shift changes and intensity changes where as Panel B shows only Pup bound to Mpa-S. C. Sequence of Pup showing greatest intensity changes upon titration in red. Helical binding region is indicated by coil.
Pup is not degraded when bound to Mpa-proteasome complex in vivo.

To examine whether order of proteasomal assembly affects the degradation or in-vivo structure of Pup we performed a STINT-NMR experiment in which we tightly control the order of induction of Pup with its interactor complex of Msm Mpa and wild type (WT) or open gate Mtb proteasome. Open gate proteasomal core lacks inhibitory n-terminal 8 amino acidss and shows increased proteolytic activity towards its substrates.

We used Mtb proteasomal core cloned in pACYC-Duet, which places the expression of two proteasomal subunits prcA and prcB under control of IPTG inducible Lac repressor. pACYC-Duet is compatible with both pASK3+ and pRSF, which were used to overexpress Pup and Msm Mpa, respectively. In this case, IPTG induces both Mpa and proteasomal core allowing us to reconstitute mycobacterial proteasomal system in E. coli. We were unable to achieve high overexpression level of Mtb Mpa and proteasomal core in the same cell, possibly, due to genomic instability of overexpressing plasmids. There is 91% sequence identity between Msm and Mtb Mpa. Due to similar genetic content, Msm is often used in functional studies of Mtb proteasomal activity[189].

In order to analyze changes in the interaction surface of Pup we created complexes of [U-^15N] Pup- Mpa- WT proteasome and varied the order of the expression, in each case expressing Pup first then Mpa- WT proteasome or Mpa- WT proteasome followed by overexpressing [U-^15N]- Pup. In-cell $^1$H{^15N}-HSQC of free Pup was overlayed with in-cell $^1$H{^15N}-HSQC of [U-^15N] Pup-Mpa-Wt proteasome complex in order to determine changes in chemical shift and peak intensities due to the complex formation. We expected that in the presence of the Mpa- WT proteasome complex Pup’s N- terminus will be engaged by the proteasomal complex resulting in Pup’s degradation [190].
Figure 3.5. Pup is not degraded in the presence of the Mpa-S - wild type proteasome complex. A. Overlay of in-cell $^{15}$N-HSQC spectra of free Pup (red) titrated with Mpa-S wild-type proteasome complex (black) (Protocol 3). B. Overlay of in-cell $^{15}$N-HSQC of free Pup (red) and Pup bound to Mpa-S-wild type proteasome complex after primary overexpression of Mpa-S - wild-type proteasome complex (black) (Protocol 4). Panel A is indicative of an equilibrium between free and bound Pup to Mpa-S where residues S21 through Q64 reflect the greatest chemical shift and intensity changes. Panel B shows Pup completely bound to Mpa-S in the presence of the Mpa-S - wild type proteasome complex. C. Sequence of Pup showing greatest intensity changes upon titration in red. Helical region is indicated by coil.
Surprisingly, as the concentration of Mpa–WT proteasome complex increased, we observed no degradation of Pup, as monitored by the Western blot or in-cell NMR. We observed the most pronounced chemical shift changes (>0.5 ppm) and 20% of peak intensities in the following residues: G10, A23, L32, D38, L40, D41, E42, D44, D45, V46, L47, A57, V59, and G62. (Supplementary Fig. 19 and 20)

Regions of the reported helical structure of Pup showed some of the most pronounced changes in chemical shift and intensity. Since N-terminus of Pup was largely unaffected by binding we argued that the intracellular concentration of proteasomal core was significantly lower that the concentration of Mpa, consistent with Western blotting (Fig 3.6).

Figure 3.6 Order of assembly of the Mpa-Proteasome complex does not affect the binding of Pup. A. Western blot analysis of in-cell overexpression of the Pup-His7x followed by overexpression of Mpa-S-Open gate proteasome complex. PrcB is the proteasomal β-subunit, PreB is unprocessed proteasomal β-subunit. B. Reverse order of assembly of the in vivo overexpression where the Mpa-Proteasome complex is formed first followed by the
overexpression of Pup- His$_{7x}$ and monitored for 12hr. Note that in both Figs A. and B. Pup-His$_{7x}$ shows no sign of degradation in cell.

When [U- $^{15}$N]- Pup was over expressed post Mpa-WT proteasome complex induction, the most pronounced chemical shift changes occurred to residues [Q5, G10, G20, A24, A24, E34, L39, L40, D41, E42, I43, D44, D45, L47, N50, A51, E52, D53, F54 Q60] (Fig. 3.5). No degradation of Pup was observed in the western blot (Fig 3.6). Signal intensities were almost uniformly 95% broadened out for the majority of Pup residues. Thus, interaction between Pup and Mpa-WT proteasome affects not only the helical region of Pup but also Pup’s N-terminal tail. (Supplementary Fig. 21 and 22)

We rationalized that Mpa-WT proteasome may require additional factors not present in E. coli to remove inhibition for proteolysis of Pup. We substituted WT for Open gate proteasome where inhibitory ... N-terminal residues are removed to possibly circumvent this requirement. A secondary STINT-NMR was carried out in identical manner as described above to form the Mpa-Open gate proteasome complex. No Pup degradation was observed in this case. As the concentration of Mpa-Open gate proteasome complex was increased in the presence of [U-, $^{15}$N]-Pup (Fig 3.7.), the most pronounced chemical shift changes and intensity changes where in residues [A23, D38, L39 D41, D46, V46, N50, R56, A57, Q60] with most chemical shift staying within 0.01ppm of their original shift of [U-, $^{15}$N]- Pup control$^1$H{$^{15}$N}-HSQC spectrum. (Supplementary Fig. 15 and 16)As seen with the Mpa –WT proteasome complex, overlay of $^1$H{$^{15}$N}-HSQC spectrum with free Pup reveals some peaks to have not have shifted suggesting an end point titration of the complex with Pup was not reached. However, when the Mpa-Open gate complex is overexpressed first followed by [U-, $^{15}$N]- Pup and the spectra is compared to that of free Pup, the most pronounced chemical shift changes were seen in [ Q5, G9, A19, A23, L39, I43, D44, D45, V46, L47, D53, R56, V59, Q60, K61, G63]. Signal intensities were almost
uniformly 95% broadened out for the majority of Pup residues, though no Pup degradation was observed. Thus, interaction between Pup and Mpa-Open gate proteasome, identical to pup-Mpa-WT proteasome, affects not only the helical region of Pup but also Pup’s N-terminal tail. (Supplementary Fig. 17 and 18)
Figure 3.7 Pup is not degraded in the presence of the Mpa-S-Open gate proteasome complex. A. Overlay of in-cell $^{15}$N-HSQC spectra of free Pup (red) titrated with Mpa-S Open gate proteasome complex (black). B. Overlay of $^{15}$N-HSQC spectra of free Pup (red) and Pup bound to Mpa-S after overexpressing Mpa-S (Protocol 4). Panel A is indicative of equilibrium between free and bound Pup where residues between S21 and Q64 reflect the greatest chemical shift changes/intensity changes. B. Shows complete interaction remains between Pup and Mpa-S Open gate proteasome complex and Pup is not degraded. C. Sequence of Pup showing greatest intensity changes upon titration in red.
Thus, Western Blotting revealed that Pup was not degraded in either case varying both order of expression between Mpa-proteasome complex and Pup or by using either WT or Opegate proteasome. The STINT NMR results along coupled with western blotting show strong evidence that Pup’s N-terminus is engaged by the Mpa-Proteasome complex but is not degraded.

Recreation of the Pup proteasome pathway in-vitro reveals that Pup is retained and not degraded.

Proteins that are targeted for degradation in eukaryotes are tagged with Ubiquitin, a small (76 aa), highly conserved regulatory protein. One of the key characteristics of this pathway is the highly conserved mechanism of how Ubiquitin is recycled. It was suggested by (Weber-Ban et. al 2010) that unlike the Ubiquitin proteasome pathway Pup-GGE is degraded both when coupled to target proteins and when uncoupled. To confirm that Pup-GGQ remained in solution and was not being degraded we recreated the Pup degradation assay as reported by (Weber-Ban et. al 2010) using WT and Open gate proteasomes. (Supplementary Fig. 8)
**Figure 3.8** Pup-GGQ is not degraded in the presence of the Mpa-Proteasome complex in vitro. A. Western blot of the degradation assay using Mpa-S (0.3 µM), Open gate proteasome (0.3 µM) with Pup-His$_{7x}$ (10 µM) in the presence of 5mM ATP. PrcB is the proteasomal β-subunit, PreB is the unprocessed proteasomal β-subunit. B. *In-vitro* degradation assay using Mpa-S (0.3µM), wild type proteasome (0.3µM) with Pup-His$_{7x}$ (10 µM) in the presence of 5mM ATP. Pup is not shown to be degraded in either Panels A. and B. suggesting Pup may be recycled instead of degraded.
Figure 3.9 Pup-GGQ is not degraded in the presence of the Mpa-S-Proteasome complex *in vitro* upon varying the concentrations of complex. A. Western blot analysis of the degradation assay using Mpa-S (.10µM), Open gate proteasome (0.3µM) and Pup-His$_{7x}$(15 µM) lanes 1-5, followed by degradation assay using Mpa-S (.15µM), wild-typeproteasome (0.20µM) and Pup-His$_{7x}$(15 µM) lanes 6-10. B. Western blot analysis of the degradation assay using Mpa-S (.40µM), Open gate proteasome (0.6µM) and Pup-His$_{7x}$(10 µM), lanes 1-5 followed by degradation assay using Mpa-S (.40µM), wild type proteasome, (0.20µM) and Pup-His$_{7x}$(15 µM). No degradation of Pup GGQ was evident in any experiment.
To verify that purified Mtb proteasome is active we used a known proteasomal substrate, Suc-LLVY-Amc, that fluoresces upon cleavage (Fig 3.10). To confirm that the increase in fluorescence was due to proteasomal cleavage and not from contaminating proteolytic activity we quench the proteasomal activity with Bortezomib, a known specific proteasome inhibitor which forms a tetrahedral adduct over catalytic Thr10 in the prcB subunit. Indeed, in the presence of Bortezomib fluorescence emission over a period of time was drastically reduced (Fig 3.11).

Figure 3.10 The Open gate and wild type proteasome show activity when mixed with 7-amino-4-methylcoumarin (AMC). This peptide is a fluorogenic substrate for the 20S proteasome. Cleavage of the AMC peptide by the proteasome generates strong fluorescence. Proteasomes were incubated at 37 °C with 20 μM Suc-LLVY-Amc using a 96 well black plate in a fluorescence spectrophotometer with continuous stirring. The release of Amc was monitored by fluorescence at $\lambda$ ex = 360 nm, $\lambda$ em =440 nm. Both proteasomes reactions show increase in fluorescence after 1 hr incubation indicating proteasome activity.
Figure 3.11. Proteasome activity is inhibited by the addition of Bortzeomib (BTZ). Inhibition of the proteasome was accomplished by adding (1.0 nmol) of Bortzeomib to each reaction. Fluorescence was monitored using a 96 well black plate in a fluorescence spectrophotometer with continuous stirring. The release of Amc was monitored by fluorescence, \( \lambda_{ex} = 360 \text{ nm}, \lambda_{em} = 440 \text{ nm} \). Reactions showed a decrease in fluorescence upon addition of Bortzeomib confirming that the increase in fluorescence was due to proteasomal activity. Initial fluorescence reaction is shown in blue while inhibited fluorescence is shown in red.

Upon initiation of the degradation reaction of Pup-GGQ by Mpa-proteasome with 5mM ATP, samples were then taken and quenched with SDS loading buffer over the course of enzymatic reaction (Fig 3.9). We used both WT and Open gate proteasome and yet observed no degradation of Pup even after an extended period of time. This in-vitro result is consistent with our in-vivo results obtained for E. coli reconstituted proteasomal system.
Discussion

Assembly of macromolecular machinery in the presence of its ligands in the crowded cytosol presents a complicated system to be studied by atomic resolution techniques. We used STINT NMR to map interactions of prokaryotic ubiquitin-like protein Pup with mycobacterial proteasomal machinery expressed in *E. coli*. A functional mycobacterium pupylation pathway has been successfully reconstructed in *E.coli* (Cerda et al 2011). *E. coli* crowded cytosol provides a prokaryotic environment for structural study of the proteasomal function without complications of additional factors that may specifically interact with this system.

In-vitro studies showed that Pup is a disordered protein possessing a C terminal region of transient helical structure. As in the case of alpha-synuclein, physiological conditions may result in seemingly disordered protein acquiring stable secondary and even tertiary structure. We observed only minor changes in the in-cell NMR spectrum of Pup as compared to its spectrum in the lysate. This result suggests that Pup does not possess a stable secondary structure in the cytosol. Since Pup works as an anchor for proteasome system, with N-terminus assuming an extended structure [188], Pup’s disorder may be important for its function.

We demonstrated that Pup interacts with hexameric proteasomal ATPase, Mpa, intracellularly and determined that the interaction surface of Pup does not vary with order of expression between Pup and Mpa. This in-cell result was expected since Mpa subunits bind at least three order of magnitude tighter to each other than Pup. Consistent with in-vitro studies, Pup binds to Mpa mostly via its C-terminal helical region. We also identified additional short region of N-terminal tail Arg8-G12 affected by Mpa binding. *In-vitro* titration of Mpa into [U-, 15N] Pup resulted in gradual uniform broadening of the peaks from residues S21 to Q64[188]. Dissociation constant of Mpa-Pup was determined to be submicromolar. We did not observe
gradual decrease in peak intensities for all the residues in the helix. The result is consistent with Pup being in exchange between Mpa and free in the cytosol. Since the intracellular concentration of Pup was at least 10 uM, intracellular Pup should be predominantly bound to Mpa. This result suggests that the presence of cytosol decreased the apparent affinity of Pup for Mpa, possibly due to decreasing non-specific interactions between Pup and Mpa.

In-cell NMR spectrum of $[^{15}\text{N}]$ Pup expressed first and Mpa-proteasome second and Mpa-proteasome first and $[^{15}\text{N}]$ Pup second are clearly different. Since proteasome has to be matured in the cell for 12 hours, at the end of $[^{15}\text{N}]$ Pup first and Mpa-proteasome second sequential expression we will have a mixed population of the cells containing either Pup or Pup and Mpa, or Pup Mpa and proteasome. NMR signal will be an average of these spectra. In this case, the spectrum of $[^{15}\text{N}]$ Pup first and Mpa-proteasome second is very similar to that of $[^{15}\text{N}]$ Pup in complex with Mpa presenting the interaction between Mpa and helical region of Pup. When Mpa-proteasome was expressed first and $[^{15}\text{N}]$ Pup second we have cells containing both $[^{15}\text{N}]$ Pup and Mpa and proteasome. In this case, $[^{15}\text{N}]$ Pup spectrum was completely broadened. This was not due to Pup degradation since we detected Pup in Mpa-proteasome expressing cells by using the Western blot.
Figure 3.11B STINT-NMR reveals an equilibrium of events. In-cell NMR spectrum of $[U\cdot^{15}N]$ Pup expressed first and Mpa-proteasome second and Mpa-proteasome first and $[U\cdot^{15}N]$ Pup second are clearly different. Since proteasome has to be matured in the cell for 12 hours, at the end of $[U\cdot^{15}N]$ Pup first and Mpa-proteasome second sequential expression we will have a mixed population of the cells containing either Pup or Pup and Mpa, or Pup Mpa and proteasome. NMR signal will be an average of these spectra. In the case were Pup is expressed second no maturation time is needed and we seen only type of cell where the NMR signal is arising.
In the cells expressing Pup, Mpa, and proteasome, the amount of proteasome is significantly lower than Mpa or Pup. In this case, Mpa-Pup complex will be a predominant species. Nevertheless, the in-cell NMR spectrum of Mpa-Pup complex in the cells expressing non-stoichiometric amounts of proteasome is different from that of the cells expressing only Mpa and Pup; the presence of proteasome results in the complete broadening of the peaks from N-terminal tail of Pup. Pup-Mpa complex appears to be stabilized by the presence of non-stoichiometric amount of proteasome. Since proteasome binds to Mpa with low affinity, we postulate that transient binding of proteasome to Pup-Mpa complex results in the N-terminal tail of Pup to be occluded in the Mpa central cavity leading to the complete broadening of Pup’s peaks.
Figure 3.12 Pup gets degraded by the Mpa-proteasome only upon deamidation. Pup-GGQ gets deamindated by Dop in the presence of ATP and becomes engaged by the Mpa-proteasome complex and in turn degraded. Pup-GGE gets engaged by the Mpa-proteasome but is not degraded indicative that the deamidated Pup is required as a precursor for degradation.

Unlike Ubiquitin, which is recycled in the eukaryotic proteasome, Pup-GGE either alone or upon conjugation to a substrate is degraded and not recycled in mycobacterium. Pup-GGQ is a precursor molecule, which is converted to Pup-GGE before being ligated to a substrate destined to degradation (Fig 3.12). Our STINT-NMR experiments present a dynamic picture of the fate of Pup-GGQ inside a cell containing Mpa and core proteasome: In the absence of proteasome, C-terminal and N-terminal regions of Pup-GGQ interacts with Mpa only weakly; non-specific
interactions are blocked by the cytosol. In the presence of a non-stoichiometric amount of proteasome, Pup-GGQ is completely bound to Mpa with C terminus binding to the mouth of hexameric Mpa and N-terminus falling into Mpa’s central cavity. Unlike Pup-GGE, Pup-GGQ is a not a substrate for the proteasomal degradation, consistent with its function as a precursor molecule. The charge difference of glutamine to glutamate in Pup possibly acts as an anchor, which interrupts its occlusion to the proteasome catalytic chamber. Reconstruction of the interactions between the mycobacterial Mpa-proteasome complex and Pup inside a cell at atomic resolution allowed us to look at intracellular processes, which are not accessible by *in-vitro* investigations.
Supplementary Information:

Signal Transduction in RAGE: Solution Structure of C-terminal RAGE (ctRAGE) and its Binding to mDia1.

Supplementary Figure 1. ctRAGE does not interact with a model membrane formed by DMPC/CHAPSO bicelles. Overlay of the $^{15}$N-HSQC spectra of 300 μM [$^{15}$N] ctRAGE (red) in 0.5 % DMPC/CHAPSO bicelles dissolved in NMR buffer (10 mM potassium phosphate [pH 6.5], 100 mM NaCl, 0.02% (w/v) NaN₃, in 100%D₂O) and 500 μM [$^{15}$N] ctRAGE (black) in NMR buffer. No peaks changes were observed. The peaks at 8.06 ppm in proton dimension are from the unedited CHAPSO amide. The spectra were acquired on a 500 MHz Bruker Avance III NMR spectrometer at 298K.
Supplementary Figure 2. The tertiary structure of ctRAGE is stabilized by electrostatic interactions between R5, E10, E11 and R12. Strip slices from $^{15}$N-1H NOESY spectrum of $[U-^{15}$N] ctRAGE that shows the through space proton-proton connectivities of the amide protons of Q3, R5, E10, E11 and R12. The connectivities suggest the presence of a stable ctRAGE structure. The spectrum was collected on a Bruker Avance II 700 MHz NMR spectrometer at 298 K. The data were processed using Topspin 2.1 and analyzed by CARA software.
Supplementary Figure 3. Stereoview of the ctRAGE structure. Backbone is shown by using ribbons. Residues are labeled at alpha carbons. Stereoview was constructed by using Molmol (Koradi et al, 1996).
Supplementary Figure 4. The R5AQ6A mutation leads to significant structural changes in the folded region of ctRAGE. A. Overlay of the $^{15}$N-HSQC spectra of [U-$^{15}$N] wild type (black) and R5A6QA ctRAGE (red). B. Chemical shift changes between wild type and R5AQ6A ctRAGE amide protons, $\delta_{H1}$ and $\delta_{H2}$, and nitrogens, $\delta_{N1}$ and $\delta_{N2}$ respectively, are calculated as $\sqrt{(\delta_{H1} - \delta_{H2})^2 + (\delta_{N1} - \delta_{N2})^2 / 4}$. Notice that the changes affect residues that are located away from the mutation site. Changes in the chemical shifts of R5 and Q6, which are absent in R5AQ6A ctRAGE, were set to 1 ppm for ease of viewing.
Supplementary Figure 5. R5AQ6A ctRAGE does not form a complex with mDia1 FH1. Overlay of the $^{15}$N-HSQC spectra of free [$U^{-15}$N] R5AQ6A ctRAGE (black) and the FH1-[$U^{-15}$N] ctRAGE complex (red). The spectra were collected on a Bruker Avance III 500 MHz NMR spectrometer at 298K. To form the mDia1-R5AQ6A ctRAGE complex, 0.5 mM of unlabeled mDia1 FH1, in NMR buffer (10 mM potassium phosphate [pH 6.5], 100 mM NaCl, 0.02% (w/v) NaN$_3$, in 90%/10% H$_2$O/D$_2$O) was added into 100 μM [$U^{-15}$N] R5AQ6A ctRAGE to yield ctRAGE:mDia1 FH1 molar ratios of 1:2. There are no significant changes in the peak positions or intensities of the amide protons and nitrogens of [$U^{-15}$N] R5AQ6A ctRAGE. This result strongly suggests that mDia1 does not interact with R5AQ6A ctRAGE.
Supplementary Figure 6. Wild-type primary murine aortic smooth muscle cells were subjected to electroporation with construct encoding GFP. In other sections (lane 1), DAPI was used to visualize the nuclei; lane 2 reveals GFG+ SMCs and lane 3 represents the merged image. From these studies, an estimated 65-70% transfection efficiency of primary SMCs was achieved.
Supplementary Figure 7. Vector- and double mutant transfected primary murine aortic smooth muscle cells were treated with RAGE ligand S100B (10 µg/ml) or non-RAGE ligand PDGF (10 ng/ml) and tritiated thymidine incorporation was measured as described in Materials and Methods. The results represent the mean ± SD of the mean and are representative of triplicate experiments. * indicates p<0.05.
Supplementary Information:

Fate of Pup during *Mtb* proteasome proteolysis as studied by in-cell NMR spectroscopy.

**Supplementary Figure 8.** Pup-GGQ is not degraded in the presence of the Mpa-T-Proteasome complex in vitro. A. In vitro degradation assay using Mpa-T (0.3μM), Open gate proteasome (0.3μM) with Pup-His (20 μM) in the presence of 5mM ATP as monitored by Coomassie staining. PrCB is the proteasomal β-subunit, PreB is unprocessed proteasomal β-subunit. B. Invitro degradation assay using Mpa-S (0.3μM), wild type proteasome (0.3μM) with Pup-His (10 μM) in the presence of 5mM ATP as monitored by Coomassie staining. Pup-GGQ is not shown to be degraded in both A. and B. suggesting Pup may be recycled instead of degraded.
Supplementary Figure 9. Chemical shift changes in the $^{15}$N-HSQC spectra of in-cell relative to in-vitro Pup. During structure comparison experiments and in-cell titration experiments, we measured the change in the chemical shifts of amide nitrogens and covalently attached amide protons according to the equation $\Delta = \sqrt{\delta_H^2 + (0.25)\delta_N^2}$, where $\delta_{H(N)}$ represents a change in hydrogen and nitrogen chemical shifts. Based on this comparison of chemical shift we conclude that in-cell Pup does not contain any structure induced by macromolecular crowding.

Supplementary Figure 10. Intensity difference between the $^{15}$N-HSQC spectra of in-cell and $^{15}$N-HSQC of in vitro Pup. Intensities were calculated by integrating each peak relative to cytosol protein peaks and using, $I = (V_{in\ cell} - V_{in\ vitro})/V_{in\ cell}$, where $V_{in\ vivo}$ is the volume of an individual peak in the $^{15}$N-HSQC in cell spectrum of Pup and $V_{in\ vitro}$ is the volume of individual peaks on the $^{15}$N-HSQC of in-cell Pup. Positive intensities reflect the sharpening of due to the decrease in viscosity of the lysate between spectra.
Supplementary Figure 11. Mpa-S binds to two distinct regions on Pup upon titration in-cell. Chemical shift changes in the $^{15}$N-HSQC spectra of Pup with Pup titrated with Mpa-S. The change in the chemical shifts of amide nitrogen and covalently attached amide protons was calculated according to the equation $\Delta = \sqrt{\delta_H^2 + (0.25)\delta_N^2}$, where $\delta_{HN}$ represents a change in hydrogen and nitrogen chemical shifts. Regions R8-G12 and S21-Q60 are the most perturbed residues due to Mpa-S binding. Region S21-Ala51 reflects the induced $\alpha$-helix of Pup arising from binding to Mpa-S.

Supplementary Figure 12. Intensity difference in the $^{15}$N-HSQC spectra of in-cell free Pup with $^{15}$N-HSQC spectra of Pup titrated with Mpa-S in-cell. Intensities were calculated by integrating the volume of each residue and using, $I = (V_{Free} - V_{Titrated})/V_{Titrated}$, where $V_{Free}$ is the volume of an individual peak on the $^{15}$N-HSQC unbound spectrum of Pup. $V_{Titrated}$ is the volume of individual peaks on the $^{15}$N-HSQC of the titrated spectrum. Positive intensities denote residues decreasing in intensity. Regions E4-R8 and D38-Q64 shows the greatest increase in intensity due to Mpa-S binding.
Supplementary Figure 13. Mpa-S binds to two distinct regions on Pup upon titration in vivo. Chemical shift changes in the $^{15}$N-HSQC spectra of free Pup with Pup titrated with Mpa-S. The change in the chemical shifts of amide nitrogen and covalently attached amide protons was calculated according to the equation, $\Delta = \sqrt{\delta_H^2 + (0.25)\delta_N^2}$, where $\delta_{\text{HN}}$ represents a change in hydrogen and nitrogen chemical shifts. Regions R8-G12 and S21-Q60 reflect most perturbed residues due to Mpa-S binding. Region S21-Ala51 reflects the induced $\alpha$-helix of Pup arising from binding to Mpa-S.

Supplementary Figure 14. Intensity difference in the $^{15}$N-HSQC spectra of free Pup with $^{15}$N-HSQC spectra of Pup over expressed second post Mpa-S over expression in-cell. Intensities were calculated by integrating the volume of each residue and using, $I = (V_{\text{Free}} - V_{\text{Titrated}})/V_{\text{Titrated}}$, where $V_{\text{Free}}$ is volume of an individual peak on the $^{15}$N-HSQC unbound spectrum of Pup and $V_{\text{Titrated}}$ is the volume of individual peaks on the $^{15}$N-HSQC of the titrated spectrum. Two regions upon titration increase in intensity due to Mpa-S binding, E4-R8 and residues, V46, N50, E52, F54 Y58, G64, toward the C terminus of Pup.
Supplementary Figure 15. Mpa-S-Open gate complex interacts with the C-terminus of Pup. Chemical shift changes of the $^{15}$N-HSQC spectra of free Pup with the $^{15}$N-HSQC spectra of Pup titrated with Mpa-S-Open gate complex second in-cell. The change in the chemical shifts of amide nitrogen and covalently attached amide protons was calculated according to the equation, 

$$
\Delta = \sqrt{\delta_H^2 + (0.25)\delta_N^2},
$$

where $\delta_{\text{HN}}$ represents a change in hydrogen and nitrogen chemical shifts. Shifts in the region of D38-Q60 is indicative of Pup binding to the Mpa-S-Open gate complex within the $\alpha$-helix region of Pup induced by Mpa-S binding interaction.

Supplementary Figure 16. Intensity difference in the $^{15}$N-HSQC spectra of free Pup with $^{15}$N-HSQC spectra of Pup expressed first followed by over expression of the Mpa-S-Open gate complex. Intensities were calculated by integrating the volume of each residue and using $I = (V_{\text{Free}} - V_{\text{Titrated}})/V_{\text{Titrated}}$, where $V_{\text{Free}}$ is volume of an individual peak on the $^{15}$N-HSQC unbound spectrum of Pup and $V_{\text{Titrated}}$ is the volume of individual peaks on the $^{15}$N-HSQC of the titrated spectrum. Pup shows strong intensities increases within its C-terminus and the Mpa-S-Open gate complex, however the -terminus seems to be left unengaged when Pup is overexpressed first followed by Mpa-S-Open gate complex expression.
Supplementary Figure 17. Mpa-S-Open gate complex interacts with the C-terminus of Pup. Chemical shift changes of the $^{15}$N-HSQC spectra of free Pup with the $^{15}$N-HSQC spectra of Pup titrated with Mpa-S-Open gate complex in-cell. The change in the chemical shifts of amide nitrogen and covalently attached amide protons was calculated according to the equation, 

$$\Delta = \sqrt{\delta_H^2 + (0.25)\delta_N^2},$$

where $\delta_{HN}$ represents a change in hydrogen and nitrogen chemical shifts.

Supplementary Figure 18. Expression of Mpa-S-Open gate complex first followed by Pup induction second is indicative of complete interaction of Pup. Intensity difference between the $^{15}$N-HSQC spectra of free Pup with $^{15}$N-HSQC spectra of Pup expressed second in the presence of overexpressed Mpa-S-Open gate complex. Intensities were calculated by integrating the volume of each residue and using, $I = (V_{Free} - V_{Titrated}) / V_{Titrated}$. where $V_{Free}$ is volume of an individual peak on the $^{15}$N-HSQC unbound spectrum of Pup and $V_{Titrated}$ is the volume of individual peaks on the $^{15}$N-HSQC of the titrated spectrum.
Supplementary Figure 19. Pup’s N-terminus interacts with the Mpa-S-wild type proteasome complex. Chemical shift changes in the $^{15}$N-HSQC spectra of free Pup with Pup titrated with Mpa-S-wild-type complex second in-cell. The change in the chemical shifts of amide nitrogen and covalently attached amide protons was calculated according to the equation,

$$
\Delta = \sqrt{\delta_H^2 + (0.25)\delta_N^2},
$$

where $\delta_{HN}$ represents a change in hydrogen and nitrogen chemical shifts. The most pronounced shifts were found in region D38-V46, indicative of interaction of Pup’s $\alpha$-helical region with the complex.

Supplementary Figure 20. Expression of Mpa-wild type proteasome complex second indicates only partial interaction with Pup. Intensities were calculated by integrating the volume of each residue and using, $I = (V_{\text{Free}} - V_{\text{Titrated}})/V_{\text{Titrated}}$, where $V_{\text{Free}}$ is volume of an individual peak on the $^{15}$N-HSQC free spectrum of Pup and $V_{\text{Titrated}}$ is the volume of individual peaks on the $^{15}$N-HSQC of the titrated spectrum. Intensities changes were greatest at the N-terminus between residues D38- E40. Partial binding of Pup reflects and equilibrium between free and bound Pup.
Supplementary Figure 21. Pup shows increased surface interaction in the presence of the Mpa-S-wild type proteasome complex when complex is expressed second in-cell. Comparison of the $^{15}$N-HSQC spectra of free Pup with the $^{15}$N-HSQC spectra of Pup expressed second in the presence of the Mpa-S-wild-typeproteasome. The change in the chemical shifts of amide nitrogen and covalently attached amide protons was calculated according to the equation, \[ \Delta = \sqrt{\delta_H^2 + (0.25)\delta_N^2}, \] where \( \delta_{HN} \) represents a change in hydrogen and nitrogen chemical shifts. Regions in both the C-terminus D14-Q26 and the N-terminus D38-V46 contain the most pronounce chemical shift changes.

Supplementary Figure 22. Expression of Mpa-S-wild type proteasome complex second is indicative of Pup’s full interaction with the Mpa-S-wild type complex. $^{15}$N-HSQC spectra of Pup is obliterated in the presence of Mpa-S—wild type proteasome complex when complex is overexpressed first in-cell. Intensities were calculated by integrating the volume of each residue and using, \[ I = (V_{\text{Free}} - V_{\text{Titrated}})/V_{\text{Titrated}}, \] where $V_{\text{Free}}$ is volume of an individual peak on the $^{15}$N-HSQC unbound spectrum of Pup and $V_{\text{Titrated}}$ is the volume of individual peaks on the $^{15}$N-HSQC of the titrated spectrum.
References


[27] G. Cornilescu, F. Delaglio, A. Bax, Protein backbone angle restraints from searching a database for chemical shift and sequence homology, J Biomol NMR, 13 (1999) 289-302.


[168] C.D. Collier, J.R. Johnson, The Escherichia coli K-12 metJ193 allele contains a point mutation which alters the hydrophobic pocket responsible for in vitro binding of S-


