New tools to study amyloid fibrils and intrinsically disordered proteins in vitro and in vivo

Jacqueline D. Washington

University at Albany, State University of New York, jwashington22@yahoo.com

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

Recommended Citation
https://scholarsarchive.library.albany.edu/legacy-etd/1040

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.
New Tools to Study Amyloid Fibrils and Intrinsically Disordered Proteins *In Vitro* and *In Vivo*

by

Jacqueline D. Washington

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

2013
Acknowledgements

My doctoral dissertation consists of several projects and focuses on the use of high resolution techniques and molecular biology methods, to understand structural interactions in vitro and in vivo. The projects presented in this dissertation are an accumulation of the completed work done during my tenure in the Shekhtman Research Lab within the Department of Chemistry. Although I am the sole author of this dissertation, this work would not have been possible without the patience and guidance of many.

I would like to thank Professor Alexander Shekhtman for his guidance, patience and support throughout this experience. His passion for science has made me into the scientist I am today. His dedication often motivated me to work harder to reach my projected goals even when faced with adversity. I would like to thank Dr. Sergey Reverdatto for his patience and the sharing of his knowledge of Biochemistry and Molecular Biology. I would be remiss if I didn’t thank Dr. Andres Y. Maldonado for his friendship and all of his guidance. I don’t know how I would have managed if not for his presence. Much of this work would have not been possible without Dr. David Burz, who has provided wonderful knowledge and suggestions.

I would also like to thank my group members, especially Rachael Hartz for her assistance and friendship. I would like to thank all the graduate students in the Department of Chemistry who have encouraged me and lent a listening ear from time to time, especially NaKesha Smith.

Last, but certainly not least, I would like to thank my family for the love and support throughout my academic career. I would like to thank my parents, Frances and
Charles Washington, whom I love so much. They always encouraged me to get an education and taught me to not be afraid and follow my heart no matter where it takes me. To Matthew Cobbert, who helped me get to this point with encouragement and support, thank you and I love you. And to all the Maxwell and Washington family members, thank you for your love and support. I love all y’all!
Abstract

Amyloid fibrils are β-sheet-rich protein aggregates commonly found in the organs and tissues of patients with various amyloid-associated diseases. The structure of insulin fibrils was characterized by deep ultraviolet resonance Raman (DUVRR) and Nuclear Magnetic Resonance (NMR) spectroscopy combined with hydrogen-deuterium exchange. Our new approach of combining NMR and Raman spectroscopy with molecular dynamic simulations for characterizing amyloid fibrils provided exclusive knowledge about fibril structure at amino acid residue resolution.

We review the basics of in-cell NMR spectroscopy with emphasis on applications that are relevant to the structure-based drug discovery process. Applications such as determining three-dimensional (3D) structures de novo, mapping interaction surfaces, and measuring changes in protein structure due to post-translational biochemical modifications, folding states, metabolic processing and protein maturation under physiological or near physiological conditions. The latest methods for incorporating target molecules into the intracellular milieu at sufficient concentrations for NMR spectroscopy are detailed along with the physiological relevance and pitfalls of each technique. Lastly, specific examples of drug-protein interactions identified by using in-cell NMR are presented.

Intrinsically disordered proteins (IDPs) or unstructured segments within proteins play an important role in cellular physiology and pathology. We use a combination of peptide aptamers selected by using the yeast-two-hybrid scheme and in-cell NMR to identify high affinity binders to transiently structured IDP and unstructured segments at atomic resolution. The method is validated by using peptide aptamers selected against the
prokaryotic ubiquitin-like protein, Pup (an intrinsically disordered protein), of the mycobacterium proteasome. We discover two peptide aptamers that bind to opposite sites of a transient helix in Pup that have vastly different effects on the survival of *Mycobacterium bovis* (*M. bovis*) BCG.

Drug discovery is based on the ability to specifically enhance or impede biological activity by binding to a single target protein or complex. We use SMILI-NMR along with STINT-NMR to screen a library for small drug molecules that impede the interaction between Pup and Mpa, an interaction that is essential for Mtb proliferation. We identify 12 compounds similar to previously discovered Mtb drugs and test them in a Mtb functional assay to analyze their effects on cell growth.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-SYN</td>
<td>α-Synuclein</td>
</tr>
<tr>
<td>AD</td>
<td>Alzheimer’s Disease</td>
</tr>
<tr>
<td>GAL4-AD</td>
<td>GAL4 Activating Domain</td>
</tr>
<tr>
<td>AFM</td>
<td>Atomic Force Microscopy</td>
</tr>
<tr>
<td>AUI</td>
<td>Ataxin 3</td>
</tr>
<tr>
<td>AUR A</td>
<td>Aureobasidin A</td>
</tr>
<tr>
<td>BCG</td>
<td>Bacille Calmette-Guèrin</td>
</tr>
<tr>
<td>BTZ</td>
<td>Bortezomib</td>
</tr>
<tr>
<td>CFU</td>
<td>Colony Forming Units</td>
</tr>
<tr>
<td>CI2</td>
<td>Chymotrypsin Inhibitor 2</td>
</tr>
<tr>
<td>CLIPs</td>
<td>Combinatorial Library of Improved Peptide aptamers</td>
</tr>
<tr>
<td>CPPTat</td>
<td>Tat Protein of HIV-1</td>
</tr>
<tr>
<td>CV</td>
<td>column volume</td>
</tr>
<tr>
<td>DETA-NO</td>
<td>2,2-(Hydroxynitrosohydrazino)-Bis-Ethanamine</td>
</tr>
<tr>
<td>DLS</td>
<td>Dynamic Light Scattering</td>
</tr>
<tr>
<td>DMF</td>
<td>Dimethylformamide</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl Sulfoxide</td>
</tr>
<tr>
<td>DNA-BD</td>
<td>GAL4 DNA Bindnig Domain</td>
</tr>
<tr>
<td>DOP</td>
<td>Deaminase/Depypylase of Prokaryotic Ubiquitin-like Protein</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>DUBs</td>
<td>Ubiquitin-Specific C-terminal Proteases</td>
</tr>
<tr>
<td>DUVRR</td>
<td>Deep Ultraviolet Resonance Raman Spectroscopy</td>
</tr>
<tr>
<td>fALS</td>
<td>Familial Amyotrophic Lateral Sclerosis</td>
</tr>
<tr>
<td>FITC</td>
<td>Fluorescein Isothiocyanate</td>
</tr>
<tr>
<td>FKBP</td>
<td>FK506 Binding Protein</td>
</tr>
<tr>
<td>FRB</td>
<td>FKBP-Rapamycin-Binding domain of mTOR</td>
</tr>
<tr>
<td>GB1</td>
<td>Streptococcal Protein G B1-domain</td>
</tr>
<tr>
<td>GL5</td>
<td>5-(5-methyl-2-(methylthio)thiophen-3-yl)-1,3,4-oxathiazol-2-one</td>
</tr>
<tr>
<td>HA</td>
<td>Hemagglutinin</td>
</tr>
<tr>
<td>HD</td>
<td>Huntington’s Disease</td>
</tr>
<tr>
<td>H/D</td>
<td>Hydrogen/Deuterium Exchange</td>
</tr>
<tr>
<td>HRP</td>
<td>Horseradish peroxidase</td>
</tr>
<tr>
<td>HRS</td>
<td>Hepatocyte Growth Factor-regulated Tyrosine Kinase Substrate</td>
</tr>
<tr>
<td>hSOD1</td>
<td>Zinc Superoxide Dismutase 1</td>
</tr>
<tr>
<td>HSQC</td>
<td>Heteronuclear Single Quantum Coherence</td>
</tr>
<tr>
<td>HTS</td>
<td>High Throughput Screening</td>
</tr>
<tr>
<td>IDPs</td>
<td>Intrinsically Disordered Proteins</td>
</tr>
<tr>
<td>IgG</td>
<td>Immunoglobulin G</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl 1-Thio-β-D-Galactopyranoside</td>
</tr>
<tr>
<td>M. bovis</td>
<td>Mycobacterium bovis</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamics</td>
</tr>
<tr>
<td>MLSMR</td>
<td>Molecular Libraries Small Molecule Respository</td>
</tr>
<tr>
<td>MPA</td>
<td>Proteosomal ATPase</td>
</tr>
<tr>
<td>Msm</td>
<td>Mycobacterium smegmatis</td>
</tr>
</tbody>
</table>
Mtb  Mycobacterium tuberculosis  
MWCO  Molecular Weight Cutoff  
NIH/NCI  National Institutes of Health/National Cancer Institute  
Ni-NTA  Nickel-Nitrilotriacetic Acid  
NMR  Nuclear Magnetic Resonance Spectroscopy  
NOESY  Nuclear Overhauser Enhancement Spectroscopy  
PAFA  Proteasome Accessory Factor A  
PA  Peptide Aptamer  
PAGE  Polyacrylamide Gel  
PBCs  Periodic Boundary Conditions  
PBS  Phosphate Buffered Saline  
PD  Parkinson’s Disease  
PI  Propidium Iodide  
PME  Particle-Mesh Ewald  
PTMs  Post-Translational Modifications  
Pup  Prokaryotic Ubiquitin-like Protein  
SDM  Synthetic Dropout Media  
SDS-PAGE  Sodium Dodecyl Sulfate Polyacrylamide Gel  
SLO  Streptolysin O  
SMILI-NMR  Screening Small molecule Interactor Library using In-cell NMR  
SOFAST-HMQC  Selective Optimized Flip-Angle Short-Transient heteronuclear multiple quantum coherence  
SPC  Single Point Charge  
STAM2  Signal Transducing Adapter Molecule 2  
STINT-NMR  In-cell NMR-based method for mapping the STructural INTeractions  
Tβ4  Thymosin β4  
TCEP  tris(2-carboxyethyl) phosphine  
TFA  Trifluoroacetic Acid  
Thio  Thioredoxin  
TOCSY  Total Correlation Spectroscopy  
trpL  Tryptophan Leader Sequence  
UASs  Upstream Activating Sequences  
UBQ  Ubiquitin  
UBQ-CPP_Tat  Ubiquitin contained a C-terminal CPP fusion protein  
UCH  Ubiquitin Carboxyl-Terminal Hydrolyses  
UIM  Ubiquitin Interacting Motif  
UIPs  Ubiquitin-Interacting Proteins  
Y2H  Yeast-Two-Hybrid  
YPDA  Yeast Peptone Dextrose Adenine  
X-α-gal  X-α-galactosidase
List of Figures and Tables

Figure 1.1 AFM images of insulin fibrils (A) before H/D exchange (B) and two days after incubation in D₂O, pD* 1.9 at 25 °C.
Figure 1.2 DLS data of insulin fibrils under different solvent conditions.
Figure 1.3 The majority of H/D exchange-protected amino acids are located in the B-chain of insulin.
Figure 1.4 Graph illustrating that the protection degree of each amino acid residue in the insulin monomer.
Figure 1.5 Disulfide bonds preserve their conformation upon insulin fibrillation.
Figure 1.6 Backbone conformations of fibrillar and monomer insulin are very similar in the vicinity of disulfide bridges.
Figure 1.7 The secondary structure of insulin changes dramatically from being mostly α-helical in the native protein to highly β-sheet-rich in the fibrillar form.
Figure 1.8 Equilibrated structure of an insulin dimer from 150-ns MD simulations in aqueous solution.
Figure 1.9 Snapshots of insulin dimers from MD simulations.
Figure 2.1 Interaction surface maps of Ubiquitin-ligand complexes.
Figure 2.2 Ubiquitin-ligand complexes in E. coli.
Figure 2.3 Atox1-cisplatin complexes in E. coli.
Figure 2.4 (Top) Ternary FKBP-FRB-ligand complexes in E. coli. (Bottom) Yeast assay for biological activity of the dipeptide, A-E.
Figure 3.1 Plasmids Used for PA Selection and Analysis.
Figure 3.2 Compounds used in BCG assay.
Figure 3.3 Overview of Yeast-Two-Hybrid Systems.
Figure 3.4 Peptide Aptamer Sequences identified during Y2H Screen and Isotherms from Fluorescence Anisotropy.
Figure 3.5 PA-1 and PA-3 affects Pup in-cell.
Figure 3.6 Chemical shifts differences and differential peak broadening from in-cell interactions between PA-1 and Pup.
Figure 3.7 Chemical shifts differences and differential peak broadening from in-cell interactions between PA-3 and Pup.
Figure 3.8 PA-1 and PA-3 bind to opposite sides of the Pup helix.
Figure 3.9 Only few Pup residues are affected by the interaction with PA-7.
Figure 3.10 Chemical shifts differences and differential peak broadening from in-cell interactions between PA-7 and Pup.
Figure 3.11 Only residues in the PA loop bind Pup.
Figure 3.12 Only the PA-7 loop is involved in Pup binding.
Figure 3.13 Positive Expressions of PAs in BCG.
Figure 3.14 Visible Biological Log Reduction on BCG Assay.
Figure 4.1 ¹H(¹⁵N)-HSQC spectral differences between of free Pup and Pup-Mpa complex.
Figure 4.2 Matrix Method of screening chemical libraries.
Figure 4.3 In-cell NMR spectra of [U-¹⁵N] Pup-Mpa complex changed in presence of NIH mixtures.
Table 4.1 NIH compounds found during first round of screening NIH library against in-cell [U-¹⁵N] Pup-Mpa complex.
TABLE OF CONTENTS

Acknowledgements ........................................................................................................ ii
Abstract ......................................................................................................................... iv
List of Abbreviations ...................................................................................................... vi
List of Figures and Tables ............................................................................................. viii
Preface .......................................................................................................................... 1

CHAPTER ONE
DISULFIDE BRIDGES REMAIN INTACT WHILE NATIVE INSULIN
CONVERTS INTO AMYLOID FIBRILS

Introduction .................................................................................................................... 5
Materials and Methods ................................................................................................ 9
  Fibril Formation ......................................................................................................... 9
  NMR Experiments ..................................................................................................... 9
  Raman Experiments ................................................................................................. 10
  Dynamic Light Scattering ......................................................................................... 11
  Atomic force microscopy ......................................................................................... 11
  Computational Methods .......................................................................................... 12
  Computational Modeling ......................................................................................... 13
Results ........................................................................................................................... 13
  Determination of amino acid protection ................................................................. 13
  The conformations of disulfide bonds ..................................................................... 20
  Secondary structure of the fibril core ....................................................................... 24
  MD simulations of β-sheet-rich dimer formation .................................................... 27
Conclusions .................................................................................................................. 32

CHAPTER TWO
IN-CELL NMR SPECTROSCOPY TO STUDY PROTEIN DRUG-INTERACTIONS

Introduction .................................................................................................................... 36
In-Cell NMR Spectroscopy ........................................................................................ 37
  Cell types .................................................................................................................. 39
Expression and purification of aptamers for in vitro assays ........................................86
Expression and purification of Pup for in vitro assays ................................................88
Fluorescence titrations ..................................................................................................89
Sequential overexpression of $^{15}$N-Pup and aptamers for STINT-NMR .................89
NMR Spectroscopy .......................................................................................................91
M. bovis growth assay .................................................................................................91
Results ..........................................................................................................................93
Selection of peptide aptamers that interact with Pup ....................................................93
Peptide aptamers bind to α-helix and C-terminal region of Pup ..................................97
PA-1 inhibits M. bovis BCG growth .............................................................................108
Discussion ...................................................................................................................112

CHAPTER FOUR
IN-CELL NMR SCREENING (SMILI-NMR) OF NATIONAL CANCER INSTITUTE (NCI) DRUG LIBRARY LEADS TO MTB INHIBITION

Introduction ..................................................................................................................116
Materials and Methods ..............................................................................................118
Mtb strain ......................................................................................................................118
Plasmids Utilized ........................................................................................................118
Sequential overexpression of $^{15}$N-Pup and Mpa .....................................................119
NMR Spectroscopy ......................................................................................................120
Mtb growth assay .......................................................................................................121
Results ..........................................................................................................................123
The formation of Pup-Mpa complex ..........................................................................123
Screening of NIH/NCI Library Identifies Drug Interactors .........................................125
Mtb growth inhibition assay .......................................................................................132
Discussion ...................................................................................................................132

References ..................................................................................................................134
**Preface**

Understanding the structural organization of amyloid fibrils and intrinsically disordered proteins can be beneficial while searching to drugs to successfully treat diseases associated with protein misfolding or IDPs. Thus, developing tools that give insight into the molecular dynamics involved in these interactions is vitally important. Using DUVRR in combination with NMR spectroscopy to characterize amyloid fibrils can offer detailed information about protein folding and fibrillar structure. NMR spectroscopy encompasses many techniques, such as STINT-NMR and SMILI-NMR that can monitor protein-protein and protein-drug interactions *in vitro* and *in vivo*. This dissertation is a combination of works in which new tools are developed to study proteins related to diseases *in vitro* and *in vivo*.

Chapter one is based on our published work entitled “Disulfide Bridges Remain Intact while Native Insulin Converts into Amyloid Fibrils” in PloS ONE and describes our new approach of using DUVRR, NMR spectroscopy combined with hydrogen-deuterium exchange, and molecular dynamics simulations to characterize the structure of insulin amyloid fibrils. We reveal all three disulfide bonds of native insulin remain intact during aggregation process, *i.e.* withstanding scrambling. We show three out of four tyrosine residues were packed into the fibril core, and phenylalanine was located in the unordered parts of insulin fibrils. In addition, using all-atom MD simulations, the disulfide bonds were confirmed to remain intact in the insulin dimer, which mimics the fibrillar form of insulin.

Chapter two is based on our published review entitled “In-Cell NMR Spectroscopy to study protein drug-interactions” as Chapter 4 in New Developments in
NMR No. 2: New Applications of NMR in Drug Discovery and Development. This chapter covers all of the latest methodologies and applications of in-cell NMR. The importance of cell type and viability on in-cell NMR studies is discussed, along with specific isotopic labeling schemes used to detect and resolve in-cell NMR protein resonance peaks. The use of tightly regulated promoters for protein over-expression remains the method of choice for prokaryotic cells; however, new techniques, like microinjection and cell-penetrating peptides, have been developed for delivering target molecules to eukaryotic cells at physiological concentrations. Studies that have illuminated the properties of target proteins relevant to the drug discovery process and in-cell NMR-based strategies to drug discovery are highlighted.

Chapter three is based on our ‘in press” work entitled “Caught in Action: Selecting Peptide Aptamers Against Intrinsically Disordered Proteins in Live Cells” were we use yeast-two-hybrid in combination with STINT-NMR to identify high affinity binders to an IDP. Typically, low cellular concentration, multiple binding partners, frequent posttranslational modifications and the presence of multiple conformations make it difficult to characterize IDP interactions in intact cells. Both the selection and characterization of peptide aptamers take place inside the cell where only physiologically relevant conformations of IDPs are targeted. This technology can be applied to study the elusive action of IDPs under near physiological conditions.

Chapter four describes the direct application of SMILI-NMR and STINT-NMR to screen a large drug library to find drugs that hinder complex formation in-cell. The method entails screening a small-drug library via the matrix method against Pup and Mpa, which forms a complex critical to protein degradation in Mtb, to observe NMR
spectral changes. We found molecules that are similar in structure to known promising Mtb drugs.
CHAPTER ONE

DISULFIDE BRIDGES REMAIN INTACT WHILE NATIVE INSULIN CONVERTS INTO AMYLOID FIBRILS

Dmitry Kurouski¹, Jacqueline Washington¹, Mehmet Ozbil², Rajeev Prabhakar²,
Alexander Shehtman¹, Igor K. Lednev¹*

¹ Department of Chemistry, University at Albany, State University of New York, Albany,
New York, United States of America

² Department of Chemistry, University of Miami, Coral Gables, Florida, United States of America

Published in PLoS ONE in 2012
Introduction

Protein aggregates play an important role in living cells due to their ubiquity. Aggregation of proteins results in the formation of long, unbranched β-sheet-rich structures, commonly known as amyloid fibrils\(^1\). These fibrils are found as deposits in the tissues and organs of patients with various amyloid-associated diseases, such as Alzheimer’s disease (AD), Parkinson’s disease (PD), Huntington’s disease (HD), prion disease, and type II diabetes\(^2\)-\(^4\). There is also increasing evidence that small aggregates of misfolded proteins are most toxic and the formation of amyloid fibrils is a defense mechanism\(^2\)-\(^4\). It is known that more than 20 proteins that can aggregate to form amyloid-like fibrils. Previously, it was proposed that the ability to form amyloid fibrils is not a peculiarity of this small group of disease-related proteins, but rather, the ability to form amyloids is a generic property of the polypeptide chain\(^5\). Thus, many physiochemical properties of protein sequences, such as charge, hydrophobicity, and the tendency to form secondary structures, were extensively elucidated in recent decades to understand their relative propensities for amyloid fibril formation. One example of these properties is disulfide bonds, which are present in 65% of all secreted proteins, and in 50% of proteins involved in amyloidosis\(^6\).

The behavior of disulfide bonds upon protein aggregation has been extensively studied over the past decade\(^7\)-\(^9\). Disulfide bonds limit the way in which a protein or a peptide can aggregate into a fibril via steric restraint. For example, the reduction of intramolecular disulfide bonds in β\(_2\) microglobulin was determined to limit the formation of long fibrils upon protein aggregation\(^9,10\). In our previous work, we demonstrated that a reduction of three out of four disulfide bonds in bovine apo-α-lactalbumin leads to
significant changes in the aggregation pathways of these proteins, as well as the structure and morphology of their mature fibrils\textsuperscript{11}.

There is great interest in understanding the influence of disulfide bonds on the stability of insulin. The polypeptide hormone insulin stimulates a complex signal transduction pathway associated with glucose metabolism. The native structure of the insulin monomer is mainly helical, with two of its polypeptide chains linked by one intra-chain and two inter-chain disulfide bonds. Importantly, disulfide bonds are critical for the physiological function of insulin\textsuperscript{12}. Insulinoma and injection amyloidosis are associated with insulin aggregation\textsuperscript{13,14}. Zako \textit{et al.} showed that reducing all disulfide bonds of native insulin leads to the formation of structurally and morphologically different insulin fibrils\textsuperscript{15}. In addition to the dramatic impact on insulin stability and aggregation, disulfide bonds can contribute to free radical formation and fibrillar toxicity. In particular, Schöneich proposed that sulfur-containing amino acids cause free radical shrapnel during protein aggregation\textsuperscript{16}. However, whether disulfide bonds undergo cross-scrambling during insulin aggregation, their role in this process, and their location in the fibrillar structure remain unknown.

Insulin is present as a dimer in solution. However, only the insulin monomer is physiologically active\textsuperscript{17}. Insulin dimerization has been proposed as a key step in the amyloidogenic pathway\textsuperscript{18}. Belfort \textit{et al.} proposed that three dimers of insulin comprise the fibril precursors that function as a template for further insulin aggregation\textsuperscript{19}. Insulin fibrils are β-sheet-rich aggregates, whereas native insulin has a predominantly α-helical structure. Thus, an extensive α-helical to β-sheet refolding should occur during the fibrillation process. The elucidation of the amyloidogenesis of the insulin sequence,
which is a primary determinant in protein aggregation, has been a topic of active research in recent decades\textsuperscript{20-22}. New possibilities could be created for specific drug design to block insulin aggregation and fibril formation. Eisenberg \textit{et al.} proposed that part of the B-chain sequence, LVEALYL, is the smallest segment responsible for the initiation of insulin aggregation\textsuperscript{18}. However, this segment has also been determined to terminate protein aggregation. Previous studies by Sawaya \textit{et al.} have demonstrated that several other sequences, such as LYQLEN (residues A\textsuperscript{13}-A\textsuperscript{18}) and VEALYL (residues B\textsuperscript{12}-B\textsuperscript{17}), also modify protein aggregation and form amyloid fibrils\textsuperscript{23,24}. In addition, point mutations have been found to either delay or prolong the lag phase of insulin fibrillation\textsuperscript{21,22}. However, whether the studied amino acid fragment is located in parts of the unordered fibril or forms the core spine remains elusive.

Hydrogen-deuterium (H/D) exchange is a valuable tool for characterizing protein structure, solvation, and water exposure when combined with NMR, mass spectrometry, and vibrational spectroscopy. Coupling NMR with H/D exchange has been demonstrated to be a powerful method for determining the amino acid motif involved in β\textsubscript{2} microglobulin fibril formation\textsuperscript{25}. Deep UV resonance Raman spectroscopy combined with H/D has also been shown to be a very powerful tool for fibril core characterization\textsuperscript{26,27}. In an amino acid residue, the main chain NH group and O-, N-, and S-bound protons exchange easily, whereas carbon-bound hydrogens do not. In the hydrophobic core or strongly hydrogen-bonded secondary structures of proteins, the H/D exchange rates are strongly reduced due to the shielding of exchangeable sites. Previously, the hydrophobic fibril core was demonstrated to be highly resistant to H/D exchange\textsuperscript{27}. The current model of amyloid fibrils postulates that a highly hydrophobic
cross-β core is flanked by unordered parts. Taking this model into account, one can expect that hydrogen-deuterium exchange of these fibril structures will result in the proton exchange only in unordered parts, whereas the cross-β core remains protonated.

Herein, using the combination of deep ultraviolet resonance Raman (DUVRR) and Nuclear Magnetic Resonance (NMR) spectroscopy with H/D exchange, we determined the parts of the insulin sequence that form the fibril core and are present in the unordered parts. We observed that at least two B-chain segments, B3-B7 and B10-B18, remain highly protonated under H/D exchange and most likely form the fibril core. Surprisingly, we did not observe any highly protonated segments in the insulin A-chain that were longer than two amino acid residues. We also found that one cysteine residue of each disulfide pair is located in the hydrophobic fibril core, whereas the other residue sticks out of the fibril core and is most likely located in the unordered parts of the fibril. This discovery demonstrates that fibril disulfide bonds remain intact with the same molecular conformation as in native insulin, even after an extensive conversion of the α-helical structure to a fibrillar β-sheet. One can envision that during protein aggregation, cysteine disulfide bonds extend out into the aqueous media, whereas tyrosines are packed inside the fibril cross-β core. We performed MD simulations in aqueous solution to model the conversion of mainly α-helical monomers into primary β-sheet dimers. Our results indicate that the monomer aggregation process occurs via a zipper-like mechanism as previously proposed by Eisenberg and co-workers. The complete melting of α-helices and the formation of a significant amount of β-sheets occur, although all three disulfide bonds of native insulin remain intact.
Materials and Methods

Fibril Formation

Bovine insulin was purchased from Sigma-Aldrich (I5500). Insulin fibrillation was performed by growing insulin (60 mg/ml) in HCl, pH 1.9 at 65 °C overnight as previously described. The amyloid fibrils were washed with HCl, pH 1.9 and centrifuged for 30 min at 12,000 g at 25 °C. The supernatant was removed and the process was repeated twice. The insulin fibrils were then redispersed in HCl, pH 1.9 and lyophilized. To prepare a protonated sample, 28 mg of lyophilized powder was dissolved in d6-dimethyl sulfoxide and 0.05% Trifluoroacetic acid (TFA) for NMR analysis. To prepare a deuterated sample, 25 mg of lyophilized powder was exposed to D2O for 7 days at 20.5 °C, followed by lyophilization. Deuterated lyophilized amyloid fibrils were dissolved in d6-DMSO and 0.05% TFA for NMR analysis. The final concentrations of the protonated and deuterated amyloid fibrils were 9 mM and 10 mM, respectively.

NMR Experiments

All samples were placed in a 5-mm NMR tube, and the experiments were conducted on a Bruker AM-500 spectrometer with a z-axis gradient cryoprobe. The probe temperature was maintained at 30 °C. 1H, 1H-TOCSY and 1H, 1H NOESY spectra were collected with a mixing time of 45 ms and 150 ms, respectively, to optimize magnetization transfer. Spectra were collected using the Watergate pulse sequence for water suppression. All spectra were processed using TOPSPIN 2.1 (Bruker, Inc). In t1 and t2 dimensions, 4096 and 512 points were collected, respectively. The 2D data sets were apodized by a sine-bell and Fourier transformed. NMR chemical shift assignments were made using CARA software.
To date, the bacterial recombinant expression of mature insulin is not possible. Homonuclear NMR is the only method available for assigning protons of insulin. The chemical shift assignments of dispersed insulin amyloid fibrils were assigned based on known assignments of the insulin monomer dissolved in 65% H$_2$O, 35% d$_3$-acetonitrile, and 0.05% TFA$^{31}$. Because the $^1$H,$^1$H-TOCSY spectrum of monomeric insulin was very similar to that of amyloid fibrils, we used the insulin monomer during chemical shift assignment experiments. To match the chemical shifts of insulin under different buffer conditions, 1 mM of insulin monomer was dissolved in H$_2$O, 35% d$_3$-acetonitrile, and 0.05% TFA, and the solvent was gradually changed to d$_6$-DMSO and 0.05% TFA. The changes in the NMR spectra of monomeric insulin were monitored by $^1$H,$^1$H-TOCSY.

*Raman Experiments*

*Non-resonance Raman spectroscopy*. Insulin fibrils were lyophilized and the resulting insulin protein powder was placed onto alumina foil. A Renishaw inVia confocal Raman spectrometer equipped with a research-grade Leica microscope, 20x long-range objective (numerical aperture of 0.35), and WiRE 2.0 software was used for non-resonance Raman spectroscopy. A 785-nm-wavelength laser was used, and the laser power was reduced to approximately 11.5 mW to avoid sample photo-degradation.

*Deep UV resonance Raman spectroscopy (DUVRR)*. DUVRR spectra (197-nm excitation) were collected using a home-built Raman spectrometer as previously described$^{32}$. A spinning quartz NMR tube with a magnetic stirrer inside was used for sampling. Raman scattering was dispersed and recorded using a homebuilt double monochromator coupled to a liquid-nitrogen-cooled CCD camera (Roper Scientific, Inc.). All reported Raman spectra were an average of at least three independent measurements.
GRAMS/AI 7.0 software (Thermo Galactic, Salem, NH) was used for data processing. The application of hydrogen-deuterium exchange combined with DUVRR spectroscopy for structural characterization of the fibril core has been previously described\(^{33}\). Samples (1 mL) of fibrils dispersed in water were centrifuged at 14000 \(\times g\) for 30 min. The precipitate was subsequently divided into two parts, each of which was washed two times with either D\(_2\)O or H\(_2\)O by subsequent spinning-redispersion. The washed dispersions (in D\(_2\)O and H\(_2\)O) were used for Raman spectroscopic measurements.

*Dynamic Light Scattering*

Solutions of insulin protein and insulin fibrils disintegrated by DMSO/TFA were analyzed using Dyna Pro Titan DLS (Wyatt Technology Corp.). Acquired data were analyzed using DYNAMICS V6 software (Wyatt Technology Corp.).

*Atomic force microscopy*

Fibril solution was centrifuged at 12,000 \(g\) prior to the deposition to remove non-aggregated protein. A gelatinous pellet was diluted with HCl, pH 1.9 or DCl, pD* 1.9 solution with a 1:400 dilution factor (V/V). A drop of this solution was placed onto freshly cleaved mica in AFM fluid chamber and incubated for 2 min followed by removing of the solution excess. To avoid mica surface drying, 2 ml of distill water were placed on the top of the mica. AFM scanning was performed immediately in tapping mode using MFP-3D\textsuperscript{TM} Bio Asylum Research microscope (Asylum Research, CA, USA) with Olympus TR400PSA tips.
Computational Methods

Using the GROMACS program\textsuperscript{34,35}, all MD simulations were performed utilizing the GROMOS force field GROMOS96 53A5\textsuperscript{36}. In the simulations, the starting structures were placed in a large cubic box (9.0 x 7.0 x 7.0 Å\textsuperscript{3}) to avoid artificial interactions with their images in the neighboring boxes created by the application of periodic boundary conditions (PBCs). The box was filled with single point charge (SPC) water molecules. The GROMOS96 force field and SPC water model have been successfully employed to explore protein dynamics in several recent studies (cite 1-3 from the letter of response). Some water molecules were replaced with sodium and chloride ions to neutralize the system and to simulate an experimentally used ion concentration of 150 mM. Subsequently, the starting structures were energy-minimized with a steepest descent method that used 3000 steps. The results of these minimizations produced the starting structure for the MD simulations. Subsequently, the simulations were performed with a constant number of particles (N), pressure (P) and temperature (T), \textit{i.e.} the NPT ensemble). The SETTLE algorithm\textsuperscript{37} was used to constrain the bond lengths and angles of the water molecules, and the LINCS algorithm\textsuperscript{38} was used to constrain the bond length of the peptide bond. Long-range electrostatic interactions were calculated by the Particle-Mesh Ewald (PME) method\textsuperscript{39}. A constant pressure of 1 bar was applied with a coupling constant of 1.0 ps. Peptides, water molecules and ions were coupled separately to a bath at 300 K with a coupling constant of 0.1 ps. The equation of motion was integrated at each 2-fs time step. Umbrella pulling, as implemented in the GROMACS package, was applied to unfold the monomeric structure. In this method, a harmonic potential is applied between the center of mass of two groups. A standard pulling rate and a force constant of
0.002 nm/ps and 1600 kJ/mol nm², respectively, were used. The tools available in the GROMACS program package and the YASARA program were used for analyzing the trajectories and the simulated structures.

**Computational Modeling**

The monomer X-ray structure of the bovine insulin hexamer dimer was crystallized from bovine insulin hexamer (PDB ID: 2ZP6). The insulin monomer was abstracted from the X-ray structure of bovine insulin hexamer (PDB ID: 2ZP6). This structure was simulated in water without any constraints for 100 ns. In the next step, umbrella pulling was applied on the N- and C-termini to unfold this structure. The unfolded structure was then used to prepare a model for the dimer. In the model for the 150-ns simulation, both monomers were oriented to form a zipper-like conformation. The most representative structure obtained from this simulation was further utilized to develop 10-ns MD simulations for the insulin hexamer. In this model, the B6L-B20G and A5Q-A21N fragments of each monomer were truncated.

**Results**

**Determination of amino acid protection**

After the termination of insulin fibrillation, mature insulin fibrils, separated from un-aggregated protein, were re-dispersed in D₂O, pD* 1.9 at 25 °C. Deuterium atoms were substituted for hydrogen atoms in fibril unordered parts, while the highly hydrophobic core remained protonated. Our microscopic observation of insulin fibrils before and after H/D exchange did not show any noticeable changes in their morphology (Fig. 1.1). The exchanged fibril solution was then lyophilized and re-dissolved in a buffer.
composed of 99.95% DMSO and 0.05% TFA, which disintegrates the fibril structure into protein monomers (Fig. 1.2) without changing the protonation state of amide protons. The resulting protein was then analyzed by homonuclear NMR spectroscopy. The unexchanged amides from the amino acids that were localized in the fibril core were detected, and the amino acid residues that exchanged a proton for a deuterium were “invisible.” To determine the degree of protection for each residue that remained protonated, the peak intensities of insulin that was incubated with and without D$_2$O were plotted against the residue number (Fig. 1.3 and 1.4). The following three thresholds were established to illustrate the degree of protection: yellow, $I_{D2O}/I_{H2O} \geq 0.75$; red, $0.75 > I_{D2O}/I_{H2O} \geq 0.675$; and blue, $0.675 > I_{D2O}/I_{H2O} \geq 0.60$ (Fig. 1.3A and B).
**Figure 1.1** (A) AFM images of insulin fibrils before H/D exchange (B) and two days after incubation in D$_2$O, pD* 1.9 at 25 °C

**Figure 1.2** DLS data of insulin disintegrated (A) fibrils by DMSO/TFA and (B) protein in HCl, pH 2.0. (A) Insulin monomer, which has a radius of 0.07 nm, takes 99.9% of the solution mass. (B) Insulin protein remains as a dimer, with a radius ~2.5 nm, in HCl, pH 2.0$^{43}$. 

15
Figure 1.3 The majority of H/D exchange-protected amino acids are located in the B-chain of insulin. (A) The amino acids in the 3D structure and (B) primary sequence of insulin fibrillar monomer are colored according to the following degree of H/D exchange protection: yellow, $I_{D2O}/I_{H2O} \geq 0.75$; red, $0.75 > I_{D2O}/I_{H2O} \geq 0.675$; and blue, $0.675 > I_{D2O}/I_{H2O} \geq 0.60$. Gray (in the 3D structure) and white colors (in the primary sequence) indicate unprotected (exchanged) amino acids. (C) 2D $^1H,^1H$-TOCSY spectrum of insulin fibrillar monomer in DMSO and 0.05% TFA at 30 °C after a 7-day D$_2$O exchange at room temperature. NMR data were acquired at room temperature on a Bruker Avance III 500 MHz NMR spectrometer equipped with a cryoprobe.
Figure 1.4 Graph illustrating the intensity differences between deuterated (I$_{D2O}$)/protonated (I$_{H2O}$) monomer protection degrees of each amino acid residue in the insulin monomer.
Because a native source of insulin was used, we could not isotopically enrich insulin with NMR-active nuclei, such as $^{13}$C and $^{15}$N, which help facilitate the chemical shift assignment process. Therefore, 2D $^1$H,$^1$H-total correlation spectroscopy (TOCSY) and nuclear Overhauser effect spectroscopy (NOESY) were used for the sequence-specific assignment of the chemical shifts of protons in the amino acid residues of the protein. $^1$H,$^1$H TOCSY experiments were used to determine the types of amino acid residues that were present. $^1$H,$^1$H NOESY experiments were used to place residues within the protein primary sequence for sequence-specific assignments.

The spectral dispersion of insulin in the amide proton region was limited, ranging from 7.1 ppm to 8.8 ppm (Fig. 1.3C), which is a strong indication that insulin is largely devoid of tertiary structure under the NMR buffer conditions. This relatively narrow spectral region leads to the significant spectral overlap of proton resonances, complicating the chemical shift assignment. When analyzing the fibril monomer, we unambiguously assigned 35 of the 49 amino acid insulin residues resolved in 2D $^1$H,$^1$H-TOCSY spectra. However, 14 residues (A chain: Cys$^6$, Cys$^7$, Ser$^9$, Ser$^{12}$, Tyr$^{14}$, Leu$^{16}$, Asn$^{18}$, Cys$^{20}$, B chain: Phe$^{22}$, Gly$^{29}$, Ser$^{30}$, Gly$^{41}$, Gly$^{44}$, Phe$^{45}$) were missing and could not be assigned due to either spectral overlap or extreme line broadening caused by the intermediate exchange between conformers in solution. It is also possible that amino acid residues located in unordered fibril parts exchanged with deuterium at rates that were too fast to be detected by NMR spectroscopy.

Only two amino acid doublets, Ile$^2$-Val$^3$ and Val$^{10}$-Cys$^{11}$, in the A-chain were observed to be protected to a medium and low extent (60-75%). In the rest of the A-chain, four single amino acids, Gln$^{15}$, Glu$^{17}$, Tyr$^{19}$ and Asn$^{21}$, were observed to have high
(over 75%) protection, and Leu$^{13}$ was observed to have medium (67.5 to 75%) protection. Interestingly, these amino acid residues alternate with the following unprotected residues, which have even numbers: Ser$^{12}$, Tyr$^{14}$, Leu$^{16}$, Asn$^{18}$ and Cys$^{20}$. One might expect that the protected amino acid residues from the A-chain form the core spine, whereas the rest of the residues (57 %) are most likely located in the unordered parts of the fibril.

In the B-chain, most amino acid residues (73%) are protected. We identified two segments [(Asn$^{3}$, Gln$^{4}$, His$^{5}$, Leu$^{6}$ and Cys$^{7}$); (His$^{10}$, Leu$^{11}$, Val$^{12}$, Glu$^{13}$, Ala$^{14}$, Leu$^{15}$, Tyr$^{16}$, Leu$^{17}$, Val$^{18}$ and Cys$^{19}$)] that remain protonated under H/D exchange. Close to the N-terminus, three amino acids (Phe$^{25}$, Tyr$^{26}$ and Thr$^{27}$) along with Lys$^{29}$ and Ala$^{30}$ are also protonated to medium and high extents. Among the protonated amino acid residues of the B-chain, there is an LVEALYLV segment predicted by Eisenberg to be the main contributor to the fibrillar core formation$^{18}$. We also found that the following N-termini of both chains are protected: Asn$^{21}$ of the A-chain and Ala$^{30}$ of the B-chain to a medium and a high extent, respectively. However, the C-termini of both chains remain unprotected, suggesting that the protein C-terminus is located outside the core, whereas the N terminus takes part in fibrillar core formation. Our calculations show that in both A- and B-chains, 61% of the insulin sequence remains protonated and located in the fibril core. We also determined that three of four tyrosine residues were protected and were most likely present in the hydrophobic fibril core. However, another aromatic amino acid, phenylalanine, remained mostly unprotected (only one of three amino acid residues is protected) in the fibril structure.

Three disulfide bridges have been shown to play a vital role in the stability of the insulin monomer. Taking into account the dramatic perturbation of the insulin secondary
structure from an α-helix into a β-sheet upon protein aggregation, the stability of the disulfide bridges during this process was investigated. Prior to this study, there was no experimental evidence about the scrambling or stability of these bridges\textsuperscript{7,8}. Our results show that insulin fibrils have an intriguing organization of cysteine such that one of each cysteine pair is protected (Cys\textsuperscript{11}, Cys\textsuperscript{7} and Cys\textsuperscript{19}), and the other is not (Cys\textsuperscript{6}, Cys\textsuperscript{7}, and Cys\textsuperscript{20}). We found that Cys\textsuperscript{11}, Cys\textsuperscript{7} and Cys\textsuperscript{19} have a higher degree of protection of 60\% for H/D exchange and are most likely located in the hydrophobic fibril core. These data suggest that each unprotected cysteine may “follow” its protected partner, which is integrated into the β-sheet core, during secondary structure changes, leaving the disulfide bond intact. Therefore, we hypothesized that the unordered parts of the fibril contain cysteine and are rich in disulfide bonds. One might speculate that the specific location of cysteine may play a role in intertwining proto-filaments and proto-fibrils. As a result, Raman spectroscopy was used to investigate the conformations of disulfide bridges in native insulin and insulin fibrils.

The conformations of disulfide bonds

There are three disulfide bonds that maintain the structure of the insulin monomer. Their scrambling and stability upon fibrillation have been extensively studied\textsuperscript{7,8,13}. Raman spectroscopy is a unique technique used in the structural characterization of protein disulfide bond conformations. Structural information can be obtained regarding the internal rotation of C-C-S-S-C-C bonds that are present in the following conformations: gauche-gauche-gauche (g-g-g), gauche-gauche-trans (g-g-t), and trans-gauche-trans (t-g-t)\textsuperscript{46}. 

20
**Figure 1.5** Disulfide bonds preserve their conformation upon insulin fibrillation.

Raman spectra of native insulin (red) and insulin fibrils (black) have a peak at 510 cm$^{-1}$, corresponding to the gauche-gauche-gauche (g-g-g) conformation of disulfide bonds (schematically represented in the inset).
Using non-resonance Raman spectroscopy with excitation at 785 nm, we determined that the predominant gauche-gauche-gauche (g-g-g) conformation for all fibril disulfide bonds is identical to that of disulfide bonds in the native protein (peak at 510 cm\(^{-1}\), Fig. 1.5).

These data indicate that all three disulfide bonds do not break, and all three keep the predominant gauche-gauche-gauche conformation of the C-C-S-S-C-C segment. Insulin aggregates and preserves its disulfide bonds by squeezing one cysteine residue in each disulfide bond outside the fibril core to the flexible unordered parts. Based on this observation, we hypothesized that the surface of the insulin fibrils should be rich with cysteine residues and disulfide bonds, which may play a significant role in the high free radical activity that is associated with sulfur atoms\(^{16}\). Additionally, the NMR data support the conclusion that disulfide bonds do not scramble during insulin aggregation. The scrambling of disulfide bonds would lead to changes in the orientation of the protein backbone around these cysteine residues, resulting in large chemical shifts for the affected residues. We observed that the amino acid peak positions within the vicinity of cysteine residues in the 2D \(^1\)H,\(^1\)H-TOCSY spectra of an insulin fibril monomer and native protein are similar, indicating that the disulfide bonds did not scramble upon protein aggregation (Fig. 1.6). Our NMR and Raman data provide new insights about insulin fibril surface organization, which may serve as a basis for the design of therapeutic drugs.
Figure 1.6 Backbone conformations of fibrillar and monomer insulins are very similar in the vicinity of disulfide bridges. Overlay between $^1$H-$^1$H-TOCSY spectra of 1 mM fibrillar (red) and monomer (black) insulins in d$_6$-DMSO and 0.05% TFA at 30 °C. The amino acids located within 4 residues from insulin cysteins are labeled. NMR data were acquired on Bruker Avance III 500 MHz NMR spectrometer equipped with the cryoprobe.
Secondary structure of the fibril core

Deep UV resonance Raman spectroscopy has been demonstrated to be a powerful tool for the characterization of the amyloid fibril structure\textsuperscript{47}. In particular, DUVRR spectroscopy combined with hydrogen-deuterium exchange has been utilized in the structural characterization of the fibril core\textsuperscript{27}. A typical protein DUVRR spectrum is dominated by amide bands, which characterize the polypeptide backbone conformation. In addition, the spectrum displays aromatic amino acid bands, which provide information about their local environment\textsuperscript{48}. We observed a gradual increase in Raman band intensity with incubation time for C\textsubscript{\alpha}–H (1390 cm\textsuperscript{-1}) as well as Amide I and II modes, indicating β-sheet formation. A significant decrease in tyrosine band intensity was also observed, indicating the changes in the local environments of tyrosine residues during protein aggregation (Fig. 1.7A).

Previous studies of native insulin conformation have indicated that tyrosines have a hydrophilic environment\textsuperscript{43,49}. The decrease in the intensity of aromatic bands in the DUVRR spectrum of insulin fibrils could indicate that tyrosine residues are in a more hydrophobic environment than in the native protein. However, a significant change in the phenylalanine (1000 cm\textsuperscript{-1}) band intensity was not observed, indicating that the local environments of phenylalanine residues do not change. These data corroborate our NMR results on the aromatic amino acid environments.
Figure 1.7 The secondary structure of insulin changes dramatically from being mostly α-helical in the native protein to highly β-sheet-rich in the fibrillar form. The local environment of tyrosine residues changes simultaneously during hydrophilic to hydrophobic protein aggregation. (A) DUVRR spectra of bovine insulin at pH 2.0, 25 °C (red) incubation solution after 30 min (black) and one hour (blue) of heating at 70 °C. Simultaneously, approximately 75% of the insulin sequence is packed into the cross-β-core, and the remainder forms unordered parts of fibrils. (B) Deep UV resonance Raman spectra of insulin fibrils in H2O (blue), after deuteration (red), and a spectrum of cross-β-sheets (black). The contribution of aromatic amino acids is quantitatively removed by subtracting the spectra of phenylalanine and tyrosine.
In an unordered protein spectrum, Mikhonin and Asher showed that H/D exchange caused a downshift of the amide II DUVRR band from 1555 to 1450 cm$^{-1}$ (Amide II’) and the virtual disappearance of the amide III band$^{50}$. Figure 1.7B illustrates the corresponding spectral changes in the DUVRR signature of insulin fibrils upon deuteration. The deuteration of mature insulin fibrils resulted in a slight decrease in Amide II band intensity, indicating that the amount of fibril protein available for deuteration was very small. This finding is also corroborated by the relatively small intensity of the Amide II band in the fibril spectrum compared with that of unfolded insulin completely exposed to H/D exchange. The unfolded state of insulin was achieved by dissolution of bovine insulin in D$_2$O, pD 1.0 and brief heating at 95 °C for several minutes.

Based on our observations, insulin fibrils mainly consist of highly organized cross-β-sheets with high resistance to H/D exchange. According to Asher and coworkers, the position of the Amide III$_3$ band corresponds to the Ψ dihedral angle. We observed that insulin fibrils have a single peak centered at 1226 cm$^{-1}$, which, according to Asher’s semi-empirical approach$^{50}$, corresponds to β-sheet conformation characterized by a Ψ dihedral angle of 134.5°.
**MD simulations of β-sheet-rich dimer formation**

According to our experimental data, in addition to tertiary changes, dramatic secondary structure changes occurred upon insulin aggregation, although the disulfide bonds remained intact. An obvious question to address is whether the types of structural perturbations that take place in the insulin monomer during fibrillation to satisfy both of these criteria. To demonstrate the possibility of the conversion of the mainly α-helical monomer into the predominantly β-sheet-rich fibril form, we performed all-atom MD simulations in aqueous solution on both monomeric and dimeric forms of insulin. The latter was constructed to represent only the fibrillar form of insulin and the existence of other non-structured oligomeric states was ignored. These simulations provided both the structure and location of each amino acid residue at the atomic level. The accuracy of the simulated structures was validated by comparing them with the available experimental DUVVR and NMR data reported in this study. The structures of the full-length insulin were investigated in the following sequence: monomer → dimer. The most representative structure obtained from the simulation of the monomeric form was used to develop models for the dimeric form. The equilibrated structure of the full-length insulin monomer derived from a 100-ns MD simulation was found to contain mostly α-helical (49.0%) and small β-sheet (3.9%) character (Table 1.1).

<table>
<thead>
<tr>
<th>Structure</th>
<th>α-helix character</th>
<th>β-sheet character</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monomer</td>
<td>49.0%</td>
<td>3.9%</td>
</tr>
<tr>
<td>Dimer</td>
<td>0.0%</td>
<td>43.1%</td>
</tr>
</tbody>
</table>

**Table 1.1** The α-helix and β-sheet character of simulated insulin monomer and dimer.
This structure is well folded and stabilized by a large number of hydrogen bonds and hydrophobic interactions between A- and B-chains, including their N- and C-termini. In order to form a dimer, the structure must substantially unfold to associate with another monomer (Fig. 1.8). As discussed in the "computational procedure" section, this unfolding is accomplished by applying a constant force on the N- and C-termini of the monomer. During the unconstrained simulations of the dimer, both insulin monomers undergo significant structural changes.

The time evolution of the 150-ns MD simulation of the dimer shows that there is a gradual increase in the β-sheet character compared with that of the monomer (Table 1.1). In the first 5 ns of the simulation, the β-sheet content enhanced sharply (3.9 → 31.4 %). Subsequently, the following slow increase occurred: 34.3, 39.2, 45.1 and 46.1% at 10, 15, 20 and 120 ns, respectively (Fig. 1.9). The β-sheet character largely remained unchanged in the 120-150-ns time period. By contrast, the α-helix character was significantly reduced from 49.0% to 0.0% (Table 1.1). Notably, within the dimer, the N- and C-termini of the A- and B-chains of both monomers did not interact with each other. However, in fibrils, these regions most likely associate through β-sheet interactions with other dimers, increasing the content of this secondary structure.
**Figure 1.8** Equilibrated structure of an insulin dimer from 150-ns MD simulations in aqueous solution. The A- and B-chain of the second monomer are labeled A’ and B’, respectively. Intact disulfide bridges are shown in green.
**Figure 1.9** Snapshots of insulin dimers from MD simulations. The β-sheet character of each structure is shown in parentheses.
Dimerization occurs through a zipper-like mechanism in which the B10-B18 regions of monomers form two sides of the zipper. The formation of this type of structure has previously been observed in the oligomerization of short fragments of insulin and other amyloidogenic peptides. The information provided by the MD simulations can be combined with the measured H/D exchange data to elucidate the structure of the insulin fibril. In the structure of the zipper derived from simulations, Leu$^{11}$, Leu$^{15}$ and Leu$^{17}$ residues of one monomer interact with their counterparts in the second monomer through hydrophobic interactions. In addition, two tyrosine residues (Tyr$^{19}$ and Tyr$^{16}$ of A and B chains, respectively) interact with each other through π-π interactions. Due to His$^{10}$, Leu$^{11}$, Val$^{12}$, Glu$^{13}$, Ala$^{14}$, Tyr$^{16}$, Leu$^{17}$, Val$^{18}$, Cys$^{19}$, and Gly$^{20}$ residues forming the hydrophobic fibril core, these residues must be protected under H/D exchange. The location and orientation of all these residues in the simulated structure were found to be in excellent agreement with current and previously reported experimental data. In addition, the protection of free Gln$^{15}$, Tyr$^{19}$ and Asn$^{21}$ residues of the A-chain and the Phe$^{25}$-Pro$^{28}$ segment of the B-chain is also in accord with the collected DUVVR and NMR data. Furthermore, as suggested by the experimental data, all three disulfide bonds [Cys$^{6}$(A)-Cys$^{11}$(A), Cys$^{7}$(A)-Cys$^{7}$(B), and Cys$^{20}$(A)-BCys$^{19}$(B)] remained intact in the dimer. Based on these calculations, we confirmed that the tertiary and secondary structures of insulin can dramatically change during oligomerization without breaking disulfide bonds.

Based on these results, we concluded that the dimeric structure of full-length insulin is a good model for elucidating several key structural properties of amyloid fibrils.
Dimerization appears to be a critical step in fibrillation. After dimerization, fibrils can grow through the stacking.

**Conclusions**

Our new approach of combining NMR and Raman spectroscopy with MD simulations for characterizing amyloid fibrils has provided exclusive knowledge about fibril structure. With single-residue resolution, we determined the amino acid residues that form the fibril core in addition to those that are located in the unordered parts of the fibril. We found that most of the sequence from the B-chain of insulin is highly protected from H/D exchange, including a segment previously described by Eisenberg\(^{18}\). However, we did not find any long (not longer than two amino acids) protected segments in the A-chain. Moreover, the A-chain was observed to have the following intriguing order of protection: starting from Cys\(^{11}\) to Asn\(^{21}\), with alternating protected and unprotected amino acid residues.

We demonstrated that three out of four tyrosine amino acid residues packed into the cross-β-sheet during insulin aggregation, whereas another aromatic amino acid, phenylalanine, remained in the unordered parts. Based on NMR data, we determined that the following B-chain residues are highly protected: Phe\(^{25}\), Tyr\(^{26}\), and Thr\(^{27}\); Lys\(^{29}\) and Ala\(^{30}\). In addition, we found that both amino acids at the insulin C-termini were unprotected, whereas both amino acids at the N-termini remained highly protonated, packing into the cross-β-core. The location and orientation of these residues and secondary structures of the Phe\(^{25}\)-Pro\(^{28}\) region of the B-chain were supported by structures derived from MD simulations. Furthermore, the structures showed that all three disulfide bonds remained intact in the dimer, which models the fibrillar form of insulin.
Together with the determination of the amino acid sequence that directly participates in the association and fibrillation of insulin dimers, we discovered a unique organization of six insulin cysteine residues, supporting the presence of intact disulfide bonds and their lack of scrambling during insulin aggregation. These results indicate that cysteine residues localized on the fibril surface may play a direct role in free radical formation, which has been previously described for sulfur atoms in proteins\(^6\).

Twenty out of 51 amino acids in the insulin sequence demonstrated complete H/D exchange. We determined that 10 residues are hydrophilic and 10 are hydrophobic (Table 1.2). Thus, our preliminary estimation of the insulin fibril surface based on NMR data analysis of unprotected amino acids indicates that it is equally hydrophilic (polar) and hydrophobic (nonpolar). The importance of identifying solvent-exposed residues in fibrils is underlined by the fact that the fibrillar surface is one of the major sources of fibrillar toxicity.

**Table 1.2** Hydrophilic and hydrophobic amino acids that were completely accessible for H/D exchange.

<table>
<thead>
<tr>
<th>Hydrophilic (polar)</th>
<th>Hydrophobic (nonpolar)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A’ chain</strong></td>
<td></td>
</tr>
<tr>
<td>E4, Q5, C6, C7, C20</td>
<td>G1, A8, L17</td>
</tr>
<tr>
<td>S9, S12, Y14, N18</td>
<td></td>
</tr>
<tr>
<td><strong>B’ chain</strong></td>
<td></td>
</tr>
<tr>
<td>S9</td>
<td>F1, V2, G8, G20, G23, F24, P28</td>
</tr>
</tbody>
</table>

**Acknowledgments**

We thank Arghya Barman for his assistance in performing the MD simulations.
Author Contributions

Conceived and designed the experiments: DK AS IKL. Performed the experiments: DK JW. Analyzed the data: DK JW MO RP AS IKL. Contributed reagents/materials/analysis tools: RP. Wrote the paper: DK RP AS IKL. Molecular dynamics simulations: MO RP.
CHAPTER TWO

IN-CELL NMR SPECTROSCOPY TO STUDY PROTEIN DRUG-INTERACTIONS

Jacqueline Washington, David S. Burz, Alexander Shehtman*

Department of Chemistry, University at Albany, State University of New York, Albany, New York, United States of America

Published as Chapter 4 in New Developments in NMR No. 2, New Applications of NMR in Drug Discovery and Development in 2013
Introduction

In-cell NMR spectroscopy is valued for its ability to shed light on molecular structures and interactions under physiological conditions. One of its greatest potentials lies in the area of drug discovery where it can reveal molecular interactions at the atomic level at all stages of structure-based drug discovery. Drug therapy is based on the ability to specifically enhance or impede biological activity by binding to a single target protein or protein complex. 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 NMR spectroscopy\textsuperscript{52}, these processes can now be studied within a cellular environment.

By using in-cell NMR spectroscopy the structure of the target protein and drug candidate can be deduced through uniform and/or selective isotopic labeling and both can be monitored simultaneously through the use of differential isotopic labeling; in addition structural changes as a result of protein maturation processes or post-translational modifications (PTMs) can be analyzed. Non-specific and specific target interactions are distinguished through concentration-dependent chemical shift analyses and signal broadening, which also helps identify the interacting surface of the target, and the affinity of the interaction can be approximated. Techniques for incorporating isotopically labeled proteins into both prokaryotic and eukaryotic cells provide a means for drug delivery as well. Control experiments easily distinguish intracellular signals from extracellular, facilitating the identification of valid drug-protein interactions. Finally, high throughput techniques are increasingly able to screen larger numbers of drug candidates without the need for extensive in vitro experimentation.
In this review, the basics of in-cell NMR spectroscopy are described with emphasis on applications that are relevant to the structure-based drug discovery process. Applications such as determining 3D structures *de novo*, mapping interaction surfaces, and measuring changes in protein structure due to post-translational biochemical modifications, folding states, metabolic processing, and protein maturation under physiological or near physiological conditions. The latest methods for incorporating target molecules into the intracellular milieu at sufficient concentrations for NMR spectroscopy are detailed along with the physiological relevance and pitfalls of each technique. Lastly, specific examples of drug-protein interactions identified by using in-cell NMR are presented.

**In-cell NMR spectroscopy**

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 as well as changes due to biochemical modifications. This sensitivity can be exploited to screen for the binding of small drug-like molecules to selected targets in a cellular environment. These interactions alter molecular surfaces and can result in tertiary and quaternary conformational changes, all of which are reflected by changes in the chemical shifts of these nuclei.

Molecular interactions form the basis of biological activity. One of the most difficult challenges in studying protein interactions in cells arises from the fact that proteins bind many types of ligands, from protons and other ions, to small drug-like molecules and ligands and large macromolecules such as nucleic acid polymers,
membranes 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 dissect the differences between in-cell and in vitro studies of protein-protein interactions, the interactions between expressed proteins and other intracellular molecules can be divided into two parts: Specific interactions with affinities lower than 10 μM and non-specific interactions with affinities much greater than 10 μM.

To detect specific interactions by using in-cell NMR spectroscopy interactor molecules have to be present in a stoichiometric ratio with the target protein so that the bulk of the labeled population is bound. 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. Under these conditions, the contribution of undesired specific interactions between the target protein and intracellular binding partners to the resulting in-cell NMR spectrum can generally be ignored. Interactor molecules are later introduced at concentrations that are comparable to that of the target protein, thereby giving rise to specific interactions inside the cell that are detectable by in-cell 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\textsuperscript{53}. Because the concentration of molecules inside eukaryotic and prokaryotic cells are very similar to each other, the in-cell NMR spectra of cytosolic proteins acquired in either cell type will also be very similar.
**Cell types**

The choice of cell is a key element in the drug screening process. Methods exist for delivering labeled target molecules to the cytosol of both prokaryotic and eukaryotic cells. However both cell types offer benefits and drawbacks. The delivery of the candidate drugs to the cell is also a critical feature. Finally, cell viability throughout the selection process must be considered.

**Prokaryotic cells**

Early work in the field of in-cell NMR utilized prokaryotes as host cells. Prokaryotic cells, specifically *E. coli*, were exploited because they are easy to handle and grow very rapidly. Proteins can be uniformly labeled with NMR-active isotopes, primarily $^{13}$C and $^{15}$N, and over-expressed to high enough intracellular levels to yield high quality heteronuclear single quantum coherence (HSQC) spectra with little or no interfering background. Moreover, the ability to selectively label proteins and grow cells in deuterated medium provides the capability to study high molecular weight proteins in their native cellular environment. By appropriately engineering the transcriptional machinery of prokaryotic cells, post-translational modifications can be studied under conditions in which the effects of PTMs on protein structure can be examined without competing reactions, in effect, turning the bacteria into "cellular test tubes". Although *E. coli* are well-suited for simulating an intracellular environment to study eukaryotic proteins, in that they present the lowest potential for intrinsic binding partners, these cells lack the inherent ability to affect post-translational modifications and provide compartmentalization for selective activity.
**Eukaryotic cells**

In-cell NMR in eukaryotic cells is hampered by 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. Early in-cell NMR spectroscopy using eukaryotic cells was limited to very large and mechanically manipulable cells, such as *Xenopus laevis* oocytes\(^{54-56}\); in these studies, isotope-labeled molecules were microinjected into these cells. Later studies developed methods to introduce isotopically labeled proteins into a range of cell types including human embryonic kidney (293F) and HeLa cells\(^{57,58}\). The delivery methods included using active transport of isotope-labeled molecules into the cell by linking them to cell penetrating peptides\(^{57}\), and, more recently, using a pore-forming bacterial toxin, streptolysin O (SLO)\(^{58}\), to allow target molecules to passively diffuse into the cell interior before the pores are sealed. The primary problem of attaining a sufficiently high intracellular concentration of target to perform in-cell NMR has been largely overcome by these newer methods.

**Cell viability**

In-cell NMR is an effective tool for studying proteins in their native environment. However the long experimental times required to collect in-cell spectra can lead to a loss of cell viability. One consequence of reduced cell viability is 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 intracellular protein\(^{59}\). Another important factor that limits the acquisition time of in-cell NMR experiments is cell lysis, which also results in leakage of labeled target from the cells\(^{60}\). Stabilizing cells by using
known cell protectants, such as glycerol or sucrose\textsuperscript{60} may also extend the in-cell NMR acquisition time. Control experiments are performed to ensure that NMR signal arise from proteins located inside the cell.

Typically, following data acquisition, the cells are sedimented and an $^1\text{H}^{15}\text{N}$-HSQC spectrum of the supernatant is collected. Little or no signals above noise level is a good indication that the NMR signals were intracellular in origin. Li \textit{et al.}\textsuperscript{61} collected a spectrum of an 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 $\alpha$-synuclein ($\alpha$-SYN), and the 7 kDa globular protein chymotrypsin inhibitor-2 (CI2) showed that $\sim$20\% of the CI2 leaks from the cell, while $\alpha$SN remains in-cell with only small metabolites found in the spectrum of the supernatant. To improve the in-cell spectrum of $\alpha$-SYN, Li \textit{et al.}\textsuperscript{61} used an alginate encapsulation method\textsuperscript{62} to stabilize the cells, which, in turn, yielded a clean $^1\text{H}^{15}\text{N}$-HSQC for $\alpha$-SYN. These results show that each protein must separately evaluated to optimize the in-cell signal and minimize protein leakage.

Xie \textit{et al.}\textsuperscript{63} 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 \textit{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. To minimize cellular degradation during their study of the Tau protein, Bodart \textit{et al.}\textsuperscript{54} suspended Xenopus oocytes in a 20\% Ficoll solution, which allowed the cells to remain stable overnight\textsuperscript{64}. Control experiments showed no Tau in the extracellular medium. These studies demonstrated methods for
extending the lifetimes of cells in the NMR tube, a critical prerequisite for collecting data at physiological concentrations.

Another method used to assess bacterial cell viability is colony plating, in which 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\textsuperscript{65}.

**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 isotopic 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 because bacterial growth is significantly reduced during protein over-expression. The protein resonance peaks of the resulting $^1$H[$^{15}$N]-HSQC\textsuperscript{66} spectra are adequately resolved provided that the protein is expressed to a sufficiently high concentration within the cell.
Another backbone labeling scheme uses auxotrophic bacterial strains\textsuperscript{67,68} 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\textsuperscript{67} 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 $[U^{-13}\text{C}]$ 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\textsuperscript{69}. 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\textsuperscript{69}.

Other labeling techniques, such as selective labeling of individual amino acids or the use of fluorine probes\textsuperscript{70} may also be used in drug screening applications. These schemes will only be effective when the probe is close enough to the binding site to be affected by changes in the chemical environment.
In-cell delivery of protein target molecules

Proper preparation of samples for in-cell NMR spectroscopy is critical for acquiring high quality in-cell spectra. To study drug-protein interactions it is critical that the concentration of labeled nuclei be high enough to provide well-resolved resonances for unambiguous identification. This often requires concentrations that exceed normal physiological levels by an order of magnitude or more. Such high concentrations can be problematic for assessing the efficacy of a drug-protein interaction, which would occur at much lower target protein concentrations. While the use of tightly regulated promoters for protein over-expression remains the method of choice for prokaryotic cells, new techniques for delivering target molecules to eukaryotic cells at physiological concentrations that provide a good signal to noise ratio have been developed. These methods have their advantages and disadvantages, but all have proven effective in facilitating the acquisition of reproducible in-cell NMR spectra.

Endogenous over-expression

Endogenous 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 for generating samples for use in-cell NMR spectroscopy. The most popular and convenient method for producing labeled targets for in-cell NMR spectroscopy utilizes an inducible plasmid in bacterial cells to over-express isotopically labeled protein. The concentration of over-expressed protein can be controlled by using promoters in which level of transcription is proportional to the concentration of the inducing molecule (e.g. the arabinose $P_{BAD}$ and rhamnose $P_{RHA}$). Other methods for controlling the levels of
intracellular concentration of over-expressed protein include varying the induction time and using plasmids with greater or lesser copy number. Polyacrylamide gel electrophoresis (PAGE) or Western blot analyses are typically employed to estimate the intracellular concentration of over-expressed target proteins. Endogenous over-expression in eukaryotic cells does not produce sufficiently high levels of target protein to perform in-cell NMR; in these cells, exogenous delivery is required to attain high levels of target protein.

**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. The only significant contribution to background arises from the natural abundance of $^{13}\text{C}$ (1.1%) when employing such labeling. 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.*$^{56}$ used a robotic microinjection device to administer precise quantities of streptococcal protein G B1 domain (GB1)$^{73}$ into *X. laevis* oocytes.$^{74}$ GB1 was uniformly $[\text{U}-^{15}\text{N}]$ 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 $^{15}\text{N}$-GB1 obtained in extracts was
virtually identical to that obtained \textit{in vitro} using purified protein over the concentration range examined.

Next, \textit{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 \textit{in vitro} by using purified protein. 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.

There were no discernible 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 shows the feasibility of performing in-cell NMR spectroscopy in \textit{X. laevis} oocytes and \textit{Xenopus} egg extracts for high-resolution NMR spectroscopy in eukaryotic cells.

Bodart \textit{et al.} \cite{54} studied the neuronal protein, Tau, inside \textit{X. laevis} oocytes, an environment in which it is not normally found \cite{75}. 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 $[U^{15}$N$]$ labeled in and purified from bacteria \cite{76,77}. The protein was microinjected into \textit{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 \textit{in vitro} \cite{78}, 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\textsuperscript{78}. 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.

To optimize the reproducibility of spectra acquired from different in-cell preparations the same solution was injected into the equivalent site on oocytes selected in a qualitatively reproducible manner. They concluded that while the overall spectral profile is improved, 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.

\textit{Cell-penetrating peptides}

Inomata \textit{et al.}\textsuperscript{57} utilized a heretofore untested procedure for introducing isotopically labeled proteins into eukaryotic cells by using cell penetrating peptides to transduce human HeLa cells. Uniformly [$U^{15}$N] labeled protein was covalently tagged or conjugated with a cell-penetrating peptide derived from the Tat protein of HIV-1 (CPP\textsubscript{Tat})\textsuperscript{79}, and incubated with human HeLa cells and pyrenebutyrate. Pyrenebutyrate mediates the translocation of CPP-tagged proteins into the cytosol\textsuperscript{80}. 
Experiments were performed using a Ubiquitin (Ubq) that was mutated at three sites (L8A, I44A, V70A)\textsuperscript{81} to preclude binding with ubiquitin-interacting proteins (UIPs). The $[U,^{15}\text{N}]$ labeled Ubq contained a C-terminal CPP fusion (Ubq-CPP\textsubscript{Tat}). Following transduction, a well-resolved $^1\text{H}\{^{15}\text{N}\}$-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 Ubq. Control experiments showed that Ubq-CPP\textsubscript{Tat} was cleaved between G76 and D77, presumably by endogenous ubiquitin-specific C-terminal proteases (DUBs)\textsuperscript{82} to yield free Ubq.

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\textsuperscript{83}. 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 Ubq-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. Ubq-CPP\textsubscript{Tat} was also transduced into monkey COS-7 cells, demonstrating the versatility of this method.

The in-cell spectrum of wild-type Ubq showed extensive peak broadening relative to the in-cell spectrum of Ubq and included the G76 cross-peak observed for cleaved
Ubq. This indicated that at least some of the wild-type Ubq existed in a C-terminally unconjugated state. The differences in peak intensity between the two spectra likely reflect the interaction of wild-type Ubq with endogenous proteins, since the mutated residues prevented binding with UIPs. A similar effect was seen in in-cell NMR experiments performed by using X. laevis oocytes\textsuperscript{55}.

Other methods of CPP-linked transduction were tested. CPPs linked to cargo proteins by using disulfide bonds\textsuperscript{84} are cleaved in the cytosol by autonomous reduction. The $^1$H-$^{15}$N-HSQC spectrum of the 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 Ubq.

A final delivery method was used to demonstrate the feasibility of studying drug-protein interactions. A $[^{15}$N] labeled fusion protein consisting of Ubq containing an N-terminal CPP\textsubscript{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\textsubscript{Tat}-Ubq 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\textsubscript{Tat}-Ubq 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 *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 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.

**Membrane permeabilization**

Ogino *et al.* 58 used the bacterial toxin streptolysin O (SLO) to deliver $^{15}$N-labeled actin-sequestering protein, thymosin β4 (Tβ4), into human embryonic kidney (293F) host cells. SLO forms 35 nm diameter pores on cholesterol-containing plasma membranes, allowing molecules up to 150 kDa to diffuse into and out of the cell; the pores are then repaired by adding Ca$^{2+}$ into the cytosol. Tβ4 conjugated with fluorescein isothiocyanate (FITC) and propidium iodide (PI), which stains cells with damaged plasma membranes were used, in conjunction with flow cytometry, to assess the efficiency of pore formation and resealing, and to estimate the intracellular concentration of Tβ4. In-cell NMR was then used to directly observe the protein in living cells.

Most of the SLO-treated cells were FITC-positive and PI-negative relative to untreated cells, indicating that the formation of the pores and subsequent resealing were successful. These cells provide a good in-cell NMR signal resulting exclusively from intracellular protein. FITC-positive and PI-positive cells were also observed; these cells produce undesired, extracellular NMR signals due to leakage from unrepaired pores. The
presence of FITC-negative cells indicates that the formation of pores was unsuccessful, and would result in no intracellular NMR signal.

The concentration of SLO was optimized to achieve a pore-forming efficiency of 50% with resealing efficiency of 70-80%. Higher concentrations of SLO resulted in more FITC-positive cells, but also more PI-positive cells. Confocal fluorescence microscopy and a sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE) revealed that the Tβ4 protein was evenly distributed throughout the cytoplasm and the nucleus, the cells remained viable, and Tβ4 did not undergo proteolytic degradation. By comparing the fluorescent signals of lysates from cells treated with and without SLO the intracellular concentration of Tβ4 was estimated to be 50 µM after incubating SLO-permeablisized cells with 1 mM Tβ4. Assuming 2 x 10⁷ cells containing 50 µM Tβ4 in 200 µL of NMR buffer, the authors estimate that the final concentration of Tβ4 in the NMR sample would be 4 µM.

In-cell \(^1\)H\(^{15}\)N-HSQC spectra were collected on cells suspended in CD293 medium containing 20% D₂O and 30% RediGrad, a colloidal medium used to stabilize the cell suspension\(^{54}\). After the experiment, the cells were sedimented and a HSQC spectrum of the supernatant showed that almost all of the in-cell signals were intracellular. The in vitro spectrum of Tβ4 differed from the in-cell spectrum; the chemical shifts were consistent with those of chemically synthesized Tβ4 with an N-terminal acetylation modification. Interaction with G-actin was ruled out as the no significant line broadening was observed and the affected residues were clustered near the N-terminus of Tβ4. Endogenous Tβ4 is known to be enzymatically acetylated in the cytosol\(^{85}\). The results show that this PTM occurs in 293F cells.
The study demonstrates an alternative way to introduce isotopically labeled protein into mammalian cells using SLO reversible permeabilization and Ca\(^{2+}\) resealing. The method is advantageous because it is applicable to many types of cells, does not require any synthetic PTMs or special equipment.

The methods described for delivering target proteins for in-cell NMR spectroscopy meet the requirements for accurately assessing drug-protein interactions. Intracellular concentrations of \(\leq 50\ \mu\text{M}\) can be typically achieved. Each target protein must be individually evaluated for optimum signal to noise ratio, leakage, folded state and interactions with other endogenous molecules before being used in the drug screening process.

**Methods and Applications**

Identifying a drug-protein interaction is merely the first step of a complex process. The structural details and consequent activity resulting from such an interaction depends on the stage of maturation of the target protein including its folded state and other structural characteristics arising from metabolic processing and PTMs, and finally the structures of the unbound and bound states. This section describes studies that have illuminated the properties of target proteins that are relevant to the drug discovery process.

*STructural INTeractions (STINT-NMR)*

The primary criterion for determining drug-protein interactions entails verifying that a specific interaction is taking place. Burz *et al.*\(^{86,87}\) developed an in-cell NMR-based method for mapping the structural interactions (STINT-NMR) that underlie protein
complex formation. The method entails sequentially expressing a target protein within a single bacterial cell in a time-controlled manner\(^8\) and monitoring its interaction with another protein or small molecule by using in-cell NMR spectroscopy\(^9\). The resulting NMR data provide a complete titration of the interaction and define the structural details of the interacting surface of the target protein at atomic resolution. Unlike the case when interacting molecules are simultaneously over-expressed in labeled medium, in STINT-NMR spectral complexity is minimized because only the target protein is labeled with NMR-active nuclei, which leaves the interactor molecule(s) cryptic.

The target protein, whose NMR structure must be known, is first overexpressed on uniformly labeled \([U^{\text{15}}\text{N}]\) medium to yield a high-resolution, isotope-edited \((^1\text{H}\{^{15}\text{N}\})\) HSQC backbone spectrum of the target protein inside the bacterial cells. The growth medium is changed, and the unlabeled interactor is overexpressed or delivered to the cytosol. 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 \([U-^{15}\text{N}]\) Ubq followed by over-expression of either one of two Ubq ligands containing the Ubiquitin Interacting Motif (UIM)\(^90\): a 28-amino acid peptide from Ataxin 3 (AUIM; \(~4\) kDa) or the Signal-Transducing Adapter Molecule (STAM2\(^91\); \(~50\) kDa) in unlabeled medium. AUIM binds Ubq \textit{in vitro} with \(~230\) \(\mu\)M affinity\(^92,93\) and STAM2, which contains two Ubq 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 \(^1\text{H}\{^{15}\text{N}\}\)-HSQC spectrum of Ubq, which was maintained at a single concentration, changed as the concentration of interactor was increased (Fig. 2.1). The chemical shift changes were mapped onto the three-dimensional structure of Ubq. 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 Ubq, demonstrated that the chemical shifts changes result from specific interactions and not merely over-expression of the interacting molecule.
**Figure 2.1 Interaction surface maps of Ubiquitin-ligand complexes.** Interaction surface of Ubq mapped onto the three-dimensional structure of Ubq (PDB code 1D3Z\textsuperscript{94}). 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 Ubq 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). Ubq ligands are indicated in each panel. Reproduced from Burz and Shekhtman\textsuperscript{87}. 
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 molecules. 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.

Protein Maturation

Evaluating the state of maturation of a target protein is critical to evaluating its interaction with drug candidates. Different structures may bind small molecules with varying affinities and the resulting structural complexes may exhibit profoundly different activities. Banci et al.\textsuperscript{95} used in-cell NMR to characterize the maturation of human copper, zinc superoxide dismutase 1 (hSOD1) in \textit{E. coli}. Apo-hSOD1 has been linked to familial amyotrophic lateral sclerosis (fALS), a fatal motor neurodegenerative disease\textsuperscript{95-97}. In its immature form, hSOD1 lacking one Zn\textsuperscript{2+} ion and one catalytic Cu\textsuperscript{+} ion per subunit and having a misfolded structure is believed to play critical role in ALS pathology\textsuperscript{98}. NMR experiments were performed to determine the folded state of apo-hSOD1 in the cytosol, to characterize zinc binding both in-cell and \textit{in vitro}, and to monitor how disulfide bond formation affects protein self-association.
To identify the folded state of apo-hSOD1, uniformly labeled protein over-expression was induced in metal-free $^{15}$N-labeled M9 medium and $^1$H{$^{15}$N}-Selective Optimized Flip-Angle Short-Transient heteronuclear multiple quantum coherence (SOFAST-HMQC)\textsuperscript{99} in-cell spectra were collected. Under favorable relaxation conditions, $^1$H{$^{15}$N}-SOFAST-HMQC can lead to up to ten times decrease in the NMR acquisition time, as compared to $^1$H{$^{15}$N}-HSQC, with minimal loss in signal sensitivity. The in-cell spectrum contains peaks between 7.5 and 8.5 ppm, implying that the protein is largely unfolded in the cytosol. Following in-cell NMR the cells were lysed and the spectrum was re-acquired. A few broadened amide peaks in the lysate spectrum indicate the presence of some structure; however most of the peaks correspond to unfolded protein. The lysate spectrum is similar to that of monomeric apo-hSOD1 (E,E-hSOD1$^{\text{SH-SH}}$), which contains reduced cysteines, and exhibits many of the peaks corresponding to the unstructured regions seen in the in-cell spectrum.

To determine if the protein is completely unfolded or partially unfolded inside the cell, the in-cell spectrum was compared to that of E,E-hSOD1$^{\text{SH-SH}}$ denatured with guanidinium chloride \textit{in vitro}. The \textit{in vitro} spectrum exhibits peaks only in the region expected for unfolded protein, but with different chemical shift values from that of the in-cell spectrum. For example, the cross peak of the side chain of W32, which is located in a β-strand of the hSOD1 β-barrel, has a chemical shift of ~10.3 ppm in-cell and ~10.1 ppm for \textit{in vitro} denatured protein. Furthermore, the in-cell chemical shift of W32 is the same as in the lysate spectrum suggesting that apo-E,E-hSOD1$^{\text{SH-SH}}$ is not completely unfolded in the cytosol and that cross-peaks attributed to folded protein are broadened due to chemical or conformational exchange phenomena.
To rule out that the loss of signal was due to chemical or conformational exchange, $^1$H($^{15}$N)-SOFAST-HMQC spectra were collected at 500 MHz (315 K) and at 800 MHz (288 K & 298 K) to monitor the change in the signal to noise ratio, which is expected to increase with increasing magnetic field strength and temperature. No increase of the S/N ratio was observed suggesting that exchange phenomena were not responsible for the loss of signals.

Signal loss or line broadening may also be due to oligomerization or interactions with other cellular components. The formation of soluble apo-hSOD1 oligomers in vitro leads to the complete loss of signals in the HSQC spectrum. Therefore the presence of in-cell signals combined with the recovery of missing signals upon cell lysis indicates that no apo-hSOD1 oligomers are forming in the cytosol. Because apo-hSOD1 is present in at high intracellular concentrations, it is capable of engaging in a myriad of non-specific interactions. Each interaction places apo-hSOD1 in different chemical environments with correspondingly different chemical shifts. The resulting cellular anisotropy gives rise to inhomogeneous broadening in which cross-peaks from folded regions exhibit a larger chemical shift dispersion that is undetectable by NMR.

Binding of zinc to a native site on the protein is critical for maturation of hOSD1. To understand the metal binding properties of hSOD1, protein over-expression was induced in $^{15}$N-labeled M9 medium supplemented with 10 µM to 1 mM ZnSO$_4$. The in-cell $^1$H($^{15}$N)-SOFAST-HMQC spectrum of hSOD1 expressed in zinc medium was different from that of in-cell apo-hSOD1. The appearance of dispersed peaks and the broadening of some signals in the unfolded region of the spectrum indicate that in cell hSOD1 binds zinc when zinc is in excess relative to the total amount of protein over-
expressed in the cell. The in-cell spectrum is very similar to that of \textit{in vitro} E,Zn-hSOD1^{SH-SH}, (the species with one zinc bound to the zinc binding site) and not with that of \textit{in vitro} Zn,Zn-hSOD1^{SH-SH} (the non-physiological species with two zinc ions bound to both metal binding sites). Amide signals from G61 and T135, located near the metal binding sites and indicators of binding site occupancy, have chemical shift values that are close to those of the \textit{in vitro} E,Zn-hSOD1^{SH-SH} spectrum and further from the signals corresponding to the \textit{in vitro} spectrum of Zn,Zn-hSOD1^{SH-SH}. The results demonstrate that intracellular hSOD1 exhibits greater selectivity for metal ion binding than \textit{in vitro} because only the native zinc binding site is occupied in-cell when the concentration of zinc ions is in molar excess. This is unlike what is observed at physiological pH \textit{in vitro} in which sub-stoichiometric amounts of zinc give rise to a mixture of three species, E,E-hSOD1^{SH-SH}, E,Zn-hSOD1^{SH-SH} and Zn,Zn-hSOD1^{SH-SH}, and when 2 equivalents of zinc are present per protein subunit, only Zn,Zn-hSOD1^{SH-SH} is formed. Following the in-cell experiments, the cells were washed with metal-free medium and NMR spectra were collected on the cleared lysates. Only the Zn,Zn-hSOD1^{SH-SH} species was detected, indicating that there was an excess of zinc still present inside the cells and zinc was bound to the hSOD1 copper binding site.

To determine the redox state of the cysteines, $^1$H\{15N\}-SOFAST-HMQC spectra were acquired on both in-cell and \textit{in vitro} samples grown in metal-free media. Monitoring the $^1$H-15N signals of selectively labeled cysteines, showed that in apo-hSOD1, all four are in a reduced state in the cytosol and in cell lysates. Following exposure to air, oxidized C57 and C146 displayed chemical shifts consistent with an oxidized state and identical to those of \textit{in vitro} dimeric E,E-hSOD1$_{S}$$_{95}$. Upon reduction by adding
dithiothreitol (DTT), C57 and C146 exhibit peaks consistent with the reduced state observed in-cell and in cell lysates. When the same analysis was performed for cells and lysates in which hSOD1 was expressed in the presence of Zn(II), the C57 and C146 residues of E,Zn-hSOD1 were found to be reduced in the cytosol.

Disulfide bond formation is tightly linked to dimerization. The spectra of $^{15}$N-cysteine labeled protein provide evidence on the assembly state of hSOD1 in the cytosol. In vitro, E,E-hSOD1$^{SH-SH}$ is monomeric, while E,E-hSOD1$^{S-S}$ and E,Zn-hSOD1$^{SH-SH}$ are homodimers. The in vitro NMR spectra of both E,E-hSOD1 and E,Zn-hSOD1 exhibit a large chemical shift difference for C146 between the oxidized and reduced states, because C146 is directly involved in disulfide formation. C6 is used as a marker to indicate the state of assembly because C6 is located close to the interaction surface of the homodimer. Upon E,E-hSOD1 dimerization and consequent disulfide bond formation, the chemical environment of C6 changes. For E,Zn-hSOD1, which is dimeric regardless of the oxidation state, the chemical shift of C6 remains unchanged. The chemical shift of C6 for both in-cell E,E-hSOD1$^{SH-SH}$ and E,Zn-hSOD1$^{SH-SH}$ species matches that of the corresponding in vitro species, which suggests that, in the cytosol, E,E-hSOD1$^{SH-SH}$ is monomeric and E,Zn-hSOD1$^{SH-SH}$ is dimeric.

This study investigated the maturation of human superoxide dismutase 1. The processes observed included determining the folded and assembly states of the protein, the redox state of key cysteine residues as an indicator of quaternary structure, and zinc binding to native and non-native sites on the protein. The work demonstrates that linked functions necessary for protein maturation can be identified and dissected from one another by using in-cell NMR spectroscopy. The results provide insight into the myriad
of structures that are available for drug binding interactions that may affect the onset of fALS.

**Metabolic processing of proteins**

The effect of metabolic processing on in-cell NMR spectra was examined by Sakai *et al.* Reduced spectral quality resulted from interactions involving the target protein due to endogenous enzymatic activity and binding of metal ions. The results highlight the effect of protein-protein interactions on NMR spectra and the structure and activity of the resulting species. Uniformly \^[U-15]N\] labeled Ubq and its derivatives were expressed and purified in bacteria. The purified protein was microinjected into *X. laevis* oocytes to a maximum final concentration of ~100 μM. The magnitude of the signal for the \[^1^H\{15^N\}\]-HSQC spectrum of wild-type Ubq injected into the oocytes was very weak. Control experiments verified that the Ubq spectrum originated from an intracellular environment and not from labeled protein that had leaked from the cells into the surrounding medium.

Experiments performed using a series of mutant Ubq, in which residues implicated in binding UIPs, L8, I44 and V70, were changed to alanines. The results showed that the in-cell \[^1^H\{15^N\}\]-HSQC spectrum of Ubq could be largely recovered when the UIP interface is perturbed. The perturbation disrupts protein-protein interactions, which decreases peak broadening and resolves the spectrum. Ubq molecules carrying single mutations affecting the UIP binding site (L8A-D77, I44A-D77 & V70A-D77) partially restored the in-cell Ubq \[^1^H\{15^N\}\]-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 Ubq results from its
interacting with endogenous UIPs, preventing a sufficient in-cell concentration of free Ubq for analysis by NMR spectroscopy.

The Ubq used in this study contained a C-terminal D77 residue (Ubq-D77), whereas mature Ubq has a G76 in that location\textsuperscript{101}. The D77 protein, therefore, mimics a Ubq 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 Ubq-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 Ubq aldehyde, which specifically inhibits UCH\textsuperscript{101}, the in-cell spectrum of $^{15}$N-Ubq-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 Ubq processing. These observations suggest that in-cell NMR spectroscopic analyses of metabolic processing may be possible under select conditions.

Sakai \textit{et al.}\textsuperscript{55} also microinjected [U-,\textsuperscript{15}N] labeled Calmodulin into oocytes. The in-cell $^1\text{H}(^{15}\text{N})$-HSQC spectrum of Calmodulin, acquired without Ca\textsuperscript{2+} in the buffer, resembled that of apo-Calmodulin acquired \textit{in vitro}, except the peaks were broader, indicating that the majority of the in-cell Calmodulin was Ca\textsuperscript{2+-free}. When an excess of Ca\textsuperscript{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\textsuperscript{2+}-bound Calmodulin were identified. Since Ca\textsuperscript{2+}-bound Calmodulin is more likely to interact with downstream effector proteins\textsuperscript{102,103} than apo-Calmodulin, the reduced spectral quality observed in the presence of Ca\textsuperscript{2+} suggests interactions between injected and endogenous proteins.

*Regulation of Post-translational modifications (In-cell biochemistry)*

PTMs can alter the structure, reactivity and activity of proteins. Each resulting conformation has the potential to interact with drug-like molecules differently to produce different outcomes. Burz and Shekhtman\textsuperscript{104} developed an in-cell methodology to introduce 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\textsuperscript{86,87}. By creating specific populations of post-translationally modified target proteins, the contribution of these modifications to drug-protein interactions can be quantitatively evaluated.

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 to identify changes in the interaction surface of Ubq resulting from these post-translational modifications. STAM2 and Hrs are components of an endocytic pathway present in eukaryotic cells. Each binds Ubq via an UIM\textsuperscript{105}. In addition, STAM2 has a VHS domain capable of binding Ubq. 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\textsuperscript{106}. 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”.

To study Ubq binding to STAM2 and Hrs, [\textsuperscript{15}N]Ubq 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 Ubq revealed no changes in the interaction surface when bound to non-phosphorylated or phosphorylated Hrs (Fig. 2.2). The spectrum of Ubq 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. 2.2). 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\textsuperscript{107}.

A similar result was obtained for the interaction between Ubq and the STAM2-Hrs heterodimer: Ubq 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. 2.2). 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 Ubq-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 PTMs in an environment that normally lacks the ability to provide such modifications, *i.e.* bacterial cells, affords an opportunity to examine the effects of PTMs on protein structure without competing reactions, thus tailoring the ability of the target to interact optimally with the desired drug candidate. The methodology can be applied to any stable target molecule and may be extended to include other post-translational modifications.
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] Ubq and 0 hr (black), 2 hr (red) and 3 hr (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] Ubq (black) and $[U-^{15}$N] Ubq-AUIM complexes at a molar ratio of 1:1 (red) and 1:2 (blue). Reproduced from Burz et al.$^{86}$. 
De novo structure determination

The exact conformation of an intracellular protein or protein complex may be quite different from what is observed in vitro, and the structural changes may affect the resulting activity. Therefore, determining the structure of a drug-protein complex in-cell is critical to the drug discovery process. 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\textsuperscript{53} employed a novel non-linear sampling scheme\textsuperscript{54,75,76} to solve the de novo structure of TTHA1718, a 66 amino acid, putative heavy-metal binding protein from T. thermophilus. By combining this scheme with maximum entropy processing\textsuperscript{108}, 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 an intracellular 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 fresh samples prepared for each experiment; 86% of Hα, 71% of Hβ and 34% of the aliphatic $^1$H/$^{13}$C side chain resonances were identified. NMR spectra collected in vitro using purified protein was used to confirm assignments from in-cell NMR. After 6 hours of data acquisition colony plating tests showed that 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 reference \(^{1}H\{^{15}N\}\)-HSQC spectrum 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 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\(^{109}\). The final calculated structure and was similar to the structure determined \textit{in vitro} for purified TTHA1718, and had a root mean square 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 work was possible because of innovations that allow rapid data collection and unambiguous identification of long-range NOE interactions based on selective labeling of methyl groups. Although eukaryotic cell stability is still a limiting factor, in-cell structure determination in eukaryotic cells may be possible since labeled proteins can be introduced into \textit{Xenopus} oocytes at concentrations up to 0.7 mM\(^{56}\). In-cell determination of the structure of drug-protein complexes will provide a linchpin to understanding the mechanisms of drug activity.
Drug screening

Efficient delivery of drug candidates to the intracellular medium is a necessary component of the drug discovery process. In addition to monitoring the target protein and identifying potential structures for drug design, the ability to control the concentration of drug candidate species during in-cell screening is key to evaluating drug-protein interactions and the effectiveness of potential drug therapies. This section details in-cell NMR-based strategies that apply to the drug discovery process including the delivery of drugs to intracellular targets, chemical shift mapping of changes in the target or drug upon binding and, a high-throughput method to rapidly screen small molecule libraries for compounds capable of strengthening or weakening protein-protein interactions within a biomolecular complex.

In-cell delivery of drug molecules

Efficient delivery of drugs to the intracellular medium and the ability to control the concentration of drug candidate species during in-cell screening is key to evaluating drug-protein interactions and potential drug therapies. Arnesano et al.\textsuperscript{110} used solution and in-cell NMR spectroscopy to explore intracellular drug delivery and the interaction of cisplatin (cis-[PtCl\(_2\)(NH\(_3\))\(_2\)]) with Atox1, a human copper chaperone that mediates Cu(I) delivery to copper-transporting P-type ATPases\textsuperscript{111,112}. Cisplatin is a platinum based anticancer drug that forms a stable complex with DNA to interfere with DNA replication and transcription processes\textsuperscript{113}. Cellular uptake of Pt drugs is tightly connected to Cu transport\textsuperscript{110,114} by virtue of surface-exposed metal binding sites, which are used to bind metallo-drug compounds. Atox1 has a \(\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4\) ferredoxin-like structure and a metal-
binding site, CxxC, located in the loop between $\beta_1$ and $\alpha_1$, which is highly conserved among metallochaperones and soluble Cu-transport ATPases$^{112,114}$.

To identify the residues of Atox1 that are involved in Pt coordination $^1$H$^{15}$N-HSQC NMR spectra were collected on a purified $[^{15}$N] Atox1 in the absence and presence of unlabeled cisplatin. Chemical shift changes were localized to the CxxC Cu(I)-binding region of Atox1 (Fig. 2.3). $^1$H$^{66}$-HSQC NMR spectra, in which cysteine residues were selectively labeled, revealed fast-to-intermediate exchange between free and coordinated Atox1 in the presence of cisplatin, due to the complete broadening of cross-peaks Cys12 and Cys15. To monitor the fate of the amine ligands released by cisplatin upon interaction with Atox1, $^1$H$^{15}$N-HSQC NMR spectra were collected on $[^{15}$N] cisplatin in the absence and presence of unlabeled Atox1. In the presence of Atox1 the amine signal decreased as the reaction progressed, indicating the release of ligand.

In-cell $^1$H$^{15}$N-SOFAST-HMQC spectra$^{99}$ collected on $[^{15}$N] Atox1 over-expressed in E. coli was similar to the in vitro spectrum and consistent with a reduced apo state for Atox1. Cross-peaks for the in-cell spectrum were broadened relative to the in vitro spectrum of purified protein. Controls indicate that there was no leakage of protein from the cells. Purified Atox1 is sensitive to oxidation and must be handled with precautions. Typically Atox1 is handled in the presence of a reducing agent such as DTT or tris(2-carboxyethyl) phosphine (TCEP), however, these materials are good ligands for Pt(II); this problem was overcome by observing Atox1 in-cell where the intracellular environment provided suitable conditions for the preserving the protein in its active form.
To determine whether Pt drugs could enter the cell and interact with Atox1, \([U-^{15}\text{N}]{\text{Atox1}}\) was over-expressed in *E. coli* grown in the presence of Pt drugs. Cisplatin concentrations >10 \(\mu\text{M}\) inhibited cell growth thus setting an upper limit to the extracellular concentration used in these experiments. The uptake of cisplatin was estimated by atomic absorption spectroscopy to be \(~10\%\). In-cell, \(^1\text{H}\{^{15}\text{N}\}\text{-SOFAST-HMQC}\) spectra were collected on cells grown with and without cisplatin. The largest spectral changes were observed for residues that comprise the CxxC metal-binding motif, Cys12, Cys15 as well as Ala16. The changes are similar to those observed for the cisplatin-Atox1 interaction *in vitro* suggesting that cisplatin entered the cell and was bound to Atox1 (Fig. 2.3) without inducing any major structural re-arrangements.

This work highlights the advantages of in-cell NMR spectroscopy to probe intracellular drug delivery by monitoring a known drug-protein interaction. The binding-competent form of the protein target was identified and the fate of the ligand upon binding was elucidated. In addition, the intracellular milieu provided a stable environment for the otherwise labile drug, confirming its efficacy *in vivo*. 
**Figure 2.3** Atox1-cisplatin complexes in *E. coli*. (A) Overlay of $^1$H{$^{15}$N}-HSQC spectra of free [U-$^{15}$N] Atox1 in the absence (red) and in the presence (blue) of 1 mol equiv of cisplatin 4 hr after mixing. (B) Overlay of $^1$H{$^{15}$N}-SOFAST-HMQC spectra of *E. coli* after 4 hr induction of [U-$^{15}$N] Atox1 in the absence (red) and in the presence (blue) of 10 μM cisplatin. Cys12, Cys15, and Ala16 cross-peaks are indicated with arrows. Reproduced from Arnesano *et al.*$^{110}$. 
Screening of small molecule interactor library (SMILI-NMR)

The ability to rapidly screen large numbers of candidate compounds, i.e. high throughput screening (HTS), presents one of the major challenges to the drug discovery process. HTS in combination with in-cell NMR requires minimal sample preparation and eliminates the need for extensive protein purification. The technique takes advantage of the fact that drug candidates can pass through the cell membrane to interact with the cytosolic target. Furthermore, screening a Small Molecule Interactor Library (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.

Xie et al.\textsuperscript{63} developed an in-cell NMR spectroscopy-based screening procedure, 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\textsuperscript{86,87} technology to produce biomolecular complexes inside the cell in which one of the constituent proteins is uniformly $[U^{\text{-15}}\text{N}]$ 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. In complex with rapamycin, a macrolide antifungal antibiotic currently undergoing clinical trials for a variety of cancer treatments, FKBP binds to FRB. When [U-15N] FKBP is over-expressed in bacterial cells, the 1H{15N}-HSQC spectrum shows no well-resolved peaks, implying that the single species is part of a large complex and therefore invisible to NMR. When unlabeled FRB is then over-expressed in the same cells, the NMR spectrum of FKBP becomes evident but only at the highest FRB concentrations, indicating the formation of a complex. Similar results were obtained when [U-15N] FRB and unlabeled FKBP were sequentially over-expressed in the same cells. These observations demonstrated that creating a proper protein complex is necessary for high-resolution studies.

Adding rapamycin to the cell suspension results in visible changes in the 1H{15N}-HSQC-spectrum of [U-15N] FKBP indicating the formation of a high-affinity ternary complex between FKBP-rapamycin and FRB (Fig. 2.4). Adding rapamycin to cells overexpressing labeled FKBP in the absence of FRB or labeled FRB in the absence of FKBP does not produce an NMR spectrum. In each case, co-expression of the second protein is required to generate an in-cell NMR spectrum.

A dipeptide chemical library composed of 17x17 dipeptides was chosen to provide a collection of compounds that are capable of interacting with the target molecule at a detectable level. These compounds provide suitable starting points for subsequent optimization into credible drug candidates and are considered as potential sources of novel lead structures. Dipeptides have been shown 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\textsuperscript{120}, and (4) dipeptides can pass through the cell membrane and interact with the target protein directly \textit{in vivo}\textsuperscript{122,123}.

A standard procedure called the matrix method, in which compounds located in one row or one column of a matrix plate are mixed and tested, is used to screen the library. Individual mixtures are examined for their ability to change the in-cell NMR spectrum of the FKBP-FRB bio-complex. Samples exhibiting similar spectral changes, located at the intersection of rows and columns, are used in the second round of screening to deconvolute and validate the initial findings. In this way, a matrix of 289 (17x17) compounds can be screened by examining 34 (17+17) samples.
**Figure 2.4** (Top) Ternary FKBP-FRB-ligand complexes in *E. coli*. 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 (black) (A) in the presence of 150 μM rapamycin (red) and (B) the presence of 5 mM A-E (red). (Bottom) Yeast assay for biological activity of the dipeptide, A-E. Isogenic haploid yeast strains (*S. cerevisiae*) that express (*FPR1*) or lack (*frp1*) the FKBP proline isomerase were grown for 3 days on YPD medium. (C) Control plate. (D) 100 μM rapamycin. Expression of FKBP allows the formation of a toxic FKBP-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 FKBP. Reproduced from Xie *et al.*^63^.
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. 2.4). 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 (Fig. 2.4).

SMILI-NMR provides an important means to bridge the gap between biochemical identification of small ligands capable of interfering with target bio-complexes and the biological activity resulting from the inhibition of cellular processes by these ligands. The extension of SMILI-NMR to eukaryotic cells, in which compartmentalization may be important for drug-protein interactions, will rely on the development of new techniques for monitoring protein targets in eukaryotic cells.

Conclusions

In-cell NMR offers many possibilities for drug development as improvements in instrumentation and new procedures for introducing target proteins and drug candidates into prokaryotic and eukaryotic cells are developed. In-cell NMR reveals how the
intracellular environment influences molecular structure and how, in turn, these structures affect protein-drug interactions. Understanding the details of such interactions can lead to the design of new molecules capable of binding with high affinity to proteins at different stages of spatial and temporal maturation. Indeed, it is at this level of specificity where the efficacy of drug activity is best optimized, and it is through the use of in-cell NMR that these properties are best examined.
CHAPTER THREE

CAUGHT IN ACTION: SELECTING PEPTIDE APTAMERS AGAINST INTRINSICALLY DISORDERED PROTEINS IN LIVE CELLS

Jacqueline Washington¹, Eric A. Smith², Sergey Reverdatto¹, David S. Burz¹, Kathleen A. McDonough², Alexander Shekhtman¹*

¹ Department of Chemistry, University at Albany, State University of New York, Albany, New York, United States of America

² Wadsworth Center, New York State Department of Health, Albany, New York, United States of America
Introduction

As much as a 5% of prokaryotic and, up to 21% of eukaryotic genomes, encode proteins or large segments of proteins that lack a stable secondary or tertiary structure\textsuperscript{124}. These intrinsically disordered proteins (IDPs) or disordered regions can be highly conserved between species and are often functional\textsuperscript{125-127}. IDPs have the ability to bind to multiple proteins and play roles in the assembly of macromolecular arrays\textsuperscript{128,129}. IDPs are implicated in neuropathology, autoimmune diseases and cancers and are considered to be promising targets for drug therapy\textsuperscript{130}. Despite wide recognition of IDPs and the nascent secondary structures that exist in the free state, the molecular mechanisms underlying the functions of IDPs are not fully understood.

There is a limited arsenal of molecular tools that can be applied to study interactions of IDPs in a physiological milieu\textsuperscript{130,131}. Due to the intrinsic flexibility of IDPs, the powerful methodology of using small molecules to block surfaces engaged in protein-protein interaction can only be applied to well-folded IDPs binding partners. Since many IDPs fold upon binding to their cognate targets\textsuperscript{127,128,132}, site directed mutagenesis, which creates local changes in the chemical structure of the IDP, cannot be effectively used to interrogate the usually large IDP-target interaction surfaces\textsuperscript{128,133}.

Recently we introduced a Combinatorial Library of Improved Peptide aptamers (CLIPs) containing more than $3 \times 10^{10}$ peptide aptamers, as a tool to isolate site-specific binders to select cellular targets\textsuperscript{134}. Peptide aptamers\textsuperscript{135,136} (PA) are small $\sim 12$ kDa proteins that consist of a randomized 8 amino acids peptide sequence inserted into the loop of a modified thioredoxin platform. The modification of the thioredoxin platform\textsuperscript{134}
improved solubility and decreased the tendency of PAs to oligomerize, creating molecules with drug-like binding abilities\textsuperscript{137}.

PAs are selected against a target by using the yeast-two-hybrid (Y2H) technique\textsuperscript{135,136} due to their ability to bind a broad range of cellular, viral and bacterial target proteins with high specificity and strong affinity\textsuperscript{137}. During Y2H selection, individual PAs from CLIPs do not compete with each other for binding sites; thus, it is possible to select PAs with vastly different affinities for closely located or even overlapping target sites. CLIPs technology is an attractive molecular tool to analyze the functional consequences of nascent structures present in free IDPs due to the \textit{in vivo} nature of target selection and comprehensive coverage of the potential binding surfaces.

In-cell NMR is used to analyze nascent structures of IDPs inside live cells at atomic resolution\textsuperscript{138,139}. STINT-NMR is an in-cell technique developed to study protein-protein structural interactions inside live cells\textsuperscript{86,87}. We used STINT-NMR to define the interaction surfaces between a small IDP, the prokaryotic ubiquitin-like protein, Pup\textsuperscript{140}, and selected PAs in the crowded cytosol. Unlike its eukaryotic counterpart, Ubq, Pup is unstructured\textsuperscript{141}. Since both selection and structural characterization of PA-IDPs complexes are made inside cells, we expect that this technology will be highly relevant to identify functional complexes between IDPs and its binding targets.

The \textit{Mycobacterium tuberculosis}, Mtb, proteasomal system is an attractive target for drug therapy against latent Mtb infection since it provides defense against the reactive nitrogen intermediates generated against bacteria inside phagosomes\textsuperscript{142,143}. Indeed, potent proteasomal inhibitors, Bortezomib and GL5, increase sensitivity of Mtb to nitro-oxidative stress\textsuperscript{143}. The Mtb proteasomal system consists of: the Mtb core particle, a
proteasomal ATPase, Mpa, Pup which targets proteins for degradation, the Pup ligase, PafA, and Pup deaminase/depupylase, Dop. We used the Mtb proteasomal system to illustrate how the combination of CLIPs and STINT-NMR technologies allows us to select specific binders to Pup and use these molecules to analyze functional interactions of Pup inside mycobacteria.

Materials and Methods

Reagents and chemicals

Restriction enzymes and Taq DNA Polymerase and Phusion Polymerase were from New England BioLabs. Pfu DNA Polymerase was from Agilent Technologies. All other chemicals used were reagent grade or better.

Yeast strains

Matchmaker Gold yeast-hybrid system 2 and vectors, pGBK7 and pGAD77, were obtained from Clontech. Yeast strain Y2HGold (MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2:: GAL1_uas–Gal1_tata–His3, GAL2_uas–Gal2_tata–Ade2, URA3:: MEL1_uas–Mel1_tata, AUR1-C MEL1) and pJ69-4a (trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2::GAL1-HIS3, GAL2-ADE2, met2::GAL7-lacZ) were used for peptide screening and grown in yeast peptone dextrose adenine (YPDA) growth medium. Four reporter genes, AUR1-C, ADE2, HIS3, and MEL1, under the control of GAL4 upstream activating sequences (UASs) and TATA boxes, were used to select PAs and eliminate false positives. AUR1-C confers strong resistance to the highly toxic antibiotic, Aureobasidin A (Aur A). ADE2 and HIS3 provide metabolic
selection, and *MELI*, which encodes α-galactosidase, is used for X-α-galactosidase (X-α-gal) selection.

**Plasmid Construction**

The peptide aptamer (PA) library generated in our lab was used to screen against Pup\textsuperscript{134}. The library contains approximately $3\times10^{10}$ peptide aptamers cloned into a pGADT7 to yield a pGADT7-thioredoxin-peptide aptamer library construct, in which an 8-amino-acid randomized peptide is inserted into a constrained loop of D26A, K57Q thioredoxin\textsuperscript{134}. The thioredoxin is fused in frame with amino acids 768-881 of the GAL4 activating domain (AD). The plasmid has a hemagglutinin (HA) epitope tag, a 2μ origin, and confers ampicillin resistance for selection in *E. coli* and LEU2 for selection in yeast.

DNA encoding Pup (64 amino acids) was PCR amplified from *Mtb* genomic DNA using *Taq* DNA Polymerase and oligonucleotides 5’-TTTTTTCATATGGCGCAAGAGCAGACCAAGC-3’ and 3’-TTTTTTGTCGACTCAGTCCGCCCTTTTGG-5’, which contain flanking 5’-*NdeI* and 3’-*SalI* restriction sites. The restriction-digested PCR product was ligated into expression vector pGBK7 to yield pGBK7-Pup in which *Pup* is fused in frame with amino acids 1-147 of the GAL4 DNA binding domain (DNA-BD) (Fig. 3.1). The expressed protein contains a N-terminal c-Myc epitope tag and confers kanamycin resistance for selection in *E. coli* and TRP1 for selection in yeast.

DNA encoding Mpa (613 amino acids) was PCR amplified from *Msm* genomic DNA using Phusion DNA Polymerase and oligonucleotides 5’-AAAAAAGGTACCATTGAGTAGTCTAGA-3’ and 3’-AAAAAAGGTACCAGGTACTGGCCCAG-5’, which contains flanking 5’ and 3’-*KpnI*
restriction sites. The restriction-digested PCR product was ligated into pGADT7 to yield pGADT7-Mpa in which Mpa is fused to a GAL4 activating domain (AD). The expressed protein contains a C-terminal HA epitope tag and ampicillin resistance for selection in E. coli and LEU2 for selection in yeast (Fig. 3.1).

DNA encoding each individual peptide aptamer was PCR amplified from pGADT7 using Taq DNA Polymerase and oligonucleotides 5’-CGGGGCCATCCTCGTCGCTTTCTGG-3’ and 3’-GGTTTTGATCGATGTTCAGTTGTGCA-5’. Pfu DNA Polymerase and PCR products were used to clone peptide aptamers into a modified pBAD expression vector, which confers ampicillin resistance. The resulting plasmids, pBAD-PA, express N-terminal 6xHis-tagged scaffold protein, thioredoxin (Fig. 3.1).

DNA encoding each peptide aptamer was PCR amplified using Phusion DNA Polymerase and oligonucleotides 5’-TTTTTTTCATATGGGCCATAAAATTATT-3’ and 3’-CCCAAGCTTACGCCAGTTAGCGTCG-5’, which contain flanking 5’-NdeI and 3’-HindIII restriction sites. The restriction-digested PCR product was ligated into pVV16 to yield pVV16-PA. The plasmid is constitutively active and confers kanamycin resistance for selection in E. coli and mycobacterial strains. Each aptamer was handled in the same manner (Fig. 3.1).
Figure 3.1 Plasmids Used for PA Selection and Analysis.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Size (kb)</th>
<th>Selection</th>
<th>Promoter</th>
<th>Vector Origin</th>
<th>Protein Expressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGBKT7-Pup</td>
<td>7.4</td>
<td>TRP1</td>
<td>(truncated) ADH1</td>
<td>E. coli: pUC</td>
<td>Pup-GAL4-Binding Domain</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yeast: 2 μ</td>
<td></td>
</tr>
<tr>
<td>pGADT7-Mpa</td>
<td>9.7</td>
<td>LEU2</td>
<td>ADH1</td>
<td>E. coli: pUC</td>
<td>Mpa-GAL4-Activating Domain</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yeast: 2 μ</td>
<td></td>
</tr>
<tr>
<td>pBAD-Thio-PA</td>
<td>4.3</td>
<td>Ampicillin</td>
<td>araC</td>
<td>E. coli: pBR322</td>
<td>Thioredoxin-Peptide aptamer</td>
</tr>
<tr>
<td>pVV16-Thio-PA</td>
<td>6.1</td>
<td>Kanamycin</td>
<td>hsp60</td>
<td>E. coli: pUC19</td>
<td>Thioredoxin-Peptide aptamer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mycobacterial: pAL5000</td>
<td></td>
</tr>
</tbody>
</table>
Verification of reporter gene activation

Plasmid pGBKT7-Pup (bait) was transformed into yeast strain Y2HGold and plasmid pGADT7-Mpa (prey) was transformed into yeast strain Y187. Transformants were mated on YPDA agar plates and mated transformants were selected on synthetic dropout media (SDM) lacking tryptophan and leucine, SDM lacking tryptophan, leucine, and histidine, SDM lacking tryptophan, leucine, and adenine, SDM lacking tryptophan and leucine supplemented with X-α-gal, and SDM lacking tryptophan and leucine supplemented with Aur A, to verify the activation of reporter genes by these constructs.

Library transformation and screening

Plasmid pGBKT7-Pup was used as bait to screen a CLIPs by using the Matchmaker Gold yeast-hybrid system 2 protocol. The CLIPs was transformed into yeast strain Y2HGold harboring pGBKT7-Pup by using the lithium acetate method. Transformants were selected initially for growth on ade- trp- leu- plates. Positive clones were streaked on trp- leu- his- plates; ade- trp- leu- plates; trp- leu- plates supplemented with X-α-gal; and trp- leu- plates supplemented with Aur A, to test the stringency of the interaction with Pup.

Expression and purification of aptamers for in vitro assays

For fluorescence titration experiments, E. coli strain DH10B was transformed with pBAD-PA and grown in M-505 auto-inducing medium containing 0.25% L-Arabinose, and 150 mg/L of carbenicillin. All PA plasmids were treated in the same manner. For [U-15N] labeling of PAs, E. coli strain Origami B (Novagen) was transformed with pBAD-PA and grown in N-505 auto-inducing medium containing
2.66 g/L of [U-^15^N] ammonium chloride as the sole nitrogen source, 0.25% L-Arabinose, and 150 mg/L or carbenicillin. Cultures were grown at 37 °C for 20-26 h. Cells were harvested, re-suspended in extraction buffer (50 mM sodium phosphate, pH 7.0, 300 mM sodium chloride, 6 M guanidine hydrochloride) and sonicated. The lysate was cleared by centrifugation, and the supernatant was loaded onto a pre-equilibrated TALON column (Clontech). The column was washed with 10 column volumes (CV) of extraction buffer, 10 CV of Extraction buffer containing 0.4% octyl phenol ethoxylate (Triton X-100, Baker), and again with 10 CV of Extraction buffer. Protein was eluted with 10 CV of 1X Imidazole Elution buffer (45 mM sodium phosphate, pH 7.0, 270 mM sodium chloride, 5.4 M guanidine hydrochloride, 150 mM imidazole). Eluted protein was concentrated to 2-2.5 mL by using an Amicon Ultra-15 centrifugal filter (Millipore) with a 3 kDa molecular weight cutoff (MWCO), and supplemented with 50 µL of 0.5 M EDTA, 6 mL of TSP buffer (20 mM sodium phosphate, pH 7.5, 100 mM sodium thiosulfate) and 1 mL of glycerol. Protein samples were dialyzed against 1 L of 50 mM sodium phosphate, pH 8.0, 100 mM sodium chloride, 0.5 M guanidine hydrochloride, 2 mM EDTA, and 10% glycerol, for 5-10 hours at 4 °C in a 1 kDa MWCO dialysis bag (Spectrum Laboratories, Inc.). Samples were further dialyzed against 2 L of TSP buffer supplemented with 10% glycerol, for 10 hours at 4 °C. Finally, peptide aptamers were concentrated by using an Amicon Ultra-15 centrifugal filter with a 3 kDa MWCO, exchanged into TSP buffer and stored on ice. The peptide aptamers were estimated to be >95% pure by Coomassie-stained SDS-PAGE.
Expression and purification of Pup for in vitro assays

*E. coli* strain BL21(DE3) Codon+ (Novagen) was transformed with pTM-Pup, which encodes a tryptophan leader sequence (trpL) before Pup\(^{147}\). Cells were grown to an OD\(_{595}\) of 0.5 in minimal medium (M9 salts, 10 μM calcium chloride, 1mM thiamine hydrochloride, 2 mM magnesium sulfate, and 0.4% glucose) containing 35 mg/L of kanamycin, induced with 0.5 mM isopropyl 1-thio-β-d-galactopyranoside (IPTG), and grown overnight at 37 °C. Cells were harvested, re-suspended in 20 mM HEPES-Na, pH 7.0 buffer containing 8 M urea, and sonicated. The lysate was cleared by centrifugation, and the supernatant was loaded onto a nickel-nitrilotriacetic acid (Ni-NTA) agarose column (Qiagen). The column was washed with 10 CV of wash buffer (20 mM HEPES-Na buffer, pH 7.0, containing 8 M urea), and the protein was eluted with elution buffer (20 mM phosphate buffer, pH 4.0 containing 8 M urea). Eluted protein was concentrated to 2-2.5 mL by using an Amicon Ultra-15 centrifugal filter (Millipore), with a 3 kDa MWCO. The concentrated sample was dialyzed against 1 L of 10 mM sodium phosphate buffer, pH 6.5, for 3-4 h at 4 °C in a 1 kDa MWCO dialysis bag (Spectrum Laboratories, Inc.). Formic acid was added to the resultant sample to a final concentration of 70% and the N-terminal 6xHis-tag and trpL\(^{147}\) were removed by cyanogen bromide cleavage at room temperature for 1 h. The sample was cleared by centrifugation and placed in a 1 kDa MWCO dialysis bag (Spectrum Laboratories, Inc.) for dialysis against 4 L of 10 mM potassium phosphate buffer, pH 7.0 at room temperature. The dialysis buffer was changed every 3-4 hours or until a precipitation formed. The sample was cleared by centrifugation, concentrated by using an Amicon Ultra-15 centrifugal filter with a 3 kDa MWCO and loaded onto an anion exchange column (Amersham Biosciences). The
protein was eluted with 30 mL of 10 mM potassium phosphate buffer, pH 7.0, using a 0-1 M NaCl gradient. The fractions containing eluted protein were concentrated by using an Amicon Ultra-4 centrifugal filter with a 3 kDa MWCO and the final protein was stored on ice. Protein was estimated to be >95% pure by Coomassie-stained SDS-PAGE.

**Fluorescence titrations**

Native tryptophan fluorescence experiments were conducted using a Horiba Jobin Yvon Fluorolog spectrofluorometer equipped with a Perkin Elmer 4 x 4 mm quartz cuvette. 100 nM aptamer solutions were individually titrated with Pup from 0.01-2 μM in 500 μL of TSP buffer. The excitation and emission wavelengths were 280 nm and 352 nm, respectively. Dissociation constants, $K_d$, were estimated from the changes in peak fluorescence intensities as a function of Pup concentration. Data were fit to the equation, 

$$\frac{(F-F_0)}{F_{\text{max}}} = \frac{[\text{Pup}]}{(K_d+[\text{Pup}])}$$

where $F$ is the fluorescence intensity at a given Pup concentration, $F_0$ is the fluorescence intensity of the blank, and $F_{\text{max}}$ is the maximum fluorescence intensity by using Prism 5 (GraphPad Software).

**Sequential over-expression of $^{15}$N-Pup and aptamers for STINT-NMR**

Plasmids pRSF-Pup and pBAD-PA were co-transformed into *E. coli* strain BL21(DE3) Codon+ (Novagen) for sequential over-expression. Two methods were used to evaluate the interaction between Pup and the peptide aptamers. The first method entailed over-expressing and labeling Pup followed by the over-expression of unlabeled PA. In the second, unlabeled PA was overexpressed followed by the over-expression and labeling of Pup. Each PA was handled in the same manner.
For $[U-^{15}\text{N}]$ labeling, cells were grown in M9 minimal medium containing 100 mg/L of ampicillin, 35 mg/L of kanamycin, and 1 g/L of $[U-^{15}\text{N}]$ ammonium chloride as the sole nitrogen source. Pup expression was induced with 0.5 mM IPTG at an $OD_{595}$ of 0.5 and grown for 2 h at 37 °C. A 100 mL sample of culture was collected, centrifuged, washed twice with 50 mL of 10 mM potassium phosphate buffer, pH 6.5, and washed twice 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 of $[U-^{15}\text{N}]$ Pup was collected to assess the extent of over-expression. The remainder of the culture was centrifuged, washed twice with M9 salts, re-suspended in fresh unlabeled minimal medium containing 100 mg/L of ampicillin and 35 mg/L of kanamycin. PA expression was induced with 0.25% L-arabinose (w/v) and grown overnight at 37°C. A 100 mL sample of culture was collected, centrifuged, washed twice with 50 mL of 10 mM potassium phosphate buffer, pH 6.5, and washed twice with 1 mL of 10 mM potassium phosphate buffer, pH 6.5, containing 10% glycerol, and stored at -80 °C for subsequent NMR analysis.

To over-express PAs, cells were grown in M9 minimal medium containing 100 mg/L of ampicillin and 35 mg/L of kanamycin. PA expression was induced with 0.25% L-arabinose (w/v) at an $OD_{595}$ of 0.5, and grown overnight 37 °C. The culture was harvested, centrifuged, washed twice with M9 salts, and re-suspended in fresh minimal medium containing 100 mg/L of ampicillin, 35 mg/L of kanamycin, and 1 g/L of $[U-^{15}\text{N}]$ ammonium chloride as the sole nitrogen source. Pup over-expression was induced with 0.5 mM IPTG, and grown for 2 h at 37 °C. A 100 mL sample of culture was collected,
centrifuged, washed twice with 50 mL of NMR buffer, and re-suspended in 500 μL of NMR buffer and 10% D₂O for NMR analysis.

**NMR Spectroscopy**

NMR experiments were performed on Bruker Avance III spectrometers equipped with a cryoprobe, operating at ¹H frequencies of 500 MHz and 700 MHz. All spectra were collected at 298 K, which yielded high quality spectra of Pup. We used a Watergate version of the ¹H{¹⁵N}-edited HSQC experiment recorded with 32 transients as 512 complex points, apodized with a squared cosine-bell window function and zero-filled to 1k points prior to Fourier transformation. The corresponding sweep widths were 12 and 35 ppm in the ¹H and ¹⁵N dimensions, respectively. The spectra were processed by using the program TOPSPIN (Bruker, Inc) and the program CARA³⁰,¹⁴⁸ was used for spectral analysis. We calculated the changes in the chemical shifts (ΔΩ) of amide nitrogens and covalently attached amide protons using the equation, \( \sqrt{(ΔΩ_{NH})^2 + (0.25*ΔΩ_{N})^2} \), where \( Ω_H \) and \( Ω_N \) represent hydrogen and nitrogen chemical shifts, respectively. We calculated the differential broadening was as: \((I/I_{high})_{control} - (I/I_{high})_{complex}\), where \( I \) and \( I_{high} \) is the intensity of each peak and intensity of highest peak in spectrum that did not decrease, shift or broaden during titration, respectively. \( I_{high} \) is the same peak in control and complex.

**M. bovis growth assay**

Plasmid pVV16-PA was transformed into *M. bovis* strain, Bacille Calmette-Guèrin (BCG). Approximately 1.5 μg of DNA was added to a 50 μL aliquot of cells in a 0.2 cm electroporation cuvette and DNA was introduced using a Gene Pulser (Bio-Rad) set at 2.5 kV, 25 μF, with the pulse controller resistance set at infinity. Following
electroporation, the sample was immediately diluted with 2 mL of mycomedia\textsuperscript{149} (liquid 7H10 medium (Difco) supplemented with 0.5% [v/v] glycerol, 10% [v/v] oleic acid-albumin-dextrose-catalase (Difco) and 0.05% [v/v] Tween-80) and incubated at 37 °C overnight. Cells were plated on Middlebrook 7H10 plates supplemented with 25 μg/mL of kanamycin and incubated for 3 weeks at 37 °C. Colonies were grown in 5 mL mycomedium for 5-7 days at 37 °C for PCR screening and seed stocks. Each PA was introduced in the same manner.

To monitor the growth of BCG during the expression of each PA, cultures at 2.5*10\textsuperscript{6} Colony Forming Units (CFU)/mL containing 25μg/mL of kanamycin under various conditions were grown for 4 days at 37 °C. Bortezomib (Btz) (Sigma-Aldrich) and 5-(5-methyl-2-(methylthio)thiophen-3-yl)-1,3,4-oxathiazol-2-one (GL5) were used as controls for positive inhibition (Fig. 3.2). Vector only BCG was used as a negative control. Btz (50 μM) and GL5 (50 μM) were assayed in the presence and absence of 2,2-(hydroxynitrosohydrazino)-bis-ethanamine (DETA-NO) (Fig. 3.2). BCG was assayed in the presence of 50 μM DETA-NO. Each PA was assayed in the presence of 50 μM DETA-NO. PA-1 was also assayed in the presence of 50 μM Btz and 50 μM DETA-NO. Triplicates of each culture were grown and serial dilutions ranging from 10\textsuperscript{-2} to 10\textsuperscript{-5} were plated in triplicate and incubated for 3 weeks at 37 °C for enumeration of surviving CFU.

BCG cultures were grown seven days to late log phase for Western blot analysis. Cells were collected, washed with phosphate buffered saline (PBS) buffer containing protease inhibitor, re-suspended in Tris/SDS buffer, and sonicated. Sonicated sample were frozen and thawed 10 times and the sonication was repeated. Lysates were heated for 5 mins at 37 °C after adding SDS and β-mercaptoethanol. Lysates were cleared by
centrifugation and stored at -80 °C for subsequent analysis. Lysates were immunoblotted and probed for thioredoxin with an anti-thioredoxin monoclonal antibody (Novagen). Horseradish peroxidase (HRP)-conjugated goat anti-mouse immunoglobulin G (IgG) (Amersham Pharmacia Biotechnology) was used to identify sites of binding by the primary antibody. Blots were incubated in SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific) for 15 mins and imaged on a ChemiDoc XRS+ System (Bio-Rad) by using Quantity One software.

![Image](image.png)

**Figure 3.2** Compounds used in BCG assay.

**Results**

*Selection of peptide aptamers that interact with Pup*

Yeast two-hybrid selection (Fig. 3.3) was used to identify PAs that bind specifically to Pup *in vivo*. Pup fused to the Gal4 DNA binding domain (BD) was used as bait, and randomized PA library, CLIPs, was the prey. *E. coli* thioredoxin protein fused to the GAL4 transcriptional activation domain, served as a scaffold to present 8-amino acid peptides of randomized sequence constrained in a thioredoxin aptamer loop\textsuperscript{134,135}. Approximately $1.8 \times 10^5$ transformants were screened and three clones, designated as PA-
1, PA-3, and PA-7, were found to activate all Y2H markers, *ADE2, HIS3, MEL1*, and *AUR1*-C, which is an indication of molecular interactions between PA and Pup.

DNA encoding the interacting PAs was PCR amplified from the yeast and sequenced (Fig. 3.4A). PA-1 contains nonpolar aliphatic and two aromatic residues, tryptophan and histidine. PA-3 and PA-7 are similar in sequence, and contain positively charged and aromatic residues. To estimate the binding affinity of each PA for Pup *in vitro*, fluorescence titrations were conducted using bacterially overexpressed and purified components. Fluorescence signals were quenched as Pup was added (Fig. 3.4B). The resolved K_d’s are in the nanomolar range.
**Figure 3.3 Overview of Yeast-Two-Hybrid Schemes.** (A) The target protein, Pup, fused to a DNA binding domain (BD) is expressed in yeast, which contains one or several reporter genes. The reporter genes are constructed to contain binding sites for the BD in their promoters. The reporter gene is not activated if the activation domain (AD) is not present. (B) The expression of an interacting protein, Mpa, fused to a transcriptional AD leads to the formation of a transcription factor, activating transcription of the reporter gene downstream to the DNA-BD. (C) Pup-BD is expressed concurrently with randomized peptides aptamers displayed in a constrained conformation in a scaffold protein, thioredoxin (Thio), linked to an AD. Yeast expressing target-binding peptides is selected due to the activation of reporter genes. In contrast, the growth of yeast clones expressing peptides that do not interact with the target protein will be suppressed.
Figure 3.4 Peptide Aptamer Sequences identified during Y2H Screen and Isotherms from Fluorescence Anisotropy. (A) Table of peptide aptamer sequences selected through Y2H screening by binding to Pup. (B) Fluorescence Anisotropy results by titrating Pup into each PA. The changes in the fluorescence intensity were plotted vs. the concentration of Pup. The binding buffer at pH 7.5 contained 100 mM Na$_2$S$_2$O$_3$ and 20 mM Na$_3$PO$_4$ and the experiments were carried out at 25°C. The data represents mean values and are fitted as described under Materials and Method. The fluorescence signal was quenched by the addition of Pup for each PA.
Peptide aptamers bind to the α-helix and C-terminal region of Pup

To identify the residues affected by PAs binding to Pup, [U-15N] Pup and each PA were sequentially over-expressed off compatible plasmids in E. coli for in-cell NMR (STINT-NMR) analysis. An in-cell [U-15N] Pup-PA complex was formed by first over-expressing Pup in uniformly labeling [U-15N] medium, replacing the medium with unlabeled medium, and over-expressing the PA. The order of expression was reversed to assess how the in-cell conformation of the complex is affected by the presence of each PA.

STINT-NMR reveals chemical shift changes and differential peak broadening of residues within the target protein that are in direct contact with the ligand. The 1H{15N}-HSQC spectrum of free Pup was overlaid with that of the [U-15N] Pup-PA complex to identify these changes. For clarity of presentation, chemical shift changes above 0.08 ppm and differential peak broadening greater than 60% were considered to be significant. Pup residues affected by PA-1 binding are G10, G11, G12, G13, R29, E30, L32, E34, D37, L40, V46, E48, and E49 (Fig. 3.5A and Fig. 3.6A); for PA-3 binding are R8, G10, G11, G12, G13, A23, E27, R29, E34, E35, D37, L40, D41, E42, D44, D45, V46, E48, E49, D53, F54 and Q60 (Fig. 3.5B and Fig. 3.7A); and for PA-7 binding are K7, G12, R28, T33, E34, E35, D44, V46, F54 (Fig. 3.9A & 3.10A). There were no significance differences in chemical shifts due to the order of expression of PA-1, PA-3 or PA-787.
Figure 3.5 PA-1 and PA-3 affects Pup in-cell. $^1$H($^{15}$N)-HSQC spectra of E. coli after 2 h of [$U$-$^{15}$N] Pup over-expression (red), overlaid with spectra (black) obtained from E. coli after 2 h of [$U$-$^{15}$N] Pup over-expression followed by 16 h over-expression of (A) PA-1 and (B) PA-3. Due to $^{15}$N editing of the experiment, only backbone amide protons and nitrogens of Pup are present in the spectrum. Most peaks do not change their positions reflecting that only a subset of Pup residues interact with each PA (bold). The sharp peaks in the spectra, which correspond to various metabolites of [$U$-$^{15}$N] ammonium ion, are indicated with asterisks.
**Figure 3.6** Chemical shifts differences and differential peak broadening from in-cell interactions between PA-1 and Pup. (A) Amide chemical shift differences for all assigned residues between free and PA-1 bound \([U-^{15}\text{N}]\) Pup. The chemical shift difference was calculated as: \(\sqrt{(\Delta\Omega_{\text{NH}})^2+(0.25*\Delta\Omega_{\text{N}})^2}\), where \(\Delta\Omega_{\text{NH}}\) and \(\Delta\Omega_{\text{N}}\) are the changes in the amide proton and nitrogen chemical shifts (ppm), respectively. (B) Differential peak broadening for all assigned residues in free and PA-1 bound \([U-^{15}\text{N}]\) Pup. The differential broadening was calculated as: \((I/I_{\text{high}})_{\text{control}} - (I/I_{\text{high}})_{\text{complex}}\), where \(I\) and \(I_{\text{high}}\) is the intensity of each peak and intensity of highest peak in spectrum that did not decrease, shift or broaden during titration, respectively. \(I_{\text{high}}\) is the same peak in control and complex.
Figure 3.7 Chemical shifts differences and differential peak broadening from in-cell interactions between PA-3 and Pup. (A) Amide chemical shift differences for all assigned residues between free and PA-3 bound [U-$^{15}$N] Pup. The chemical shift difference was calculated as: $\sqrt{(\Delta \Omega_{\text{NH}})^2 + (0.25 \times \Delta \Omega_{N})^2}$, where $\Delta \Omega_{\text{NH}}$ and $\Delta \Omega_{N}$ are the changes in the amide proton and nitrogen chemical shifts (ppm), respectively. (B) Differential peak broadening for all assigned residues in free and PA-3 bound [U-$^{15}$N] Pup. The differential broadening was calculated as: $(I/I_{\text{high}})_{\text{control}} - (I/I_{\text{high}})_{\text{complex}}$, where $I$ and $I_{\text{high}}$ is the intensity of each peak and intensity of highest peak in spectrum that did not decrease, shift or broaden during titration, respectively. $I_{\text{high}}$ is the same peak in control and complex.
Pup is an intrinsically disordered protein that forms a helical structure spanning residues 21-51 when bound to Mpa\textsuperscript{150}. We mapped the affected residues of Pup onto the crystal structure of the Pup-Mpa complex\textsuperscript{150} to show how each PA interacts with Pup. For PA-1 binding, the C-terminal half of the Pup helix is most strongly perturbed (Fig. 3.8A and C); this region corresponds to the Mpa binding surface of Pup. For PA-3 binding, the largest chemical shift changes also occur at the C-terminus of the Pup helix (Fig. 3.8B and D); however, the region affected is on the opposite side of Pup helix, which does not interact with Mpa. For PA-7 binding, fewer residues of the helical structure are affected than for PA-1 or PA-3 (Figure 3.9B and C) and these residues are largely localized to the non-Mpa binding side of the helix. This is expected due to the sequence similarity between PA-3 and PA-7.
Figure 3.8 PA-1 and PA-3 bind to opposite sides of the Pup helix. (A) The structure of Pup, based on the Pup-Mpa complex (PDB code 3M9D\textsuperscript{150}), shows that residues on the side of the α-helix that interacts with Mpa are most strongly affected by the binding of PA-1. (B) The same structure, horizontally rotated 180°, shows that residues on the opposite side of the helix are affected by PA-3 binding. Residues with chemical shift changes >0.08 ppm are in red and those with >60% differential peak broadening are in yellow. Pup residues affected by PA-1 (C) and PA-3 (D) binding are mapped onto the Pup-Mpa complex\textsuperscript{150}. Note: residues in (A), (B), (C), and (D) are hidden from view due to image orientation. The image of Pup-Mpa was constructed based on the Pup-Mpa structure by using Modeller\textsuperscript{151}.

Figure 3.9 Only a few Pup residues are affected by the interaction with PA-7. (A) $^1$H-$^1$N-HSQC spectra of *E. coli* after 2 h of [$U-{^{15}N}$] Pup over-expression (red), overlaid with spectra (black) obtained from *E. coli* after 2 h of [$U-{^{15}N}$] Pup over-expression followed by 16 h of over-expression of PA-7. Due to $^{15}$N editing of the experiment, only backbone amide protons and nitrogens of Pup are present in the spectrum. Most peaks do not change their positions reflecting the only a subset of Pup residues interact with PA-7 (bold). (B) Structure of Pup, based on the Pup-Mpa complex (PDB code 3M9D$^{150}$), shows that a few residues on the side of the α-helix that interacts with Mpa affected by the binding of PA-7. Residues with chemical shift changes $>0.08$ ppm are in red and those with $>60\%$ differential peak broadening are in yellow. (C) Residues affected by PA-7 are mapped onto the Pup-Mpa complex. Note: residues in (B) and (C) are hidden from view due to image orientation. The image of Mpa-Pup was constructed based on the Pup-Mpa structure by using Modeller$^{151}$.
Figure 3.10 Chemical shifts differences and differential peak broadening from in-cell interactions between PA-7 and Pup. (A) Amide chemical shift differences for all assigned residues between free and PA-7 bound [U-^{15}\text{N}] Pup. The chemical shift difference was calculated as: \( \sqrt{(\Delta \Omega_{NH})^2 + (0.25 \cdot \Delta \Omega_N)^2} \), where \( \Delta \Omega_{NH} \) and \( \Delta \Omega_N \) are the changes in the amide proton and nitrogen chemical shifts (ppm), respectively. (B) Differential peak broadening for all assigned residues in free and PA-7 bound [U-^{15}\text{N}] Pup. The differential broadening was calculated as: \( (I/I_{\text{high}})_{\text{control}} - (I/I_{\text{high}})_{\text{complex}} \), where \( I \) and \( I_{\text{high}} \) is the intensity of each peak and intensity of highest peak in spectrum that did not decrease, shift or broaden during titration, respectively. \( I_{\text{high}} \) is the same peak in control and complex.
To identify the residues of each PA that are involved in binding to Pup, we performed NMR titration experiments using bacterially expressed and purified \([U-^{15}\text{N}]\) PAs and bacterially expressed and purified Pup. 100 μM \([U-^{15}\text{N}]\) PA was titrated with 100 μM unlabeled Pup in 2 steps to yield \([U-^{15}\text{N}]\) PA to Pup molar ratios of 1:1 and 1:2, respectively. The peaks broadening observed in the \(^1\text{H}\{^{15}\text{N}\}\)-HSQC spectra (Fig. 3.11 and 3.12) revealed that only the residues located in the aptamer loop of the PA interact with Pup. The thioredoxin scaffold of the PA is largely unaffected by Pup binding, supporting the idea that thioredoxin provides a neutral platform for aptamer loop presentation\(^{136}\).
**Figure 3.11 Only residues in the PA loop bind Pup.** $^1$H-$^{15}$N-HSQC spectra of purified [U-$^{15}$N]-PA (red), overlaid with spectra (black) obtained after titrating with unlabeled purified Pup, (A) PA-1 and (B) PA-3. Each insert shows the residues from the PA loop. Most peaks do not change their position or intensities reflecting the fact that thioredoxin is neutral PA scaffold. Only a subset of PA residues from the Pup loop, exhibit substantial or complete broadening of peaks, indicating Pup-PA loop interactions. Due to $^{15}$N editing, only backbone amide protons and nitrogens of PA are present in the spectra.
**Figure 3.12** Only the PA loop is involved in PA-7-Pup binding. $^1$H($^{15}$N)-HSQC spectra of purified [U-$^{15}$N]-PA (red), overlaid with spectra obtained after titrating with purified Pup (black). Insert shows the residues from the PA-7 loop. Most peaks do not change their position or intensities reflecting the fact that thioredoxin is a neutral PA scaffold. Only a subset of PA-7 residues from the PA loop, exhibit substantial or complete broadening of peaks, indicating Pup-PA-7 interaction. Due to $^{15}$N editing, only backbone amide protons and nitrogens of PA-7 are present in the spectrum.
PA-I inhibits M. bovis BCG growth

BCG, the vaccine strain of M. bovis, was used in a cell-based assay to analyze the PA effect on cell survival\textsuperscript{152}. To examine whether the PAs inhibited cell growth, we cloned each PA into a constitutively active shuttle vector, pVV16 (BEI Resources), which confers kanamycin resistance. BCG transformed with either null-pVV16 or pVV16-PA (Fig. 3.1) were grown in BCG specific media\textsuperscript{149}. We mimicked the nitro-oxidative stress, which limits the replication of BCG and Mtb, by adding nitric oxide donor, DETA-NO to the cultures. We also grew BCG in the presence of Btz and GL5, known mycobacterial proteasome inhibitors\textsuperscript{143,153}, to compare the effects of growth inhibition to each PA.

BCG cultures were grown under specified conditions for 4 days, and aliquots were plated to determine the live bacteria count. Cultures were then grown for additional 3 days to accumulate detectable amounts of protein for Western blot analysis. Expression of each PA was verified by Western blotting with an anti-thioredoxin monoclonal antibody and native thioredoxin was not detected under the control conditions (Fig. 3.13).
Figure 3.13 Positive Expressions of PAs in BCG. Expression of PAs in BCG was verified by Western blot analysis. Lysates from *M. bovis* BCG cultures were probed for thioredoxin. Thioredoxin was not detected in the control cultures (lanes 1-6) and was detected at ~12 kDa in the cultures in which PAs were expressed (lanes 7-10).
We determined the reduction of BCG growth under specific conditions to assess the extent of inhibition \textit{in vivo}. The biological log reduction of the relative number of live microbes is expressed in terms of colonies forming units, CFU/mL. After incubating plates for 3 weeks, we observed a dramatic, 100-fold or a $2 \log_{10}$, reduction in CFU/mL for BCG expressing PA-1 grown under DETA-NO stress (Fig. 3.14A and B).

Nitro-oxidative stress alone, provided by DETA-NO, had no effect on the growth of BCG (Fig. 3.14B). Cell growth was inhibited by the presence of proteasome inhibitors GL5 (1.0 log) and Btz (1.0–1.5 log), and, importantly, BCG viability was further reduced by the combination of GL5 and DETA-NO (2.5–3.0 log) or Btz and DETA-NO (2.0–2.5 log). These results corroborate previous studies that proteasome inhibitors limit Mtb growth under nitro-oxidative stress.

Importantly, cell viability in the presence of PA-1 and DETA-NO was comparable to that of Btz and DETA-NO, suggesting that both Btz and PA-1 lead to enhancement of BCG sensitivity to nitro-oxidative stress. In contrast, PA-3 and PA-7, which interact with a segment of Pup that does not bind Mpa, in combination with DETA-NO, had only minimal effects on growth (0.5 log reduction) (Fig. 3.14B). Overexpression of PA-1 in the presence of Btz and DETA-NO led to no further growth inhibition beyond that of Btz and DETA-NO alone, which is consistent with PA-1 and Btz affecting the same pathway. These assays demonstrate that despite the uniformly high binding affinity of the selected PAs (Fig. 3.3B), the specificity of the interaction with Pup leads to distinct physiological outcomes.
**Figure 3.14 Visible Biological Log Reduction on BCG Assay.** (A) BCG plates after 3 weeks of growth showed 100-fold reduction in colonies due to expression of PA-1. (B) Inhibition of BCG growth under conditions the reduce proteasome function. 50 μM of Btz, GL5, and DETA-NO were used in designated conditions. The arrow indicates the strength of the initial inoculum. The dashed line is the lower limit of detection.
Discussion

Using CLIPs technology we isolated three peptide aptamers that bind to Pup with nanomolar affinity (Fig. 3.4B). The binding sites of the selected PAs are close to or even overlap each other. This fact highlights a critical advantage of Y2H selections$^{135}$. PAs selected to bind to a given target do so without competition between the molecules in the library. These PAs would not be selected by using *in vitro* selection schemes, such as phage, yeast, or ribosomal display, where all molecules in the library compete for the same target and, consequently, only the highest affinity aptamers remain after multiple cycles of selection$^{154}$.

Despite the fact that Pup is unstructured in the cytosol, the interaction surface between peptide aptamers and Pup, identified by STINT-NMR, suggests that, in the cytosol, Pup presents partially folded conformation.

It is important to emphasize that both Pup and PAs have to be over-expressed well above their physiological concentrations to be visible by in-cell NMR. Under these conditions, the cellular binding partners of Pup do not play a significant role in complex formation and can be largely ignored$^{155}$. We also expect that Pup folding in the crowded cytosol of *E. coli* will be similar to that which occurs in mycobacteria.

In spite of similar binding affinities, the peptide aptamers identified in this screen have vastly different functional effects on the survival of *M. bovis* BCG. Expression of PA-1 leads to a 100-fold inhibition, but PA-3 and PA-7 led to less than a 5-fold decrease in survival rate of DETA-NO treated cells. These effects are likely due to the different binding sites of PAs to their target, Pup. Indeed, Pup is known to form a helical structure when bound to its target, Mpa; only PA-1 interaction surface overlaps with the Mpa
helical binding surface on Pup, whereas PA-3 and PA-7 bind to the opposite side of the helix (Fig. 3.8 and 3.9)

There are two other enzymes that can bind to Pup, Pup ligase (PafA) and depupylase/deamidase, Dop. PA-1 can potentially affect binding to either of these complexes. Inhibition of Pup binding to PafA or Dop should lead to the inability to pupylate proteins, thus incapacitating the proteasomal function. Since both PA-3 and PA-7 do not appreciably affect the survival rate, we suggest that Dop and PafA interact with Pup via the same surface that Pup uses to bind to Mpa.

Both Btz and GL-5 proteasomal inhibitors were shown to substantially increase Mtb sensitivity to nitro-oxidative stress. The growth reduction caused by PA-1 with DETA-NO was comparable to that observed with Btz and DETA-NO and only 0.5 log lower than that of GL5 and DETA-NO treatment (Fig. 3.14B). This result suggests that PAs can be as potent as small molecules for inhibiting Mtb growth.

Drugs are widely used to inhibit enzymatic activities. It is very important to know whether these drugs have off-target effects. Here we showed that known inhibitors of the Mtb proteasome combined with PA-1 do not synergistically inhibit growth, suggesting that they act along the same biological pathway. At the same time, both GL5 and Btz are more effective than PA-1 alone in reducing survival rate. This suggests that besides the proteasome, GL5 and Btz inhibit other enzymes important for Mtb survival.

The technology demonstrated in this work permits both screening and structural characterization of protein interactions in the crowded environment of the cytosol. This is important for IDPs in which local structures may be partially stabilized due to
macromolecular crowding. Thus, selected PAs are already directed against physiologically relevant conformations, which may not be present during \textit{in vitro} selection and characterization of aptamers.

Functionally, IDPs and unstructured segments within proteins facilitate protein-protein interactions, serve as flexible linkers between folded domains or provide convenient sites for post-translational modifications. Site-specific binding of PAs to IDPs or unstructured segments allows us to uniquely affect these functions, either by blocking access to the protein interaction sites and the sites of post-translational modification, or by decreasing the flexibility of critical sites in the linkers. The availability of PAs selected to bind to specific sites on IDPs allows unprecedented opportunity to analyze the functional consequences of IDP interactions with atomic precision in intact cells.

\textbf{Acknowledgements}

This work was supported by NIH grant GM085006 to AS.
CHAPTER FOUR

IN-CELL NMR SCREENING (SMILI-NMR) OF NATIONAL CANCER INSTITUTE DRUG LIBRARY LEADS TO MTB INHIBITION

Jacqueline Washington¹, Eric A. Smith², Sergey Reverdatto¹, David S. Burz¹, Kathleen A. McDonough², Alexander Shekhtman¹*

¹ Department of Chemistry, University at Albany, State University of New York, Albany, New York, United States of America

² Wadsworth Center, New York State Department of Health, Albany, New York, United States of America
Introduction

NMR-based *in vitro* screening, utilized in drug discovery programs\(^{158-161}\), has become a powerful method for identifying and analyzing small drug-like compounds that specifically bind to attractive drug targets. One proven approach to the drug discovery process entails HTS of compound libraries. However, finding inhibitors of complex formation proved to be difficult for HTS in *in vivo* assays due to the difficulty in distinguishing small molecule interactions and effectiveness\(^{162,163}\).

We used a NMR spectroscopy-based screening procedure, screening of Small Molecule Interactor LIbrary by using in-cell NMR (SMILI-NMR), to screen a large drug library for compounds capable of weakening protein-protein interactions within a biomolecular complex\(^{63}\). SMILI-NMR is uniquely positioned between high throughput screening techniques capable of screening hundreds to thousands of protein complexes a day and atomic resolution structural methods for studying protein-ligand complexes that require significantly longer periods of time.

SMILI-NMR utilizes STructural INTeractions by using in-cell NMR (STINT-NMR) technology\(^{86,87,164}\), 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, we can directly observe the formation of high-affinity complexes or changes in structure induced by the binding of a small drug-like molecule that disrupts or enhances the stability of the complex. By combining the advantages of SMILI-NMR and STINT-NMR, we can analyze physiologically relevant conformations of target proteins or complexes, with minimal sample preparation, at a low cost and relatively fast.
In this way, STINT-NMR serves as a direct assay for protein-protein interactions, facilitating high throughput screening. SMILI-NMR technology offers unique advantages for the screening of therapeutic agents against chosen disease-related protein targets. We can selectively screen for drug-like molecules that block the interaction surface that is critical for the normal functioning of the biocomplex. This greatly reduces the likelihood of compensatory mutations or suppressor gene products arising that may restore the function of the drug-compromised biocomplex.

We used a well-studied system of two interacting proteins, prokaryotic ubiquitin-like protein (Pup) and the proteasomal ATPase (Mpa), because it is an attractive system for drug therapy against latent Mtb infection\textsuperscript{165}. Since the proliferation of Mtb relies on its ability to resist nitro-oxidative stress due to the immunity of the host via nitric oxide synthase\textsuperscript{142}, it was suggested that Mtb contains defenses to resist eradication by nitric oxide\textsuperscript{142,166,167}. The Pup and Mpa interaction plays a key role in the prokaryotic protein degradation\textsuperscript{140,168}. Pup, a 64 amino acid IDP\textsuperscript{141}, modifies and targets mycobacterium proteins to the proteasome for degradation after recognition by the amino-terminal coiled-coil domains of Mpa\textsuperscript{169}. Mpa-proteasome complex unfolds and degrades substrates tagged with Pup in response to oxidative stress\textsuperscript{170,171}.

The goal of this study is to identify drug molecules that alter the binding of Pup and Mpa through the screening of a library of molecules containing diverse structures. Thus, we chose the National Institutes of Health/National Cancer Institute (NIH/NCI) library, which contains compounds with significant variations in chemical structure, to screen against an IDP and its known binding partner. A total of 1597 compounds are contained in this library, which was obtained from the Molecular Libraries Small
Molecule Respository (MLSMR). This library contains a diverse set of compounds that are amenable to forming structure-based hypotheses. Thus, these molecules are relatively rigid, with 5 or fewer rotatable bonds, having a tendency to be planar, 1 or less chiral centers. Moreover, they have pharmacologically desirable features, *i.e.* did not contain: obvious leaving groups, weakly bonded heteroatoms, organometallics, polycyclic aromatic hydrocarbons, etc.

**Materials and Methods**

*Mtb strain*

Attenuated Mtb strain mc^{2}7000 (*Mtb* H37 Rv ΔRD1::GFP and ΔpanCD) was obtained from Kathleen McDonough from the Division of Infectious Diseases at Wadsworth Center. mc^{2}7000 is the direct unmarked derivative of mc^{2}6030 with a removed hygromycin-cassette.

*Plasmids Utilized*

All plasmids used were previously constructed in our lab.

Pup-pASK3+-His_{7x} expresses Pup-His_{7x} 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.

Mpa-S-pRSF expresses Mpa extracted from *Mycobacterium smegmatis* (*Msm*). The plasmid expresses Mpa from a T7 promoter/lac operator (PT7/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.
PA-pVV16 expresses a PA constrained within a thioredoxin loop. The plasmid is constitutively active and confers kanamycin resistance for selection in *E. coli* and Mtb.  

*Sequential over-expression of $^{15}$N-Pup and Mpa*

Plasmids Pup-pASK3+ and Mpa-pRSF were co-transformed into *E. coli* strain BL21(DE3) Codon+ (Novagen) for sequential over-expression. Our method entailed over-expressing and labeling Pup followed by the over-expression of unlabeled Mpa. For $[U-^{15}\text{N}]$ labeling, cells were grown in M9 minimal media containing 100 mg/L of ampicillin, 35 mg/L of kanamycin, and 1 g/L of $[U-^{15}\text{N}]$ ammonium chloride as the sole nitrogen source. Pup expression was induced with 0.2 mg/mL of anhydrotetracycline in dimethylformamide (DMF) at an OD$_{595}$ of 0.7 and grown for 1 h at 37 °C. A 100 mL sample of culture was collected, centrifuged, washed twice with 50 mL of NMR buffer (10 mM potassium phosphate buffer, pH 6.5), washed twice with 1 mL of NMR buffer containing 10% glycerol, and stored at -80 °C for subsequent NMR analysis. A control sample of $[U-^{15}\text{N}]$ Pup was collected to assess the extent of over-expression. The remainder of the culture was centrifuged, washed twice with M9 salts, re-suspended in fresh unlabeled minimal medium containing 100 mg/L of ampicillin and 35 mg/L of kanamycin to a final OD$_{595}$ of 0.7. Mpa expression was induced with 0.5 M IPTG and grown for 8 h at 37°C. The cells was harvested and washed twice with 50 mL of NMR buffer containing 10% glycerol. The cell pellet was re-suspended to a final OD$_{595}$ of 95 with NMR buffer containing 10% glycerol, aliquoted into 250 uL aliquots, centrifuged, and stored at -80 °C for subsequent NMR analysis.
For NMR sample preparation, the cells harboring \([U^{-15}N]\) Pup and unlabeled Mpa were thawed on ice for 10 mins and washed twice with 1.5 mL of NMR buffer, and re-suspended in 500 μL of NMR buffer with 10% D₂O. NMR experiments were performed on Bruker Avance III spectrometer equipped with a cryoprobe, operating at \(^1H\) frequency of 500 MHz. All spectra were collected at 298 K, which yielded high quality spectra of Pup. We used a Watergate version of the \(^1H\{^{15}N\}\)-edited HSQC experiment recorded with 32 transients as 512 complex points, apodized with a squared cosine-bell window function and zero-filled to 1k points prior to Fourier transformation. The corresponding sweep widths were 12 and 35 ppm in the \(^1H\) and \(^{15}N\) dimensions, respectively. The spectra were processed by using the program TOPSPIN (Bruker, Inc) and the program CARA\(^{30,148}\) was used for spectral analysis.

We prepared sample matrices for the primary screening of the Diversity Set III NCI/NIH library of over 1500 compounds. The compounds, solubilized in neat DMSO at a 10 mM concentration, are stored at -80 °C. To preserve the integrity of the compounds, each original 96-well plate was thawed only once to make a matrix plate and a duplicate plate. For each matrix plate, we prepared 18 mixtures of compounds by adding compounds located in a row or column. For example, one plate contains 70 compounds: 8 compounds in a column and 10 compounds in a row. The compounds located in row 4 were mixed and tested.

Before starting the library screening, we analyzed the effects of DMSO on the \(^1H\{^{15}N\}\)-HSQC spectrum of \([U^{-15}N]\) Pup and unlabeled Mpa. We varied the final concentrations of DMSO from 2.5 – 12.5% and added the appropriate amount of NMR
buffer and 10% D2O to result in 500 μL. After determining the 7.5% DMSO concentration was needed for the control spectrum, the library screening was begun. Library mixtures were added individually at 50 μM final concentrations to E. coli cells over-expressing [U-15N] Pup and unlabeled Mpa.

After the primary screen, the samples exhibiting similar spectral changes, located at the intersection of rows and columns, will be used in the second round of screening; thus, the compound at the intersection will be titrated individually at a final concentration of 50 μM into E. coli cells over-expressing [U-15N] Pup and unlabeled Mpa.

*Mtb growth assay*

Plasmid pVV16-PA and pVV16 were transformed into Mtb strain mc27000. Approximately 1.5 μg of DNA was added to a 50 μL aliquot of cells in a 0.2 cm electroporation cuvette and DNA was introduced using a Gene Pulser (Bio-Rad) set at 2.5 kV, 25 μF, with the pulse controller resistance set at infinity. Following electroporation, the sample was immediately diluted with 2 mL of mycomedium (Middlebrook 7H10 medium supplemented with 25 μg/mL of kanamycin, 0.5% glycerol, 10% oleic acid-dextrose-catalase, 0.05% Tween 80, 0.5% tyloxapol, and 0.2% casaminoacids, and 0.024 mg/mL pantothenate) and incubated at 37 °C overnight. Cells were plated on mycomedia (minus tyloxapol) agar and incubated for 3 weeks at 37 °C. Colonies were grown in 5 mL mycomedia cultures for 5-7 days at 37 °C for PCR screening and seed stocks. Each PA was introduced in the same manner.

To monitor the growth of mc27000 in the presence of each NIH/NCI compound and PA, cultures at 2.5*10^6 CFUs/mL will be grown for 4 days at 37 °C under various conditions. Btz (Sigma-Aldrich) and GL5 will be used as controls for positive inhibition
and kanamycin resistance mc²7000 will be used as a negative control. Btz (50 μM) and GL5 (50 μM) will be assayed in the presence and absence of DETA-NO. mc²7000 will be assayed in the presence of 50 μM DETA-NO. Each NIH compound at a final concentration of 50 μM and each expressed PA will be assayed in the absence and presence of 50 μM DETA-NO. PA-1 will be assayed in the presence of 50 μM Btz and 50 μM DETA-NO. Triplicates of each culture will be grown and serial dilutions ranging from 10⁻² to 10⁻⁵ will be plated in triplicate and incubated for 3 weeks at 37 °C for enumeration of surviving CFU.

Mtb cultures will be grown for seven days to late log phase for Western blot analysis. Cells will be collected, washed with PBS buffer containing protease inhibitor, re-suspended in Tris/SDS buffer, and sonicated. Sonicated sample will be frozen and thawed 10 times and the sonication was repeated. Lysates will be heated for 5 mins at 37 °C after adding SDS and β-mercaptoethanol. Lysates will be cleared by centrifugation and stored at -80 °C for subsequent analysis. Lysates will be immunoblotted and probed for thioredoxin with an anti-thioredoxin monoclonal antibody (Novagen). HRP-conjugated goat anti-mouse IgG (Amersham Pharmacia Biotechnology) will be used to identify sites of binding by the primary antibody. Blots will be incubated in SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific) for 15 mins and imaged on a ChemiDoc XRS+ System (Bio-Rad) by using Quantity One software.
Results

The formation of Pup-Mpa complex

To demonstrate the efficacy of SMILI-NMR, we setup the in-cell model system for STINT-NMR by sequentially over-expressing $[^{15}\text{N}]\text{Pup}$ and unlabeled Mpa in *E. coli* using two compatible plasmids, Mpa-pRSF and Pup-pASK3+, with tightly controlled inducible promoters to express the proteins individually. STINT-NMR reveals chemical shift changes of residues within the target protein that are in direct contact with the ligand. In-cell $[^{15}\text{N}]\text{Pup}$-Mpa complexes were formed by over-expressing Pup in uniformly labeled $[^{15}\text{N}]$ medium, replacing the medium with unlabeled medium, and over-expressing Mpa. We performed in-cell $^1\text{H}\{^{15}\text{N}\}^{150}$-edited HSQC experiments on cells over-expressing $[^{15}\text{N}]\text{Pup}$ and over-expressing unlabeled Mpa. The spectrum of free $[^{15}\text{N}]\text{Pup}$ showed well-resolved peaks, implying that the Pup is the only species in the cells. The spectrum of $[^{15}\text{N}]\text{Pup}$ and unlabeled Mpa showed chemical shift changes and peak broadening, indicating the formation of Pup-Mpa complexes. The two spectra were overlaid to determine changes in chemical shifts and peak intensities (Fig. 4.1).

We expected the overlay to reveal the interaction of $[^{15}\text{N}]\text{Pup}$ and Mpa. The overlay shows extreme peak broadening coupled with slight chemical shifting, which both are due to protein-protein interaction and molecular crowding. Since binding affinity of Pup for Mpa was estimated to be submicromolar$^{172}$, there is selective peak broadening in the helical region of Pup residues: R8, A23, L39, D41, E42, I43, D45, V46, D53, F54, V55, V59, Q60, and K61 (Fig. 4.1).
Figure 4.1 $^1$$H$($^{15}$N)-HSQC spectral differences between of free Pup and Pup-Mpa complex. $^1$$H$($^{15}$N)-HSQC spectra of *E. coli* obtained after 1 h of $[U-^{15}$N] Pup over-expression (red), overlaid with $^1$$H$($^{15}$N)-HSQC spectra of *E. coli* obtained after 1 h of $[U-^{15}$N] Pup followed by sequential 8 h over-expression of Mpa (black). Due to $^{15}$N editing of the experiment, only backbone amide protons and nitrogens of Pup are present in the spectrum. The labeled peaks showed peak broadening and slight chemical shifting. The sharp peaks in the spectra, which correspond to various metabolites of $[U-^{15}$N] ammonium ion, are indicated with asterisks.
We chose to screen the NIH/NCI Diversity Set III library (http://dtp.nci.nih.gov) composed of over 1500 compounds against the Pup-Mpa system to provide a collection of compounds that are capable of interacting at a detectable level. The NIH/NCI diversity set contains 1597 small molecules for their high chemical and pharmacological diversity from the entire collection of 140,000 compounds at NCI. Each compound is dissolved in DMSO at a 10mM concentration. This library was considered useful as a potential source of novel drug candidates because: first, these small biologically relevant molecules can pass through bacterial cell membranes and interact with the target protein directly in vivo \textsuperscript{122,123} and second, a library of this size provides a data set that spans a broad range of physicochemical properties.

We used a standard screening procedure, which we call the matrix method\textsuperscript{63}, in which library compounds located in one row or one column of a matrix plate were mixed and tested (Fig. 4.2). Individual mixtures were examined for their ability to change the in-cell NMR spectrum of the $[U^{15}\text{N}]$ Pup-Mpa complex. Samples exhibiting similar spectral changes, located at the intersection of rows and columns, will be used in the second round of screening to validate the initial findings. In this way, we can screen a library of over 1500 compounds by examining 342 (18 samples per plate x 19 plates) mixtures.
**Figure 4.2 Matrix Method of screening chemical libraries.** A library containing over 1500 compounds is screened by examining individual mixtures located in the first row (black arrows) and first column (red arrows) of a matrix plate. Mixtures that result in similar changes in the in-cell NMR spectrum, located at the intersection of rows and columns (shaded boxes), will be used in the second round of screening. For example: C3 and F9 of this plate would be further analyzed, individually.
To establish the effects DMSO would have on each $^1$H{$^{15}$N}-HSQC spectra, we collected $^1$H{$^{15}$N}-HSQC spectra on E. coli cells with [U$^{-15}$N] Pup-Mpa complexes under the following DMSO concentrations: 2.5%, 5%, 7.5%, 10%, and 12.5%. We determined that 7.5% DMSO was the appropriate amount needed as a control.

108 mixtures were screened, in which each compound is at a final 50 μM concentration, in cells with [U$^{-15}$N] Pup-Mpa complexes. We compared the control spectra to the spectra recorded on [U$^{-15}$N] Pup-Mpa complexes in the presence of each matrix mixture. We expected to see changes in the in-cell NMR spectrum of the protein complex when at least one of the compounds in the mixture binds to the interaction surfaces involved in complex formation. Majority of the mixtures screened showed no interactions with the complex, i.e. the $^1$H{$^{15}$N}-HSQC spectra did not change upon adding these mixtures. However, 24 mixtures showed peak shifting and broadening. For illustration, the control spectrum was overlaid with the spectrum of [U$^{-15}$N] Pup-Mpa complex titrated with mixtures #1 and #2 (Fig. 4.3). For mixture #1, residues Q5, R8, G9, S21, G25, E27, T33, T36, D44, L47, D53, V55, Y58, V59, and K61 experienced extreme broadening and residues Q5, D44, V55, Y58, and V59, also, underwent a slight chemical shift (Fig. 4.3A). For mixture #2, residues Q5, G9, T36, L39, I43, L47, N50, A51, E52, F54, V55, R56, A57, V59, K61, and G62 underwent peak broadening and residues G9, I43, L47, A51, E52, F54, V55, A57, and G62, also, shifted slightly (Fig. 4.3B). Majority of those residues affected are known to be in contact with Mpa.
Figure 4.3 In-cell NMR spectra of [U-^{15}N] Pup-Mpa complex changed in presence of NIH mixtures. $^1$H{$^{15}$N}-HSQC spectra of E. coli with [U-^{15}N] Pup-Mpa complexes with 7.5% DMSO in the buffer (red) and a $^1$H{$^{15}$N}-HSQC spectra of E. coli with [U-^{15}N] Pup-Mpa complexes in the presence of mixture #1 (A) and mixture #2 (B). Due to $^{15}$N editing of the experiment, only backbone amide protons and nitrogens of Pup are present in the spectrum. The affected residues are labeled.
After cross-referencing the appropriate rows and columns of the 24 mixtures, 12 compounds were found in the intersections for the second round of screening (Table 4.1). All the compounds found have at least one aromatic ring; however, most are multi-ringed structures. Most of the compounds contain nitrogen species, *i.e.* amines, pyrroles, pyridyls, and guanidines, and two of the compounds are alcohols (Table 4.1). Each compound will be titrated individually into cells at a final 50 μM concentration. We expect that the compounds will show similar broadening and peak shifting as observed in the first round of screening.
<table>
<thead>
<tr>
<th>Compound Name</th>
<th>NSC Number</th>
<th>Chemical Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-(3-Pyridyl)benzothiazole</td>
<td>37812</td>
<td><img src="image1.png" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>4-(2-Benzothiazoly)-1,3-benzenediol</td>
<td>33005</td>
<td><img src="image2.png" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>purpurogallin</td>
<td>35676</td>
<td><img src="image3.png" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>3,3’-Pyridil</td>
<td>91516</td>
<td><img src="image4.png" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>thermospine</td>
<td>21725</td>
<td><img src="image5.png" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>2,2'-benzene-1,4-diylidiguanidine</td>
<td>20586</td>
<td><img src="image6.png" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>7H-Pyrrole[3,2-(\phi)]quinazoline-1,3-diamine</td>
<td>339578</td>
<td><img src="image7.png" alt="Chemical Structure" /></td>
</tr>
</tbody>
</table>
Table 4.1 NIH compounds found during first round of screening NIH library against in-cell \([U^{15}\text{N}]\) Pup-Mpa complex. During the first round of screening, each compound was analyzed at 50 μM final concentrations. Majority of the compounds are multi-ringed structures that contain nitrogen species.

<table>
<thead>
<tr>
<th>Compound Name</th>
<th>NSC Number</th>
<th>Chemical Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-(ethoxymethyl)-2-methylsulfonyl-1H-pyrimidin-4-one</td>
<td>77422</td>
<td><img src="image1.png" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>3,5,7-Trimethyl(1,2,4)triazolo(4,3-a)pyridine</td>
<td>79010</td>
<td><img src="image2.png" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>1-hydroxyphenazine</td>
<td>88882</td>
<td><img src="image3.png" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>1-oxa-4-azaspiro[4.4]nonane-3,3-dimethanol</td>
<td>106045</td>
<td><img src="image4.png" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>3-(3-Methoxy-isoxazol-5-y)-propionic acid</td>
<td>303800</td>
<td><img src="image5.png" alt="Chemical Structure" /></td>
</tr>
</tbody>
</table>
**Mtb growth inhibition assay**

Due to the increasing emergence of drug resistance in Mtb, we will perform our growth assay in non-virulent Mtb (mc²7000) to analyze the effect each NCI compound has on cell survival. We will, also, compare the growth effects of NCI compounds, PA-1, and PA-3. Thus, PA-pVV16 and null-pVV16 were transformed into mc²7000. We will grow mc²7000 in the presence of Btz and GL5 known mycobacterial proteasome inhibitors, to compare the effects of growth inhibition to PAs and NCI compounds. We will mimic the nitro-oxidative stress that limits the replication of Mtb by adding nitric oxide donor, DETA-NO to the cultures. 5 mL cultures will be grown under specified conditions for 4 days, and aliquots will be plated for CFU plating. Cultures will then be grown for an additional 3 days to accumulate sufficient limits of protein for Western blot analysis. Total cell lysates will be prepared by the sonication/freeze-thaw method and probed for thioredoxin with an anti-thioredoxin monoclonal antibody. HRP-conjugated goat anti-mouse IgG will be used to identify sites of binding of the primary antibody. Expression of each PA will be verified by Western blotting.

We will determine the reduction of Mtb growth under specified conditions to assess the extent of inhibition *in vivo*. Biological log reduction shows the relative number of live microbes eliminated, which is expressed in terms of colonies forming units, CFU/mL.

**Discussion**

Our use of SMILI-NMR screening presents a unique approach to monitor changes at an amino acid level in the interaction surface caused by a small drug interfering with complex formation. It is a comparatively simple and fast method that offers validity that
the small molecule is capable of penetrating the cell membrane and specifically engaging the target molecule. Using SMILI-NMR in combination with STINT-NMR, we successfully identify a dozen compounds as suitable starting points for subsequent optimization into credible drug candidates. Those compounds will be used in a Mtb growth assay to evaluate the functionality of each drug on cell survival.

Due to the fact that several of the drugs discovered here have functional groups similar to promising TB drugs, there is potential for future use in animal models. For example, NSC 339578 is comparable to diarylquinoline, which is known to drug-sensitive and drug-resistant Mtb in vitro\textsuperscript{173}. NSC 88882, 37812, 33005, and 79010 are azoles that are analogous to phenothiazines, which deplete membrane energy in Mtb\textsuperscript{174}.

The use of a Mtb-based functional assay to identify drug candidates is necessary and efficient when dealing with infectious diseases. Drug screens that involve the importance of physiological conditions, i.e, acidic pH and and nitric oxide stress, is significant for identifying drugs that shorten tuberculosis treatment. Studies, such as this, could identify drugs that have synergistic effects. It would be of interest to identify molecules that synergize with PA-1, shown to reduce BCG growth under nitric oxide stress.
References


