A survey of the current drug screening techniques to obtain rational design and study drug-target interactions

Stephen Dansereau
University at Albany, State University of New York, sdansereau@albany.edu

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

Part of the Biochemistry Commons

Recommended Citation

This Master's Thesis 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.
A Survey of the Current Drug Screening Techniques to Obtain Rational Design and Study Drug -Target Interactions

by

Stephen Dansereau

A Thesis
Submitted to the University at Albany, State University of New York
In Partial Fulfillment of
the Requirements for the Degree of

Master of Science

College of Arts and Sciences
Department of Chemistry

2016
# TABLE OF CONTENTS

Abstract

Overall Introduction .

**Chapter 1. Functional Assays that Use Fluorescence Readout**

1.1) Introduction

1.2) Green fluorescent protein (GFP) as a cellular marker and drug-delivery tool

1.3) Bioluminescent properties of luciferase for drug screening and analyzing drug efficacy

1.4) Bodipy as a diverse tool for probing proteins and assisting in drug delivery

1.5) Conclusions

**Chapter 2. Using X-Ray Crystallography to Visualize Drug-Target Interactions**

2.1) Introduction

2.2) Drugs crystallized to G protein-coupled receptor kinases (GRKs)

2.3) Drugs crystallized to phosphatidylinositol phosphates (PI4P)

2.4) Drugs crystallized to the GTPase-associated region (GAR) of bacterial ribosomes

2.5) Employing microcrystallography (MXC) to study membrane-proteins

2.6) Drugs crystallized to OmpF via x-ray microbeams

2.7) Conclusions

**Chapter 3. In-cell NMR based drug-screening**

3.1) Introduction

3.1.1) Nuclear magnetic resonance (NMR) spectroscopy

3.1.2) NMR chemical shifts and resonance assignments

3.1.3) Protein-structure determination from NMR spectroscopy
3.1.4) In-cell NMR spectroscopy ..............................................................97
3.1.5) Studying large proteins with isotopic labeling ......................................................103
3.1.6) Biopharmaceutical analysis via in-cell NMR spectroscopy ................................104
3.1.7) Using in-cell NMR spectroscopy to study protein-ligand interactions ............106
3.1.8) STD-NMR spectroscopy ..............................................................................108
3.1.9) STINT-NMR spectroscopy ..........................................................................113
3.2) Prokaryotic applications ...............................................................................119
3.2.1) NMR of proteins ........................................................................................119
3.2.1.1) In-vitro NMR solution structures of prokaryotic proteins ............................120
3.2.1.2) In-cell NMR studies of prokaryotic proteins ..............................................128
3.2.2) Protein-protein interactions studied by NMR spectroscopy ..........................129
3.2.2.1) In-vitro NMR solution structures of protein-protein interactions .............129
3.2.2.2) In-cell NMR studies of protein-protein interactions ..................................131
3.2.3) Protein-drug interactions studied by NMR spectroscopy ..............................137
3.3) Eukaryotic applications ..................................................................................141
3.3.1) In-yeast NMR ............................................................................................141
3.3.2) In-insect NMR .............................................................................................149
3.3.3) In-cell NMR of Xenopu ..............................................................................150
3.3.4) In-cell NMR of human cells .......................................................................152
3.4) Conclusions ....................................................................................................161
Overall Conclusions ..............................................................................................161
REFERENCES .................................................................................................162
# TABLE OF FIGURES

## Chapter 1
- Figure 1.1 ........................................................................................................... 5
- Figure 1.2 ........................................................................................................... 9
- Figure 1.3 ......................................................................................................... 11
- Figure 1.4 ......................................................................................................... 12
- Figure 1.5 ......................................................................................................... 15
- Figure 1.6 ......................................................................................................... 19
- Figure 1.7 ......................................................................................................... 23
- Figure 1.8 ......................................................................................................... 25
- Figure 1.9 ......................................................................................................... 33
- Figure 1.10 .................................................................................................... 36
- Figure 1.11 .................................................................................................... 39
- Figure 1.12 .................................................................................................... 44

## Chapter 2
- Figure 2.1 ......................................................................................................... 50
- Figure 2.2 ......................................................................................................... 53
- Figure 2.3 ......................................................................................................... 54
- Figure 2.4 ......................................................................................................... 58
- Figure 2.5 ......................................................................................................... 63
- Figure 2.6 ......................................................................................................... 66
- Figure 2.7 ......................................................................................................... 69
- Figure 2.8 ......................................................................................................... 72
- Figure 2.9 ......................................................................................................... 76
- Figure 2.10 ...................................................................................................... 78
- Figure 2.11 ...................................................................................................... 80
- Figure 2.12 ...................................................................................................... 86
- Figure 2.13 ...................................................................................................... 92

## Chapter 3
- Figure 3.1 ......................................................................................................... 94
- Figure 3.2 ......................................................................................................... 101
- Figure 3.3 ......................................................................................................... 110
- Figure 3.4 ......................................................................................................... 112
- Figure 3.5 ......................................................................................................... 115
- Figure 3.6 ......................................................................................................... 118
- Figure 3.7 ......................................................................................................... 121
- Figure 3.8 ......................................................................................................... 125
- Figure 3.9 ......................................................................................................... 127
- Figure 3.10 ..................................................................................................... 130
- Figure 3.11 ..................................................................................................... 132
- Figure 3.12 ..................................................................................................... 134
Abstract

Different techniques have been developed over the years for the purpose of studying proteins and understanding their functions. Early techniques typically employed bioluminescence or fluorescence such as the firefly protein luciferase and the jellyfish green fluorescent protein (GFP), respectively, to localize proteins within the cell. X-ray crystallography has also provided valuable structural details of many different proteins in vitro. Yet, nuclear magnetic resonance (NMR) spectroscopy offers the most realistic insight into proteins' physiologic structures and how proteins function in their native, cellular environments.

In-cell NMR spectroscopy is a technique that can exploit the nuclei of proteins in order to discern their structures and binding interactions in atomic resolution. This requires use of a chemical shift library or reference spectrum respectively, to which the chemical shifts of the protein under study can be compared. The results offer real time analyses of proteins in their native, crowded cellular environments, making for more accurate data. The implications of unraveling a protein's structure are many, including understanding protein misfolding associated with pathologies, identifying a suitable therapeutic target, determining a biomolecules active site, studying protein-substrate interactions, all of which contribute to the rational design of drugs. These valuable insights are especially important in the treatment of both pathogenic and genetic diseases.
**Overall Introduction:**

Recent advances in molecular biology research techniques and automation have allowed for greater intricacy in studying chemical biology and drug discovery by means of high-throughput screening (HTS) (Thorne, 2010). Through the transformation of chemical and biological information into a detectable signal, researchers are able to observe intracellular processes and structures (Yong, 2014). These visualization techniques have implications in a range of fields including biology, biochemistry, clinical diagnosis, and drug design and discovery (Yong, 2014). In fact, HTS is a widely used technique in both academia and industry that vastly increases the drug discovery process (Pawel, 2012).

Numerous different libraries including combinatorial chemistry, genomics, protein, and peptide libraries are screened via HTS assays at a rate of a few thousand compounds per day, greatly cutting the costs of drug development (Pawel, 2012): Up to 10,000 compounds per day can be screened using typical HTS while this number increases to 100,000 compounds per day using ultra-high throughput screening (UHTS) (Pawel, 2012).

The multi-step procedure itself requires target identification, reagent preparation, compound management, assay development, and a high-throughput screening library whereby an immense repertoire of chemical compounds are tested against different biological targets in welled plates (Pawel, 2012). The target molecule's activity is measured in a functional assay while nonfunctional assays are limited to observing drug-target binding observations (Pawel, 2012). Both are detected using fluorescence, scintillation proximity assays (SPA), and luminescence (Pawel, 2012).

The screening process consists of a generic primary screening and a more specific secondary
screen if a “HIT”, or positive result, is reported (Pawel, 2012). Each biological or biochemically oriented secondary screening assay can be either homogenous or the more specific and complex heterogenous (Pawel, 2012). Thus, cell-based assays can provide insight into the target's cellular activity (Pawel, 2012). Examples of molecules targeted by cellular microassays include small molecules, polymers, and antibodies (Pawel, 2012).

Once a target is determined to be relevant to a disease, modulators (ie. agonists/antagonists, channel blockers/openers) of this target are identified (Pawel, 2012). Next, an appropriate assay that will allow observation of the target is chosen (Pawel, 2012). HTS is particularly useful at this step because, as mentioned, it allows for the target to be very efficiently tested against a massive number of compounds (Pawel, 2012).

Compounds shown to be selective towards the target will then undergo further study (Pawel, 2012). Once this “lead” compound is optimized, it is ready for entry into the final stage of drug discovery – human trials. Thus, by screening thousands of drug candidates, HTS greatly speeds up the drug discovery process (Pawel, 2012).

Despite this improved surveillance, reporters have been known to bias results in unintended ways (Thorne, 2010). An understanding of these drawbacks as well as carrying out appropriate follow-up experiments is necessary to give credibility to the HTS results (Thorne, 2010).
CHAPTER 1

Functional assays that use fluorescence readout

1.1) Introduction.

Several imaging techniques are currently at our disposal for detecting various biological parameters within live cells in terms of both environmental factors and protein activity. Exploring such variables as pH, ion concentration, and enzymatic activity has provided insight into possible drug targets as well as conveying useful details in the rational design of drugs. For instance, the overexpression of certain enzymes or receptors is associated with many pathologies, and understanding their structures and cellular processes can lead to the development of more effective therapeutics. Further, the efficacy of test compounds against pathogens can be detected based upon the fluorescence activity within each cell. Currently, three main fluorescence or bioluminescence techniques are available, each with their own unique advantages towards biological research and clinical studies.

1.2) Green fluorescent protein (GFP) as a cellular marker and drug delivery tool.

First isolated in 1961 from the jellyfish *Aquorea reigi*, green fluorescent protein (GFP) is a light emitting 20 kDa protein that has a blue glow in the presence of trace amounts of calcium (Shimomura, 2005). Bioluminescence is an effect of the energy transfer from the calcium-activated photoprotein *aequorin* to GFP, and this process is specific to the photogenic cells basal of the jellyfish umbrella as either a communication or defense mechanism (Kain, 1999). Since its discovery, GFP has been found in various other species including the colenterates *Hydrozoa abelia* and *Phialidium* as well as the anthozoa *Renilla* (Arun, 2005). Of the different species from which GFPs can be extracted, those of
*Renilla* and *Aequorea* have been most thoroughly characterized biochemically (Arun, 2005), making them powerful tools in many areas of research including cell biology, medicine, and physiology (Arun, 2005; Park, 2016).

The GFP chromophore is composed of a p-hydroxybenzenylidene-imidazolone subunit “formed from cyclization/autoxidation within a tripeptide unit of the polypeptide sequence consisting of 238 amino acids (Liederman, 2007).” X-ray crystal structures have revealed the location of the chromophore to be within the 11-stranded beta-barrel along a coaxial helix where it is shielded from the aqueous solvent surrounding the protein (Liederman, 2007), as illustrated in Figure 1.1. This affinity between the chromophore and the beta-barrel center is reinforced by the extensive hydrogen-bond network offered by the protein backbone and side chains (Liederman, 2007).
Figure 1.1: The 11-stranded beta-barrel houses the GFP chromophore. The GFP chromophore is an intrinsic component to the protein backbone (Meech, 2009). Its formation involves posttranslational modifications among three amino acid residues – Ser65, Tyr66, and Gly67 (Meech, 2009). Once the protein is correctly positioned within the beta-barrel of the protein, it undergoes a cyclization and subsequent oxidation reaction (Meech, 2009). This yields an HBDI with an extended pi-system, and it is shielded within the beta-barrel from the environment (Meech, 2009).
Employing GFP in cell-based assays is an effective technique for observing real-time analysis of molecular events in living cells, and variants of GFP have enabled both high-throughput and high-content screening (Kain, 1999; Park, 2016). For instance, fusion of GFP with proteins provides information about their locations and movements within living cells and tissues (Arun, 2005). In this way, GFP can be used as a protein tag against a variety of proteins and targeting sequences in both prokaryotic and eukaryotic cells (Kain, 1999). The advantages of GFP fusion over traditional antibody staining are its enhanced sensitivity, resolution, and efficiency: “the GFP tag eliminates the need for the fixation, cell permeabilization, and antibody incubation steps (Kain, 1999).”

Further, GFP variants with different spectral properties can be used to determine characteristics of subcellular compartments such as size, shape, mobility, as well as dynamic changes to proteins during or in response to external stimuli (Arun, 2005). Also, fluorescence resonance energy transfer (FRET) has been used to study protein-receptor interactions, giving insight into homo- and heterodimerization and oligomerization (Arun, 2005). Another advantage of GFP variants are their potential to serve as cellular biosensors, including the ability to report intracellular pH, calcium levels, halide ion concentrations, protein-protein interactions, enzyme activity, and to indicate the redox states in living cells (Hanson, 2002).

Detecting redox potential is relevant to drug therapy against pathogens such as *Mycobacterium tuberculosis* (Mtbc) because it can overcome its host's immune system by secreting the antioxidant mycothiol (MSH) (Bhaskar, 2014). The plentiful quantity of reducing equivalents given off by MSH protects the pathogen from the usual oxidative stress associated with infection (Bhaskar,
Using redox-sensitive fluorescent probes, real-time measurements of mycothiol redox potentials can be determined (Bhaskar, 2014). In turn, this information can be used to identify anti-TB drugs that accelerate oxidative stress in cells infected with MH (Bhaskar, 2014). In addition to studying redox potentials, GFP can be used to evaluate changes in cellular pH. This requires the use of dual emission GFP variants, which have been used to measure differing pH levels within various cellular compartments (Hanson, 2002).

Dual emission GFP variants exhibit an excitation spectrum with two major absorption bands that conveys information about the internal ground-state equilibrium between the neutral and the anionic forms of the chromophore (Hanson, 2002). For example, the excited state of the neutral chromophore fluoresces blue light whereas the anionic chromophore gives off green fluorescence (Hanson, 2002). The latter occurs as a result of the fast internal proton transfer from the excited state of the neutral chromophore, yielding an anion (Hanson, 2002). This excited-state proton transfer (ESPT) is outlined in Figure 1.2.
Figure 1.2: Excited-State Proton Transfer (ESPT). GFP can emit light at either 395 nm or 480 nm depending on the ionic state of the chromophore – protonated (A) or deprotonated (B) (Meech, 2009). Fluorescence is generated when excited state A* undergoes an excited-state proton transfer (ESPT), causing it to decay to the anionic ground state B (Meech, 2009). This feeds into the cycle by yielding the reprotonated A state, which can then jump to the excited state again and undergo another ESPT to produce fluorescence (Meech, 2009). The intermediate state I* connotes the emissive deprotonated state that ultimately relaxes to the ground state A (Meech, 2009).

In order to detect changes in cellular pH, certain modifications to this proton transfer network need to be made that will result in pH sensitive GFP variants (Hanson, 2002).

Specifically, “these variants retain the two absorption bands characteristic of wild type, but the ratio of the neutral and anionic populations depends strongly on pH (Hanson, 2002).” One variant in particular, deGFP4, has proven to be a successful intracellular pH indicator in mammalian cells (Hanson, 2002). Its emission for all excitation wavelengths increases proportionately to increases in pH due to the dependence of the chromophore on hydrogen bonding (Hanson, 2002). The two main absorption peaks of GFP at 398 nm and 478 nm indicate the chromophore and its conjugate base due to the increasing intensity of the absorption band at higher pH values (Liederman, 2007).

This pH dependence on the spectrum clearly demonstrates the existence of a ground-state acid-base equilibrium within the chromophore (Liederman, 2007). Likewise, the excitation emission spectra is similarly influenced by pH in that lower pH values are associated with a larger blue shift of the band position (Liederman, 2007). For instance, a pH decrease from 7.9 to 6 results in a width increase from 950 cm\(^{-1}\) to 1100 cm\(^{-1}\) (Liederman, 2007). These emission bands are depicted by the I* (1100 cm\(^{-1}\)) and B* (950 cm\(^{-1}\)) bands, and demonstrate how selective mutations in GFP can change its ESPT, thus, allowing it to serve as a pH marker (Liederman, 2007). Another tool used to study pH changes in living cellular compartments is using the GFP variant blue fluorescent protein (BFP) as a reporter (Elsliger, 1999).

The different spectral properties of fluorescent proteins, such as BFP, are mainly due to their chemically distinct chromophores, which are formed via self-catalytic modifications of internal amino acids (Sarkisyan, 2015). This diverse variety of chromophores is illustrated in Figure 1.3. The cyclization and oxidation of the protein backbone, shown in Figure 1.4, results in an extended
aromatic system capable of absorbing and emitting visible light (Liederman, 2007). It is noteworthy that the central chromophore-forming residue of all natural fluorescent proteins is tyrosine because this tyrosine residue can be readily substituted with another residue to form a GFP variant that emits a different wavelength (Liederman, 2007). For example, exchanging the tyrosine with a histidine or phenylalanine residue produces blue fluorescence (Liederman, 2007).
Figure 1.3: Chemically distinct chromophores with unique spectral properties. The many different chromophores that can be built within the GFP beta-barrel are illustrated here (Mishin, 2015). These compounds are formed from self-catalyzing posttranslational modifications within the protein's amino acid sequence (Mishin, 2015).

Figure 1.4: Mechanism of the autocatalytic cyclization and autoxidation. The GFP's chromophore, p-HOBDI, of the jellyfish *Aequorea victoria* is formed after a post-translational autocatalytic cyclization and an autoxidation within a tripeptide unit within the protein (Tolbert, 2012). The chromophore is housed within the 11-stranded β-barrel of the protein core, secluded from solvent (Tolbert, 2012). This orientation is reinforced by an extensive network of hydrogen bonding (Tolbert, 2012).

Another factor that influences spectral properties of GFP is the protonation state of the chromophore, which is dictated by the reversible interactions with surrounding amino acids (Liederman, 2007). Illustrating this concept is the shift in absorption maxima from 396 nm to 475 nm upon deprotonation of the GFP chromophore, and the 80-100 nm red shift in in response to the deprotonation of different tyrosine-based chromophores in both green and red fluorescent proteins (Liederman, 2007).

BFP is useful in detecting intracellular pH because it is only viable between pH 5 and pH 7, thus providing information about slight acidity (Elsliger, 1999). Other GFP variants are also used to determine the pH of alkaline organelles such as the mitochondria, the cytosol, and the Golgi (Elsliger, 1999). The production of different pH sensitive variants requires an understanding of the different protonation states of the chromophore (Elsliger, 1999). This entails determining which two titratable groups of the chromophore pi-system is protonated: the hydroxyl group at the phenolic end or the imidazolinone ring nitrogen of the heterocyclic end (Elsliger, 1999). GFP and its variants are preferable over traditional pH indicators because of their specificity to particular cellular organelles and live-cell imaging capabilities (Elsliger, 1999). Other cellular phenomena such as changes in ionic strength, redox potential, and specific binding events can also be studied via the unique proton transfer pathways of GFP variants that are sensitive to proton structure (Hanson, 2002).

With so many useful applications, variants of GFP have been designed to have special and unique characteristics for the purpose of improving imaging techniques (Arun, 2005). These enhanced characteristics include improved brightness, shifted excitation and emission spectra, organelle-specific targeting, reduced half-life, and responsiveness to environmental changes (Arun, 2005). Moreover, GFP and its variants can be used in living cells and have broad ramifications in the drug discovery
process (Arun, 2005).

In a protein visualization technique called a strand (or fragment) complementation assay, proteins ready for cellular delivery are labeled with GFP fragments (Schmidt, 2015). As these proteins labeled with GFP fragments accumulate within a particular cellular location, the fragments of GFP beta-barrels on neighboring peptides join together resulting in green fluorescence (Schmidt, 2015). These proteins can be either the drugs themselves or vectors designed to carry the drug as in cell-penetrating peptides (CPP) (Schmidt, 2015).

CPPs are primarily used to encourage the transport of drugs to their desired locations (Milech, 2015). One drawback to using these vectors is their tendency to get trapped in endosomes where their accumulation can be toxic (Milech, 2015). Therefore, it is necessary to use a diagnostic assay that discriminates between cytoplasmic delivery and endosomal entrapment and is sensitive enough to detect CPPs at very low concentrations (Milech, 2015).

To meet these requirements, a self-assembling split-GFP protein-solubility assay, illustrated in Figure 1.5, was used to detect cytoplasmic uptake (Milech, 2015). This is accomplished by expressing the larger GFP1-10 protein fragment in the cytosol while fusing the GFP S11 protein fragment to the CPP moiety (Milech, 2015). When the CPP GFP S11 fragment unites with the GFP1-10 fragment in the cytosol, the newly formed complex will fluoresce (Milech, 2015). By contrast, those CPPs that find their way to the endosome will not fluoresce because the larger fragment is not present there (Milech, 2015).
Figure 1.5: A self-assembling split-GFP protein-solubility assay. The joining of the larger GFP1-10 fragment in the cytosol with the CPP GFP S11 fragment, as it reaches the cytoplasm, results in green fluorescence (Milech, 2015). Such fluorescence does not occur if the CPP erroneously arrives in the endosome because it lacks the larger fragment (Milech, 2015). Thus, the fluorescence serves as a means to verify that the CPP drug vector arrived at its target location (Milech, 2015).

This self-assembling split-GFP protein assembling assay, also known as split-endosomal complementation escape (SEE), provides a means to identify effective peptide sequences capable of penetrating cell membranes (Milech, 2015). “Such cell-penetrating peptides find numerous uses in targeting intracellular proteins and carrying diverse cargos into cells including peptides, small molecules, DNA, RNA, and proteins (Milech, 2015).” Finding these CPPs that are able to be efficiently uptaken by the cytosol will mark an advance in intracellular drug research (Milech, 2015).

One way to improve this technique is to develop GFP variants with high quantum efficiency or faster assembly times, both of which will result in greater sensitivity in detecting complemented fluorophores (Milech, 2015). Also, this assay could be used to measure the cellular uptake of any molecule conjugated to the complementation component (Milech, 2015). For instance, fusion with the bifurcated endosomal agent TAT would allow the measurement of functional CPP uptake, which is involved in the pharmacological delivery of biotherapeutics (Milech, 2015). Thus, self-assembling split-GFP assays coupled to CPPs provide a means of imaging in-vivo intracellular and site-specific drug delivery (Milech, 2015).

Another method of drug delivery that employs GFP for imaging is the loading of specific therapeutic proteins or peptides onto hydrogels or hydrogel microparticles (Bertz, 2013). Understanding this delivery system is important because although many protein pharmaceuticals have great potential in curing various debilitating diseases, they readily disperse or degrade under physiological conditions (Bertz, 2013). However, with current development and improvements upon drug-delivery systems (DDSs), there is increasing potential that these therapeutic peptides will reach their target locations in tact (Bertz, 2013). One widely used technique called fluorescence microscopy has been used to study the delivery efficacy of a hydrogel system, including the molecular interactions
between a protein and the polymeric hydrogel (Bertz, 2013).

Using fluorescence microscopy in conjunction with fluorescence recovery after photobleaching (FRAP), translational diffusion behavior of the hydrogel microparticle can be examined (Bertz, 2013). GFP possesses several advantages as a marker over other fluorescent molecules due to its protein structure: GFP engages in the same binding interactions with the hydrogel as the protein-drug; GFP’s small size approximates that of protein drugs and antibodies being carried; and, the GFP chromophore within the beta-barrel center is well protected from the harsh physiologic environment (Bertz, 2013).

FRAP studies have demonstrated that GFPs remain quite immobile against the hydrogel system while certain hydrogels diffuse more quickly than others (Bertz, 2013). A low mobility and prolonged release time are required to ensure that all of the drug is deposited at the desired location (Bertz, 2013). Further characterization of hydrogels and measurements of protein mobility using FRAP are necessary to improve this DDS (Bertz, 2013).

With this myriad of applications in pharmaceutical research, GFP offers a practical means of visualizing drug delivery via different vehicles such as CPPs and hyrogels. The goal is to identify a vehicle that will efficiently deliver the drug to the target location; and, observing where it is directed within the cell along with an understanding of the environmental conditions, such as pH and redox potential, of each cellular subcompartment will provide clues as to how to modify the delivery system for more optimal transport and protein release. GFP and its variants have been invaluable tools in identifying all of these cellular and vector parameters, and future work in cell biology and medicine should employ these fluorescence molecules to delve further into intracellular drug delivery methods.

1.3) Bioluminescent properties of luciferase for drug screening and analyzing drug efficacy.

Bioluminescence is a light emitting process that occurs in both prokaryotes and eukaryotes
(Hosseinkhani, 2011), and the purpose ranges from distracting predators to attracting prey and mates (Conti, 1996). Although the specifics of this bioluminescence reaction vary across species, including bacteria, fungi, algae, fish, squid, shrimp, and insects, the reaction is essentially a two-step oxidation process with molecular oxygen (Hosseinkhani, 2011; Inouye, 2002; Tafreshi, 2007). The result of this reaction is the conversion of chemical energy into light via the luciferase catalyzed oxidation of a luciferin protein (Inouye, 2002; Conti, 1996). For example, the bioluminescence system of fireflies (Lampyridae) involves luciferin, ATP, Mg$^{2+}$, and a luciferase enzyme, which has been extensively studied (Inouye, 2002).

Firefly luciferase is a monoxygenase that catalyzes the emission of yellow-green light from the substrate D-luciferin, Mg$^{2+}$, ATP, and oxygen (Inouye, 2002), as it is written out in Figure 1.6. This reaction begins with the formation of luciferase-bound luciferyl adenylate (D-LH2-AMP) in the presence of Mg$^{2+}$ and ATP with the release of inorganic phosphate (Inouye, 2002). The specific site that is adenylated is the carboxyl group of D-luciferin (D-LH2) (Inouye, 2002).
1.6: Mechanism of Luciferase Bioluminescence. The firefly luciferase is a decarboxylating and ATP-hydrolyzing enzyme that acts on the substrate luciferin \( (\text{LH}_2) \) in a reaction that requires ATP, oxygen, and metallic cation (Marques, 2009). (Panel A) The first step of the bioluminescence reaction involves \( \text{D-LH}_2 \) reacting with ATP:

\[
\text{Luciferase + D-LH}_2 + \text{ATP-Mg}^{2+} \leftrightarrow \text{Luciferase} \cdot \text{D-LH}_2 \cdot \text{AMP} + \text{Pp}_1 \cdot \text{Mg}^{2+}
\]

This reaction proceeds through an \( S_N^2 \) nucleophilic displacement, involving the C-4 thiazole group of \( \text{D-LH}_2 \) and the phosphate groups of ATP (Marques, 2009). In this manner, the carboxyl oxygen nucleophilically attacks the electrophilic alpha-phosphoryl group of ATP with the inorganic pyrophosphate acting as a leaving group (Marques, 2009). The remaining adenylate \( (5' \cdot \text{AMP}) \) is transferred to \( \text{D-LH}_2 \) creating a luciferyl-adenylate \( (\text{LH}_2 \cdot \text{AMP}) \) intermediate (Marques, 2009). Because AMP is a good leaving group, the intermediate is conducive to subsequent reactions (Marques, 2009). (Panel B) The \( \text{LH}_2 \cdot \text{AMP} \) intermediate next undergoes oxidation and decarboxylation reactions,
resulting in the release of a photon (Marques, 2009):

$$\text{Luciferase-D-LH}_2\text{-AMP + O}_2 \leftrightarrow \text{Luciferase +AMP + CO}_2 + \text{oxyluciferin + photon}$$

Specifically, ATP increases the acidity so that the intermediate loses a proton, resulting in a carbanion (Marques, 2009). The carbanion subsequently undergoes nucleophilic attack by molecular oxygen, a reaction facilitated by the displacement of AMP (Marques, 2009). The resulting strained four-member ring (luciferin dioxetanone) is high in energy and will spontaneously break up generating CO$_2$ and oxyluciferin (Marques, 2009). Because oxyluciferin is a singlet in the excited state, its decay to the ground state releases a photon (Marques, 2009).

The next step of the bioluminescence reaction is the oxygenation of LH2-AMP with molecular oxygen to produce oxyluciferin, adenosine monophosphate (AMP), and carbon dioxide (CO$_2$) (Inouye, 2002). A photon is released upon the relaxation of oxyluciferin from its excited state to the ground state, resulting in the emission of light (Inouye, 2002). These photons can express any of the visible spectrum, such as yellow-green (fireflies) and green-orange (click beatles), and varies among species (Tafreshi, 2007). But, the mechanism by which these colors are expressed remains unknown (Hosseinkhani, 2011).

Several mechanisms have been proposed to account for the different colors of bioluminescence (Tafreshi, 2007). For instance, the solvent effect is believed to manipulate the differences in energy between the ground and excited states and, subsequently, maximize luminescence (Tafreshi, 2007). Also, the oxyluciferin keto-enol tautomerization is facilitated by the basic residues of luciferase interacting with oxyluciferin: the enol form favors green-yellow light while the keto form favors red light (Tafreshi, 2007). This would explain the red shift observed in acid environments of pH-sensitive luciferases (Tafreshi, 2007). There could also be a color-dependence on the active site conformation, which dictates the freedom of rotation of the oxyluciferin thiozolinic rings (Tafreshi, 2007), and may affect the rotation of the excited luciferin along the C$_2$-C'$_2$ bond (Tafreshi, 2007).

Another factor that influences the bioluminescence color of luciferases is the microenvironment of the active enzyme active site (Hosseinkhani, 2011; Tafreshi, 2007). As stated earlier, acidic conditions favor the emission of red light (Hosseinkhani, 2011). An additional variable affecting bioluminescence color is the luciferase structure itself, of which the oxyluciferin anion (OL-) is believed to have spectroscopic properties (Hosseinkhani, 2011; Tafreshi, 2007). Luciferase has been
shown to modulate emission color by directing the resonance-based charge delocalization of the anionic keto form of the oxyluciferin-excited state (Tafreshi, 2007).

This structure-color dependence was further reinforced by the experimental insertion of an Arg$^{365}$ residue into luciferase, which caused changes in the enzymes kinetic and structural properties (Tafreshi, 2007). In particular, the additional arginine resulted in a green shift, a bimodal spectrum, greater pH-sensitivity, changes in the conformation backbone, and faster decay time (Tafreshi, 2007). Each of these factors demonstrated the structure-emission relationship among luciferase, and a greater understanding of the relationship between luciferase structure and color-emission requires studying the structure of the luciferase enzyme itself as well as how it functions.

The firefly luciferase is a 62 kDa protein that folds into two distinct domains, the topology of which is shown in Figure 1.7 (Conti, 1996). The tertiary structure consists mainly of a compact domain containing a distorted antiparallel beta-barrel and two beta-sheets that are flanked by alpha-helices (Conti, 1996). The C-terminal domain extends from residues 440-544 to form a small separate alpha+beta domain, while the N-terminal domain consists of two beta-sheet subdomains (Conti, 1996). The latter two subdomains “are assembled to form a five-layered alph-beta-alpha-beta-alpha tertiary structure so that the two alpha-helices are sandwiched between the sheets and the other helices are packed against the outer faces (Conti, 1996). The tilt in these two beta-sheets brings the C-terminus of the parallel strands closer together (Conti, 1996).
**Figure 1.7: The topology of luciferase.** (Panel A). The enzyme luciferase contains a tertiary structure made up of many right-handed antiparallel beta-sheets to form beta-barrels (Conti, 1996). These subdomains are supported by the antiparallel arrangement of alpha-helices on each side of the sheets (Conti, 1996). (Panel B). While the composition of one N-terminal domain is mostly continuous (residues 77-222 and 399-405), the other subdomain is discontinuous (Conti, 1996).

The shared topology between the N-terminal beta-sheets are made up of eight strands that display right handed twists (Conti, 1996). One beta-sheet contains five parallel and three antiparallel beta-strands with six associated helices, and mostly occupies a single portion of the polypeptide chain (77-222) (Conti, 1996). The other beta-sheet contains six parallel and two antiparallel beta-strands along with six helices (Conti, 1996). Unlike the first beta-sheet, this is built from non-continuous portions of the polypeptide chain (Conti, 1996). “The core of each beta-sheet consists of parallel strands joined to alpha-helices with standard right-handed cross-over connections, resulting in the arrangement of helices on either side of the sheet, antiparallel to the strands (Conti, 1996).”

Like the N-terminal domain, the C-terminal domain contains an assortment of beta-sheets, beta-strands, and alpha-helices (Conti, 1996). Specifically, there are two short antiparallel beta-strands, a three-stranded mixed beta-sheet, and three helices against the sides (Conti, 1996). The resulting fold forms an alpha+beta structure (Conti, 1996), and the two domains are separated by a wide cleft (Conti, 1996).

The primary sequence of luciferase closely resembles those of peptide synthetases and acyl CoA ligases, all of which require ATP and the carboxyl moiety of their substrates (Conti, 1996). In addition, all three of these highly conserved enzymes share an active site, shown in Figure 1.8.
Figure 1.8: The conserved active site of luciferase. The luciferase enzyme contains a common active site with two other enzymes it closely resembles – peptide synthetases and acyl CoA ligases (Conti, 1996). The majority of the highly conserved residues lie in the core of the beta-barrel of the C-terminal domain, though many are found in different motifs of the enzyme (Conti, 1996). A few conserved residues around the active site are Tyr401, Ser420, and Glu344, illustrated above (Conti, 1996).

The G-x-x-x-x-G-K-[STG] motif connects the first beta-strand and alpha-helix of the beta-barrel and is responsible for binding the phosphate group from either ATP or GTP (Conti, 1996). Of these conserved residues, Lys206 is exposed to the solvent and is oriented towards the cleft separating both domains (Conti, 1996). The active site architecture is further supported by hydrogen bonding of Ser198 to a carboxylate oxygen of Glu344 (Conti, 1996). Only 2.9 angstrom away, Tyr401 assists in phosphate binding by interacting with the adenine ring of ATP (Conti, 1996). Also, the arrangement of the residues on the surface of the two domains causes them to come together and sandwich the substrate (Conti, 1996). This provides a favorable environment for the light emitting reaction, detailed earlier, because the exclusion of water from the active site in this way prevents intermolecular quenching of the excited-state product (Conti, 1996). The resulting bioluminescence is a necessary means of converting analyte concentrations to signal intensities in most modern immunoassays (Oyama, 2015).

The joining of the photoprotein to a test compound or cellular target allows for highly sensitive and rapid recordings of cellular or pharmaceutical activity (Oyama, 2015). For instance, luciferase markers have implications in identifying compounds with therapeutic potential in tissue repair, which can occur through fibrosis or regeneration (Gordon, 2015). The latter is preferable because it restores the tissue to normal function, while the former results in scar tissue devoid of any physiologic function (Gordon, 2015). Therefore, the inhibition of fibrosis with new compounds should encourage tissue regeneration, and can be observed by transfecting cells with a firefly reporter gene (Gordon, 2015). Coupling this assay to high-throughput screening allows for an analysis of large libraries of uncharacterized compounds with anti-fibrosis potential (Gordon, 2015). This massive identification of potential compounds is necessary due to the development of drug-resistance.
A consequence of long-term pharmacological EFN-beta treatment is the development of anti-drug antibodies (Hermanrud, 2016; Cludts, 2013). Specifically, neutralizing antibodies (NAbS) bind IFN-beta and prevent the therapeutic from binding to its cell surface type I IFN-receptor (Hermanrud, 2016; Cludts, 2013). However, luciferase-mediated detection of NAbS, by means of cellular transfection with a firefly luciferase reporter gene, serves as an indicator of drug efficacy (Hermanrud, 2016). Nevertheless, the emergence of multidrug-resistant parasites has created a formidable obstacle in eradicating malaria, a life-threatening disease prevalent in the tropics and subtropics (Che, 2012). Overcoming this obstacle requires utility of high-throughput screening (HTS) assays for rapid identification of antimalarial drugs (Che, 2012). One HTS technique involves a cell-based luminescent assay for antimalarial drugs based on a transgenic \textit{P. falciparum} line that expresses high levels of firefly luciferase (Che, 2012).

The development of a stable transgenic parasite line with high-level luciferase expression has advanced into a cell-based HTS luminescent assay for antimalarial drug discovery. Already, using this luciferase-based HTS assay, hundreds of compounds have been identified with inhibitory activity against the parasite (Che, 2012; Lucantoni, 2013). For example, placing a luciferase reporter gene under the control of the same promotor as the gametocytogenesis-specific gene Pfs16 was used to screen against the asexual stages of \textit{P. falciparum} (Lucantoni, 2013). It was found that targeting the gametocytes with test compounds on days one to four of gametocytogenesis allowed for the detection of longer-acting compounds with greater specificity, which adheres to the WHO recommendations of treating patients with 3-day courses of antimalarial therapies (Lucantoni, 2013).

Despite some shared drug-targets (eg. Hemozoin) shared between these early-stage gametocytes and the asexual stages of \textit{P. falciparum}, this luciferase-based HTS stresses the unique
threat posed by antimalarial compounds to gametocytes (Lucantoni, 2013). This underscores the importance of compound screening on gametocytocidal activity and the implications this has on developing novel transmission-blocking antimalarial drugs (Lucantoni, 2013). However, the applications in which luciferase-based screens and assays can be used is limited by their narrow variety (Mofford, 2014).

As mentioned, D-luciferin is the firefly's natural substrate for its luciferase; however, synthetic luciferins can create new wavelengths to image more transparent tissues (Mofford, 2014). Substituting the 6'-hydroxyl group of D-luciferin with a 6'-amino group results in a red shift of the light emitted (Harwood, 2011). For example, the synthetic luciferin CycLuc1 more readily accesses the luciferase enzyme than D-luciferin, making it more useful for imaging studies (Mofford, 2014). Further, synthetic aminoluciferins modeled after CycLuc1 and CycLuc2 greatly increased the total photon flux of near-IR light from live luciferase-expressing cells over D-luciferin (Mofford, 2014). The above examples demonstrate that bioluminescence can be improved through chemical modifications to luciferin (Mofford, 2014). For instance, the dramatically improved bioluminescent ability of CycLuc1 allows for the imaging tumor cells using 20-200-fold less substrate than D-luciferin (Mofford, 2014). Also, this synthetic luciferin's greater sensitivity allows for deep brain imaging and detection of protease activity via bioluminescence that was previously unattainable (Mofford, 2014). A wider variety of luciferase-assays will translate into more approaches to screen for potential drugs. One case in particular is the search for a cure to tuberculosis.

Tuberculosis (TB), caused by Mycobacterium tuberculosis (Mtb), is a serious public health issue inflicting millions of people each year (Yuriko, 2015). Current treatment involves taking medications over the course of many months, which as stated earlier, encourages the development of drug-
resistance, resulting in multi-drug resistant Mtb (MDR-Mtb) and, thus, complicating TB therapy (Yuriko, 2015). Owing to the danger posed by this disease to human health, it is necessary to develop new TB drugs that can be administered over a short term with the capability of controlling MDR-Mtb (Yuriko, 2015). However, this requires overcoming the challenges of Mtb's slow growth rate, high virulence, and its contagiousness (Yuriko, 2015).

Using the firefly luciferase with a nonvirulent recombinant Bacillus Calmette-Geurin (rBCG) serves as a safe and effective way of detecting viable cells in response to infection (Yuriko, 2015). Since ATP is a reagent in the luciferase catalyzed bioluminescence reaction, only living cells will have the ability to become luminescent (Yuriko, 2015). Thus, “the number of photons released by the enzymatic activity of intracellular luciferase can be correlated with the viability of the cells (Yuriko, 2015).” This means of testing for drug-resistant Mtb employs a luciferase-based reporter phage, in which an rBCG was equipped with a multi-plasmid vector and a promoter (Yuriko, 2015). This construct allowed for the constant expression of luciferase and Mtb-specific genes for the ultimate purpose of screening for TB drug candidates (Yuriko, 2015).

Introduction of TB drugs such as LVFX, RFP, SM, or INH into a 96-well plate, each containing the rBCG-MDP1-luc construct does not result in any increase in luciferase activities (Yuriko, 2015). By contrast, the non-treated culture's luciferase activity increases dramatically over a four day cultivation period, indicating that these bacterial cells were still alive (Yuriko, 2015). In this way, the number of viable bacteria (CFU) counted after 3 to 4 weeks of culture coincides with the luciferase activity (Yuriko, 2015).

“After the rBCG-MDP1-luc was grown on the culture plate, each drug was added and the luciferase activity was monitored for the following 96 h (Yuriko, 2015).” All treated cultures displayed
decreases in luciferase activity and CFUs, though the decreases varied among the treatments (Yuriko, 2015). For instance, the DNA synthesis inhibitor LVFX proportionately decreased both luciferase activity and CFU count, whereas treatment with RFP and SM caused rapid decreases in luciferase activity without lowering CFUs (Yuriko, 2015). The disparity between luciferase activity and CFU was even more pronounced after treatment with the cell wall inhibitor INH (Yuriko, 2015). This luciferase-based assay-system is able to detect the effect of all types of TB drugs within 2 to 4 days, though it exhibits greater sensitivity to compounds that inhibit replication, transcription, and translation than inhibitors of cell wall synthesis (Yuriko, 2015).

This rBCG-MDP1-luc construct can also be used to screen for active compounds against TB (Yuriko, 2015). Many extracts from the popular antinomycetes-based antibiotics have been shown to be potent against Mtb in various dilutions (Yuriko, 2015). In fact, 10,080 extracts from actinomycetes were screened for possible TB drugs (Yuriko, 2015). Increasing the dilution of these extracts narrowed the field of potential drug candidates from 609 extracts at a 1:100 dilution to 41 extracts at 1:200 dilution (Yuriko, 2015). Finally, extract 1904-1 was chosen out of the 18 extracts found potent after a 6,400-fold dilution for further characterization as a TB drug-candidate.

Comparison of the bacteriocidal activity of the 1904-1 extract against rBCG-MDP1-luc with those of currently available anti-TB drugs revealed an MIC of 0.125 micg/ml, which was more effective than SM and LVFX and similar to that of INH (Yuriko, 2015). The MIC of this extract was also determined against Mtb H37Rv and the two clinical MDR-Mtb strains, MDR-Mtb5 and MDR-Mtb44, all increased to 0.5, 2.0 and 7.5 micg/ml (Yuriko, 2015). This potency also expanded to one other species of *Mycobacterium, Mycobacterium smagmatis*, but not to any other tested bacterial species (Yuriko, 2015).
Testing the bacteriocidal effect of 1904-1 in Mtb-infected macrophages revealed that the substance of 1904-1 is effective against intracellular Mtb, and NMR studies determined this compounds identity (Yuriko, 2015). Further, the genome analysis demonstrated that the target of this active substance is clpcl (Yuriko, 2015). This overall experimental procedure illustrates how “a drug screening system using a rBCG expressing luciferase is useful for discovering TB drug candidates in a timely fashion (Yuriko, 2015)”, which will provide new compounds to combat MDR-Mtb. Moreover, luciferase-based screening assays may be an effective techniques in the advancement of drug discovery (Yuriko, 2015).

In summary, the bioluminescent activities of many species of prokaryotes and eukaryotes have been analyzed and applied for research purposes in molecular and cell biology and to advance drug-screening techniques. The unique characteristics of bioluminescence such as the substrates involved have been exploited for the purpose of understanding many cellular processes and evaluating drug efficacy. For instance, certain modifications to the natural D-luciferin substrate can be made to cause a shift in its emission color, which is useful in observing more transparent tissues such as those situated deep in the brain. Another reagent, ATP, can be used to measure the vitality of infected cells following drug treatment, which eventually leads to the isolation of a test compound that can progress towards a potential drug candidate. These multifaceted approaches of using luciferase-based assays to observe both cell processes and drug activity offers a promising tool in the development of new drugs and the identification of novel drug-targets (Yuriko, 2015).

1.4) Bodipy as a diverse tool for probing proteins and assisting in drug delivery.

The propensity of previous fluorophores to decompose in response to prolonged light exposure and the interference of genetically engineered fluorescence with normal protein function
has called for the synthesis of a new type of fluorescent molecule (Benniston, 2009; Dowada, 2015). A certain half-porphyrin-difluoroboron complex named Bodipy, short for borondifluorodipyrromethene, has demonstrated greater resistance to light-mediated degradation and has been utilized in a diverse array of applications (Benniston, 2009). The structure of BODIPY (4,4-Difluoro-4-borata-3a,4a-diaza-s-indacene) is drawn in Figure 1.9, and is a prominently used and highly fluorescent dye discovered in 1968 (Yong, 2014; Descalzo, 2008). Its many qualities include intense absorption, high fluorescence quantum yield, narrow absorption and emission bands, insensitivity to polarity and pH (making them stable in a physiological environment), chemical robustness, good photo-stability, high solubility, and resistance towards self-aggregation (Yong, 2014; Descalzo, 2008; Bergstro, 2002).
Figure 1.9: The structure of Bodipy. Bodipy (4,4-Difluoro-4-borata-3a,4a-diaza-s-indacene), or borondifluorodipyrromethene, consists of half-porphyrin ring and a difluoroboron complex (Benniston, 2009). The substituents R along with their substitution pattern and the group at the meso position influence the dyes photophysical properties (Benniston, 2009).

Bodipy derivatives can be readily synthesized in high yields from pyrrole-based starting materials, and their photophysical properties are dependent upon three things: the substituents R, their substitution pattern, and the type of group at the meso position (Benniston, 2009). Specifically, it is the electron withdrawing/donating groups that influence the dye's spectral properties, which are typically highly colored with narrow absorption bands (Benniston, 2009). Moreover, a highly conjugated Bodipy core will exhibit near-infrared (NIR) wavelengths (Benniston, 2009). Other characteristics of these dyes are: the NIR wavelengths exhibited by a highly conjugated core; the minimal effects that solvents exert on absorption spectra; intense color emissions and low Stoke shifts (SS), suggesting that permanent dipole moments do not differ between the ground state and the excited state; low energy triplet state mitigating the risk of degradation via self-produced singlet oxygen; and, a redox chemistry which is manipulated by the level of substitution at the dipyrrromethene core (Benniston, 2009; Descalzo, 2008).

Biological molecules of interest can be targeted with this fluorescent dye by adding certain functional groups to the Bodipy core (Zheng, 2008). In particular, various aryl substituents have been added to the meso position, imparting certain photophysical properties (Zheng, 2008). This change in characteristics is due to the sterical configurations between the aryl group at the meso position and the Bodipy core (Zheng, 2008). Also, the presence of certain beta-substituents can prevent the free rotation of the phenyl group and, in turn, reduce the loss of energy from the excited state via nonirradiative molecular relaxation (Zheng, 2008).

Other substituents can increase the dihedral angle between the Bodipy core and the phenyl group causing a blue-shifted absorption and emission spectra, as do methyl groups (Zheng, 2008). The addition of an even bulkier naphthalenyl group causes a bathochromic shift in the absorption and
emission spectra due to the increased angle between the naphthenyl group and the indacene plane (Zheng, 2008). By contrast, the introduction of the five-membered heteroaromatic furan ring at the meso position decreases the dihedral angle and results in a red-shifted linear absorption and emission spectra (Zheng, 2008).

These conformationally restricted Bodipy dyes are all able to image cells in both the visible and near-infrared (NIR) spectra (Zheng, 2008). In doing so, these probes can be readily conjugated with biomacromolecules and used to image cellular processes and deep tissue over a long period of time with little photodamage (Zheng, 2008). Such NIR derivatives offer a convenient solution to the current problem of fluorescent damage imaging probes can cause to living tissues (Yong, 2014).

The growing uses of Bodipy in biological research and medical applications has necessitated the search for new fluorescent NIR dyes (Descalzo, 2008). Such dyes fall within the boundaries of the biological window, between 650 and 900 nm, in which absorption of water, scattering of light, autofluorescence, and absorption by tissues and cells are all minimal (Descalzo, 2008). These attributes improve imaging techniques by increasing the dye's brightness and photostability while lowering the risk of tissue damage (Descalzo, 2008). This photostability of Bodipy is provided by the rigidity of the boron bridge, which prevents cis-trans isomerization and twisting, common deactivation pathways among other chromophores (Descalzo, 2008). Many Bodipy derivatives, illustrated in Figure 1.10, have been engineered to display bathochromic shifts, with absorption maxima between 495 and 525 nm and emission maxima between 505 and 545 nm (Descalzo, 2008).
Figure 1.10: Different substituents R and meso groups. The different substituents R and meso groups attached to the porphyrin ring of Bodipy influence its photophysical properties and can cause a bathochromical shift in the spectral bands (Descalzo, 2008). Modifications have been made to these portions of the dye to create derivatives that absorb approximately 500 nm and include: aryl substitution at positions 3 and/or 5 (2 and 5); ethynlphenyl substitution (3); styryl substitution (4); vinyl substitution at the 2 and 6 positions; aryl substitution at the 1, 3, 5, and 7 positions; exchange of the meso-(8)-carbon for a nitrogen atom together with 1,3,5,7-tetraaryl substitution (6); and, aromatic ring fusion (7), by means of increasing pi-conjugation (Descalzo, 2008).


In order to lower the dyes wavelength to the infrared region, many derivatives have been
developed such as styryl-substituted BODIPYs, aromatic units-fused BODIPYs, conformationally constrained BODIPYs, and azo-BODIPYs (Yong, 2014). Such low wavelength fluorescence is established by pi-conjugation, which increases electron delocalization (Yong, 2014). Three methods accomplishing this are functionalization at the alpha-, beta-, and meso-sites of BODIPY, fusion of either pi-extended pyrrole units or aromatic units, and replacement of the meso-carbon by an imine-type nitrogen atom (Yong, 2014).

Bodipy derivatives experience altered fluorescence and emission wavelengths in response to environmental conditions such as pH, ion concentration, and polarity (Descalzo, 2008; Bergstro, 2002), and recognition of these properties has lead to a new detection method called DDEM (Bergstro, 2002). This DDEM method works by “monitoring donor-donor energy migration within a pair of chemical and photophysically identical groups (Bergstro, 2002),” which has implications in studying protein structure (Bergstro, 2002).

The G protein estrogen receptor (GPER) is pivotal in many physiologic processes that use estrogen including an array of transduction pathways and biologic responses in both normal and malignant cells (Papalia, 2015). In fact, the GPER has been correlated with various cancers including of the breast, endometrium, and ovary (Papalia, 2015). An appropriate imaging probe, such as Bodipy, would decipher the process of GPER-mediated signaling and offer insight into drug design (Papalia, 2015). Bodipy is also capable of imaging other cellular components such as ions including Fe$^{3+}$, Cd$^{2+}$, Cu$^{2+}$, and other proteins (Moriarity, 2014).

A Bodipy derivative with a bromobenzydioxolyl substituent is an effective probe for imaging tumors because of its NIR fluorescence (Papalia, 2015). This dye binds within the narrow cleft of the GPER bonding pocket and participates in different types of interactions with both the hydrophobic and
hydrophilic residues along the alpha-helices, as depicted in Figure 1.11 (Papalia, 2015).
**Figure 1.11: Bodipy binding to GPER.** The GPER binding pocket is a deep cleft composed of alpha-helices that are covered with both hydrophobic and hydrophilic residues (Papalia, 2015). The Bodipy dye is able to interact with the residues of this binding pocket through various types of interactions and emit NIR fluorescence (Papalia, 2015).

Specifically, Bodipy engages in hydrogen bonding with Gln138, the carboxyl group of Pro303, a halogen bond with Glu115 and hydrophobic interactions with Val116, Leu137, Met141, Phe206, and Phe208 (Papalia, 2015). Upon binding and excitation at its absorption band, Bodipy exhibits an intense fluorescence peak at 520 nm (Papalia, 2015). These binding interactions and subsequent NIR fluorescence demonstrates the utility of Bodipy in selectively imaging the GPER in order to unravel the signaling process and binding interactions that take place at the receptor binding site (Papalia, 2015).

Understanding the hydrophobic interactions between proteins is paramount to exploring molecular recognition processes, biological functions, and many protein misfolding diseases (Dorh, 2015). Of the many interactions required for protein-folding and stability, hydrophobic interactions are the most prominent and play important roles in many intermolecular interactions such as protein-ligand interactions and receptor recognition (Dorh, 2015). However, point mutations or oxidative damage can increase the surface hydrophobicity of proteins leading to a plethora of age-related proteinopathies (Dorh, 2015). The development of probes for detecting protein-surface hydrophobicity should provide valuable insight into improving the design of drugs specific to such hydrophobic proteins (Dorh, 2015).

Different Bodipy-based hydrophobic sensors (Hpsensors) have been used to measure protein hydrophobicity, and their hydrophobic-sensing abilities can be tweaked via various aryl substitutions at the meso position (Dorh, 2015). This aryl substitution at the meso position of the Bodipy core increases dye sensitivity to solvent polarity and protein hydrophobicity, and the aqueous solubility of these dyes are enhanced by substituting the beta-positions with 2-methoxyethylamine (Dorh, 2015). This presence of three different donor aryl groups (NH$_2$, NHAc, and OCH$_3$) at meso position significantly enhance the fluorescence intensity of the HPsensors in response to binding hydrophobic
proteins (Dorh, 2015). In addition, the more nucleophilic aryl groups of these HPsensors display greater hydrophobic sensing abilities, including the differentiation of varying degrees of hydrophobicity among proteins (Dorh, 2015).

The enhanced fluorescence can be accounted for by the restricted rotation at the meso position and the 2-methoxyethylamine preventing peripheral binding and dispersal of fluorescence (Dorh, 2015). Also, these dyes show a red shift in excitation (561-569nm) and emission (577-587nm), probably due to the increased pi-conjugation of the chromophore (Dorh, 2015). This strong signal-to-noise ratio means that this probe can be used to detect even minute quantities of hydrophobic protein, making this technique a model for the detection of other molecules and types of interactions (Dorh, 2015).

Over one-hundred different such proteins exist in the blood plasma, and changes in this protein content can serve as an indicator or even the cause of disease (Marfin, 2016). However, these changes in plasma and serum protein concentrations can be monitored using fluorescent probes (Marfin, 2016). Of these fluorescent probes, Bodipy dyes are the most relevant to such measurements because their hydrophobicity makes them useful markers of lipids and lipophilic substances (Marfin, 2016).

The relative fluorescent intensity of these bodipy-protein interactions is based upon the concentration of plasma proteins; higher concentrations result in greater fluorescence intensity (Marfin, 2016). By contrast, low plasma protein concentration leads to the formation of nonfluorescent molecular complexes and subsequent static quenching (Marfin, 2016). Although the specific binding details vary with the functional groups attached to the dye, the derivatives generally embed into a hydrophobic pockets of the proteins (Marfin, 2016). Further, these interactions are
dynamic at concentrations of protein greater than 1 g/L, but static at lower concentrations (Marfin, 2016). Also, the more stable dyes were able to interact with both hydrophobic and hydrophilic residues, each fluorescing a different color (Marfin, 2016). Such binding characteristics demonstrate the usefulness of Bodipy probes in biologic studies (Marfin, 2016).

Despite the utility of Bodipy dyes penetrating cell membranes, they tend to aggregate within subcellular membranes due to their lipophilicity (Kowada, 2015). Nevertheless, these far-red and NIR BODIPY derivatives offer a promising means of bioimaging (Yong, 2014). Of the many mechanisms to highlight their target, this dye uses photo-induced electron transfer (PET), intramolecular charge transfer (ICT), Forster resonance energy transfer (FRET), or coupled mechanism to name several (Yong, 2014). In particular, PET is a non-invasive in vivo molecular imaging technique that allows for the evaluation of biological molecules (Kondo, 2015). Already, this technique has made biological and clinical impacts in the detection of proteolytic matrix metalloproteinases (MMPs) and other proteins and receptors using synthetic Bodipy derivatives (Kondo, 2015; Brizet, 2014). For instance, MMPs are involved in various cellular processes, and observing their activity in cancer cells should help in designing appropriate therapies (Kondo, 2015).

Bodipy dyes serve as important probes for the rapid and accurate localization of biomarkers characteristic of malignant cancers (Wang, 2015). Enzymes in general are particularly telling biomarkers because of their high specificity and ability to rapidly catalyze reactions (Wang, 2015). One such enzyme, gama-Glutamyltranspeptidase (GGT), is a useful biomarker because the cysteine formed by its catalyzed reaction is necessary for the growth and survival of cancer cells (Wang, 2015). Thus, the abundance of cysteines resulting from overexpression of GGT is beneficial to cancer cells and can even lead to the formation of several types of cancers including cervical and ovarian (Wang,
For these reasons, detection of GGT serves as both diagnostic and therapeutic tools against cancer (Wang, 2015). Also, employing a Bodipy-GSH adduct should advance our understanding of how GGT works in complex biological systems and assist in the rational design of drugs specific to this target (Wang, 2015).

One strategy of anti-cancer drug-delivery involves the insertion of nanoparticle-based drugs inside self-assembled, Bodipy tagged micelles (Liu, 2015). Of the many attributes characteristic of drug-incorporated micelles, including increased solubility, metabolic stability, minimal side effects, elevated bioavailability, and efficacy of the incorporated therapeutic agent, their prolonged systemic circulation lifetimes and enhanced permeability and retention effect encourages their accumulation in cancer cells (Liu, 2015). The exact location of drug release is controlled by the cellular environment and tracked using fluorescence techniques. One environmental stimuli to which micelles can be programmed to dissociate in response to is pH, and Bodipy derivatives can be tagged to these micelles and function as pH indicators (Liu, 2015). This incorporation of Bodipy into the micelle-drug complex enables real-time quantification of pH changes in live cells from the time the micelle was administered to the release of a cancer-adverse antibiotic such as Doxorubicine (Dox) (Liu, 2015).

Dox is held within the micelle core through electrostatic interactions; however, upon entering an acidic environment such as the endosomes or lysosomes of cancer cells, its carboxylic groups become protonated (Liu, 2015). This disrupts the interactions holding the drug to the micelle so that upon the simultaneous dissociation of the micelle, the drug is released into the cell for its anti-cancer activity to take effect (Liu, 2015). This process is illustrated in Figure 1.12.
Figure 1.12: Bodipy detection of the pH-controlled delivery of drugs from micelles. The cancer-adverse antibiotic Doxorubicin (Dox) inserts into the core of the polymer micelle carrier via electrostatic interactions (Liu, 2015). Incorporation of Bodipy in this complex allows for real-time quantification of pH changes in live cells, as the carboxylic groups of the anthracycline-based drugs become protonated and the micelle dissociates once reaching an acidic environment (Liu, 2015). This technique has proven very effective against HeLa cells, as the acidic environment of endosomes and lysosomes speeds this drug release phenomenon once the micelle is brought into the cancer cell (Liu, 2015).

This pH-sensitive, micelle-mediated mechanism of drug delivery is a promising technique to target various cancer cells, as it illustrates the utility of Bodipy derivatives coupled to micelles as just one of many applications this dye has in biologic research and pharmaceutical therapy.

In summary, the synthetic dye Bodipy offers a reliable means of imaging a wide range of cellular targets owing to its many assets. Its high photostability, lack of interference in normal protein function, and flexible emission spectra place this dye ahead of other fluorescence imaging techniques. A wide of assortment of derivatives, each containing their own unique photophysical properties, can be synthesized by simply changing the functional groups at the meso and beta positions of the fluorophore core. This change in photophysical properties and emission spectra is due primarily to the restriction in rotation from steric hindrance, an expanded pi-conjugation, and the presence of certain electron withdrawing/donating groups. Derivatives have been synthesized that emit in the NIR spectrum, thus, allowing for the imaging of more transparent tissue and minimizing the fluorescent damage done to living tissues. Further, specific cellular proteins can be targeted via the addition of certain functional groups. In particular, Bodipy is a useful clinical tool in the diagnosis of cancer, as it has been used to detect the overexpression of certain proteins, including GPERs, GGTs, and MMPs, characteristic of various cancers. Moreover, the dye is an important therapeutic tool in the delivery of drugs, such as the cancer-adverse antibiotic Dox, to their cellular targets in cancer cells. Therefore, the dye's hydrophobicity coupled to the insertion of certain functional groups that give the dye protein specificity allows for the study of protein folding and binding interactions, which provides valuable insight in the rational design of drugs.

1.5) Conclusions.

The naturally occurring fluorescence or bioluminescence proteins, GFP and luciferase, are
expressed in a wide range of species in both prokaryotes and eukaryotes, and their exploitation has served as valuable research tools in the fields of biology and medicine. These proteins offer several conveniences over traditional antibody staining, and are able to partake in live cell imaging. Such imaging imparts knowledge of alterations within the cellular environment as well as several processes involving different molecules. The detection of overexpressed enzymes or receptors associated with pathology has clinical implications in that it clarifies appropriate drug targets and may improve the rational design of drugs against these targets. Moreover, these imaging techniques can be used to study the efficacy of certain drugs or drug-delivery processes, as in the case of GFP surveillance over hydrogel-mediated drug-delivery or luciferase detection of drug-resistance antibodies from the pathogen Mtx.

The synthetic Bodipy dye, along with its derivatives, offers a more cell-friendly, non-protein-disruptive alternative. The photophysical and emission properties of this synthetic dye can be altered by changing the functional groups at the meso and beta positions of the fluorophore core. This has even enabled emissions in the desired NIR range, which further minimizes fluorescent tissue damage and allows for the imaging of more transparent tissues. Other characteristics of this dye such as its improved photostability and hydrophobicity allow for prolonged imaging of hydrophobic protein regions. Consequently, protein folding and binding interactions can be studied in greater detail with implications in the drug discovery process.
CHAPTER 2

Using x-ray crystallography to visualize drug-target interactions

2.1) Introduction.

Various molecules that play crucial roles in bacterial and viral cell signaling and replication have been proposed as targets in the rational design of antibacterial and antiviral drugs. The identification of novel drug targets is of paramount importance in warding off parasitic intruders because of the parasite's ability to develop resistance to current treatments. Effective development of drugs requires a detailed understanding of those targets' structures and mechanisms of function. One tool that offers such insight is x-ray crystallography.

2.2) Drugs crystallized to G protein-coupled receptor kinases (GRKs).

G protein-coupled receptor kinases (GRKs) desensitize active G protein-coupled receptors as a means of regulating cell signaling (Homan, 2015). This is accomplished through the phosphorylation of activated G-protein-coupled receptors (GPCRs), among other proteins, resulting in receptor desensitization and trafficking (Komalov, 2015). Left uninhibited, GPCRs become activated by extracellular signals, which are then translated into cytoplasmic signaling via a small heterotrimeric G protein with various downstream effects (Homan, 2015).

Humans contain seven GRK subfamilies (GRK1-7), all of which contain a kinase domain in their catalytic core (Homan, 2015). The distinguishing feature of each subfamily is the C-terminal region, which is involved in membrane localization (Homan, 2015). Examples of these differences include the prenylation site of the GRK1 subfamily, the pleckstrin homology (PH) domain of GRK2, and GRK4's amphilic helix (alpha-CT) (Homan, 2015). The opposite N-terminus better orients the molecule for optimal bonding interactions (Homan, 2015). GRK2 and GRK5 are two ubiquitous GRKs that play roles...
in cardiovascular disease, potentially making them important drug targets in the fight against cardiovascular disease (Homan, 2015).

GRK5 is uniquely separated from its cytoplasmic counterparts in the nucleus via G_q signaling (Komalov, 2015). Understanding GRK5 is of particular importance owing to its role in many diseases such as cardiovascular disease, cancer, and various neurological and metabolic disorders (Komalov, 2015). Of these, the role of GRK5 in cardiovascular disease is most widely studied, but has also been linked to type 2 diabetes, prostate tumor growth and metastasis, Parkinson disease with dementia, and Alzheimer disease (Komalov, 2015; Homan, 2016).

Cardiac hypertrophy refers to an increase in cardiac mass and can be a normal component of physiology as occurs during growth and pregnancy (Naskar, 2015). However, this increased mass can be pathological when it occurs in response to erroneous signals (Naskar, 2015) such as the overexpression of GRK5 (Gold, 2012). GRK5 is a class II histone deacetylase (HDAC) kinase that lies downstream of the nodal hypertrophic signal transducer, Gq (Gold, 2012).

When overexpressed, it facilitates heart growth by activating the myocyte enhancer factor 2 (MEF2) (Gold, 2012). In fact, cardiomyocytes require GRK5 to mediate cardiac hypertrophy following stress, and decreasing the level of GRK5 in the heart is a means of avoiding pathologic cardiac growth and heart failure – the leading cause of death in the western world (Gold, 2012). With so many disease implications, it is likely that GRK5 is a valuable target in fighting these diseases (Komalov, 2015).

Overexpression of GRK5 in mice has resulted in dramatically increased beta-adrenergic receptor desensitization, decreased cardiac output and contractility, exaggerated hypertrophy, and early heart failure (Homan, 2015). Such overexpression is believed to pertain to the accumulation of this kinase in the nucleus as a result of residues 388-395, which are located in the large lobe of the kinase domain.
and contain a nuclear localization signal (Homan, 2015; Gold, 2013). This localization is further dependent upon the binding of Ca\(^{2+}\)-CaM at the N-terminus (Homan, 2015), occurring only after selective adaptation of hypertrophic Gq-coupled receptors (Gold, 2013). Within the nucleus, GRKF phosphorylates histone deacetylase-5 (HDAC5), a myocyte enhancer factor-2 (MEF2) repressor, which is not exported from the nucleus and leads to transcription of MEF2-associated hypertrophic genes (Homan, 2015). In this way, the accumulation of GRK5 in the nucleus causes cardiac hypertrophy (Homan, 2015). However, a mutation within the amino-terminus of GRK5 that inhibits CaM binding will, in turn, halt the movement of GRK5 in the nucleus – even during times of GRK5 overexpression (Gold, 2013). This highlights the amino-terminus as a potential drug target to compete with the binding of CaM (Gold, 2013).

Studying GRK5 as a drug target entails understanding the mechanisms that mediate GRK function, and any changes in conformation dictated by its crystal structure, depicted in Figure 2.1, could explain the cause of pathology (Komalov, 2015). The crystal structure of GRK5 revealed its various domains including an N-terminal alpha-helical, calmodulin/PIP\(_2\) binding and RH domains, and central kinase domain, and a second calmodulin/PIP\(_2\) binding domain (Komalov, 2015). Since crystal structures tend to adopt the lowest energy level and most stable conformation, it should not be assumed that this structure is what appears in-vivo (Komalov, 2015).
Figure 2.1: The crystal structure of GRK5. The crystal structure of GRK5 contains an RH domain, a central kinase domain, and two calmodulin/PIP$_2$ domains (Komalov, 2015). Further, a larger N-terminal lobe and smaller C-terminal lobe constitute the kinase domain while the GRK5 C-lobe contains a two stranded antiparallel beta-sheet (Komalov, 2015).

The crystal structure of GRK5 bound to the ATP analogue, AMP-PNP, demonstrated the similarity of this enzyme with other GRKs and protein kinases in general (Komalov, 2015). A larger N-terminal lobe and smaller C-terminal lobe constitute the kinase domain while the GRK5 C-lobe contains a two stranded antiparallel beta-sheet (Komalov, 2015). The active site appeared near the RH domain (Komalov, 2015). Two lobes of alpha-helices make up the GRK RH domain and their two ends join near the N- and C-lobes to form a V-shape (Komalov, 2015). This puts the RH and kinase domains of GRK5 in close contact so that they allosterically regulate one another (Komalov, 2015).

Size-exclusion chromatography experiments revealed the molecular mass of GRK5 as roughly 70 kDa, indicating the monomeric state (Komalov, 2015). Despite evidence of GRK5's monomeric existence, it is thought to be capable of dimerizing (Komalov, 2015). Further, the intermolecular autophosphorylation of GRK5 in the presence of phospholipids suggests that GRK5 dimerization is linked to phospholipid binding (Komalov, 2015).

The kinase domain of GRK5 exists in a partially closed conformation with a notably narrow catalytic cleft (Komalov, 2015). It also has a C-tail that spans its C-lobe, active site, and N-lobe while playing a regulatory role (Komalov, 2015). One component of the active site cleft, AST, regulates nucleotide coordination and may assist in ADP release upon completion of the catalytic cycle (Komalov, 2015). In this way, the AST fragment of GRK5's fully ordered kinase domain C-tail is able to play a nucleotide gate role through the direct interaction with nucleotides in the catalytic cleft (Komalov, 2015).

GRK5 employs several membrane localization mechanisms to position itself in close proximity to its integral membrane G-protein coupled receptors substrates (Komalov, 2015). These include C-terminal prenylation, palmitoylation, pleckstrin homology domain binding to acidic phospholipids and membrane-associated betagamma-subunits, and electrostatic interactions with phospholipids, all of
which rely upon two clusters of basic/hydrophobic residues within the N- and C-terminal regions in order to interact with phospholipids (Komalov, 2015).

After anchoring, GRK5 diffuses across the membrane to reach an activated GPCR (Komalov, 2015). Immediately following phosphorylation of its target receptor, GRK5 dissociates from the membrane via autophosphorylation (Komalov, 2015). However, this entire catalytic process is impeded by calcium-mediated binding of CaM to GRK5, which is then prevented from binding to the membrane (Komalov, 2015). Moreover, CaM redirects GRK5 from the membrane to a different set of cytoplasmic or nuclear substrates. Collectively, knowledge of the crystal structure and function of GRK5 provides insight into the development of novel inhibitors to these molecular targets.

One compound in particular, CCG215022, selectively inhibited both GRK2 and GRK5 in only nanomolar concentrations of this inhibitor (Homan, 2015). Moreover, this inhibitor was much stronger than the already established drug paroxetine while being administered at substantially lower concentrations (Homan, 2015). As shown in Figures 2.2 and 2.3, a 2.4 angstrom crystal structure of the GRK5·CCG215022 complex revealed the active site to which this inhibitor binds (Homan, 2015).
Figure 2.2: The active site of GRK5. A 2.4 angstrom crystal structure of the GRK5·CCG215022 complex revealed the active site to which this inhibitor binds (Homan, 2015). Characteristic to its design, the inhibitor's 2-pyridylmethyl amide side chain resides within the hydrophobic subsite of the kinase's active site (Homan, 2015). Three additional hydrogen bonds, including one with the catalytic lysine, strengthen this binding (Homan, 2015). This complex has a relatively large degree of domain closure and P loop conformation that can interfere with the binding of this kinase to its corresponding GPCR (Homan, 2015).

Figure 2.3: The crystal structure of the GRK5·CCG215022 complex. The GRK5·CCG215022 complex has a relatively large degree of domain closure and P loop conformation that can interfere with the binding of this kinase to its corresponding GPCR (Homan, 2015). Specifically, the inhibitor covers 384 angstroms of the kinase's surface area through the binding of the inhibitor's imidazole ring in the adenine subsite (Homan, 2015). In this way, two hydrogen bonds are formed with the backbone atoms in the hinge of GRK5 (Thr-264 and Met-266) and “its fluorophenyl ring packs into the polyphosphate subsite, a hydrophobic pocket created between the P-loop and the catalytic Lys-215 (Homan, 2015).” Two more hydrogen bonds are formed between the inhibitor's 2-pyrdadylmethyl amide group and the kinase's amide of Phe-197 in a P loop (Homan, 2015). This causes a 2.5 angstrom displacement of Gly-195,196 away from the active site, leaving the side chain of Asp-329 in the large lobe (Homan, 2015). A third hydrogen bond arises between the pyridine nitrogen and the
side chain of Lys-215, and it is believed that these hydrogen bonds collectively occlude the active site of the target kinase, thus, inhibiting its activity (Homan, 2015).

By engaging in hydrophobic interactions between the inhibitor’s 2-pyridylmethyl and the target kinase, 80 angstroms of the kinase’s active site is covered (Homan, 2015). This sufficiently blocks the kinase from binding to its corresponding GPCR, thus alleviating certain cardiac issues (Homan, 2015).

Several peptides that mimic the domains of either GPCRs or GRks are known to selectively inhibit GRks (Baameur, 2014). Sequences that inhibit GRK5 include helices 3, 9, and 10 (alpha 3, alpha 4, and alpha 10) in the RH domain, the N-terminus, the intracellular loop 1 (iL1) of the beta2-adrenergic receptor (beta2AR), and the alpha transducin C-tail (TCT) (Baameur, 2014). Unfortunately, these peptide mimics showed inconsistent selectivity for GRK5 and was not always in accordance with structural predictions (Baameur, 2014). Yet, it was demonstrated that two TAT-tagged peptides, TAT-alpha9-R169A and TAT-TCT, inhibited certain GRK activities including isoproterenol-stimulated GRK phosphorylation of beta2AR as well as isoproterenol and forskolin stimulation of AC activity (Baameur, 2014).

The oftentimes lack of selectivity and poor pharmacokinetic properties behind the development of pharmacologic inhibitors has been an obstacle to designing such drugs (Homan, 2016). Nevertheless, this adversity is being overcome by expanding our understanding of these kinases’ structures and functions, which has aided in the development of novel compounds. High-throughput screening techniques have rapidly identified compounds with favorable pharmacokinetic properties (Homan, 2016). By this means, alternative scaffolds that selectively target GRK subfamilies are currently being developed (Homan, 2016).

2.3) Drugs crystallized to phosphatidylinositol phosphates (PIPs).

Phosphatidylinositol (PI) synthesis and metabolism provides an effective means of regulating a wide range of cellular activities (Weixel, 2005). PI metabolism is controlled by phophatidylinositol
kinases (PIKs), which are responsible for synthesizing phosphatidylinositol phosphates (PIPs) by phosphorylating PIs within the cell membrane (Boura, 2015), and also in a number of subcellular compartments (Weixel, 2005). Following this phosphorylation at the 3, 4, or 5 position of the inositol ring, the subsequent PIPs serve as membrane markers (Boura, 2015). In particular, PI4K synthesizes PI4P, the prominent lipid of the Golgi and trans-Golgi network (Boura, 2015), by phosphorylating PI at the D-4 position of the inositol ring to produce phosphatidylinositol 4-phosphate (PI4P) – a crucial element in regulating traffic along the biosynthetic pathway (Weixel, Kelly). Also existing in the plasma membrane and endosomes, PI4P is the most abundant monophosphoinositide in the eukaryotic cell and serves as the precursor to other PIs (Mejdrava, 2015). Moreover, hydrolysis of PI(4,5)P2 begins a downstream signaling cascade that is an important means of signal transduction (Boura, 2015). This role of PI4P is especially vital to the Golgi of both yeast and humans, making it highly conserved (Boura, 2015).

In fact, there are two families of these conserved PI4Ks found in every eukaryote called types II and III (Boura, 2015). The former is stably associated with the membrane and assists with intracellular trafficking, endosome mobility, the regulation of actin polymerization, and interacts with clathrin adaptor complexes (Boura, 2015; Mejdrava, 2015). Two subtypes of this kinase exist in humans, II-alpha and II-beta, both of which are similarly activated via palmitoylation of their CCPCC motifs (Boura, 2015). The first subtype is most active in humans, producing roughly 50% of PI4P, and playing roles in vesicle and endosomal tracking and Wnt signaling (Boura, 2015).

The latter family, consisting of two subtypes of soluble cytoplasmic PI4Ks, are recruited to the membrane via protein-protein interactions (Boura, 2015; Mejdrava, 2015). Namely, Efr, YPP1/TTC7, and TMEM150 recruit PI4K IIIalpha for PI4P synthesis while the Golgi proteins ACBD3 and GBF1/Arf
recruit PI4K IIIbeta to replenish the Golgi pool of PI4P (Mejdrava, 2015). Both isoforms of types III PI4K function mainly in the cytoplasm by providing the PI4P for the kinase that synthesizes the cytoplasmic pool of PI(4,5)P$_2$ (Boura, 2015). PI4K IIIbeta, in particular, can recruit effector proteins to the membrane while, itself, can be recruited to endosomal and Golgi membranes (Boura, 2015).

Certain viruses have the ability to confiscate host PI4Ks to produce “membrane webs” contributing to its viral replication machinery (Mejdrava, 2015). Because genetic inactivation of type III PI4Ks is lethal to the parasite, PI4Ks are believed to be potential drug targets (Mejdrava, 2015). Existing inhibitors are specific to PI4K IIIalpha; targeting PI4K IIIbeta requires more structural data to better understand its structure-function relationship (Mejdrava, 2015).

Crystal structures of inhibitor and ATP bound to PI4K IIIbeta reveals the inhibitor's mode of action: the inhibitor binds to the ATP binding site where the adenine portion would otherwise bind, thus, preventing the lipid phosphorylation reaction (Mejdrava, 2015). Further, this binding distorts the ATP binding site in a way that Lys$^{564}$ adopts a conformation conducive to hydrogen bonding with the inhibitor's methyl group off an aromatic side chain (Mejdrava, 2015). Docking studies have shown that inhibitors can be designed that specifically target this binding site, and the amide side chain of lysine has proven pivotal because it decreases the enthalpy of the interaction between the enzyme and inhibitor (Mejdrava, 2015).

PI4P IIIalpha was shown to be a necessary host factor for HCV replication (Bianco, 2012). Such pathogens reorient the host's PI4P pathway to supporting their own replication (Bianco, 2012). During this process, the levels of PI4P in the “membranous web” rise in proportion to the drop in levels in the plasma membrane (Bianco, 2012). Fortunately, a promising lead, 4-anilino quinazoline, inhibits PI4P formation by inhibiting PI4 IIIalpha, and therefore, provides an effective route for antiviral therapy.
Crystal structures obtained via small angle x-ray scattering and molecular simulations have revealed the similarity between PI4Ks IIalpha and IIIbeta, and illustrated by Figure 2.4 (Boura, 2015). For instance, the Armadillo repeat domain of III-alpha is identical in both structure and function to the helical domain of III-beta (Boura, 2015). Unique to III-alpha are its three helical domains and large unstructured N-terminus, which tethers to the plasma membrane (Boura, 2015).
Figure 2.4: PI4K Ialpha and IIIbeta. The ribbon structures above depict helical domains of PI4K III-beta and II-alpha, each of which imparts the ability to non-enzymatically recruit certain effector molecules to the cell membrane (Boura, 2015).

Inhibition of both subfamilies of PI4K has implications in various diseases, namely, silencing of the PI4K II-alpha gene disrupts the Wnt signaling cascade potentially leading to carcinogenesis (Boura, 2015). Such overexpression causes increased angiogenesis, and it was found that suppressing this kinase dramatically decreased tumor growth in mice (Boura, 2015). By contrast, PI4K III-beta aids in the replication of hepatitis C virus (HCV) via its contribution to the reorganization of cell membranes to form extensively phosphorylated membranous webs (Boura, 2015). PI4K III-beta is also believed to contribute to the replication of HCV, only to a lesser extent, as siRNA screens have shown genotype dependent arrest of HCV replication (Boura, 2015). However, PI4K III-beta “appears to be indispensable for the replication of viruses from the Picornviridae and Coronaviridae families (Boura, 2015),” as well as for many picornaviruses, which “hijacks PI4K III-beta through the Golgi residing ACBD3 protein (Boura, 2015).” This suggests the utility in targeting PI4K III-beta for anti-viral therapy (Boura, 2015).

Cardioviruses belong to a group of viruses called Picornaviridae and adapt their host’s intracellular membrane architecture to purposes of viral replication (Dorobantu, 2015). A central feature of viral genome replication is the formation of viral replication organelles, which work to recruit certain host molecules to their intracellular membrane to carry out the replication process (Dorobantu, 2015). Certain viruses such as enteroviruses utilize the Golgi-localized PI4K IIIbeta while cardioviruses override the ER localized PI4K IIIalpha to produce PI4P-enriched Ros (Dorobantu, 2015). This dependence on host PI4K was demonstrated by both siRNA silencing experiments and pharmaceutical inhibition (Dorobantu, 2015), and this same dependence has been illustrated in experiments with poliovirus-infected cells (Arita, 2014). One small difference between EMCV and HCV is the protein they use to recruit PI4K IIIalpha: the EMCV’s nonstructural protein 3alpha is required to
recruit PI4K Illalpha while HCV uses NS5B for this purpose (Dorobantu, 2015). Once bound, PI4P assists in the viral genome replication cycle by recruiting oxysterol-binding protein (OSBP) to the membrane, and this protein subsequently delivers cholesterol to ROs (Dorobantu, 2015).

Thus, PI4P lipids and cholesterol are necessary components in the formation of ROs and, subsequently, viral genome replication (Dorobantu, 2015). The extensive overlap between viral genome replication in EMCVs and HCVs indicate functional convergence within this process among different species and may be useful for the rational design of drug inhibitors against multiple types of viruses (Dorobantu, 2015).

Along the anti-bacterial route of treatment, knockdown of PI4K II-alpha remarkably decreases the number of bacteria that invade cells, implying that bacteria require a PI4K II-alpha generated pool of PI4P (Boura, 2015). For instance, siRNA mediated depletion of PI4K II-alpha greatly diminished *Chlamydia trachomatis* formed inclusion bodies and production of bacterial progeny (Boura, 2015). This suggests that PI4K II-alpha inhibitors could serve as useful antibodies (Boura, 2015), and some test compounds are shown in **Figure 2.5**.
Figure 2.5: Antibiotic test compounds that inhibit PI4K. Several derivatives of the known PI4K inhibitor, enviroxime, have shown promise as potential antibiotics due to their higher specificity for PI4K III-beta, which inhibits its activity by binding to the portion of the ATP binding active site (Boura, 2015). Through this inhibition the bacteria are deprived of their PI4P₂ pool necessary for their remodification of the cell membrane into a membranous web as well as the production of progeny (Boura, 2015). The various compounds produced from high throughput screening such as T-OO127-HEV1 and BF738735 are effective against various pathogens, though the development of PI4K inhibitors is in its infancy (Boura, 2015).

The development of PI4K inhibitors is in its infancy, as proposed compounds proved to be neither selective or potent (Boura, 2015). Several derivatives of the known anti-picornavirus compound enviroxime have been developed through high throughput screening (Boura, 2015). One such derivative, T-00127-HEV1, was shown to selectively inhibit PI4K III-beta, but was inactive against HCV (Boura, 2015). Another derivative, BF738735, also selectively inhibited PI4K III-beta and was effective against a range of picornaviruses as well as HCV (Boura, 2015). Structural analyses have revealed that there are other PI4K III-beta inhibitors occupy the binding site for the ATP binding ring (Boura, 2015). This ATP catalytic site is located between the N-terminal and C-terminal lobes and is primarily formed by antiparallel beta-sheets (Klima, 2015). Learning the structure of this catalytic binding site in more detail should help in the future design of inhibitors (Kilma, 2015).

2.4) Drugs crystallized to the GTPase-associated region (GAR) of bacterial ribosomes.

The scant variety of cellular targets for antibiotics has been a notable impediment in the development of novel treatments (Wolf, 2014). One proposal has been to target bacterial ribosomes in order to inhibit protein synthesis (Wolf, 2014). Thiostrepton and microccocin are two highly complex polyheterocyclic molecules that are ribosomally synthesized in several bacterial strains including *Streptomyces, Bacillus, and Micrococcus* (Baumann, 2010).

Thiopeptide antibiotics inhibit bacterial protein synthesis by binding to either the bacteria's ribosome or ribosome-associated factors, and have been identified as a promising candidate against *Mycobacterium tuberculosis* (Baumann, 2010). One ribosomal antibiotic binding site in particular is the GTPase-associated region (GAR), though it has not yet been a target for antibiotic therapy (Wolf, 2014).

Rich in ribosomal proteins (r-proteins) and rRNA, the GAR binds translation factors and
stimulates their GTPase activity (Harms, 2008). Illustrated in Figure 2.6, it contains three main regions necessary for GTPase activation of transcription factors: (1) the sarcin-ricin loop (SRL) located in helix 95 (H95) of the 23S rRNA; (2) the ribosomal stalk, composed of r-proteins L10 and 4-6 copies of L7/L12; and (3) the stalk base, comprising L11 and its binding site on the 23S rRNA – H43 and H44 (Harms, 2008). This 23S rRNA contains a stretch of highly conserved 58 nucleotides from 1051-1108 (Wimberly, 1999). Its structure is noteworthy because certain regions of the stalk, such as L7/L12 and L10, are necessary for GTPase activity of translational factors (Wimberly, 1999).
Figure 2.6: The structure of ribosome's GAR. The GAR rRNA is important in stimulating the GTPase activities of translational factors. However, the thiopeptide antibiotics thiostrepton and microccocin can bind to the ends of the stem loops at A1067 and A1095 and disrupt the movement of charged tRNAs along the ribosomal sites (Wimberly, 1999). Assisting these interactions is the cooperative binding of L11, which increases the binding affinity between the RNA and thiostrepton (Wimberly, 1999).

The only known class of molecules that can bind at this site is thiopeptide antibiotics, which bind in nanomolar concentrations with high affinity but also demonstrate very low aqueous solubility (Wolf, 2014). Thiostrepton and microccocin belong to the thiopeptide family of antibiotics, and both bind to the stalk base, targeting both the r-protein L11 and 23S RNA (Harms, 2008) and inhibit GTP hydrolysis of different translation factors (Wimberly, 1999). Microccocin disrupts translation by stimulating the GTPase activities of EF-G while thiostrepton prevents the EF-G catalyzed movement of the charged tRNA along the ribosomal sites (Harms, 2008). Nevertheless, understanding their binding mode and inhibitory action could allow insights into developing effective inhibitors of ribosomal mediated protein synthesis (Wolf, 2014) as well as into the mechanism of action of thiopeptide antibiotics (Wimberly, 1999).

The GAR RNA secondary structure consists of four helical segments including the terminal stem, the 1067 stem-loop, the 1082 hairpin, and the 1095 stem-loop (Wimberly, 1999). These stems and hairpin are stacked pairwise in the tertiary structure to form two extended helical subdomains (Wimberly, 1999). The complementarity between each subdomain enables the entire GAR RNA to fold into a single compact globular domain (Wimberly, 1999). While the 1067 and 1095 stem-loops of the parallel helical domains are in close contact with one another, their protruding residues support long-range interactions between the two subdomains (Wimberly, 1999).

The ribosomal protein L11 recognizes the RNA's minor groove via the exposed 2' OH moieties, and once bound, its two globular domains facilitate the antibiotic binding to its GAR rRNA target (Wimberly, 1999). Of these, the N-terminal domain, with its smaller surface area, plays a key role in antibiotic binding (Wimberly, 1999). This domain “bridges the interface between the 1067 and 1095 stem-loops, while making only a few specific interactions with the RNA (Wimberly, 1999).” The
thiopeptide antibiotics thiostrepton and microccocin bind to the ends of the stem loops at A1067 and A1095, and the cooperative binding of L11 increases the binding affinity between the RNA and thiostrepton (Wimberly, 1999). Specifically, the antibiotics bind to the cleft between the RNA and the proline-rich helix of the N-terminal domain (Wimberly, 1999), and the crystal structure of thiostrepton (TS) bound to GAR has been analyzed for the purpose of designing new structural classes of compounds (Wolf, 2014).

Crystal structures have shown that these thiopeptides bind to the GAR’s 23S-L11 ribosomal subunit, which “is a loose and large macromolecular interface that is composed of rRNA and protein (Wolf, 2014)” and is located on the “stalk base” of the GAR (Baumann, 2010). In doing so, they block a cooperative binding region formed by the rRNA and a ribosomal protein (Baumann, 2010). The specific binding site has been narrowed down to the nucleobases A1067 and A1095 at the tips of helices 43 and 44 of the 23S rRNA and a proline-rich helix in the L11 proline (Baumann, 2010). In fact, alterations in nucleotides 1067A and 1095 through direct interactions with either thiostrepton or microccocin impedes normal function of the ribosomal A site (Rosendahl, 1994). Figure 2.7 depicts the crystal structure of this binding interaction.
Figure 2.7: Thiopeptide binding to 23S rRNA and L11. Thiopeptide antibiotics primarily target the 23S rRNA complex and the ribosomal L11 protein of the “stable base” of the GTPase-associated region (GAR) (Baumann, 2010). Specifically, they bind to the nucleobases A1067 and A1095 at the tips of helices 43 and 44 of the 23 rRNA in addition to a proline-rich helix in the L11 protein (Baumann, 2010).

Source: Baumann, Sascha. Molecular Determinants of Microbial Resistance to Thiopeptide Antibiotics.

Owing to the large surface area characteristic of dual macromolecule interfaces, understanding the ligand's flexible binding interactions with its target entails observation of optimal binding orientations (Wolf, 2014). However, the three-dimensional structure of these relatively large thiopeptide ligands and their binding conformations with GAR remain unsolved (Wolf, 2014).

To determine these unknowns, an integrated docking and molecular dynamics (MD) approach was used to calculate free energies of binding (Wolf, 2014). This method of MD-refined target structure in conjunction with medium resolution crystal structure allowed for the identification of important structural binding elements within both the GAR's rRNA-protein binding site and the ligand structures (Wolf, 2014).

A crystal structure of TS bound to the 50S ribosomal subunit was observed, though the position of the residues' side chains and TS remain undefined (Wolf, 2014); however, x-ray crystal structures of thiopeptides in the 50S subunit illustrated that these antibiotics bind to a cleft-like cavity between the L11 protein and the 23S rRNA (Baumann, 2010). Further, the molecular recognition of the RNA/protein target complex depends on the local RNA structure, namely, residues A1067 and A1095 (Baumann, 2010). To enhance molecular docking studies, a modified crystal structure of 23S-L11 ribosomal subunit was created whereby the protein L11 and helices H43-H44 of the 23S rRNA were extracted and the voids were filled with the missing valencies and hydrogen atoms (Wolf, 2014). This new structure, called rRNA·L11, showed stark resemblance to its original crystal structure upon relaxation to its local potential energy minimum (Wolf, 2014).

The crystal structure of the 23S-L11 binding site revealed an inverted binding mode where the molecule's tail is inserted into the cleft between the rRNA and the protein, as illustrated in Figure 2.8 (Wolf, 2014). Such an orientation is due to the disparities in shapes between the ligand and receptor.
(Wolf, 2014). A cluster of low bond energy interactions of the ligand to the MD-refined receptor conformation gave insight into experimental binding affinities, providing valuable insight into the location and binding mode of the ligand (Wolf, 2014).
Figure 2.8: Crystal structure of the 23S·L11 binding site. Illustrated is the crystal structure of the modified 23S·L11, rRNA·L11, bound to its ribosomal binding site via inserting its tail into the cleft between the rRNA and the protein (Wolf, 2014). Once a cluster of different binding affinities was observed, the low bond energy interactions were observed as a reference to the future design of ribosomal inhibitors (Wolf, 2014).

In this way, the MD simulations showed native-like binding modes of thiopeptide models to their receptors as well as their docking affinities (Wolf, 2014). The future design of compounds will require the modification of certain residues shown to be important in the TS macrocycle to encourage more optimal binding (Wolf, 2014).

2.5) Employing microcrystallography to study small membrane proteins.

Since membrane proteins comprise a significant portion of cell membranes and are the most common drug targets, it is important to have effective techniques of studying them (Liu, 2014). A recently developed technique, serial microcrystallography (SMX), employs many small, high-diffraction-quality crystals in order to obtain the atomic structures of proteins that are otherwise too small for conventional x-ray crystallography (Gruner, 2015; Hunter, 2014). Since crystals developed in this technique are typically only a few microns in size, a large number are required to obtain the structure of an entire protein (Gruner, 2015). Special considerations of this technique include using low divergent beams so that more x-rays can be fitted in a microbeam, minimizing radiation damage to the crystallized protein, and eliminating unnecessary x-ray background noise (Gruner, 2015). A subtype of this technique, called x-ray free-electron laser (XFEL), uses even smaller crystals than traditional microcrystallography (Liu, 2014). This very small crystal size allows the structure to be obtained at room temperature with minimal radiation damage (Liu, 2014; Ibrahim, 2015). With improved contemporary techniques, the use of a large number of microcrystals offers insight into the atomic structure of small proteins, including cellular membrane components (Gruner, 2015).

2.6) Drugs crystallized to OmpF porins via x-ray microbeams.

The outer membrane of bacteria serve as an important protective barrier against the harsh outside environment (Kefala, 2012). Despite the utility of this protective barrier, the bacteria must
allow for the passage of certain molecules such as nutrients and cell signals (Kefala, 2012). This is accomplished by the various porins that coat the bacterial outer membrane (Kefala, 2012), comprising 60-90% of its surface (Efremov, 2012). All of these porins consist of a common structure of 16-stranded beta-barrels that associate in tightly packed trimers (Kefala, 2012; Efremov, 2012). The beta-strands are connected by short strands, T1-T8, on the “smooth” periplasmic surface and by long loops, L1-L8, on the “rough” extracellular surface (Efremov, 2012). The result is the formation of very stable trimers densely packed into two-dimensional hexagonal crystals (Efremov, 2012). An example is the OmpF porin: a 340 amino acid protein that forms a voltage-gated channel spanning the outer membrane and allows for the diffusion of small polar molecules (Kefala, 2012). The pore dimensions are approximately 7x11 angstrom – wide enough to accommodate molecules smaller than 600 Da – and favors the passage of molecules with a positive charge (Kefala, 2012).

One such bacteria in which OmpF is found is the typhoid causing pathogen Salmonella typhi (Balasubramaniam, 2012). Fortunately, these porins have been exploited in antibiotic therapy via their potent surface antigens or acting as a passage for antibiotics (Balasubramaniam, 2012). Low molecular weight antibiotics are able to pass through porins of the bacterial outer membrane (Tran, 2014). However, bacterial resistance to antibiotics occurs a few different ways: reduction of porins; expression of new, mutated versions of the targeted porins, or upregulation of porins not targeted by antibiotics (Balasubramaniam, 2012). Because the structural and functional properties of E. coli’s porins are well known, it has been a useful model in studying the biophysical and mechanistic parameters of antibiotic uptake through porins (Balasubramaniam, 2012).

It is known that the transport of antibiotics through porins requires appropriate interactions with key residues within the porin, and a mutation in any of these residues can disrupt interactions
between the porin and the antibiotic (Balasubramaniam, 2012). For this reason, the crystal structures
of porins from various bacteria have been analyzed, and a greater understanding of the atomic detail
of porins can aide in the development of vaccines and antibiotics (Balasubramaniam, 2012).

*S. typhi*’s OmpF, illustrated in **Figure 2.9**, closely resembles other general diffusion porins, and
“each monomer barrel has a 16-stranded anti-parallel beta-sheet defining an aqueous channel that
spans the outer membrane (Balasubramaniam, 2012).” The periplasmic end of the barrel has eight
short beta hairpin turns (T1-T8) while the extracellular side has eight long loops (L1-L8), with loop L3
bending into the center of the barrel narrowing the pore opening (Balasubramaniam, 2012; Kefala,
2012). A salt-bridge is formed between adjacent monomers via the N-terminal methionine and C-
terminal tryptophan (Balasubramaniam, 2012).
The sixteen stranded anti-parallel beta-sheets that compose the beta-barrel monomers of *S. typhi*’s OmpF form an aqueous channel that spans the outer membrane (Balasubramaniam, 2012). The monomers depicted in Panel A form a trimer that is stabilized by hydrophilic interactions while 16 beta-sheets and eight loops depicted in Panel B are all involved in maintaining the protein’s structural integrity (Balasubramaniam, 2012; Chaptal, 2016). The trimers are then stacked along the c-axis as jagged columns with interactions between their smooth and rough sides, exhibiting extremely close crystal packing (Chaptal, 2016). This sturdy protein structure is necessary because OmpF is exposed to the harsh conditions of the outside environment (Chaptal, 2016).


Within this beta-barrel structure, loops L4-L8 are closely packed and partially occlude the
barrel entrance (Balasubramaniam, 2012; Efremov, 2012). Loop L2 is the “latching” loop because it participates in the physical connection between the two monomers through an array of hydrophobic interactions, including many hydrogen bonds with residues of the adjacent monomer (Balasubramaniam, 2012).

Conversely, residues 3I, 13L, 39I, 41F, 57W, 76V, and 341F engage in hydrophobic interactions between each monomer and, in doing so, greatly contribute to the trimer's stability (Balasubramaniam, 2012). These hydrophobic interactions, shown in Figure 2.10, are characteristic of other trimeric proteins as well, and responsible for these proteins' robustness in the presence of chaotropic agents and non-phenolic solvents (Balasubramaniam, 2012). The location of such hydrophobic residues rests largely on two belts on the outer surface of the S. typhi OmpF (Balasubramaniam, 2012). At opposite sides of the barrel, each belt is coated with mostly phenylalanine and tyrosine residues oriented with the hydroxyl groups facing the hydrophobic regions on either side of the belt (Balasubramaniam, 2012).
Figure 2.10: Hydrophobic interactions within the trimer. The structural integrity of the trimer is maintained largely via the hydrophobic interactions along the interface between each monomer (Balasubramaniam, 2012).

Crystal structures of *S. typhi* monomers, shown in Figure 2.11, have demonstrated the uniqueness of this porin in that there is a slight asymmetry among the monomers that make up the trimer (Balasubramaniam, 2012). Owing to the different environments around each monomer, various numbers of detergent and solute molecules coalesce around them (Balasubramaniam, 2012). The only contact between the monomers within the pore is from B285G of loop L7 and C144G of turn T4 at either end of the barrel (Balasubramaniam, 2012).
**Figure 2.11: Asymmetric monomers of *S. typhi*.** Because of the slight asymmetries between the three monomers of *S. typhi*’s OmpF trimers, each resides within different chemical environments as depicted by the varying number of detergent and solvent molecules around each monomer (Balasubramaniam, 2012). Panel C illustrates the only point of direct contact within the porin between monomers: B285G of loop L7 and C144G of turn T4 are join either end of the barrel (Balasubramaniam, 2012).

By contrast, the trigonal crystal of *E. coli* OmpF revealed abundant hydrophobic interactions along the bands within the trimer, but any tetragonal crystals showed very different surface topography of the trimer (Balasubramaniam, 2012). Uncovering such discrepancies may prove a valuable tool in understanding the types of molecules able to interact with bacterial porins (Balasubramaniam, 2012). The ramifications of this also extend to the rational design of drugs that will effectively block porins.

One of the less well understood mechanisms by which bacteria become resistant to antibiotics involves decreasing the permeability of the outer membrane (OM) (Ziervogel, 2013). This outer membrane permeability is dependent upon the presence of general diffusion porins, which, as already mentioned, allow for non-specific transport of charged and zwitterionic nutrient molecules (Ziervogel, 2013). An example of reduced OM permeability to mediate antibiotic resistance is when *E. coli* stops the production of two main general-diffusion porins, OmpF and OmpC (Ziervogel, 2013). Consequently, antibiotics are no longer able to pass through the OM and kill bacteria (Ziervogel, 2013). A better understanding of how OM permeability effects antibiotic uptake is necessary for improving antibiotic therapy (Ziervogel, 2013).

Studies of a range of bacterial porin crystal structures including *Klebsiella pneumonia* OmpK36, *E. coli* OmpC, OmpF, and PhoE, and *Salmonella typhi* OmpF have all revealed a homotrimer consisting of a conserved, antiparallel 16-stranded structure (Ziervogel, 2013). Other notable features of the crystal structure include the extracellular loop (L2) contacting the adjacent monomer to stabilize the trimer and a second loop (L3) that folds back into the middle of the barrel to form the “constriction zone”, which limits the permeation of molecules by size (Ziervogel, 2013). Consequence of the
disparate acidities between acid residues on L3 and basic residues on the adjacent beta-barrel wall, a strong transverse electrical field is created with implications in ion permeation and solute transfer (Ziervogel, 2013).

Molecular dynamics (MD) simulation studies allowed by these high resolution crystal structures has shown that the zwitterionic ampicillin molecule stably binds to the OmpF constriction zone (Ziervogel, 2013). However, di-anionic carbenicillin binded to the extracellular side of the OmpF constriction zone on the positively charged pore wall, thus, partially blocking the channel wall (Ziervogel, 2013). In conjunction with single channel recording experiments, these MD simulations of antibiotic permeation suggest that only zwitterionic antibiotics interact specifically and favorably near the OmpF constriction zone (Ziervogel, 2013).

Such binding of zwitterionic antibiotics to the constriction zone paradoxically increases antibiotic permeation, implying that this binding is necessary to facilitate efficient antibiotic transfer across the OmpF porin (Ziervogel, 2013). Although it is plausible that the binding of a compound to the large pore not interfere with ion transfer, it is unknown whether such binding facilitates or impedes antibiotic transfer across the OmpF porin (Ziervogel, 2013).

Crystallization of OmpF to three antibiotics – ampicillin, carbenicillin, and estapenem – and MD simulations showed mixed results among the location to which each antibiotic bound, illustrated in Figure 2.12, and the effect this binding had on ionic current (Ziervogel, 2013). In particular, ampicillin bound perpendicular to the pore axis on the extracellular side of the constriction zone and effectively blocked the ionic current through the OmpF channel (Ziervogel, 2013). By contrast, carbenicillin and estapenem bind parallel to the pore axis near the periplasmic mouth or extracellular loops without blocking ionic current (Ziervogel, 2013).
Figure 2.12: Interactions between antibiotics and OmpF. (Panel A) A 2.0 angstrom resolution crystal structure was obtained of ampicillin bound within the extracellular pore of OmpF (Ziervogel, 2013). Specifically, ampicillin bound approximately 6 angstroms above the OmpF constriction zone within the larger extracellular pore vestibule (Ziervogel, 2013). The orientation of ampicillin is perpendicular to the channel axis, “in a configuration that allows a number of stabilizing interactions with OmpF residues on both sides of the pore (Ziervogel, 2013).” The negatively charged carboxylate group of ampicillin is positioned between two arginine residues (R167, R168) and is also capable of hydrogen bonding with the S125 side chain's hydroxyl group and the backbone's nitrogen groups (Ziervogel, 2013). In addition, water molecules at either terminus aid in this hydrogen bonding (Ziervogel, 2013). The phenyl ring of ampicillin is located in a hydrophobic pocket near several aromatic residues (Y22, Y32, F118) where it participates in pi stacking (Ziervogel, 2013). (Panel B) The crystal structure of the OmpF-carbenicillin complex measures 2.3 angstrom resolution, with the carbenicillin oriented parallel
to the channel axis and 10 angstrom below the constriction zone (Ziervogel, 2013). The phenyl group of carbenicillin faces the extracellular side of the pore near Y14, thus, allowing the carboxylate group to interact with the basic residues (K16, R42, R82) directly under the constriction zone (Ziervogel, 2013). Enhanced stability is provided by hydrogen bonding between a carbenicillin nitrogen moiety and E62 (Ziervogel, 2013). The position of OmpF residues remain unchanged in response to carbenicillin binding, as demonstrated by only an RMSD value of 0.26 angstrom among residues that interact with the compound (Ziervogel, 2013). (Panel C) The crystal structure of the OmpF-ertapenem complex was observed at 1.9 angstrom resolution, with ertapenem bound roughly 17 angstrom above the constriction zone where it adhered closely to the extracellular loops on one side of the OmpF pore (Ziervogel, 2013). Although ertapenem is able to adopt various conformations due to its two planar structures connected by a sulfur atom, it was predicted that the ertapenem carboxylate group hydrogen bonded with R168 (Ziervogel, 2013).

MD simulations from the X-ray structures revealed that these compounds remained bound to its binding site via their hydrogen bond networks as already mentioned (Ziervogel, 2013). A reduction in conductance across the OmpF porin was also observed because the compounds bind to the dielectric regions of the pore, thus, increasing the high ionization energy barrier (Ziervogel, 2013). The binding of ampicillin caused the largest decrease in ion selectivity whereas carbenicillin resulted in greater cation selectivity (Ziervogel, 2013).

Non-equilibrium steered molecular dynamics (SMD) revealed the path each compound takes along the porin, illustrated in Figure 2.13. Both ampicillin and carbenicillin enter the pore via the extracellular mouth of the OmpF channel, after which they make their way to their respective binding sites (Ziervogel, 2013).
Figure 2.13: Antibiotic's permeation path through OmpF. (Panels A,C) Ampicillin enters the extracellular mouth of the OmpF channel near residues E29, G33, T241, K243, and N246 and progresses to residues Y32, D121, S125, R167, and R168 of its extracellular binding site (Ziervogel, 2013). In this way, ampicillin adopts its perpendicular orientation in relation to the channel axis with the positively charged amine near the acidic L3 and the carboxylate group near the basic residues of the beta-barrel wall (Ziervogel, 2013). (Panels B,D) Carbenicillin enters the channel near residues T165, N246, T247, and S248 of the extracellular mouth (Ziervogel, 2013). From there, carbenicillin makes its way to the beta-barrel wall where it interacts with S125, R167, and R168, and remains parallel to the channel axis (Ziervogel, 2013).

Bacteria have a mechanism to shield themselves from antibiotics by downregulating larger porins such as OmpF while increasing the expression of smaller porins such as OmpC (Cama, 2015). In this way, the bacteria filters out the antibiotic during the passage of smaller nutrients and cell signaling molecules. Overcoming this adaptation to antibiotic treatment calls for a deeper understanding of antibiotic permeation across the OM and accumulation within the periplasm (Cama, 2015). Experiments on the antibiotics fluoroquinolone and norfloxacin have demonstrated that both pH and electrical gradient affect this process of antibiotic uptake, and future studies should emphasize both porin structure and environmental conditions (Cama, 2015).

2.7) Conclusion.

Defending the host from bacterial or viral invasion requires a deep understanding of both eukaryotic and prokaryotic cellular structures and functions owing to the diverse array of molecules that are required for the health of the host and the parasite. For instance, pathogens are able to hijack their host's PI membrane lipids in order to support their own membranous webs, constructed to aid in replicating the pathogens' genomes. Likewise, certain prokaryotic structures such as their translational machinery and porins assist in replication and nutrition. The goal of understanding these and other molecules in greater detail as it pertains to drug therapy is to, first, identify them as potential drug targets, and second, to develop inhibitors with high selectivity for the desired targets.

One technique used for this purpose of improving drug design is x-ray crystallography, which creates a three-dimensional structure of target molecules for experimental observation. These images offer intricate architectural observations including the tertiary and quaternary structures of the proteins and offer insight into the forces holding the integrity of these structures. More importantly, as it pertains to drug therapy, the crystal structures enable the visualization of the protein's active site,
which is very relevant to the rational design of highly selective and potent inhibitors. In the case of PI4K, some inhibitors were shown to bind to the ATP binding site, and visual studies have conferred the molecular details of the binding interactions that impart high affinity and selective binding between the active site and the inhibitor.

For other structures such as porins, the crystal structure provides information about the binding interactions that occur as the drug journeys through the membrane pore. These interactions can be exploited to create a compound that either can easily permeate across the membrane or bind to the porin and block the escape of cell signals and the entry of nutrients. As for OmpF, a series of hydrogen bonds hold the antibiotics within the pore opening so that the ionic current across is disrupted along with the flow of other important molecules such as nutrition and cell signals.

Aside from fighting off invasion of pathogens, drug therapy can also be administered to control the regulation of certain molecules required for homeostasis. Sometimes a molecule can be overexpressed if the body's normal means of regulation go awry, resulting in harmful and even fatal conditions. For instance, overexpression of GR5Ks have implications in a number of diseases, most notably, pathogenic cardiac hypertrophy. However, crystal structures have revealed that inhibitors can be developed with the ability to bind to a certain hydrophobic pocket on the kinase, resulting in its downregulation and subsequent alleviation of the ailment.

Thus, the use of x-ray crystallography has enabled the identification of a vast array of cellular and molecular targets for drug therapy as well as the development of highly selective and potent drug compounds. Past studies of test compounds have have been promising, and the utilization of newer techniques such as microcrystallography and XFEL offer a much more refined means of observation. Accumulating even more knowledge on the structure and function of various molecular targets will
translate into even greater potency and specificity in the ration design of drugs.

CHAPTER 3

In-Cell NMR Based Drug Screening

3.1 Introduction:

3.1.1 Nuclear magnetic resonance (NMR) spectroscopy.

Unraveling a protein's structure is a prerequisite for studying its folding, design, and structure-function relationship as well as the design of therapeutics (Carlisle, 2007). One prominent technique used to discern a protein's structure in atomic detail is nuclear magnetic resonance (NMR), and insights about parts of the protein's structure can be made from the Chemical Shift Index (CSI) method (Carlisle, 2007). In this way, NMR studies identify the various chemical shifts, or signals in particular frequency ranges, that constitute the atoms of a compound or biomolecule such as a protein (Schmidt, 2012). Once obtained, the chemical shifts must be assigned to their respective atoms in the protein, a process that uses a set of different multi-dimensional spectra to compare the chemical shifts across peaks (Schmidt, 2012). The chemical shifts are then aligned with their respective atoms, and important insights can be gained about the structure of the protein (Schmidt, 2012).

3.1.2 NMR chemical shifts and resonance assignments.

CSI determines the presence of alpha-helices or beta-sheets based on the direction and extent of chemical shifts from a reference point – typically a random coil of a particular amino acid – to the alpha-proton of the same amino acid (Carlisle, 2007). A chemical shift of 0.10 ppm or greater upfield of the reference point indicates an alpha-helix secondary structure whereas a downfield shift of such intensity signals the presence of a beta-sheet (Carlisle, 2007). This supposition is strengthened by both
the magnitude of the chemical shift and the consistency of these shifts across neighboring amino acids (Carlisle, 2007). However, these reference points, referred to as alpha-proton random coil chemical shift values, of each amino acid fluctuates with changes in both the solution environment (ie. pH, urea concentration, and temperature) and the molecular environment (ie. peptide length, the presence of a tertiary cap, and adjacent amino acids) (Carlisle, 2007). For example, the presence of certain amino acids causes upfield shifts in neighboring alpha-protons, and this shift is more pronounced when the neighboring amino acid is in the N-terminal position (Carlisle, 2007).

Comparisons between a designated chemical shift, or resonance frequency within a given magnetic field, with its random coil counterpart are drawn from the chemical shift library to decipher details of the protein's secondary structure (Simone, 2009). These reference shifts from which comparisons are drawn are pulled from experimental shifts coupled to protein databases (Kosol, 2013). Reference-data stored in these libraries is obtained from random coil shifts using a series of short peptides under a variety of experimental conditions (Schwarzinger, 2001), and the referenced random coil shifts, tabulated in Table 1, are derived from the characteristic chemical shifts of amino acid residues, or nucleic acid bases in the case of nucleic acid polymers (Wishart, 2001). Chemical shifts of several nuclei are then made available from the protein-backbone resonance assignments (Schwarzinger, 2001).
Table 3.1: Reference Random coil chemical shifts defined for each amino acid taken at 25°C, pH 5.

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>$^{1}$HN</th>
<th>$^{15}$N</th>
<th>$^{1}$Hα</th>
<th>$^{13}$Cα</th>
<th>$^{1}$Hβ</th>
<th>$^{13}$Cβ</th>
<th>$^{13}$CO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>8.24</td>
<td>123.8</td>
<td>4.32</td>
<td>52.5</td>
<td>1.39</td>
<td>19.1</td>
<td>177.8</td>
</tr>
<tr>
<td>Cys(r)</td>
<td>8.32</td>
<td>118.8</td>
<td>4.55</td>
<td>58.2</td>
<td>2.93/2.93</td>
<td>28.0</td>
<td>174.6</td>
</tr>
<tr>
<td>Cys(o)</td>
<td>8.43</td>
<td>118.6</td>
<td>4.71</td>
<td>55.4</td>
<td>3.25/2.99</td>
<td>41.1</td>
<td>174.6</td>
</tr>
<tr>
<td>Asp</td>
<td>8.34</td>
<td>120.4</td>
<td>4.64</td>
<td>54.2</td>
<td>2.72/2.65</td>
<td>41.1</td>
<td>176.3</td>
</tr>
<tr>
<td>Glu</td>
<td>8.42</td>
<td>120.2</td>
<td>4.35</td>
<td>56.6</td>
<td>2.06/1.96</td>
<td>29.9</td>
<td>176.6</td>
</tr>
<tr>
<td>Phe</td>
<td>8.30</td>
<td>120.3</td>
<td>4.62</td>
<td>57.7</td>
<td>3.14/3.04</td>
<td>39.6</td>
<td>175.8</td>
</tr>
<tr>
<td>Gly</td>
<td>8.33</td>
<td>108.8</td>
<td>3.96</td>
<td>45.1</td>
<td>—</td>
<td>—</td>
<td>174.9</td>
</tr>
<tr>
<td>His</td>
<td>8.42</td>
<td>118.2</td>
<td>4.73</td>
<td>55.0</td>
<td>3.29/3.16</td>
<td>29.0</td>
<td>174.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>56.3 (pH 9)</td>
<td></td>
<td></td>
<td>30.8 (pH 9)</td>
</tr>
<tr>
<td>Ile</td>
<td>8.00</td>
<td>119.9</td>
<td>4.17</td>
<td>61.1</td>
<td>1.87</td>
<td>38.8</td>
<td>176.4</td>
</tr>
<tr>
<td>Lys</td>
<td>8.29</td>
<td>120.4</td>
<td>4.32</td>
<td>56.2</td>
<td>1.84/1.75</td>
<td>33.1</td>
<td>176.6</td>
</tr>
<tr>
<td>Leu</td>
<td>8.16</td>
<td>121.8</td>
<td>4.34</td>
<td>55.1</td>
<td>1.62/1.62</td>
<td>42.4</td>
<td>177.6</td>
</tr>
<tr>
<td>Met</td>
<td>8.28</td>
<td>119.6</td>
<td>4.48</td>
<td>55.4</td>
<td>2.11/2.01</td>
<td>32.9</td>
<td>176.3</td>
</tr>
<tr>
<td>Asn</td>
<td>8.40</td>
<td>118.7</td>
<td>4.74</td>
<td>53.1</td>
<td>2.83/2.75</td>
<td>38.9</td>
<td>175.2</td>
</tr>
<tr>
<td>Pro</td>
<td>—</td>
<td>—</td>
<td>4.42</td>
<td>63.3 (trans)</td>
<td>2.29/2.94</td>
<td>32.1 (trans)</td>
<td>177.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>62.8 (cis)</td>
<td></td>
<td></td>
<td>34.5 (cis)</td>
</tr>
<tr>
<td>Gln</td>
<td>8.32</td>
<td>119.8</td>
<td>4.34</td>
<td>55.7</td>
<td>2.12/1.99</td>
<td>29.4</td>
<td>176.0</td>
</tr>
<tr>
<td>Arg</td>
<td>8.23</td>
<td>120.5</td>
<td>4.34</td>
<td>56.0</td>
<td>1.86/1.76</td>
<td>30.9</td>
<td>176.3</td>
</tr>
<tr>
<td>Ser</td>
<td>8.31</td>
<td>115.7</td>
<td>4.47</td>
<td>58.3</td>
<td>3.89/3.87</td>
<td>63.8</td>
<td>174.6</td>
</tr>
<tr>
<td>pSer</td>
<td>8.65</td>
<td>115.5</td>
<td>4.60</td>
<td>57.4</td>
<td>4.22/4.13</td>
<td>66.8</td>
<td>174.9</td>
</tr>
<tr>
<td>Thr</td>
<td>8.15</td>
<td>113.6</td>
<td>4.35</td>
<td>61.8</td>
<td>4.24</td>
<td>69.8</td>
<td>174.7</td>
</tr>
<tr>
<td>pThr</td>
<td>8.43</td>
<td>113.4</td>
<td>4.43</td>
<td>61.8</td>
<td>4.66</td>
<td>74.0</td>
<td>175.1</td>
</tr>
<tr>
<td>Val</td>
<td>8.03</td>
<td>119.2</td>
<td>4.12</td>
<td>62.2</td>
<td>2.08</td>
<td>32.9</td>
<td>176.3</td>
</tr>
<tr>
<td>Trp</td>
<td>8.25</td>
<td>121.3</td>
<td>4.66</td>
<td>57.5</td>
<td>3.29/3.27</td>
<td>29.6</td>
<td>176.1</td>
</tr>
<tr>
<td>Tyr</td>
<td>8.12</td>
<td>120.3</td>
<td>4.55</td>
<td>57.9</td>
<td>3.03/2.98</td>
<td>37.8</td>
<td>175.9</td>
</tr>
<tr>
<td>pTyr</td>
<td>8.21</td>
<td>120.1</td>
<td>4.61</td>
<td>57.0</td>
<td>3.15/3.02</td>
<td>38.7</td>
<td>176.7</td>
</tr>
</tbody>
</table>

$^a$ Measured at 25°C, pH 5. Data given in ppm. Note that Cys(r) refers to cysteine and Cys(o) refers to cystine. Data compiled from Refs. 32, 38, and 40.

Resonances are assigned to proteins by observing their chemical shifts across a series of nucleus-variable spectra including $^{15}\text{N}$, $^{1}\text{N}$, $^{1}\text{H}_n$, $^{13}\text{C}_{\text{alpha}}$, $^{13}\text{C}_{\text{beta}}$, $^{1}\text{C}$, and $^{15}\text{N}$ (Kosul, 2013). Some of these nuclei are more sequence dependent ($^{13}\text{CO}$, $^{15}\text{N}$, $^{1}\text{H}^N$) than others ($^{\text{Calpha}}$ and $^{1}\text{H}^{\text{alpha}}$) (Carlisle, 2007), all of which comprise the protein-backbone's amide plane (Wishart, 2001). Resonance assignments have been automated for already reviewed proteins, but in most other cases these determinations are made manually (Schmidt, 2012). Standard protocol of resonance assignments is to first assign the backbone resonance and then expand outward with side chain resonance assignments (Hiller, 2008). Automated approaches assign the backbone and the $^{\text{Cbeta}}$ chemical shifts following triple resonance experiments, and many algorithms use a list of chemical shifts from specific sets of NMR spectra (Schmidt, 2012). One type of algorithm for resonance assignment called FLYA computes a map of a particular protein's expected peaks along with the magnetic transfer pathway from a set of NMR experiments (Schmidt, 2012). Comparisons between this created network and the actual peaks yield information depicting the measured peaks as belonging to certain atoms in the protein (Schmidt, 2012).

Compared to the array of approaches available for obtaining backbone resonance assignments, relatively few automated or manual methods exist for determining side chain resonance assignments (Hiller, 2008). These limited means generally function by correlating the side chains with the backbone nuclei, assuming that the backbone's C-H moieties have strong enough magnetization transfer amplitudes to appear in the same spectrum (Hiller, 2008). This simultaneous sharing of a spectrum can be ensured by either choosing an intermediate time where a majority of the peaks have sufficient transfer or by superimposing data sets measured with different mixing times (Hiller, 2008). These tend to be cumbersome processes of obtaining a sequence-specific resonance assignment of a
protein, and more efficient automated methods are in the making (Hiller, 2008). The goal of these automated techniques is to simultaneously analyze every C-H moiety of the twenty amino acid types and yield precise chemical shifts along with clear sequence-specific resonance assignments (Hiller, 2008).

Sequence specific resonance assignments of proteins can be determined using either the NOESY-based approach or the heteronuclear approach based on triple resonance experiments (Bhavesh, 2001). The latter method yields clearer results with greater efficiency and allows for the study of large proteins (Bhavesh, 2001). Of the nuclei mentioned above, resonance assignments along a protein are typically made using three nuclei ($^1H$, $^{15}N$, and $^{13}C$) in order to obtain their protein-structure (Bhavesh, 2001).

NMR studies on protein-structure are ubiquitously performed using multidimensional, multi-nuclear experiments on doubly ($^{15}N$, $^{13}C$) or triply ($^{15}N$, $^{13}C$, $^2H$) labeled protein samples (Chatterjee, 2002). “This heteronuclear approach for assignment relies on a set of standard three-dimensional experiments such as HNCA, HN(C)OCA, CBCANH, and CBCA(C)NH which employ one-bond magnetization transfers along the polypeptide chain in doubly labeled proteins ($^{13}C$ and $^{15}N$), and in most cases, the magnetization on the amide protons is detected (Bhavesh, 2001).” The resonance assignment of the HN(C)N spectrum is illustrated in Figure 3.1.
Figure 3.1: Resonance assignment of the HN(C)N spectra. The distinctive peak patterns in the different planes of the HN(C)N spectrum are used to assign NMR resonance by means of the magnetization transfer pathway: $H^N(i)\cdot^{15}N(i)(t_1)\cdot CO(i-1)\cdot C^{\alpha lpha}(i-1)\cdot ^{15}N(i,j-1)(t_2)\cdot H^N(i,j-1)(t_3)$; $t_1$-$t_3$ refer to time variables (Chatterjee, 2002). This spectrum shows $H^N$-$^{15}N$ correlations between three consecutive residues $i-1$, $i$, and $i+1$, and the specific combination of positive (+) and negative (−) signs at each plane depends upon the natures of the residues at the $i$ and $i-1$ positions (Chatterjee, 2002). Shown in this figure are the four possible combinations of + and − in the $(F_1,F_3)$ planes and the specific residue-triplets $(i-1, i+1)$ corresponding to those patterns (Chatterjee, 2002). The peak signs are polar opposite between glycine and proline since glycine lacks a beta-carbon while proline lacks an amide proton (Chatterjee, 2002). For this reason, glycine significantly contributes to the signs of the peaks whereas any peaks following a proline at the $i+1$ position will be absent (Chatterjee, 2002).

Identifying these resonance assignments depends upon the dispersion of peaks in its $^1\text{H}-^{15}\text{N}$ spectrum (Chatterjee, 2002). An overlap of peaks results in offsetting + and – intensities, which can occur if the $^{15}\text{N}$ chemical shifts of neighboring residues are too close or if the diagonals with + and – intensities experience close $^{15}\text{N}$ chemical shifts (Chatterjee, 2002).

Chemical shifts are notoriously used in structural biology for predicting regions of secondary structure in native and non-native protein-states, further unraveling complex structures, and for characterizing conformational changes associated with partial folding or binding (Cavalli, 2007). A protein's tertiary structure can also be determined by using the protein's chemical shifts in conjunction with appropriate insight into the interproton distances and the relative orientation of the different nuclei within the protein-structure (Cavalli, 2007).

3.1.3) Protein-structure determination from NMR spectroscopy.

Deciphering protein-structure from NMR analysis entails the accumulation of both the intra- and inter-residual chemical shifts from the peptide-backbone (Xiang, 2016). The shifts associated with certain atoms can then be correlated with either the inter- or intra-residual amide group, and subsequent two-dimensional packing of chemical shifts projects the sequence along the amide-backbone (Xiang, 2016).

The CSI method is widely used as a tool to detect secondary structure elements in folded proteins and of residual structure in unfolded and partly folded proteins (Schwarzinger, 2001) such as the so called intrinsically disordered proteins (IDPs). IDPs play a crucial role in many essential life processes and are found in animals, plants, and to a lesser degree in bacteria and archaea (Kosol, 2013). IDPs and intrinsically disordered regions (IDRs) are highly flexible molecules devoid of any stability in their secondary or tertiary structures (Kosol, 2013). This lack of stability confers
conformational heterogeneity, which in some cases results in transient secondary structures—upending conventional structure-function paradigm (Kosol, 2013).

Despite their structural irregularity, IDPs are microtubule-proteins that serve as important components of many biological processes including differentiation, transcription, regulatory processes, DNA condensation, mRNA processing, and apoptosis (Kosol, 2013). However, IDPs can also facilitate the onset or progression of various diseases in the event that these proteins go awry (Kosol, 2013). Such circumstances make IDPs a suitable target in the design of drugs, the rational design of which requires understanding the structure of IDPs in atomic detail (Kosol, 2013).

The dynamic nature of IDPs makes their structure difficult to study using conventional techniques such as x-ray crystallography or fluorescence microscopy (Kosol, 2013). Fortunately, in-cell NMR spectroscopy, to be discussed in greater detail in the next section, allows for the study of these proteins in solution (Kosol, 2013). The advantage of in-cell NMR over conventional NMR is the greater clarity in signals owing to the slower intramolecular relaxation rates, resulting in narrower peaks with less signal overlap (Kosol, 2013). Advances in this technique have further minimized the signal overlap, thus allowing for even larger IDPs to be studied (Kosol, 2013).

One member of IDPs are the tau proteins, key components of neurofibrillary tangles (NTFs) – a hallmark of Alzheimer's disease (AD) (Harbison, 2012). As indicated by their family, tau proteins lack a stable tertiary structure that could otherwise be isolated and crystallized (Harbison, 2012). Instead, NMR offers a convenient means of studying the details of these proteins' structures in solution (Harbison, 2012).

Chemical shifts change with the local environment of the nucleus, meaning that they depend on the primary, secondary, and tertiary structures of proteins (Kosol, 2013). Because IDPs exist in
solution as constantly changing interchanging conformers, their chemical shifts are simply the average of these conformers (Kosol, 2013). Such studies have conveyed information on the various isoforms of tau and their cellular interactions (Harbison, 2012).

The NMR spectra of tau proteins typically display sharp resonances but lack spectral dispersion, evident of these proteins' unstructured natures (Harbison, 2012). Consequently, the resonance overlap complicates observation of the full length protein (Harbison, 2012). Nevertheless, decoding the residue-specific backbone resonance assignments is a necessary precursor to understanding the detailed structural characterization of disordered proteins (Harbison, 2012). These chemical shifts of full-length tau isoforms provide insight into the protein's secondary structure propensities, which has implications in tau's binding interactions within the cell (Harbison, 2012). Another cellular application of NMR is evaluating structural genomics, a field whose goal is to determine the structures of all native proteins in a given species (Bhavesh, 2001).

The dawn of structural genomics has called for faster means of structure determination and, by default, NMR resonance assignment (Chatterjee, 2002). In-cell NMR spectroscopy presents a powerful means of performing structural and functional studies in live cells including the ability to analyze the native cell states of proteins, nucleic acids, glycans, and lipids in atomic detail (Freedberg, 2014).

3.1.4) In-cell NMR spectroscopy.

In-cell NMR spectroscopy is well suited for studying molecules within living cells because of its ability to harmlessly evaluate the structural and biochemical details of macromolecules in solution (Luchinat, 2015). This technique is especially important in understanding how the thermodynamics of protein folding are affected by the intracellular environment (Luchinat, 2015). One obstacle to in-cell
NMR analyses are the macromolecular crowding resulting from the high concentration of macromolecules in the cytoplasm, and the associated steric repulsions denies proteins access to a large portion of the cytoplasm (Luchinat, 2015). Further, these numerous nonspecific interactions between the target proteins and cytosolic components can interfere with and even prevent NMR analysis (Luchinat, 2015; Majumder, 2014). Specifically, the spectrum is characterized by peak broadening and changing chemical shifts, clouding resonance assignments, and can be altered by degradation of the target protein over time (Majumder, 2014). This ambiguity of chemical shifts necessitates an approach that can more accurately discern the perplexing signals (Majumder, 2014).

STINT-NMR, to be discussed in greater detail later, is able to study the specific high affinity interactions between a target and interactor protein by exploring the Structural INTERactions between proteins within their native environment (Majumder, 2014). This method works by overexpressing a target protein over an unlabeled medium and observing the changes in the in-cell spectrum as the binding partner is also incrementally overexpressed over the same medium (Majumder, 2014). The specific amino acids involved in binding to the target are identified by comparing the spectra of the free and final targets (Majumder, 2014). To account for any promiscuous interactions within the matrix that alter peak intensities, a variable called Single Value Decomposition (SVD) is used (Majumder, 2014).

This mathematical technique has alleviated the overestimation of the number of amino acids involved in binding to a specific target (Majumder, 2014). For example, its ability to follow consistent and specific spectral changes highlight the amino acids involved in the principle interactions, as in those between the IDP Pup and its MPA target (Majumder, 2014). By doing so, SVD STINT-NMR is able to discriminate between the cytosolic protein-protein background signals and the protein-ligand
signals of interest (Majumder, 2014). Despite the background noise that interferes with spectra, this crowded milieu is highly relevant to protein structure and dynamics because the many different interactions between proteins and the intracellular environment are responsible for the quinary interactions of proteins (Luchinat, 2015).

Because the cellular environment can influence the structural and dynamic properties of biological macromolecules, in-cell NMR spectroscopy is ideal for studying the specific factors that direct protein behavior including folding, cofactor-binding, posttranslational modifications, translocation, and degradation (Luchinat, 2015). Cellular environments are extremely dense in macromolecules with scant bulk water and are characterized by transient hydrophobic and hydrophilic interactions (Majumder, 2015). These temporary interactions constitute a fifth level of protein structural organization known as the quinary structure (Majumder, 2015).

Studying a protein's quinary structure requires in-cell techniques such as in-cell NMR spectroscopy due to the thermodynamic fragility of these protein complexes (Majumder, 2015). Experiments are carried out using an $^1$H-$^{15}$N HSQC probe, which is extremely sensitive to changes in protein structure, to collect a two-dimensional spectrum (Majumder, 2015). Two hindrances against studying protein quinary structure via in-cell NMR spectroscopy are the molecular size of cellular complexes, as already discussed, and the intrinsic cytosolic viscosity (Majumder, 2015). A certain technique called $^1$H-$^{15}$N CRINEPT-HMQC-TROSY NMR combines the assets of previous methods that compensate for high molecular weight molecules and short transverse protein relaxation signals in order to probe the quinary interactions of proteins (Majumder, 2015).

Bacterial thioredoxin is an example of a protein that requires a specialized technique to remedy the broad peaks and overlapping signals of conventional in-cell NMR spectroscopy (Majumder,
In-cell $^1$H-$^{15}$N CRINEPT-HMQC-TROSY NMR has demonstrated that Trx binds to a heterogeneous set of cellular components with high specificity, with its surface residues concentrated around the active site meaning that the Trx active site is involved in quinary interactions (Majumder, 2015). However, the absence of these residues does not impede formation of quinary structures, suggesting that other components including substrates and RNA also contribute to quinary structures (Majumder, 2015). Other related applications of in-human-cell NMR spectroscopy is the study of the folding and maturation processes of proteins in live human cells (Luchinat, 2014).

The maturation process of the human protein superoxide dismutase 1 (SOD1) is vital to normal physiologic functioning, though this protein is also linked to neurodegenerative disease (Luchinat, 2014). The aggregation of SOD1 precursors causes neurological damage, and is a consequence of an incomplete maturation process (Luchinat, 2014). In-human-cell NMR studies have revealed the source of this immaturity to be impaired zinc binding that can be overturned with hCCS and copper (Luchinat, 2014). This application demonstrates the utility of in-cell NMR spectroscopy in studying the protein misfolding pathways that underlie many neurodegeneracies (Luchinat, 2014).

A central component of in-cell NMR spectroscopy is the isotopic labeling of NMR-visible biomolecules that stand out in unlabeled, NMR-inactive environments (Freedberg, 2014). In this way, the structural and functional properties of labeled biomolecules can be selectively visualized against a background of the remaining cellular components (Freedberg, 2014). The visibility is made possible through the replacement of a biomolecule’s atomic nuclei with stable, NMR-active isotopes (Freedberg, 2014). The various cell types in which this technique is utilized is illustrated in Figure 3.2.
Figure 3.2: In-cell NMR spectroscopy applied to both prokaryotic and eukaryotic cells. Recombinant proteins can be isotopically labeled in *Escherichia coli* (Panel a), and upon induction of recombinant protein expression, $^{15}$N-, $^{13}$C-, or $^2$H-isotopes are incorporated into the newly synthesized polypeptides.
(Freedberg, 2014). Subsequent overexpression of the protein sample leads to an in-cell NMR spectrum with great clarity and a diminished presence of background labeling artifacts (Freedberg, 2014). This technique has since been applied within eukaryotes including yeast (Panel b), insects, (Panel c), amphibians (Panel e), and humans (Panels d, f, and g) (Freedberg, 2014). A combination of prokaryotic and eukaryotic applications can be implemented by which isotope labeled biomolecules are produced in *E. coli* before being introduced into eukaryotic cells via microinjection (Panel e), active endocytotic transport via the use of cell-penetrating peptide (CPP) tags, or by passive diffusion through pore-forming toxins (Panels f and g) (Freedberg, 2014).

3.1.5) **Studying large proteins with isotopic labeling.**

Despite the many inroads made by in-cell NMR microscopy in understanding the protein-structure and binding interactions for small proteins, larger proteins in solution are significantly more difficult to study owing to the high degree of signal overlap across the many resonances (Skrisovska, 2008). Further, their increased transverse relaxation rates lead to exaggerated line broadening, in turn, causing poor resolutions and sensitivities in $^1$H, $^{13}$C, $^{15}$N dimensions (Skrisovska, 2010). Extending the rate limit is advantageous because it offers a complementary study to x-ray crystallography (Skrisovska, 2010). Two ways by which this hindrance of overlapping NMR signals can be overcome is via isotope labeling of amino acids or by stereo-array isotope labeling of amino acids (Skrisovska, 2008). Both approaches are not fully adequate in that protein resonance assignments cannot be made from the former while the latter is limited to in vitro study (Skrisovska, 2008).

These shortcomings are compensated for by so called segmental isotope labeling methods, which isotopically label the proteins' segments or domains, thus simplifying the structural determination of large proteins (Skrisovska, 2008). Segmental isotope labeling relies upon the natural process of protein-splicing whereby inteins, short for internal segments, catalyze their own excision from a parent protein and subsequently recombine to form a new, native protein (Skrisovska, 2010). The selective recombination of an alpha-thioester and alpha-cysteine at the N- and C-termini is accomplished using prepared inteins engineered to selectively cleave at only one of their termini, though other binding groups have been employed in variations within this technique (Skrisovska, 2010). This same feat can be accomplished with proteolytic leader sequences destined for specific proteases or via endogenous methionyl aminopeptidase treatment targeting Met-Cys (Skrisovska, 2010). Usually, only two protein fragments are created, but in the case of larger proteins more can be
made to allow for the isotopic labeling of the proteins' internal segments (Skrisovska, 2010).

Many insights have been gained from this segmental isotope labeling approach concerning interdomain interactions within multidomain proteins, conformational changes, ligand binding, resonance assignments, and protein-structure (Skrisovska, 2010). Of these, the most common purpose is studying interdomain interactions: it can be discovered whether two or more domains interact with one another by isotopically labeling one domain while leaving the others unlabeled (Skrisovska, 2010). The degree of interaction is then inferred by comparing the isotopically labeled $^{15}$N-$^1$H spectrum with the unlabeled segments’ $^{13}$C-$^1$H spectrum (Skrisovska, 2010). A strong resemblance among the spectra suggests similar protein folding along the C-termini across the different domains meaning that the N- and C-termini of the domains cannot interact with one another (Skrisovska, 2010). These methods of using stable isotope labeling to determine the secondary through quaternary structures of proteins have implications in drug manufacturing (Arbogast, 2016), especially protein-therapeutics – the fastest segment of the pharmaceutical industry (Wang, 2014).

Inferences from protein-structure and dynamics augment drug-screening and design as the target of small molecule drugs (Wang, 2014). Moreover, proteins themselves have also been used as drugs, making up greater than 85% share of the biopharmaceutical market (Wang, 2014). Their high efficacy, specificity, and profitability makes them desirable pharmaceuticals for manufacturing and use, but their relatively large structures are challenging to characterize (Wang, 2014). In-cell NMR serves a useful purpose in characterizing the conformations of protein-therapeutics in formulations at the atomic level (Arbogast, 2016).

3.1.6) Biopharmaceutical analysis via in-cell NMR spectroscopy:

A plethora of clinically approved biopharmaceuticals have been developed over the past two
decades, raising the need to assess the folded structures of protein therapeutics (Arbogast, 2016). The unique precision of in-cell NMR spectroscopy in determining protein-structure at the atomic level through the assignment of spectral signals of individual residues allows for accurate comparisons to be made between different protein-drug species (Arbogast, 2016). Due to the impracticality of applying stable isotope ($^{13}$C, $^{15}$N, $^2$H) enrichment to protein therapeutics, one- and two-dimensional proton-($^1$H) based methods are used instead (Arbogast, 2016). Further, the natural abundance of isotopes can be employed in two-dimensional $^{15}$N and $^{13}$C correlation spectroscopy to better understand protein-therapeutic characteristics (Arbogast, 2016).

Two-dimensional $^1$H-$^{13}$C-methyl and $^1$H$^{-}$-$^{15}$N-amide correlation spectra provide valuable information on the structure of protein-therapeutics and ideally occur at their natural isotopic abundance (Arbogast, 2016). Of these two spectra, the methyl analysis is more advantageous in its natural isotope abundance because the methyl groups are well dispersed throughout primary structures, and their hydrophobic pockets provide insight into protein folding (Arbogast, 2016). Also, the greater natural abundance of $^{13}$C (1.1%) to $^{15}$N (0.37%) facilitates more favorable relaxation properties and greater sensitivity to protein-structure analyses (Arbogast, 2016).

The ultimate aim of these natural abundance isotopes spectra is to gain insight into biopharmaceutical protein-structure by comparing an experimental spectrum against a reference spectrum (Arbogast, 2016). This is accomplished by overlaying the analyte spectrum with the reference spectrum of interest (Arbogast, 2016). Any disparities in peak positions suggest deviant samples, though such interpretations must be verified with resonance assignments (Arbogast, 2016). As mentioned earlier, potential differences in environmental conditions (ie. pH, temperature, or salinity) that could influence results must also be accounted for (Arbogast, 2016). Likewise, a
definitive assessment of spectra that perfectly overlay cannot be settled on without further analysis that explore more subtle divergences (Arbogast, 2016), making abundance isotope spectra a useful tool in conjunction with the many other in-cell NMR techniques used for structural analysis. This assortment of in-cell NMR techniques, which explore in atomic resolution the structural and functional parameters of natural biomolecules and pharmaceuticals alike, are contributing to both drug-screening and the rational design of drugs. One important aspect of developing drugs with high specificity to their targets is an elaborate understanding of protein-ligand interaction.

3.1.7) Using in-cell NMR spectroscopy to study protein-ligand interactions.

NMR is one of the many tools used to study protein-ligand interactions – pivotal encounters for numerous cellular processes – and it poses several advantages over other techniques (Dyachenko, 2010). Among these unique pros are the abilities to determine affinity and specificity, identify protein binding isotopes on the protein and ligand, characterize conformational changes resulting from binding, and quantify turnover of substrates by enzymes (Dyachenko, 2010). Another prominent advantage of being able to perform NMR spectroscopy in the context of the cell is that observations can be made under physiologic conditions (Dalvit, 2009). Either the ligand or the protein may be the molecule of interest in studying these interaction with in-cell NMR spectroscopy, and such studies usually require a more specialized technique called bidimensional (2D) NMR spectroscopy or heteronuclear spectroscopy described previously (Dalvit, 2009).

As discussed earlier, heteronuclear $^1$H-$^{15}$N correlation NMR experiments allows visualization of the $^1$H and $^{15}$N nuclei in a molecule by generating a spectra containing one or more signals per amino acid (Tarrago, 2009). Subsequent signal assignments enables surveillance of binding-induced conformational changes in the protein-backbone (Tarrago, 2009). Finally, upon determination of the
protein's three-dimensional structure, the exact binding regions can be identified (Tarrago, 2009). Likewise, heteronuclear $^1$H-$^{13}$C correlation NMR experiments are highly relevant to observing amino acids with methyl-bearing side chains (Tugarinov, 2005). A very important subtype of these correlation experiments for studying protein-ligand interactions is $^1$H-detected 2D-$[^{15}$N, $^1$H]-HSQC (Bodenhausen, 1980; Kay, 1992).

The cue for protein-ligand binding events in such studies is the change in the chemical shifts of protein signals upon ligand binding (Shuker, 1996). This change in signals is referred to as the chemical shift perturbation (CSP), and it can pinpoint the exact location over which binding occurs on the protein (Shuker, 1996). Aside from changes to the amide-backbone, a more thorough understanding of protein dynamics such as binding affinity and kinetics can also be obtained from CSI experiments (Smrcka, 2010). Owing to the potential lag time between the kinetic constants of the binding event and the altering of chemical shifts of the bound and free states, the concentration of ligand is varied incrementally in certain titration techniques (Reibarkh, 2006).

The fast and slow exchange regimes are two titration methods designed to analyze the protein-ligand binding affinity and kinetics (Reibarkh, 2006). The former assumes a faster exchange between the bound and unbound form than the difference in chemical shifts while the latter assumes the opposite (Reibarkh, 2006). Normally, the chemical shift deviates in correspondence to the bound and unbound protein states, or new shifts appear while others diminish (Reibarkh, 2006). However, the intermediate regime uniquely correlates the bound and free states to their differences in chemical shifts, but the complexity of the resulting mixture of shifts makes for a difficult analysis (Reibarkh, 2006). In general, CSI experiments are not straightforward due to the ambiguity in signals between protein-ligand interactions and conformational changes within the protein induced by ligand binding.
NMR analysis of ligand binding is based upon the difference in NMR parameters, particularly the changes in the nuclear Overhauser effects (NOEs), between the bound and free states of the ligand (Neuhaus, 2000). Small molecules typically exhibit low, if any, NOEs while the larger proteins frequently display very negative NOEs along with highly efficient spin diffusion (Meiboom, 1958). The remarkable decrease in NOE occurs in response to the ligand binding its protein, and represents the larger size of the newly formed complex (Meiboom, 1958). This change in NOE behavior is referred to as the transferred NOE (trNOE), and the shift is proportional to the difference in molecular weight between the ligand and the protein-ligand complex (Meiboom, 1958).

One method that exploits this change in NOE behavior is 2D NOESY, which permits observation of transferred NOEs in order to determine the conformations of ligands bound to proteins (Ni, 1994). This approach relies upon the difference in the NOE buildup rate between the transferred NOEs and those of the free ligand (Ni, 1994). Typically, the trNOE rates range from 50 to 100 ms for binding ligands, compared to 200 to 1,000 ms for their nonbinding counterparts (Ni, 1994). From this information, the positioning of the bound ligand inside the protein's binding pocket can be determined (Ni, 1994). Another type of study called saturation transfer difference (STD) uses information from the intermolecular proton magnetization transfer (Ni, 1994).

3.1.8) STD NMR spectroscopy.

STD NMR is a ligand-based approach to NMR that is designed to detect the interaction of small-ligand molecules with large proteins (Nie, 2011). Inferences are made concerning the receptor-ligand complexes by monitoring the ligand resonances in either the free or bound states (Bhunia, 2012), as shown in Figure 3.3. STD experiments are usually run with a sizeable protein and a much
smaller ligand in large excess to more clearly observe fast binding exchanges and to quantify both the receptor-bound and free ligand (Groves, 2007; Bhunia, 2012). The corresponding 1D $^1$H NMR spectrum will exhibit both the broad peaks characteristic of the protein and narrower peaks belonging to the ligand (Groves, 2007).

The difference in two sets of experiments, on-resonance and off-resonance, yields the STD spectrum, which analyzes the ease of the ligand binding to the protein (Groves, 2007). In this way, the on-resonance experiment provides information on the protein-ligand binding event while the off-resonance experiment is used as a reference (Groves, 2007). When these two spectra are compared, only the signals of the binding molecules are visible (Groves, 2007). STD is the preferred method due to its high specificity, its ability to determine binding epitope and dissociation constant (Meyer, 2003), and perhaps most important, its ability to analyze high molecular weight therapeutic targets (Mayer, 2001).

Directly analyzing the sites of binding interactions makes NMR spectroscopy an invaluable tool to screen and validate lead compounds, and large compound libraries have been screened on the basis of the chemical change described earlier such as trNOESY, NOE pumping, and STD (Bhunia, 2012). STD NMR is an especially reliable NMR-based screening technique, and is widely used to direct receptor-small molecule interactions (Bhunia, 2012). Moreover, advanced techniques in STD NMR spectroscopy are able to detail the internal dynamics of these receptor channels in response to the binding of small molecules (Bhunia, 2012).
Figure 3.3: On-resonance and off-resonance STD experiments. On-resonance 1D $^1$H NMR spectrum uses a frequency range at which no ligand protons happen to resonate while an off-resonance spectrum uses a frequency that differentiates between the protein and the ligand (Bhunia, 2012). The saturation transfer from the protein to the ligand is then determined by subtracting the off-resonance spectrum from the on-resonance spectrum, yielding a new spectrum (Bhunia, 2012). This difference spectrum conveys information about the receptor and ligand interactions, and a subtechnique of STD NMR called group epitope mapping (GEM) allows for ligand-receptor binding interactions to be characterized at the atomic level (Bhunia, 2012).

The key parameter of STD NMR is the saturation time \( t_{\text{sat}} \), characterized by the intermolecular transfer of magnetization from the saturated protein to the ligand (Bhunia, 2012). The extent to which this saturation occurs is controlled by another parameter called the residence time \( t_{\text{res}} \) of the ligand in the binding site of the protein (Bhunia, 2012). The varying longitudinal relaxation times among the protons in the free ligand state provide a useful tool in determining a molecule's GEM (Bhunia, 2012). Protons with short relaxation times display greater STD intensities, and these short domains are sorted out for further analysis (Bhunia, 2012).

An STD pulse sequence, a schematic of which is illustrated in Figure 3.4, is acquired using a series of Gaussian pulses to facilitate the saturation transfer from receptor to ligand (Bhunia, 2012). If necessary, a greater dispersion of the STD effects can be reached using other pulse sequences, such as STD-total correlation spectroscopy (TOCSY) and STD-HSQC (Heteronuclear Single-Quantum Coherence) (Bhunia, 2012).
Figure 3.4: STD pulse sequence. Panel (a) illustrates a greater generic pulse program scheme for STD NMR spectroscopy, with the subtraction of the on- and off-resonances being performed following each scan using phase coupling (Bhunia, 2012). Panel (b) incorporates the various NMR techniques that can enhance this STD pulse sequence acquisition (Bhunia, 2012).

STD can then be used to screen a compound library whereby isotope labeling can be used to track the affinity of each ligand towards its receptor (Bhunia, 2012). Pertaining to the field of drug discovery, this technique can likewise identify drug-targets via analyzing protein-structures and identify drugs through high-throughput screening (Bhunia, 2012).

Owing to their abilities to bind to multiple proteins and their involvement in the assembly of macromolecular arrays, IDPs have been proposed as drug-targets (Cobbert, 2015). Small peptide aptamers (PAs) are employed by a specialized technique called STINT-NMR to study protein-protein structural interactions inside the crowded cytosol (Cobbert, 2015). Some PAs have been shown to have inhibitory activity against bacterial growth, suggesting their plausibility as drug-targets, and site-specific binders can be chosen from a combinatorial library of improved peptide aptamers (CLIPs) (Cobbert, 2015). STINT-NMR is another approach to in-cell NMR spectroscopy that can map the structural interactions that are involved in the formation of protein-protein complexes (Burz, 2010).

3.1.9) STINT-NMR spectroscopy.

STINT-NMR spectroscopy monitors the interaction of two or more sequentially expressed proteins within a cell over defined time intervals, outlined in Figure 3.5, in order to understand the amino acids involved in the surface interactions (Burz, 2010). The interacting surface of the target protein is analysed upon binding of one or more interactor proteins, and the resulting NMR spectrum represents the entire titration of the interactions while conveying structural details of the interacting surfaces (Burz, 2010). The target protein with a known spectrum is first overexpressed on a uniformly labeled medium (U-$^{15}$N) to yield a high-resolution, heteronuclear single quantum coherence ($^{15}$N-HSQC) spectrum of the target protein within a cell (Burz, 2010). Next, the growth medium is changed as the unlabeled interactor protein shifts the chemical environment of the target protein's residues,
and their corresponding changes in bonding interactions is documented by the HSQC spectrum (Burz, 2010). These atomic details provide information about the molecular interactions that regulate physiologic processes (Burz, 2010).
Figure 3.5: STINT-NMR spectroscopy flowchart. Samples are obtained at different time points for STINT-NMR analysis (Burz, 2010). Both the target protein and the three interactor proteins are overexpressed at 1 h, 2.5 h, and 4 h (Burz, 2010).

STINT-NMR spectroscopy is highly sensitive to various intracellular protein-protein interactions and their ensuing structural changes (Burz, 2010). For instance, conformational changes due to posttranslational modifications (PTMs) can be studied by making PTMs on inducible plasmids in a process that is sometimes called in-cell biochemistry (Burz, 2010). Another application of STINT-NMR is high-throughput drug-screening by a related technique called SMILI-NMR (Screening of small Molecule Interior Library by In-cell NMR), which rapidly screens small molecule libraries for any molecules with drug-like capabilities. This approach is a hybrid technique of HTS methods and atomic resolution structural studies in that it is capable of screening thousands of protein complexes per day in atomic resolution (Xie, 2009).

SMILI-NMR spectroscopy relies on the formation of a ternary structure in which one of the constituent proteins is uniformly [U-15N] labeled with an NMR active nuclei (Burz, 2010). Any changes in structure to this complex induced from the binding of small drug-like molecules are detected by the in-cell NMR spectrum of the labeled protein (Burz, 2010). This combination of the in-cell nature of the assay with a high-resolution NMR readout offers a unique approach to screen small drug-like molecules, thus providing useful information on potential drugs (Burz, 2010). The efficacy of this drug screening system is demonstrated by the interactions of the FKBP and FRB immunoproteins (Xie, 2009).

Conforming to standard STINT-NMR procedure, [U-15N]-FKBP and unlabeled FRB were each overexpressed in the same cells, and an HSQC NMR spectrum was obtained (Xie, 2009). The appearance of signals on the spectrum upon overexpressing unlabeled FRB to the [U-15N]-FKBP containing medium indicates the formation of the complex (Xie, 2009). A ternary complex is then formed when these cells are titrated with the immunosuppressant drug rapamycin (Xie, 2009). The
crystal structure of this FKBP-rapamycin-FRB ternary complex, illustrated in Figure 3.6, details the limited FKBP-FRB interaction surface (Xie, 2009).
Figure 3.6: FKBP-rapamycin-FRB ternary complex. Panel (A) illustrates the possible binding regions on FKBP, the structure of the small immunosuppressant drug rapamycin, and the FRB backbone structure (Xie, 2009). Panel (B) highlights the changes in chemical shifts in 32 of 107 FKBP residues resulting from the addition of rapamycin (Xie, 2009).

The addition of rapamycin caused noticeable changes in the HSQC spectrum, including alterations in the chemical shifts in 32 out of the 107 residues in FKBP, and these deviations indicate the assembly of the ternary FKBP-rapamycin-FRB complex (Xie, 2009). Likewise, addition of another immunosuppressant, ascomycin, resulted in ternary complex formation with similar changes in the NMR spectrum (Xie, 2009). It should be noted that the binding of both proteins are necessary to produce a spectrum upon drug binding (Xie, 2009). A chemical library is then screened to find compounds capable of binding the biological target (Xie, 2009). Using a dipeptide library is ideal because cellular transport systems easily take up dipeptides, meaning that they can readily meet their protein target invivo (Xie, 2009). Any changes in the NMR spectrum of the protein complex upon introduction of a compound indicates binding interactions, and the residues involved can be specified (Xie, 2009). In this way, SMILI-NMR can be sued to screen small libraries of compounds for potential drugs, and serves as another useful tool for drug-screening purposes (Xie, 2009).

3.2) Prokaryotic applications:

In-cell NMR spectroscopy serves a useful purpose in studying the structures of the various proteins across bacterial species that contribute to their pathogenicities. The aim of such structural studies is to better understand the specific binding interactions that occur on certain regions of the proteins, as well as to unravel the downstream signaling that results from these binding interactions. Finally, once the atomic details of the protein targets are adequately identified, appropriate therapeutics can be designed that will inhibit the proteins activity and consequently stymie the pathogenicity of the bacteria.

3.2.1) Proteins.

Proteins are essential to many cellular processes (Shi, 2015), and one field, called structural
genomics, aims to determine the protein structures throughout the entire genome to gain insights into the mechanisms of protein folding and function (Leonor, 2003). Structural determination via in-cell NMR offers a reliable means of screening these many proteins’ structures in high-resolution, and the in-vitro solution NMR structures of many prokaryotic proteins have already been identified (Leonor, 2003). In turn, this knowledge can be used to identify potential drug targets against certain bacteria such as the tuberculosis causing *Mycobacterium tuberculosis*.

### 3.2.1.1) In-vitro NMR solution structures of prokaryotic proteins.

One important protein of study is the *mycobacterium tuberculosis* resuscitation promoting factor Rpf, which enables the bacteria to emerge from dormancy and proliferate (Maione, 2015). Its peptidoglycan hydrolase activity accomplishes this feat by altering the mechanical properties of the cell wall to facilitate cell division and the release of anti-dormancy factors (Maione, 2015). Understanding these proteins’ internal dynamics is highly relevant to its mechanisms of function including protein-protein and protein-ligand interactions (Maione, 2015). Thus, differences in dynamics between the five homologues (RpfA-E) of this family can confer information about the functional properties of resuscitation factors (Maione, 2015).

**In-vitro solution NMR spectroscopy revealed the structure of the catalytic domain of RpfC (RpfCc) as a globular fold of 69 amino acids extending from Trp9 to Ala78 containing two disordered N and C terminal tails (Maione, 2015).** This domain also presents a helix bundle, shown in Figure 3.7, comprised of four α-helices joined by flexible loops (Maione, 2015).
Figure 3.7: In-vitro NMR solution structure of RpfC. The structure of the catalytic domain of RpfC (RpfC), as determined by solution NMR spectroscopy, consists of a 69 amino acid globular fold and a helix bundle of four joined α-helices (Maione, 2015). The above illustration are taken from four different angles, Panels (A) - (D) (Maione, 2015).

This helix contains a catalytic glutamate residue (Glu16), and the presence of a conserved hydrophobic pocket near the C-terminus suggests that such an environment arouses the catalytic ability of all members of this enzyme family (Maione, 2015). Likewise, most of the backbone along the catalytic domain is also conserved, though with some variations creating differences in specificity (Maione, 2015). Adjacent to the RpfCc catalytic site lies a disulfide bond between two cysteine residues on the protein surface, where it serves as a modulator of protein function by inducing conformational changes in the hydrophobic catalytic cleft (Maione, 2015). The atomic details gained from in-vitro solution NMR spectroscopy concerning this conserved, hydrophobic catalytic pocket present a potential therapeutic target to prevent the anti-dormancy activities of these Rpf proliferation factors and, by extension, protect against tuberculosis (Maione, 2015).

In-vitro solution NMR spectroscopy has also been useful studying the interactions within the general secretion (sec) pathway of Gram-negative Eubacteria such as *Escherichia coli* (Grady, 2012). This system is necessary to transport many proteins of the cytoplasmic membrane, periplasm, and outer membrane from their ribosomal origin (Grady, 2012). A signal recognition particle (SRP) and a SecA ATPase are two proteins required for bacterial translocation, the former of which recognizes the SecYEG channel while the latter powers the protein through (Grady, 2012). Understanding the nature of these interactions could broaden our understanding of an important component of protein synthesis (Grady, 2012).

SecA has a conserved promotor fold made up of six domains including two nucleotide-binding domains (NBD-I and NBD-II), a preprotein crosslinking domain (PPXD), the helical scaffold domain (HSD), the helical wing domain (HWD), and a carboxyl-terminal linker domain (CTL) (Grady, 2012). Identifying the regions on this protein that interact with signal peptides is the subject of great interest.
and has been explored by forming an NMR structure of the *E. coli* SecA (Grady, 2012).

With the help of Forster resonance energy transfer (FRET), five amino acid residues have been shown to be indispensable for the SecA signal recognition step (Grady, 2012). One of these residues, Ser226, forms a stem connecting the PPX2 and NBD-I domains, while the other four residues, Val310, Ile789, Glu806, and Phe808, belong to the PPXD domain (Grady, 2012). The binding surface recognizes both the aminoterminal basic and hydrophobic core regions of signal peptides, and a few of these residues (Ser226, Ile789, and Glu806) play dual roles in signal recognition and translocation of the preprotein into the protein-conducting channels (Grady, 2012). These atomic details present potential drug targets for disrupting the bacterial translocation process.

The pathogen *Staphylococcus aureus* relies on a cell-density sensing mechanism called quorum sensing QS for virulence, and this process is controlled by autoinducing peptides (AIPs) and their corresponding transmembrane AgrC receptors (Tal-Gan, 2016). Upon reaching a sufficiently high concentration, the AIP signal binds to and activates the extracellular sensor domain of the AgrC receptor, causing a signal cascade to initiate the transcription of QS genes (Tal-Gan, 2016). This makes the AIP-AgrC binding event central to *S. aureus* virulence, meaning that attenuating this interaction could impede its virulence (Tal-Gan, 2016). The synthesis of AIP mimetics to block native AIP binding has been proposed as a possible therapeutic (Tal-Gan, 2016).

The development of AIP mimetics is complicated by the wide variation in AIP ligands across Staphyloccocal species, with four different groups present in *S. aureus* alone (Tal-Gan, 2016). These individual AIPs not only encourage QS within their own species, but function as natural inhibitors of QS other bacterial species nearby (Tal-Gan, 2016). Knowledge of the different AIPs could aide in the development of mimetics that specifically target *S. aureus* in a mixed species population. Solution
NMR spectroscopy has determined the structures of four native AIP signals in *S. aureus* as well as several synthetic AIP analogues (Tal-Gan, 2016).

The goal of these structural studies is to determine motifs that activate or inhibit AgrC receptors and then characterize their binding interactions (Tal-Gan, 2016). These include two structural motifs, in particular, that modulate the activity of the AgrC receptors: a hydrophobic “knob” is essential for binding interactions while a hydrophobic “anchor” on the N-terminus activates the receptor (Tal-Gan, 2016). Absent of this N-terminal “anchor”, the hydrophobic “knob” works to inhibit the receptor activity (Tal-Gan, 2016). Acquiring an understanding of the specific residues constituting these motifs will aid in the design of therapeutic mimetics that can prevent QS in *S. aureus* (Tal-Gan, 2016).

The pili of *Streptococcus pneumonia* contribute to this pathogen’s virulence by enabling it to adhere to the respiratory tract’s epithelium and helping it colonize the nasopharynx (Gentile, 2011). The three proteins that constitute the pneumococcal pilus are named RrgA, RrgB, and RrgC, and their polymerization via the three sortase enzymes SrtC1, SrtC2, and SrtC3 creates enough rigidity to protect the pili from proteolytic cleavage (Gentile, 2011). Moreover, the four domains of RrgB, named D1, D2, D3, and D4 and shown in Figure 3.8, are covalently linked by intramolecular isotopic bonds between Lys and Asn residues, further contributing to the pili’s stability (Gentile, 2011).
Figure 3.8: Schematic structure of pili. The three proteins that constitute the pneumococcal pili are joined by intermolecular isotopic bonds while the four domains within each protein are similarly connected via intramolecular isotopic bonds (Gentile, 2011). This covalent bonding network offers structural stability and protection from proteolytic cleavage (Gentile, 2011).

NMR structure-analysis has shown that Lys-183 of D1 is responsible for the intermolecular isopeptide bond formation during pilus polymerization (Gentile, 2011). An aim of studying these three pilus proteins is to develop a protein-vaccine against these harmful respiratory pathogens (Gentile, 2011).

Three of the four RrgB domains protect against bacteremia, especially D1 and D4, meaning that immunization with D1 and to a lesser extent D4 will induce an antibody response that will protect against the pathogen (Gentile, 2011). NMR spectroscopy reveals that the in-vitro solution structure of D1, illustrated in Figure 3.9, contains an IgG-like β sandwich fold and secondary structure elements including seven parallel and antiparallel β-strands arranged in two sheets (Gentile, 2011). These strands are held together via hydrophobic interactions and a salt bridge between Lys-41 and Glu-143 (Gentile, 2011). The absence of any intramolecular isopeptide bonds in D1 offers structural flexibility that allows this domain to play a role in the specific antigen-antibody recognition process, which would explain its superior protection capability (Gentile, 2011). It has been proposed that this structural flexibility enables this domain under conformational changes to adapt its structure to interact more strongly with antibodies and increase the affinity and specificity of the antigen-antibody recognition process (Gentile, 2011). An understanding of D1’s involvement in antigen-antibody recognition coupled to the overall structural features of pilus proteins is necessary to rationally design protein-based vaccines to protect against pneumococcal disease (Gentile, 2011).
Figure 3.9: In-vitro NMR solution structure of D1. Panel (A) presents the secondary structure of D1, which consists mostly of β-sheets but also contains an α-helix and a β-hairpin (Gentile, 2011). The topology diagram in Panel (B) depicts the outlay of the secondary structure elements (Gentile, 2011). Source: Gentile, M.A. (2011) Structural and functional characterization of the Streptococcus pneumoniae RrgB pilus backbone D1 domain. J. Biol. Chem., 286(16), 14588-97.
3.2.1.2) In-cell NMR studies of prokaryotic proteins.

The high concentration of macromolecules in the cytoplasm causes an excluded volume of proteins due to molecular crowding (Luchinat, 2016). This crowding effect encourages protein folding as demonstrated by in-cell NMR studies of FlgM, an IDP from *Salmonella enterica*, in the cytoplasm of *E. coli* (Luchinat, 2016). Its folding rendered the amide crosspeaks of the C-terminal part of the protein undetectable in the NMR spectra, while those at the N-terminus remained unfolded and were therefore detectable (Luchinat, 2016). The spectrum resembled that of FlgM interacting with the transcription factor $\sigma^{28}$ in vitro, as the C-terminus exchanges with a folded conformation bound to $\sigma^{28}$ (Luchinat, 2016). However, measurements of the amide hydrogen exchange rates showed that protein disorder does not exist in the crowded cellular environment (Luchinat, 2016).

This crowded environment also causes broadened resonance assignments owing to the slower tumbling of the protein than occurs in aqueous buffers (Luchinat, 2016). Broadening occurs to a lesser extent in IDPs than in folded proteins such as the globular proteins, which tumble too slowly for detection via in-cell NMR (Luchinat, 2016). For instance, in-cell NMR of a fusion construct of folded ubiquitin and IDP $\alpha$-synuclein is only able to detect the unfolded backbone resonances of $\alpha$-synuclein in the bacterial cytosplasm (Luchinat, 2016). The slower tumbling and consequent line broadening are due not only to the increased viscosity and excluded volume effects, but more so from weak interactions with cytoplasmic components (Luchinat, 2016). As discussed earlier, the diversity of protein-environment interactions is responsible for the quinary structure of proteins (Luchinat, 2016). Moreover, interactions between proteins and their intracellular environment are intricately involved in many protein processes including initial folding, cofactor binding, posttranslational modifications, translocation and degradation (Luchinat, 2016).
3.2.2) Protein-protein interactions.

Underlying nearly all biologic processes are binding interactions between proteins, and these protein-protein interactions occur with a high degree of specificity (Ames, 2016). Studying these binding interactions in atomic detail is relevant to understanding the mechanisms by which various cellular processes occur and designing protein-specific drugs.

3.2.2.1) In-vitro NMR solution structures of protein-protein interactions.

*Vibrio cholerae* is responsible for causing the small intestine cholera, during which *V. cholerae* genes code for virulence factors including proteases (Edwin, 2015). These enzymes cause damage by degrading tissue barriers and cellular matrix components (Edwin, 2015). In particular, the metalloprotease PrtV belongs to the M6 peptidase family and consists of four domain types: the N-terminal domain, the catalytic active M6 domain, two polycystic kidney disease domains, and the PKD2 domain (Edwin, 2015). PKD domains are common among prokaryotes and eukaryotes and, as extracellular parts of proteins, are involved in protein-protein interactions (Edwin, 2015). Although previously a point of speculation, NMR structures of the N-terminal domain have shown that this domain may contribute to PrtV function by providing a binding region for membrane proteins (Edwin, 2015).

The $^1$H-$^{15}$N HSQC NMR spectrum of the N-terminal domain indicates a partially folded disordered protein (Edwin, 2015), and the NMR structure is illustrated in Figure 3.10. The random coil chemical shifts of non-uniform intensities of the first eight and last 32 residues suggest the region is disordered, while the highly dispersed chemical shifts with uniform intensities of the remaining 39 middle residues indicate a folded core structure (Edwin, 2015). The core’s secondary structure consists of two helices connected by a loop region, and these helices are held together by
hydrophobic interactions between conserved amino acids (Edwin, 2015).
Figure 3.10: In-vitro NMR solution structure of the PrtV N-terminal domain. Panel (A) illustrates that the N-terminal domain consists of a folded core in the midst of a disordered structure, and Panel (B) depicts the two helices and one loop that comprise the core's secondary structure (Edwin, 2015). Panel (C) demonstrates the variance in electrostatic potential among the surfaces presented in Panel (B) (Edwin, 2015). The loop carries a negative charge while the N- and C-terminal surfaces are positively charged (Edwin, 2015).

Although the function of the N-terminal domain is unknown, its NMR structure demonstrated a highly conserved sequence among members of the genus Vibrio (Edwin, 2015). These conserved and positively charged lysine residues may play a role in binding membrane lipoproteins and glycoproteins (Edwin, 2015).

3.2.2.2) In-cell NMR studies of protein-protein interactions.

In-cell NMR spectroscopy is unique from in-vitro techniques such as solution NMR spectroscopy because the protein structures and interactions are determined in the context of the various other molecules that constitute the cellular milieu (Sakakibaro, 2009). Further, these interactions between proteins and other cellular molecules tend to be specific, making the atomic details and consequences of these interactions the subject of much interest (Sakakibaro, 2009). So far, in-cell NMR spectroscopy has been used to study protein-protein interactions inside E. coli, as well as the IDPs discussed in the introduction (Sakakibaro, 2009). Sakakibaro and colleagues have used this technique to characterize the 66 amino acid heavy-metal binding protein TTHA1718 from Thermus thermophilus HBB overexpressed in E. coli cells (Sakakibaro, 2009).

Backbone resonance assignments were made using six three-dimensional triple-resonance NMR spectra (Sakakibaro, 2009). Combining this 3D data revealed the expected backbone resonances with the exception of the three residues of the metal-binding loop, and these backbone resonance assignments were compared with the in vitro NMR solution structure (Sakakibaro, 2009). Illustrated in Figure 3.11, the NMR solution structure deviated from the structure determined in vivo at the residues adjacent to the region with missing $^1$H-$^{15}$N correlation peaks (Sakakibaro, 2009).
Figure 3.11: NMR solution structure of TTHA1718. Panels (a) through (e) represent the NMR solution structure of TTHA1718 that was used as a comparison to the structure determined by in-cell NMR spectroscopy (Sakakibaro, 2009).

Using methyl protonation at methionine residues as an in-cell NMR probe, 31 out of 40 side-chain methyl $^1$H and $^{13}$C resonances were assigned on TTHA1718, with 148 NOEs involved in methyl groups assigned in total (Sakakibaro, 2009). The results of this study are illustrated in Figure 3.12, and three-dimensional structural studies revealed the dynamic nature of the heavy-metal binding loop, which changes in response to interactions with metal ions in the E. coli cytosol (Sakakibaro, 2009).

Having successfully determined in high resolution 3D structure of the TTHA1718 protein in vivo, the next step was to prove the feasibility of analyzing larger proteins using this technique (Sakakibaro, 2009). This feat was accomplished by identifying the sequential backbone connectivities and sequential $^{1}H$-$^{1}H$ NOEs of rat calmodulin expressed in E. coli (Sakakibaro, 2009). These results suggest a useful new means of investigating protein conformations at atomic resolution and how they are influenced by both molecular and ionic interactions within the crowded cytosol (Sakakibaro, 2009).
Figure 3.12: NOE compilation of TTHA1718 in live *E. coli* cells. Panel (a) indicates the assignments of methyl group residues Ala, Leu, and Val from the methyl regions of a selectively methyl-protonated sample (Sakakibaro, 2009). Panel (b) displays the $^{13}$C-$^{13}$C cross-sections corresponding to the $^1$H frequencies of the methyl groups from the spectrum (Sakakibaro, 2009). Panel (c) shows the $^1$H-$^1$H cross-sections corresponding to the $^{15}$N frequencies of the backbone amide groups from the spectrum (Sakakibaro, 2009). Panel (d) illustrates the topology of the β-sheet structure in TTHA1718 (Sakakibaro, 2009).

One application of STINT-NMR in prokaryotes is to study the interactions of the ubiquitin-like protein Pup with a certain proteasome complex consisting of Mpa – a mycobacterium proteasome ATPase – and a core particle (CP) (Maldonado, 2013). This proteasome complex is responsible for imparting resistance of *Mycobacterium tuberculosis* against the immune system's reactive nitrogen intermediates (Maldonado, 2013), and thus causing a lethal infection in a mammalian host (Wang, 2010). The proteasome CP is composed of α- and β-subunits, which form four homo-oligomeric 7-member rings (α7β7β7α7). The β-subunits are sandwiched between the α-subunits forming a gated channel that allows entry of unfolded substrates and exit of cleaved particles from the proteolytic chamber (Maldonado, 2013). Structural studies have proposed that Mpa closely interacts with the proteasome CP and augments the binding, unfolding, and translocation processes of the proteasome complex (Maldonado, 2013). After searching an Mtb genomic library, it was found that a ubiquitin-like molecule, Pup, binds Mpa (Maldonado, 2013). This 64 amino acid protein tags mycobacterium proteins for proteolytic degradation, and these so-called pupylated proteins interact with Mpa (Maldonado, 2013).

The interactions between Mtb proteasome and Pup can be examined by overexpressing the α- and β-subunits of the Mtb proteasome core particle in *E. coli* (Maldonado, 2013). In particular, the C-terminal Pup-GGQ glutamine interacts with the Mpa-proteasome complex, and this interaction surface has been characterized using STINT-NMR spectroscopy (Maldonado, 2013). The peak broadening associated with overexpression of Pup-GGQ in the presence of Mpa suggests that several regions of Pup-GGQ interact with Mpa (Maldonado, 2013), illustrated in Figure 3.13. These interacting regions include most of the C-terminal region, part of the α-helix, and part of the N-terminal region of Pup-GGQ.
Figure 3.13: Pup-GGQ interacting with Mpa. Panels (A) and (B) overlay Pup-GGQ overexpression spectrum with the Pup-GGQ and Mpa overexpression spectrum at 4 h and 16 h respectively (Maldonado, 2013). The binding interactions take place across most of the C-terminal region including residues D53, F54, V55, A57, V59, K61, and Q64; part of the α-helix including residues L39, F42, I43, D44, D45, L47, E49, and A51; and, part of the N-terminal region such as the K7 residues (Maldonado, 2013).

Source: Maldonado, A.Y. (2013) Fate of Pup inside the Mycobacterium proteasome studied by in-cell NMR. PloS One, 8(9), e74576.
This binding demonstrates the necessity of the interactions between Mpa and the proteasome CP in regulating proteasomal degradation (Maldonado, 2013). Also, drawing the differences between how the mammalian and prokaryotic proteasome degradation systems work will assist in the therapeutic design of Mycobacterium-specific proteasome inhibitors as anti-tuberculosis drugs (Wang, 2010).

3.2.3) Protein-drug interactions.

Another type of proteolytic enzymes common among prokaryotes and eukaryotes that contributes to the pathogenicity of Clostridium difficile is the zinc-containing metalloprotease (Rubino, 2015). A special subset of metalloproteases produced by human pathogenic microorganisms are called zincins, which contain a consensus amino acid sequence HExxH (Rubino, 2015). Two histidine residues act as zinc ligands while glutamate functions as a catalytic base (Rubino, 2015). This gram-positive bacterium is a major cause of antibiotic-resistant diarrhea worldwide (Rubino, 2015). Virulence targeting factors are believed to be promising therapeutic strategies and have been the focus of researching new targets for drug design (Rubino, 2015).

One potential target is the zincin Zmp1, which selectively cleaves two adhesins named CD2831 and CD3246 (Rubino, 2015). The in-cell NMR solution structures of Zmp1 in the substrate-free and substrate-bound states has provided information on substrate recognition and specificity (Rubino, 2015). Likewise, testing this protease against a large number of peptides has elucidated the mechanism by which C. difficile interacts with its human host as well as potential protease inhibitors against this pathogen (Rubino, 2015).

The three-dimensional solution structure of Zmp1, shown in Figure 3.14, is composed of a nine-helix bundle over a four mixed-stranded β-sheet and a long loop, which itself consists of three short helices (Rubino, 2015). The aromatic and aliphatic residues of the core engage in hydrophobic
interactions, and stacking contacts are made within the side chains of the helix bundle (Rubino, 2015).

For example, hydrophobic interactions occur between the aromatic rings of Phe191 and Tyr195, both of α7, and the aliphatic side chain of Ile64 (Rubino, 2015). Similarly, stacking contacts are made between Tyr194 of α7 and Tyr212 and Tyr214 of α9 (Rubino, 2015).
Figure 3.14: In-vitro NMR solution structure of Zmp1. Panel (A) and (B) illustrate the secondary structure characteristics of the protease Zmp1 (Rubino, 2015). These elements include nine α-helices packed against four mixed-stranded β-sheets and a long loop composed of three short helices (Rubino, 2015). Panel (C) shows the amino acids involved in binding zinc at the active site (Rubino, 2015).

The NMR structure has also revealed that the active site contains conserved residues for substrate recognition, and that the regions involved in substrate recognition have structural flexibility (Rubino, 2015). The dynamic changes in the binding pocket are believed to contribute to structural specificity (Rubino, 2015). It is the binding of the substrate that then organizes the conserved residues of the enzyme appropriately for proteolytic cleavage to occur (Rubino, 2015). Two other regions, identified by heteronuclear NMR, are not involved in substrate binding or cleavage but may play a role in substrate specificity (Rubino, 2015). Developing inhibitors that directly block the active site or cause allosteric inhibition by binding peripheral regions may impede Zmp1 activities, and in the process, stymie the pathogenicity of C. difficile (Rubino, 2015).

Membrane-bound proteins, such as G-protein coupled receptors (GPCRs), are intriguing targets in drug design (Claasen, 2005). Because these proteins often lose their structure when removed from their cellular environment, it is necessary to study them in their natural membrane environment (Claasen, 2005). One such technique that allows in vivo study of protein-ligand recognition events is saturation transfer difference (STD) NMR, and it has been used to analyze membrane-bound proteins when reintegrated into liposome membranes (Claasen, 2005). Surface glycoproteins make up a high percentage of membrane proteins, and the integrin \( \alpha_{IIb}\beta_3 \) is ubiquitous on platelets, recognizing proteins and peptides presenting the peptide motif RGD (Claasen, 2005).

STD NMR studies of the pentapeptide inhibitor cyclo(RGDfV) binding to the integrin \( \alpha_{IIb}\beta_3 \) reveal specific binding interactions between these two proteins, though not a complete epitope mapping and identification of all signals (Claasen, 2005). To solve for these unknowns, an STDD-filter is used in which an STD reference spectrum is subtracted from an experimental spectrum, and the new difference spectrum only shows resonances of the inhibitor peptide such as cyclo(RGDfV).
(Claasen, 2005).

These studies divulge the high affinity of cyclo(RGDfV) to integrin $\alpha_{IIb}\beta_3$ on native platelets (Claasen, 2005). The strongest STD effects are observed for Arg and Phe, while the protons with the alkyl side chain of Arg exhibit decreased saturation, suggesting the presence hydrophobic interactions near the peptide backbone (Claasen, 2005). Thus, the STDD method is very conducive to studying the interactions involving both membrane and submembrane proteins, including binding events of pharmaceutical inhibitors (Claasen, 2005). Many drug targets are seven helix transmembrane-spanning proteins such as GPCRs, and STDD NMR spectroscopy offers a promising technique to learn the atomic details of their structures (Claasen, 2005). The knowledge gained from such studies will contribute to the rational design of drugs (Claasen, 2005).

3.3) Eukaryotic applications:

Since the time in-cell NMR was first used to study proteins inside *E. coli* and then other prokaryotic cells, its applications have been expanded to eukaryotes (Hamatsu, 2013). Early studies of eukaryotic proteins consisted of injecting proteins into Xenopus laevis oocytes or eggs, and later on CPPs have been used to deliver proteins into mammalian cells (Hamatsu, 2013). These approaches have allowed for in vivo observations to be made concerning post-translational modifications, protein-ligand interactions, and protein-folding stability in eukaryotic cells (Hamatsu, 2013).

3.3.1) Yeast.

It is generally more challenging to apply in-cell NMR to eukaryotic cells using the overexpression techniques employed in prokaryotes because eukaryotes typically lack strong promotors necessary for producing sufficiently high amounts of isotopically labeled proteins (Bertrand, 2012). One exception is yeast because they contain bacteria-like promotors, and this overexpression
technique has been used to study yeast ubiquitin in both the cytosol and vesicles (Bertrand, 2012). Observing the differences in protein conformation between cytosolic ubiquitin and compartment-bound ubiquitin is relevant to the drug screening process (Bertrand, 2012).

Because the yeast cytosol occupies a majority of the intracellular space, ubiquitin tumbling is unaffected, thus, yielding a high quality in-cell spectrum (Bertrand, 2012). Illustrated in Figure 3.15, the in-cell NMR spectrum acquired 24 h post-induction demonstrates the dispersion of backbone amide protons from 6.5 ppm to 10 ppm, indicating that ubiquitin is natively folded (Bertrand, 2012). This also suggests that a significant portion of ubiquitin is free to tumble inside yeast cells (Bertrand, 2012). Moreover, the fact that many peaks are broadened or missing suggests that ubiquitin interacts with large intracellular complexes, a plausible conclusion since eukaryotic proteins contain many proteins with ubiquitin binding domains (Bertrand, 2012). However, due to the high degree of background noise between 7.5 ppm to 8.5 ppm, a range essential to study protein function, a significant portion of the amide proton spectrum was obscured (Bertrand, 2012).
Figure 3.15: In-cell NMR spectra of yeast ubiquitin. (Panel A) In-cell NMR spectrum obtained after 24 h post-induction without nitrogen starvation (Bertrand, 2012). (Panel B) In-cell NMR spectrum obtained 48 h post-induction without nitrogen starvation (Bertrand, 2012). (Panel C) In-cell NMR spectrum obtained 48 h post-induction with 6 h nitrogen starvation (Bertrand, 2012). (Panel D) Lysate in-cell NMR spectrum 48 h post-induction without nitrogen starvation (Bertrand, 2012).

The background noise is remedied by briefly nitrogen starving the cells following overexpression of the target protein while small metabolites containing $^{15}$N isotope are added to the target protein to enhance its peaks (Bertrand, 2012). Subsequent cell lysis experiments, which results in the protein being released from the vesicles, demonstrate that the spectrum of the released protein is identical to that of folded ubiquitin (Bertrand, 2012). This indicates that ubiquitin is maintained in its functional folded form within the vesicle, and this finding has important ramifications in drug-screening and therapy (Bertrand, 2012). For instance, knowledge of the cytoplasmic distribution of ubiquitin and other protein targets via in-cell NMR is important to understanding conditions that improve drug-target interactions in drug-screening (Bertrand, 2012).

Gobular proteins are more difficult to analyze via in-cell NMR due to their slow tumbling rate within the crowded cytoplasmic milieu and consequent peak broadening (Popovic, 2015). However, the globular portions of proteins can still be distinguished between the unfolded portions, and members of the frataxin family have been studied using such an approach (Popovic, 2015). This family stretches across a diverse array of prokaryotic and eukaryotic species, and is essential to iron-sulfur cluster biogenesis (Popovic, 2015). Its highly conserved 100 to 120 amino acid sequence contains a globular $\alpha\beta$ fold, and unique to eukaryotic frataxins is an N-terminus that allows import into the mitochondria following cytoplasmic production (Popovic, 2015). The yeast ortholog Yfh1 has been studied to elucidate whether the protein remains in its folded, functional form as it enters the mitochondria or if it becomes unfolded (Popovic, 2015).

In-cell NMR studies, as shown in Figure 3.16, have been carried out on the yeast ortholog Yfh1 to determine that structure and functional behavior of the N-terminal tail upon entry into the mitochondria (Popovic, 2015). The appearance of nine resonances otherwise absent from the
noninduced sample indicates that these peaks belong to the protein, and peaks have been assigned to residues of the unstructured N-terminus (Popovic, 2015). Residues E53 to V61 of this N-terminus constitute the mitochondrial import signal, which can be clearly observed in the difference spectrum, and indicates that this region of the protein is seemingly unfolded (Popovic, 2015).
Figure 3.16: In-cell NMR spectra of Yfh1. Panel (A) represents a spectrum of whole cells without induction (Popovic, 2015). Panel (B) shows a spectrum of cell after protein induction (Popovic, 2015). Panel (C) displays the different spectrum from which visible peaks corresponding to the N-terminus can be observed (Popovic, 2015).

In vitro solution structures and NOE analysis demonstrated the presence of a local fold or bend around V63, just after the N-terminal region but before the globular region (Popovic, 2015). This demarks the length of the flexible region along the N-terminus until V61, and indicates the importance of the unfolded, flexible region with regard to mitochondrial import from the cytoplasm (Popovic, 2015). The solution structures and chemical shifts of Yfh1 are shown in Figure 3.17 as a representation of the unfolded mitochondrial import signal along the N-terminal region in the context of the entire protein (Popovic, 2015).
Figure 3.17: In-vitro solution structures and chemical shifts of Yfh1. Panels (A) and (B) depict the structure of Yfh1, which as determined by chemical shift analysis in Panel (C) and 3D $^{15}$N NOESY analysis (Popovic, 2015). Residue P62, the boundary between the unfolded region and the local fold along the N-terminal region, is marked with an asterisk (Popovic, 2015).

Thus, the utility of in-cell NMR in studying intracellular proteins has provided evidence that the N-terminal region of Yfh1 is unfolded before and during mitochondrial entry, offering insight into the specific mechanism behind this process (Popovic, 2015).

3.3.2) Insects.

Insect model systems have the ability to produce and fold proteins of higher eukaryotic species correctly, making them a desirable system for obtaining $^{13}$C/$^{15}$N-labeled proteins (Hamatsu, 2013). One example of this application is the eukaryotic in-cell NMR in the S19 cell/baculovirus system, which consisted of four proteins: Streptococcus protein G B1 domain, GB1; *T. thermophilus* HB8 TTHA1718; rat calmodulin, CaM; and human HAH1 (Hamatsu, 2013). Illustrated in Figure 3.18, the 2D $^1$H-$^{15}$N HSQC spectrum of sf9 cells expressing GB1 contains high resolution cross-peaks with chemical shifts nearly identical to the in-vitro spectrum, indicative of proper folding of GB1 in the sf9 cells (Hamatsu, 2013). Also displayed in this figure are selected $^1$H-$^{15}$N strips from the HNCA, HN(CO)CA, and HNCO spectra, all in which 44 out of 56 backbone $^1$H, $^{13}$N$\alpha$, and $^{15}$N resonances have been assigned to GB1 in living cells (Hamatsu, 2013). A minor disparity in chemical shifts between in vitro and in-cell spectra around the loop regions may be an effect of viscosity and molecular crowding in the cytosol (Hamatsu, 2013).
Figure 3.18: In-cell NMR spectrum of GB1 in sf9 cells. Depicted in this figure are the difference spectrum obtained from a 2D $^1$H-$^{15}$N HSQC spectrum of $^{15}$N-labeled sf9 cells expressing GB1 and selected $^1$H-$^{13}$C strips extracted from the 3D HNCA, HN(CO)CA, and HNCO spectra of sf9 cells expressing GB1 (Hamatsu, 2013).

The NOESY spectrum of GB1-expressing sf9 cells reveals well-resolved cross-peaks, which are comparable to those of the in vitro spectrum (Hamatsu, 2013). This in-cell NMR approach using the sf9 cell/baculovirus system provides a reliable means of studying proteins that are either difficult to isolate or challenging to express in prokaryotes (Hamatsu, 2013). This approach is also applicable to eukaryotic proteins that require posttranslational modifications such as mammalian membrane proteins (Hamatsu, 2013).

Drosophila is another valuable model organism owing to the resemblances of many of its proteins and cellular processes to those of humans, and in-cell NMR offers a practical means of probing for biomarkers indicative of either disease or trauma (Righi, 2014). The end goal is to gain knowledge that will contribute to the rational design of drugs (Righi, 2014).

3.3.3) Xenopus.

Another model organism for cellular studies is Xenopus laevis, and the large size of their oocytes enables simple and inexpensive microinjection of proteins for study (Thongwichian, 2012). The diverse array of proteins taken from different species still retain their functionalities within oocytes (Heikkila, 2007). Numerous studies have shown that microinjecting proteins and other biomolecules into oocytes is a useful technique for studying a target protein’s structure and function within an intracellular environment (Heikkila, 2007). This unique cellular environment of the oocyte eliminates background noise frequently encountered in other cell types, though they still retain general physical properties of eukaryotic cells such as macromolecular crowding and cellular viscosity. Both folded and natively unfolded proteins have been analyzed in Xenopus oocytes using in-cell NMR spectroscopy (Thongwichian, 2012).

The unique “protein expression vessel” is commonly used to study membrane proteins,
whereby an mRNA encoding a membrane protein is microinjected into the oocyte (Sakai, 2006). Following injection, the mRNAs become translated into proteins, and additional $^{15}$N-labeled proteins are injected into the cell as to create a more uniform in-cell NMR spectrum (Sakai, 2006). However, the resultant $^1$H-$^{15}$N NMR spectrum of ubiquitin projects only a few cross-peaks, a weakness that may be due to extensive protein interactions between ubiquitin and proteins in the oocytes (Sakai, 2006). This theory is supported by the fact that much more intense cross-peaks were observed by the protein calmodulin, nearly double the size of ubiquitin, meaning that cytosolic viscosity cannot be the cause of ubiquitin's weak signal intensities (Sakai, 2006). These results also underpin the concept that the structural and functional behavior of proteins is influenced by its intracellular environmental conditions (Sakai, 2006). Just as oocyte microinjection has been used to examine the in-cell NMR spectra of ubiquitin and calmodulin, it should be able to study HSPs and their interactions with other proteins (Heikkila, 2007).

3.3.4) Humans.

Inomata and colleagues developed a way to observe in-cell NMR spectra of $^{15}$N-labeled proteins delivered to cells via CPPs (Inomata, 2009). This procedure entails fusing the CPP sequence of Tat from HIV-1 to the carboxy terminus of a human ubiquitin derivative containing the substituents at Leu 8, Ile 44, and Val 70 (Inomata, 2009). The resulting fusion protein is called Ub-3A-CPP$_{\text{Tat}}$ and is uniformly labeled with $^{15}$N before being incubated with human HeLa cells (Inomata, 2009). The 2D $^1$H-$^{15}$N correlation spectrum, illustrated in Figure 3.19, exhibits well-resolved cross-peaks in a pattern indicative of a stably folded and homogenously dispersed protein (Inomata, 2009). These peaks were confirmed to be bona fide by comparison to the in vitro spectrum (Inomata, 2009).
Figure 3.19: In-cell NMR spectra of a Ub-3A-CPP$_{\text{Tat}}$ fusion protein. The in-cell NMR spectrum of a Ub-3A-CPP$_{\text{Tat}}$ fusion protein (Panel a) is compared to an in vitro reference (Panel b) (Inomata, 2009). Electrophoresis confirms the site of cleavage (Panel c) (Inomata, 2009). There is a uniform distribution of the protein throughout the cytosol and nucleus of HeLa cells (Panel d) (Inomata, 2009).

An intense signal at the position corresponding to the C-terminal Gly 76 of mature ubiquitin in the in-cell spectrum coupled to the absence of the cross-peak of the C-terminal residue of CPP\textsubscript{Tat} suggests that the Ub-3A-CPP\textsubscript{Tat} fusion protein is cleaved between Gly 76 and Asp 77 (Inomata, 2009). It is believed that an endogenous ubiquitin-specific C-terminal protease is responsible for this cleavage, and this cleavage of Ub-3A-CPP\textsubscript{Tat} has been further verified by electrophoretic analysis of transduced fluorescently labeled protein (Inomata, 2009). The uniform distribution of the protein throughout the cytosol and nucleus of HeLa cells indicates that these proteins somehow detach from their CPP carriers, which are prone to binding to membranes and cytosolic components (Inomata, 2009).

The in-cell NMR spectra of specific proteins can be obtained by linking CPPs to cargo proteins via disulfide bonds, which are cleaved within the cytosol (Inomata, 2009). These spectra provide valuable information on the conformation, dynamics, and functions of proteins in the intracellular environment (Inomata, 2009). In this case, the intense signal observed at the position corresponding to the C-terminal Gly76 suggests that the CPP is cleaved from Ub-CPP\textsubscript{Tat} and that ubiquitins exist in a C-terminally unconjugated form in cells (Inomata, 2009).

Another application of in-cell NMR spectroscopy is studying interactions between proteins and small drug-like compounds, as occurs during the drug screening process (Inomata, 2009). Shown in Figure 3.20, the in-cell NMR spectrum of the protein target FKBP12 with the immunosuppressants FK506 and rapamycin in HeLa cells is an example (Inomata, 2009). The goal is to observe the efficiency of drug delivery and identify protein-drug interactions in atomic detail (Inomata, 2009). In-cell spectroscopy is an important means of reaching this end because the intracellular structure and behavior of the target protein can be accounted for in the design of drugs and in observing the drug-protein interactions (Inomata, 2009).
Figure 3.20: In-cell NMR spectra of FKBP12 with the immunosuppressants FK506 and rapamycin in HeLa cells. Panels (a) and (c) present the in-cell spectra of FKBP12 with FK506 and rapamycin while Panels (b) and (d) show the in vitro comparisons (Inomata, 2009).

In-cell NMR experiments have been performed on the IDP α-synuclein (α-syn) in order to better understand the structure and dynamics of this protein in mammalian cells (Theillet, 2016). This is important because of the protein's propensity to forming aggregates, which are detrimental to the cell (Theillet, 2016). To complicate matters, α-syn adopts different conformations in different types of neuronal cells, and each aggregate in unique ways (Theillet, 2016). Controversy also exists around the native structure of the protein: does it exist as a monomer or tetramer (Theillet, 2016)? Resolving this issue will contribute to our understanding of the mechanisms behind amyloid formation (Theillet, 2016).

To account for this diversity in α-syn form and location, in-cell NMR experiments were carried out by Theillet and colleagues in both neuronal and non-neuronal cells (Theillet, 2016). These include non-neuronal A2780 and HeLa cells and neuronal B65, SK-N-SH, and RCSN-3 cells (Theillet, 2016). The $^{15}$N isotop-enriched α-syn proteins were delivered in defined amounts into cultured mammalian cells (Theillet, 2016). A 1D $^{15}$N-filtered in-cell NMR spectrum was recorded on each specimen in order to correlate the effective NMR concentrations of isotope-enriched αSyn in different cell samples with concentrations determined by semi-qualitative western blotting, as shown in Figure 3.21.
Figure 3.21: In-cell NMR spectra of α-synuclein. Panel (a) depicts a schematic of intracellular protein concentrations versus effective NMR concentrations as well as semi-quantitative western blotting of the different cell types to approximate the concentration of protein delivered to each cell (Theillet,
Panel (b) presents the 2D $^1$H-$^{15}$N in-cell NMR spectra of $\alpha$Syn in A2780 and SK-N-SH cells as well as in N-terminally acetylated $\alpha$Syn in buffer (Theillet, 2016). Panel (c) displays the site-selective signal attenuations of N-terminal $\alpha$Syn residues (Theillet, 2016).


The 2D $^1$H-$^{15}$N in-cell NMR spectra of αSyn in these cells closely resembled the disordered reference state of the isolated protein including average backbone amide chemical shift changes of less than 0.01 ppm (Theillet, 2016). From this analysis, it can be concluded that the structure of monomeric αSyn remains largely unchanged in different intracellular environments (Theillet, 2016). Also deduced from the in-cell spectra is the existence of αSyn in the N-terminally acetylated form under physiologic conditions, as evidence by both the decrease in signal intensities for the first ten residues of αSyn and the similarity of these peak positions to those of the N-terminally acetylated protein (Theillet, 2016).

The dynamic properties of acetylated αSyn has been characterized in different cellular environments by measuring in-cell NMR peak intensity changes and backbone amide relaxation patterns in A2780 and SK-N-SH cells and in artificially crowded solutions (Theillet, 2016). Evidence shows that mainly signals from N- and C-terminal residues become attenuated, and that αSyn's interactions within the cytoplasm are characterized as electrostatic at the C-terminus and hydrophobic at the N-terminus (Theillet, 2016). Both types of interactions cause transient binding events with cellular membranes, further contributing to this protein's in-cell NMR peak broadening (Theillet, 2016). The effect of these binding interactions within the cell is confirmed by the disappearance of the line broadening characteristics as soon as the cell becomes lysed in diluted buffer solution (Theillet, 2016). Moreover, although intramolecular contacts result in loosely packed αSyn structures, the protein has not been shown to aggregate spontaneously (Theillet, 2016). Still, it is believed that the hydrophobic contacts and transient binding to cellular membranes may ignite the large conformational arrangements necessary for protein aggregation to occur (Theillet, 2016). Understanding these atomic details more clearly would be an asset to the rational design of drugs (Theillet, 2016).
3.4) Conclusions:

In-cell NMR spectroscopy is an important upgrade over the more traditional solution NMR spectroscopy in that it allows for the structure and function of a protein to be studied in its native cellular environment. This offers a more realistic depiction of a protein's molecular behavior in terms of binding interactions, folding, and mechanism of action because such an analysis accounts for the many transient and nonspecific interactions that take place within the crowded cytoplasm of both prokaryotic and eukaryotic cells. In this way, a variety of proteins from both domains of life have been scrutinized, and the ability to express foreign proteins in model organisms while retaining their functions is an additional advantage in deciphering their characteristics. The knowledge gained from atomic-resolution observations of target proteins within the cell provides important insights into disease progression as well as the rational design of drugs.

Overall Conclusions:

The techniques used to study proteins for the purpose of high-throughput drug-screening have evolved over the years from different methods of fluorescent readout such as GFP, luciferase, and bodipy to approaches that acquire atomic details of target proteins including x-ray crystallography and the more refined technique in-cell NMR spectroscopy. Equipped with their own unique advantages, the development of each method has enhanced the ability to investigate the structural and functional properties of target proteins. Obtaining a better understanding of a protein's characteristics, especially within its natural cellular environment, contributes greatly to the identification of disease biomarkers, protein-targets for drug-therapy, and the rational design of drugs themselves.
References:


12798-806.


Maldonado, A.Y. (2013) Fate of Pup inside the Mycobacterium proteasome studied by in-cell NMR.
**PloS One, 8(9), e74576.**


Schmidt, Samual. *Detecting Cytosolic Peptide Delivery with the GFP Complementation Assay in the Low*


Wimberly, Brian. A Detailed View of a Ribosomal Active Site: The Structure of the L11-RNA Complex.


