1-1-2012

Development of an in-cell förster resonance energy transfer technique to study protein structure inside living cells

Victoria Lynn Murray
Wayne State University,

DEVELOPMENT OF AN IN-CELL FÖRSTER RESONANCE ENERGY TRANSFER TECHNIQUE TO STUDY PROTEIN STRUCTURE INSIDE LIVING CELLS

by

VICTORIA LYNN MURRAY

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2011

MAJOR: BIOCHEMISTRY AND MOLECULAR BIOLOGY

Approved by:

_____

Advisor Date

_____

_____

_____

_____

DEDICATION

I would like to dedicate my dissertation to my mother, Maria Bennecke, and husband, Joseph Murray.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor, Dr. Jianjun Wang, for the excellent training, support and advice he provided me during my Ph.D. studies.

I would also like to thank my committee members, Dr. Timothy Stemmler, Dr. Chunying Li and Dr. Jian-Ping Jin for their advice and guidance.

I'm thankful to all of my former and current colleagues: Snow, Arun, So-Young, Yunhuang, Yonghong, Chris, Nan, Yuefei, Anthony, Sam, Winston, Kristi and Carrie and the many others in Dr. Wang's laboratory. I'm especially thankful for the training and advice given to me over the years from Wentao, Leo and Qianqian.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2x YT	2 times yeast extract tryptoone
5-HT	5-hydroxy-L-tryptophan
ApoAI	Apolipoprotein AI
ApoE	Apolipoprotein E
ASA	Accessible surface area
BP	β-propeller/EGF domain of LRP5/6
CET	Cryo-electron tomography
CFP	cyan fluorescent protein
CPP	cell-penetrating peptide
Cryo-EM	Cryo-electron microscopy
DMEM	Dulbecco's Modified Eagle Medium
DTT	Dithiothreitol
ER	Endoplasmic reticulum
ERAD	ER-associated degradation
FlAsH	Fluorescein-based arsenical hairpin
FP	Fluorescent protein
FRET	Förster resonance energy transfer
GFP	Green fluorescent protein
HSQC	Heteronuclear single quantum coherence
IAEDANS	5-[2-[(2-Iodo-1-oxoethyl)amino]ethylamino]-1-naphthalenesulfonic acid
IPTG	Isopropyl-β-D-thiogalactoside

KAN	Kanamycin
LCAT	Lecithin:cholesterol acyl transferase
LDL	Low-density lipoprotein
LDLR	Low-density lipoprotein receptor
LRP	Low-density lipoprotein receptor-related protein
MESD	Mesoderm development protein
MRI	Magnetic resonance imaging
NMR	Nuclear magnetic resonance
NOESY	Nuclear overhauser and exchange spectroscopy
OD	Optical density
OD_{600}	Optical density at 600 nm
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein data bank
PEI	Polyethyleneimine
PET	Positron emission tomography
RAP	LDL receptor associated protein
ReAsH	Resorufin-based arsenical hairpin
SDS	Sodium dodecyl sulfate
SMF	Small molecule fluorophore
SMoC	Small molecule mimics
UPR	Unfolded protein response
YFP	Yellow fluorescent protein

CHAPTER 1

REVIEW OF THE LITERATURE

1.1 Research Goals

The research goal of my thesis is to develop an in-cell fluorescence technique to study the *in situ* protein structure and function inside living mammalian cells. To achieve this goal, our rationale is to label a protein of interest *in vitro* with small molecule fluorophores, either at specific residues or at surface located lysine residues. We will then deliver this fluorescence labeled protein into the correct intracellular compartment of living mammalian cells using our newly developed QQ-protein delivery technology. This will generate a special mammalian cell population that only contains one fluorescent labeled protein, whereas the other intracellular background proteins are unlabeled. We anticipate that this strategy will allow us to study structure, intracellular trafficking and *in situ* function of this fluorescent labeled protein inside the living cell using either confocal fluorescence imaging or fluorescence spectroscopy.

To gain structural information of a protein inside living cells, we will develop an in-cell FRET technique that allows us to measure the distance between the specifically labeled fluorescence donor and acceptor at atomic resolution. Our strategy is to utilize the intrinsic fluorescence of the tryptophan residue in a protein as the fluorescence donor and label the protein with a fluorescence acceptor at a cysteine residue with a thiol-reactive small molecule fluorophore, such as IAEDANS. This provides a fluorescence donor and acceptor with minimal disturbance of the structure of the

protein of interest. In addition, this also generates a specifically labeled protein with site-specific labeled fluorescence donor and acceptor, allowing us to accurately measure the distance between the donor and acceptor at atomic resolution. Thus, this technique enables us to gain structural information of a protein inside cells at atomic resolution.

To solve the problem of disturbance of protein-protein interactions between the protein of interests and intracellular proteins that also contain tryptophan residues, we will label the protein of interest with a tryptophan analog, 5-hydroxy-L-tryptophan (5-HT), to generate a specific 5-HT labeled protein. Since a 5-HT labeled protein can be excited at 310 nm to generate emission spectra, where minimal excitation of regular tryptophan residues at this wavelength, this will allow us to eliminate any contribution of intracellular proteins to the FRET due to protein-protein interactions. Thus, the in-cell FRET measurement using this special labeled protein sample will be only intra-protein FRET between the labeled 5-HT donor and the acceptor. Therefore, this strategy will allow us to obtain accurate distance measurements between the fluorescence donor and acceptor within the labeled protein.

An important aspect of this project is to develop an innovative FRET technique that allows us to obtain multiple distances within a protein via a single FRET measurement. To achieve this goal, we will prepare a protein sample that contains multiple fluorescence donors (multiple 5-HTs) and a single acceptor. We will develop a fluorescence technique for FRET measurement using the synchronous fluorescence spectroscopy for these FRET measurements of these protein samples for the purpose to simultaneously obtain multiple distances via a single FRET measurement.

In summary, this thesis pioneers a novel in-cell fluorescence technology to study the *in situ* structure, intracellular trafficking and functions of a protein of interests inside the cells. This is a very challenging project, since the intracellular environment is extremely complex. However, any success of this project shall generate an innovative in-cell fluorescence technology that enables the study of the cellular physiology of roles played by a protein of interest. This review of literature begins with a look at available high-resolution structural biology techniques as well as low-resolution cell biology techniques in order to explain why fluorescence spectroscopy was the chosen method for this study.

1.2 High-resolution structural biology and Low-resolution cell biology

After the invention of the microscope, scientists in the cell biology field have consistently pushed the optical resolution limits (Figure 1-1). With the advent of the light microscope, one could study at the tissue level and visualize individual cells. As the lenses improved, the subcellular compartments became apparent. Until the utilization of fluorescent proteins in the 1990s, individual proteins could not be elucidated within cells or even when purified and in solution. Other techniques were developed to determine the structure of a protein at the atomic level.

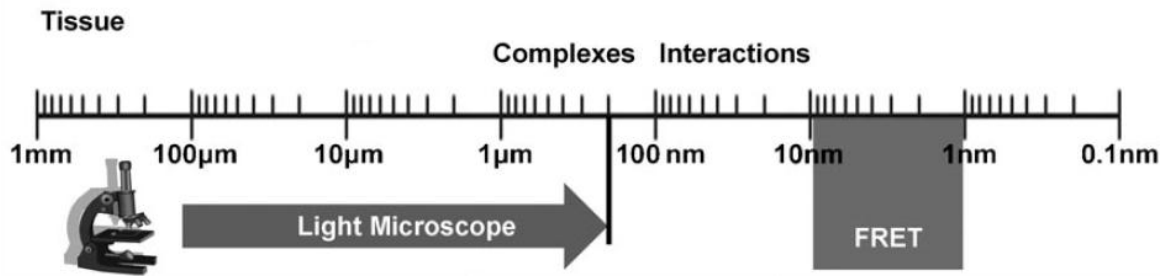


Figure 1-1: Optical resolution scale from the tissue level to atomic resolution. Reprinted from Sun 2011 [1].

1.2.1 High-resolution structural biology techniques

Structural biology is a combination of methods from the fields of molecular biology, biochemistry and biophysics. Various technologies have been developed to elucidate protein structure and the top three methods are X-ray crystallography, nuclear magnetic resonance (NMR) and cryo-electron microscopy (cryo-EM).

1.2.1.1 X-ray crystallography

The field of X-ray crystallography has a long, rich history. Several milestones were achieved in the early 20th century as shown in Table 1-1 [2, 3].

Table 1-1: Important milestones in the history of X-ray crystallography

Year	Achievement	Achieved by
1912	Observation of X-ray diffraction	Friedrich, Knipping & von Laue
1912-13	Bragg's law developed	Bragg
1912-50s	Minerals, organic and small biological molecules structures solved used X-ray crystallography	
1934	First X-ray diffraction picture taken of pepsin protein	Bernal
1958	First protein structure solved by X-ray crystallography	Kendrew & Perutz
1970-80s	Utilization of synchrotron X-ray radiation	
1990s	MAD Phasing	Hendrickson

The first step to obtaining X-ray data is the production, purification and crystallization of the target protein. The most important part of any protein crystallography studies is protein crystallization. Without perfect crystals of protein (or any other biological samples) it is impossible to carry out any crystallographic structural studies. The aim of protein crystallization is to produce well-ordered protein mono-crystals without any inclusion and large enough to diffract X-Ray beam. Despite very wide knowledge about protein crystallization it is still impossible to predict any conditions for protein crystallization. The protein crystallization process is still empiric and the biggest part of success is hidden in the hand and experience of the scientist who performed the protein crystallization and pure luck. However, recent structural genomics projects developed high-throughput screening methods using robots for

different crystallization conditions significantly accelerate the success rate of protein crystallization [4, 5].

Once a decent sized crystal is produced, it can be mounted on a goniometer and gradually rotated while being bombarded with X-rays, producing a diffraction pattern of regularly spaced spots known as reflections. However, the information at this point only contains the intensities of the diffracted rays. In order to continue with structure calculation, the phase must be determined as this information is lost during the X-ray data collection.

The phase can be determined in a number of ways. The first method is isomorphous replacement where a crystal is soaked in a solution containing heavy atoms or co-crystallized with that heavy atom. The native crystal X-ray diffraction is then compared to the crystal containing the heavy atom and the differences allow for phase determination [6]. The second method is molecular replacement. Many proteins are similar to other proteins with a known structure. The new protein's structure can be solved by using the intensities and phases of the known structure and can give a general orientation guideline [7]. The third and most commonly used method is multi-wavelength anomalous dispersion [8]. Typically, selenomethonine is incorporated into the protein during translation. Data is collected at a variety of wavelengths, the various diffraction patterns are compared allowing for phase determination. Since its introduction in 1990, this has become the most standard method for solving X-ray crystal structures [9].

Once the phase has been determined, electron density maps can be calculated. At this point, the atoms of the known molecules can be fit into the electron density map. X-ray data processing is a key stage for protein crystallography. The final quality of X-Ray data is related with the data processing procedure, which includes integration of the crystallographic data and scaling. The first step between experimental X-ray data and protein structure is the structure solution procedure. Structure solution can be based on the many different techniques, depending on the data available. The final stage of protein crystallography is the structure refinement. At this stage it is necessary to do a lot of visual graphics work with the model together with structure refinement [10]. Figure 1-2 displays a summary of all the steps needed to determine protein structure from a crystal: (1) crystallization, (2) collect x-ray diffraction data, (3) determine phase and create electron density map, (4) fit protein and produce an atomic model.

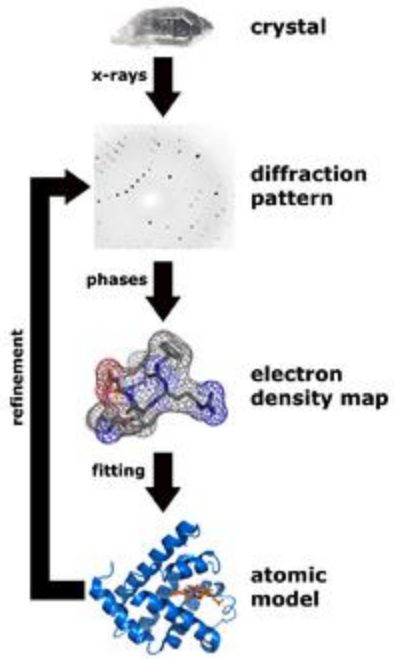


Figure 1-2: Summary of steps needed to determine structure of protein based on x-ray crystal data.

X-ray crystallography suffers from a few drawbacks when using this technique to determine protein structure. First, flexible regions of a protein cannot be seen in an electron density map and therefore cannot be seen in the final structure. Second, occasionally there are artifacts from the crystallization process such as protein dimerization when the protein in solution does not dimerize.

1.2.1.2 Nuclear magnetic resonance (NMR) spectroscopy

Nuclear magnetic resonance spectroscopy, most commonly known as NMR spectroscopy, is a research technique that exploits the magnetic properties of certain

atomic nuclei to determine physical and chemical properties of atoms or the molecules in which they are contained. It relies on the phenomenon of nuclear magnetic resonance and can provide detailed information about the structure, dynamics, reaction state, and chemical environment of molecules [11].

Most frequently, NMR spectroscopy is used by chemists and biochemists to investigate the properties of organic molecules, though it is applicable to any nucleus possessing spin. This can range from small compounds analyzed with 1-dimensional proton or ^{13}C NMR to large proteins or nucleic acids using 3 or 4-dimensional techniques.

The first step to collecting NMR data is the expression and purification of the target protein. Depending on the size of the protein, it can be single or double-labeled with ^{13}C and/or ^{15}N. For large proteins, they can be triple-labeled with ^{2}H, ^{13}C and ^{15}N. Once the protein has been purified and a powder is obtained, it is then dissolved in buffer at around a 1 mM concentration. The solution is placed in a slim, cylindrical tube and taken to the NMR spectrometer. The main subsystems of an NMR spectrometer are: (1) superconducting magnet (300-950 mHz), (2) probe, (3) pulse programmer and rf transmitter, (4) receiver and (5) data acquisition and processing computer [12]. When placed in a magnetic field, the NMR active nuclei ($^{1}H/^{13}C/^{15}N$) absorb electromagnetic radiation at a frequency characteristic of the isotope. The resonant frequency, energy of the absorption and the intensity of the signal are proportional to the strength of the magnetic field [11-13].

There is a variety of NMR experiments that can be performed and depending on the size of the protein. One-dimensional ^{1}H-NMR spectral analysis is only useful for elucidating the structures of small organic molecules. More complex samples require two and three-dimensional analysis using ^{13}C and ^{15}N nuclei. Two-dimensional analysis basically measures the spectra of two different nuclei in a sample and plots them against each other. The different local electronic environments of each of the nuclei within a folded structure give rise to unique spectral shifts. Each crosspeak represents a unique proton-nitrogen pair within the protein. Well-folded proteins typically give better resolved spectra with distinct peaks. Unstructured proteins give spectra with poor resolution between peaks. This type of analysis is useful as a rapid screen for the "foldedness" of a given peptide/protein and can be used to screen for optimal conditions for good "foldedness" [13].

Complex protein samples will have significant spectral overlap with 2D-NMR, thereby complicating the downstream assignments and structure determination. Three-dimensional analysis extends the spectra into one further dimension, to resolve these ambiguities. An example of a typical experiment is the 3D ^{1}H-^{15}N NOESY correlation spectrum. This experiment, like the 2D NMR spectroscopy described above, correlates amide hydrogens to their resident nitrogens. The experiment extends into another ^{1}H spectral dimension and contains ^{1}H-^{1}H NOE correlations between the amide hydrogens and other nearby hydrogens [14].

Another example of a 3D-NMR experiment is the 3D HNCACB (^{1}H-^{15}N, ^{13}Cα-^{13}Cβ). This experiment can give unambiguous assignments of residues to the NMR shifts. Each

amide ^{1}H-^{15}N pair is coupled to the ^{13}C nuclei in backbone of its own residue and those of the one before it in the primary sequence. The result of this phenomenon is that each amide strip contains four shifts (two strong peaks for its resident carbons and two weak peaks for the carbons of the residue before it). Shifts from alpha vs. beta carbons are identified by having opposite signs. These data strips can then be used to determine which residues correspond to which shifts, a process called sequential assignment. Once the assignment is complete, other NMR experiments like the 3D-NOESY-correlation can be performed to determine which atoms are nearby in space (< 6Å), the basis of NMR solution structure determination. Figure 1-3 illustrates all the steps needed to determine protein structure based on an NMR sample: (1) isotopically labeling protein samples, (2) NMR data collection using specific pulse sequences depending on the 2-D or 3-D experiment performed, (3) process data and assign peaks with NMR software, (4) perform structural calculations and produce an NMR structure of the protein.

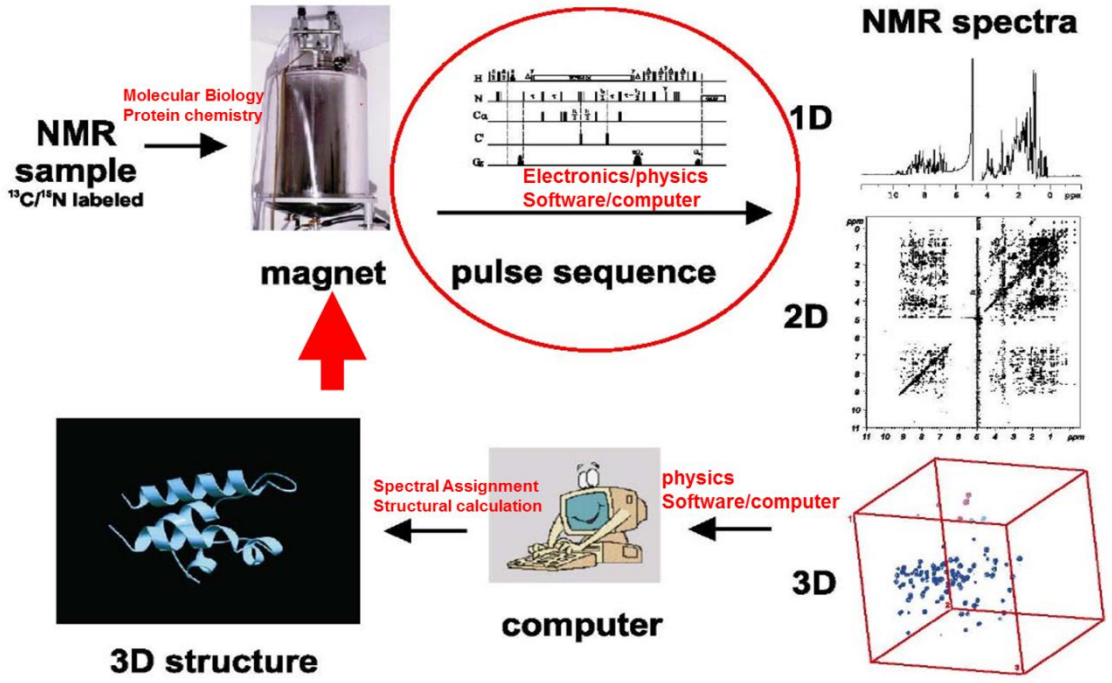


Figure 1-3: A summary of all the steps needed to determine a protein structure using NMR spectroscopy.

NMR can be complimentary to or more advantageous than X-ray crystallography in several ways. First, comparisons can be made between a protein's structure that has been crystallized and the same protein solved by solution NMR. Second, NMR can provide structural information for proteins that do not crystallize. Third, the solution conditions can typically vary widely allowing a protein to be studied under various conditions that include changes in pH and denaturing conditions. On the other hand, proteins tend to crystallize under strict conditions and changing those conditions may result in the protein not crystallizing. Finally, NMR is conducive to the study of protein dynamics [13]. One major drawback for NMR is that it is limited to smaller proteins (<

40 kD) for high-resolution NMR structural determination because there is too much spectral overlap and significantly enhanced linewidths for large proteins that compromise signal-to-noise.

1.2.1.3 Cryo-electron microscopy

There are times when a protein or protein complex is too difficult to form a consistent crystal and too large for solution NMR techniques, cryo-electron microscopy (cryo-EM) presents a possible solution. Cryo-electron microscopy is a form of transmission electron microscopy (EM) where the sample is studied at cryogenic temperatures (generally liquid nitrogen temperatures) [15].

The popularity of cryo-electron microscopy stems from the fact that it allows the observation of specimens that have not been stained or fixed in any way, showing them in their native environment, in contrast to X-ray crystallography, which generally requires placing the samples in non-physiological environments, which can occasionally lead to functionally irrelevant conformational changes. In practice, the resolution of cryo-electron microscopy maps is not high enough to allow for unambiguous model construction on the basis of EM maps only, and models obtained by protein crystallography are used to interpret the cryo-EM maps [16]. However, the resolution of cryo-EM maps is improving steadily, and some virus structures obtained by cryo-EM are already at a resolution that can be interpreted in terms of an atomic model.

A version of electron cryo-microscopy is cryo-electron tomography (CET) where a 3D reconstruction of a sample is created from tilted 2D images. Electron tomography is

comparable to medical tomographic techniques like CAT, PET and MRI in the sense that it provides a 3D view of an object, yet it does so at a cellular scale and with nanometer resolution. Electron tomography has the unique ability to visualize molecular assemblies, cytoskeletal elements and organelles within cells. The three-dimensional perspective it provides has revised our understanding of cellular organization and its relation with morphological changes in normal development and disease. Cryo-electron tomography of vitrified samples at cryogenic temperatures combines excellent structural preservation with direct high-resolution imaging [15]. Current resolutions of ET systems are in the 5-20 nm range, suitable for examining supra-molecular multi-protein structures, although not the secondary and tertiary structure of an individual protein or polypeptide.

1.2.1.4 Limitations and challenges of current structural biology techniques

Each of the major methods used to determine protein structures has its own unique limitations and disadvantages. X-ray crystallography has generated more structures in the PDB database than any other technique and accounts for 87% of the total structures. X-ray crystallography has a long history and has become a high-throughput method for obtaining structures. However, it still suffers from a few disadvantages: crystal formation is more of an art than science, membrane proteins are extremely difficult to crystallize, artifacts from crystallization may negatively affect structure, and flexible regions are not visible.

NMR accounts for 12% of the structures in the PDB database. Since NMR is solution-based, the key advantage of protein NMR is the option to study protein dynamics under changing conditions. Proteins that have been unable to crystallize have been studied using NMR. However, for the most part, NMR is still only used to study proteins under 30 kD, despite available techniques to study larger proteins.

Cryo-EM, or CET, is advantageous in the fact that it can provide a snapshot into subcellular level structure. In combination with other structural techniques, Cryo-EM can aide in the final structure determination of proteins. Cryo-EM is especially helpful with modeling structures for membrane proteins. The main disadvantages to this technique are its nanometer resolution and frozen state of the sample which could produce artifacts.

1.2.2 Low-resolution cell biology techniques

Although there have been numerous advances in the field of cell biology, it is still considered low-resolution as protein structures cannot be elucidated. Listed below are various technologies available that have pushed the lower boundaries of optical resolution. The majority of the techniques are fluorescence based.

1.2.2.1 Fluorescence imaging

Fluorescence imaging is a broad category that covers many available fluorophores: small organic dyes, quantum dots and fluorescent proteins. There are a number of techniques available, including immunolabeling and genetic tagging.

Depending upon the fluorophore chosen and technique used, a number of passive and active applications can be elucidated from mammalian cells. Some of the passive applications include protein expression and localization in primary cells and fixed tissues, Active applications include protein diffusion and trafficking, conformational changes, protein-protein interactions, protein synthesis and turnover, manipulation of protein activity, seeing endogenous enzyme activity [17, 18].

Small organic dyes or fluorophores (< 1 kD) can covalently bond to a protein and have been designed to be much brighter, photostable and have reduced self-quenching. Since they do not have the capability to specifically bind to a target protein within a cell, they must be first attached to secondary antibodies that will specifically bind to primary antibodies for the target protein [17].

Quantum dots are inorganic nanocrystals that are thought to be superior to small organic dyes in several ways. One of the most obvious differences is brightness due to the high extinction coefficient combined with a comparable quantum yield to fluorescent dyes. The second difference is their high photostability, allowing much less photobleaching [19]. It has been estimated that quantum dots are 20 times brighter and 100 times more stable than traditional fluorescent reporters [20].

The improved photostability of quantum dots, for example, allows the acquisition of many consecutive focal-plane images that can be reconstructed into a high-resolution three-dimensional image [21]. Another application that takes advantage of the extraordinary photostability of quantum dot probes is the real-time tracking of molecules and cells over extended periods of time [22]. Antibodies, streptavidin,

peptides, nucleic acid aptamers, or small-molecule ligands can be used to target quantum dots to specific proteins on cells [23-25].

1.2.2.2 Immunohistochemistry

The goal of immunohistochemistry is to provide color and contrast to microscopic images. The field uses different techniques to accomplish the specific labeling of biological structures. Histochemists pioneered the use of small-molecule cellular stains, labeled molecules such as antibodies, and enzyme mediated detection and signal amplification. Historically, however, histochemistry involves the imaging of fixed cells and tissues. The advent of genetic manipulation techniques has greatly expanded histochemical methods to living cells [26].

The fundamental concept behind immunohistochemistry is the demonstration of antigens (Ag) within tissue sections by means of specific antibodies (Abs). Once antigen–antibody (Ag-Ab) binding occurs, it is demonstrated with a colored histochemical reaction visible by light microscopy or fluorochromes with ultraviolet light [27].

Immunolabeling is a technique that involves identifying endogenous proteins with primary antibodies, followed by detection with secondary antibodies conjugated with small organic dyes or quantum dots. Alternatively, the primary antibodies can be directly conjugated with a fluorophore itself. Both methods depend on the specificity of the primary antibody [17].

1.2.2.3 Fluorescent proteins

The Green Fluorescent Protein was first discovered in 1962 alongside aequorin, a chemiluminescent protein from *Aequorea* jellyfish [28]. Since then, it has been studied in numerous biochemical ways including crystallization and the phenomenon of energy transfer. However, the major breakthrough did not occur until 1992, when Prasher et al cloned the gene and then by Chalfie et al and Inouye et al who demonstrated that the protein fluoresced when expressed in other organisms [29-32].

The Green Fluorescent Protein was first crystallized in 1974 and diffraction data collected in 1988, but it was not until 1996 that two independent groups determined the structure of the protein [32-34]. Figure 1-4 shows the two crystal structures, Panel A by Ormo et al and Panel B by Yang et al.

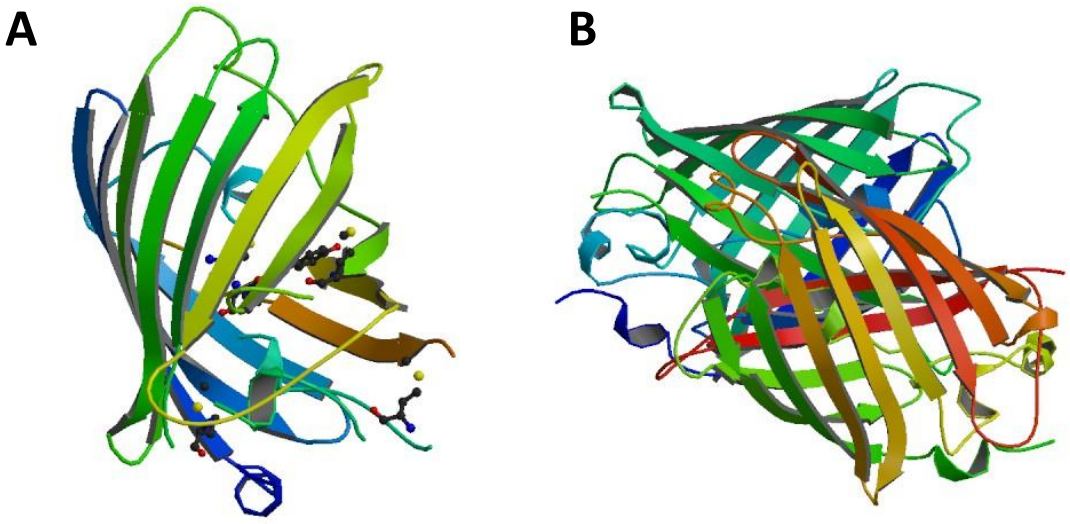


Figure 1-4: Panel A shows X-ray crystal structure of GFP solved by Ormo et al, PDB# 1EMA. Panel B shows X-ray crystal structure of GFP solved by Yang et al, PDB# 1GFL.

The structure is an 11-stranded β-barrel with an α-helix running through the middle of the structure and attached to the chromophore. Based on the X-ray crystal structure, it became clear that the chromophore is buried almost perfectly in the center of the protein. It also became obvious that there was no way to truncate the protein in order to make it smaller.

The GFP has a major peak excitation at 395 nm and a minor peak excitation at 475 nm, with corresponding peak emissions at 508 nm and 503 nm respectively. As a result of various mutations, different variants of green fluorescent protein can be produced that are excited and emit at different wavelengths and can be classified as yellow, cyan and blue fluorescent proteins [32].

Although the chromophore is protected inside the barrel of the protein and is typically insensitive to changes in the chemical environment, there are a few overall stability issues that inevitably affect fluorescence. It was found that after raising the temperature of the solution above room temperature, wild type GFP was unable to properly fold. Subsequent mutations of GFP allow for it to be folded properly at 37 °C [32]. In addition, changing to a high pH decreases excitation amplitude at 395 nm and gains amplitude at 470 nm, although pH > 11 is physiologically irrelevant and only interesting biochemically.

There are quite a few techniques available to study intracellular process in live and fixed cells using fused fluorescent proteins. There are a number of passive applications that utilize fluorescent proteins as spatial or temporal markers. The protein

trap strategy fuses fluorescent proteins to a library of cloning DNA sequences and based on their localization pattern, interesting proteins can be cloned and identified. Fusion of fluorescent proteins can also yield information about the production of small molecule messengers and gene activity/transcripts [18].

Fluorescence can serve as a temporal marker in several ways. First, fused proteins can serve as an indicator of gene expression. In addition, they can provide a temporal history by using a specifically designed fluorescent protein, DsRed, that changes from green to red fluorescence over a 24-hour period. The ratio of green to red fluorescence will provide the temporal history of the promoter activation [18]. Accumulation and degradation of fluorescent protein fused substrates can provide dynamic information about intracellular complexes. Finally, protein diffusion and trafficking can be monitored over time with the use of fused fluorescent proteins [17].

1.2.3 Is an in-cell structural biology technique possible?

With all of the available structural and cell biology techniques available, it is still impossible to obtain structural information from proteins inside living cells. X-ray crystallography cannot be used since that requires a solid crystal consisting of a purified protein at a high concentration. Cyro-EM is excluded for similar reasons. Although a pure sample is not required, there is the problem of flash freezing a sample to cryogenic temperatures that may result in artifacts. In addition, although in theory cryo-EM can reach Angstrom resolution, in practice, it still has nanometer resolution.

In-cell NMR is still in the infant stages of development. There are several challenges to overcome concerning in-cell NMR [35, 36]. The first challenge is the introduction of an isotopically labeled protein into the cell. Protein transduction is highly inefficient with currently available methods and using the cells themselves to produce the isotopically labeled protein results in native proteins being labeled as well. The second problem is the millimolar concentration requirement. In order to achieve a good signal to noise ratio, the concentration of protein must be high to the point it becomes physiologically irrelevant.

As far as the available cell biology techniques are concerned, they all share the common problem of low-resolution. Of course, there have been advances made with single molecule studies and super-resolution fluorescence imaging. But the fact remains, that many of these techniques are in their infant stages and/or require expensive equipment not available to most labs [37, 38]. Fluorescent proteins are not sufficient to calculate intra- or intermolecular distances with angstrom resolution.

Figure 1-5 illustrates the differences between fluorescent proteins and small molecule fluorophores in terms of sensitivity to changes in the environment. In the top half of the figure, a fluorescent protein is shown fused to an unfolded target protein. Since the chromophore is shielded by the barrel of the protein and is far removed from any changes that the target might be undergoing, be it folding, post-translational modification or protein-protein interactions, the fluorescent protein is insensitive to these changes. Small molecule fluorophores on the other hand are small enough to be bonded to the target protein itself, in various spots if necessary. Since they are bonded

to the sidechain of the target protein, any changes in its chemical environment caused by folding or modifications can be detected.

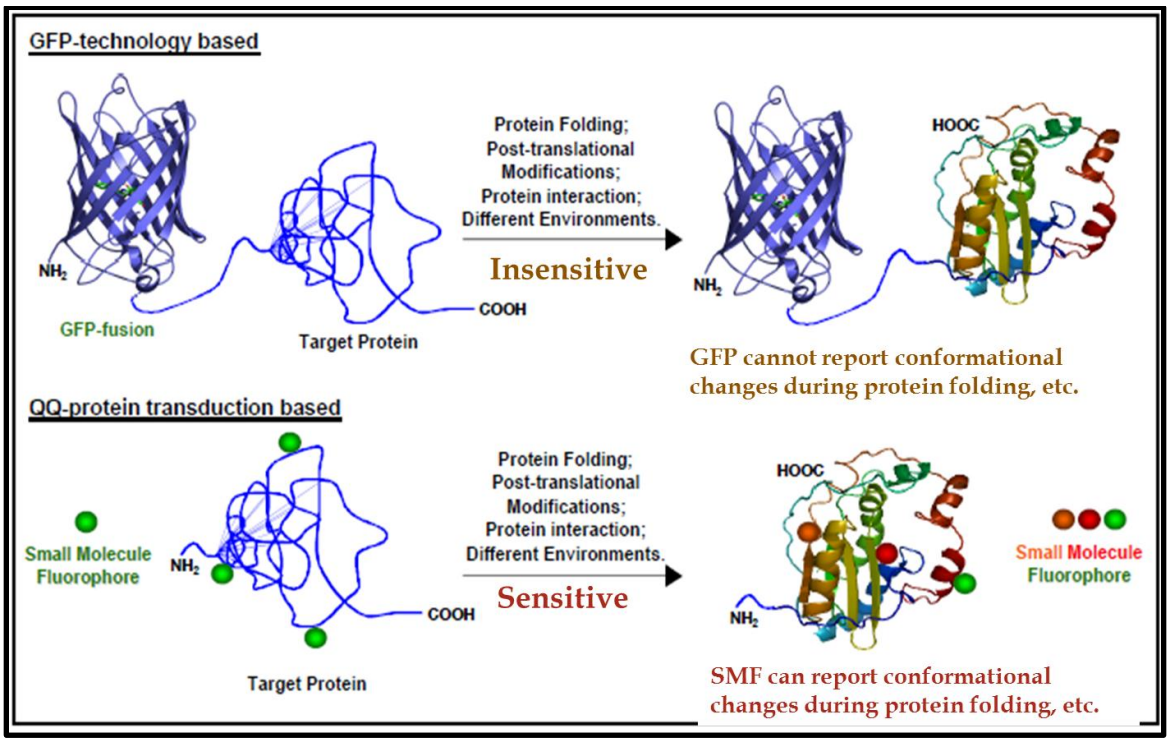


Figure 1-5: Comparison between GFPs and SMFs concerning sensitivity of fluorophore to changes in its environment.

In order for atomic-resolution information to be gained from inside living cells and have the technology be easily accessible to the average structural biology lab, we need to utilize a few unrelated techniques. First, bacteria could be used to produce large quantities of recombinant protein. This will be followed by specific labeling with small molecule fluorophore donor and acceptors. Next, the protein will be transfected into living mammalian cells. This will generate live mammalian cells that contain only

one protein that is labeled with a fluorescence donor and acceptor. Such a specifcallly labeled mammalian cell allows us to perform FRET experiments. We think that by introducing exogenously produced proteins labeled with small molecule fluorophores into living cells, it will be possible to obtain structural information using FRET. The next review section takes a closer look at the history and applications of FRET.

1.3 Förster resonance energy transfer (FRET)

1.3.1 History and mathematics behind FRET

Theodore Förster published the first paper concerning Förster (or Fluorescence) resonance energy transfer in 1946 [39]. FRET is defined as the physical phenomenon in which there is a non-radiative transfer of energy via long-range dipole-dipole coupling between a donor and acceptor molecules [40]. There are three basic requirements in order for FRET to occur (Figure 1-6):

(1) The donor and acceptor molecule must be within 1-10 nm of each other (Figure 1-6, Panel A).

(2) The emission spectrum of the donor must overlap the absorption spectrum of the acceptor (Figure 1-6, Panel B).

(3) The dipoles of the donor and acceptor molecule cannot be perpendicular (Figure 1-6, Panel C).

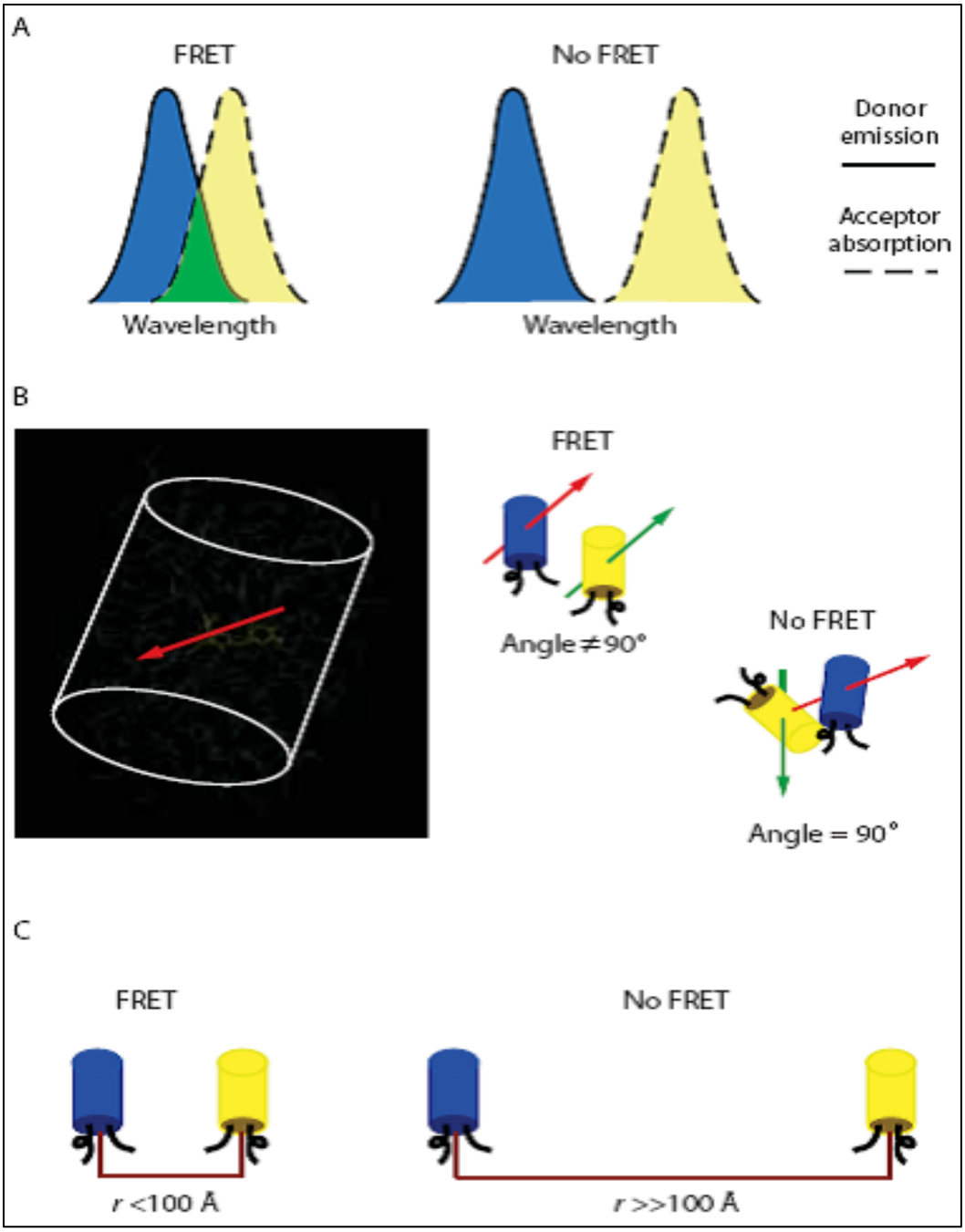


Figure 1-6: Panel A – donor emission spectrum must overlap acceptor absorbance spectrum. Panel B – donor and acceptor dipoles must not be perpendicular to one another. Panel C – donor and acceptor molecules must be between 10 and 100 Å of one another.

The rate of transfer (k_t) is proportional not only to the distance between the donor and acceptor (r), but also the unperturbed lifetime of the donor (τ_o) and the Förster distance (R_o – the distance at which energy transfer is 50%):

[1] $$k_t = \frac{1}{\tau_o} * \left(\frac{R_o}{r}\right)^6$$

The efficiency of energy transfer (E) is a quantitative measure of the number of quanta that are transferred from donor to acceptor. E is also known as the quantum yield of energy transfer. To determine the efficiency of energy transfer, we can multiply the rate of transfer (k_t) by the first excited singlet state lifetime (τ):

[2] $$E = k_t * \tau = \frac{R_o^6}{R_o^6 + r^6}$$

where $\tau_o^{-1} = k_f + k_{nr} + k_{isc} + k_{pb}$

and $\tau^{-1} = \tau_o^{-1} + k_t$

We can also determine E by measuring the steady-state donor fluorescence intensity from a sample containing only the donor and another sample containing the donor-acceptor pair [40].

[3] $$E = 1 - \frac{Q_{DA}}{Q_D}$$

Where Q_D is the quantum yield of the sample containing only a donor and Q_{DA} is the quantum yield of the sample containing both the donor and acceptor.

By setting equation [2] equal to equation [3] and solving for r, we can now calculate the distance between the donor and acceptor on our protein:

[4] $r = R_o \left(\frac{1}{1 - \frac{Q_{DA}}{Q_D}} - 1 \right)^{1/6}$

These calculations can be greatly simplified by using a donor-acceptor pair with a known Förster distance (R_o). Calculations are still possible without having a known R_o value, but you must use equation [5] to determine this value:

[5] $R_o^6 = c_o \kappa^2 J n^{-4} k_f \tau_o$

Where c_o = 8.8 x 10^{-28} nm, κ^2 is the orientation parameter, typically valued at $\frac{2}{3}$, J is the overlap integral of the donor emission and acceptor absorption spectra, n^{-4} is the refractive index, typically valued between $\frac{1}{3}$ and $\frac{1}{5}$, and k_f is the radiative deactivation rate constant [41].

When deciding on a donor-acceptor pair, the effective range must be taken into consideration. Organic fluorophores and fluorescent proteins typically have an R_0 value around 5 nm yielding an effective range of 3 – 8 nm which corresponds to the 5 – 95 % range of E where changes can still be detected sensitively [38]. However, due to experimental limitations, such as noise, the useful range is around 4 – 7 nm. Therefore, when choosing a donor-acceptor pair with a known R_0 value, it is necessary to calculate the effective/useful ranges to ensure the likelihood of obtaining observable FRET efficiencies.

Another consideration when deciding upon a donor-acceptor pair is the type of FRET experiment to be performed. There are four general categories FRET experiments can fall into: (1) methods that monitor changes in donor fluorescence, (2) methods that examine changes in acceptor fluorescence, (3) methods that simultaneously measure

changes in both donor and acceptor fluorescence and (4) methods that monitor changes in the orientation of the fluorophores [42]. Depending upon which type of experiment is chosen, the photostability of the fluorophore may need to be high or low.

1.3.2 Discovery of fluorescent proteins (FPs) and FRET applications

After it was demonstrated that the GFP could be fused to target proteins and expressed inside other organisms, a number of both intramolecular and intermolecular FRET applications became possible [32]. One active application of intramolecular FRET monitors conformational changes and involves sandwiching a protein between two fluorescent proteins, typically CFP and YFP [17]. For example, a target protein may be oriented in a way that keeps CFP and YFP more than 80 Å from one another or oriented in a certain way, but upon binding of a ligand or phosphorylation there is a subsequent conformational change that brings the two fluorescent proteins closer together or changes the orientation of the chromophores and FRET becomes observable.

Another active application of intramolecular FRET is the monitoring of protease activity within a cell. This is achieved by linking two fluorescent proteins with a short region with the known protease cut site. At first, FRET should be observed since the two fluorescent proteins are near to one another, but depending upon the rate of protease activity, the FRET signal should slowly lessen as the active site is cut and the two fluorescent proteins drift apart from one another [18].

One of the most common active applications of intermolecular FRET is the study of protein-protein interactions, provided the fluorescent proteins get within 6 – 8 nm of

one another [17, 18, 42]. Two target proteins are independently fused with different fluorescent proteins (i.e., CFP and YFP) and if FRET is observed it can be assumed that the two proteins interact with one another. The big problem with studying intermolecular FRET is that the ratio of donor and acceptor expression is no longer fixed since they are not fused to the same protein. Endogenous proteins can also interact with the fused proteins and reduce the amount of target proteins available for FRET [18].

1.3.3 Limitations of FPs

There are several limitations of the fluorescent protein technology. First, GFP contains only one fluorophore and it is not very bright. For low expressing fusion proteins, this becomes a problem since the fluorescence signal is weak and difficult to detect. Another limitation is the innate environmental insensitivity of the chromophore caused by the shielding of the barrel structure. This prevents detection of minor changes of the proteins environment as it is trafficking or undergoing conformational changes [32]. The fluorescent proteins large size and possibility that fusion may affect the protein's function is another consideration [17].

When performing FRET experiments there are additional concerns. First, colocalization does not necessarily indicate protein-protein interaction. Resolution limits of conventional light microscopy prevents accurate determination of nearness versus protein-protein interactions. FRET efficiencies measured in cells are often an

ensemble measurement arising from both specific protein-protein interactions and from random associations [42].

1.3.4 Advantages of small molecule fluorophores in FRET

Small molecule fluorophores are advantageous to use in FRET experiments for several reasons [37]. As mentioned previously, these fluorophores are small (<1 kD) and covalently attached to proteins without disturbing the structure. There are also much brighter than fluorescent proteins are more easily detected when performing FRET experiments. Since they are so small, accurate distance measurements can be calculated. Another major advantage to small molecule fluorophore is that they can be attached to small proteins that would otherwise be overwhelmed by a large fluorescent protein.

1.3.5 Synchronous fluorescence scanning spectroscopy

Synchronous scanning fluorescence spectroscopy is a technique that takes advantage of the ability to vary both the excitation and emission wavelengths simultaneously during data collection. In this technique, the fluorescence signal is recorded when excitation and emission wavelengths are simultaneously scanned keeping in between a fixed wavelength interval (called the offset value, $\Delta\lambda$) throughout the spectrum. As a result, the selectivity for individual fluorescent components is considerably improved; additionally, much more information on mixtures of fluorescent compounds is gained [43, 44].

The tryptophan residues of a protein could be considered “fluorescent components” as they are each in their own unique environment and could possibly have unique spectral patterns. The immediate chemical environment of tryptophan residues, such has being involved in hydrogen bonding or hydrophobic interactions, can have effects on its fluorescent properties. These interactions could result in slight red or blue shifts of its emission spectrum, causing different wavelengths of the FRET-peaks for different tryptophan donor/IAEDANS acceptor pairs [45].

It might become possible to create a protein with multiple donors and a single acceptor and obtain information about each donor/acceptor pair from one measurement. These wavelength differences may possibly allow us to assign the FRET-peaks to an individual FRET donor/acceptor pair, if these synchronous scanning fluorescence spectra can separate the individual FRET-peaks.

1.3.6 Can we use a small molecule fluorophore labeled protein for in-cell FRET measurements to obtain atomic resolution distances between fluorescence donors and acceptors?

Ideally, in order to obtain structural information from proteins inside living cells, using fluorescence spectroscopy or imaging, they need to be labeled with small molecule fluorophores since fluorescent proteins are large, independently folded proteins separate from the target protein. However, currently there is no available technique that allows specific labeling of a particular protein inside the cells with small molecule fluorophores. Small molecule fluorophores may specifically adhere to a single

amino acid *in vitro*, like cysteine or lysine, but they are unable to specifically label a target protein while ignoring the other intracellular proteins. A few advances have been made, with respect to targeting small molecule fluorophores to specific proteins within a cell (i.e., “FlAsH” and “ReAsH”) but this requires a genetic insertion of a specific sequence into the target protein that these fluophores can target [17].

On the other hand, if there was a way to label purified target protein *in vitro* and introduce them efficiently into the target compartment of living mammalian cells at a sufficient concentration, then performing in-cell FRET measurements should be possible. Since proteins only perform their function inside the correct intracellular compartment, the target capability of the protein delivery into the correct intracellular compartment is essential for this approach. In-cell FRET would require specific excitation of a donor molecule on the target protein, followed by an observable emission of the acceptor molecule also on the target protein. In recent years, there have been several advances made in the field of exogenous protein delivery making this application possible.

1.4 Exogenous protein delivery inside mammalian cells

The ability to deliver exogenous proteins into the specific intracellular compartment of mammalian cells is of great interest for several reasons. First, protein based therapy would be available for patients suffering from genetic diseases resulting in non-functional proteins. It would be more beneficial and less risky than viral gene delivery. In my case, I would like to selectively label bacterially produced proteins with

probes and then deliver this specifically labeled protein into the correct intracellular compartment of mammalian cells and perform live cell fluorescence spectroscopic studies. This allows me to gain structural information of the labeled proteins within live mammalian cells.

1.4.1 The CPP-based protein delivery technology and its limitations

Protein transduction is a technique that delivers proteins into living cells. It emerged after the discovery of the cell penetrating peptides (CPP) [46, 47]. These CPPs are small peptides with the ability to enter cells *via* an unconventional way, although their transduction mechanism is still debatable [48, 49]. Fusion of a CPP with proteins/DNAs/RNAs allows their intracellular delivery [48-50]. Protein transduction was galvanized by a report on the ability of CPP to deliver β-galatosidase to multiple tissues, including the liver, spleen, lung, heart and brain, in mice [51]. Efforts have been made to pursue non-peptide protein delivery reagents. Polyethyleneimine (PEI) is found to have the ability to deliver protein and DNA intracellularly [8, 52]. A small-molecule mimic (SMoCs) of CPP has been reported to have a similar protein delivery property [53].

Despite these notable successes, protein delivery technology has yet to become commonplace for biomedical applications [46, 47]. The CPP-fused proteins share common problems. The CPP-fusion changes protein sequence and intracellular proteases likely degrade the delivered proteins, if they are not folded properly, before they reach their target intracellular compartment. The CPP-fusion also lacks targeting capability to specific intracellular compartments, significantly restricting their

applications. It remains unknown if the intracellular folding machinery can refold the CPP-delivered bacterial expressed proteins. Blobel's "signal theory" guides the fate of endogenous proteins, dictating their intracellular locations and trafficking [54, 55]. Questions remain regarding whether the CPP-delivered proteins follow the same intracellular trafficking/secretion pathway inside cells. These are critical questions regarding the physiological relevance of protein delivery technology.

1.4.2 The QQ-protein delivery technology

Recently, our lab has developed a QQ-reagent based protein delivery technology that has solved the problems related to the CPP-based technology [Li 2011]. The QQ-protein delivery has several novel features, including non-covalently association with proteins, protection from intracellular protease degradation and the target capability to specific intracellular compartments. These features enable the delivered proteins to be indistinguishable from their endogenous counterparts by the cell machinery. It is further demonstrate that the intracellular folding machinery properly refolds the delivered proteins and the refolded proteins follow the same trafficking pathway as their endogenous counterparts. Indeed, QQ-protein delivery provides new tools in cell biology studies, allowing one to introduce specific labeled proteins inside the cells for high-resolution biophysical studies of these proteins at the molecular level.

The QQ-reagent is a cocktail of several commercially available compounds, making this protein delivery technology easily accessible and affordable. The QQ-protein delivery only requires an incubation step of the QQ-modified proteins with cells. It

enables delivering multiple proteins either simultaneously or consecutively. This allows one to study protein-protein interaction in a consecutive way once these proteins are labeled in different methods and consecutively QQ-delivered into the correct intracellular compartment of the mammalian cells. The second novel feature of the QQ-reagent is its high delivery efficiency and the delivered proteins can be easily detected by a SDS-PAGE. Although the mechanism of the QQ-protein delivery is unknown, this is a significant result since high delivery efficiency is critical for applications of a protein delivery technology.

The intracellular toxicity of the QQ-protein delivery is minimal. Another novel feature of the QQ-reagent is the non-covalent modification feature, allowing for dissociation of the QQ-reagent from the delivered proteins inside cells. The QQ-reagent protects the delivered proteins from intracellular protease degradation, solving the major problem of the current protein delivery techniques. Since the delivered proteins are coated with non-covalently associated QQ-reagent, they are initially camouflaged from the intracellular proteases and cellular machinery even if they are misfolded. Once inside the cells, the non-covalent association nature permits QQ-reagent dissociation from proteins, thus, the QQ-delivered proteins become “naked” proteins similar to the endogenous proteins. Therefore, the QQ-delivered proteins can reach their target compartment based on their sequence localization signals. This allows the cellular folding machinery to properly refold the misfolded, QQ-delivered proteins inside the mammalian cells. More importantly, this design of non-covalent association allows the

QQ-protein delivery to have a targeting capability to specifically deliver proteins into their intracellular destinies.

These novel features of the QQ-protein delivery enable the delivered proteins to be indistinguishable from the endogenous proteins by the cell machinery. Once inside cells, the cell's machinery functions as if the QQ-delivered proteins were the endogenous counterparts. The development of this critical technology is the necessary step needed to introduce a model protein labeled with a FRET donor and acceptor pair to obtain in-cell FRET measurements. The model protein, MESD, is reviewed in the following section.

1.5 MESD

1.5.1 LDLR Family

The low-density lipoprotein (LDL) receptor (LDLR) family members are utilized in many biological processes, including cholesterol homeostasis, neuronal migration and pattern formation during development. LDLR members act in two main ways – as endocytic receptors or as receptors in signaling pathways. Variant LDLRs or their ligands can contribute to several major human diseases, including hypercholesterolemia, atherosclerosis, bone diseases and developmental and neurodegenerative disorders such as Alzheimer's disease. In the Wnt/Wingless (Wg) signaling pathway, the mammalian LDLR-related proteins 5 and 6 (LRP5/LRP6) are essential co-receptors for binding to Wnts, controlling many aspects of animal development [56].

1.5.2 The modular organization of the LDLR family

The structural organization of the LDLR family is complex and contains several common modular structures, including: (1) complement-type repeats (Type A repeats), (2) epidermal growth factor repeats (EGF) and (3) YWTD repeats which display a six-blade β-propeller structure. The receptors are anchored in the plasma membrane by a single trans-membrane domain, followed by a cytoplasmic domain that contains signal sequences for endocytosis and interaction motifs, such as the NPXY motifs, for binding to cytoplasmic adaptors and scaffolding proteins.

The modular structures are complex and provide a challenging task for the cell in regards to their correct folding. Specifically, they contain many cysteines which form specific intramolecular disulfide bonds which contribute to the overall 3-dimensional structure; in addition to the hydrophobic residues of the ligand binding region that could potentially cause aggregation. Failure to achieve the native conformation or protein aggregation can cause many diseases. The presence of molecular chaperones in the ER prevent these problems from occurring.

1.5.3 Specific chaperones

Like most transmembrane and secreted proteins, LDLR family members enter the secretion pathway as they are translated by endoplasmic reticulum (ER)-associated ribosomes. These proteins fold and mature inside the ER and traffic from the ER through the Golgi apparatus on their way to the cell membrane. Effective quality

control systems in mammalian cells ensure that only the properly folded and matured proteins can be exported from the ER and the misfolded proteins will be retained in the ER and removed by ER-associated degradation (ERAD) [57]. Numerous chaperones and folding enzymes function in the ER to ensure proper folding and maturation of most proteins [58]. For the LDLR family, two specialized chaperones have been identified.

The receptor-associated protein (RAP) is an ER-resident chaperone that is necessary for efficient folding and ER export of some LDLR family members [59]. RAP also escorts the receptor trafficking from the ER to the Golgi and prevents premature association of ligands that are also expressed in the same compartment with the nascent receptor and thus may interfere with proper folding and trafficking of the receptors [57].

Another specialized chaperone, termed *mesoderm development protein* (*MESD*) in the mouse and *boca* in the fly, is essential for the Wg/Wnt signaling [60, 61]. Flies or mice with nonfunctional *boca* or *mesd* gene die during embryogenesis and display phenotypes that are consistent with an inability to transduce a Wg/Wnt signal. Interestingly, MESD/Boca does not function as a direct component of the Wg/Wng signaling pathway. Instead, it functions as a molecular chaperone inside the ER specifically for proper folding and export of the LDLR family, including LRP5 and LRP6 [57]. In the absence of MESD, LRP5/LRP6 fails to reach the cell surface and remains sequestered as insoluble aggregates due to misfolding.

1.5.4 MESD: Sequence, structure, mechanism

Our laboratory recently solved the NMR structure of MESD (Figure 1-7), which includes two structural domains: a central core domain and a C-terminal flexible helical domain [62]. Mutagenesis and functional data indicates that the central core domain is the chaperone domain which promotes LRP5/6 folding and maturation inside the ER, whereas the C-terminal flexible helical domain is the escort domain that safe-guards the properly folded receptor trafficking from the ER to the Golgi.

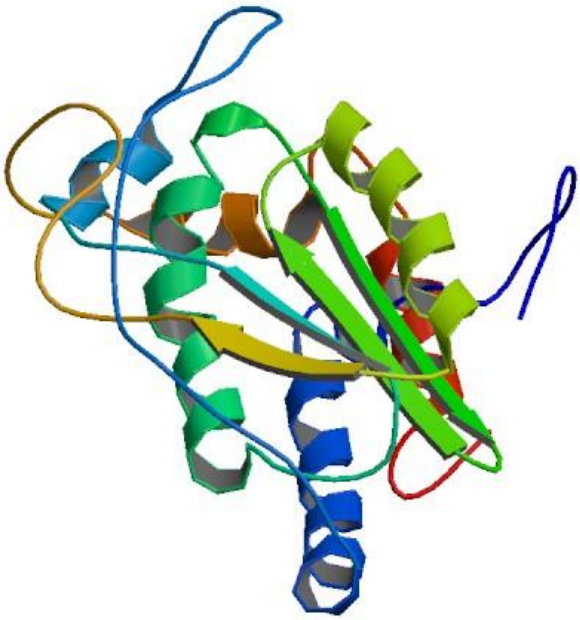


Figure 1-7: NMR structure of MESD (solved by Chen 2011, PDB# 2KGL).

Looking at Figure 1-8, we can see that residue W32 is highly conserved amongst all species. Mutation of this residue is fatal and development does not get past the

embryonic stage. In addition, all species have conserved lysine residues in both the C- and N-termini that are important for MESD's chaperone and escort functions. Finally, each species ends in an ER retention signal (KDEL for the fly and *C. elegans* species and RDEL for mammals) [62].

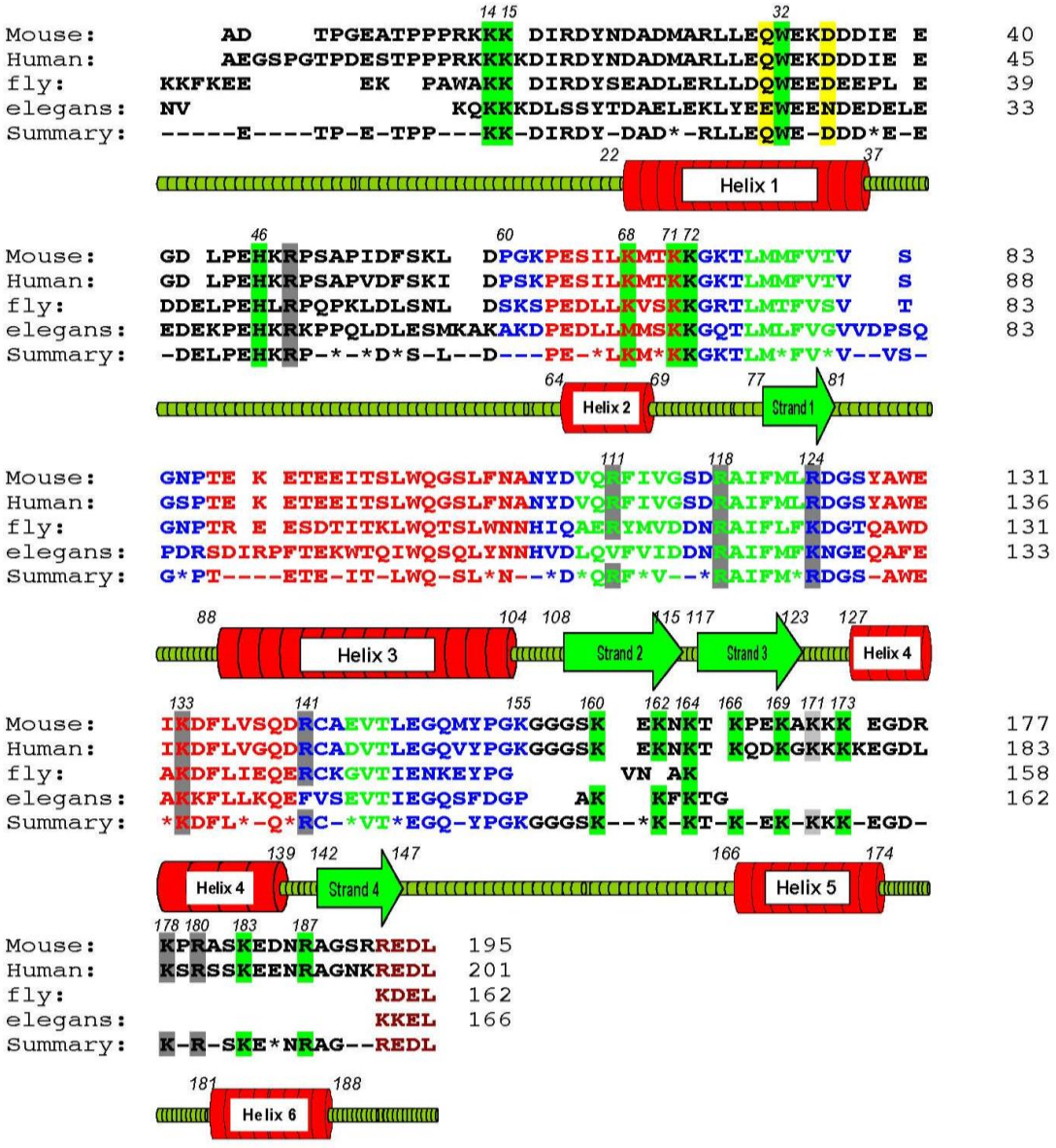


Figure 1-8: Sequence Alignment of MESD from Four Species. The secondary structure of MESD is displayed under the sequence based on our NMR structure. In the sequence

alignment, the critical receptor-binding residues in both the central core domain and the C-terminal flexible helical domain are highlighted in green. For residues that are involved in the binding pocket of W32 are highlighted in yellow. The other conserved Lys/Arg residues are highlighted in grey. The sequence of MESD(60-155) is shown in blue with the helical residues colored in red and β-strand residues in green.

1.5.5 MESD promotes folding of BP domain of LRP5 and LRP6

A cell biology picture of the MESD/LRP5/6 pathway, shown in Figure 1-9 suggests that it is the specialized chaperone MESD that determines proper folding of the BP domain of LRP5/6, whereas the BP domain strategically regulates structural switches of the two structural domains of MESD in a unique fashion to ensure both proper folding and safe trafficking of the receptor along the secretory pathway, as well as the ER-retrieval of MESD protein. Chen et al suggest that the escort function may be a recent evolutionary acquisition of these chaperones, since Boca, the Drosophila ortholog of MESD, lacks the C-terminal escort domain. By strategically placing fluorescent probes to serve as donor and acceptors, it might be possible to prove the exact point of interaction between MESD and the BP domain of LRP6/6.

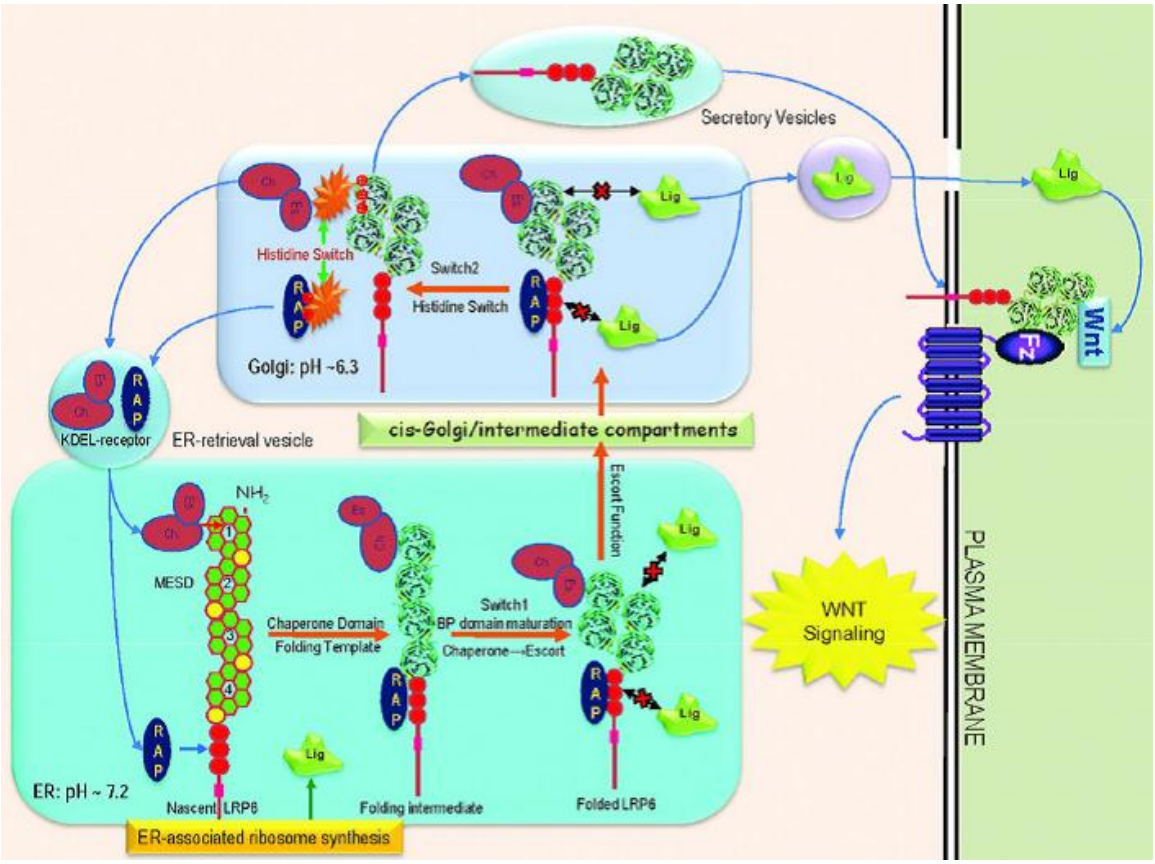


Figure 1-9: The MESD/LRP5/6 Pathway. The rigid chaperone domain (Ch) of MESD binds to the newly synthesized BP domain of LRP6, serving as a folding template. After BP domain properly folded, MESD switches the binding from the rigid chaperone domain to the flexible escort domain (Es), safely guarding the mature receptor traveling from the ER to the Golgi, preventing premature ligand (Lig) binding. The acidic environment of the Golgi activates the histidine switch in the BP domain that leads to the dissociation of MESD from the receptor. MESD will be retrieved back to the ER by the KDEL-receptor. The properly folded receptor will be properly post-translationally modified and further reach the cell membrane for activation of canonical WNT pathway. RAP goes through a

similar cycle for promoting the ligand-binding domain folding and trafficking as described before [57].

1.5.6 Unanswered biological questions

1.5.6.1 MESD ER concentration versus LRP5/LRP6 folding

As mentioned previously, chaperones and folding enzymes are present in the ER to prevent misfolding and subsequent degradation by ERAD. During the unfolded protein response (UPR), protein translation is halted and there is upregulation of the production of molecular chaperones. It is unknown at what concentration MESD exists in the ER of cells or to what level it is elevated during the UPR. How much MESD is necessary to ensure LRP5/LRP6 are folding efficiently? Currently, there are no methods available to determine the intracellular concentration of a protein inside live cells.

1.5.6.2 MESD ER and Golgi structures versus NMR structure

MESD has two distinct domains that serve as a chaperone and then as an escort. It is thought that the change of environments from the ER to the Golgi causes the change from chaperone to escort [62]. Since the Golgi is a more acidic environment than the ER it is possible that there is a structural change. There are inhibitors available that will inhibit transport from the ER to the Golgi and vice versa. It is possible, therefore, to create a batch of mammalian cells with a strict population of MESD located in the ER only and a separate population of MESD located in the Golgi only. Measurements could

be taken in these separate compartments to determine if there is a difference in distance calculations indicated a structural change between the ER and Gogli.

1.5.6.3 MESD serves as a model protein for our technique to address unanswered questions

MESD serves as an excellent model protein for this thesis due to a number of reasons. First, it has three tryptophan residues that can act as FRET donors. By using a native residue, there will be no need to add an additional probe to serve as the donor. It is also easy to mutate one or two tryptophan residues at a time to reduce the number of donors, since the simplest measurement involves one donor and one acceptor. Second, the sequence contains one cysteine residue. There are a number of commercially available small molecule fluorophores that are thiol reactive and will covalently bind to the cysteine residue and serve as the FRET acceptor. Third, the MESD sequence contains an "RDEL" ER-retention signal. This ensures that after the protein has been delivered inside the mammalian cell using the QQ-protein delivery technique, it will traffic to the ER and remain there.

Once several basic methodologies have been established, we can also use the MESD protein to answer several questions pertaining to MESD's structure and function within the ER and Golgi. The first step into answering these questions is the production of a large quantity of pure protein to be used in these structural studies. Bacterial expression is still one of the best methods for protein production on a large scale.

1.6 Bacterial protein expression

1.6.1 Recombinant protein production

Recombinant protein expression allows for the production of large quantities of protein to be used for laboratory study or industrial purposes, typically therapeutics. Depending upon the final purpose of the protein, scientists choose a host organism to insert recombinant DNA into that will ultimately produce their target protein in a controlled fashion. There are many options available for host organisms, including both prokaryotic and eukaryotic cells, and there are advantages and disadvantages of each type of host.

1.6.1.1 Prokaryotic expression systems

Bacterial expression systems are the most attractive organisms for heterologous protein production for a number of reasons. The ability to reproduce rapidly and grow at high densities results in high yield protein production, as compared to eukaryotic systems. In addition, bacteria can generally be grown in inexpensive medium and a large amount of protein can be produced in one day. Bacterial hosts typically have well characterized genetics that can be manipulated to create even better hosts. Finally, due to all of the mentioned advantages, there are a multitude of commercially available products, including bacterial strains, bacterial expression vectors and protein purification materials, to aid in heterologous protein production utilizing bacterial expression systems [63-66].

Escherichia coli remains the most commonly used bacterial host for all the above mentioned reasons. Expression of eukaryotic proteins inside a prokaryotic cell does not always result in a properly folded and active protein. However, techniques to optimize heterologous protein overproduction in *E. coli* have been explored that significantly enhanced the yield of the foreign eukaryotic proteins. Two recent excellent reviews summarized these optimizations (Table I in Refs. 2 and 3). Some empirical "rules," for host strain selection, plasmid copy numbers, promoter selection, mRNA stability, and codon usage, have been derived from these optimizations that can be used to guide the design of expression system and to limit the unpredictability of protein expression in E. coli [64, 65]. However, an important optimization is cell growth conditions and media, which seems to be target protein dependent and there does not seem to be any empirical rules reported to date in this aspect [5].

Although *E. coli* is the most popular host, other bacteria have been utilized as well. Kay Terpe provided an informative review of all the various bacterial systems that have been used over the years [66]. Gram-positive bacteria such as *Bacilli* strains, *Streptomyces*, and *Staphylococcus carnosus* have been used since they typically have lower protease activity, high secretion capacity and have produced protein in a soluble form when *E. coli* was unable to do so. Gram-negative bacteria such as *Caulobacter crescentus*, *Methylobacterium extorquens*, *Pseudomonas fluorescens* and *Ralstonia eutropha* have been used to take advantage of their unique properties. These properties include the ability to secrete target proteins into media, growth on a single low cost substrate, production of highly enriched proteins with stable isotopes and

remarkably high yields of target proteins. These bacteria provide several advantages over an *E. coli* system, but due to their rare use, *E. coli* still proves to be a more economical choice.

1.6.1.2 Eukaryotic expression systems

In contrast, eukaryotic expression systems are utilized when a full length, properly modified and active protein is required. A eukaryotic host, such as mammalian cell culture systems, has the necessary machinery to properly fold and post-translationally modify target proteins. Mammalian cells lines produce proteins most closely resembling those made by human cells in the body. However, maintenance of mammalian cell lines is complicated, costly and protein yield is typically low [4, 67, 68]. In recent years, yeast systems such as *Saccharomyces cerevisiae* and *Pichia pastoris* have been utilized with moderate success. Yeast has the advantages of both prokaryotic and eukaryotic expression systems. They rapidly grow and are of lower maintenance and cost, like prokaryotic systems, while having the added advantage of improved folding and post-translational modifications [67]. However, there are problems that include but are not limited to hyperglycosylation and reduced secretion [68].

Protein production can also be performed in insect cells using the baculovirus system. The target gene is inserted into the baculovirus genome and insect cells or larvae are infected with the mutated virus [69]. Protein production is similar to that of mammalian cells, in that the insect cells will properly fold and post-translationally modify the protein. The main advantage of the insect system is the robust nature and

inexpensive cell culture [4].There are a few disadvantages of the insect system that prevent it from being more widely used. Specifically, there are complications from the baculovirus that lead to proteolytic activity and impaired protein production [68].

Filamentous fungi and plants have also been utilized for recombinant protein productions with limited success. Filamentous fungi have a more advanced post-translational modification complex, similar to mammals. However, little is known about their genome and metabolic pathways that there has been successful production of only a few proteins [68]. Plants cells have also been used sparingly as they have problems with glycosylation and proteolysis [64].

1.6.1.3 Cell-free expression systems

Another alternative to bacterial expression is the utilization of cell-free systems. In order to create a cell-free system, cells of bacterial, plant or mammalian origin are grown to a certain optical density. Afterwards, the cells undergo a series of centrifugation and lysis ending in the isolation of the subcellular protein producing machinery [70]. Cell-free systems are advantageous for a number of reasons. First, they solve the problem of eukaryotic proteins that are toxic to bacterial cells. Second, they have increased production since all the metabolic resources will be singly focused on protein production [71]. Finally, cell-free systems that have been isolated from eukaryotic cells have the ability to post-translationally modify the target protein [4]. Another recent advancement has shown that by addition chaperonins and glutathione redox buffer improved the folding yields of recombinant proteins produced with a

bacterial cell-free system [71]. Many of the problems associated with cell-free systems, such as low protein yield, have been solved by optimizing lysate preparation and experimental conditions. The last remaining hurdle is the expensive nature of purchasing commercially available kits or creation of in-house lysates and reaction buffers [4, 70, 71].

1.6.2 Heterologous protein expression in *Escherichia coli*

Once the decision has been made to produce a target protein in a lab, several decisions must be made. First, should the target protein be produced in bacteria, yeast or mammalian cells? Second, what type of vector should be used? It is suggested that *E. coli* be used as the expression host when first attempting expression due to its rapid results and inexpensive nature [5]. In order to express recombinant proteins inside of bacteria, a gene construct must be made that includes the target gene inserted in a plasmid containing a regulated promoter region. The plasmid also contains an affinity tag that will be attached to the target protein that allows for purification after expression.

The pET expression vector is a powerful system developed for the cloning and expression of recombinant proteins in *E. coli.* Based on the T7 promoter-driven system originally developed by Studier and colleagues, Novagen's pET System has been used to express thousands of different proteins [72-74]. In pET vectors, target genes are cloned under control of strong bacteriophage T7 transcription and translation signals, and expression is induced by providing a source of T7 RNA polymerase in the host cell. The

host cell typically has the T7 RNA polymerase gene in the host chromosome under *lacUV5* control and therefore is not expressed until an inducer is added. The *lacUV5* promoter is a mutated version of the *lac* operon promoter with decreased basal activity. Inducing expression of the T7 RNA polymerase with lactose, or a lactose analog like IPTG, results in over-expression of the target gene.

In addition to efficient vectors, several bacterial strains have been developed to facilitate large quantity production of recombinant proteins and they are commercially available. For example, a BL21(DE3) bacterial strain, developed by Brookhaven National Laboratory, knocks out intracellular protease expression inside this bacterial strain [75]. This solved the problem associated with protease digestions of the expressed recombinant protein, significantly enhancing the yield of recombinant proteins. To overcome codon bias and intracellular toxicity of the recombinant proteins, several BL-21(DE3) bacterial strains, including BL-21(DE3)pLys and BL-21(DE3) CodonPlus, have been developed [65, 66].

If the target protein is produced without any biophysical probes, the bacteria are grown in rich medium to a desired optical density. The inducer is added and the culture continues to grow for a specific period of time. The cells are harvested and the protein can be purified based on the affinity tag used.

1.6.3 Protein production in minimal medium

Structural biologists often need to specifically label recombinant proteins with biophysical probes for their studies. Proteins may be labeled with selenomethionine for

X-ray crystallography studies or stable isotopes ($^{2}H/^{13}C/^{15}N$) for NMR studies amongst other possibilities. Using any expression system besides a bacterial expression system is highly inefficient and expensive.

The standard protocol for specifically labeling proteins with biophysical probes involves first growing bacterial cells in a rich medium, followed by transfer to minimal medium. Minimal medium has one specific carbon source (glucose) and a single nitrogen source (ammonium chloride or ammonium sulfate) in addition to various salts, minerals and metals. The culture is grown until a desired density is reached, then an inducer is added that allows for expression of the target protein. However, expression in minimal medium often results in lower protein yields as compared to growth and expression in rich medium. This is especially true when attempting to make triple-labeled proteins for NMR studies, which requires growth in D_2O-based minimal medium.

Minimal medium is also necessary to specifically label proteins with amino acid analogues or isotopically labeled amino acids. The problem of "scrambling" arises since bacteria have the ability to synthesize all twenty amino acids. To ensure that bacteria do not synthesize their own amino acids, auxotrophic bacterial strains are used. Auxotrophic bacterial strains usually contain some type of genetic mutation inhibiting biosynthesis of a specific amino acid(s). Minimal medium supplemented with the specific amino acids must be used even when working with an auxotrophic strain since rich medium will contain the standard amino acids.

1.6.4 Auto-induction method

Recently, Studier introduced an auto-induction bacterial expression method, which provides several advantages over the standard IPTG induction method, including:

(1) Achieving a high cell density (leading to higher target protein production and
(2) Minimal handling as there is no need to monitor cell growth for induction [76].

In order for auto-induction to occur, the growth medium must contain varying amounts of glucose, glycerol and lactose. The bacteria will utilize glucose initially as its energy source and naturally repress induction despite the presence of lactose. Once the glucose has been depleted, induction will occur since the bacteria will use glycerol as its energy source and lactose will remove the repressor protein that is preventing induction. An inducer molecule, like IPTG, is not needed since the lactose metabolite, allolactose, is the native molecule necessary to remove the repressor protein from the lac operon. This method is extremely low maintenance, as it is only necessary to inoculate the culture medium and wait for the culture to saturate.

Studier and others devised various recipes for different auto-induction medium, depending on the final use of the target proteins, including auto-induction minimal medium for proteins to be used for structural studies. This method has been used to prepare $^{13}C/^{15}N$ double-labeled proteins for NMR studies and selenomethionine-labeled proteins for X-ray crystallographic studies, both produced a moderate yield of target

proteins (~40 mg/L) [77, 78]. Our lab also developed a modified recipe for the production of triple-labeled proteins ($^{2}H/^{13}C/^{15}N$) in D_2O-based auto-induction minimal medium [79].

When using the auto-induction method to produce isotopically labeled proteins in our lab, we noticed a few problems. First, a high cell density culture did not guarantee high yield protein production. Second, when expressing proteins in the auto-induction minimal medium, expression times could take up to several days, especially with D_2O-based medium. Another issue was the expensive ^{13}C-glycerol that is a key ingredient in auto-induction method. All of these problems lead us to re-evaluate our expression method and develop a method that produced a consistently high yield of target proteins [79].

1.6.5 High cell-density bacterial expression method

In order to solve the problems we encountered while using the auto-induction method, we developed a bacterial expression method that maintains the advantage of the tightly controlled induction by IPTG and utilizes both rich and minimal media to achieve a very high cell density for production of a very high yield of recombinant proteins. Unlike the auto-induction method that incorporates minimal medium, our high-cell-density method does not require long time durations for achieving a high cell density, which is much more time efficient. This method starts with a cell culture grown in rich medium that allows for a significantly enhanced initial cell density at the OD_{600} values of 3–7 before IPTG induction, depending on the rich medium used. After

switching the cells into the minimal medium, the bacterial cells are cultured at a previously optimized temperature for 1.0–1.5 h and induced with IPTG for protein expression. With both auto-induction and the high cell density IPTG-induction methods, the final cell density before cell harvest can reach to OD_{600} of 10–20, resulting in very high yields of protein production [79].

1.6.6 Three critical protocols to ensure a very high yield production of pure recombinant proteins

During the course of developing our new hybrid expression method, we developed three critical protocols that would ensure consistent, high yield protein production. First, expression must begin with a proper starting culture. Typically, starting cultures are grown in rich medium overnight ending in a saturated culture that will be diluted in minimal medium for expression. We found that the best results were obtained when a starting culture was only allowed to grow to mid-log phase of its growth curve. Second, double colony selection must be performed to ensure a colony with a stable, high expressing plasmid has been chosen. It is standard practice to screen colonies after transformation to choose a high expressing colony, but we found that performing a second round of selection resulted in a more reliable glycerol stock. Third, optimization of induction temperature, time and IPTG concentration must be performed on every protein as all of the variables are protein dependent and can vary widely.

1.6.7 High level expression as used in my research

In order to perform *in vivo* FRET experiments, my target protein must be labeled with an amino acid analogue which requires growth in minimal medium. The first advantage to using our new hybrid method is the repeatability of the results. We found that after performing double colony section, we never had a problem repeating the high yield expression as compared to the auto-induction method. The second advantage is the extremely high yield of protein production directly due to the high cell density of the culture. Third, since we first grow the culture in rich medium before transferring to minimal medium, we can obtain a high yield of protein production typically in a single day. The auto-induction method using minimal medium can take up to three or four days. The hybrid method we developed provides us with consistent, high yield target protein. High level expression ensures that I could efficiently produce enough protein to be used in both the *in vitro* and *in vivo* experiments.

1.7 Selective labeling of amino acids in recombinant proteins

1.7.1 Uses for selective labeling of amino acids

In structural biology, two of the atomic resolution techniques take advantage of selective labeling. When performing X-ray crystallography, the multi-wavelength anamolous diffraction [8] method requires the target protein to be labeled with an atom that will absorb or scatter X-rays in a specific way [80]. Most commonly this is achieved by incorporating selenomethionine (Se-Met) into the protein. Se-Met is identical to the

original amino acid structure except for the replacement of the sulfur atom with selenium. Selenium shares chemical properties with sulfur, but causes unique diffraction patterns at various wavelengths. Incorporation of Se-Met into a protein does not change the protein's structure or function. However, the selenium atom that has been incorporated into the amino acid helps determine the phase of the crystal structure by measurable changes in the diffraction pattern and therefore allows for structure determination [80]. Prokaryotic expression systems are the most commonly used to produce proteins labeled with Se-Met and can typically achieve 90-100% Se-Met incorporation. Yeast and insect cells systems have been used, but Se-Met incorporation is much lower and the selenomethione can be toxic to the host cell [9].

NMR experiments require target proteins to be isotopically labeled with isotopes containing NMR active, nuclear half spins (i.e., ^{1}H, ^{13}C, ^{15}N and ^{19}F). Depending upon the size of the protein and what structural information is desired, a protein can be uniformly labeled with general isotopes, like ^{13}C and ^{15}N. This is achieved by growing the bacterial culture in minimal medium that contains a ^{13}C-carbon source, typically ^{13}C-glucose, and a ^{15}N-nitrogen source, typically ^{15}N-ammonium chloride. This ensures that all the carbon and nitrogen atoms have an active spin that can be utilized by specific pulse sequences.

If the target protein is over 25 – 30 kD, suffers from severe spectral overlap or if you want to only see specific changes in a particular region of a protein, like an active site, the protein can be selectively labeled with isotopic amino acids. There are several

methods available to achieve this goal. Originally, selective labeling was achieved by growing the auxotrophic bacterial cells in a defined minimal medium, excluding the nitrogen source and supplementing with all 20 amino acids including the labeled amino acid of interest [81]. More recently, cell-free systems are being widely used as well as alternative bacteria, like *Brevibacillus choshinensis* [82]. Selective "unlabeling" is another alternative which allows the protein to be uniformly labeled with ^{13}C and ^{15}N with the exception of certain unlabeled amino acids that are supplemented in the medium [83]. Spectra collected from a selectively unlabeled sample is compared to a uniformly labeled sample, the missing crosspeaks can then be assigned.

1.7.2 Potential problems for amino acid selective labeling

Whether the protein will be used in an X-ray experiment or an NMR experiment, there are a few problems involved with selective labeling. First, the protein yield will be reduced when grown in the required defined minimal medium. This is especially true when selectively labeling with amino acid analogues. The analogues do not fit as well into the specific pocket of the tRNA synthetase.

A second problem is the fact that *E. coli* can produce all 20 amino acids and the biosynthetic pathways are redundant at times (Figure 1-10), this can lead to one of two results: (1) the breakdown of the labeled amino acid and incorporation into another amino acid, called "scrambling" or (2) the production of the native amino acid from other supplemented native amino acids. The first result occurs either because a

particular amino acid is a direct metabolic precursor to another amino acid or is susceptible to aminotransferase [84]. For example, if you were trying to label a protein with an isotopic asparagine residue, that residue could be broken down and incorporated into aspartate (and all the other amino acids it is related to) or the bacteria could use the normal aspartate that is provided in the defined minimal medium and produce non-labeled asparagine itself.

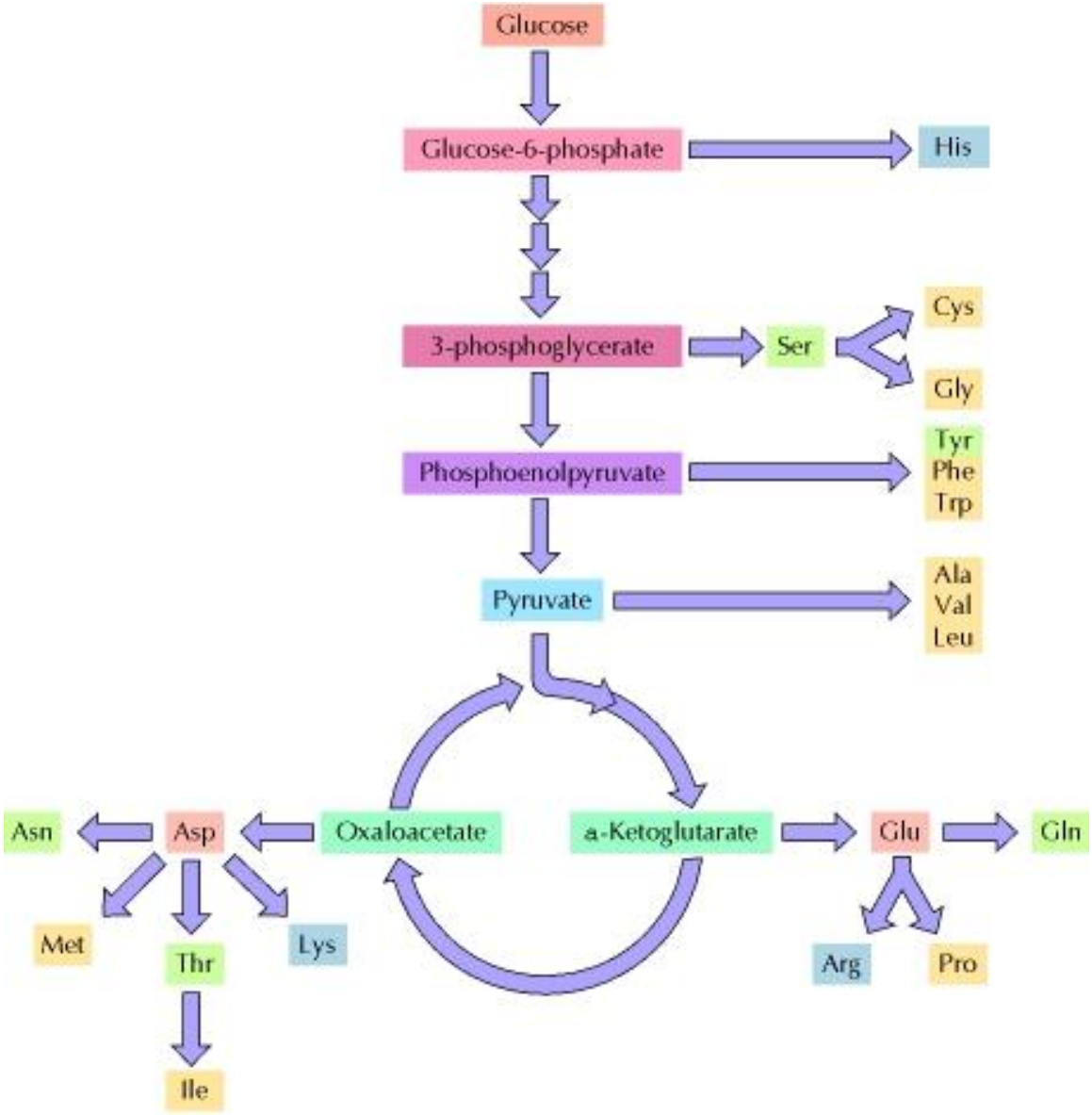


Figure 1-10: Amino acid biosynthesis pathways [85].

1.7.3 Auxotrophic bacterial strains

One way to prevent scrambling is to inhibit the biosynthetic production of the amino acid in question during induction [86]. However, inhibition is not 100%; therefore incorporation of the amino acid analogue is not 100% but can be as high as

greater than 90% [9]. In order to achieve 100% incorporation and prevent scrambling, auxotrophic bacterial strains have been developed by various labs [9, 84].

Auxotrophic bacteria strains are typically created by altering the genome in some way that an important enzyme in an amino acid's biosynthetic pathway is rendered nonfunctional. Depending upon where the enzyme acts in the pathway, several amino acids could be affected or a single amino acid is affected. Using the asparagine residue as a simple example, it was found that by creating a lesion in the two genes (*asnA* and *asnB*) responsible for converting aspartate into asparagine, a bacterial strain auxotrophic for asparagine could be created [84].

When using auxotrophic bacterial strains, the bacteria culture is grown in minimal medium supplemented with the necessary amino acids, but the standard carbon and nitrogen source can once again be used. Since the bacteria cannot produce the amino acid it is auxotrophic for, there is no chance of scrambling the label.

1.8 Summary of the literature review

In order to perform protein structural studies using biophysical techniques, pure protein samples have to be made and these proteins have to be specifically labeled with probes. In many cases, bacterial expression of proteins in minimal media with labeled components significantly reduced protein yield and in some cases, result in no protein production. In this literature review, I reviewed current literature of protein production using different host systems, including mammalian cells, insect cells, yeast and bacterial cells. I then focused on bacterial protein expression methods with emphasis on our

newly developed high cell-density method and the three practical protocols that ensure this high cell-density bacterial expression method to routinely produce large quantities of pure recombinant proteins.

In order to develop a novel method to study protein structure inside living cells, I then reviewed the literature of current structural biology techniques, including X-ray crystallography, NMR and cryo-EM which are atomic resolution structural biology techniques and fluorescence imaging and spectroscopy which are lower resolution techniques. I discussed the advantages and disadvantages of each structural biology technique and pointed out the possibility of applications for in-cell structural biology studies of proteins using these techniques. With this discussion, I concluded that the fluorescence technique may provide the best solution to study protein structure inside the living cells under a physiological concentration. I further discussed the challenges in developing an in-cell fluorescence technique that can be used to study protein structure inside living mammalian cells under a physiological concentration.

The first challenge is to generate a pure, specifically labeled mammalian cell population that contains only one protein, which is fluorescently labeled. I reviewed the current in-cell fluorescence labeling techniques in the literature and discussed the problems of using GFP for in-cell structural biology studies. I further pointed out that small molecule fluorophores might solve the problem since these fluorescence probes are sensitive to the changes in the chemical environment of the labeled protein, thus they could to report any structural changes of this labeled protein inside living cells.

However, there is no method in the literature that could specifically label a protein inside mammalian cells with these small molecule fluorophores.

To solve this challenge, I proposed a novel strategy that labels a protein with a small molecule fluorophore and then delivers this labeled protein into the correct intracellular compartment of a living cell for structural studies of this protein. I further proposed to use a tryptophan analog, 5-hydroxy-L-tryptophan (5-HT), as the fluorescence donor for this proposed strategy, since 5-HT labeled protein might allow us to separate the labeled protein from intracellular background proteins that contain regular tryptophans. To achieve this novel strategy, I reviewed the literature of protein labeling techniques with small molecule fluorophores and with specific amino acid analogs. I also reviewed current protein delivery techniques and pointed out that our newly developed QQ-protein delivery technique might serve as a physiological relevant protein delivery technique to achieve this proposed novel strategy.

To gain protein structural information inside the cells under a physiological condition, I proposed to develop an in-cell FRET technique to measure the distance between a specifically labeled fluorescence donor and acceptor either within a protein or between proteins. I reviewed the literature of the current FRET theory and its *in vitro* applications. I further proposed to use a protein called MESD as a model protein for my study. I reviewed the literature of MESD with the focus on the unsolved questions about this protein for its structure and biological functions.

CHAPTER 2

HIGH CELL DENSITY IPTG-INDUCTION BACTERIAL EXPRESSION FOR PRODUCTION OF GRAM/LITER PURE RECOMBINANT PROTEINS

2.1 Introduction

In order to perform *in vivo* FRET experiments, the target protein must be labeled with an amino acid analogue which requires growth in minimal medium with auxotrophic bacterial strains. High level expression in minimal medium ensures that I could efficiently produce enough proteins to be used in both the *in vitro* and *in vivo* FRET experiments.

Bacterial expression systems are the most attractive organisms for heterologous protein production for a number of reasons. The ability to reproduce rapidly and grow at high densities results in high yield protein production, as compared to eukaryotic systems. In addition, bacteria can generally be grown in an inexpensive medium and a large amount of protein can be produced in one day. Bacterial hosts typically have well characterized genetics that can be manipulated to create even better hosts. Finally, due to all of the mentioned advantages, there are a multitude of commercially available products, including bacterial strains, bacterial expression vectors and protein purification materials, to aid in heterologous protein production utilizing bacterial expression systems [63-66].

Escherichia coli remains the most commonly used bacterial host for all the above mentioned reasons. Expression of eukaryotic proteins inside a prokaryotic cell does not always result in a properly folded and active protein. However, techniques to optimize heterologous protein overproduction in *E. coli* have been explored that significantly enhanced the yield of the foreign eukaryotic proteins [87].

Structural biologists often need to specifically label recombinant proteins with biophysical probes for their studies. Using any expression system besides a bacterial expression system is highly inefficient and expensive. However, bacterial expression in minimal medium often results in lower protein yields as compared to growth and expression in rich medium such as LB and 2X YT.

However, minimal medium is necessary to specifically label proteins with amino acid analogues or isotopically labeled amino acids. The problem of "scrambling" arises since bacteria have the ability to synthesize all twenty amino acids. To ensure that bacteria do not synthesize their own amino acids, auxotrophic bacterial strains are developed. Auxotrophic bacterial strains usually contain a certain type of genetic mutation(s) inhibiting or knocking out biosynthesis of a specific amino acid(s). Minimal medium still must be used even when working with an auxotrophic strain since rich medium will contain all the standard amino acids. Minimal medium supplemented with isotopically labeled amino acid(s) or amino acid analogues allows the auxotrophic bacterial strain to grow and produce labeled proteins without scrambling.

Recently, Studier introduced an auto-induction bacterial expression method, which provides several advantages over the standard IPTG induction method, including:

(1) Achieving a high cell density and

(2) Minimal handling as there is no need to monitor cell growth for induction [76].

However, we were unable able to obtain consistent results using this method. In addition, protein yields were not very high using the auto-induction method although the cell density could get quite high. Following the idea of auto-induction, we developed a bacterial expression method that maintains the advantage of the tightly controlled induction by IPTG and utilizes both rich and minimal media to achieve a very high cell density for production of a very high yield of recombinant proteins. Unlike the auto-induction method, our high-cell-density method does not require longer time durations for achieving a high cell density, which is much more time efficient. Most importantly, we developed several practical protocols, to ensure high yield protein production at high cell density. These protocols allow us to use regular incubator shakers and original bacterial expression vectors. Our high cell density IPTG induction method is able to produce nearly gram quantity of pure recombinant proteins from a liter bacterial cell culture.

2.2 Materials and Methods

2.2.1 Bacterial strains, plasmids and media

Seven different proteins were tested, including two different constructs of receptor-associated protein, RAP (1-210), RAP (91-323), truncation mutants of the

human apolipoprotein E, apoE (1-183) and apoE (1-214), full-length apoE, a truncation mutant of mouse apolipoprotein AI, apoAI (1-216), and full-length human apoAI. The genes of these proteins were subcloned into different expression vectors as follows: RAP (1-210)/pET30a, RAP (91-323)/pET30a, human apoE (1-183)/pET22b [88], human apoE (1-214)/pTYB1 [89], apoE/pET30a-sHT [90], mouse apoAI (1-216)/pET30a [91]and human apoAI/pET30a-sHT [92]. The pET vectors were from EMD Biosciences and the pTYB1 vector was from New England BioLabs, MA. We engineered the pET30a vector to introduce a Factor Xa site between the long his-tag and the target gene. The pET30a-sHT is also an engineered pET30a vector in which the long his-tag was replaced by a short his-tag containing a six histidine tag plus a two serine linker. The pET30a and pET30a-sHT are kanamycin resistant vectors whereas the pET22b and pTYB1 vectors are ampicillin resistant vectors. The expression vectors were transformed in to BL-21(DE3) bacterial strains.

2.2.2 Creating a proper starting culture

To find the best cell density to prepare proper starting cultures, we added glycerol stock to a 5 mL LB culture containing the appropriate antibiotic and measured the OD_{600} every 30 minutes for 10 hours. After plotting time vs. OD_{600}, we could determine the various phases of bacterial cell growth: lag, exponential/log and stationary. Once the mid-log phase could be determined, we will use the OD_{600} of the mid-log phase as our starting culture. We noticed that different bacterial strains display

different growth curves, thus have to be tested when a new bacterial strain is used for the first time.

2.2.3 Traditional IPTG-induction bacterial expression method

We use this expression method to either check protein expression levels of different colonies during double colony selection or serve as an expression control.

For double colony selection, we used a small-scale expression with the following procedure: 2 mL of LB media was inoculated with a single colony from a freshly transformed plate as the starting culture and cultured at 37°C. When the OD_{600} reached the middle of its growth curve (usually between 3 and 5), 50 µl of the starting culture was added to 5 mL of the minimal M9 medium to obtain an initial OD_{600} between 0.05 and 0.1. When the culture reached an OD_{600} of ~1.0, it was induced with 0.5 mM IPTG and incubated at 20°C overnight [Human apoE (1-215), full-length apoE, mouse apoAI, human apoAI, RAP (1-210), RAP (91-323)] or at 28°C for 16 - 18 h [human apoE (1-183)]. Two hundred fifty microliters of cell suspension was collected and spun down at 3300g for 5 - 10 min. The cell pellet was resuspended in 50 µL of 2X SDS gel loading buffer and heated at 90°C for 30 min. Cell debris and DNA molecules were pelleted by centrifuging at a maximum speed for 10 min in a benchtop microcentrifuge. Finally, 10 µL of the supernatant was loaded into the SDS-PAGE gel to check the expression level.

For the IPTG method as an expression control, we used a 50 mL expression with the following procedure: 10 mL of LB media was inoculated with glycerol stock (after double selection) as the starting culture and cultured at 37°C. When the OD_{600} of the

starting culture reached between 3 and 5, 500 µl of the starting culture was added to 50 mL of the minimal M9 medium to obtain an initial OD_{600} between 0.05 and 0.1. When the culture reached an OD_{600} of ~1.0, it was induced with 0.5 mM IPTG and incubated at 20°C overnight [Human apoE (1-215), full-length apoE, mouse apoAI, human apoAI, RAP (1-210), RAP (91-323)] or at 28°C for 16–18 h [Human apoE (1-183)]. The cells were harvested and the cell pellet was used for protein purification.

2.2.4 Double colony selection

First, LB agar plates were prepared either in H_2O or in 70% D_2O (for triple-labeled protein expression). For 70% D_2O plates, the agar medium was not autoclaved, but microwaved until the agar dissolved. Three milliliters of agar was poured into a 35 mm x 10 mm petri plate (Corning, NY). Bacterial cells, either from a glycerol stock or 5 µL of a starting culture that has been diluted to an OD_{600} of ~0.05 - 0.1, were streaked onto the LB agar plates. Several colonies were picked from the plates next morning, and the expression levels of these colonies were checked using the traditional IPTG induction expression. Glycerol stocks were prepared for each colony. We chose the colony with the highest protein expression and went through another round of selection, following the procedure described above. The colonies selected from the double selection were used for preparation of glycerol stocks and were stored in a –80°C freezer. Once the double colony procedure is completed and high protein production is achieved, we recommend making at least 10 glycerol stocks to be stored in a –80°C freezer. Our

experience is that these glycerol stocks can be used even after two years storage and will still produce a similar high protein yield.

2.2.5 High-cell density IPTG-induction bacterial expression method

This expression method uses rich medium to achieve an initial high cell density before switching to minimal medium for expression. We started bacterial expression using a rich medium, such as LB or 2X YT, at 37°C. Once the cell density reached a cell density that was in the middle of its exponential phase, we switched the cell culture by gently spinning down cells and resuspending the pellet into the same volume of minimal medium. After switching the medium, we cultured bacterial cells for another 1.0 - 1.5 hours without adding IPTG, at the optimized temperature that is used for the cell culture after IPTG induction. During this period, the OD_{600} of the cell culture should increase by 0.5 - 1 unit. IPTG was then added to induce protein production. The cell culture was incubated at the same temperature for a period that is optimized for different proteins before cell harvest. Usually, we found that the OD_{600} value at the end of the cell culture increased by 2 to 3 fold compared with the OD_{600} value from the start of IPTG induction. Therefore, before harvesting the cells, the bacterial culture can reach an OD_{600} of 10 - 15 with LB as the starting rich medium and an OD_{600} of 15 - 20 using 2X YT. This is about a 5 to 10 fold increase in OD_{600} compared to that of the regular IPTG induction bacterial expression in minimal medium.

2.2.6 Optimization of various conditions

Another important step for high level protein production using high-cell density bacterial expression is to optimize the expression conditions such as culture temperature, IPTG concentration and induction time. These steps are critical for the initial expression of a protein using the high cell density expression method.

2.2.6.1 Temperature Optimization & Time Course

We prepared several 10 mL starting cultures in a rich medium at 37°C. Usually three cultures were started to check induction temperatures at 15, 20 and 37°C; however, if space and time permitted, we also checked induction temperatures of 18, room temperature and 28°. Once the optimal OD_{600} was reached, we gently spun down the culture and resuspended the cell pellet in 10 mL minimal M9 medium. We placed each flask in the appropriate incubator shaker and let the culture grow for 1 – 1.5 hours. At this point, we checked the OD_{600} to make sure the culture had adapted to the minimal medium. An OD_{600} increase of 0.5 to 1.0 units is a good indication that the bacterial cells have adapted well to the minimal M9 medium. We then added 0.5 mM IPTG to each culture and put the flasks back into the appropriate incubator shakers.

For the cultures growing at temperatures below 25°C, we allowed the cultures to grow overnight (14 – 16 hours). The following morning, we began collecting 500 µL samples and monitoring the OD_{600} every two hours for the remainder of the day (collected samples 14 – 28 hours after induction). For the cultures growing at temperatures above 25°C, we collected 500 µL samples and monitored OD_{600} every 2

hours after induction for a total of 8 hours. Once all the samples had been collected, we performed SDS-PAGE analysis to determine at which temperature the bacteria produce the highest protein yield.

During the time course, we closely monitored the pH of the medium, OD_{600} and protein yield. This allows us to obtain the best pH of the starting minimal medium, best cell culture temperature after IPTG induction and the best time to harvest bacterial cells for the highest protein yield.

2.2.6.2 IPTG concentration optimization

Once the optimal induction temperature and time were determined, we will determine the optimal IPTG concentration. We prepared several 2 mL starting cultures in a rich medium at 37°C. Once the optimal OD_{600} was reached, we gently spun down the culture and resuspended the cell pellet in 2 mL minimal M9 medium. We placed each tube in the incubator shaker set at the optimized induction temperature and let the culture grow for 1 – 1.5 hours. At this point, we checked the OD_{600} to make sure the culture had adapted to the minimal medium. We then added various amounts of IPTG: 0.1, 0.25, 0.5, 0.75 and 1.0 mM. We put the tubes back into the appropriate incubator shaker and let the culture grow for the optimized induction time. Finally, we collected 500 µL from each culture and performed SDS-PAGE analysis to determine the optimal IPTG concentration.

2.2.7 Aeration and medium pH

A bacterial culture grown to a high cell density results in a decrease in available dissolved oxygen and a large release of metabolites that will lower the culture pH [76]. In order to increase the available dissolved oxygen, we tested various culture volumes to flask size ratios, in addition to shaker speed. We also added NaOH to increase the starting medium's pH to various levels to determine the best starting pH.

2.3 Results

2.3.1 High cell density IPTG induction bacterial expression method

Figure 2-1 shows a comparison of the procedures and final OD_{600} before cell harvest for three different bacterial expression methods used in our studies. The traditional IPTG induction method we used in the laboratory uses minimal medium for bacterial expression. This is because we frequently prepare isotopically labeled proteins for NMR studies, which requires minimal medium with ^{13}C-glucose and $^{15}NH_4Cl$ in either H_2O for double-labeled proteins or in D_2O for triple-labeled proteins.

As Figure 2-1 indicates, the final OD_{600} of the traditional IPTG induction expression before cell harvest is usually about 2 – 3. We tested auto-induction expression for both unlabeled and triple-labeled proteins. Direct application of Studier's protocols using the C750501 recipe lead to inconsistent results [76]. For some proteins, the yield was a two to three-fold increase compared with the traditional IPTG method, whereas for other proteins, bacteria either did not grow or only a poor yield was obtained. This is especially true when we grow bacteria in D_2O for triple-labeling, which

is not very surprising because different growth patterns for bacteria in D_2O and H_2O are expected. In addition, we frequently observed a phenomenon during auto-induction experiments: using minimal media, the OD_{600} reached quite high levels (usually 8–20), but no protein production was observed. Due to these inconsistences, we developed the third bacterial expression method. The hybrid high cell density method uses rich medium, such as LB and 2X YT, to reach a high cell density before IPTG-induction. We then switch the culture medium by gently spinning down the cells and resuspending to an equal volume of minimal medium.

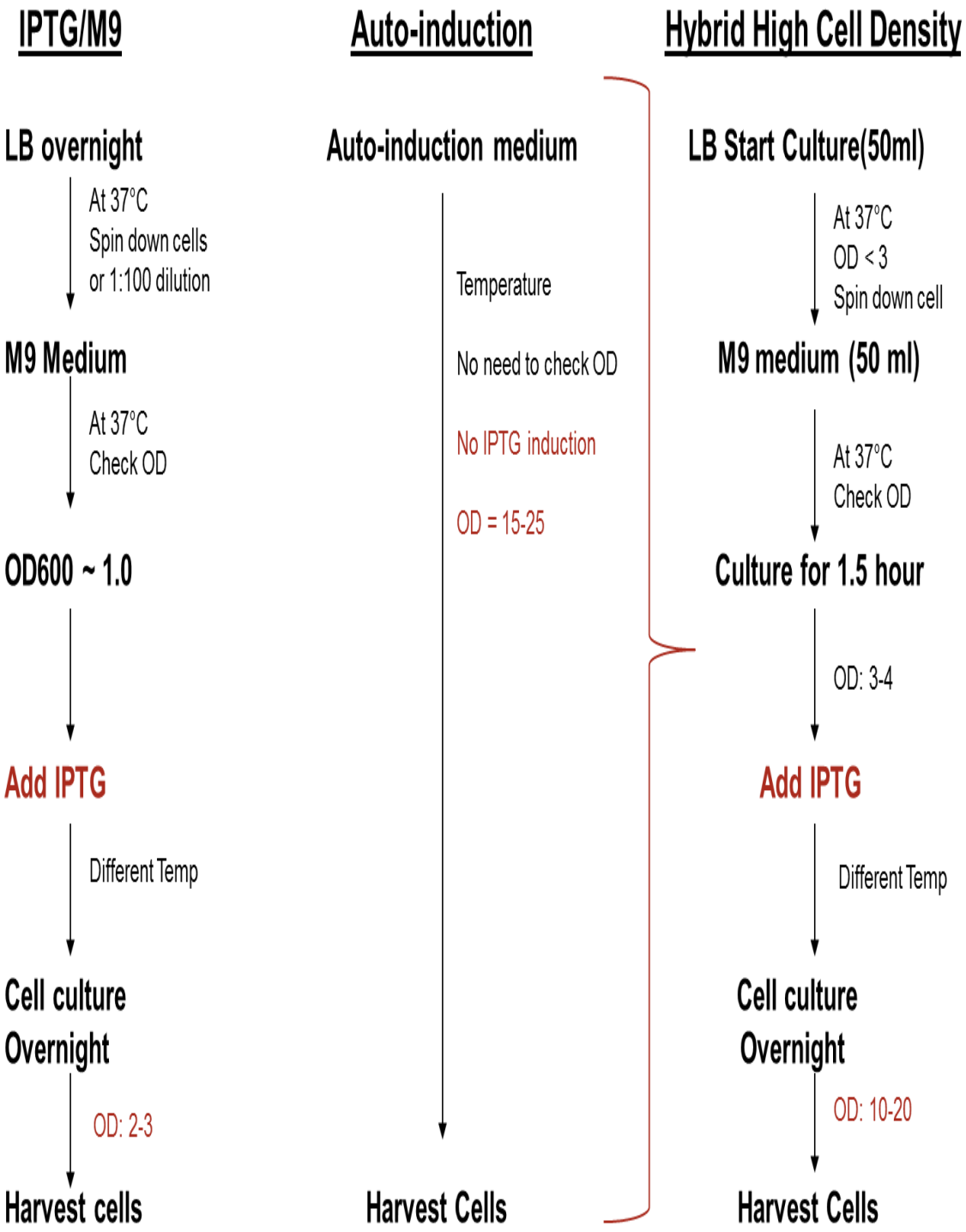


Figure 2-1: A schematic diagram of thee expression methods used in this study.

A similar method was reported previously by Cai et al. and by Marley et al. for making double-labeled protein with ^{13}C-glucose and $^{15}NH_4Cl$ [93, 94]. The procedure carried out by Cai et al. used a fermentor with a carefully controlled O_2 level and pH, whereas our method uses a regular incubator shaker that is commonly used in many laboratories for bacterial expression [93]. Marley et al. generated a cell mass with unsaturated LB medium (OD_{600} = 0.7) [94]. They then concentrated the suspension (2X, 4X, and 8X) and transferred the bacteria into isotopically labeled minimal medium for expression. The cells were incubated for 1 h at 37°C to allow for the discharge of unlabeled metabolites and then induced with IPTG. They discovered that the 4X concentrated LB medium conferred maximal protein expression.

We found that a bacterial culture should not be saturated in the rich medium, because a saturated bacterial expression would not result in a high yield of protein production [76]. Instead, the OD_{600} of the bacterial cell culture in the rich medium should be an intermediate value, preferably in the middle of its log phase, to ensure high level protein expression. This will also avoid the problems associated with cells going into stationary phase, such as induction of proteases [5]. For example, our experience suggested that an OD_{600} at 3–5 in LB medium and an OD_{600} at 5–7 in 2X YT medium were adequate before switching to minimal medium. After switching the medium, the bacterial cells were cultured at a previously optimized temperature for another 1.0 – 1.5 hours before IPTG induction, to allow bacterial cells to adjust to minimal medium and to the new culture temperature. Using this high cell density method, we can easily achieve IPTG induction within a ‘‘normal’’ working day, making

this method time efficient when comparing with the auto-induction method. For isotopic labeling of proteins using the high cell density method, a slightly longer period of medium exchange time, such as 1.5 – 2.0 hours, at a lower temperature might be preferred, because this not only allowed for the clearance of the unlabeled metabolites but also slowed down the bacterial growth during the exchange period, preserving the labeled nutrients for protein synthesis after IPTG induction. At the end of this short period of medium exchange time, the OD_{600} of cell culture should increase, normally, by ~ 0.5 – 1.0 units. After IPTG induction, the bacterial cells are cultured at an optimized temperature for an optimized time period before harvest. With this method, the final cell density before harvest can reach OD_{600} of 10–20, which significantly enhances the protein yield.

It is important to point out that there is no guarantee that a high cell density cell culture results in a high protein yield. As we described earlier, several drawbacks occur at high cell density bacterial expression, including plasmid loss, reduced medium pH, and limited dissolved molecular oxygen, causing either no protein production or a low protein yield. Indeed, when we initially worked with high cell density bacterial cultures, we frequently encountered a situation that even though cell density became quite high, the protein yield was either very low or no protein production could be seen at all. In addition, the protein expression yield was not always repeatable. We sometimes obtained an intermediate protein yield when we started with a freshly transformed colony. The other times we obtain a very low protein yield or no protein production at all even with a freshly transformed colony. When we started expression with freshly

prepared glycerol stocks, most of the times we only obtained a very low protein yield. To solve these problems, we further developed the following practical protocols that ensure repeatable very high yield protein production using these high cell density bacterial expression methods.

2.3.2 A proper starting culture

The general practice in most labs is to make a starting culture by growing an overnight culture using rich medium, such as LB, at 37°C. We observed that an overnight starting culture in rich medium at 37°C usually reached saturation by the next morning. A saturated overnight culture might result in plasmid instability because of the basal leakage of the T7 expression system [95]. This usually resulted in a poor yield of target protein. We found the best time to utilize a starting culture, with the least amount of plasmid instability, was in the middle of the exponential growth phase. For all expressions hereafter, we grew a starting culture in a rich medium (H_2O or D_2O based) for several hours at 37°C until the OD_{600} was in the middle of its exponential log phase; typically an OD_{600} between 3 and 5 for cultures grown in LB medium and 5 and 7 for cultures grown in 2x YT medium. For example, Figure 2-2 shows the growth curve for LCAT (lecithin:cholesterol acyl transferase) using a pET30a-sHT vector inside BL21(DE3) cells. The optimal OD_{600} for LCAT is ~2.5 and is reached after 7 hours of growth in a 50 mL cell culture. If the culture was grown in D_2O based rich medium, the growth rate was much slower. This method for obtaining a proper starting culture is used in both the traditional IPTG method and our new high cell density method.

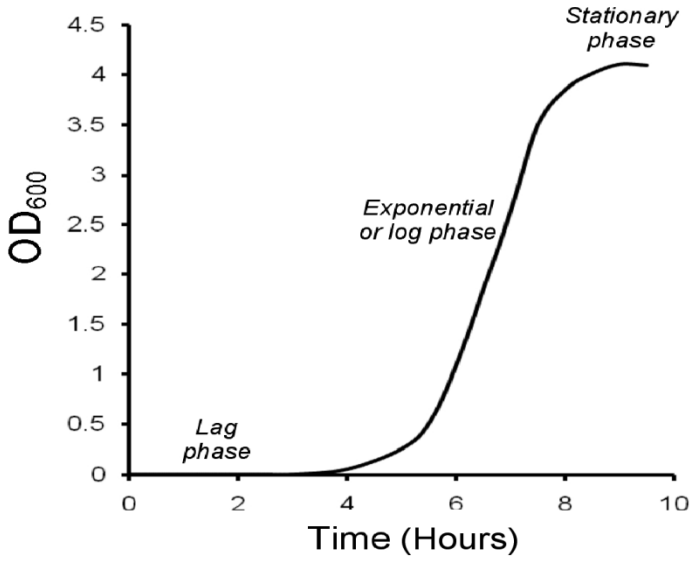


Figure 2-2: Plot of *E. coli* growth with the middle of the log phase at an OD_{600} value of ~ 2.

2.3.3 Double colony selection

We observed that colony selection was one of the most important factors for high level protein production using high density bacterial expression methods. This is especially true for bacterial expression in D_2O when making triple-labeled proteins. As a common laboratory practice for high level protein production of proteins, we routinely select high level expressing colonies. However, we often found that a low yield of protein was obtained using the glycerol stock made with a selected colony, even though this glycerol stock previously produced high yield protein. Such a situation happened quite often when we worked with human proteins that were toxic to the bacterial cells. This situation is also often observed when bacterial expression is carried out in D_2O. To solve this problem, we have developed a double colony selection protocol. In this protocol, the LB medium was inoculated with a single freshly transformed colony for a

starting culture, which was grown to an OD_{600} of 0.7–0.9. The medium was then spread onto a plate, followed by selection of colonies from the plate. The selected colonies were checked for protein expression levels using the traditional IPTG-induced expression. After expression, 200–500 μL of cell suspension was spun down and the cell pellet was treated with SDS loading buffer for 20 min at 70°C. An SDS-PAGE was carried out to check the expression level. Only those colonies that displayed high level expression will be used for the second selection. The second selection repeated the aforementioned procedure. If this double colony selection is used for selecting high level protein expression colonies in D_2O, we will carry out all the aforementioned experiments in D_2O, including D_2O plates.

With this double colony selection procedure, we were able to select several colonies for high level expression of a protein, whereas our previous experiments showed very low protein production in D_2O. An example can be seen in Figure 2-3, showing SDS-PAGEs of expression levels of apoE (1-215), using an apoE (1-215)/pTYB1 expression vector, in D_2O before (Panel A), during (Panel B), and after (Panel C) double selections. Panel A shows a lower protein production yield for all four colonies that were picked from a freshly transformed plate. However, colony 2 seemed to give a higher protein expression level (Lane 2), thus was selected for the next round of colony selection. Panel B shows a comparison of three different colonies from colony 2 selected in Panel A (Lanes 1–3) and another three different colonies from a double-colony selection (Lanes 4–6), suggesting that Colony 3 (Lane 3) gave the best protein expression level after the first-colony selection. Using Colony 3, we made a plate and picked three

more colonies, Colonies 4–6. The second colony selection indicated that Colony 6 gave the best protein expression level. With Colony 6, we made another plate and picked six colonies.

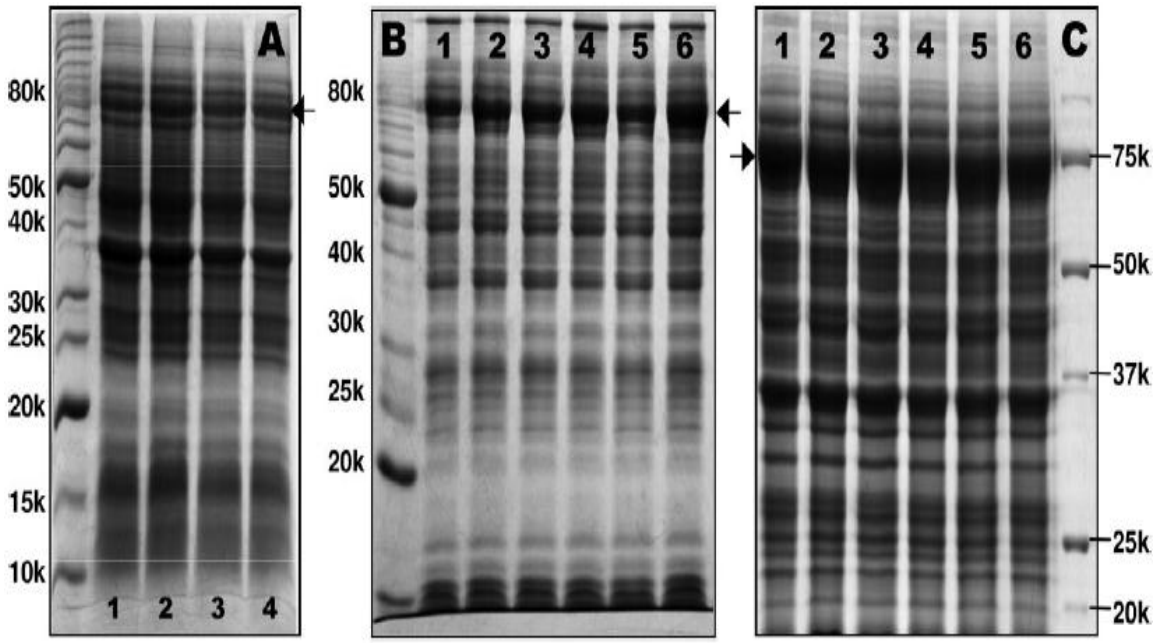


Figure 2-3: SDS-PAGEs of protein expression of apoE (1-215)/pTYB1 in D_2O before (Panel A), during (Panel B), and after (Panel C) double-colony selections. Arrows indicate the expected protein band (~80 kDa, apoE (1-215) + intein + CBD). Panel A shows four different colonies before colony selection. Panel B shows results of three different colonies selected from the single-colony selection (Lanes 1–3) and another three colonies selected from the double-colony selection (Lanes 4–6). The second-colony selection was based on Colony 3 (Lane 3) in the single-colony selection, because this colony gave a higher protein production. Panel C shows the results of six colonies from the double-colony selection, indicating a high protein expression level of all six colonies. Molecular weight markers are labeled with kDa.

It clearly demonstrates that all six colonies after double selection indeed solved the problem of low expression of apoE (1-215), resulting in a very high-level expression

of target protein in D_2O (Panel C). In contrast to single colony selection, a glycerol stock prepared using a colony from double colony selection can pass on for many generations and always give a consistent reproducible high level protein production. Therefore, a double colony selection procedure is recommended for colony selection of high level expressing colonies.

2.3.4 Optimization of bacterial expression in D_2O

Protein expression and purification is a routine practice in many NMR labs, but it is not uncommon to see a drastic reduction in protein yield when isotopically labeling the proteins, especially when D_2O must be used. If the bacteria did not grow well in 70-99% D_2O, we found that training the bacteria to adapt to D_2O-based medium and performing double colony selection upon D_2O plates solved the problem. First, we performed double colony selection (as outlined in section 2.2.4) utilizing 70% D_2O plates. Next, we trained the bacteria by picking a colony off a D_2O plate and starting a 5 mL bacterial culture of LB medium in 25% D_2O. Once the OD_{600} of the culture reached 1.0 at 37 ^{o}C, we transferred 100 µL of the cell culture into 5 mL of LB medium in 50% D_2O. The starting OD_{600} of this new culture is about 0.1. We let the cell culture grow at 37 ^{o}C until the OD_{600} reached 1.0 and transferred 100 µL of the cell culture into 5 mL of LB medium in 75% D_2O. The culture grew at 37 ^{o}C until the OD_{600} reached between 2 and 3. At this point, we set some aside to be used to make glycerol stock and the remaining culture was used as a starting culture for expression. Figure 2-3 shows an example of double colony selection on D_2O plates. After training the bacteria, we

consistently obtained 15 mg triple-labeled protein per 50 mL cell culture, as opposed to 0.6 mg per 50 mL cell culture using the traditional IPTG method.

2.3.5 Optimization of high cell-density IPTG-induction bacterial expressions

2.3.5.1 Temperature and Time optimizations after IPTG-induction

Another important step for high level protein production using a high cell density bacterial expression culture is to optimize expression conditions, such as culture temperature and the time after IPTG induction. This step is critical for the first time expression of a new protein using the high cell density expression method. First, we carry out time courses at different temperatures, such as 15, 20, 23 (room temperature), 28, 30, and 37°C. We closely monitor the following parameters: OD_{600}, pH, and target protein production. We normally make a 10 mL culture, either D_2O or H_2O-based, for the temperature and time course. To check target protein yield, we take 500 µL from the culture every 2 hours after IPTG induction (depending on the temperature), spin down, treat the cell pellet with SDS loading buffer for 30 min at 90°C, and take 10 µL to run SDS-PAGE. As an example, Figure 2-4, Left Panel shows an SDS-PAGE gel of a time course of triple-labeled human apoAI expression in D_2O at room temperature and Panel B shows a Western blot of the same time course. At each time point, we also checked pH of the expression medium and OD_{600}. This figure clearly demonstrates the importance of the time course, indicating that either apoAI does not have enough time to be expressed under 30 hours or the expressed apoAI starts to degrade after 40 hours, both resulting in low protein production. In contrast, bacterial

expression at 36 hours gave the highest protein yield as confirmed by the Western blot in the right panel of Figure 2-4.

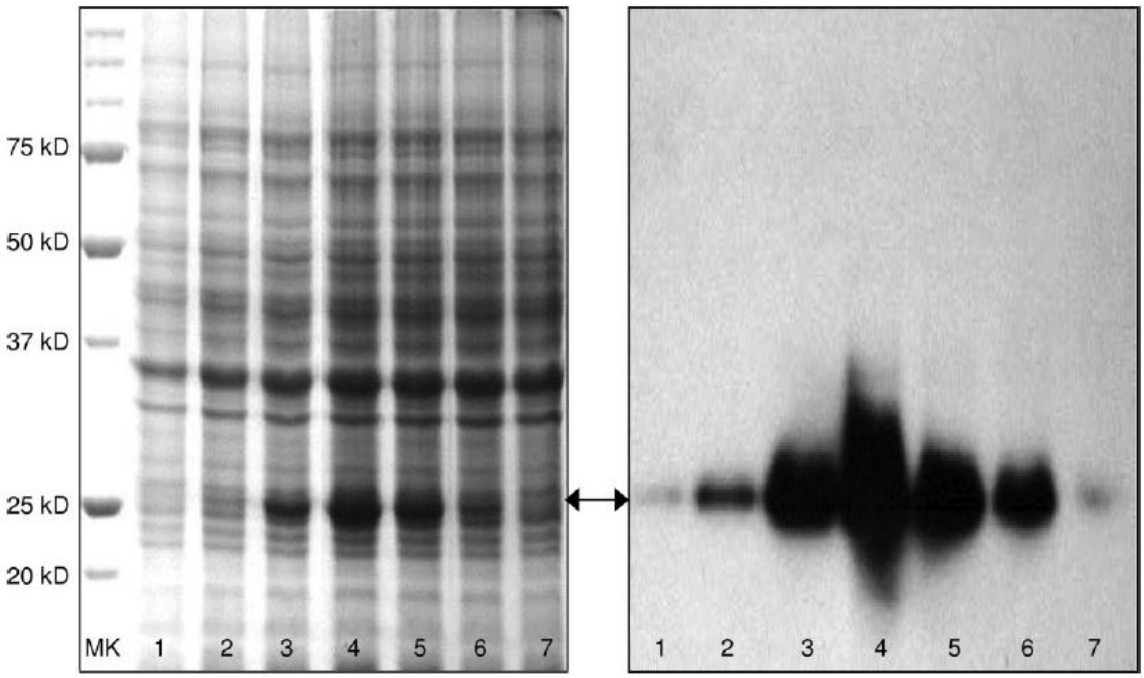


Figure 2-4: Left Panel: An SDS-PAGE showing a time course of triple-labeled human apoAI expression in D_2O at room temperature. The expected apoAI band is indicated by an arrow. Lane 1:24 h, Lane 2:28 h, Lane 3:32 h, Lane 4:36 h, Lane 5:40 h, Lane 6:44 h, and Lane 7:54 h. Right Panel: Western blot of the same time course using an anti-human apoAI monoclonal antibody, 5F6.

Table 2-1 lists the OD_{600}, pH, and protein yield at each time point, suggesting that OD_{600} has indeed reached its maximum at 36 hours (OD_{600} = 9.1), resulting in the highest protein yield. In contrast, the pH of the expression drops from the starting pH.7.2 to 6.01 after 36 hours. Further reduction of pH may lead to significant instability of the plasmid, resulting in plasmid loss and significant reduction of the protein yield.

Table 2-1: Parameters of the time course of human apolipoprotein A-I expression

Time	*24h*	*28h*	*32h*	*36h*	*40h*	*44h*	*54h*
OD_{600}	2.5	3.9	7.2	9.1	8.4	8.0	8.1
pH	6.6	6.5	6.3	6.0	6.0	6.0	6.1
Protein yield	-	+	++	+++	++	++	-

For the seven proteins we tested, we found that different proteins require different temperatures for the optimized yield. For example, we expressed the two fragments of RAP, RAP (1-210) and RAP (91-323) at 37°C. For human apoE N-terminal domain, apoE(1-183), expression was carried out at 28°C. For apoE(1-215)/pTYB1, optimal expression temperature was 20°C after IPTG-induction and for the two apoAI proteins, experiments at room temperature provided the best yields. Nevertheless, time course experiments at different temperatures allow us to quickly optimize expression conditions for a high level production of proteins.

2.3.5.2 Optimized media

2.3.5.2.1 ^{13}C-glucose optimization for high cell density bacterial expression

High density bacterial cells require more nutrition in the minimal medium, which usually uses NH_4Cl as the nitrogen source and glucose as the carbon source. For making isotope-labeled protein, we use $^{15}NH_4Cl$ and ^{13}C-glucose to replace normal NH_4Cl and glucose for double-labeling the proteins. We intended to optimize both $^{15}NH_4Cl$ and ^{13}C-

glucose amounts for a low cost of production of isotope-labeled proteins. In most cases, our laboratory used 0.2 – 0.4% of ^{13}C-glucose and 0.1% of $^{15}NH_4Cl$ for regular IPTG induction bacterial expression. We found that this recipe did not work well with a high cell density IPTG induction method, simply because of limited nutrition in the minimal medium, which limited bacterial cells to reach a high cell density and significantly reduced protein yield. We optimized different nutrition in the minimal medium for high cell density expression (Table 2-2).

Table 2-2: Optimized High Cell Density Minimal Medium

50 mM	$Na_2HPO_4 \cdot 7H_2O$
25mM	KH_2PO_4
10mM	NaCl
5mM	$MgSO_4$
0.2mM	$CaCl_2$
0.25X	Trace metals
0.25X	Vitamins
0.1%	NH_4Cl or $^{15}NH_4Cl$
1.0%	Glucose or ^{13}C-glucose

As an example, Figure 2-5 shows glucose optimization of high cell density IPTG induction expression of human apoE in D_2O. These expressions started with a glycerol

stock of apoE after double colony selection and are carried out with an optimized time course and temperature.

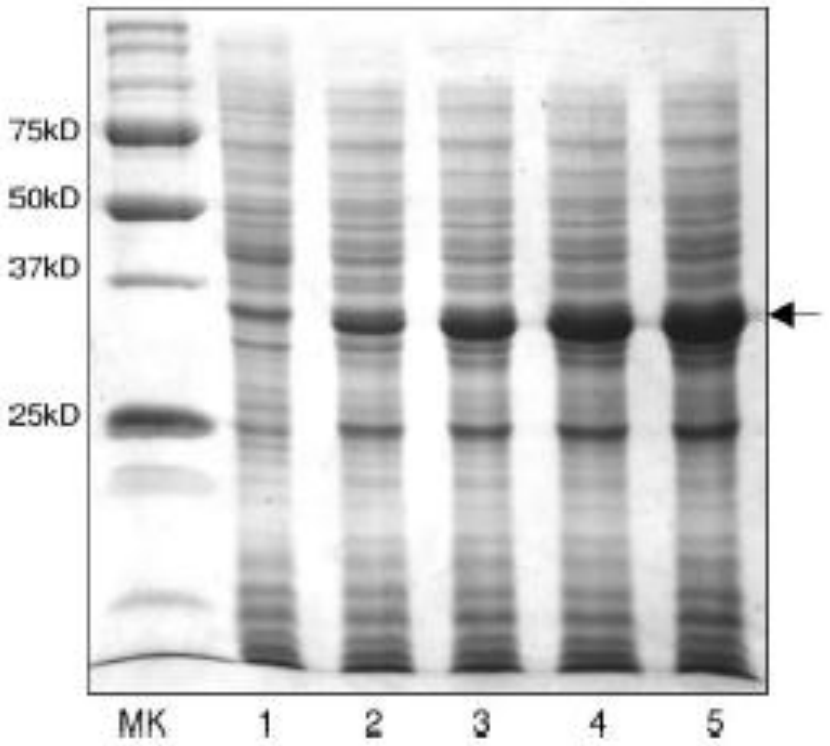


Figure 2-5: A 12% SDS-PAGE of glucose optimization of human apoE expression in D_2O at 20°C using high cell density IPTG induction bacterial expression: Un-induced (Lane 1), with 0.4% (Lane 2), 0.6% (Lane 3), 0.8% (Lane 4), and 1.0% glucose (Lane 5). Molecular weight marker is shown in left lane. Small-scale time course experiments with different glucose concentrations were also carried out to find the optimum protein expression time after induction of the culture.

A cell culture containing 0.4% glucose could reach an OD_{600} of only 4.2 and the expression yield was also low. Increasing the glucose concentration increases the culture cell density and protein yield. With 1.0% glucose, the OD_{600} reached to 7.4 and protein yield seems enhanced by ~10-fold (Table 2-3). Marley et al. previously showed

that increasing the amount of glucose to 0.8% only improved the protein yield modestly. They suggested that glucose concentration is not a critical factor in enhancing protein yield. Our results seemed to be different, indicating that the amount of glucose is critical for high cell density IPTG induction expression. Our high cell density expression is based on several optimizations as described earlier, which may make a significant difference. We believe that at a high cell density bacterial culture, more nutrients, especially the carbon source, are required for healthy cell growth, thus, the culture can reach to a high cell density, resulting in a higher protein production.

Table 2-3: Glucose Optimization of High Cell Density IPTG Induction Bacterial Expression of Human apoE in D_2O

Glucose (%)	*0.4*	*0.6*	*0.8*	*1.0*
OD_{600} at IPTG induction	2.0	2.5	3.0	3.3
OD_{600} at harvest	4.2	5.1	6.1	7.4
pH at harvest	6.7	6.6	6.4	6.3
Time after IPTG induction	12	19	23	36
Protein yield	+	++	+++	++++

2.3.5.2.2 Optimization of the starting pH of the M9 medium

Bacterial cultures grown to a high cell density typically result in a drastic drop in pH. This drop in pH is largely due to the increased glucose concentration in the minimal

medium and subsequent metabolism of the glucose. Medium with a low pH may cause stress to the bacterial cells, which results in plasmid loss from the high density bacterial cells [95]. We chose to modify the pH of the culture medium to 8.2 using NaOH, so that the medium has a larger buffer capability. Our result confirmed that an enhanced pH of the expression medium indeed helped to control medium pH at high cell density, thus at the end of bacterial expression, pH maintained at pH >6.0 for the cell culture of OD_{600} between 8 and 14 (Table 2-1).

2.3.6 Culture volume: flask size ratio – proper aeration

To solve the problem of limited availability of dissolved oxygen in the high cell density culture, we used a smaller expression volume with a larger culture flask. For example, a 250 mL high cell density expression culture was divided into 5 x 50 mL cultures, and each culture used a 250 mL flask. This resulted in better aeration in a small culture compared with a large culture, thus more molecular O_2 will be dissolved in the medium. In addition, the shaker speed was kept for 175 – 225 rpm. Our results were similar to Studier's findings with his auto-induction studies [76].

2.3.7 A very high yield of pure recombinant proteins

With these protocols, we routinely produced 14–25 mg of triple-labeled proteins and 17–34 mg of unlabeled proteins from 50 mL cell cultures for all the proteins we tested. Table 2-4 lists the final yields of unlabeled and triple-labeled proteins using high cell density bacterial expressions and compared with the yields of the traditional IPTG

induction methods, which is also fully optimized, in a 50 mL cell culture, suggesting a 9 to 85 fold enhancement in protein yield.

Table 2-4. Final yields of unlabeled and triple-labeled proteins: high cell-density vs traditional IPTG method.

Protein	***High cell density[b] (mg)***	***IPTG[b] (mg)***	***M.W. [45] (Dalton)***	***M.W. (MS) (Dalton)***	***%D[c]***
Triple-labeled					
RAP(1-210)	20 ± 3	0.5	33,801	33,525 ± 195	~92
RAP(91-323)	25 ± 3	0.8	36,633	36,376 ± 200	~93
ApoE(1-183)[a]	18 ± 4	2	22,866	22,686 ± 116	~89
Mouse apoAI(1-216)	15 ± 2	0.8	28,014	27,732 ± 125	~90
Human apoAI	14 ± 1	0.6	32,814	32,401 ± 150	~88
Unlabeled					
Human apoAI	34± 1	1.0			
Human apoE	17± 2	0.2			

M.W.: molecular weight.

[a] ApoE(1–183) was expressed in 40% D_2O, the rest are expressed in 99.7% D_2O.

[b] High cell density (50 mL culture volume): high-cell-density expression methods, including auto-induction and high cell-density IPTG-induction; IPTG: the optimized traditional IPTG-induced expression. We repeated the expressions at least three times for all proteins, the yield shown is the average ± standard deviation.

[c] Estimated percentage of deuteration, assuming 100% ^{13}C and ^{15}N-labeling. For apoE(1–183), the %D is the estimated percentage of deuteration based on 40% D_2O. For the other four proteins, the %D is the estimated percentage of deuteration based on 99.7% D_2O.

It is worth noting that we repeated the expressions of each protein more than three times and Table 2-4 gives the average yields with standard deviations. This indicates that the protocols described earlier produce a consistent, high level, triple-labeled protein production and that is always reproducible. Table 2-4 also gives the mass spectroscopic data of the triple-labeled protein, indicating that the efficiency of deuteration for triple-labeled protein using high cell density expressions. Overall, the deuteration efficiency is around 90% if we assume that the ^{13}C and ^{15}N-labeling are 100%. This is because we used 99.7% D_2O and ^{13}C-glycerol or ^{13}C-glucose (not deuterated) in the high cell density expressions. For the apoE (1-183) case, we only used 40% D_2O and ^{13}C-glycerol or ^{13}C-glucose (not deuterated), the 89% deuteration level was based on 40% D_2O (Table 2-4). This result is comparable with the deuteration efficiency of the traditional IPTG induction expression with single-labeled ^{13}C-glucose.

2.3.8 Protein structure integrity is maintained using high cell density IPTG-induction bacterial expression

To confirm the efficiency of triple-labeling by high cell density expressions, we carried out NMR experiments of these proteins. Figure 2-6 shows an example of the ^{1}H-^{15}N HSQC experiments of human apoE (1-183), for which the NMR samples were obtained using high cell density IPTG induction expression (Panel C), auto-induction expression (Panel B), and traditional IPTG induced expression (Panel A).

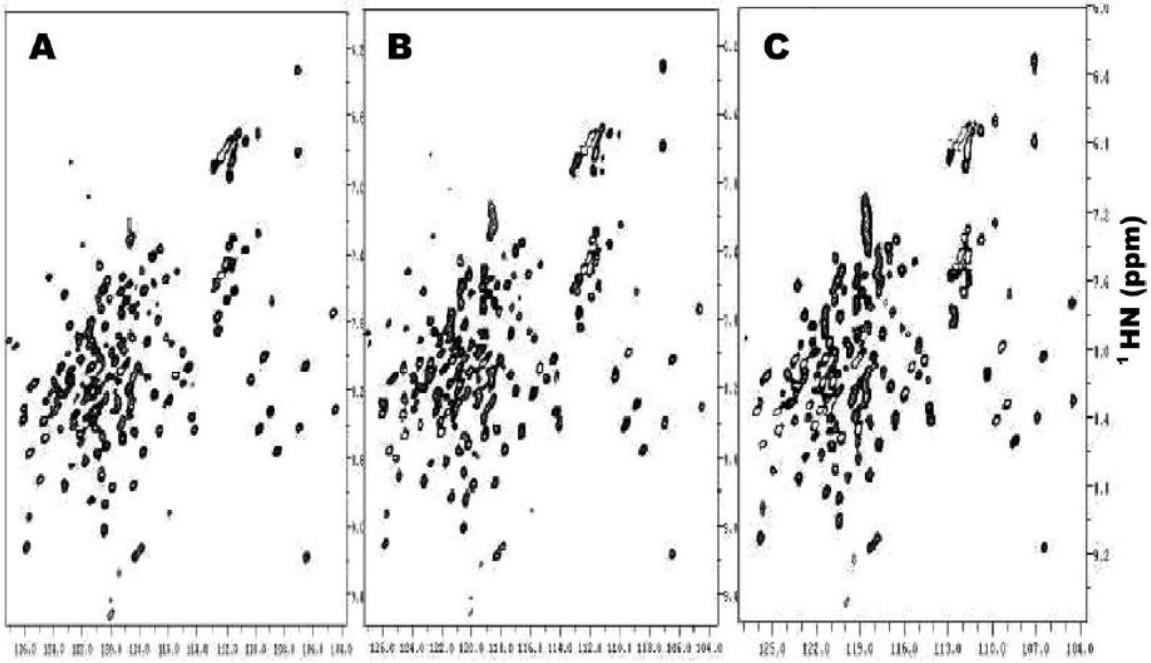


Figure 2-6. ^{1}H-^{15}N HSQC spectra of triple-labeled human apoE (1-183) obtained using high cell density IPTG induction expression (Panel C), auto-induction expression (Panel B), and traditional IPTG induction expression (Panel A). All three samples contained 1.0 mM triple-labeled human apoE (1-183) in 100 mM phosphate buffer, 10 mM EDTA, 0.5 mM NaN_3, 90 mM DTT, and 0.02 mM DSS, pH 6.80. The spectra were collected at 30°C on a 600 MHz NMR spectrometer with a cold probe.

This figure demonstrates that all three NMR samples produce an identical HSQC spectrum. In addition, with the proteins obtained using our high cell density IPTG method for apoE and apoAI, we have carried out NMR studies allowing us to completely assign NMR spectra of lipid-free apoE, lipid-free mouse apoAI (1-216), and human apoAI/preβHDL [92, 96, 97]. In addition, we also determined NMR structures of lipid-free apoE (1-183) and mouse apoAI (1-216) (manuscript in preparation)[98]. Thus, we conclude that the high cell density expression produces a very high yield of triple-labeled, well folded proteins for NMR studies (14–25 mg/50 mL for triple-labeled

proteins; 17–34 mg/50 mL for unlabeled proteins). Table 2-4 also indicates that the principles we described here for optimization of high cell density bacterial expression methods can be directly applied to other proteins, including membrane proteins in either H_2O or D_2O to obtain very high level production of target proteins.

2.4 Conclusions

We have developed a hybrid bacterial expression method that combines the advantages of the traditional IPTG method and Studier's auto-induction method. We have combined the tight control allowed by a timed IPTG induction and the high cell density nature of the auto-induction method. To achieve this, we developed three critical protocols to ensure consistent and high level protein production:

(1) A proper starting culture must be used before transfer to minimal medium which ensures plasmid stability. By transferring a culture growing in rich medium in the middle of its growing phase, we have eliminated the problems caused by allowing a culture to reach saturation (low pH, plasmid instability, antibiotic degradation).

(2) Double colony selection must be performed to generate a reliable glycerol stock needed for future high level expressions. This is especially critical when expressing proteins in D_2O or labeling with amino acid using auxotrophic bacterial strains.

(3) Optimization of induction temperature, time and IPTG concentration is critical since all of these variables are protein dependent. Very few of the proteins we tested optimized in the same manner – some proteins required low induction temperature and a long induction time, while others required high induction temperatures for a short period of time.

After performing these three critical protocols, we have found that we can obtain consistent high level expression for all of the proteins we tested in our lab. Compared to the traditional IPTG induction method we used previously, we observed a 9 to 85 fold enhancement in protein production:

(1) Unlabeled proteins: 17-34 mg/50 mL cell cultures

(2) Triple-labeled proteins: 14-25 mg/50 mL cell cultures

In summary, following the practical protocols developed in this study, this high cell-density IPTG-induction bacterial expression method allows for production of gram quantity of pure recombinant proteins from one-liter bacterial expression.

2.5 Discussion

E. coli offers a mean for the rapid and economical production of recombinant proteins. In recent years, the number of recombinant proteins used for therapeutic

application increased dramatically. These demands drive the development of a variety of strategies for achieving high level bacterial expression of proteins using *E. coli*. Optimizations in expression vector design, gene dosage, promoter strength (transcription regulation), mRNA stability, translation initiation and termination, *E. coli* host strain design, and codon usage have been performed, which result in significant enhancement of protein production and different commercial products [63], such as the pET expression vectors and pLysS plasmid by EMD Biosciences. The pLysS plasmid carries the gene for T7 lysozyme, which is a natural inhibitor of T7 RNA polymerase and serves to suppress basal expression of T7 RNA polymerase prior to induction, thus stabilizing recombinants encoding target proteins that may also affect cell growth and viability [63, 66]. In addition, empirical selection yields E. coli strains that are superior to the traditional BL21(DE3) host strain by overcoming the toxic effects associated with the overproduction of membrane and globular proteins under T7 transcriptional control [66, 87]. Based on the BL21(DE3) strain, another strain was developed to overcome the problem of rare codons. Many eukaryotic proteins may contain codons that are not typically used in *E. coli*, therefore the BL21 CodonPlus-RIL and BL21 CodonPlus-RP strains were engineered to enhance expression of eukaryotic proteins that use these rarely used codons in *E. coli* [65, 66].

In contrast, optimization of bacterial expression conditions seems to be protein dependent [99]. The general consideration is to increase cell density of bacterial expression for the purpose of enhancing recombinant protein production. Much of the efforts have been centered on enhancement of cell density in a fermentation setting,

rather than in a general laboratory setting, because bacterial expression conditions, such as O_2 level, pH, and nutrients, can be much better controlled using a fermentor to achieve a high cell density [95, 100]. In contrast, these expression conditions are difficult to control using a regular incubator shaker, and thus a much lower cell density can be achieved using this general laboratory setting.

We aimed to develop a novel bacterial expression method for high yield recombinant protein production in a general laboratory setting using a regular incubator shaker without changing the expression vector. The biggest challenge to achieve this is to obtain high yield recombinant protein production using a high cell density bacterial expression since bacterial cells experience stress at a high cell density using a regular incubator shaker which does not control the O_2 level, pH and nutrients of the expression medium. To achieve high cell density of bacterial expression in a general laboratory setting, we first utilized the auto-induction method developed by Studier [76]. We obtained inconsistent results which sometimes produced a higher yield (2 – 5 fold enhancement), but many times did not produce recombinant protein at all. We found that although the auto-induction method produced much higher cell density as compared with traditional IPTG-induction method in minimal medium, this high cell density often caused the plasmid to be dropped from the bacterial cells, thus no protein production was observed. This is because high cell density of bacterial expression often changed medium conditions, such as acidity and limited dissolved oxygen, causing significant stress to the bacteria cells. Under such a stressed condition, the expression vector could be easily dropped by the bacterial cells.

This result encouraged us to develop a better high cell density bacterial expression method, so that we can maintain a high cell density without plasmid drop and also take the advantages of the tightly controlled IPTG induction for better recombinant protein expression. This will allow us to take advantage of the optimized features of the commercially available bacterial expression vectors for high yield production of recombinant proteins. We developed a hybrid bacterial expression method that utilizes rich medium to achieve a high cell density before IPTG induction, while maintaining the advantage of the tightly controlled induction by IPTG in minimal medium. Our hybrid expression method allows us to reach a high cell density with a final OD_{600} that is 5 to 10 fold higher than that of the regular IPTG induction method.

High cell density culture systems, especially under the non-fermentation, laboratory conditions, frequently suffer from several drawbacks, including plasmid loss, limited availability of dissolved oxygen, and increased carbon dioxide levels in the medium which causes significant reduction of medium pH [63, 101]. These problems often cause a low or even no protein production with a high cell density culture. Indeed, we frequently observed a low protein production from a high cell density bacterial expression before we implemented our protocols to solve these problems. The common practices in general laboratories to solve these problems are as follows: selecting high expressing colonies and optimizing growth temperatures and time. We found that these common practices sometime produce inconsistent results that are not always repeatable. This is especially true for protein expression in D_2O for production of the triple-labeled proteins.

We have designed several modifications to these common laboratory practices specifically for bacterial expression at high cell densities in a routine laboratory setting. The major modifications include:

(1) Double colony selection

(2) Proper preparation of a starting culture

By performing a double colony selection, we ensure that we have chosen a stable colony with a high level of expression. We found that it is not enough to only perform one round of colony screening. In the past, we have seen inconsistent expression results from a single glycerol stock based on one round of colony selection. Once we started performing double colony selection on all proteins, we obtained much more consistent results.

Many labs start with an overnight culture in rich medium before switching to minimal medium for expression. A saturated culture reaches a high cell density overnight and therefore experiences a drop in pH and limited dissolved oxygen. This stress can cause some of the bacteria in the culture to drop the plasmid. Thus, by using this overnight culture as the starting culture for expression, the end result will be lower protein production since many of the bacteria have already dropped their plasmids. By monitoring the growth of the starting culture and capturing the bacterial cells in their mid-log phase, we have eliminated the stressors that will ultimately cause plasmid loss.

Using these modifications, we have obtained repeatable very high yield of protein production for all the proteins tested, especially for triple-labeled proteins in D_2O. Our data indicated a 9 to 85 fold enhancement of protein yields. Importantly, such a high protein yield used the same DNA constructs and the same bacterial strains that we previously used for regular IPTG induction method. This provides a critical advantage of our method/protocols – a simple optimization in bacterial expression conditions can result in 9 to 85 fold enhancement of protein yields. Indeed, we routinely obtain 14–25 mg of triple-labeled proteins and 17–34 mg of unlabeled proteins from a 50-mL cell culture for all the proteins tested.

CHAPTER 3

IN VITRO FRET MEASUREMENTS, FRET PEAK ASSIGNMENT AND SELECTIVE LABELING OF MESD WITH 5-HYDROXY-L-TRYPTOPHAN

3.1 Introduction

The goal of this thesis is to develop an in-cell fluorescence technique that allows for measurement of the distances between fluorescence acceptors and donors within a protein or between two proteins inside the correct intracellular compartment of living cells. The successful achievement of this goal will allow us to obtain high-resolution structural information within a protein or between two proteins inside the cells, one key step towards high-resolution structural biology of proteins inside the living cell.

To achieve this goal, we will apply the fluorescence resonance energy transfer (FRET) technique to the specifically labeled proteins inside the cells. Our rationale is to specifically label the protein(s) of interest in the test tube with a small molecule fluorophore and then deliver the labeled protein(s) into the correct intracellular compartment of living cells for in-cell FRET measurement. The QQ-protein delivery technique can specifically deliver a protein to its intracellular destiny based on its signal sequence [102, 103]. This will result in special mammalian cells that contain a fluorescence labeled target protein with unlabeled intracellular endogenous proteins as the background. The FRET measurement will be performed on this specifically labeled protein and the calculated FRET-distance will be between the donor and acceptor of the

protein(s) of interest, thus, high-resolution structural information of a protein inside living cells can be obtained using this novel approach.

To minimally disturb the protein, we intend to use tryptophan residues of a protein as the fluorescence donor and IAEDANS as the acceptor, which is a thiol-reactive fluorophore and can bind to a cysteine residue [104]. We use MESD, a specialized chaperone protein for the LDLR super family members, as the model protein for this approach [57, 62]. MESD contains one cysteine (C142) and three tryptophan residues (W32, W98 and W130). We prepared several tryptophan mutants, including single and double tryptophan mutants. We first labeled MESD single tryptophan mutants with IAEDANS at C142, and performed *in vitro* FRET measurement to verify our FRET technique. Our lab recently determined the NMR structure of MESD and we could accurately measure the distances between the C142 bound IAEDANS and the tryptophan residues. Our data indicated that our FRET-measured distances are nearly identical to the distances in the NMR structure. This confirmed validity of our *in vitro* FRET technique.

We also explored a fluorescence approach to determine multiple FRET-distances with a single FRET-measurement using MESD samples that contain multiple fluorescence donors (tryptophans) and a single acceptor. The key steps of this approach are to separate FRET signals of these multiple donors to the acceptor and to assign the separate FRET signals as the FRET peaks that come from each individual tryptophan residue. We utilized synchronous fluorescence scanning technique that has the ability to separate individual FRET peaks. We were also able to assign these separated FRET-

peaks when we compared single and double tryptophan mutants with the wild-type MESD that contains three tryptophan residues. This is an important step towards development of a FRET-technique for efficient measurement of inter-residue distances within a protein or between proteins inside living cells, thus allowing us to develop a high-resolution in-cell structural biology technique to study protein structures and structural changes caused by intracellular events.

To eliminate intermolecular FRET contributions between the tryptophan of intracellular background proteins and the labeled IAEDANS inside the protein of interest, we will specifically label MESD with 5-hydroxy-L-tryptophan (5-HT) using an auxotrophic bacterial strain. 5-HT labeled proteins can be excited at 310 nm, whereas proteins with regular tryptophan residues cannot be excited at this wavelength [105]. This ensures that the in-cell FRET measurement at 310 nm is from the labeled 5-HT in MESD and IAEDANS labeled at C142, whereas the tryptophan residues within intracellular background proteins will not be excited, thus to eliminate any possible protein-protein interactions. This chapter also describes our results of specific labeling of MESD with 5-HT and IAEDANS.

In summary, this chapter works on technical development of a successful in-cell FRET technique for simultaneous measurement of multiple distances using a FRET protein sample with multiple tryptophan donors and one IAEDANS acceptor. Our goal is to develop a structural biology technique to study protein structure and structural changes inside living cells at high-resolution.

3.1.1 FRET-based distance calculations

FRET is defined as the physical phenomenon in which there is a non-radiative transfer of energy via long-range dipole-dipole coupling between a donor and acceptor molecules [40]. There are three basic requirements in order for FRET to occur:

(1) The donor and acceptor must be within 10 – 100 Å of each other.

(2) Emission spectrum of the donor must overlap with the absorption spectrum of the acceptor.

(3) The dipoles of the donor and acceptor molecule cannot be perpendicular.

The rate of transfer (k_t) is proportional not only to the distance between the donor and acceptor (r), but also the unperturbed lifetime of the donor (τ_o) and the Förster distance (R_0: the distance at which energy transfer is 50%). The efficiency of energy transfer (E) is a quantitative measure of the number of quanta that are transferred from donor to acceptor. E is also known as the quantum yield of energy transfer. To determine the efficiency of energy transfer, we can multiply the rate of transfer (k_t) by the first excited singlet state lifetime (τ):

[1] $$E = k_t * \tau = \frac{R_0^6}{R_0^6 + r^6}$$

We can also determine E by measuring the steady-state donor fluorescence intensity from a sample containing only the donor and another sample containing the donor-acceptor pair [40].

[2] $E = 1 - \frac{Q_{DA}}{Q_D}$

Where Q_D is the quantum yield of the sample containing only a donor and Q_{DA} is the quantum yield of the sample containing both the donor and acceptor.

By setting equation [1] equal to equation [2] and solving for *r*, we can now calculate the distance between the donor and acceptor on our protein:

[3] $r = R_0 \left(\frac{1}{1 - \frac{Q_{DA}}{Q_D}} - 1 \right)^{1/6}$

3.1.2 Donor-Acceptor Pair

These calculations can be greatly simplified by using a donor-acceptor pair with a known Förster distance (R_0). For example, the tryptophan-IAEDANS donor-acceptor pair is used in my study and has an R_0 = 22 Å with an effective range of 13.2 – 35.2 Å.

Another consideration when deciding upon a donor-acceptor pair is the type of FRET experiment to be performed. Since distances will be calculated based on differences in intensity, the donor-acceptor pair must be bright and photostable. Small molecule fluorophores are advantageous to use in FRET experiments for several reasons. The fluorophores are small (<1 kD) and covalently attached to proteins with a minimal disturbance to the protein structure. They are also much brighter than fluorescent proteins and are more easily detected when performing FRET experiments. In contrast to GFP which is a large protein (~ 28 kD), small molecule fluorophores are

less than 1 kD, therefore, accurate distance measurements between the donor and acceptor can be accurately calculated.

Ideally, in order to obtain structural information from proteins inside living cells, they need to be labeled with small molecule fluorophores. However, currently there is no *in vivo* method available to only label a protein of interest without labeling the intracellular background proteins. This imposes a major challenge in specific labeling of a protein inside cells for fluorescence studies of a protein of interest.

On the other hand, if we can label the purified target protein *in vitro* and then deliver the labeled protein efficiently into the correct intracellular compartment of living mammalian cells at a sufficient concentration, this will allow us to generate a specific labeled mammalian cell population for in-cell FRET measurements. In recent years, there have been several advances made in the field of exogenous protein delivery making this approach possible.

3.1.3 Synchronous scanning fluorescence spectroscopy

Synchronous scanning fluorescence spectroscopy is a technique that takes advantage of the ability to vary both the excitation and emission wavelengths simultaneously during data collection. In this technique, the fluorescence signal is recorded when excitation and emission wavelengths are simultaneously scanned keeping in between a fixed wavelength interval (called the offset value, $\Delta\lambda$) throughout the spectrum. As a result, the selectivity for individual fluorescent components is

considerably improved; additionally, much more information on mixtures of fluorescent compounds is gained [43, 44].

The tryptophan residues of a protein could be considered “fluorescent components” as they are each in their own unique environment and could possibly have unique spectral patterns. The immediate chemical environment of tryptophan residues, such has being involved in hydrogen bonding or hydrophobic interactions, can have effects on its fluorescent properties. These interactions could result in slight red or blue shifts of its emission spectrum, causing different wavelengths of the FRET-peaks for different tryptophan donor/IAEDANS acceptor pairs [45].

It might become possible to create a protein with multiple donors and a single acceptor and obtain information about each donor/acceptor pair from one measurement. These wavelength differences may possibly allow us to assign the FRET-peaks to an individual FRET donor/acceptor pair, if these synchronous scanning fluorescence spectra can separate the individual FRET-peaks.

3.1.4 MESD as a model protein

In order to test the ability to collect FRET experiments from inside living cells, MESD was chosen as the model protein. MESD is a specialized chaperone essential for the folding of LRP5/LRP6 which are the critical co-receptors of the Wg/Wnt signaling pathway [60, 61]. It functions as a molecular chaperone inside the ER specifically for proper folding and export of the LDLR family, including LRP5 and LRP6 [57]. In the

absence of MESD, LRP5/LRP6 fails to reach the cell surface and remains sequestered as insoluble aggregates due to misfolding.

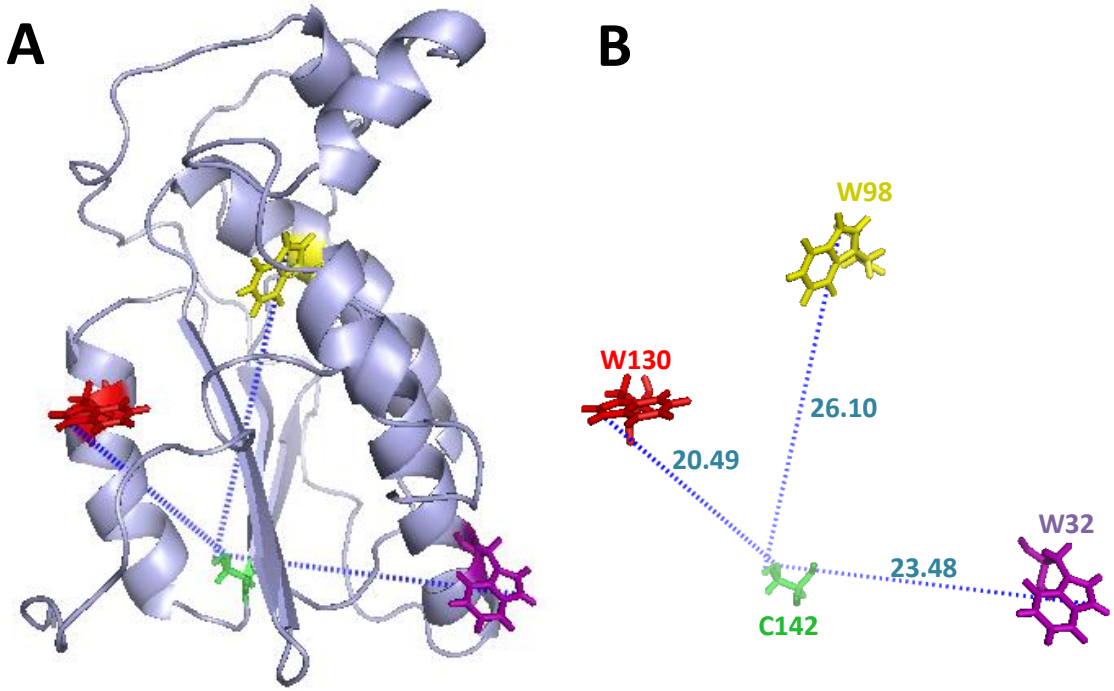


Figure 3-1: Panel A shows the ribbon structure of MESD with W32 highlighted in purple, W98 highlighted in yellow, W130 highlighted in red and C142 highlighted in green. Panel B is a slightly enlarged version of Panel A without the backbone ribbon structure. The blue dashed lines and corresponding numbers indicate the distance between the $N_{\varepsilon 1}$ atom of the tryptophan residues and the S_{γ} atom of the cysteine residue. This figure is based on the NMR structure of MESD (PDB code: 2KGL) [62].

MESD serves as an excellent model protein for a number of reasons. First, it has only three tryptophan residues that can act as the FRET donors (Figure 3-1). By using a native amino acid of MESD as the fluorescence donor, there will be no need to label the protein with an additional fluorescence probe to serve as the donor. This will also cause minimal structural changes of MESD for the fluorescence labeling. It is also easy to mutate one or two tryptophan residues at a time to reduce the number of donors, since the simplest measurement involves one donor and one acceptor. Second, the sequence contains one cysteine residue. There are a number of commercially available small molecule fluorophores that are thiol reactive (e.g., IAEDANS) and will covalently bind to the cysteine residue and serve as the FRET acceptor. Third, the MESD sequence contains an "RDEL" ER-retention signal. This ensures that after the protein has been delivered into the ER of living mammalian cells, it will stay in the ER and traffic between the ER and Golgi. This is critical for our in-cell FRET-measurement since we would like to measure the FRET distances between the donor and acceptor of MESD in its physiological intracellular locations, thus making our FRET-measured distances physiologically relevant.

3.1.5 5-Hydroxy-L-Tryptophan labeling and fluorescence spectroscopy

Since tryptophan is the chosen donor for the future in-cell FRET experiments, it is essential to develop a novel method to distinguish the tryptophan of MESD from all the other tryptophans of the intracellular background proteins. A tryptophan analogue, 5-hydroxy-L-tryptophan, was chosen based on its different fluorescent spectral properties

from normal tryptophan [105]. The 5-HT analogue has an extended absorbance spectrum that allows for specific excitation at 310 nm (Figure 3-2), whereas under this wavelength, the native tryptophan residues are not excited.

There have been a number of methods introduced that have allowed for the incorporation of tryptophan analogues. Some labs have simply synthesized peptides using the analogue [106]. Other labs have created new orthogonal pairs (analog-RS and $\mathrm{tRNA}_{\mathrm{XYZ}}^{\mathrm{analog}}$) and taken advantage of rare codons or amber stop codons [107, 108]. Another method involves an OXYPRO promoter that induces protein production once oxygen levels have fallen below a certain level [109]. Several labs have used auxotrophic bacteria to incorporate 5-HT and in order to selectively label MESD with 5-HT with near 100% incorporation, we chose to utilize a bacterial strain auxotrophic for methionine and tryptophan, DL21 (DE3) – a generous gift from Dr. Carl Frieden.

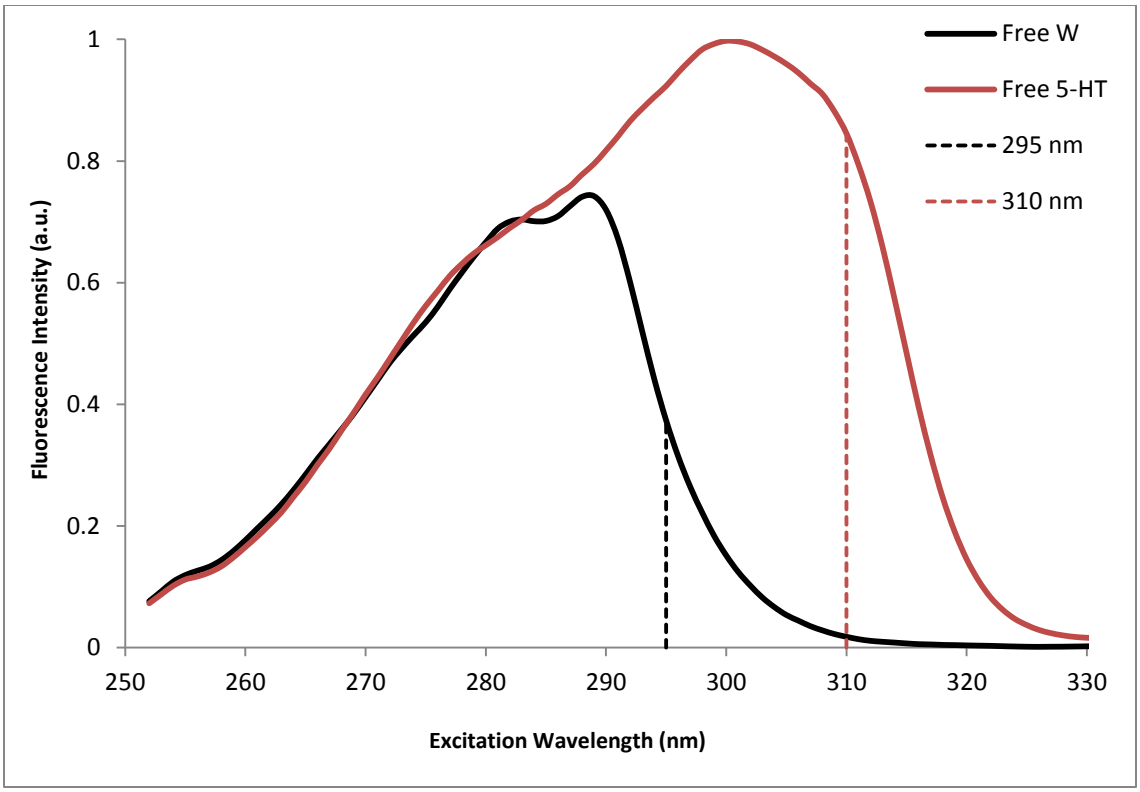


Figure 3-2: Absorbance spectra of free tryptophan versus 5-hydroxy-L-tryptophan (5-HT) (λ_{em} = 336 nm).

3.1.6 Can multiple FRET-distances be obtained from a single FRET-measurement?

The current *in vitro* FRET technique usually uses a protein that is labeled with a single donor and a single acceptor, thus this technique can only generate a single distance between a donor and an acceptor of the protein with each FRET measurement [110]. This is the major bottleneck for developing the FRET technique as a high-resolution structural biology tool. Indeed, the FRET-measured distances are accurate at angstrom resolution, this technique certainly has the potential to determine a protein structure at atomic resolution.

To determine a protein structure at atomic-resolution using the FRET technique though, hundreds of FRET-distances are required for structural simulation using the FRET-measured distances as the experimental restraints. This would require an enormous effort including mutagenesis, fluorescence labeling and FRET-measurement, representing the major challenge that limits this technique from becoming a high-resolution structural biology technique.

Comparing to the current atomic resolution structural biology techniques including X-ray crystallography, NMR and cryo-EM, the FRET technique has several major advantages for physiologically relevant in-cell experiments. First, FRET is a sensitive technique and only requires a micromolar protein concentration which is similar to the intracellular protein concentration. In contrast, NMR requires a millimolar concentration that is not a physiologically relevant concentration. Second, the FRET-technique does not have the line-broadening problem of large proteins under an intracellular environment, whereas NMR does. This is another key property that distinguishes FRET from NMR for in-cell experiments. Indeed, the macromolecular crowding effect of the intracellular environment significantly broadening the NMR signals to significantly reduce the signal-to-noise, making in-cell NMR a very challenging subject [111]. Third, although large amounts of mutagenesis and fluorescence labeling are required, the FRET-measurements are quick and multiple scans of the same sample take less than ten minutes. This is the third advantage of FRET for in-cell experiments. Since we have to collect biophysical experiments in a tube or cuvette, the longer the experiments are, the more cell viability will become a problem. A few hours inside a

cuvette without the optimized cell culture conditions, the mammalian cells will still remain healthy. Thus the experiment data collected will be true reflection of the protein structure inside the cells under physiological conditions.

In this thesis, we also address a key question: Can we obtain multiple FRET-distances via a single FRET-measurement? We will explore new techniques to achieve this goal. Usually, a protein contains multiple tryptophan residues. Our strategy is to develop a technique that allows us to obtain multiple distances between these tryptophan donors and a single acceptor. If we can achieve this goal, we can simply change the single cysteine position to prepare new samples for new FRET-measurements. This approach will allow us to minimize the mutagenesis work and to efficiently obtain many FRET-based distances used for structural simulation to generate protein structure inside the cell at atomic resolution. This shall have a revolutionary impart on structural biology.

3.2 Materials and Methods

3.2.1 Strain, plasmid and mutants

The MESD gene was subcloned into a pET30a vector from EMD Biosciences. We engineered the pET30a vector to introduce a Factor Xa site between the long his-tag and the MESD gene. The pET30a-sHT is also an engineered pET30a vector in which the long his-tag was replaced by a short his-tag containing a six histidine tag plus a two serine linker. The pET30a and pET30a-sHT are kanamycin resistant vectors. The expression vectors were transformed in to BL-21(DE3) and DL-41(DE3) bacterial strains.

MESD contains three tryptophans at residues 32, 98 and 130. In order to make double tryptophan mutants of MESD, each tryptophan residue was mutated to an alanine using the Site Directed QuickChange Mutagenesis Kit (Stratagene). The resulting mutants are identified as MESD_W32A, MESD_W98A and MESD_W130A as the double tryptophan mutants. In order to make the single tryptophan mutants of MESD, the MESD_W32A mutant was used as a template to mutate W98 to A98, resulting in MESD_W32/98A. The MESD_W32A mutant was also used as a template to mutate W130 to A130, resulting in MESD_W32/130A. The final single tryptophan mutant was made by using the MESD_W98A template to mutate W130 to A130, resulting in MESD_W98/130A. Finally, a “no tryptophan” mutant was made using MESD_W32/98A as a template and mutating W130 to A130 resulting in MESD_W32/98/130A. Table 3-1 lists all the mutants and their corresponding number of tryptophan residues.

Table 3-1: MESD and Tryptophan Mutants

Mutant	***Referred to as***	***# TRP***	***TRP positions***	***Mutations***
MESD_wt	WT	3	W32, W98, W130	No mutation
MESD_W32A	32A	2	W98, W130	W32A
MESD_W98A	98A	2	W32, W130	W98A
MESD_W130A	130A	2	W32, W98	W130A
MESD_W32/98A	130W	1	W130	W32A, W98A
MESD_W32/130A	98W	1	W98	W32A, W130A
MESD_W98/130A	32W	1	W32	W98A, W130A
MESD_W32/98/130A	NoW	0	No TRP	W32A, W98A, W130A

3.2.2 Protein expression and purification

In order to produce unlabeled MESD, glycerol stock was added to 50 mL LB broth and 50 µL KAN in a 250 mL flask and placed in a 37 °C incubator shaker. The culture was grown until OD_{600} reach ~ 2 – 2.5. The culture was spun down at 5,000 x *g* for 7 minutes and the supernatant was poured out. The conical tube was inverted for 1 minute on a paper towel to remove all traces of LB broth. To resuspend the pellet, 50 mL of minimal M9 medium was added to the pellet and poured into a new 250 mL flask. The flask was placed in a 37 °C incubator shaker for 1 hour. The OD_{600} was measured to confirm growth, then 0.5 mM IPTG was added to the culture and placed back in the 37 °C for 4 hours. The cells were harvested by spinning the culture down in two 50 mL conical

tubes at 10,000 x *g* for 10 minutes. The supernatant was removed and the pellet was either store in a -80 °C freezer or used immediately for purification.

In order to purify the MESD proteins, the cell pellets collected after harvesting were re-suspended in 20 mL 1X binding buffer (recipe modified slightly from His-Bind Resin manual and all buffers containing 6 M urea). The solution was sonicated 3X for 1 minute each time at 10 V. The lysate was spun down at 10,000 x g for 10 minutes. The supernatant was poured into a second container and stored on ice. The pellet went through two more rounds of sonication adding 10 mL of 1X binding buffer each time to resuspend the pellet. The clear lysate was loaded twice onto the previously charged and equilibrated His-Bind resin. The resin was then rinsed with 100 mL 1X binding buffer and 100 mL 1X wash buffer (25 mM imidazole). Finally, the protein was eluted with 60 mL 1X elution buffer (200 mM imidazole).

The elution was poured into a dialysis bag (MWCO 10,000 kD) and placed in 4 L of distilled water containing ~20 mM $NaHCO_3$. The solution stayed on dialysis for at least three days with three water changes per day. Once dialysis was complete, the solution was poured into a thick glassed beaker and place in a small container of liquid nitrogen to freeze. Once frozen, the beaker was placed on a lyophilizer and only the powder of the protein remained. The protein was weighed and a small sample was taken to check the purity of the protein powder.

3.2.3 Labeling MESD with IAEDANS

The protocol provided by Invitrogen was used to label MESD with IAEDANS. First, MESD protein powder was dissolved in buffer (25 mM NaCl, 25 mM sodium-phosphate buffer, 10 mM EDTA, pH 7.6) at a concentration of ~ 80 µM (2 mg/ml). To disrupt the intermolecular disulfide bonds that have possibly formed, 20X molar excess of TCEP (1.6 mM) was added to the solution and was placed on a rocker at room temperature for 1 hour. Next, the solution was divided evenly into two microcentrifuge tubes and were labeled "unlabeled" and "IAEDANS". To the IAEDANS tube, 40X molar excess of IAEDANS (3.2 mM) was added. Both tubes were wrapped completely in foil and placed on the rocker at room temperature for at least 2 hours. Finally, to both tubes, 20X molar excess of DTT (1.6 mM) was added to stop the reaction. The tubes were once again wrapped in foil and placed on the rocker at room temperature for 1 hour. In order to remove the free IAEDANS, the solution underwent size exclusion gel chromatography with a molecular weight cut off of 10 kD.

We optimized the concentration of IAEDANS to ensure efficient labeling. During this optimization, step 2 of the protocol (20X molar excess was suggested) was changed and various amounts of IAEDANS were added: 20, 30, 40 and 50X molar excess. Once the free IAEDANS was removed, emission scans were collected on each sample. The samples were excited at 290 and 336 nm and the emission range collected was 300 – 600 nm and 350-600 nm, respectively.

3.2.4 *In vitro* FRET

After labeling the protein with IAEDANS, the sample protein concentration must be determined as they need to be the same in order to correctly calculate FRET distances. BCA or Lowry assays are used to determine concentrations and the samples are diluted with buffer to equilibrate the concentrations between the two samples.

Next, emission scans are collected on each sample by pipetting 350 µL of the protein solution into a 4-sided quartz cuvette. The cuvette is placed into the QuantaMaster-6 Spectrometer (Photon Technology International, South Brunswick, NJ) and the data is collected using the Felix32 software provided by PTI. After opening a new data acquisition file, "Emission Scan Method" is selected and the following information is inputted: excitation wavelength is set at 295 nm, emission range is set at 305 – 600 nm, step size is set at 0.5 nm, integration time is set at 0.5 sec and average is set at 5. The buffer spectrum is collected for baseline correction.

3.2.5 Synchronous scanning fluorescence spectroscopy of IAEDANS labeled MESD and its tryptophan mutants

The samples that were prepared for the *in vitro* FRET experiments were also used for the synchronous scanning experiments. Using the Felix32 software package, new data acquisition file is opened, "Synchronous Scanning Method" is selected and the following information is inputted: excitation wavelength range is set at 240 – 370 nm, emission range is set at 450 – 580 nm, step size is set at 0.5 nm, integration time is set at 0.5 sec and average is set at 5. The buffer spectrum is collected for baseline correction.

3.2.6 MESD expression using auxotrophic bacteria and 5-HT labeling

In order to produce MESD selectively labeled with 5-hydroxy-L-tryptophan (5-HT), glycerol stock (DNA transformed into DL41 (DE3) competent cells) was added to 500 mL M9 minimal medium supplemented with 25 mg/L of MET and TRP and allowed to grow overnight for at least 20 hours. After overnight growth, the OD_{600} reached ~ 1.0 before spinning down the culture at 5,000 x *g* for 7 minutes. The supernatant was removed and the container was inverted for 1 minute on a paper towel. The pellet was re-suspended with 25 mL 1X PBS and vortexed briefly in order to remove any remaining TRP. After spinning down and removing the supernatant, the pellet was re-suspended with 500 mL defined M9 minimal medium (not containing any glucose, ammonium chloride or tryptophan, but supplemented with 50 mg/L of the other 19 amino acids). The re-suspended culture was poured into a 2 L flask and placed back in the 37 °C incubator shaker for 2 hours. After 2 hours, 25 mg 5-HT was added to the culture and placed back into the incubator shaker for an additional hour. At this point, 0.5 mM IPTG was added to the culture and grew for 5 hours at 37 °C. Cells were harvested by spinning the culture down at 10,000 x *g* for 10 minutes. After removing the supernatant, the pellet was either stored in the -80 °C freezer or used immediately for purification. The purification procedure is the same as unlabeled MESD (3.2.2).

3.2.7 Characterizations of 5-HT labeled MESD

In order to confirm the incorporation of 5-HT into the protein, fluorescence emission and excitation scans of both 5-HT labeled and unlabeled MESD protein samples were collected using a fluorescence spectrometer. Protein powders of MESD containing regular tryptophan and 5-HT were dissolved in buffer (25 mM NaCl, 25 mM sodium phosphate buffer, 10 mM EDTA, pH 7.6) at a concentration of 50 µM. Using a 4-walled quartz cuvette, 350 µL of the protein solution was added and the cuvette was placed in the fluorescence spectrometer. Using the Felix32 software package, new data acquisition file is opened, "Excitation Scan Method" is selected and the following information is inputted: excitation wavelength range is set at 250 – 320 nm, emission wavelength point is set at 336 nm, step size is set at 0.5 nm, integration time is set at 0.5 sec and average is set at 5. The buffer spectrum is collected for baseline correction.

3.2.8 Assignment of the FRET peaks

Standard emission scans were collected using MESD mutants containing multiple donors and one acceptor (IAEDANS) with an excitation wavelength set at 295 nm. After baseline correction, analysis of the emission spectra revealed that the emissions of the multiple tryptophan donors coalesced into one uniform spectrum and effects from individual donors could not be determined. Synchronous scanning, on the other hand, is designed to detect subtle differences amongst a mixture of fluorescent components in a solution and is therefore used to collect fluorescence spectra of these MESD samples.

After collecting all the data, it was imported into an excel file and baseline corrected, by subtracting the buffer spectrum from each protein spectrum. Next, the data was divided into three subgroups: W32, W98 and W130 groups. For example, the W32 group includes: MESD_wt, MESD_W98A, MESD_W130A, and MESD_W98/130A. The "no tryptophan" mutant spectrum was included in each subgroup to be used as a reference spectrum.

After generating a chart containing all the protein spectra, each region of the spectra was analyzed. The first region includes the emission wavelength ranging from 465 – 485 nm, this region coincides with an excitation wavelength range of 255 – 275 nm. Both tyrosine and phenylalanine residues may be excited by this wavelength range. Although typically the emission intensity of both tyrosine and phenylalanine is much less than that of tryptophan, there are 4 tyrosine residues and 7 phenylalanine residues in MESD. Their combined emission could be exciting IAEDANS and might explain the first region of this spectra.

The second region includes the emission wavelength ranging from 485 – 510 nm, which coincides with an excitation wavelength range of 275 – 300 nm. This excitation range fully overlaps the absorption range of tryptophan. Therefore, this region may be the "FRET peak" region. Once a donor is excited, its emission can excite the acceptor and cause the acceptor to emit its own emission. The peak emission of an acceptor caused by excitation from a donor's emission is called the "FRET peak". When comparing the FRET peak of MESD samples containing one donor and one acceptor using an excitation wavelength of 290 nm, we found that the corresponding FRET peaks

display maximum intensity at 480 – 490 nm (Figure 3-4). The FRET peaks observed in region 2 of the synchronous scanning spectra overlap the same region found in Figure 3-4, providing additional spectroscopic confirmation for our assignment.

It is this region that can be used to assign tryptophan residues. By comparing all the different mutants and wild-type proteins, subtle changes in this region of the spectra should be due to the contributions to the FRET peaks of different tryptophan residues. Therefore, such a comparison may allow us to assign the FRET peaks of each individual tryptophan residue.

The first comparison is between the “no tryptophan” mutant and the single tryptophan mutants of MESD. The peak in the single tryptophan mutant spectra is assigned as the tryptophan residue that the mutant contains. This is done for each single tryptophan mutant, allowing us to tentatively identify the FRET peak position of each tryptophan residue. By identifying the tentative FRET peak of each tryptophan residue, these FRET peak assignments can be confirmed using double tryptophan mutants. This can be done by comparisons of the FRET peaks of the double tryptophan mutants in the FRET region. For example, a double tryptophan mutant containing W32 and W130 should contain two individual FRET peaks of W32 and W130, if there is no FRET peak overlap between these two tryptophans. An observation of the FRET peaks of W32 and W130 in the FRET spectrum of this double tryptophan mutant will be a confirmation of assignment of the FRET peaks of W32 and W130. This comparison will also allow us to tentatively assign the FRET peak for the second tryptophan residue.

The third region includes the emission wavelength ranging from 510 – 580 nm with a peak emission at 546 nm. After subtracting the $\Delta\lambda$ of 210 nm, that leaves a λ_{ex} = 336 nm. Since the peak excitation wavelength of IAEDANS is exactly 336 nm, therefore, this third emission region is a result of the direct excitation of IAEDANS.

3.3 Results

3.3.1 MESD tryptophan mutants

Wild-type MESD was successfully mutated into single and double tryptophan mutants containing varying number and placement of tryptophan residues using Stratagene's Site Directed QuickChange Mutagenesis Kit. The ds-plasmid DNA was sent out for sequencing and the results were confirmed. Each plasmid was successfully transformed into bacterial strain BL21 (DE3) and underwent double colony selection [79]. Once a stable colony was formed, the high cell density method was used for expression and an affinity chromatography column was used for purification. Figure 3-3 shows purified protein samples of wild-type MESD and all the MESD tryptophan mutants.

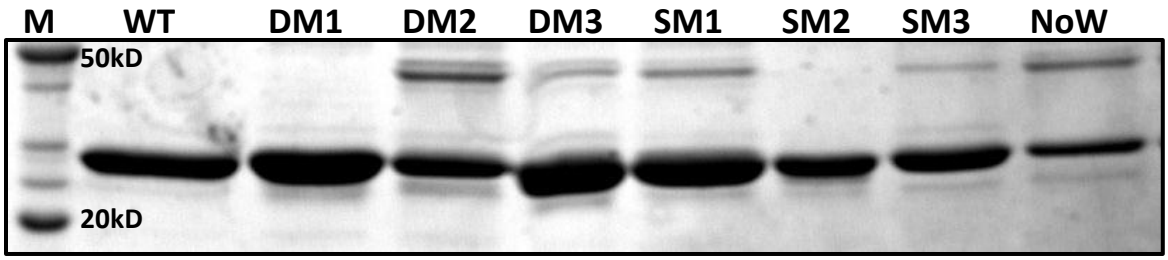


Figure 3-3: 12% SDS-PAGE of 8 MESD proteins. M – Marker, WT – wild-type MESD, DM – double tryptophan mutant (1: W32A, 2: W98A, 3: W130A), SM – single tryptophan mutant (1: W98/130A, 2: W32/130A, 3: W32/98A) and NoW: no tryptophan mutant.

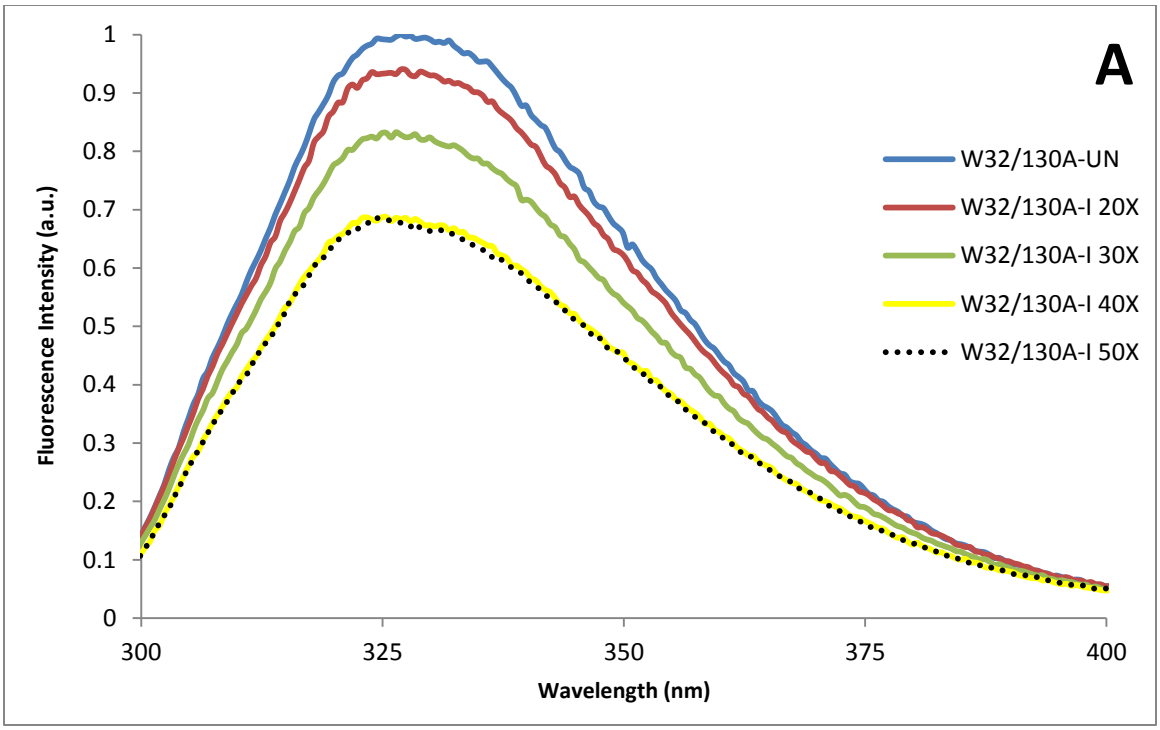
3.3.2 Optimization of IAEDANS-labeling

During the initial experiments with IAEDANS labeled MESD, widely varying spectra were obtained for the same mutant and the distance calculations were far off from the observed distances between the donor and acceptor observed in the NMR structure of MESD. Invitrogen supplied a protocol for labeling proteins with IAEDANS, but suggested optimizations may be necessary. The concentration of each reagent for each step was increased two-fold, therefore the new protocol used 20X molar excess of TCEP, 40X molar excess IAEDANS and 20X molar excess of DTT. The FRET spectra of the mutants labeled in this manner displayed consistent FRET spectra. Using these FRET spectra, the distances between the single tryptophan donor and IAEDANS acceptor provided consistent distances that are close to the observed distances between the various tryptophan residues and C142 in the NMR structure of MESD. To eliminate any artifacts of this approach, multiple FRET spectra were collected and used to calculate the FRET-distances. The calculated distances were nearly the same as the distance in the

NMR structure of MESD with small variations. Based on this result, the TCEP and DTT concentrations were maintained at 20X molar excess.

In order to ensure that all the MESD proteins were labeled with IAEDANS, various concentrations of IAEDANS were tested to determine the optimal amount of IAEDANS that was needed when labeling. Samples were treated with: 20, 30, 40 and 50X molar excess of IAEDANS and emission scans were collected in the range of 300 – 600 nm and 350 – 600 nm with the excitation wavelength (λ_{ex}) equal to 290 and 336 nm, respectively. When λ_{ex} = 290 nm, we will observe the tryptophan emission peak (decrease in intensity) and when λ_{ex} = 336 nm, we will observe the IAEDANS emission peak (increase in intensity).

As seen in Panel A of Figure 3-4, the tryptophan emission peak continues to decrease until the IAEDANS concentration is 40X molar excess of the MESD concentration. After that point, the intensity of the peak does not change indicating that 40X molar excess of IAEDANS sufficiently labels all the MESD proteins in the sample. Panel B shows a similar response to direct excitation of IAEDANS. The IAEDANS emission peak continues to increase until the IAEDANS concentration is 40X molar excess of the MESD concentration. From this point, all labeling of MESD with IAEDANS used 40X molar excess of the label.



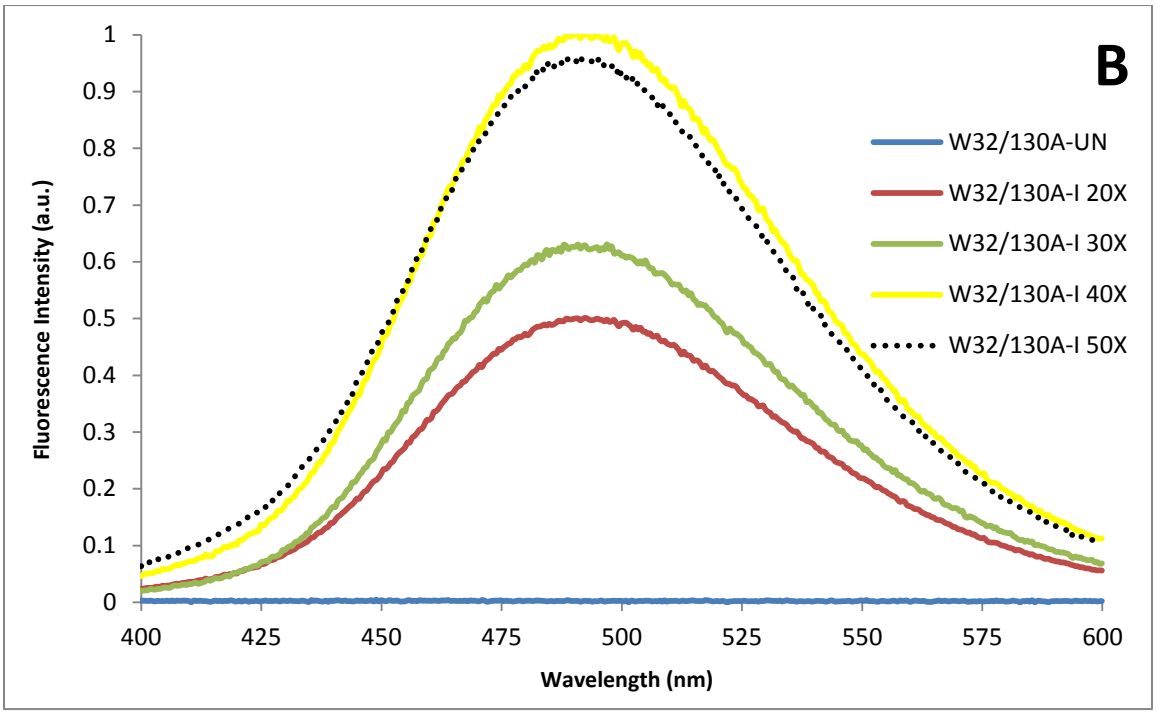


Figure 3-4: Panel A shows emission range 300 – 400 nm at λ_{ex} = 290 nm. Panel B shows emission range 400 – 600 nm at λ_{ex} = 336 nm. W32/130A-UN designates the unlabeled protein while W32/130A-I designates the IAEDANS labeled protein.

3.3.3 *In vitro* FRET measurements and calculation of FRET-distances

One method to calculate intramolecular distances between a donor and acceptor within a protein is to determine the intensity difference of the donor emission peak with a protein containing only a donor (tryptophan) and a protein containing both a donor and acceptor (tryptophan-IAEDANS). The two samples were prepared for each mutant and their concentrations were equilibrated since intensity is also concentration dependent. Emission scans were collected in the range of 305 – 600 nm with an excitation set at 295 nm. The same fluorescence spectra were collected using the buffer that was used to dissolve the fluorescence labeled protein, serving as the baseline.

After the data has been collected, it is exported to an excel file. Each of the collected FRET spectra was baseline corrected by removing the buffer spectrum from each mutant's spectrum. After baseline correction for each mutant, the tryptophan emission peaks, both labeled (Q_D) and unlabeled (Q_{DA}), are found for each mutant. The FRET-distances were calculated based on the following equation:

$r = R_0 \left(\frac{1}{1 - \frac{Q_{DA}}{Q_D}} - 1 \right)^{1/6}$, where R_0 for the tryptophan-IAEDANS pair is equal to 22 Å.

Figures 3-5, 3-6 and 3-7 show the spectra for the three single tryptophan MESD mutants as well as their distance calculations.

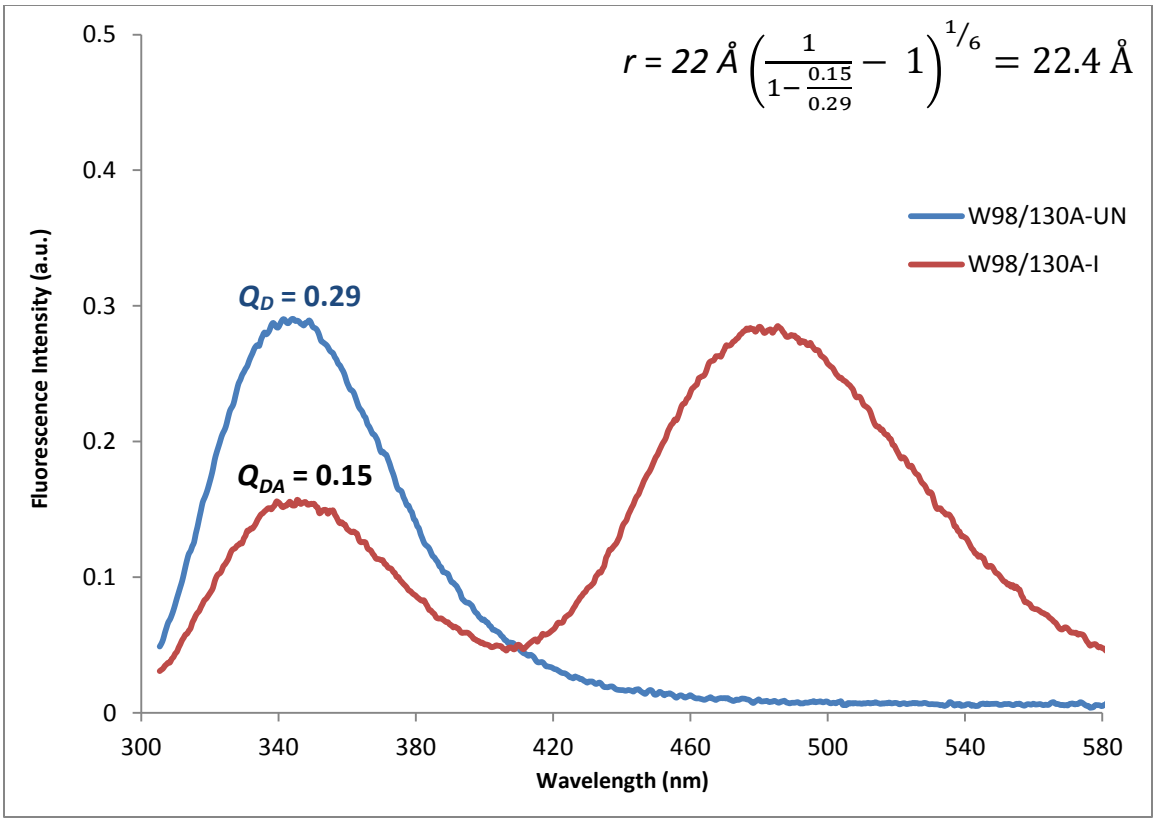


FIGURE 3-5: Emission scan of MESD_W98/130A with a λ_{ex} = 295 nm and an emission range of 305-600 nm. The tryptophan emission peaks are represented by Q_D for the unlabeled protein and Q_{DA} for the IAEDANS labeled protein.

Figure 3-5 shows an example of emission scan for MESD_W98/130A. The blue line represents the emission of the unlabeled protein, while the red line represents the emission of the IAEDANS labeled protein. The expected distance based on the NMR structure is 23.5 Å, which is quite close to the R_0 value of 22 Å. Since the R_0 value is equal to the distance at which 50% of the donor's emission is captured by the acceptor molecule, we would expect to see the donor intensity decrease by half. In fact, we do observe this halving of the intensity as shown above.

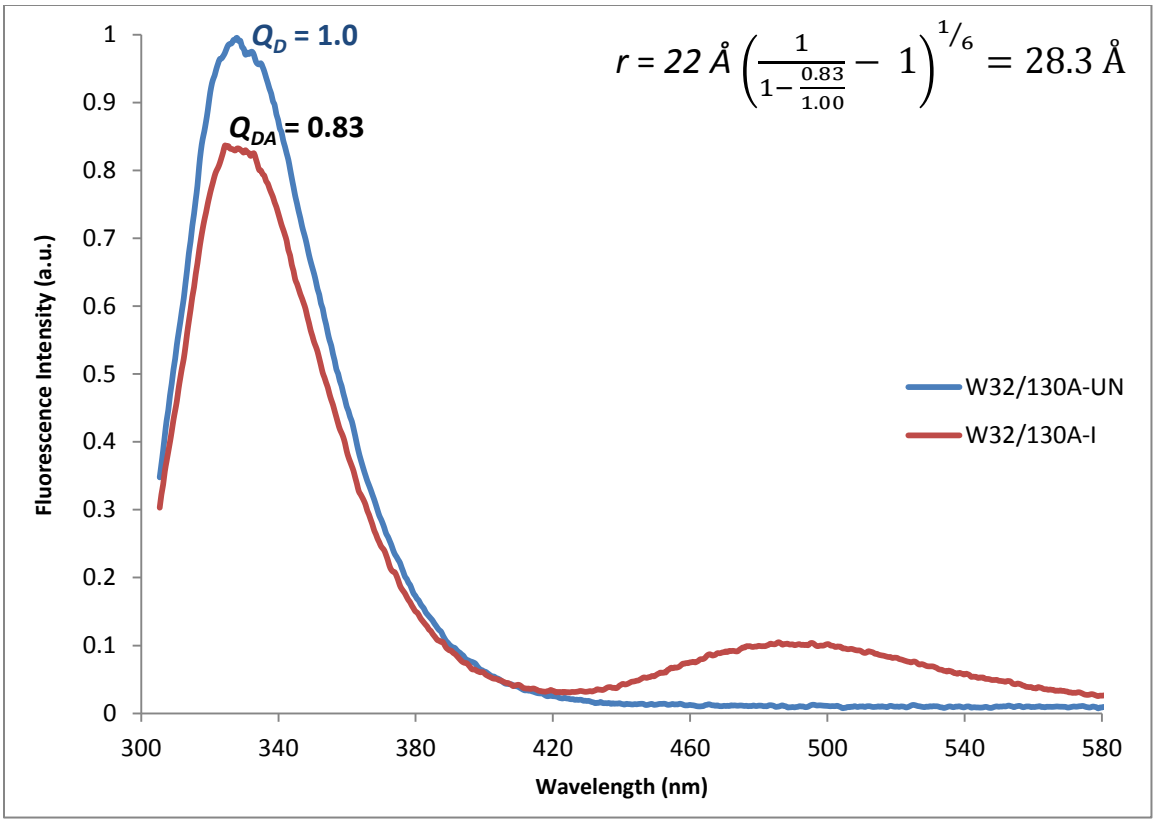


Figure 3-6: Emission scan of MESD_W32/130A with a λ_{ex} = 295 nm and an emission range of 305-600 nm. The tryptophan emission peaks are represented by Q_D for the unlabeled protein and Q_{DA} for the IAEDANS labeled protein.

Based on the observed NMR distance of 26.1 Å, we expect to see the least change in intensity as the W98 residue is furthest from residue C142. Compared to figures 3-5 and 3-6, we do in fact see the least change in intensity and thus, the largest calculated distance.

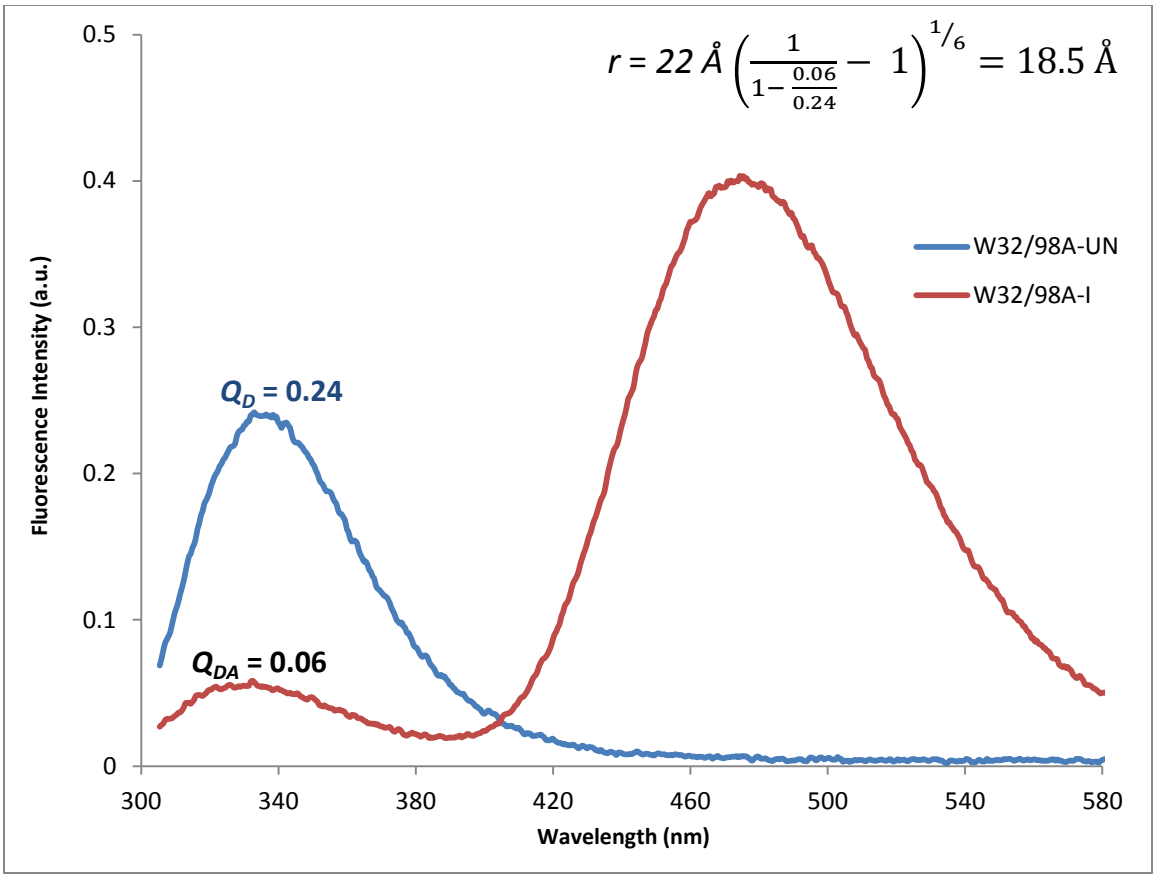


Figure 3-7: Emission scan of MESD_W32/98A with a λ_{ex} = 295 nm and an emission range of 305-600 nm. The tryptophan emission peaks are represented by Q_D for the unlabeled protein and Q_{DA} for the IAEDANS labeled protein.

Since residue W130 is expected to be the closest to residue C142, with an expected distance based on the NMR structure of 20.5 Å, the acceptor molecule should accept most of the W130 emission. As shown above, we see a 4-fold decrease in intensity and the shortest calculated distance.

Table 3-2 lists the calculated distances for each mutant with a standard deviation based on six separate experiments. The calculated distances are similar to the observed distances from the NMR structure. The NMR structure did not have an IAEDANS label, which could account for the ~ 2 Å differences seen. In addition, the addition of the IAEDANS label could slightly alter the structure. A simple ^{15}N-HSQC NMR experiment could indicate whether or not there are any structural changes after fluorescence labeling.

TABLE 3-2: Calculated distances versus observed distances

Mutant	***Calculated Distance***	***Observed Distance (NMR)***	***Difference***	***Percent Error***
W32/98A	18.5 ± 0.4 Å	20.5 Å	-2.0 Å	11%
W32/130A	28.4 ± 0.8 Å	26.1 Å	+2.3 Å	8%
W98/130A	22.3 ± 0.7 Å	23.5 Å	-1.2 Å	5%

3.3.4 Assignment of FRET-peaks using synchronous scanning fluorescence spectroscopy

Synchronous scanning fluorescence spectrometry takes advantage of the ability to vary both the excitation and emission wavelengths simultaneously during the data collection. In this technique, the fluorescence signal is recorded when excitation and emission wavelengths are simultaneously scanned keeping in between a fixed

wavelength interval (called the offset value, $\Delta\lambda$) throughout the spectrum. As a result, the selectivity for individual components is considerably improved; additionally, much more information on mixtures of fluorescent compounds is gained [43, 44].

In the case of MESD, we have mutants that contain two donors in addition to the wild-type MESD which contains three donors. The utilization of synchronous scanning was explored to determine whether the FRET peaks can be separated for each donor to acceptor using a multiple donor/one acceptor protein sample.

3.3.4.1 Tryptophan residues of MESD have unique environments

Typically, synchronous scanning is used to help identify various fluorescent components within a mixture. The tryptophan residues of MESD are each in their own unique environment and could possibly have unique spectral patterns. Indeed, based on the NMR structure of MESD, we found that each tryptophan has a distinct surface exposure value and forms possible hydrogen bonds with nearby residues (Table 3-3). The immediate chemical environment of tryptophan residues can have effects on its fluorescent properties resulting in slight red or blue shifts of its emission spectrum, causing different wavelengths of the FRET-peaks for different tryptophan donor/IAEDANS acceptor pairs [45]. These wavelength differences may possibly allow us to assign the FRET-peaks to an individual FRET donor/acceptor pair, if these synchronous scanning fluorescence spectra can separate the individual FRET-peaks.

Table 3-3 illustrates the uniqueness of each environment surrounding the tryptophan residues. Residue W32 is highly exposed to the solution, but does not seem

to participate in any hydrogen bonds. Residue W98 is deeply buried and has one potential hydrogen bond with the carbonyl oxygen acting as the acceptor. Residue W130 is in an intermediate state of surface exposure and potentially participates in two hydrogen bonds in which both backbone and sidechain atoms are involved.

Table 3-3: Environment of MESD tryptophan residues

Residue	ASA*	Possible H-Bonds (Donor:Acceptor)
W32	0.78	--
W98	0.13	SER 101 (O_γ—C_β) : **TRP 98** (O—C)
W130	0.43	ASP 134 ($O_{\delta 2}$—C_γ) : **TRP 130** (O—C)
		TRP 130 ($N_{\varepsilon 1}$—C_γ): ASP 108 ($O_{\delta 2}$—C_γ)

*ASA: Accessible surface area.

3.3.4.2 Optimization and organization of the synchronous scanning fluorescence spectra

In order to determine the best offset value ($\Delta\lambda$) to use for the synchronous scans of MESD, an extensive amount of emission ranges were tested. The most interesting data and clearly defined peaks came from using an excitation range of 240 -370 nm with a $\Delta\lambda$ = 210 nm, therefore the collected emission range was 450 – 580 nm. Once the emission and excitation ranges were determined, synchronous scans could be collected for wild-type MESD and all its mutants.

In order to perform FRET-peak assignment, the collected spectra are divided into three subgroups. Each subgroup contains all the spectra for proteins containing a particular tryptophan residue. As an example, one subgroup would contain all the spectra for proteins containing the W32 residue: MESD_wt, MESD_W98A, MESD_W130A, and MESD_W98/130A. The “no tryptophan” mutant spectrum was included in each subgroup to be used as a baseline of sorts.

3.3.4.3 Separation of the FRET-peak for each TRP donor-IAEDANS acceptor in MESD

Figures 3-8, 3-9 and 3-10 show the spectra obtained from these synchronous scanning experiments. In each figure, Panel A shows the full emission spectra collected for each protein with three distinct regions of peaks. Panel B is a zoomed-in spectra showing the emission region of 480 – 510 nm and Panel C shows the difference spectra after the “NoW” mutant spectra has been subtracted from the MESD protein spectra, treating it similar to a baseline.

Focusing on the emission range of 480 – 510 nm, we believe that this corresponds to the “FRET peak”. The FRET peak is the IAEDANS emission observed based on its excitation by tryptophan. The emission range of 480 – 510 nm coincides with an excitation range of 270 – 300 nm which does in fact cover tryptophan’s absorbance spectrum. When looking for FRET peaks, the first comparison is between the “no tryptophan” mutant and the single tryptophan mutant. The new peak observed in the single tryptophan mutant spectra, as compared to the “no tryptophan” mutant, is assigned for this tryptophan residue. By identifying the FRET peak for each tryptophan

residue, we can tentatively assign these single tryptophan residues. This tentative assignment can be verified by comparisons with the double tryptophan mutants for confirmation. By the time wild-type MESD is analyzed, the FRET spectra should contain all three FRET peaks for the three tryptophan residues in the MESD protein.

3.3.4.4 Assignment of the FRET peak for each tryptophan residue of MESD

Starting with Figure 3-8, the spectra are shown for all MESD and mutants containing residue W32 as well as the no tryptophan mutant. Panel A shows the full spectra, while Panel B zooms in on the FRET peak region. To create Panel C, the spectra of the no tryptophan mutant was subtracted from each mutant containing W32.

First, looking at the single tryptophan mutant (MESD_W98/130A) that contains W32, there is one broad peak at 502 nm and has been assigned and labeled as W32. Next, looking at the double tryptophan mutant containing W32 and W130, there are two distinct peaks. The first peak is at 495 nm and the second is also around 502 nm. If we assume that the peak at 502 nm is W32, then the peak at 495 nm must be from W130. Labeling the 495 nm peak as W130 also makes sense since the intensity of the W130 should be higher than both W32 and W98 as it is the closest to the IAEDANS labeled C142 residue. Due to the close proximity, IAEDANS should capture more of W130's emission and have a more intense FRET-based emission. After examining all the spectra, it is noted that whenever the W130 residue is present in a protein, there is a more intense peak around 495 nm. Finally, after examining the wild-type spectra, there are once again two distinct peaks. Residue W98 could have an overlapping FRET peak

with either W130 or W32. After examining the other spectra in Figures 3-9 and 3-10, it has been determined that W98 has an overlapping FRET peak to W32 and has been labeled as such.

The same process was repeated for each subgroup of proteins. In Figure 3-9, residue W98 has a FRET peak at 503 nm, which indeed is very similar to the W32 FRET peak. The FRET peak strategy allows us to confidently assign the FRET peaks for all three tryptophan residues (Table 3-4). Although the data is conclusive, further exploration into this FRET assignment technique may be fruitful. However, another challenge to overcome would be the ability to calculate distances based on changes in FRET peaks and not changes in donor emission.

Table 3-4: FRET peak assignment

Tryptophan Residue	W32	W98	W130
FRET Peak (nm)	502	503	495

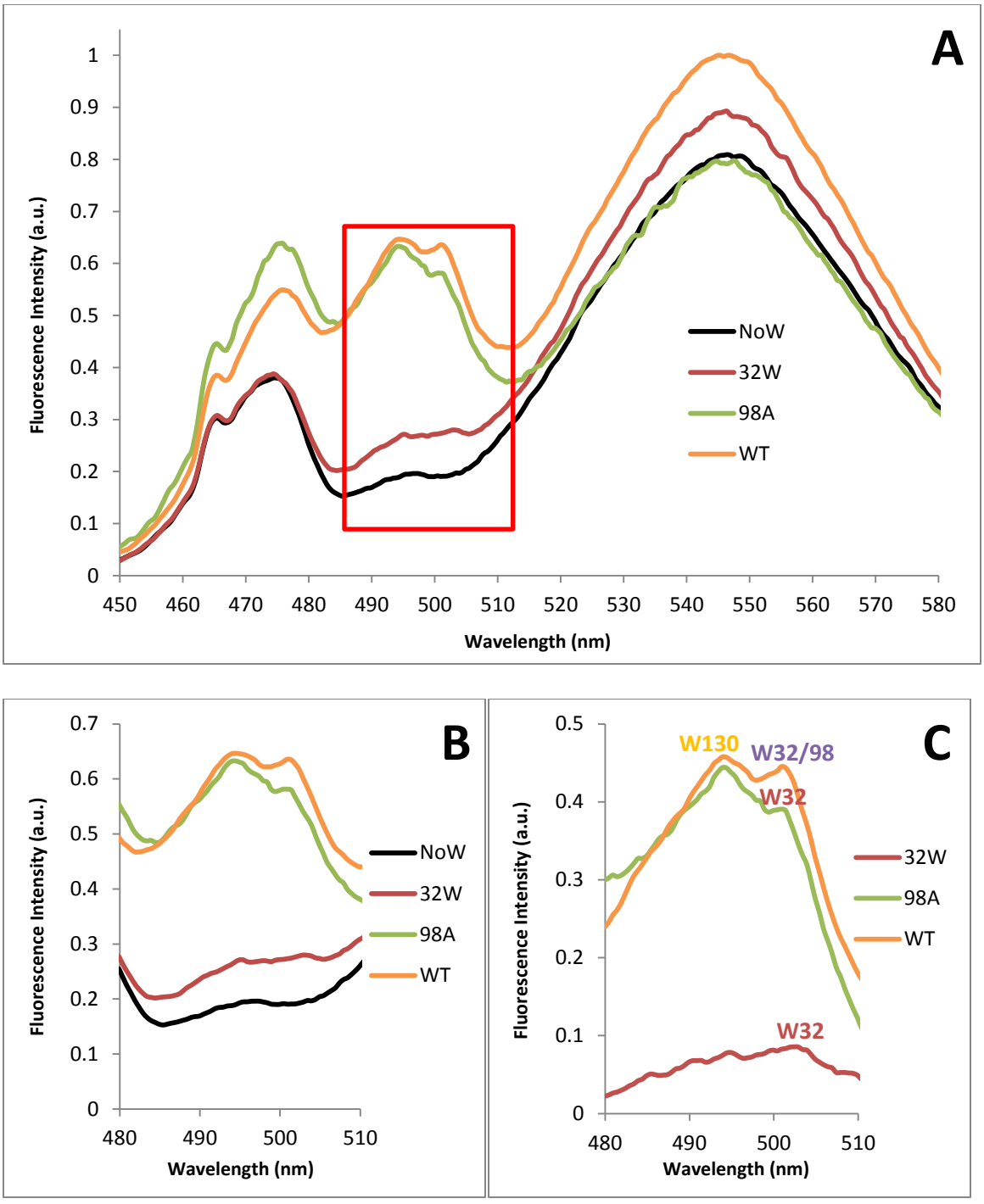


Figure 3-8: Synchronous scans for all MESD proteins containing residue W32. Panel A is full spectrum. Panel B is zoomed in on the FRET peak region. Panel C is the difference spectra with peak assignments.

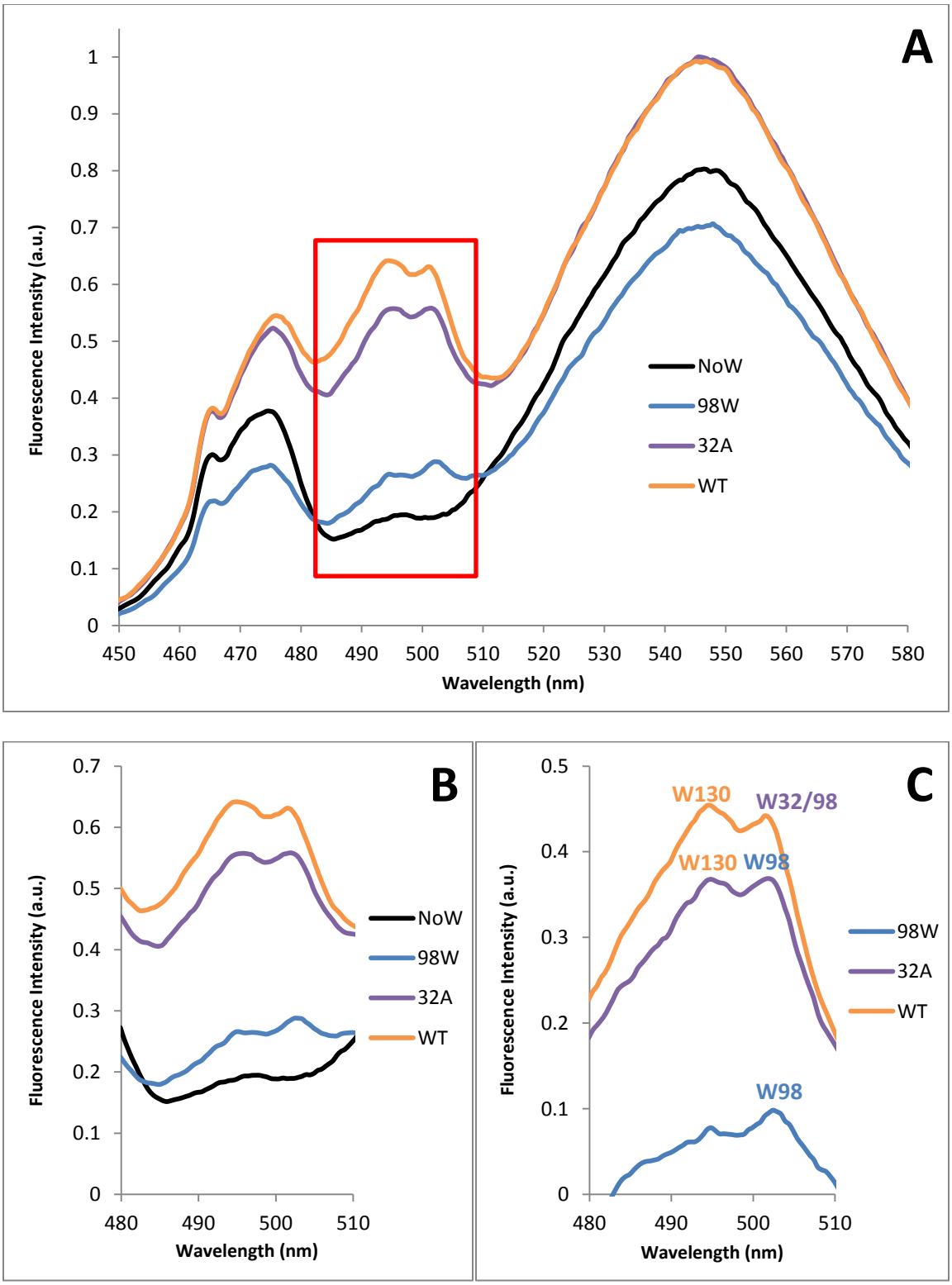


Figure 3-9: Synchronous scans for all MESD proteins containing residue W98. Panel A is full spectrum. Panel B is zoomed in on the FRET peak region. Panel C is the difference spectra with peak assignments.

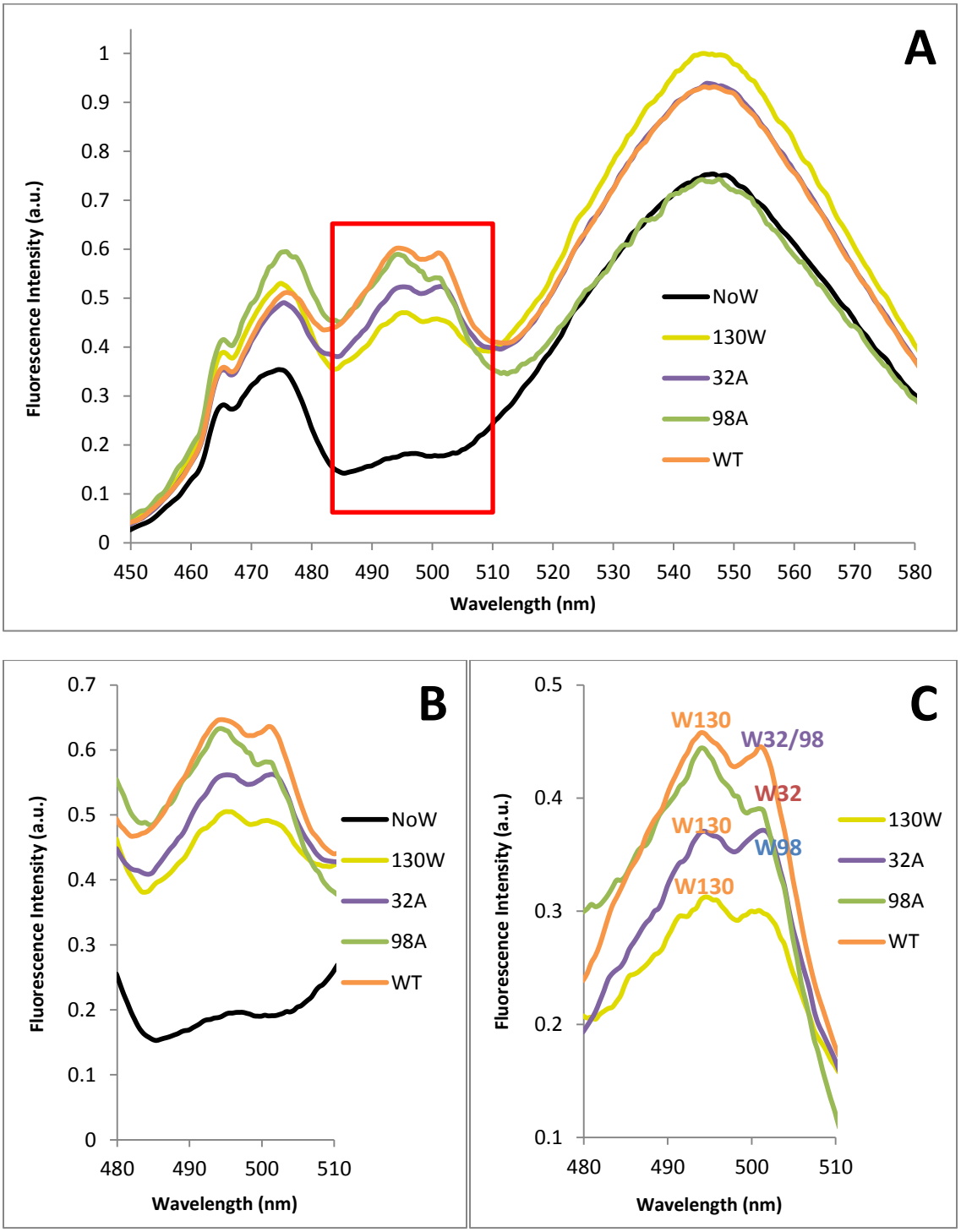


Figure 3-10: Synchronous scans for all MESD proteins containing residue W130. Panel A is full spectrum. Panel B is zoomed in on the FRET peak region. Panel C is the difference spectra with peak assignments.

3.3.5 MESD bacterial expression using auxotrophic bacterial stain: DL41 (DE3).

The double mutant plasmids were transformed into the DL41 (DE3) auxotrophic bacterial strain. After transformation, the colonies were screened using double colony selection for each mutant. Both transformation and double colony expression utilized rich medium and regular tryptophan. Figure 3-11 shows an example of expression of two of the three double mutants after double colony selection. The second and third lanes compare a non-induced and induced expression of MESD_W32/130A. The fourth and fifth lanes compare a non-induced and induced expression of MESD_W32/98A.

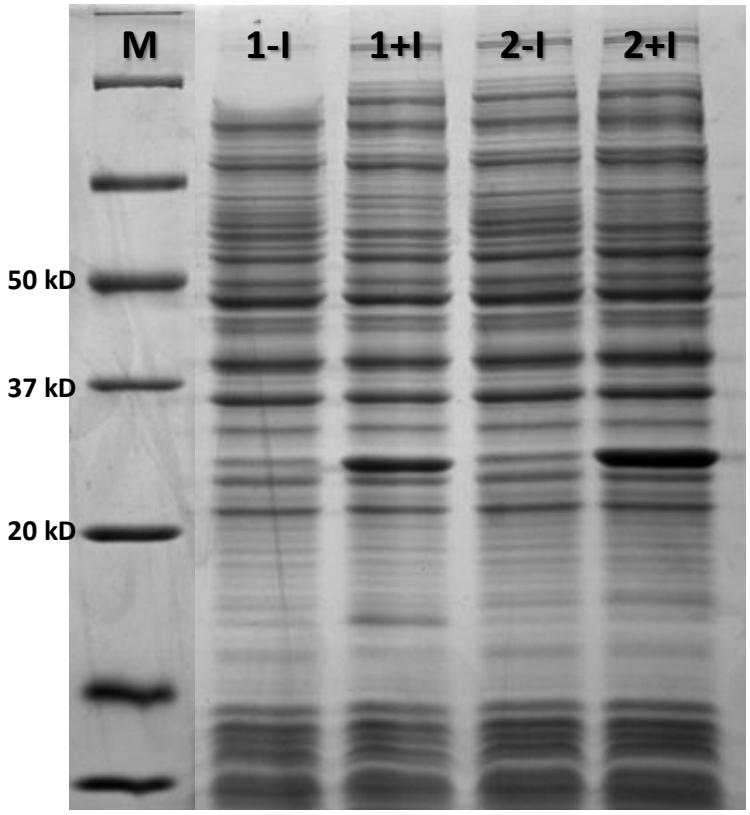


Figure 3-11: 12% SDS-PAGE of samples taken after expression. M = Marker. 1 = W32/130A, 2 = W32/98A, -I = uninduced and +I = induced with IPTG.

3.3.6 Optimization of 5-HT-labeling of MESD

The initial attempts to selectively label MESD with 5-HT were performed using the high-cell density method. The medium was supplemented with MET and 5-HT, but the protein yield was quite low after performing SDS-PAGE analysis. After several failed attempts using the high-cell density method, the traditional IPTG method was used for expression, but the yield was even lower. We then focused back on the high-cell density method with the following modifications.

First, instead of growing the starting culture in rich medium, the starting culture was grown in a minimal medium containing glucose, ammonium chloride, MET and TRP (both amino acids at 25 mg/L) for ~20 hours overnight. The following morning, the OD_{600} was ~ 1.0 and the culture was spun down at 5,000 x g for 7 minutes. The pellets were re-suspended in 1X PBS and vortexed briefly in order to remove as much regular TRP as possible. After spinning down the cells again and removing the supernatant, the pellets were re-suspended in the defined medium listed in Table 3-5, with the exception of 5-HT. The flask was placed back in the 37 °C shaker for two hours in order to "starve" the cells and to use all the remaining TRP in its reserves. After the two hour starvation period, 5-HT was added and the flask was placed back in the shaker for 30 minutes. At this point, 0.5 mM IPTG was added to induce protein production. The bacterial cells were cultured for 5 hours at 37 °C before harvesting. Removing the rich medium source and starving the cells seemed to be most beneficial for 5-HT incorporation.

Table 3-5: Defined medium for 5-HT labeling

Component	***Amount per L***
5x M9 Minimal Salts (-NH_4Cl)	200 mL
1M $MgSO_4$	2 mL
1M $CaCl_2$	0.1 mL
1000x Trace Metals	0.25 mL
Kanamycin (30 mg/mL)	1 mL
19 standard amino acids (-W)	50 mg
5-hydroxy-L-tryptophan	50 mg

To ensure that the 5-HT was being taken up by the bacteria, the uptake of 5-HT was monitored by taking samples of the medium every hour. Furthermore, it also tested whether or not adding additional boosts of 5-HT made a difference in expression levels and incorporation of 5-HT into MESD. Figure 3-12 shows two spectra of the uptake experiments. In Panel A, starting at 0 hours (the point at which IPTG was added to induce culture), a significant decrease in intensity was observed after 1 hour. At the 2-hour point, only a slight intensity decrease was observed, suggesting that the uptake of 5-HT into the bacterial cells had significantly slowed down (light blue dotted line). At the 2 hour mark, additional 5-HT was added and a new medium sample was taken (dark red line). Samples were again collected after 3 and 4 hours, and decreases in intensity

where noted for both time points, with a similar pattern as we observed for the first few hours.

Generally, significant 5-HT uptake was observed during the first hour, but significantly slows during the second hour. After the fourth hour, additional 5-HT was added and another medium sample was taken (dark green line). There was a decrease in intensity seen at the 5 hour mark (light green line), but there was not a notable decrease at the 6 hour mark (black dotted line) , again repeating the same pattern as the first four hours. This data indicate that the bacterial cells seem to uptake the 5-HT during the first hour significantly, however, this uptake is inhibited somehow during the second hour. Based on this result, we believe that it is critical to add more 5-HT during expression for efficient labeling of the MESD protein with 5-HT. Panel B was a control experiment of the medium containing 5-HT without any bacterial cells to see if intensity decreased for some other reason besides uptake. No changes in intensity were noted for the 6 hour period. Based on the results shown in Figure 3-12, the bacteria do not seem to have a problem with the uptake of 5-HT, however, additional amount of 5-HT have to be added every 2 hours during bacterial expression to efficiently label MESD with 5-HT.

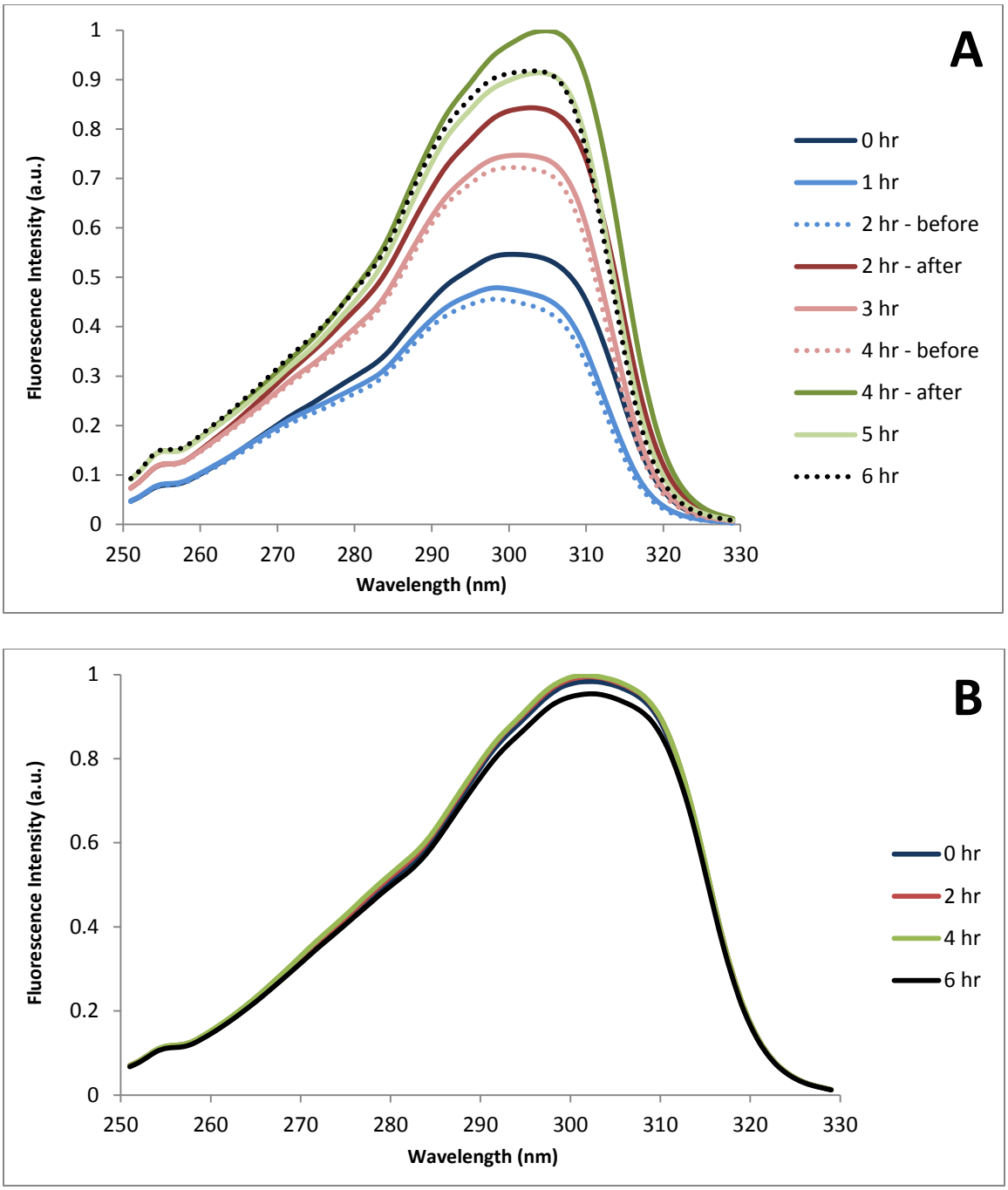


Figure 3-12: Absorption spectra of culture medium with (Panel A) and without (Panel B) bacteria. Excitation range = 250 – 330 nm, λ_{em} = 336 nm.

3.3.7 Characterizations of the 5-HT labeled MESD

In order to determine if 5-HT had been incorporated into MESD, fluorescence absorption spectra were collected for the labeled proteins. If MESD has been labeled with 5-HT, there should be a noticeable shoulder on the right half of the spectra. Figure 3-13 shows two absorbance spectra of a protein produced before (Panel A) and after labeling optimization (Panel B). It is clear in Panel B that there is a shoulder and the protein can be excited at 310 nm, indicating that 5-HT was efficiently incorporated into MESD. Under this condition, we estimated that about 50% of MESD was labeled with 5-HT.

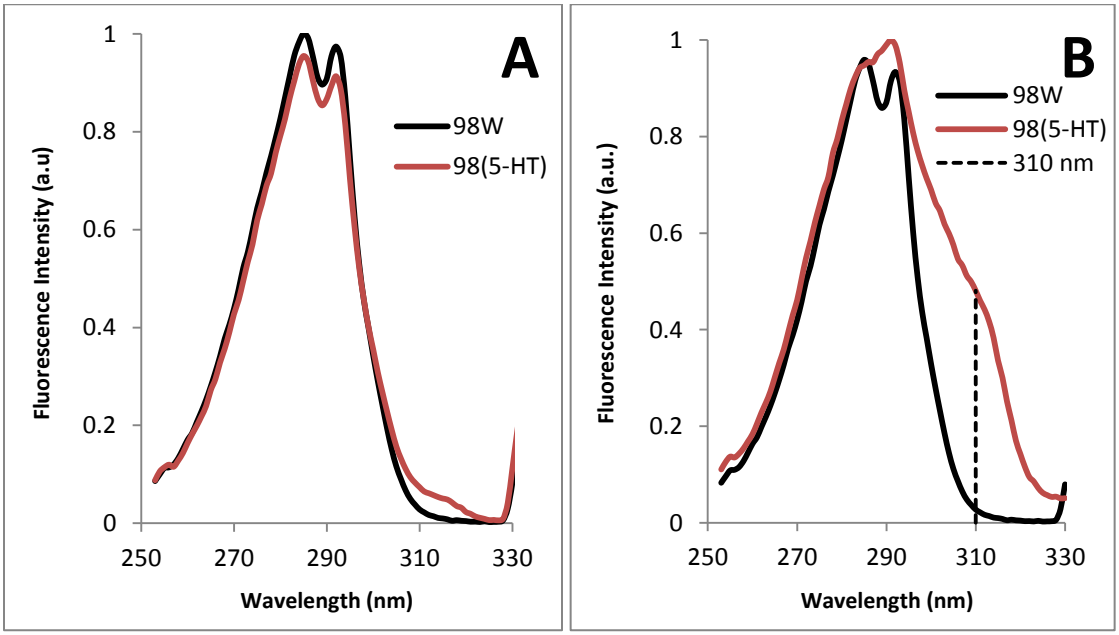


Figure 3-13: Absorbance spectra (λ_{em} = 332 nm). Panel A – before optimization of expression conditions. Panel B – after optimization of expression conditions.

Figure 3-14 shows the fluorescence emission spectra of the same protein with λ_{ex} = 295 nm and λ_{ex} = 310 nm. The protein expressed with regular tryptophan can be excited at λ_{ex} = 295 nm, as shown by the black solid line. However, exciting that same protein at λ_{ex} = 310 nm results in a nearly flat line (black dotted line).

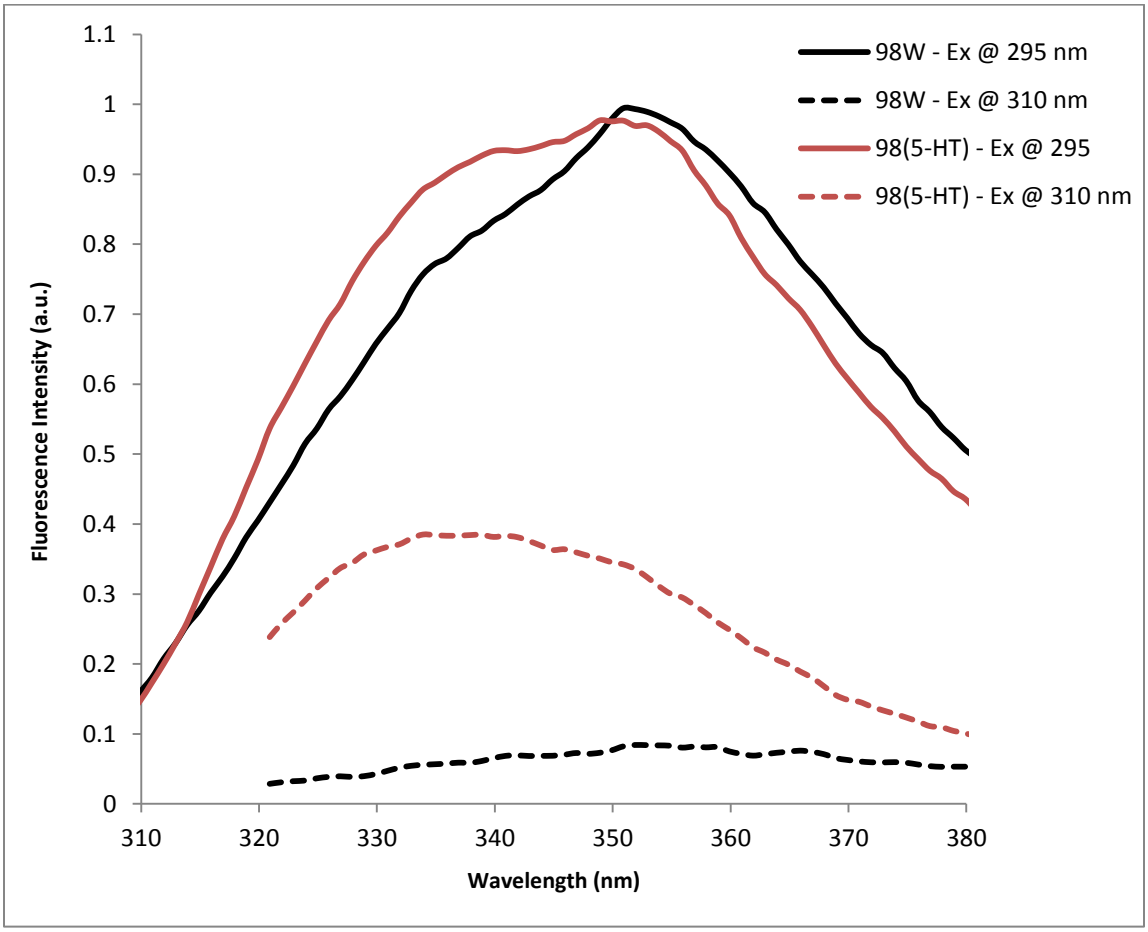


Figure 3-14: Emission spectra of MESD_W32/130A with λ_{ex} = 295 nm and λ_{ex} = 310 nm.

After examining the protein that has been labeled with 5-HT, Figure 3-14 demonstrates that it also can be excited at λ_{ex} = 295 nm (red solid line). As opposed to the protein expressed with regular tryptophan however, we see that this protein can be

excited when λ_{ex} = 310 nm (red dotted line), although the intensity is much lower. Once the labeling efficiency of MESD with 5-HT is optimized, the excitation intensity should be significantly enhanced.

3.4 Conclusions

In order to perform in-cell FRET experiments, a protein must be labeled in such a way that will allow for specific excitation of the target protein inside the cells without excitation of the intracellular background proteins. Our rationale is to use tryptophan residues as the fluorescence donor and IAEDANS labeled on the cysteine residue as the fluorescence acceptor to minimize any potential structural perturbation of the protein of interest. We have chosen MESD as the model protein for our studies since this protein contains three tryptophan residues (W32, W98 and W130) that can be used as the FRET donors and a single cysteine at residue 142 which allows for labeling with a small fluorophore, IAEDANS, as the fluorescence acceptor. MESD is also a good choice since it is an ER-resident protein and should remain in the ER and Golgi during the FRET experiments. After a few rounds of mutagenesis, three double W-to-A mutants were made that contain one tryptophan residue each (single tryptophan mutants), since the traditional FRET-based distance calculations require one donor and one acceptor within MESD. We also prepared three double tryptophan mutants of MESD and these mutants allows us to explore a methodology that possibly permits us to calculate two FRET-distances from a single FRET experiment. Furthermore, this methodology may also allow

us to calculate three FRET-distances from a single FRET experiment using wild-type MESD that contains three tryptophan donors.

To verify our *in vitro* FRET-technique, we performed *in vitro* FRET experiments of the single tryptophan mutants after labeling each mutant with a thiol-reactive probe (IAEDANS) at C142 of MESD. We compared the distances calculated from the FRET experiments to the distances observed on the NMR structure. Our calculated FRET-based distances of the three single tryptophan mutants are nearly identical to the distances of these tryptophan residues to C142 observed in the NMR structure of MESD, demonstrating that the validity of our FRET-based distance calculations using the strategy established. This allows us to verify the FRET technique first works *in vitro* before moving to an in-cell experiment.

In addition to verifying the standard *in vitro* FRET-based distance calculations, we also explored the ability to calculate distances based on multiple donors and one acceptor. Synchronous scanning fluorescence spectroscopy has been used to identify multiple fluorescent components contained in a mixture by varying both the excitation and emission wavelengths simultaneously during the data collection. However, this technique has not been used on proteins to detect multiple tryptophan residues. Typically, tryptophan residues in a protein are in their own unique chemical environments and their emissions may vary slightly depending upon the nature of the differences of those chemical environments. After optimizing the excitation and emission ranges, as well as the offset value ($\Delta\lambda$), spectra were collected of the wild-type MESD protein and all of the MESD tryptophan mutants. After careful analysis, FRET-

peaks were assigned for all three tryptophan residues of MESD. We believe further exploration of this technique could yield interesting results leading to the possibility of calculating multiple distances based on a single FRET experiment. The success of this strategy will significantly reduce the amount of work involved in sample preparation during FRET-distance calculations, holding the potential to push the FRET technique into an atomic resolution structural biology tool.

Once the *in vitro* data showed that the FRET measurements were similar to the NMR structural data, we moved on to the next step of incorporating a tryptophan analogue, 5-HT, that would allow for specific excitation of the target protein without excitation of the background cellular proteins inside living cells. Several modifications of the minimal medium used and the high cell density method resulted in MESD mutant proteins that displayed fluorescent properties similar to free 5-hydroxy-L-tryptophan, demonstrating that our current optimized method efficiently labels MESD with 5-HT. This success allows us to pursue in-cell FRET experiments to study protein structure inside the living cells, which is the main focus of Chapter 4.

3.5 Discussion

While the overall goal of this thesis is to develop an in-cell FRET technique for protein structural determination in living cells, this chapter focused on verifying several methods and ideas *in vitro* first. First, the accuracy of the FRET technique to calculate distances between a donor and acceptor within a protein must be validated. By

comparing the FRET-based distances to the NMR protein structure, this will confirm that FRET is a valid structural tool at atomic resolution.

Since MESD contains three tryptophan residues, a series of mutagenesis was performed that resulted in a total of seven mutants: 3 single tryptophan mutants, 3 double tryptophan mutants and a "no tryptophan" mutant. MESD contains a single cysteine residue, so once the proteins were expressed and purified, they were labeled with a thiol-reactive fluorophore, IAEDANS, to act as the acceptor. IAEDANS has a peak excitation that overlaps tryptophan's peak emission. Emission scans were collected for each IAEDANS labeled double mutant (one donor: one acceptor) as well as the buffer which was used for baseline correction. Finally, distances were calculated based on changes in donor emission intensity. Comparing the calculated distances to the NMR protein structure, they were found to be within 2 Å of each other. Therefore, the FRET technique was validated as an atomic resolution structural technique.

Traditional *in vitro* FRET-based distance calculations within a protein require one donor and one acceptor, resulting in a measured distance between that donor and acceptor. Currently, it is impossible to use FRET to calculate multiple distances within a protein containing multiple donors or acceptors. In order to make in-cell FRET a feasible structural technique, we explored a FRET technique that obtains multiple distances from a single measurement using protein samples that contain multiple donors and a single acceptor.

When performing an emission scan on a protein sample labeled with multiple donors and a single acceptor, the emission of each donor coalesces into a single uniform

spectra. Synchronous scanning fluorescence spectroscopy has the ability to detect multiple fluorescent components within a single mixture. The donors in the case of MESD are native tryptophan residues and each residue resides in its own unique chemical environment. It might be possible to detect slight changes in tryptophan emission and IAEDANS excitation using synchronous scanning fluorescence spectroscopy.

Extensive optimizations were done to ensure that the best excitation and emission ranges were chosen for these studies. The excitation range of 240 – 370 nm with an offset value of 210 nm which leads to an emission range of 450 – 580 nm displayed the most interesting results. Upon examination, three regions of peaks are clearly defined. Region 1 includes emissions 465 – 485 nm, region 2 includes emissions 485 – 510 nm and region 3 includes emissions 510 – 580 nm. Region 1 has been explained as the result of exciting the 11 tyrosine and phenylalanine residues of MESD and their combined effect to excite IAEDANS. Region 3 is the IAEDANS emission as a result of direct excitation of the fluorophore.

We are most interested in region 2 as this is the FRET peak region and we can clearly see the most differences between the different spectra in this region. Each MESD protein is of equal concentration, so the only difference between each protein is the amount of tryptophan residues present. By careful comparisons of the fluorescence emission spectra of these samples, we identified the FRET regions of the multiple tryptophan donors to a single IAEDANS acceptor, allowing us to observe different FRET peaks. By comparing the FRET regions of the wild-type MESD with the single and double

tryptophan mutants of MESD, we assigned the FRET peak for each tryptophan residue. The FRET peak resulting from residue W130 was found to be around 495 nm, while residues W32 and W98 were found to be around the 502 – 503 nm range. Although this technique needs further optimization, it represents the first step for measurement of multiple distances from a single FRET experiment. In our opinion, this is the key step towards development of a FRET technique into a high-resolution structural biology tool to study protein structure inside the cells.

For future experiments, the collection of region 1 and region 3 is unnecessary. The information that is necessary for structural purposes is found within region 2 – the FRET region. When designing experiments, the excitation range of the synchronous scan should cover the full absorption range of the donor. In turn, the emission range of the synchronous scan should be centered around the peak emission of the acceptor.

Finally, since tryptophan is being used as a donor, intracellular background proteins containing tryptophan could result in interference when performing in-cell FRET experiments. Since both *in vitro* FRET experiments were validated and explored, the next task was to selectively label MESD with 5-HT that would ensure specific excitation of the 5-HT labeled protein without excitation of background cellular proteins that contain regular tryptophan residues. An auxotrophic bacterial strain, DL41(DE3), was used that is auxotrophic for methionine and tryptophan. After various attempts to express the proteins within this strain, it was found that an altered high-cell density method lead to the best expression and incorporation of 5-HT into MESD. Incorporation of 5-HT into the protein was confirmed by performing excitation and emission scans

demonstrating that the protein could be excited at λ_{ex} = 310 nm. Although, incorporation has not reached 100%, further optimizations of the expression conditions should yield an MESD protein with full incorporation of 5-HT. Once full incorporation of 5-HT has been achieved, a ^{15}N-HSQC NMR experiment should be collected and compared to wild-type MESD to ensure that there are no major structural changes after addition of the amino acid analogue.

This chapter has laid the foundation for future development of an in-cell FRET methodology. MESD serves as a good model for *in vitro* work and will continue to be a good model for future in-cell work. Since MESD is an ER-resident protein due to its "RDEL" ER retention signal, once delivered inside living cells using QQ-protein delivery, it will remain in the ER and Golgi, making any information gained physiologically relevant. Also, since the *in vitro* FRET-based distance calculations are comparable to the NMR protein structure, the FRET-technique is a feasible option for measuring distances within a live cell.

In regards to the synchronous scanning fluorescence data and analysis, this illustrates the potential that FRET can be potentially used as an in-cell structural biology technique at atomic resolution. This synchronous scanning fluorescence technique has never been used for proteins, but in principle it should be able to detect the subtle differences of tryptophan's emissions based on each residue's unique chemical environment. In our analysis, we were able to visualize differences between different protein spectra in the FRET-peak region of the emission spectra. The other two regions did not display any differences in the shape of the spectra, but since the FRET-peak

region is dependent upon the number of tryptophan residues present in each protein, it is here that we see differences.

Further optimization and analysis of synchronous scanning fluorescence spectroscopy using protein samples that contain multiple fluorescence donors and one single acceptor could yield interesting results. Indeed, our results shown in this chapter suggest that the methodology developed in this chapter a potential to calculate multiple distances based on a single FRET experiment. In addition, expression and selective labeling of MESD with 5-HT needs to be further optimized until greater than 80% incorporation is achieved. This will be beneficial for future in-cell work to guarantee the intensity is strong enough to allow for usable FRET data. However, while there is still room for improvement, we feel that a strong enough foundation has been laid so that we can move on to developing an in-cell FRET methodology.

CHAPTER 4

DEVELOPMENT OF AN IN-CELL FRET TECHNIQUE:

OPTIMIZATIONS OF IN-CELL EXPERIMENTAL CONDITIONS

4.1 Introduction

The ultimate goal of this thesis project is to develop an in-cell fluorescence technique that allows for measurement of the distances between fluorescence donors and acceptors within a protein or between two proteins inside the correct intracellular compartment of living cells. The successful achievement of this goal will allow us to obtain high-resolution structural information within a protein or between two proteins inside living cells, a key step towards high-resolution structural biology of proteins inside the living cell.

Development of this methodology is extremely challenging. Proteins only perform their biological functions inside the correct intracellular compartments and different intracellular compartments may display different chemical environments. In addition, the intracellular environment is highly crowded and harbors an intricate network of biological activities simultaneously due to protein interactions with their partners. Different intracellular events, such as folding, post-translational modification, protein interactions, intracellular trafficking and secretion, may cause different protein concentrations and structures in different compartments [112]. This complex cellular environment suggests that protein concentration and structure inside living cells are

spatiotemporal dependent, making high-resolution structural studies of proteins inside living cells extremely challenging.

Fluorescence spectroscopy has been used on cell cultures before, but has not been used in an attempt to collect high-resolution structural information of a target protein inside living cells [17, 29, 37, 38, 41, 42]. This project is challenging for a number of reasons. First, the intracellular protein concentration of a target protein may vary in different intracellular compartments. However, these concentrations must be known in order to accurately calculate distances based on the current *in vitro* FRET approach [41]. Currently, it is impossible to determine the intracellular concentration of a protein inside a specific intracellular compartment of living cells. Second, a protein may display different conformations in different compartments due to changes in the chemical environment of these intracellular compartments. This causes different structural populations of a protein in different intracellular compartments, further complicating structural studies of a protein in living cells. Currently, there is no structural biology technique that allows us to study a protein structure within a specific intracellular compartment. An alternative is to deliver the labeled protein into its destiny compartment for structural studies. The next challenge is to deliver the target protein into a specific intracellular compartment. In order for the structural information obtained to be physiologically relevant, the target protein should be located where it functions.

Finally, the biggest challenge involves the complexity of the intracellular environment. Since there can be up to 300 – 400 mg/mL of protein inside a living cell,

macromolecular crowding will have an effect on a protein's structure and function [113]. The protein interaction network must also be considered as protein-protein interactions may also have an effect on a protein's structure. Finally, there is the potential for different sub-populations of a protein in different cellular compartments within a cell. In the case of MESD, it traffics from the ER to the Golgi as it acts as a chaperone and escort protein for LRP5/6 [62]. The structure of MESD within the ER may differ from the structure within the Golgi.

The concept behind an in-cell structural technique also poses several technical challenges that we have sought to overcome. First, the minimum concentration needed for a protein to collect FRET spectra is in the low micromolar range. While this is much closer to being physiologically relevant as compared to the required millimolar concentrations needed for NMR studies, it is still probably higher than what would normally be found in a cell, depending on the protein. Another technical challenge is the ability to deliver an exogenous labeled protein to the correct intracellular compartment of a living cell. Cell-penetrating peptides and their counterparts do not have this targeting capability. The QQ-protein delivery technique developed by our lab can deliver exogenous proteins to the correct intracellular compartment and solves this technical challenge.

The final technical challenges involve overcoming the complexity of the intracellular environment. With regards to macromolecular crowding, it should not have an effect on the fluorophores' ability to act as a donor or acceptor within the target protein. If the macromolecular crowding has an effect on the protein's structure,

this will be detected by the FRET measurements and provide insight into the *in situ* structure of MESD within living cells. As to the possibility of protein-protein interactions affecting the FRET measurements due to intermolecular FRET, we propose to utilize 5-HT-labeling to solve this problem, since the 5-HT-labeled protein can be specifically excited at 310 nm whereas the background intracellular proteins with regular tryptophan remain un-excited [105]. This allows us to separate 5-HT-labeled proteins from the intracellular background proteins, eliminating any possible intermolecular FRET and providing a major advantage of our FRET-measurement under native intracellular conditions. Thus, this FRET-measurement is only from the intramolecular FRET while the protein of interest is actually interacting with its binding partners under native intracellular conditions. In regards to the potential of different sub-populations of the protein within a living cell, this can be overcome by utilizing several cell biology techniques such as knockout and knockdown techniques. For example, the siRNA technique can be used to knockdown interaction partners like LRP5/6 for MESD protein. In addition, adding an inhibitor to prevent traffic from the ER to the Golgi would ensure all the MESD would be located in the ER only.

Our proposed methodology is to use an auxotrophic bacteria strain to produce a 5-HT labeled protein, ensuring that only the target 5-HT-labeled protein will be excited within a mixture of background unlabeled intracellular proteins [105]. We will use MESD as a model protein for this methodology development. A thiol-reactive small molecule fluorophore was chosen since it will bond to the single cysteine residue of MESD and serve as the fluorescence acceptor. The next two steps in this methodology

development is to deliver this specifically labeled protein to the correct intracellular compartment of a living cell and determine the intracellular protein concentration. In the case of MESD, it needs to be delivered to the ER of the living cell. The QQ-protein delivery technique serves as the protein transduction method to achieve this goal [102, 103].

4.1.1 Cell-penetrating peptides

Protein transduction is a technique that delivers proteins into living cells. It emerged after the discovery of the cell penetrating peptides (CPP) [46, 47]. These CPPs are small peptides with the ability to enter cells *via* an unconventional way, although their transduction mechanism is still debatable [48, 49]. Fusion of a CPP with proteins/DNAs/RNAs allows their intracellular delivery [48-50]. Efforts have been made to pursue non-peptide protein delivery reagents. Polyethyleneimine (PEI) is found to have the ability to deliver protein and DNA intracellularly [8, 52]. A small-molecule mimic (SMoCs) of CPP has been reported to have a similar protein delivery property [53].

Despite these notable successes, protein delivery technology has yet to become commonplace for biomedical applications [46, 47]. The CPP-fused proteins share common problems. The CPP-fusion changes protein sequence and intracellular proteases likely degrade the delivered proteins, if they are not folded properly, before they reach their target intracellular compartment. The CPP-fusion also lacks targeting capability to specific intracellular compartments, significantly restricting their

applications. It remains unknown if the intracellular folding machinery can refold the CPP-delivered bacterial expressed proteins. Blobel's "signal theory" guides the fate of endogenous proteins, dictating their intracellular locations and trafficking [54, 55]. Questions remain regarding whether the CPP-delivered proteins follow the same intracellular trafficking/secretion pathway inside cells. These are critical questions regarding the physiological relevance of protein delivery technology.

4.1.2 QQ-protein delivery technique

Recently, our lab has developed a QQ-reagent based protein delivery technology that has solved the problems related to the CPP-based technology [102, 103]. The QQ-protein delivery has several novel features that make this technique advantageous to development of in-cell FRET methodology. First, the QQ-modification reagents non-covalently associate with proteins so there are no structural changes to the protein. The modification reagents also provide protection from intracellular protease degradation. The delivery system has targeting capability to specifically deliver proteins to the correct intracellular compartments. Indeed, QQ-protein delivery provides new tools in cell biology studies, allowing one to introduce specific labeled proteins inside the cells for high-resolution biophysical studies of these proteins at the molecular level.

Another advantage of the QQ-protein delivery is that it only requires an incubation step of the QQ-modified proteins with cells. Uptake of QQ-modified proteins is highly efficient and can be visualized by SDS-PAGE analysis, Western blots and live cell fluorescent imaging. Although the mechanism of the QQ-protein delivery is unknown,

this is a significant result since high delivery efficiency is critical for applications of a protein delivery technology.

The development of this critical technology is the necessary step needed to introduce a model protein labeled with a FRET donor and acceptor pair to obtain in-cell FRET measurements.

4.1.3 General description of this proposed in-cell FRET technique

Figure 4-1 displays the general design of this in-cell FRET technique. To solve the technical challenges associated with in-cell FRET technique, our rationale is to first label the proteins with 5-HT and small molecule fluorophores (SMFs) *in vitro*. We will then specifically deliver the labeled proteins into their target intracellular compartment for in-cell fluorescence FRET studies using the QQ-protein delivery technology.

We will pioneer this novel in-cell FRET technique to determine the FRET-distances within the labeled protein in the correct intracellular compartment of living cells. These measured FRET-distances provide atomic resolution structural information of the labeled protein inside the cells. Our approach is to use labeled 5-HT as the fluorescence donor and a small molecule fluorophore at a specific residue as the fluorescence acceptor for the FRET-experiments. This allows us to specifically excite the labeled protein (5-HT) at 310 nm, while the background cellular proteins (natural tryptophans) remain unexcited. This ensures the in-cell FRET measurement to be performed only between 5-HT (donor) and a site-specific SMF (acceptor) within the

labeled protein and eliminates the complications of protein interactions with the background cellular proteins.

We will first work out experimental conditions for in-cell FRET measurements using a protein sample that contains a single fluorescence donor/acceptor pair. We will explore a methodology to calculate the FRET distance based on these in-cell FRET measurements. Using cell biology techniques, such as knockout, knockdown and transgenic cells, we can manipulate gene expression of a specific protein that interacts with the labeled protein, enabling us to study the bound structure of the labeled protein to other proteins using this novel in-cell FRET-technique.

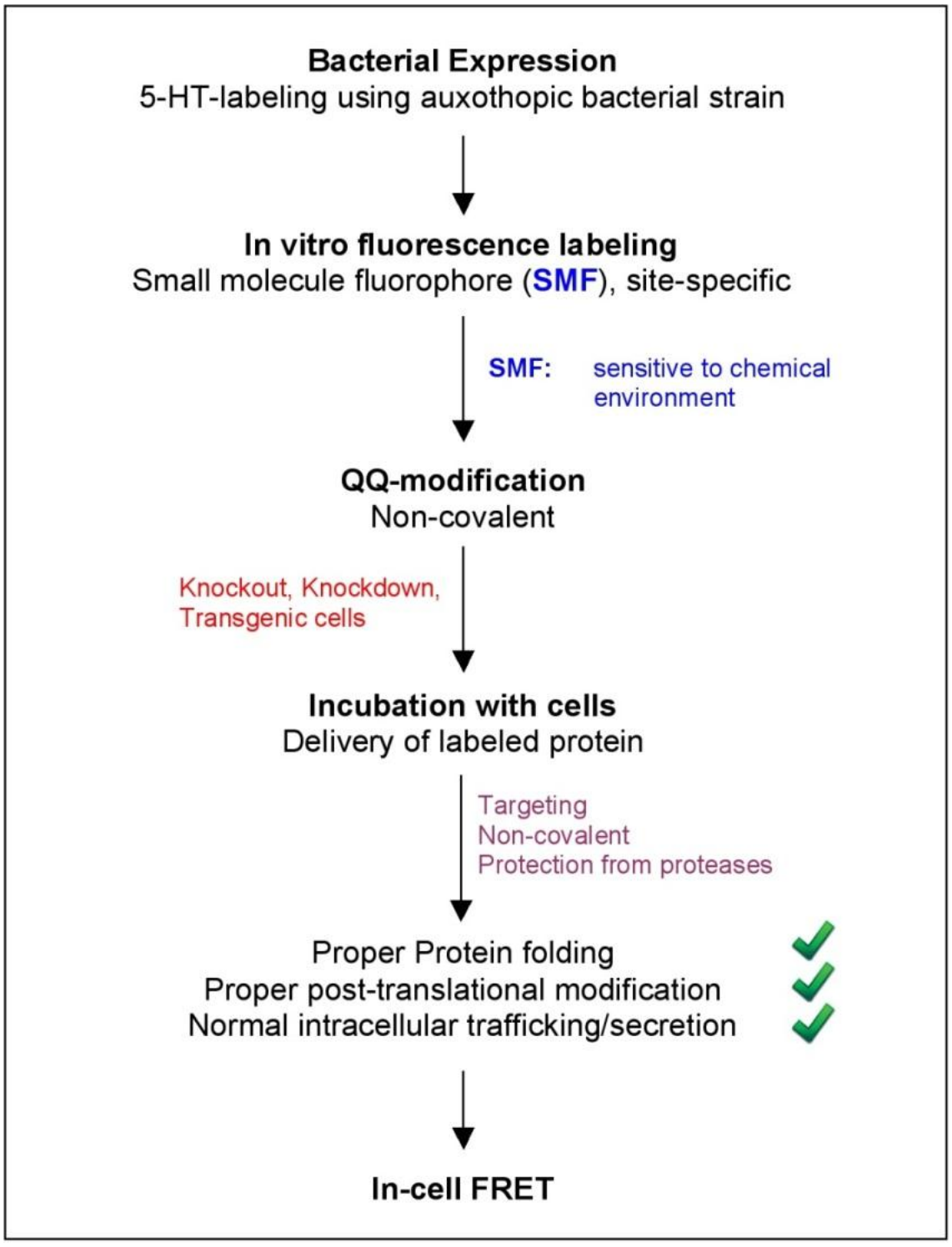


Figure 4-1: A schematic diagram of the experimental design of this in-cell FRET strategy.

We will then prepare specifically labeled protein samples that contain multiple 5-HT donors and a single acceptor for synchronized fluorescence experiments. Our results shown in Chapter 3 demonstrated that the overlapped FRET peaks of these multiple donor/single acceptor samples could be separated using this synchronized scanning fluorescence spectroscopy. Our data further indicated that the individual FRET peak could be unambiguously assigned to different donor/acceptor pair. This advance suggests a possibility for generating multiple FRET-distances from a single FRET-measurement. We will explore new FRET-methodology to achieve this goal. The success of this FRET-methodology will demonstrate a feasibility of developing FRET technique into an atomic resolution structural biology tool. In this case, we can simply make single cysteine mutants of the target protein and label them with a small molecule fluorophore as the acceptor and with multiple 5-HT as the donor. We will perform synchronized fluorescence experiments using these samples to calculate multiple FRET-distances with minimized efforts in sample preparation and FRET measurement. The generated FRET-distances will possibly allow us to generate protein structure at atomic resolution. Therefore, this approach may hold a great potential for real time atomic resolution structural determination of labeled protein at a near physiological concentration inside living cells.

4.1.4 Specific research goals of this chapter

While the overall long-term goal of this challenging project is the development of in-cell FRET methodology to calculate intramolecular distances of a protein located

within a living cell, the scope of this dissertation will focus on laying the foundation for this extremely challenging endeavor by optimizing several aspects concerning future in-cell FRET experiments. First, it must be proven that MESD can be specifically delivered into the ER of living cells with a high intracellular concentration enough for efficient FRET-measurements. Then various aspects of data collection must be optimized which includes cuvette size, determination of light scattering effects of cells, the buffer/medium used for cell suspensions, and which cell lines allow for the best detection of fluorescently labeled protein within a cell. Finally, intracellular protein concentration will be required for future in-cell FRET-based distance calculations. A method needs to be established for the determination of intracellular protein concentrations within a living cell.

The main focus of this chapter will be these optimizations of in-cell FRET experiments. We believe that these optimizations are essential for successful in-cell FRET experiments aiming at FRET-distance calculation. As the first pioneer of this challenging project, my central focus is to derive an optimized experimental condition for efficient FRET-experiments. We anticipate that the results obtained from these optimizations will lay the foundation for successful measurements using in-cell FRET in future continued studies.

4.2. Materials and Methods

4.2.1 Strain, plasmid and mutants

The MESD gene was subcloned into a pET30a vector from EMD Biosciences. We engineered the pET30a vector to introduce a Factor Xa site between the long his-tag and the MESD gene. The pET30a-sHT is also an engineered pET30a vector in which the long his-tag was replaced by a short his-tag containing a six histidine tag plus a two serine linker. The pET30a and pET30a-sHT are kanamycin resistant vectors. The expression vectors were transformed DL-41(DE3) auxotrophic bacterial strains.

MESD contains three tryptophans at residues 32, 98 and 130. Seven MESD mutants have been prepared as shown in Table 3-1, including three single tryptophan, three double tryptophan and one no tryptophan MESD mutants.

4.2.2 5-HT labeled MESD: Protein expression and purification

In order to produce MESD selectively labeled with 5-hydroxy-L-tryptophan (5-HT), glycerol stock (DNA transformed into DL41 (DE3) competent cells) was added to 500 mL M9 minimal medium supplemented with 25 mg/L of MET and TRP and allowed to grow overnight for at least 20 hours. After overnight growth, the OD_{600} reached ~ 1.0 before spinning down the culture at 5,000 x *g* for 7 minutes. The supernatant was removed and the container was inverted for 1 minute on a paper towel. The pellet was re-suspended with 25 mL 1X PBS and vortexed briefly in order to remove any remaining TRP. After spinning down and removing the supernatant, the pellet was re-suspended with 500 mL defined M9 minimal medium (not containing any glucose, ammonium

chloride or tryptophan, but supplemented with 50 mg/L of the other 19 amino acids). The re-suspended culture was poured into a 2 L flask and placed back in the 37 °C incubator shaker for 2 hours. After 2 hours, 25 mg 5-HT was added to the culture and placed back into the incubator shaker for an additional hour. At this point, 0.5 mM IPTG was added to the culture and grew for 5 hours at 37 ^{o}C. Cells were harvested by spinning the culture down at 10,000 x *g* for 10 minutes. After removing the supernatant, the pellet was either stored in the -80 °C freezer or used immediately for purification.

In order to purify the 5-HT labeled MESD proteins, the cell pellets collected after harvesting were resuspended in 20 mL 1X binding buffer (recipe modified slightly from His-Bind Resin manual and all buffers containing 6 M urea). The solution was sonicated 3X for 1 minute each time at 10 V. The lysate was spun down at 10,000 x g for 10 minutes. The supernatant was poured into a second container and stored on ice. The pellet went through two more rounds of sonication adding 10 mL of 1X binding buffer each time to resuspend the pellet. The clear lysate was loaded twice onto the previously charged and equilibrated His-Bind resin. The resin was then rinsed with 100 mL 1X binding buffer and 100 mL 1X wash buffer (25 mM imidazole). Finally, the protein was eluted with 60 mL 1X elution buffer (200 mM imidazole).

The elution was poured into a dialysis bag (MWCO 10,000 kD) and placed in 4 L of distilled water containing ~20 mM $NaHCO_3$. The solution stayed on dialysis for at least three days with three water changes per day. Once dialysis was complete, the solution was poured into a thick glassed beaker and place in a small container of liquid

nitrogen to freeze. Once frozen, the beaker was placed on a lyophilizer to obtain protein powder. The protein was weighed and a small sample was taken to check the purity of the protein powder.

4.2.3 Labeling MESD with IAEDANS

The protocol provided by Invitrogen was used to label MESD with IAEDANS, with minor modifications. First, MESD protein powder was dissolved in buffer (25 mM NaCl, 25 mM sodium-phosphate buffer, 10 mM EDTA, pH 7.6) at a concentration of ~ 80 µM (2 mg/mL). To disrupt the intermolecular disulfide bonds that have possibly formed, 20X molar excess of TCEP (1.6 mM) was added to the solution and was placed on a rocker at room temperature for 1 hour. Next, the solution was divided evenly into two microcentrifuge tubes and were labeled “unlabeled” and “IAEDANS”. To the IAEDANS tube, 40X molar excess of IAEDANS (3.2 mM) was added. Both tubes were wrapped completely in foil and placed on the rocker at room temperature for at least 2 hours. Finally, to both tubes, 20X molar excess of DTT (1.6 mM) was added to stop the reaction. The tubes were once again wrapped in foil and placed on the rocker at room temperature for 1 hour. In order to remove the free IAEDANS, the solution underwent size exclusion gel chromatography with a molecular weight cut off of 10 kD.

4.2.4 QQ-protein delivery

4.2.4.1 QQ-protein modification of MESD

First, three stock solutions were prepared that would be necessary for QQ-modification of MESD: Buffer A (8% DMSO, 1% glucose, 5 mM EDTA in 0.15 M sodium phosphate buffer, pH 8.0), Buffer B (8% DMSO, 1% glucose, 1 mg/mL 2k PEI in 0.15 M sodium phosphate buffer, pH 8.0) and Buffer C (0.5 M EDTA in 0.15 M sodium phosphate buffer, pH 8.0).

Next, 1 mg of MESD protein powder was dissolved in 500 µL Buffer A in a microcentrifuge tube and put on a rocker at room temperature for at least two hours. After 2 hours, 500 µL of Buffer B was added to the solution in the tube. The tube was then covered in foil and placed on a rocker in the cold room to modify overnight. The following morning, 1 mL of FreeStyle 293 Expression Medium (Invitrogen) was added to the modified protein solution and vortexed to mix completely.

4.2.4.2 QQ-protein delivery of MESD to mammalian cell lines

A variety of mammalian cell lines were used throughout the development of the QQ-protein delivery technology using MESD, including: GM01300 cells, HeLa cells, fibroblasts and ID8 cells. The protein delivery procedure was the same for each cell line with minor variances in cell line maintenance, such as serum percentage added to DMEM. Mammalian cells were used for QQ-protein delivery once they had reached ~ 80-90% confluency on a 75 cm^2 angled flask. The cells were prepared by removing the growth medium (DMEM + 8% FBS) and washing 3X with warmed 1X PBS. The modified

protein + FreeStyle medium solution was added to the cells and placed back in the 37 °C incubator for 2 hours. After the 2 hour loading period, the medium was removed from the flask and the cells were gently washed 3X with warmed 1X PBS solution.

To lift the cells from the flask, 500 µL of 0.25 % trypsin was added to the flask and tilted to ensure the trypsin coated the entire plate. Excess trypsin was removed before placing the flask back in the 37 °C for 2 minutes. After 2 minutes, the plate was removed from the incubator and the flask was lightly tapped on the surface the cells were attached to in order to assist lifting from the plate. To collect the cells from the flask, 3 mL of warmed 1X PBS was added to the flask to "wash" the plate and collect all the lifted cells. The lifted cells were removed via pipet and placed in a 15 mL conical tube. An additional 1 mL warmed 1X PBS was added to rinse the flask and remove any remaining cells. This rinse was also added to the 15 mL tube. The cells were immediately spun down at 2,000 x g for 3 minutes. The PBS solution was removed and the cells could now be utilized for fluorescence spectroscopy experiments.

4.2.4.3 Western blot of mammalian cell lysates

In order to perform a time course of MESD delivery into GM01300 cells, a small sample of cells were removed at one hour increments during the loading process for a total of four hours. Once the cells were removed from the flask, they were washed 3X with warmed 1X PBS with a gentle spin down of the cells following each wash (2,000 x g for 3 minutes). After removing the supernatant of the last wash, 100 µL of 2X SDS loading buffer was added to re-suspend the cells before putting the samples on a 90°C

heat block for 10 minutes. After heat shocking the samples, they were spun down at 10,000 x g for 10 minutes. The samples were loaded onto a 12% SDS poly-acrylamide gel by adding 10 µL into each lane. As a control, 0.1 mg MESD powder was dissolved in 100 µL SDS loading buffer and a 2 µL sample was added to the last lane.

After running the SDS-PAGE at 88V for 2 hours, the gel was soaked in Western transfer buffer for 20 minutes along with two filter papers and one nitrocellulose membrane cut to the same size as the gel. A semi-dry transfer was performed using 200 mA for 1.5 hours. After transfer, the membrane was soaked in 3% milk for twenty minutes on a rocker at room temperature. The primary antibody (anti-MESD, 1:3000 dilution) was added and incubated overnight in the cold room on a rocker. The following morning, the primary antibody was removed and the membrane was washed 3X with 1X PBS. After the final rinse, the secondary antibody (anti-mouse, 1:3000 dilution) was added and incubated for 1 hour on a rocker at room temperature. The secondary antibody was removed and the membrane was washed 3X with 1X PBS. Pierce ECL Western binding substrate was used to detect antibodies and expose to film.

4.2.4.4 De-glycosylation of MESD

After using the QQ-protein delivery into HeLa cells, the cells were washed 3X with warmed PBS buffer. The cell pellet was re-suspended in 1 mL PBS buffer and aliquoted equally in 4 separate microcentrifuge tubes. The cells were gently lysed by sonication. To the first tube, no enzyme was added as the control. To the second tube, 10 mU of NAase was added and gently mixed. To the third tube, 20 mU of NAase was

added and gently mixed. To the fourth tube, 40 mU of NAase was added and gently mixed. The tubes were allowed to incubate for 4 hours. After 4 hours, 10 µL of 4X SDS loading buffer was added to 30 µL of cell lysate from each tube. A Western blot was performed using these samples with an anti-MESD antibody.

4.2.4.5 Live cell fluorescence imaging

BSC-1 cells were grown to about 75% confluency and transfected with a GFP-ER marker according to the protocol provided by COMPANY. The next day, MESD was labeled with the amine-reactive fluorophore ArrayIt 640 (Arrayit Corporation, Sunnyvale, CA) according to the protocol provided by the company. After removing the free fluorophore, the fluorescently labeled MESD was QQ-modified as described in 4.2.4.1. Next, the labeled and modified protein was delivered to the BSC-1 cells expressing GFP-ER marker as described in 4.2.4.2 with a few modifications. First, the cells were only incubated with the protein for 20 minutes after which there was a 3 hour incubation period in the 37°C incubator.

After the 3 hour incubation period, the cells were imaged using a Zeiss Apo Tome microscope. Images were collected of cells using the light imaging channel, the FITC channel (detects green fluorescence) and the rhodamine channel (detects red fluorescence). The images were superimposed on one another to show the outline of the cell and the location of MESD within the cell and the ER. If the overlain images produce a yellow color, that proves MESD localizes in the ER of the cell.

4.2.5 In-cell fluorescence optimizations

4.2.5.1 Alternative fluorophores

The use of alternative fluorophores for detection of fluorescently labeled proteins within the living cells was explored. ArrayIt 640 (Arrayit Corporation, Sunnyvale, CA) as well as DyLight 488 and DyLight 649 (Pierce Biotechnology, Rockford, IL) are amine reactive and will bind to the amine group of lysine sidechains of MESD. Since MESD contains 24 lysine residues, the intensity will be much stronger and more easily detectable than an MESD protein labeled with a single IAEDANS fluorophore. MESD_wt was labeled with ArrayIt 640 and used for live cell fluorescence imaging. MESD_NoW was labeled with DyLight 488 and DyLight 649 (according to manufacturer's protocol) and absorbance and emission spectra were collected to determine the best excitation wavelength and emission range to be used for in-cell fluorescence spectroscopy optimizations.

For MESD_NoW labeled with DyLight 488 the following fluorescence spectra were collected: Using the Felix32 software package, a new data acquisition file is opened, the "Excitation Scan Method" is selected: excitation wavelength range is set at 450 – 525 nm, emission wavelength point is set at 530 nm, step size is set at 0.5 nm, integration time is set at 0.5 sec and average is set at 5. The buffer spectrum is collected for baseline correction. Next, a new data acquisition file is opened, the "Emission Scan Method" is selected: excitation wavelength is set at 507 nm, emission range is set at 510 – 640 nm, step size is set at 0.5 nm, integration time is set at 0.5 sec and average is set at 5. The buffer spectrum is collected for baseline correction.

For MESD_NoW labeled with DyLight 649 the following fluorescence spectra were collected: Using the Felix32 software package, a new data acquisition file is opened, the "Excitation Scan Method" is selected and the following information is inputted: excitation wavelength range is set at 500 – 700 nm, emission wavelength point is set at 673 nm, step size is set at 0.5 nm, integration time is set at 0.5 sec and average is set at 5. The buffer spectrum is collected for baseline correction. Next, a new data acquisition file is opened, the "Emission Scan Method" is selected and the following information is inputted: excitation wavelength is set at 654 nm, emission range is set at 600 – 750 nm, step size is set at 0.5 nm, integration time is set at 0.5 sec and average is set at 5. The buffer spectrum is collected for baseline correction.

4.2.5.2 Cuvette size

When performing *in vitro* FRET experiments, small sample volumes are placed in a 4-walled quartz cuvette with an inner chamber light path of 10 x 2 mm and a maximum volume of 500 µL (minimum volume = 250 µL). Since the light path is narrow, not as much of the sample will be excited in comparison to a larger volume cuvette with a light path that measures 10 x 10 mm. It is cost-effective to use the smaller volume cuvette, however for the purposes of methodology development of in-cell fluorescence experiments, the larger volume cuvette may provide us with a higher level of sensitivity.

First, MESD_NoW was labeled with DyLight 488 according to the manufacturer's protocol. The final sample volume was 3 mL with a protein concentration of ~ 80 µM. To collect fluorescence spectra, 300 µL of the sample was placed in the small cuvette

and placed into the QuantaMaster-6 Spectrometer (Photon Technology International, South Brunswick, NJ). The emission scan method data is collected using the Felix32 software provided by PTI with the following conditions: excitation wavelength is set at 507 nm, emission range is set at 517 – 600 nm, step size is set at 0.5 nm, integration time is set at 0.5 sec and average is set at 5. The buffer spectrum is collected for baseline correction.

The same sample is also placed into the large cuvette, with an initial volume of 3.0 mL. The same emission spectra were collected as the small cuvette. In addition, several other variable were also tested. First, the cuvette was rotated at 90° increments to confirm that position of the cuvette within the spectrophotometer yielded the same results. Next, a stir bar was added and the stirrer turned on to ensure this did not have any effect on spectra collection. Finally, a rubber stand was added to lift the position of the cuvette and allow for a minimum volume of protein sample. The absolute minimum volume necessary for data collection was determined by removing the sample in increments of 100 µL.

4.2.5.3 Time course of QQ-delivery of MESD into mammalian cells

During the QQ-delivery of DyLight 488-labeled MESD into HeLa cells, loading medium samples and cell samples were taken. The easiest way to perform a time course is to prepare multiple small dishes of the cells. HeLa cells are first grown in a 150 cm^2 angled flask to over 90% confluency. The cells are then lifted and transferred to 4

small petri dishes (35 x 10 mm) in equal aliquots. The dishes are placed back in the 37°C incubator for 24 hours to re-establish adherence to the dish.

The next day, the cells are washed 3 times with warmed 1X PBS. After preparing the cell loading medium containing the QQ-modified MESD (4.2.4.1), 1.5 mL is added to each dish. After 1 hour, one dish is removed from the incubator. A 300 µL sample of the medium is taken and the rest of the medium is removed. The cells are washed again 3 times with warmed 1X PBS. The cells are lifted and re-suspended in warmed PBS followed by immediate collection of emission spectra using FeliX32 software. The excitation wavelength is set at 507 nm, emission range is set at 510 – 600 nm, step size is set at 0.5 nm, integration time is set at 0.5 sec and average is set at 5. The PBS buffer spectrum is also collected for the baseline. The medium samples are collected in the same manner. This procedure is repeated at the 2, 3 and 4 hour time points. If necessary, longer time courses can be determined.

4.2.5.4 Buffer or medium for cell suspensions

MESD_NoW is labeled with IAEDANS first. Then, 50 µL of labeled MESD is added to 250 µL of the following solutions: 1X PBS, DMEM, DMEM + 2% FBS, DMEM + 5% FBS. A total of 8 emission spectra are collected: 4 spectra of the solutions without labeled protein and 4 spectra of the solutions with labeled protein. The excitation wavelength is set at 336 nm, emission range is set at 350 – 600 nm, step size is set at 0.5 nm, integration time is set at 0.5 sec and average is set at 5. The buffer used for IAEDANS-labeling is also used to collect a emission spectrum for baseline correction.

4.2.5.5 Determination of light scattering by mammalian cells

HeLa cells are grown to ~80% confluency in a 75 cm^2 angled flask. The medium is removed and the cells are washed 3 times with warmed 1X PBS. To lift the cells from the flask, 500 µL of 0.25 % trypsin was added to the flask and tilted to ensure the trypsin coated the entire plate. Excess trypsin was removed before placing the flask back in the 37 °C for 2 minutes. After 2 minutes, the plate was removed from the incubator and the flask was lightly tapped on the cell-attached surface in order to assist lifting from the plate. To collect the cells from the flask, 3 mL of warmed 1X PBS was added to the flask to "wash" the plate and collect all the lifted cells. The lifted cells were removed via pipet and placed in a 15 mL conical tube. An additional 1 mL warmed 1X PBS was added to rinse the flask to remove any remaining cells. This rinse was also added to the 15 mL tube. The cells were immediately spun down at 500 x g for 3 minutes at room temperature. The PBS solution was removed and then gently re-suspended with 3 mL warmed 1X PBS.

In a large cuvette, 990 µL of the cell suspension (which does not contain labeled MESD) was added followed by 10 µL of MESD_NoW labeled with DyLight 488 (80 µM). Using the FeliX32 software, "Emission Scan Method" is selected. The excitation wavelength is set at 507 nm, emission range is set at 510 – 600 nm, step size is set at 0.5 nm, integration time is set at 0.5 sec and average is set at 5. The PBS buffer spectrum is collected for baseline correction. After data collection, an additional 500 µL of the cell suspension is added to the sample and gently mixed. The same emission scan is

collected. This process is repeated 4 more times until the final volume of the cell suspension is 3.0 mL.

4.2.5.6 Cell line for detection of fluorescently labeled MESD within the cell

Several cell lines were tested to determine which allowed for the best detection of fluorescently labeled MESD within the cells. Three cell lines were used for collecting in-cell fluorescence spectra: fibroblasts, HeLa cells and ID8 cells. MESD_NoW was either labeled with IAEDANS or DyLight 488. After fluorescently labeling the protein it was QQ-modified and QQ-protein delivered into the cells. The samples were prepared for fluorescence spectroscopy by re-suspending the cell pellets in 1 mL of 1X PBS buffer. Using the FeliX32 software, the "Emission Scan Method" is selected and excitation wavelength is set at 336 nm for IAEDANS labeled MESD and 507 nm for DyLight 488 labeled MESD, emission range is set at 350 – 600 nm for IAEDANS labeled MESD and 510 – 600 nm for DyLight 488 labeled MESD, step size is set at 0.5 nm, integration time is set at 0.5 sec and average is set at 5. The PBS buffer spectrum is collected for baseline correction. Control cell samples of all cell lines are also created and undergo the same emission scan data collection.

4.2.6 Determination of intracellular protein concentration

The "no tryptophan" mutant was utilized to perform these experiments. The protein was labeled with IAEDANS. A BCA assay was performed to ensure that the protein concentration was around 80 μM (2 mg/mL). A series of emission scans were

collected by first placing 400 µL into a 4-walled quartz cuvette and setting the λ_{ex} = 336 nm and collecting an emission range of 350 – 600 nm. Five scans were collected and averaged. Next, 200 µL of the sample was removed and 200 µl of buffer was added to dilute the sample concentration by half. Again, five emission scans were collected and averaged. This series of dilutions continued until the concentration of MESD was 40 nM (1 µg/mL). After exporting the data to excel and removing the baseline, the peak of each emission was plotted against the concentration and a best-fit line was generated. To validate that the generated line was accurate, samples from previous experiments were taken and collected the same emission scan. Using the peak intensity from each sample, the concentration was calculated based on the generated equation from the best-fit line. Afterwards, a BCA assay was performed on each sample to compare the calculated concentration to the observed concentration.

4.3 Results

4.3.1 QQ-delivery of MESD inside living cells

To ensure high QQ-protein delivery efficiency for in-cell FRET experiments, verification of protein delivery had to be determined. First, cells were collected at 1 hour increments during the loading stage. A Western blot was performed on the cell lysates using an anti-MESD antibody. Figure 4-2 Panel A shows an example of a Western blot on GM01300 cell lysates during MESD cell loading, the fifth lane contains bacterial MESD as a control. In lanes 1-4 we see a strong lower band at the same level as the bacterial MESD control band, but we also see several upper bands as well. Cell lysates

from several mammalian cell lines that express endogenous MESD also show the same pattern of bands [102]. The varying intensities seen at different time points can be explained by the dynamic nature of MESD trafficking between the ER and Golgi. To prove that these upper bands are in fact glycosylated MESD bands, the cell lysates were treated with varying dosages of a deglycosylating enzyme, NAase. In Panel B, the results of this treatment are shown. In Lane 1, no enzyme was added and the three main bands remains visible after 4 hours. In Lane 2, 10 mU of NAase was added to the cell lysate and incubated for 4 hours. A noticeable decrease in the strength of the upper bands is observed in Lane 2 and is also true of Lane 3 in which 20 mU of NAase was added. In Lane 4, 40 mU of NAase was added and after the 4 hour incubation, no upper bands were observed indicating that MESD was fully deglycosylated.

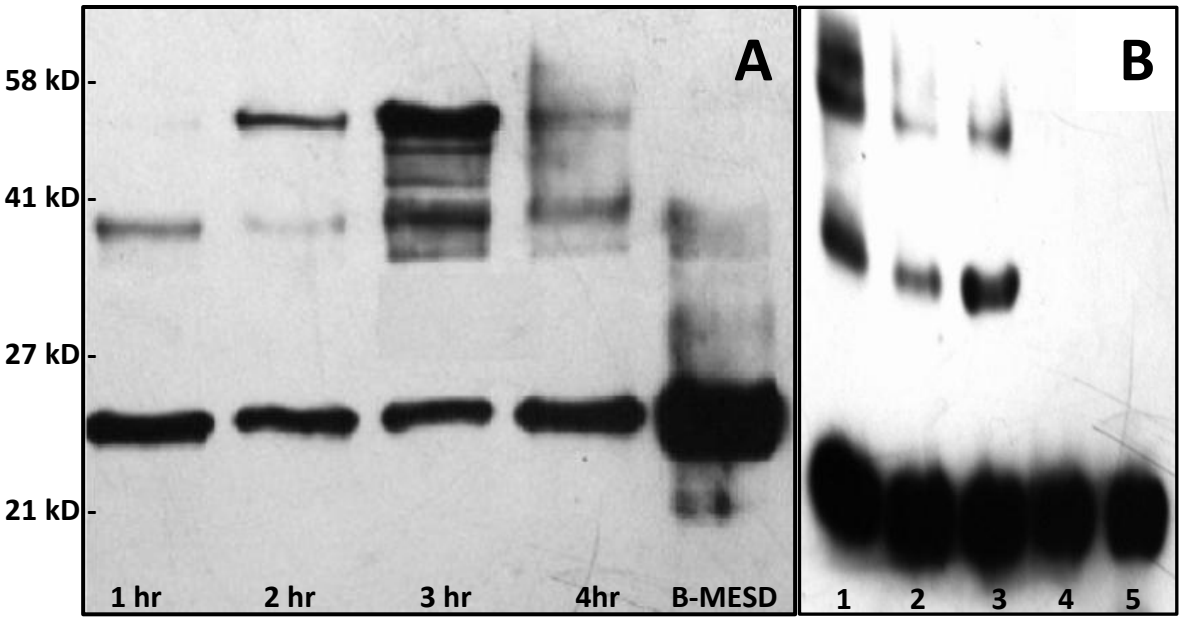


Figure 4-2: Panel A shows a Western blot of samples taken from GM01300 cell lysates during loading with QQ-modified MESD at different time points using an anti-MESD antibody. The fifth lane contains a sample of bacterial MESD as a control. Panel B shows a Western blot of HeLa cell lysate with and without different dosages of a deglycosylation enzyme, NAase. Lane 1 = no enzyme added to cell lysate. Lane 2 = 10 mU added. Lane 3 = 20 mU added. Lane 4 = 40 mU added. Lane 5 = bacterial MESD sample as a control.

In addition to performing Western blots and enzyme assays, live cell fluorescence imaging was performed to prove MESD localized in the ER. After labeling MESD with Array-It 640 and performing QQ-modification, the modified protein was incubated with BSC-1 cells for 2 hours. After removing the medium and washing the cells with warmed 1X PBS, the cells were subjected to fluorescent imaging using a Zeiss Apo Tome microscope.

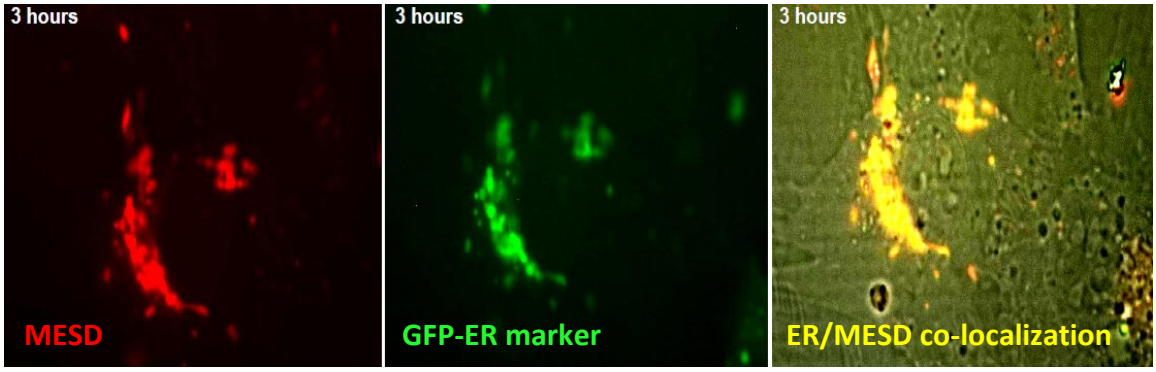


Figure 4-3: Live fluorescence images of BSC-1 cells after loading with MESD. Panel A shows the location of MESD labeled with Array-It 640. Panel B shows the location of the ER which has been labeled with GFP-ER marker. Panel C shows the merged images of the red channel, green channel and light image.

The live cell imaging results of the BSC-1 cells are shown in Figure 4-3. In Panel A, the rhodamine (red) channel was collected in order to visualize the location of labeled MESD within the cells. In Panel B, the FITC (green) channel was collected to visualize the location of the ER. Finally, in Panel C, the light image that was collected was merged with the red and green channel images. The yellow color now seen indicates that MESD is in fact located in the ER of the cells.

4.3.2 Optimizations of in-cell fluorescence spectroscopy

Since the focus of this thesis is methodology development, many factors had to be optimized. When developing a new in-cell fluorescence technique all aspects regarding the experiment must be considered which include the type of fluorophore

used for data collection, the type of cuvette a sample is placed in, the buffer/medium that the cells will be suspended in, in addition to optimizing the conditions that lead to successful delivery and detection of MESD within living cells.

4.3.2.1 Alternative fluorophores

Since MESD has 24 lysine residues as compared to a single cysteine residue, we decided to use amine-reactive small molecule fluorophores for the optimization of the experimental conditions for in-cell fluorescence spectroscopy. The ability to attach 20+ fluorophores to MESD will provide us with a brighter sample for easier detection of MESD within living cells. As we are still in the optimizing stage, using a more sensitive approach will guide us in developing a technique that will eventually be able to detect a protein labeled with a single fluorophore.

Figure 4-4 shows the absorbance and emission spectra of both DyLight 488 and DyLight 649. From this data, the best excitation wavelengths were found to be 507 nm for DyLight 488 and 654 nm for DyLight 649.

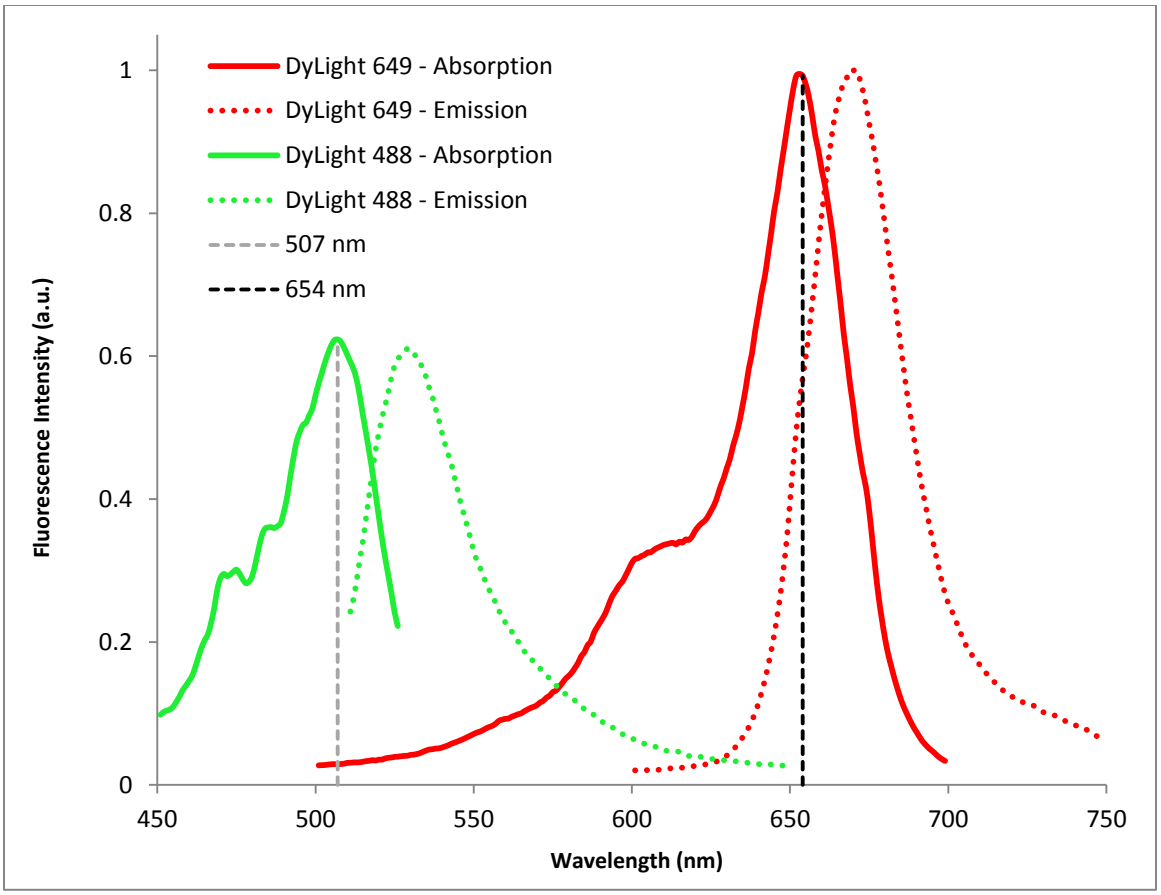


Figure 4-4: The absorption and emission spectra are displayed for both DyLight 488 (green) and DyLight 649 (red) by solid and dotted lines, respectively.

4.3.2.2 Cuvette Size

When performing *in vitro* FRET experiments, using small sample volumes is efficient and cost-effective. However, during the methodology development of an in-cell FRET experiment, the most sensitive approach must be taken. When exciting samples contained within a cuvette, the larger the light path, the more of the sample will be excited resulting in a more intense emission. To determine whether this increase in intensity was worth a more costly and labor intensive sample, a comparison between

samples placed in a small cuvette versus a large cuvette were performed. The small cuvette has a light path of 10 x 2 mm and a maximum volume of 500 µL. The large cuvette has a light path of 10 x 10 mm and a maximum volume of 4 mL. Emission spectra were collected of MESD_NoW labeled with DyLight 488 in both a large and small cuvette. As shown in Figure 4-5, Panel A, the emission spectra more than doubles when placed in the larger cuvette, indicating that the gain in intensity is worth the more costly and labor intensive sample preparation.

Once it was determined that the large cuvette would be used in future in-cell FRET experiments, other aspects of cuvette where analyzed. In Figure 4-5, Panel B, it is shown that no matter what position the cuvette is placed in the cuvette holder within the spectrophotometer, similar spectra will be collected. Next, it is also shown that by adding a stir bar and turning on the stirrer, spectra are unaffected. Finally, a rubber stand was added to the cuvette holder within in the spectrophotometer to raise the position of the cuvette allowing for a small minimum volume to be added. All of these variances did not alter the emission spectra collected. From these experiments it was determined that by adding a rubber stand to the cuvette holder, a 1 mL sample needed to be prepared and a stir bar could be used to keep the cells suspended in the buffer.

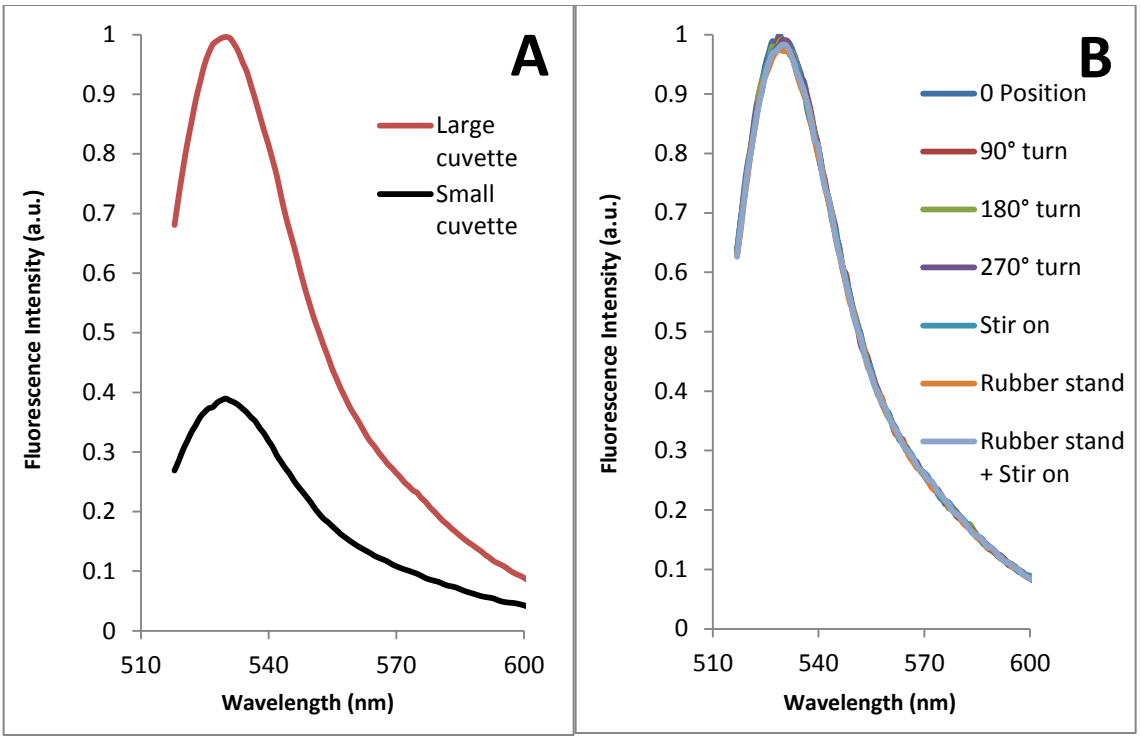


Figure 4-5: Panel A shows emission spectra of MESD_NoW labeled with DyLight 488 in a large cuvette at various volumes versus a small cuvette. Panel B shows emission spectra of MESD_NoW labeled with DyLight 488 in a large cuvette in various positions within the machine.

4.3.2.3 Time course for loading MESD

During the development of the QQ-protein delivery technique, it was clear that each protein delivery needed optimization of several variables. The concentration of protein used and ratio of modified protein solution to cell culture medium both needed to be optimized, as well as loading times. When optimizing these variables, both uptake of protein and cell viability are examined.

The MESD protein concentration ranged from 0.1 mg to 0.5 mg per mL of loading medium with no adverse effects on cell viability. However, the higher the concentration of MESD used, the more taken in by the cells resulting in brighter intensity. Figure 4-6 demonstrates the uptake of MESD by fibroblast cells. In Lane 1, QQ-modified MESD is shown. Lanes 2 – 4 represent various time points during cell loading. Lane 2 was a sample taken after adding the cell culture medium containing 5% FBS, right before adding it to the cell culture (time point zero). Lane 3 was a sample of the medium taken after 2 hours of loading. Lane 4 shows a sample taken after 4 hours of loading showing almost complete uptake of the protein by the cells. The strong upper band has been attributed to FBS that has been added to the cell culture medium. There are additional upper bands and lower bands seen as well. The lower bands do not appear to be degraded MESD as, as the sample shown in Lane 2 was taken immediately after mixing the QQ-modified protein with the cell culture medium.

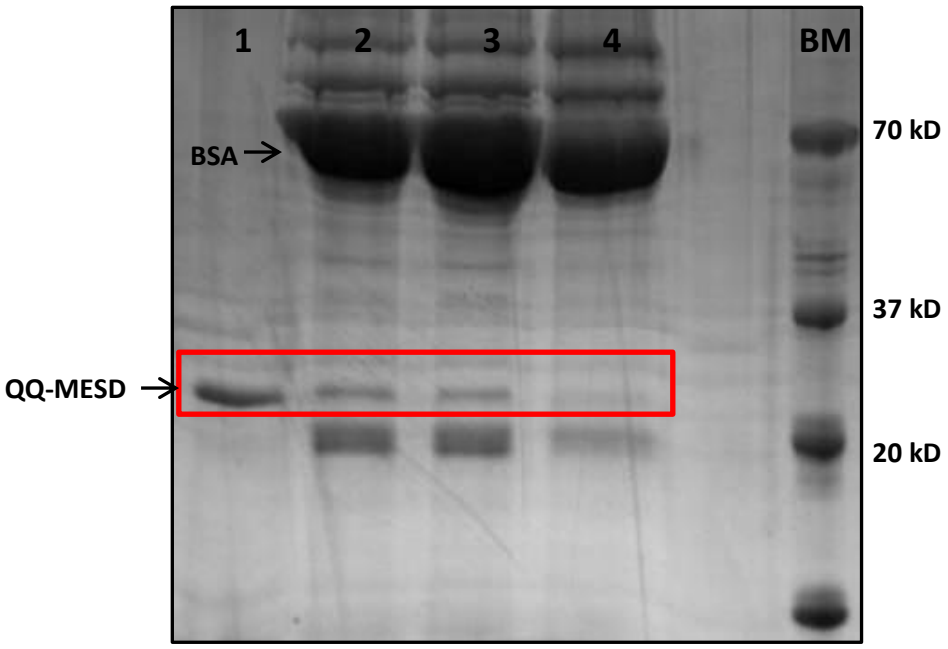


Figure 4-6: SDS-PAGE of MESD loading into fibroblast cells. Lane 1 = QQ-modified MESD before loading. Lane 2 = Sample taken from loading medium at the 0 hour point. Lane 3 = Sample taken from loading medium at the 2 hour point. Lane 4 = Sample taken from loading medium at the 4 hour point.

Since the SDS-PAGE analysis is not entirely conclusive, we also measured the fluorescence emissions of the medium and within the cell to determine the best amount of time needed for uptake of the protein. As shown in Figure 4-7, Panel A, the emission of labeled MESD continuously decreases with each hour of cell loading. Panel B displays the emission spectra collected of the washed cells suspended in PBS buffer before. Both panels indicate that the HeLa cells have taken up the maximum amount of MESD by the 6^{th} hour of loading.

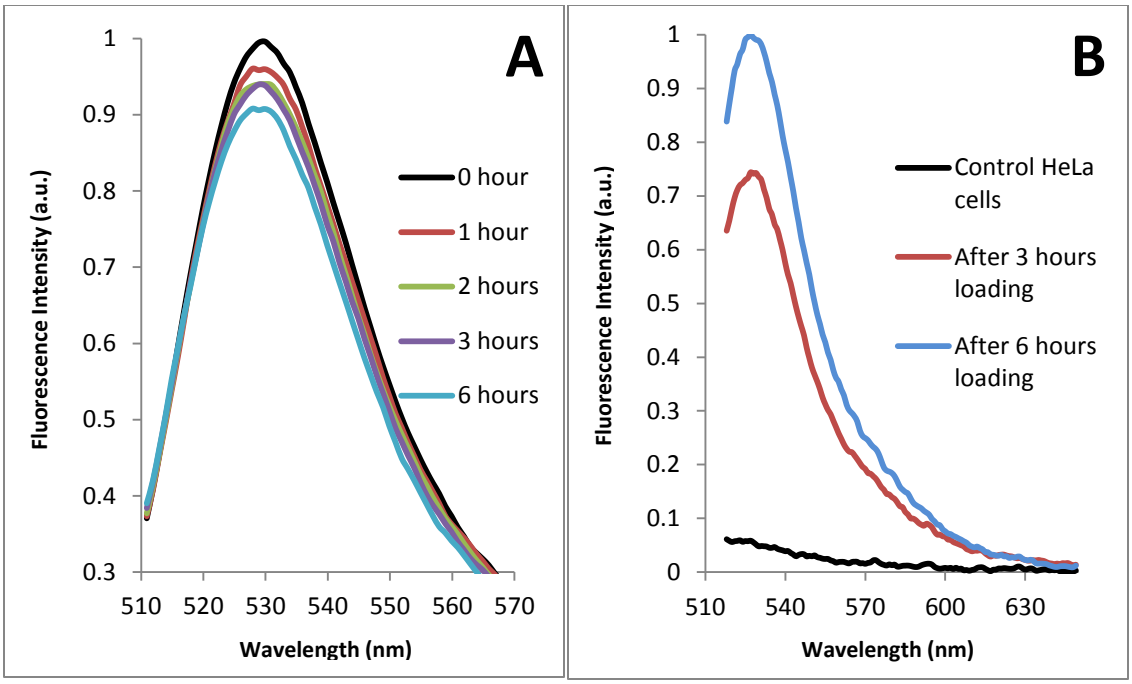


Figure 4-7: Panel A – Emission spectra collected of loading medium at various time points (0 to 6 hours) with λ_{ex} = 507 nm. Panel B – Emission spectra of HeLa cells before and after loading with MESD with λ_{ex} = 507 nm.

4.3.2.4 Medium used for in-cell fluorescence spectroscopy

Ideally, when performing in-cell FRET experiments, cell culture medium would be the best choice for the cell suspension solution in regards to maintaining healthy cells. However, the fluorescence properties of DMEM and fetal bovine serum (FBS) had to be determined. Since the final in-cell FRET experiments will be performed at λ_{ex} = 336 nm, various samples were prepared that contained MESD_NoW labeled with IAEDANS in various buffers and mediums. As shown in Figure 4-8, any sample containing DMEM displays a very intense peak that overlaps the IAEDANS emission peak.

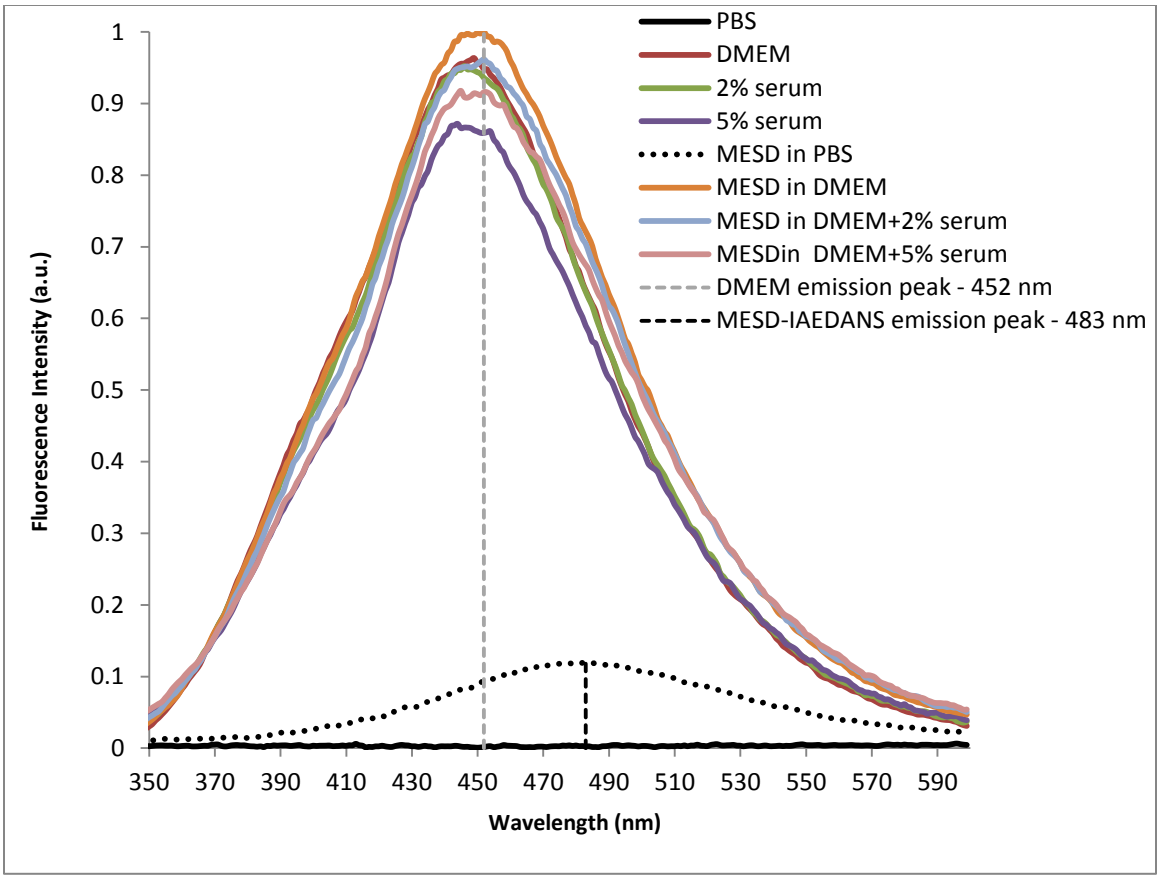


Figure 4-8: Emission spectra of 5µM MESD_NoW in various solutions: PBS, DMEM, DMEM + serum with λ_{ex} = 336 nm.

In Figure 4-9, the absorption and emission spectra of DMEM were collected. Although DMEM does not have identical excitation and emission peaks as those of IAEDANS, its intense emission peak masks the emission peak of IAEDANS due to the fact that the fluorescent components of DMEM are at a much higher concentration than that of MESD, as shown in Figure 4-8. Therefore, when performing in-cell FRET experiments, DMEM cannot be used and the cells must be suspended in a PBS solution.

Cell viability in PBS should not pose a problem since data can be collected in a manner of minutes.

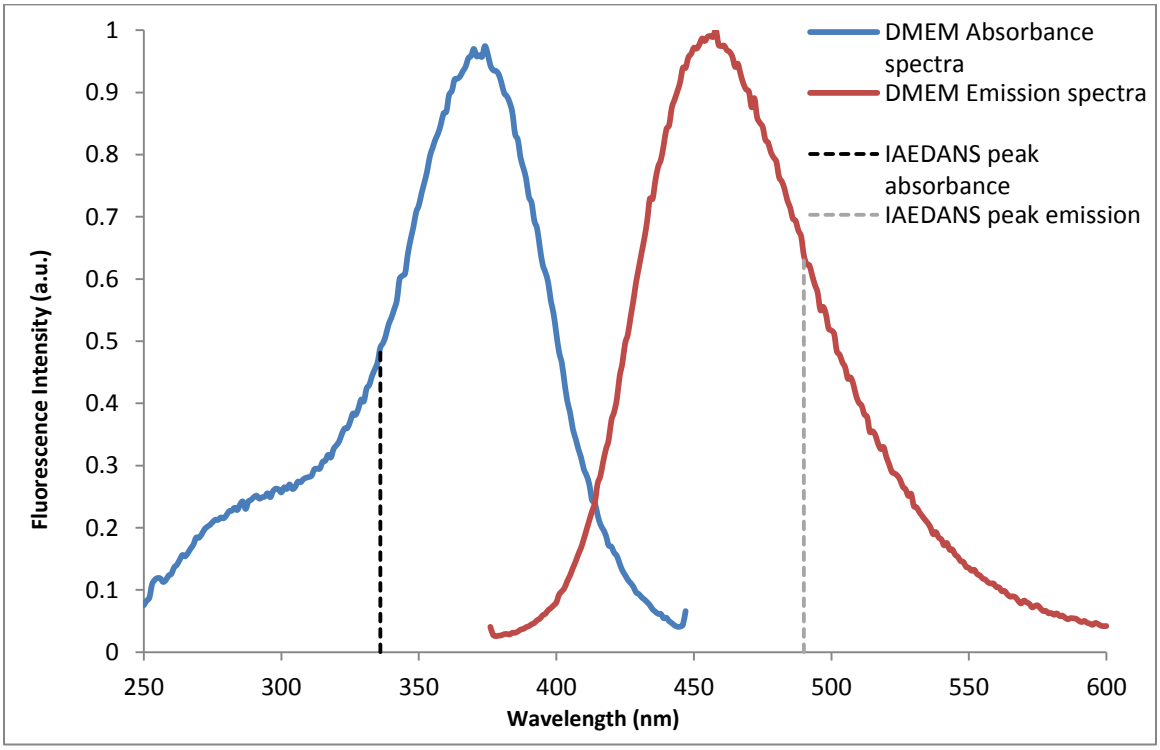


Figure 4-9: Absorbance and emission spectra of DMEM with a peak excitation wavelength of 370 nm and a peak emission wavelength of 450 nm.

4.3.2.5 Light scattering effects (Beer-Lambert Law)

Another factor to consider for in-cell FRET experiments are the effects of the cell turbidity of the solution on photon transmission through the sample. The Beer-Lambert law states that there is a logarithmic dependence between the transmission of light through a substance and the product of the absorption coefficient of the substance and the distance the light travels through the material (i.e. the path length of the cuvette).

In regards to FRET experiments, the photons that enter the cuvette will be more intense than the photons that exit the cuvette which are collected as the emission of the sample. The question now becomes, will the addition of mammalian cells to the sample effect the spectra by absorbing and scattering the photons? To answer this question, live HeLa cells were suspended in 3 mL of PBS buffer. In a large cuvette, 990 µL of the cell suspension was added followed by 10 µL of MESD_NoW labeled with DyLight 488. The emission spectrum was collected of this sample and is shown in Figure 4-10. Next, to dilute the sample, an additional 500 µL of the HeLa cell solution was added and gently mixed in the sample. Now, only the MESD concentration has been reduced while the turbidity of the sample remains the same. If the HeLa cells are absorbing light, there will be a non-linear reduction of the labeled MESD emission peaks. The sample was diluted in the same manner 4 more times and all the spectra can be seen in Figure 4-10.

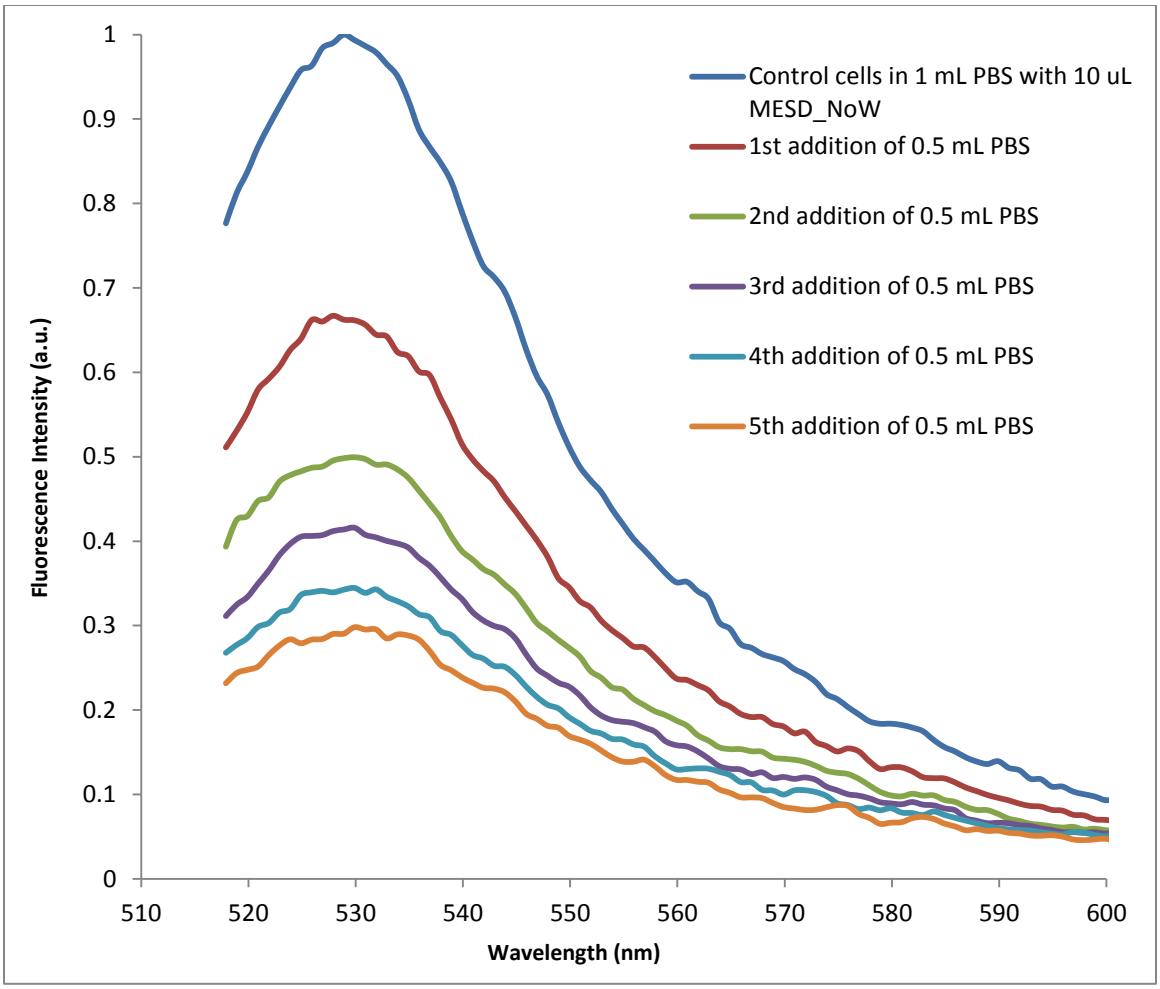


Figure 4-10: Emission spectra of a 1 mL HeLa cell suspension containing MESD_NoW labeled with DyLight 488 with an λ_{ex} = 507 nm. After the initial spectra were collected, an additional 0.5 mL of HeLa cell suspension was added to the solution until the final volume reached 3.5 mL.

After collecting all the emission spectra for each sample, the emission peak for each sample was plotted against its dilution factor. In Figure 4-11, each blue point represents an emission peak. A best fit line was added and line equation was

determined by Excel with a high R^2 value indicating a strong linear relationship amongst the peaks.

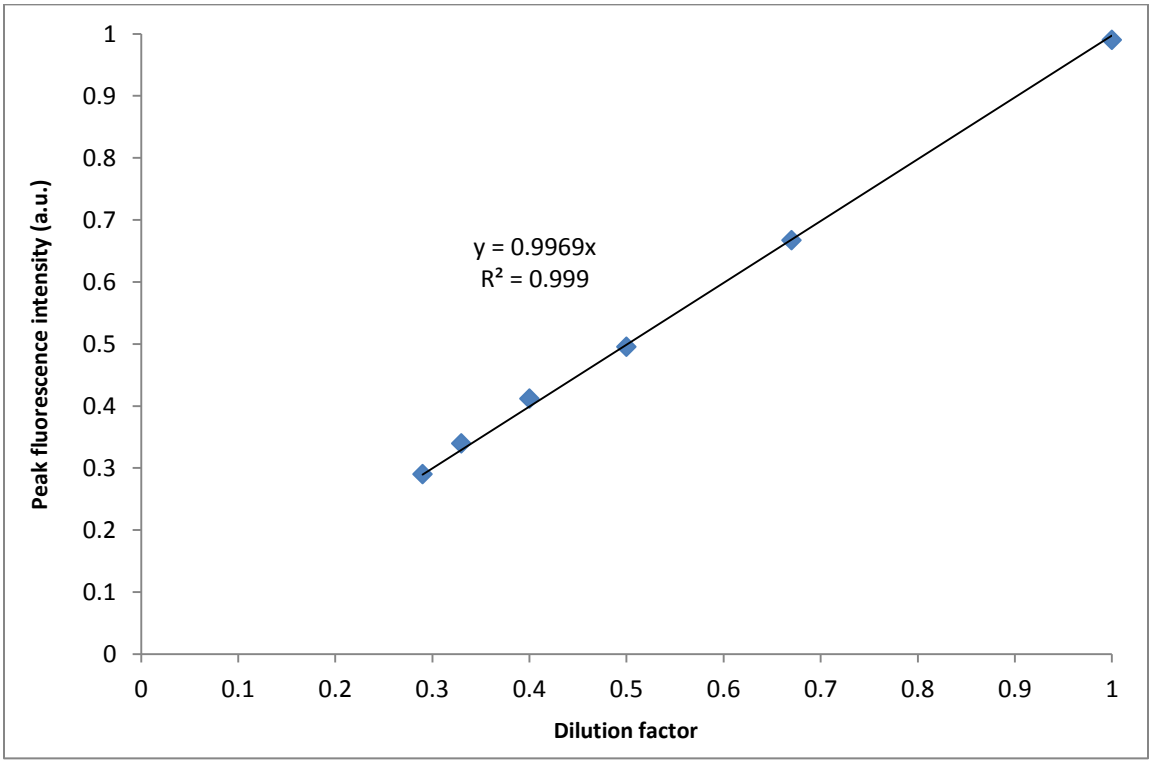


Figure 4-11: The peak intensity for each spectrum in Figure 4-9 was plotted against the dilution factor of that spectrum.

The dilution factor and peak fluorescence intensity are linearly correlated in a 1:1 ratio, which is what is expected [41]. Since a non-linear relationship is not seen, the cell suspension must not be absorbing or scattering the photons and therefore does not appear to effect fluorescence data collection.

4.3.2.6 Suitable cell line for in-cell fluorescence spectroscopy

Several cell lines were used to determine which allowed for best detection of MESD_NoW labeled either with DyLight 488 or IAEDANS. First, fibroblast cells were loaded with MESD_NoW labeled with IAEDANS. Figure 4 -12 shows the results of the emission spectra collected of control fibroblast cells and fibroblast cells containing MESD labeled with IAEDANS in Panel A. Panel B contains the difference spectrum which is obtained after subtracting the control spectrum from the MESD containing spectrum. Ideally, only the IAEDANS peak should be observed after removing the control spectrum. Unfortunately, the quality of the spectrum is poor as the intensity of the sample is quite low.

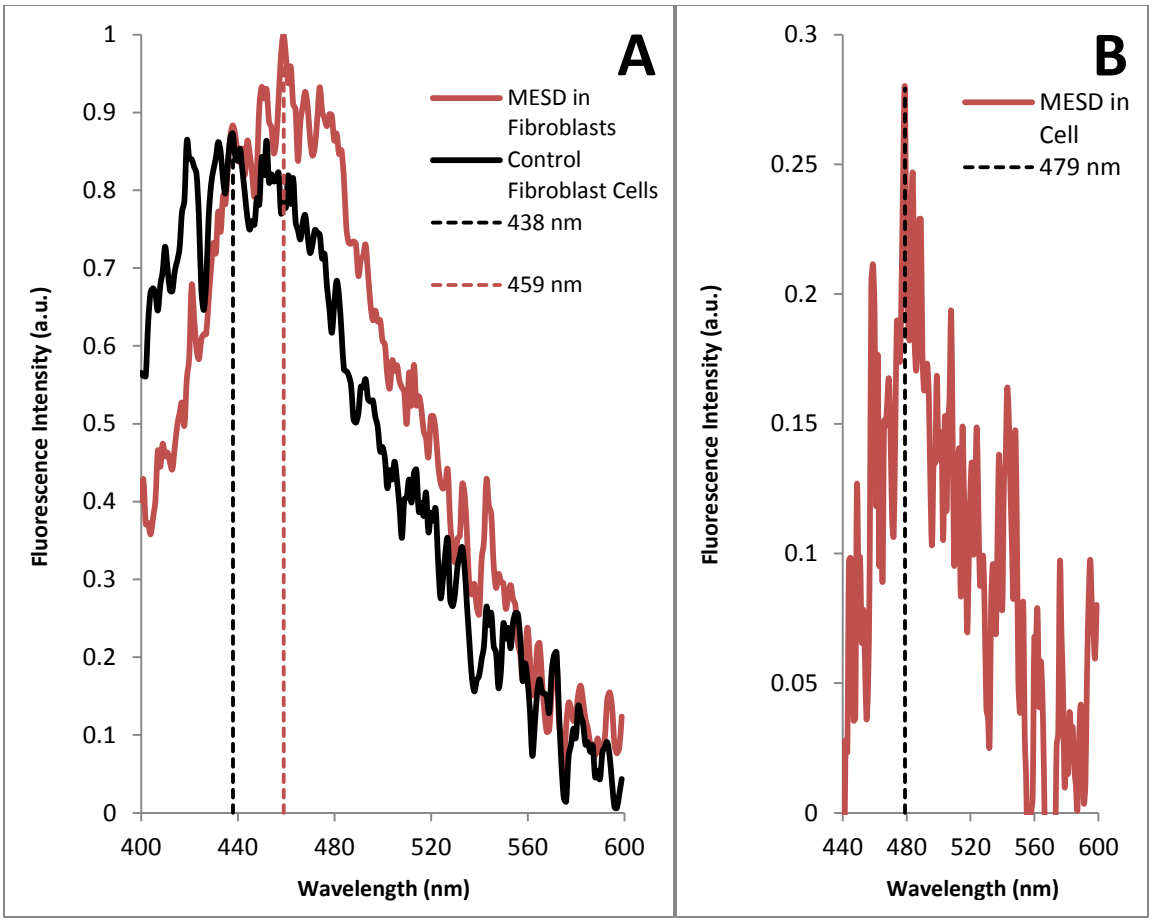


Figure 4-12: Panel A – emission spectra of control fibroblasts and fibroblasts containing MESD labeled with IAEDANS with λ_{ex} = 336 nm. Panel B – The difference spectrum which is obtained after subtracting the control spectrum from the MESD containing spectrum.

In Figures 4-13 and 4-14, HeLa cells are used when loading MESD_NoW labeled with IAEDANS (Figure 4-13) and then labeled with DyLight 488 (Figure 4-14). In Figure 4-13, Panel A, the distinction between the control spectrum and the MESD containing spectrum is much clearer as compared to fibroblast cells. In Panel B, the difference spectrum is shown after subtracting the control spectrum from the MESD containing

spectrum. A clearer peak is observed around 495 nm, which coincides with the emission peak of IAEDANS.

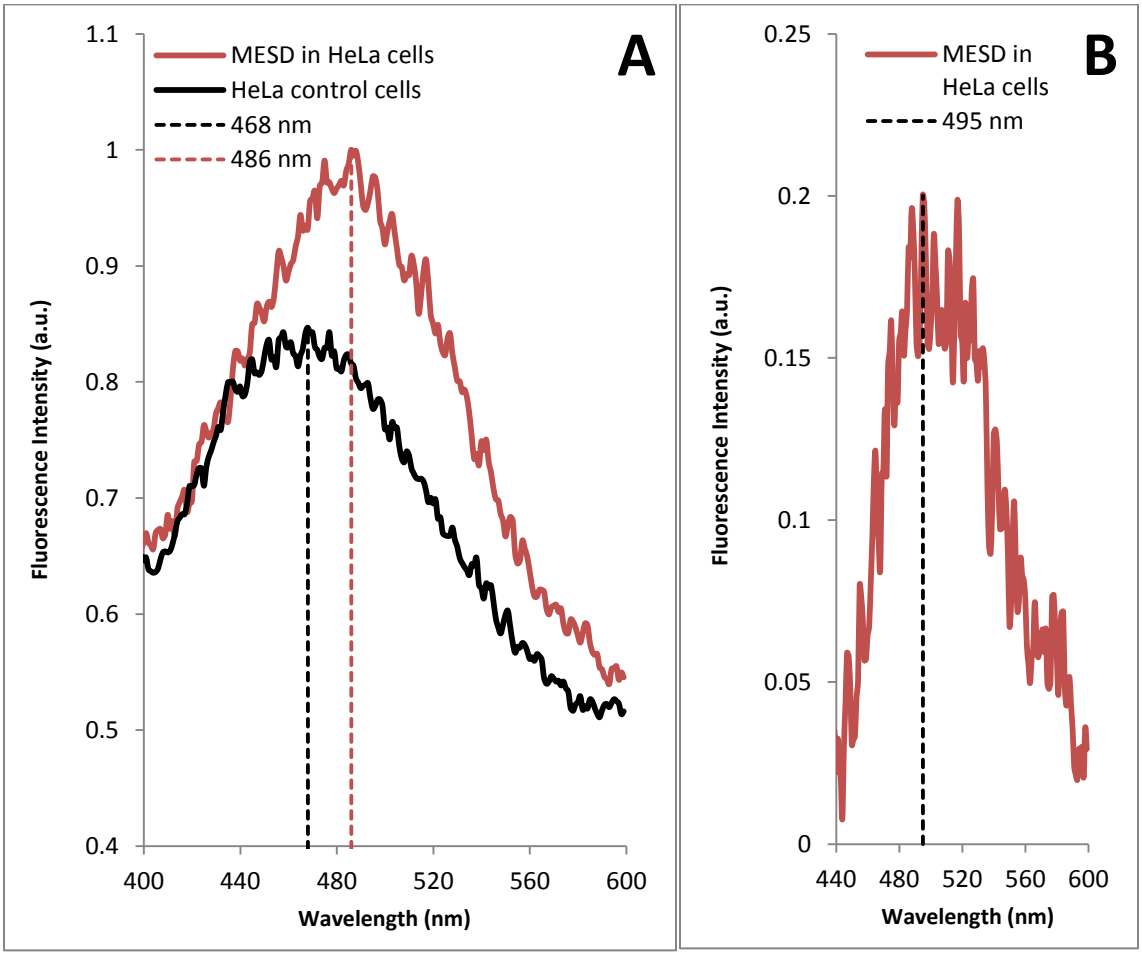


Figure 4-13: Panel A – emission spectra of control HeLa cells and HeLa cells containing MESD labeled with IAEDANS with λ_{ex} = 336 nm. Panel B – The difference spectrum which is obtained after subtracting the control spectrum from the MESD containing spectrum.

In Figure 4-14, HeLa cells were loaded with MESD labeled with DyLight 488 and the emission spectra are shown in Panel A. The distinction between the control spectrum and MESD containing spectrum are much clearer largely in part to the use of the DyLight 488 fluorophore instead of IAEDANS.

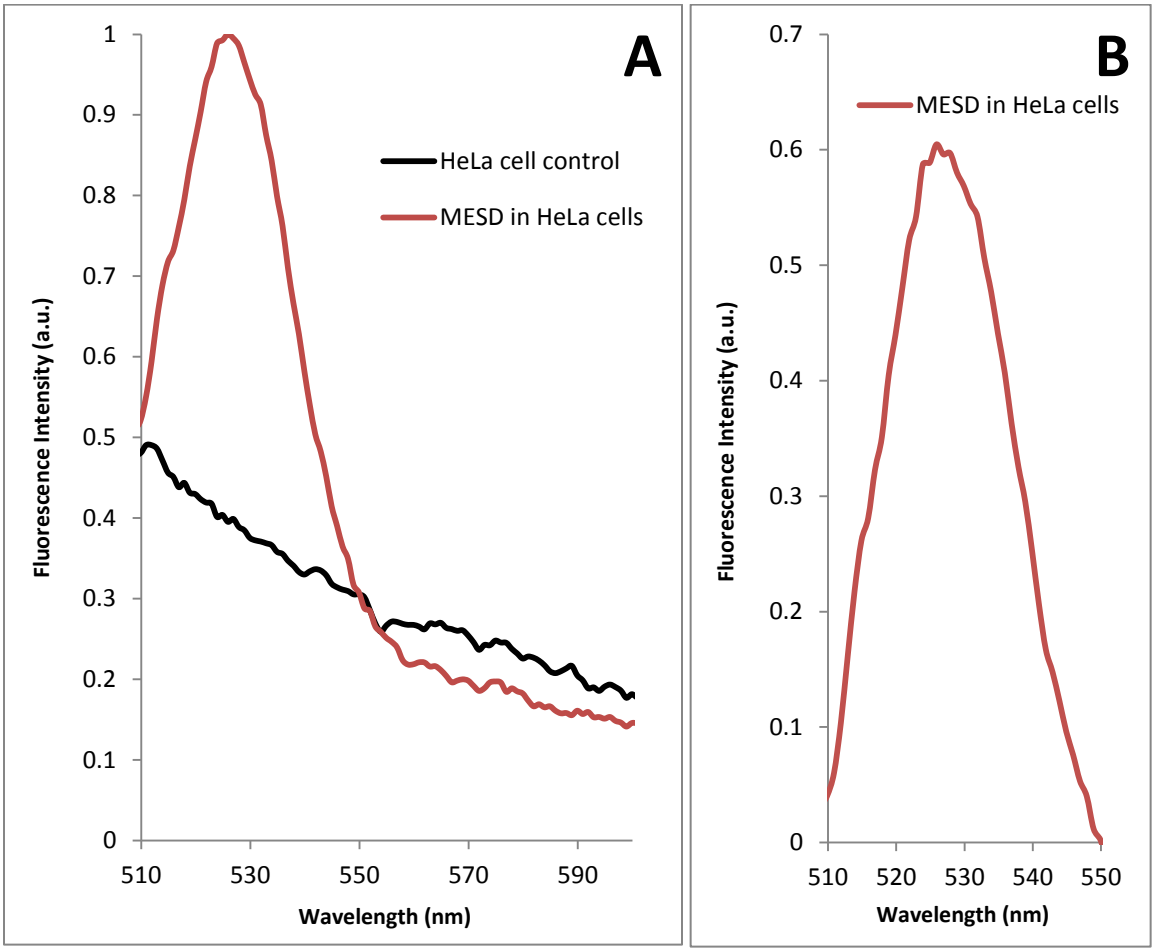


Figure 4-14: Panel A – emission spectra of control HeLa cells and HeLa cells containing MESD labeled with DyLight 488 with λ_{ex} = 507 nm. Panel B – The difference spectrum which is obtained after subtracting the control spectrum from the MESD containing spectrum.

The third cell line tested is ID8. These cells were loaded with MESD_NoW labeled with DyLight 488 and the emission spectra collected can be seen in Figure 4-15, Panel A. As with the fibroblast cells, there is not as clear distinction between the control cells and MESD containing cells as compared to the HeLa cells.

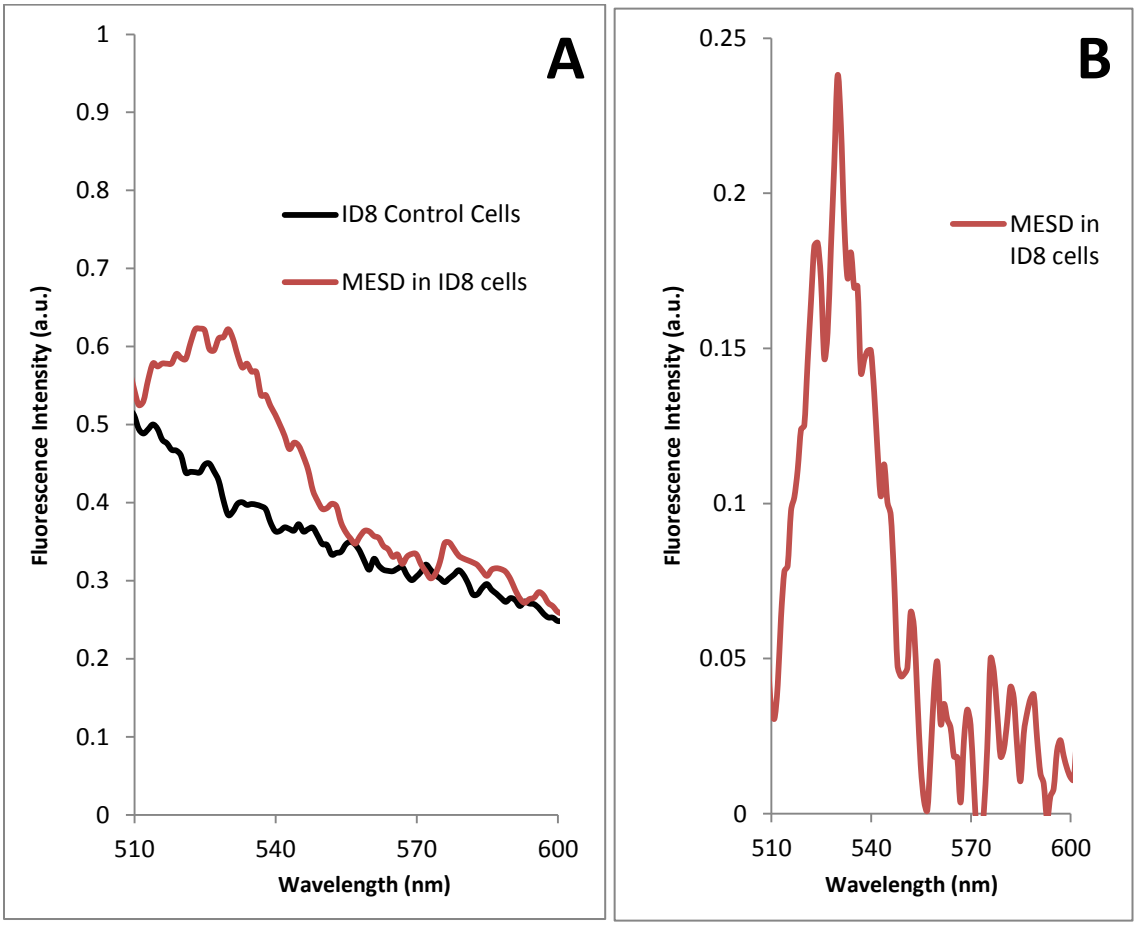


Figure 4-15: Panel A – emission spectra of control ID8 cells and ID8 cells containing MESD labeled with DyLight 488 with λ_{ex} = 507 nm. Panel B – The difference spectrum which is obtained after subtracting the control spectrum from the MESD containing spectrum.

In summary, HeLa cells seem to be the most suitable cell line for our in-cell fluorescence spectroscopic studies.

4.3.3 Intracellular protein concentration

It is crucial to know the protein concentration for each sample when for FRET-based distance calculations. When performing *in vitro* FRET experiments, protein concentration is simply determined by a quick experiment, e.g., Lowry assay, Bradford Assay or UV analysis. Currently, however, it is impossible to determine protein concentrations within living cells. Since fluorescence intensity is based on protein concentration, a standard curve should be able to be determined. MESD_NoW was labeled with IAEDANS and a set of standards were created (2000, 1000, 500… 1 μg/mL) using the BCA assay to confirm protein concentration. Emission spectra were collected for each sample and can be seen in Figure 4-16, Panel A. To calculate the standard curve line, the peak of each spectrum was plotted against the corresponding protein concentration (now in μM), shown in Figure 4-16, Panel B. A best fit line and equation were created by Excel with a strong R^2 value indicating a clear linear relationship between protein concentration and fluorescence intensity.

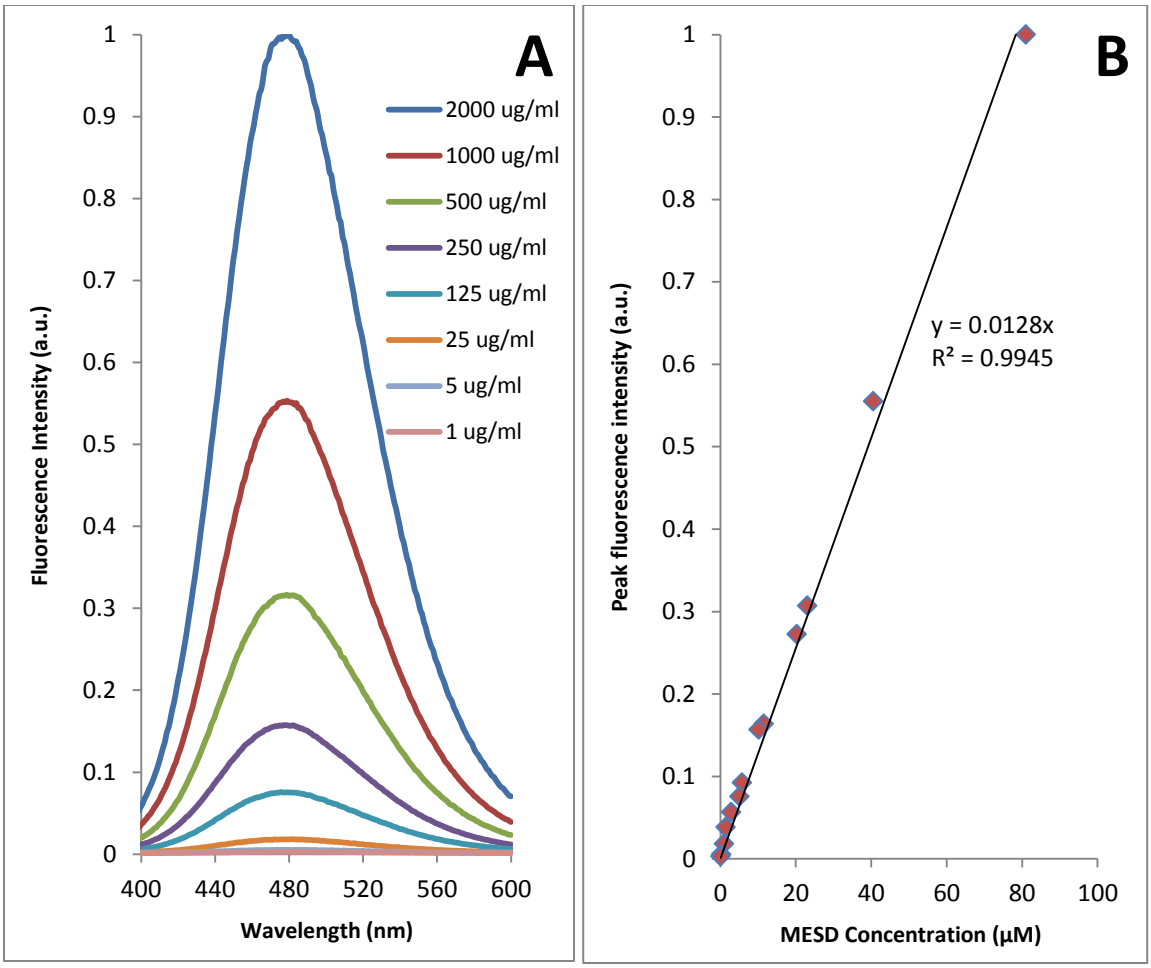


Figure 4-16: Panel A shows emission spectra of MESD_NoW labeled with IAEDANS at various concentrations with a λ_{ex} = 336 nm. In Panel B, each point represents the peak intensity of each sample and the corresponding protein concentration (µM). A best fit line (black line) and equation were determined by excel.

To confirm the accuracy of this new equation, another protein, MESD_W32/130A, was labeled with IAEDANS and emission spectra were collected at λ_{ex} = 336 nm. The peak intensity was entered into the standard curve equation and resulted in a concentration value. This value was confirmed by performing a Lowry

assay to determine the protein concentration of the sample. Table 4-1 summarizes the results of the two different methods. As can be seen below, using the standard curve to determine the protein concentration is a valid method.

Table 4-1: Protein concentration determination

Sample	Intensity @ 484 nm	Calculated (µg/mL)	Lowry (µg/mL)	% Difference
1	108100.5	**481.86**	474.5	1.6
2	52004.06	**232.69**	237.3	-1.9
3	128135.5	**570.85**	602	-5.2
4	67129.26	**299.88**	301	-0.4

These results suggest that the concentration of a protein can be estimated at a reasonable accuracy using this fluorescence method. This will provide the confidence for us to estimate intracellular protein concentration once we have optimized our in-cell fluorescence experiments.

In order to determine in-cell protein concentration a standard curve must be generated before collecting the in-cell FRET experiments. All variables must be identical, including: buffer, cuvette size, and emission scan input variables. This will ensure the most accurate data measurements.

4.4 Conclusions

In order to establish an in-cell methodology, many aspects of the experimental method must be first optimized. The rationale behind this methodology is to deliver a specifically labeled protein into the correct intracellular compartment of living mammalian cells. This protein must be shown to be taken up by the cells and delivered to the correct intracellular compartment. Using SDS-PAGE analysis and live cell fluorescence imaging, both of these aspects were confirmed. In the SDS-PAGE analysis, MESD was shown to diminish from the loading medium over a 4 hour incubation period. In the live cell fluorescence imaging, the ER was first labeled by transfection of a GFP-ER DNA marker (green fluorescence). The ArrayIt 640-labeled MESD (red fluorescence) was then QQ-delivered into the same cells; after superimposition of both the red and green images, yellow fluorescence was observed, indicating MESD is indeed located inside the ER of the living cells.

The next step was to optimize the various conditions that would lead to the optimization of in-cell FRET measurements. We examined cuvette size, the light scattering effects of cells, buffer/medium to be used for cell suspension, different cell lines used for delivery, as well optimal cell loading conditions of MESD into the mammalian cells. During the optimization process, it was decided to label MESD with an alternative fluorophore that would provide more sensitivity for these optimizations. Since MESD contains 24 lysine residues, amine-reactive fluorophores were chosen, including: ArrayIt 640, as well as DyLight 488 and DyLight 640.

The first optimization was performed using different sizes of cuvettes. Previous *in vitro* FRET experiments were performed in a small 4-walled quartz cuvette with a light path of 10 x 2 mm and a maximum volume of 500 µL. For each experiment, a minimum of 300 µL was required. With the smaller light path, a decrease in fluorescence intensity is expected as a smaller volume of the sample is being excited. Since in-cell FRET experiments will have a low concentration of protein within the cells, it might be necessary to use a larger cuvette to ensure maximum fluorescence intensity. Identical samples were placed in small and large cuvettes and emission spectra were collected. The emission collected from the larger cuvette was more than doubled in intensity which justified the use of larger cuvette for stronger fluorescence spectra. It was also found that by adding a rubber stand to the cuvette holder within the spectrophotometer, a minimum volume of 1 mL could be used for future in-cell FRET experiments. This amount of sample is only 3-fold more than what is needed for the small size cuvette, however, the fluorescence intensity remained the same as the 3 mL protein samples using the large cuvette.

The second aspect to consider is the light scattering effects of a cell suspension on FRET data collection. To determine if cells within a suspension absorb or scatter the photons used to excite the protein, a 3 mL cell suspension was used. In the first sample, 10 µL of fluorescently labeled MESD (80µM) was added to 1 mL of the cell suspension. An emission scan was collected of this initial suspension. Afterwards, increments of an additional 500 µL of cell suspension were repeatedly added which would keep the cell concentration constant, but lower the concentration of MESD. If the cells do not absorb

or scatter the photons, a linear relationship will be observed between protein concentration and fluorescence intensity. If the cells do absorb or scatter the photons, a non-linear relationship will be observed. Our results showed that the relationship was in fact linear and therefore the cells do not have a light scattering effect on the protein sample.

The next optimization was the buffer conditions used for the cell suspension for collection of in-cell FRET measurements. Ideally, to maintain the health of the cell, it would be best to suspend the cells in cell culture medium containing DMEM plus fetal bovine serum. However, our data indicated DMEM has spectral properties that overlap those of IAEDANS and therefore could not be used for future in-cell FRET experiments. The best choice is to re-suspend the cells in warmed PBS buffer and quickly perform the experiments, which is feasible since data collection only takes a few minutes.

The final optimizations regarded QQ-protein delivery of MESD and detection of fluorescently labeled MESD within various cell lines. The concentration of MESD in the loading medium varied from 0.1 mg/mL to 0.5 mg/mL. The higher concentration of 0.5 mg/mL did not have an adverse effect on cell viability and were shown to be more easily detected within the cell after collecting emission spectra of the cells. Three different cell lines were explored: HeLa, fibroblasts and ID8. HeLa cells that had MESD delivered within them displayed the most different spectra from the control cells and will allow for future in-cell FRET measurements. In the cases of fibroblasts and ID8, the control cell baseline has too much spectral overlap with the cells containing labeled MESD.

Another challenge presented for in-cell FRET measurements is the determination of intracellular protein concentrations within living cells. In order to calculated FRET-based distances, protein concentrations must be known. Currently, it is impossible to determine protein concentrations within a living cell. Since fluorescent intensity is linearly dependent upon protein concentration, a best-fit line could be determined by plotting peak intensity versus protein concentration. Using this best-fit line equation, several samples with unknown protein concentrations were used for emission scan spectra collection. The peak emission values were entered into the best-fit line equation and concentration values were determined. A Lowry assay was performed on all the samples and the concentration values obtained from the protein concentration assay were compared to those calculated from the emission spectra. The calculated values were confirmed by the observed assay values, validating this technique. This suggests that this method may be used to estimate the intracellular protein concentration of living cells.

In summary, we have optimized a number of conditions that are needed before collecting in-cell FRET measurements. First, a rubber stand must be added to the cuvette holder in the spectrophotometer which will allow for a minimum sample volume of 1 mL. A large cuvette with a light path of 10 x 10 mm should be used with a stir-bar to keep the cells suspended in the cuvette. Next, we determined that the cells do not have a light-scattering effect so FRET measurements will not be affected by the cell suspension. We also determined DMEM cannot be used for the cell suspension due to its fluorescent components, PBS must be used instead. HeLa cells were shown be the

most suitable choice for future in-cell FRET experiments as labeled protein within the cell is most clearly detected. Finally, we have proposed a method to determine intracellular protein concentration that will be critical for in-cell FRET-based distance measurements.

4.5 Discussion

The overall long-term goal of this thesis project is to develop an in-cell FRET technique as a structural biology tool to study protein structure inside living cells at high-resolution. This is a very challenging project that deserves a major research effort to overcome many technical challenges that will be encountered during development of this novel in-cell FRET technique. As the first person in the lab pioneering this project, the first step is to optimize several experimental conditions to make this in-cell FRET measurement possible. This is the main focus of this chapter. The optimizations of these experimental conditions will lay the foundation for future continued exploration.

First, it has to be shown that QQ-protein delivery can specifically deliver MESD into the ER of living mammalian cells. In addition, optimization of QQ-protein delivery has to be performed for best efficiency of protein delivery for MESD into the ER. Second, many variables may affect data collection and optimization of these variable have to be performed in order to obtain reasonable in-cell fluorescence spectra. These variables include cuvette size, the possibility of light scattering, the buffer or medium used in cell suspension and the cell line used for fluorescence spectroscopic data collection. Finally, the intracellular protein concentration of MESD within living cells must be determined

since it is important for FRET-distance calculation based on FRET-data. Currently, there is no technique available to achieve this, so a technique must be developed to enable intracellular protein concentration determination.

In order to prove that MESD was delivered inside the ER of living mammalian cells, several experiments were performed. First, MESD uptake was visualized using SDS-PAGE analysis using the samples of the loading medium before, during and after loading. it is clear that the MESD band intensity slowly decreases until it finally disappeared after 4 hours in the loading medium. Cell lysate samples were also examined during the loading process. Western blot analysis, using an anti-MESD antibody, showed distinct MESD bands at every point during the time course. Two higher molecular weight bands were also detected. By a comparison with the western blot of cell lysates of the cell lines that express endogenous MESD, it was suggested that these two higher molecular weight bands were glycosylated MESD. To verify this suggestion, the cell lysates were treated with various dosages of NAase, which is a de-glycosylating enzyme, for 4-hours at room temperature. As the dosage increased, the upper bands became weaker and eventually disappeared at 40 mU of NAase.

Glycosylation of MESD confirms that the protein has been delivered into the ER and Golgi since they are the intracellular compartments for glycosylation, suggesting that the QQ-delivered MESD follows an identical intracellular trafficking and post-translational modification as its endogenous counterpart. This provides confidence for us for the QQ-protein delivery technique used in this in-cell FRET approach. In order to visualize MESD location in the ER, the BSC-1 cells were first transfected with a GFP-ER

DNA marker for 24 hours and then incubated with QQ-modified MESD_wt labeled with ArrayIt 640 for 30-minutes. The BSC-1 cells were then subjected to live-cell fluorescence imaging. The red and green images were collected and merged, showing that MESD does in fact localize within the ER of the living cells.

Next, the various aspects that could affect in-cell fluorescence spectroscopy were examined. First, we looked at cuvette size for protein samples. *In vitro* FRET experiments can be collected in small cuvettes with short light paths and small minimum sample volumes. While it is more cost-effective to produce small sample volumes, the price is a decrease in fluorescence intensity. When performing in-cell fluorescence spectroscopy, it is more beneficial to use a large sample volume if there is a large enough increase in fluorescence intensity. We compared identical samples in small and large cuvettes and found more than doubled spectral intensity was observed using the large cuvette. By adding a rubber stand to the cuvette holder within the spectrophotometer, we were able to decrease the sample volume required for data collection to 1 mL using the large cuvette.

Next, we examined if light scattering of a cell suspension was a concern for in-cell fluorescence spectroscopy. We performed these experiments by continuously adding 500 µL increments of the same cell suspension to a sample containing a fixed amount of fluorescently labeled MESD. Since the peak intensity of the emission spectra and concentration of MESD remained in a linear relationship, this indicated the cells within the sample did not absorb and scatter the light.

Another aspect to consider for in-cell FRET measurements is the buffer or medium in the cell suspension used for in-cell FRET experiments. Ideally, cell culture medium should not display fluorescence signals, thus the collected fluorescence spectra are only from the labeled protein inside cells. In addition, cell viability is also critical for data collection. Emission spectra were collected using several cell culture medium and serum combinations. We found that DMEM displays significant fluorescence signals that overlap with those of IAEDANS. Therefore, PBS has to be used for in-cell FRET experiments to eliminate any spectral contribution from the cell culture medium. Since data can be collected in 10 minutes, cell viability in PBS should not be a concern. Another possible solution to ensure cell viability when using PBS during in-cell FRET experiments is to add glucose to the PBS solution. As long as glucose does not display any interfering fluorescent properties, this should benefit cell health.

The type of cell line used for in-cell fluorescence spectroscopy also affects data collection. Three different cell lines were tested: fibroblasts, HeLa cells and ID8. All the cell lines were incubated with fluorescently labeled MESD_NoW followed by collection of fluorescence emission spectra. Fibroblasts and ID8 control cell lines were not as distinctly different from their MESD containing counterparts as HeLa cells were. HeLa cells provided the least amount of spectral overlap and the clearest visualization of the IAEDANS and DyLight 488 emission peaks. This cell line is the better choice for in-cell FRET experiments.

Finally, to determine the intracellular concentration of a labeled protein in living cells, we proposed a solution. Since fluorescence intensity is linearly dependent upon

protein concentration, we collected emission spectra of many samples of known concentrations. The emission peaks of these spectra were plotted against the protein concentration and a best-fit line and equation were determined. This equation was validated by collecting emission spectra on fluorescently labeled protein samples of unknown protein concentrations. The calculated protein concentrations were confirmed by Lowry assays. This data indicates that the intracellular protein concentration could be possibly obtained using this method once the experimental conditions of in-cell fluorescence spectroscopy are optimized.

Based on these results, we propose that when performing in-cell FRET experiments, emission spectra from the protein standards of known concentration should be collected first that will be used to formulate a best-fit line equation. When collecting the in-cell fluorescence data, include one emission spectrum that is the result of direct excitation of the IAEDANS fluorophore. This emission peak can be used to determine the overall protein concentration of the sample. To determine a rough estimate of intracellular protein concentration, the cells can be counted using a cell counter and then dividing the concentration (μg/mL) of the sample by the number of cells (cells/mL) to estimate the intracellular concentration (μg/cell).

We were unable to calculate any in-cell FRET-based distance measurements during this thesis period. However, successful FRET-experiments require a full optimization of different parameters, which is the main focus of this chapter of my thesis. A wide range of parameters, including those in fluorescence instrumentation, protein delivery, cell lines, cell culture medium used in FRET-experiments and

intracellular protein concentration, have been optimized, providing a solid foundation for successful FRET-experiments. With the successful optimizations of all these various aspects outlined in this chapter, we believe that with a few more years of dedicated exploration into this methodology development, this technique could become a feasible atomic-resolution in-cell structural biology tool.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Conclusions

The primary objective of this thesis is to develop an in-cell structural biology technique that can be used to study protein structures inside living cells. In order to achieve this goal, many factors must be considered. First, after examining all the available atomic resolution structural biology tools (NMR, X-ray crystallography and cryo-EM), we concluded that the X-ray crystallography and Cryo-EM techniques were not suitable to study protein structure within living cells. The NMR technique, although possible for in-cell experiments, requires high intracellular concentration of labeled proteins which are much higher than physiological concentration of a protein inside the cells. Förster resonance energy transfer (FRET), on the other hand, has great potential to become a structural biology tool for studying protein structure inside living cells. FRET does not suffer from the same limitations as the other structural biology tools, such as crystallization/freezing of the sample, broadening spectral linewidths, and excessive intracellular concentration issues.

After deciding upon FRET as the chosen technique, a model protein with a known structure had to be chosen and possible donor and acceptor molecules were considered. MESD was chosen as the model protein due to our lab's expertise in working with this protein and recently determined NMR structure of MESD. MESD had attractive features when determining possible donors and acceptors. In order to keep

the protein as least perturbed as possible, tryptophan was chosen as the donor taking advantage of its intrinsic fluorescent properties. MESD contains three tryptophan residues and after two rounds of mutagenesis, single tryptophan mutants were generated. MESD also contains one cysteine residue which can be used to label with the acceptor. IAEDANS is a thiol-reactive probe that has an absorption spectra overlapping tryptophan's emission and will allow for the labeling of the protein with a single acceptor.

The final challenge to overcome when choosing the donor and acceptor was the ability to specifically excite the target protein within the living cells. The excitation of tryptophan in the target protein is not problematic *in vitro*, but once the target protein is inside living cells, all proteins containing tryptophan residues will be excited. To solve this problem, the target protein can be labeled with a tryptophan analogue (5-hydroxy-L-tryptophan) which has a broadened absorption spectrum and can be specifically excited at 310 nm. At this wavelength, regular tryptophan is not excited. Using this strategy, we could only excite the protein of interest inside the cell, while the background intracellular proteins remain un-excited for our in-cell FRET experiments.

The possibility of using FRET to measure multiple distances with a single measurement needed to be explored in order for this new technology to be valuable structural biology tool in the future. Proteins samples were generated that contained multiple donors and a single acceptor. Synchronous scanning fluorescence spectroscopy has the ability to identify multiple fluorescent components contained in a mixture by varying both the excitation and emission wavelengths simultaneously during data

collection. The tryptophan residues of MESD were shown to be in unique chemical environments which results in varying emissions spectra from each residue. After optimizing data collection procedures and careful analysis of the resulting spectra, FRET peak separation was shown to be a possibility. Assignment of peaks resulting from individual residues was elucidated. Further exploration of this technique may yield interesting results leading to the calculation of multiple distances based on a single FRET experiment.

It is possible to produce large quantities of MESD with bacteria and fluorescently label the protein *in vitro*. Once labeled, it can be modified with the QQ-reagents and delivered to the ER of living cells. The next step is to optimize the conditions necessary for successful in-cell FRET measurements. A number of factors were examined, including sample preparation, cuvette size, buffer conditions and cell line used for delivery. After careful optimization of these conditions, we could collect descend in-cell FRET spectra of MESD using HeLa cells. We believe that with the continued optimization of these conditions, a viable in-cell FRET technique can be developed into a robust structural biology tool in the near future.

5.2 Future directions

5.2.1 Proposed optimized experimental conditions for in-cell FRET experiments

Based on the results of this thesis, we can propose a methodology for in-cell FRET experiments. First, in regards to instrumentation, the QuantaMaster 6 Spectrophotometer (Photon Technology International, New Brunswick, NJ) can be used

with the addition of a rubber stand to the cuvette holder. This stand will allow the sample to be excited near the bottom of the quartz cuvette instead of the middle. A stir bar, set at a gentle speed, can be used to keep the cells suspended during data collection. The FeliX32 software provided by PTI contains necessary experiments for data collection including the “Emission Scan Method” and “Synchronous Scanning Method”.

To produce a selectively labeled protein with a single donor/acceptor pair, a protein must be produced in an auxotrophic bacteria strain and the culture medium will be supplemented with 5-hydroxy-L-tryptophan (donor). After obtaining the protein powder, it is then labeled with the small molecule fluorophore IAEDANS (acceptor). Finally, the labeled protein is QQ-modified and mixed with cell culture medium for QQ-delivery of the protein into mammalian cells.

We determined that HeLa cells loaded with fluorescently labeled protein resulted in the best spectral quality, as compared to fibroblasts and ID8 cells. Another advantage of the HeLa cell line is that it can be grown in suspension or adherent cultures. Since the in-cell FRET experiments will be performed in suspension, the results might be more physiologically relevant using a cell line that normally grows in suspension as well. After the HeLa cells have been loaded with labeled protein and properly washed, they should be suspended in 1 mL PBS buffer and have the fluorescence spectra collected immediately. If the protein used in the experiment has a single donor/acceptor pair, an emission scan can be collected in 10 minutes. A final

spectrum should be collected that directly excited IAEDANS which will be used for protein concentration determination.

After performing the in-cell FRET data collection, a set of protein standards with known protein concentration should have emission spectra collected as well. The emission peaks will be plotted against the corresponding protein concentration and a best-fit line equation will be generated. This equation can be used to determine the intracellular protein concentration of the samples used in the in-cell FRET experiments. The cells must be counted after the in-cell FRET experiment and this number can be used to calculate an intracellular concentration with units μg/mL or ng/mL.

5.2.2 Calculation of multiple FRET distances

I*n vitro* fluorescence spectroscopy of protein samples containing multiple donors and one acceptor will be used first to establish a methodology that allows for multiple distance calculations. After preparing protein samples that contain multiple donors and a single acceptor, the synchronous scanning fluorescence spectra will be optimized to ensure best separation of FRET peaks. The strategy that was outlined in chapter 3 can be used to assign an individual FRET peak to a specific donor/acceptor pair. This strategy involves comparisons of synchronous scanning spectra of a protein that has 0, 1, 2 and 3 donors, each with 1 acceptor. The next challenge to overcome is the calculation of the distance between a donor and acceptor based on changes in the FRET peak and not the donor emission peak. Theoretically, this should be possible since the distance between the donor and acceptor determines the intensity of the FRET peak.

Once this methodology has been established *in vitro*, the next step is to determine multiple distances of multiple donors to a single acceptor in a target protein within living cells. This step will lead us in the direction of making the in-cell FRET technique a high-resolution structural biology tool. If multiple distances can be calculated, then by changing the acceptor position via mutagenesis and creating a set of mutants that contain multiple donors and a single acceptor in varied position, this will allow us to obtain many distances between two residues with a minimum effort in sample preparation, possibly generating a protein structure at atomic resolution using computer simulation with these distance restraints.

5.2.3 Solving biological questions

The goal of any scientific endeavor in this field is to ultimately answer biological questions. By creating an in-cell structural biology tool, we will be able to answer a number of previously unanswerable questions. For instance, does the *in vivo* structure of a protein match the structure determine *in vitro*? Do different chemical environments of different intracellular compartments have an effect on protein structure?

To answer these questions, the in-cell FRET technique first needs to be established. Then, several cell biology techniques can be utilized to specifically address certain questions concerning protein structure within living cells. First, knockout/knockdown/transgenic techniques can be used to create special cell lines that remove interaction partners of the target protein. Using MESD as an example, siRNA

could be used to knockdown the expression of LRP5/6 which are the interacting partners. The structural information gained from this cell line can be compared to structural information of a cell line that is overexpressing LRP5/6.

Another cell biology technique that can be utilized is the use of small molecule inhibitors that will block transport from ER to the Golgi or vice versa. By using these inhibitors, we could examine structural information of MESD within the ER only and compare it to structural information of MESD within the Golgi only. MESD is known to act as a chaperone and escort with two domains. At some point, there is a structural change of MESD that switches its function from chaperone to escort. By confining the protein to a specific compartment, these changes might be elucidated.

Another use for the in-cell FRET technique could be to observe the effects of post-translational modification on a protein structure. A series of samples could be created that would first allow for the collection of structural information immediately following protein delivery. The following samples could be taken at increments of 30 minutes post-delivery. Structural changes might be seen throughout this time course indicating either post-translational modifications or trafficking through various intracellular compartments with varying chemical environments. In addition, special cell lines can be generated to knockdown or knockout specific glycosylation enzymes and these special cell lines can be used to study protein structure inside cells.

Another avenue of study using in-cell FRET is to study protein-protein interactions. Two proteins can be delivered simultaneously or consecutively using the QQ-protein delivery technique. One protein can be strategically labeled with a donor,

while the other protein is strategically labeled with an acceptor. After QQ-delivery into the cells, in-cell FRET can be performed on these proteins as they function in their correct intracellular compartment.

The overall significance of this work will be the development of a structural biology tool that can be used to determine atomic resolution protein structure in living cells under a physiologically relevant concentration. The ability to gather structural information from proteins within living cells may answer many currently unanswerable questions. Therefore, the success of this new in-cell structural biology technology may generate paradigm-shifting results for correctly understanding the intracellular events of proteins in living cells at a high spatiotemporal resolution.

REFERENCES

1. Sun, Y., et al., *FRET microscopy in 2010: the legacy of Theodor Forster on the 100th anniversary of his birth.* Chemphyschem : a European journal of chemical physics and physical chemistry, 2011. **12**(3): p. 462-74.
2. Dauter, Z., M. Jaskolski, and A. Wlodawer, *Impact of synchrotron radiation on macromolecular crystallography: a personal view.* Journal of synchrotron radiation, 2010. **17**(4): p. 433-44.
3. Jaskolski, M., *Personal remarks on the future of protein crystallography and structural biology.* Acta biochimica Polonica, 2010. **57**(3): p. 261-4.
4. Braun, P. and J. LaBaer, *High throughput protein production for functional proteomics.* Trends in biotechnology, 2003. **21**(9): p. 383-8.
5. Graslund, S., et al., *Protein production and purification.* Nature methods, 2008. **5**(2): p. 135-46.
6. Hauptman, H., *Phasing methods for protein crystallography.* Current opinion in structural biology, 1997. **7**(5): p. 672-80.
7. Adams, P.D., et al., *Recent developments in phasing and structure refinement for macromolecular crystallography.* Current opinion in structural biology, 2009. **19**(5): p. 566-72.
8. Futami, J., et al., *Intracellular delivery of proteins into mammalian living cells by polyethylenimine-cationization.* J Biosci Bioeng, 2005. **99**(2): p. 95-103.
9. Walden, H., *Selenium incorporation using recombinant techniques.* Acta crystallographica. Section D, Biological crystallography, 2010. **66**(Pt 4): p. 352-7.

10. Available from: HTTP://PROTEINCRYSTALLOGRAPHY.ORG.

11. Evans, J.N.S., *Biomolecular NMR spectroscopy*1995, Oxford ; New York: Oxford University Press. xvi, 444 p.

12. Cavanagh, J., *Protein NMR spectroscopy : principles and practice*1996, San Diego: Academic Press. xxiii, 587 p.

13. Wüthrich, K., *NMR of proteins and nucleic acids*. The George Fisher Baker non-resident lectureship in chemistry at Cornell University1986, New York: Wiley. xv, 292.

14. Cavanagh, J., *Protein NMR spectroscopy : principles and practice*. 2nd ed2007, Amsterdam ; Boston: Academic Press. xxv, 885 p.

15. Koning, R.I. and A.J. Koster, *Cryo-electron tomography in biology and medicine.* Annals of anatomy = Anatomischer Anzeiger : official organ of the Anatomische Gesellschaft, 2009. **191**(5): p. 427-45.

16. Frederik, P.M., et al., *Perspective and limitations of cryo-electron microscopy. From model systems to biological specimens.* Journal of microscopy, 1991. **161**(Pt 2): p. 253-62.

17. Giepmans, B.N., et al., *The fluorescent toolbox for assessing protein location and function.* Science, 2006. **312**(5771): p. 217-24.

18. Zhang, J., et al., *Creating new fluorescent probes for cell biology.* Nature reviews. Molecular cell biology, 2002. **3**(12): p. 906-18.

19. Michalet, X., et al., *Quantum dots for live cells, in vivo imaging, and diagnostics.* Science, 2005. **307**(5709): p. 538-44.

20. Walling, M.A., J.A. Novak, and J.R. Shepard, *Quantum dots for live cell and in vivo imaging.* International journal of molecular sciences, 2009. **10**(2): p. 441-91.

21. Tokumasu, F., et al., *Band 3 modifications in Plasmodium falciparum-infected AA and CC erythrocytes assayed by autocorrelation analysis using quantum dots.* Journal of cell science, 2005. **118**(Pt 5): p. 1091-8.

22. Dahan, M., et al., *Diffusion dynamics of glycine receptors revealed by single-quantum dot tracking.* Science, 2003. **302**(5644): p. 442-5.

23. Akerman, M.E., et al., *Nanocrystal targeting in vivo.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(20): p. 12617-21.

24. Dwarakanath, S., et al., *Quantum dot-antibody and aptamer conjugates shift fluorescence upon binding bacteria.* Biochemical and biophysical research communications, 2004. **325**(3): p. 739-43.

25. Howarth, M., et al., *Monovalent, reduced-size quantum dots for imaging receptors on living cells.* Nature methods, 2008. **5**(5): p. 397-9.

26. Lavis, L.D., *Histochemistry: live and in color.* The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society, 2011. **59**(2): p. 139-45.

27. Ramos-Vara, J.A., *Technical aspects of immunohistochemistry.* Veterinary pathology, 2005. **42**(4): p. 405-26.

28. Shimomura, O., F.H. Johnson, and Y. Saiga, *Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea.* Journal of cellular and comparative physiology, 1962. **59**: p. 223-39.

29. Chalfie, M., et al., *Green fluorescent protein as a marker for gene expression.* Science, 1994. **263**(5148): p. 802-5.

30. Inouye, S. and F.I. Tsuji, *Aequorea green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein.* FEBS letters, 1994. **341**(2-3): p. 277-80.

31. Prasher, D.C., et al., *Primary structure of the Aequorea victoria green-fluorescent protein.* Gene, 1992. **111**(2): p. 229-33.

32. Tsien, R.Y., *The green fluorescent protein.* Annual review of biochemistry, 1998. **67**: p. 509-44.

33. Ormo, M., et al., *Crystal structure of the Aequorea victoria green fluorescent protein.* Science, 1996. **273**(5280): p. 1392-5.

34. Yang, F., L.G. Moss, and G.N. Phillips, Jr., *The molecular structure of green fluorescent protein.* Nature biotechnology, 1996. **14**(10): p. 1246-51.

35. Serber, Z., et al., *In-cell NMR spectroscopy.* Methods in enzymology, 2005. **394**: p. 17-41.

36. Sakakibara, D., et al., *Protein structure determination in living cells by in-cell NMR spectroscopy.* Nature, 2009. **458**(7234): p. 102-5.

37. Fernandez-Suarez, M. and A.Y. Ting, *Fluorescent probes for super-resolution imaging in living cells.* Nature reviews. Molecular cell biology, 2008. **9**(12): p. 929-43.

38. Grecco, H.E. and P.J. Verveer, *FRET in cell biology: still shining in the age of super-resolution?* Chemphyschem : a European journal of chemical physics and physical chemistry, 2011. **12**(3): p. 484-90.

39. Forster, T., *Zwischenmolekulare Energiewanderung und Fluoreszenz.* Ann Physik, 1948. **437**: p. 55-75.

40. Clegg, R.M., *Fluorescence resonance energy transfer.* Current opinion in biotechnology, 1995. **6**(1): p. 103-10.

41. Jares-Erijman, E.A. and T.M. Jovin, *FRET imaging.* Nature biotechnology, 2003. **21**(11): p. 1387-95.

42. Vogel, S.S., C. Thaler, and S.V. Koushik, *Fanciful FRET.* Science's STKE : signal transduction knowledge environment, 2006. **2006**(331): p. re2.

43. Sikorska, E., et al., *Synchronous fluorescence spectroscopy of edible vegetable oils. Quantification of tocopherols.* Journal of agricultural and food chemistry, 2005. **53**(18): p. 6988-94.

44. Teixeira, A.P., et al., *Synchronous fluorescence spectroscopy as a novel tool to enable PAT applications in bioprocesses.* Biotechnology and bioengineering, 2011. **108**(8): p. 1852-61.

45. Vivian, J.T. and P.R. Callis, *Mechanisms of tryptophan fluorescence shifts in proteins.* Biophysical journal, 2001. **80**(5): p. 2093-109.

46. Wadia, J.S. and S.F. Dowdy, *Protein transduction technology.* Curr Opin Biotechnol, 2002. **13**(1): p. 52-6.

47. Trehin, R. and H.P. Merkle, *Chances and pitfalls of cell penetrating peptides for cellular drug delivery.* Eur J Pharm Biopharm, 2004. **58**(2): p. 209-23.

48. Deshayes, S., et al., *Cell-penetrating peptides: tools for intracellular delivery of therapeutics.* Cell Mol Life Sci, 2005. **62**(16): p. 1839-49.

49. Wadia, J.S. and S.F. Dowdy, *Transmembrane delivery of protein and peptide drugs by TAT-mediated transduction in the treatment of cancer.* Adv Drug Deliv Rev, 2005. **57**(4): p. 579-96.

50. Melikov, K. and L.V. Chernomordik, *Arginine-rich cell penetrating peptides: from endosomal uptake to nuclear delivery.* Cell Mol Life Sci, 2005. **62**(23): p. 2739-49.

51. Schwarze, S.R., et al., *In vivo protein transduction: delivery of a biologically active protein into the mouse.* Science, 1999. **285**(5433): p. 1569-72.

52. Huang, Y.C., et al., *Long-term in vivo gene expression via delivery of PEI-DNA condensates from porous polymer scaffolds.* Hum Gene Ther, 2005. **16**(5): p. 609-17.

53. Okuyama, M., et al., *Small-molecule mimics of an alpha-helix for efficient transport of proteins into cells.* Nat Methods, 2007. **4**(2): p. 153-9.

54. Blobel, G. and D.D. Sabatini, *Ribosome-membrane interactions in eukaryotic cells.* In: Biomembranes, L.A. Manson, ed., Plenum Publishing Corp., New York. , 1971. **Vol .2**: p. pps. 193-195.

55. Matlin, K.S., *When cell biology grew up.* Traffic, 2000. **1**(3): p. 291-2.

56. Huang, H. and X. He, *Wnt/beta-catenin signaling: new (and old) players and new insights.* Current opinion in cell biology, 2008. **20**(2): p. 119-25.

57. Herz, J. and P. Marschang, *Coaxing the LDL receptor family into the fold.* Cell, 2003. **112**(3): p. 289-92.

58. Stevens, F.J. and Y. Argon, *Protein folding in the ER.* Seminars in cell & developmental biology, 1999. **10**(5): p. 443-54.

59. Bu, G., *The roles of receptor-associated protein (RAP) as a molecular chaperone for members of the LDL receptor family.* International review of cytology, 2001. **209**: p. 79-116.

60. Culi, J. and R.S. Mann, *Boca, an endoplasmic reticulum protein required for wingless signaling and trafficking of LDL receptor family members in Drosophila.* Cell, 2003. **112**(3): p. 343-54.

61. Hsieh, J.C., et al., *Mesd encodes an LRP5/6 chaperone essential for specification of mouse embryonic polarity.* Cell, 2003. **112**(3): p. 355-67.

62. Chen, J.L., et al., *Two Structural and Functional Domains of MESD Required for Proper Folding and Trafficking of LRP5/6.* Structure, 2011. **19**(3): p. 313-323.

63. Baneyx, F., *Recombinant protein expression in Escherichia coli.* Current opinion in biotechnology, 1999. **10**(5): p. 411-21.

64. Jana, S. and J.K. Deb, *Strategies for efficient production of heterologous proteins in Escherichia coli.* Applied microbiology and biotechnology, 2005. **67**(3): p. 289-98.

65. Peti, W. and R. Page, *Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost.* Protein expression and purification, 2007. **51**(1): p. 1-10.

66. Terpe, K., *Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems.* Applied microbiology and biotechnology, 2006. **72**(2): p. 211-22.

67. Brondyk, W.H., *Selecting an appropriate method for expressing a recombinant protein.* Methods in enzymology, 2009. **463**: p. 131-47.

68. Schmidt, F.R., *Recombinant expression systems in the pharmaceutical industry.* Applied microbiology and biotechnology, 2004. **65**(4): p. 363-72.

69. *Baculovirus Manual*, Invitrogen, Editor.

70. Kigawa, T., et al., *Preparation of Escherichia coli cell extract for highly productive cell-free protein expression.* Journal of structural and functional genomics, 2004. **5**(1-2): p. 63-8.

71. Swartz, J.R., *Advances in Escherichia coli production of therapeutic proteins.* Current opinion in biotechnology, 2001. **12**(2): p. 195-201.

72. Rosenberg, A.H., et al., *Vectors for selective expression of cloned DNAs by T7 RNA polymerase.* Gene, 1987. **56**(1): p. 125-35.

73. Studier, F.W. and B.A. Moffatt, *Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes.* Journal of molecular biology, 1986. **189**(1): p. 113-30.

74. Studier, F.W., et al., *Use of T7 RNA polymerase to direct expression of cloned genes.* Methods in enzymology, 1990. **185**: p. 60-89.

75. Grodberg, J. and J.J. Dunn, *ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification.* Journal of bacteriology, 1988. **170**(3): p. 1245-53.

76. Studier, F.W., *Protein production by auto-induction in high density shaking cultures.* Protein expression and purification, 2005. **41**(1): p. 207-34.

77. Tyler, R.C., et al., *Auto-induction medium for the production of [U-15N]- and [U-13C, U-15N]-labeled proteins for NMR screening and structure determination.* Protein expression and purification, 2005. **40**(2): p. 268-78.

78. Sreenath, H.K., et al., *Protocols for production of selenomethionine-labeled proteins in 2-L polyethylene terephthalate bottles using auto-induction medium.* Protein expression and purification, 2005. **40**(2): p. 256-67.

79. Sivashanmugam, A., et al., *Practical protocols for production of very high yields of recombinant proteins using Escherichia coli.* Protein science : a publication of the Protein Society, 2009. **18**(5): p. 936-48.

80. Hendrickson, W.A., J.R. Horton, and D.M. LeMaster, *Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure.* The EMBO journal, 1990. **9**(5): p. 1665-72.

81. Ramesh, V., et al., *The interactions of Escherichia coli trp repressor with tryptophan and with an operator oligonucleotide. NMR studies using selectively 15N-labelled protein.* European journal of biochemistry / FEBS, 1994. **225**(2): p. 601-8.

82. Tanio, M., et al., *Amino acid-selective isotope labeling of proteins for nuclear magnetic resonance study: proteins secreted by Brevibacillus choshinensis.* Analytical biochemistry, 2009. **386**(2): p. 156-60.

83. Krishnarjuna, B., et al., *Amino acid selective unlabeling for sequence specific resonance assignments in proteins.* Journal of biomolecular NMR, 2011. **49**(1): p. 39-51.

84. Waugh, D.S., *Genetic tools for selective labeling of proteins with alpha-15N-amino acids.* Journal of biomolecular NMR, 1996. **8**(2): p. 184-92.

85. Cooper, G., *The Cell: A Molecular Approach*. 2nd ed2000, Sunderland: Sinauer Associates.

86. Van Duyne, G.D., et al., *Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin.* Journal of molecular biology, 1993. **229**(1): p. 105-24.

87. Miroux, B. and J.E. Walker, *Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels.* Journal of molecular biology, 1996. **260**(3): p. 289-98.

88. Fisher, C.A., et al., *Bacterial overexpression, isotope enrichment, and NMR analysis of the N-terminal domain of human apolipoprotein E.* Biochemistry and cell biology = Biochimie et biologie cellulaire, 1997. **75**(1): p. 45-53.

89. Zhao, W., et al., *An efficient on-column expressed protein ligation strategy: application to segmental triple labeling of human apolipoprotein E3.* Protein science : a publication of the Protein Society, 2008. **17**(4): p. 736-47.

90. Zhang, Y., et al., *A monomeric, biologically active, full-length human apolipoprotein E.* Biochemistry, 2007. **46**(37): p. 10722-32.

91. Ren, X., et al., *Engineering mouse apolipoprotein A-I into a monomeric, active protein useful for structural determination.* Biochemistry, 2005. **44**(45): p. 14907-19.

92. Ren, X., et al., *A complete backbone spectral assignment of human apolipoprotein AI on a 38 kDa prebetaHDL (Lp1-AI) particle.* Biomolecular NMR assignments, 2007. **1**(1): p. 69-71.

93. Cai, M., et al., *An efficient and cost-effective isotope labeling protocol for proteins expressed in Escherichia coli.* Journal of biomolecular NMR, 1998. **11**(1): p. 97-102.

94. Marley, J., M. Lu, and C. Bracken, *A method for efficient isotopic labeling of recombinant proteins.* Journal of biomolecular NMR, 2001. **20**(1): p. 71-5.

95. Chen, H.C., C.F. Hwang, and D.G. Mou, *High-density Escherichia coli cultivation process for hyperexpression of recombinant porcine growth hormone.* Enzyme and microbial technology, 1992. **14**(4): p. 321-6.

96. Yang, Y., D. Hoyt, and J. Wang, *A complete NMR spectral assignment of the lipid-free mouse apolipoprotein A-I (apoAI) C-terminal truncation mutant, apoAI(1-216).* Biomolecular NMR assignments, 2007. **1**(1): p. 109-11.

97. Zhang, Y., J. Chen, and J. Wang, *A complete backbone spectral assignment of lipid-free human apolipoprotein E (apoE).* Biomolecular NMR assignments, 2008. **2**(2): p. 207-10.

98. Sivashanmugam, A. and J. Wang, *A unified scheme for initiation and conformational adaptation of human apolipoprotein E N-terminal domain upon lipoprotein binding and for receptor binding activity.* The Journal of biological chemistry, 2009. **284**(21): p. 14657-66.

99. Sahdev, S., S.K. Khattar, and K.S. Saini, *Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies.* Molecular and cellular biochemistry, 2008. **307**(1-2): p. 249-64.

100. Gschaedler, A., et al., *Effects of pulse addition of carbon sources on continuous cultivation of Escherichia coli containing a recombinant E. coli gapA gene.* Biotechnology and bioengineering, 1999. **63**(6): p. 712-20.

101. Hannig, G. and S.C. Makrides, *Strategies for optimizing heterologous protein expression in Escherichia coli.* Trends in biotechnology, 1998. **16**(2): p. 54-60.

102. Li, Q., et al., *Real time investigation of protein folding, structure, and dynamics in living cells.* Methods in cell biology, 2008. **90**: p. 287-325.

103. Li, Q., Huang, Y., Murray, V., Chen, J. and Wang, J., *A Protein Delivery Technology with Targeting Capability to Specific Intracellular Compartments.* Submitting to Nature Biotechnology, 2011.

104. Haugland, R.P., et al., *The handbook : a guide to fluorescent probes and labeling technologies*. 10th ed2005, Eugene, OR: Molecular Probes. iv, 1126 p.

105. Correa, F. and C.S. Farah, *Using 5-hydroxytryptophan as a probe to follow protein-protein interactions and protein folding transitions.* Protein and peptide letters, 2005. **12**(3): p. 241-4.

106. Rogers, J.M., L.G. Lippert, and F. Gai, *Non-natural amino acid fluorophores for one- and two-step fluorescence resonance energy transfer applications.* Analytical biochemistry, 2010. **399**(2): p. 182-9.

107. Kwon, I. and D.A. Tirrell, *Site-specific incorporation of tryptophan analogues into recombinant proteins in bacterial cells.* Journal of the American Chemical Society, 2007. **129**(34): p. 10431-7.

108. Lepthien, S., et al., *In vivo engineering of proteins with nitrogen-containing tryptophan analogs.* Applied microbiology and biotechnology, 2006. **73**(4): p. 740-54.

109. Hogue, C.W., et al., *A new intrinsic fluorescent probe for proteins. Biosynthetic incorporation of 5-hydroxytryptophan into oncomodulin.* FEBS letters, 1992. **310**(3): p. 269-72.

110. Wu, P. and L. Brand, *Resonance energy transfer: methods and applications.* Analytical biochemistry, 1994. **218**(1): p. 1-13.

111. Selenko, P. and G. Wagner, *Looking into live cells with in-cell NMR spectroscopy.* Journal of structural biology, 2007. **158**(2): p. 244-53.

112. Alberts, B., *Molecular biology of the cell*. 5th ed2008, New York: Garland Science.

113. Dong, H., S. Qin, and H.X. Zhou, *Effects of macromolecular crowding on protein conformational changes.* PLoS computational biology, 2010. **6**: p. e1000833.

ABSTRACT

DEVELOPMENT OF AN IN-CELL FÖRSTER RESONANCE ENERGY TRANSFER TECHNIQUE TO STUDY PROTEIN STRUCTURE INSIDE LIVING CELLS

by

VICTORIA LYNN MURRAY

December 2011

Advisor: Dr. Jianjun Wang

Major: Biochemistry and Molecular Biology

Degree: Doctor of Philosophy

The goal of my thesis is to develop an in-cell fluorescence technique that allows for measurement of the distances between fluorescence acceptors and donors within a protein or between two proteins inside the correct intracellular compartment of living cells. The successful achievement of this goal will allow us to obtain high-resolution structural information from a protein, one key step towards high-resolution structural biology of proteins inside the living cell.

To achieve this goal, we will apply the fluorescence resonance energy transfer (FRET) technique to the specifically labeled proteins inside the cells. Our rationale is to specifically label the protein(s) of interest in the test tube with a small molecule fluorophore and then deliver the labeled protein(s) into the correct intracellular compartment of living cells for in-cell FRET measurement. The QQ-protein delivery technique can specifically deliver a protein to its intracellular destiny based on its signal

sequence. This will result in special mammalian cells that contain a fluorescence labeled target protein with unlabeled intracellular endogenous proteins as the background. The FRET measurement will be performed on this specifically labeled protein and the calculated FRET-distance will be between the donor and acceptor of the protein(s) of interest, thus, high-resolution structural information of a protein inside living cells can be obtained using this novel approach.

AUTOBIOGRAPHICAL STATEMENT

VICTORIA L. MURRAY
Email: vlmurray9@gmail.com

EDUCATION

Ph.D. in Biochemistry and Molecular Biology — Wayne State University, Detroit, MI
Expected 08/2011 — *Dissertation topic: Exploring new methods in structural biology inside live cells using fluorescence spectroscopy*

B.S. in Biology — University of Illinois at Chicago, Chicago, IL
12/2000 — Major: Biology, Minor: Chemistry

PUBLICATIONS

- **Murray V**, Huang Y, Chen J, Wang J & Li Q (2011). A novel bacterial expression method with optimized parameters for very high yield production of triple-labeled proteins. In *Protein NMR Techniques 3rd Edition*, Humana Press. Editors: Alexander Shekhtman and David S. Burz. (In press)

- **Murray V**, Chen J, Huang Y, Li Q and Wang J. (2010). Preparation of very high-yield recombinant proteins using novel high cell density bacterial expression methods. Cold Spring Harbor Protocols. doi:10.1101/pdb.prot5475

- Sivashanmugam A, **Murray V,** Cui C, Yang Y, Wang J and Li Q (2009). Practical protocols for production of very high-yield of recombinant proteins in *Escherichia coli*. *Protein Science.* **18:** 936-948

- Sivashanmugam A, Yang Y, **Murray V,** McCullough C, Li Q and Wang J (2008). Structural Basis of human high-density lipoprotein assembly at atomic resolution. An invited review for *Methods in Cell Biology*, Elsevier Inc. Editor: Bhanu Jena, **90**: 327-364

- Li Q, Huang Y, Xiao N, **Murray V,** Chen J and Wang J (2008). Real Time Investigation of Protein Folding, Structure, and Dynamics in Living Cells. An invited review in *Methods in Cell Biology*, Elsevier Inc. Editor: Bhanu Jena, **90**: 287-325

ACADEMIC HONORS & FELLOWSHIPS

- Rumble Fellowship (2007 – 2008)
- Golden Key International Honor Society (2007 – present)
- Summer Dissertation Fellowship (2011)