

4-2016

Purification and preparation of intrinsically disordered proteins for NMR spectroscopy

Kyle Chamberlain
Purdue University

Follow this and additional works at: https://docs.lib.purdue.edu/open_access_theses



Part of the [Analytical Chemistry Commons](#)

Recommended Citation

Chamberlain, Kyle, "Purification and preparation of intrinsically disordered proteins for NMR spectroscopy" (2016). *Open Access Theses*. 758.

https://docs.lib.purdue.edu/open_access_theses/758

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Kyle Chamberlain

Entitled PURIFICATION AND PREPARATION OF INTRINSICALLY
DISORDERED PROTEINS FOR NMR SPECTROSCOPY

For the degree of Master of Science

Is approved by the final examining committee:

Nikolai R. Skrynnikov

Chittaranjan Das

William A. Cramer

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification/Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Nikolai R. Skrynnikov

Approved by Major Professor(s): _____

Approved by: Timothy Zwier

04/25/2016

Head of the Department Graduate Program

Date

PURIFICATION AND PREPARATION OF INTRINSICALLY
DISORDERED PROTEINS FOR NMR SPECTROSCOPY

A Thesis

Submitted to the Faculty

of

Purdue University

by

Kyle Chamberlain

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

May 2016

Purdue University

West Lafayette, Indiana

To those who wish to understand more than what they can initially see.
May our curiosity never cease, and our understanding multiply.

... And may our spouses grant us peace from such a busy mind.

ACKNOWLEDGEMENTS

I, of course, must first thank my parents for nurturing my curiosity at a young age. When the educational system required me to memorize, my parents encouraged me to question. They pushed me to delve deeper than knowing, and instead, asked me to extrapolate. This natural curiosity was the seed that grew my passion for pursuing the unknown. Of all the lessons my parents taught me through my life, this one has given me the greatest reward.

As I entered college, I was teeming with a need to understand our molecular world and pursue a career that could marry this inventive nature with an ability to help people. I unfortunately was unsure which path –or rather, which major- would allow me to attain my goal. Luckily, the same year I began my classes in chemistry, a new professor joined the staff at the University of St. Thomas. After helping her start her undergraduate lab, Dr. Lisa Prevette became more than just my advisor; she was a true mentor. Her leadership and passion for chemistry resonated with me, and ignited my drive to become the chemist I am today. I owe much to her guidance, and would not have even pursued my Master's degree if it weren't for her.

I would like to thank my fellow students and professors at St. Thomas and Purdue University. Suffering with others is akin to tempering steel; the result is stronger than the initial iron involved, and contains other elements that strengthen it. Spending seven years learning chemistry has proven that statement, and the laughter, tears, and work other students/colleagues have shared with me during that time has left impressions upon me that have strengthened my character. Without all of you, I would not be the chemist –or man- that I am today.

Finally, to my wife and children: thank you. You have given me so much love and support through my years at Purdue. More importantly, you all have given me a new

perspective in life. I have learned to step away from focusing on a world I cannot see, and learn to enjoy the world my senses perceive. I can finally cease focusing on the goal, and instead, enjoy the journey.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vii
LIST OF FIGURES	viii
NOMENCLATURE	x
ABSTRACT.....	xi
CHAPTER 1: INTRODUCTION.....	1
1.1 General Overview of Growing and Purifying Proteins for NMR Analysis.....	1
1.2 Gene Vectors and Cell Cultures.....	2
1.2.1 Common Expression Vectors for Cloning.....	2
1.2.2 Choosing Where to Clone a Protein Sequence into a Vector	4
1.2.3 Choosing an <i>E. coli</i> Strain for Protein Expression	5
1.2.4 Expressing a Protein in Bacterial Cells.....	7
1.3 Purifying a Cell Culture	8
1.3.1 Purification of a Water Soluble Protein	8
1.3.2 Purification of a Hydrophobic Protein.....	11
1.3.3 Preparing a Sample for NMR Analysis	13
1.4 Advantages and Disadvantages to Studying Proteins in this Manner.....	15
1.5 Proteins Examined Within Thesis.....	16
1.5.1 Histone H4 Tail.....	16
1.5.2 Mucin 1	16
1.6 Objectives	18
CHAPTER 2: EXPERIMENTAL PROCEDURES.....	19
2.1 Preparing Histone H4 Tail for NMR Analysis	19
2.1.1 Purifying Histone H4 Tail	19
2.1.2 Preparing an NMR Sample of Histone H4 Tail	23
2.2 Preparing Mucin 1 for NMR Analysis.....	23
2.2.1 Developing a Mucin 1 C+TM Subunit Plasmid.....	23
2.2.2 Expressing Mucin 1 C+TM.....	25

	Page
2.2.3 Revising the Mucin 1 Plasmid and Attempting Expression.....	28
CHAPTER 3: DISCUSSION AND FUTURE WORK	33
3.1 Histone H4 Tail.....	33
3.2 Mucin 1	33
3.3 Final Remarks	34
LIST OF REFERENCES	35

LIST OF TABLES

Table	Page
1.1 Short List of <i>E. coli</i> Strains and Their Differences.....	6
1.2 Common Dissolving Chemicals for Suspending Insoluble Proteins	12

LIST OF FIGURES

Figure	Page
1.1 A Basic Layout of a Vector Plasmid.....	3
1.2 An Example of a Cloning Region Within a Vector	4
1.3 The Effect of Open Reading Frames on a DNA Sequence.....	5
1.4 General Cell Culture Protocol.....	7
1.5 Diagram of Two Common Separation Columns	10
1.6 Process of Suspending a Protein in Detergent	12
1.7 Diagram of Nanodisc Formation and Protein Encapsulation	14
2.1 Absorbance Reading of H4 Tail Eluting From Ni Column	21
2.2 SDS PAGE Gel Results of A.) Thrombin Cleavage of H4 and B.) Size Exclusion Fractions Containing H4	22
2.3 Mass Spectrometry Results for Purified H4 Tail.....	22
2.4 A General Map of Mucin 1 Sequence.....	23
2.5 pET-28a Cloning Region and Intended Insertion Sites	24
2.6 Size Exclusion Results for MUC1-C+TM.....	26
2.7 SDS PAGE gel of Size Exclusion Fractions Collected	26

Figure	Page
2.8 Mucin1 IPTG (+/-) SDS PAGE Results	27
2.9 SDS PAGE results of A.) ROSETTA cells and B.) C43(DE3) Cells tested for Mucin1 Protein Expression.....	28
2.10 IPTG +/- Results for Trx-Mucin1 protein in 8M Urea	29
2.11 SDS-PAGE of Various Detergents Suspending Trx-MUC1 Protein.....	30
2.12 SDS-PAGE Gel of Ni Column Fractions Containing Trx-MUC1.....	31
2.13 Size Exclusion Fractions Containing Trx-MUC1	32

NOMENCLATURE

NMR: Nuclear Magnetic Resonance Spectroscopy

kDa or Da: kilo Daltons or Daltons. This is equivalent to molecular weight.

His-Tag: Histidine tag. 6 histidine residues added to a protein to make a charged end.

Trx: Thioredoxin. A protein meant to fuse with other proteins to encourage expression.

IPTG: isopropylthiol- β -D-galactoside. Induces overexpression of a protein.

pET: A family of vectors with a T7 promoter region.

pUC: A family of vectors with a Lac promoter region.

LB Media: Luria Broth media. Nutritionally rich medium for bacteria.

M9 Media: rich media for bacterial growth.

FPLC: Fast Protein Liquid Chromatography

EDTA: Ethylenediaminetetraacetic acid

SDS: Sodium Dodecyl Sulfate

LDAO: Lauryldimethylamine *N*-oxide. A detergent

DPC: Diphosphocholine. A detergent

DDM: n-Dodecyl β -D-maltoside. A detergent

DTT: 1,4- Dithiothreitol. A reducing agent

DMPC: Dipalmitoylphosphatidylcholine. A Phospholipid.

MUC1-C+TM: A segment of mucin 1 containing the cytoplasmic tail and the transmembrane helix.

ABSTRACT

Chamberlain, Kyle. M.S., Purdue University, May 2016. Purification and Preparation of Intrinsically Disordered Proteins for NMR Spectroscopy. Major Professor: Nikolai Skrynnikov.

NMR spectroscopy has recently become a promising field for protein characterization and dynamic studies. As the technology and pulse sequences improve for tracking proteins, a greater demand for developing effective purification protocols to produce NMR grade protein samples have arisen. This thesis explores two proteins: histone H4 tail and Mucin 1; Two very different proteins that require different methods of expression and purification to achieve a high enough yield for NMR analysis. H4 is a water soluble protein that weighs ~2.7 kDa, and has no extinction coefficient. Since the protein is too small for many methods of expression and purification, H4 was attached to thioredoxin and a His tag, purified via Ni affinity column, the tags were cleaved with thrombin, and the resulting H4 was purified via size exclusion. Mucin 1 is a transmembrane protein, and was hard to express within *E. coli* cells. Adding a thioredoxin tag, and switching to C43(DE3) strain of cells encouraged expression of the protein. The protein was then suspended in 8M Urea, and purified via size exclusion, and thrombin cleavage to remove the thioredoxin. These two different methods of purification provide great insight on how to purify many kinds of proteins for NMR analysis, and is critical knowledge for any protein NMR specialist.

CHAPTER 1: INTRODUCTION

1.1 General Overview of Growing and Purifying Proteins for NMR Analysis

While some kinetic studies allow chemists to observe the behavior of proteins within their natural environment, many methods –including crystallography and NMR spectroscopy- require the protein of interest to be isolated in an ultrapure buffer. Pure proteins are rarely abundant enough and selective enough to be extracted from their natural eukaryotic cell, so a level of genetic manipulation is required to obtain a sample of interest.¹ The established method to achieve a high yield of pure protein is as follows: clone the genetic sequence for a protein of interest into a vector (to form a plasmid), transfer the new plasmid into a prokaryotic cell -commonly a strain if *E. Coli*-, and overexpress the protein within a cell culture (commonly using IPTG to encourage overexpression). After the protein is overexpressed, the cell contains relatively very few proteins that require extraction from the solution to produce a pure sample of the protein.^{1,2,3} This makes the job far easier to attain a sample for analysis; even in a high enough yield for NMR spectroscopic analysis.

Depending upon the protein being expressed, methods of purification can vary dramatically. Size exclusion and ion exchange columns are the gold standard for separating the protein from the rest of the molecular soup produced during expression.^{1,2,4} These processes are only effective if the charge or size of a protein varies from common proteins expressed by *E. Coli*.¹ In a case where a protein closely resembles other proteins within the cell, a histidine tag and thrombin cleavage site can be attached to the protein DNA sequence. This gives the protein a highly cationic tail that has a high affinity to a nickel column, allowing effective separation.^{1,5} Thrombin can then be added to cleave this site from the tail of the protein, and purified via size exclusion or dialysis.^{1,6} Once

again, these are just a few examples within a library of options available to allow purification of a protein.

Once a protein is purified, it can be either lyophilized or buffer exchanged into a buffer ideal for analysis. In the case of NMR spectroscopy, these buffers are carefully prepared to mimic the natural environment of the protein while, at the same time, reducing potential signal impairing effects. These buffers allow the protein (ideally) to fold properly in its native conformation, and can be analyzed via various two and three dimensional pulse sequences.

1.2 Gene Vectors and Cell Cultures

1.2.1 Common Expression Vectors for Cloning

Vectors are a viral or engineered circular plasmid which contain one or many insertion sites viable for inserting a DNA sequence.¹ A useful vector should contain these basic components: an origin region to start transcription of the vector in a bacterial organism, a promoter region to encourage a high copy volume of the sequence within the vector, a cloning/restriction site(s) for inserting a plasmid DNA sequence of interest, and a bacterial resistance marker to purify the future batch of bacterial cells containing the vector. There are many other important components comprising a vector, but only the aforementioned components will be explored in this paper.^{1,7}

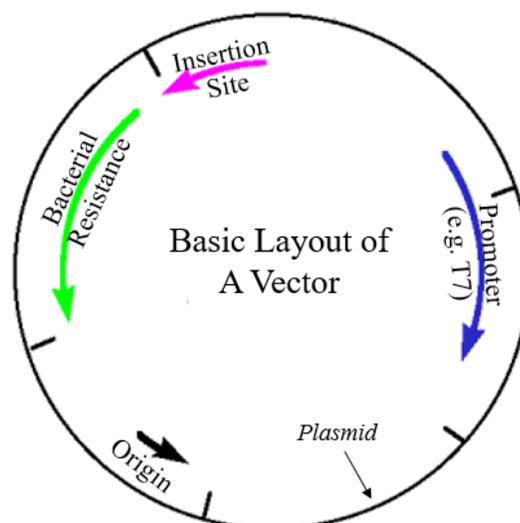


Figure 1.1 A Basic Layout of a Vector Plasmid

One of the basic vector sites used for cloning is the pUC family (pUC 8 or pUC 9 is commonly used). The plasmid is only ~2.8 Kb, contains a Lac promoter, and is resistant to ampicillin. The lac promoter is induced by the addition of *isopropylthiol- β -D-galactoside* (*IPTG*) during expression. These plasmids are easy to use, and are quite cheap to acquire. They are unfortunately, not the most efficient vector for protein expression though.^{8,9}

The engineered pET family is the usual vector of choice for protein growth within bacterial cells. It contains an f1 origin of replication to encourage production even when infected with the M13 helper phage. The T7 promoter in pET vectors is much stronger than the lac promoter in pUC vectors when induced by IPTG. The cloning sites are highly customizable, and can contain a Histidine Tag (e.g. pET 28) or thioredoxin (e.g. pET 32) on the N or C-terminal side of your insertion site to make purification easier. The resistance marker can change depending upon the class of pET vector as well.¹⁰

The pUC family is cheap and effective if the protein intended for cloning is easy to purify from the cellular soup. The pET family is much more expensive, but they are very useful for selectively aiding your protein through expression and purification, and is amenable to almost any protein available for expression in bacterial cells.

1.2.2 Choosing Where to Clone a Protein Sequence into a Vector

Many expression vectors contain multiple restriction sites available for cloning your DNA sequence into the vector. When using a vector like pET 28, the choice of restriction site can dramatically affect the conditions in which your protein is expressed. Consider the restriction sites in figure 1.2 below.

Cloning/Expression Region

```

T7 Promotor  Xba I          His Tag
...CACTATA...TCTAGAAATAAT...CATCATCATCATCATAGCA...

Xho I  His Tag          Eac I    T7 Term.
CTCGAGCATCATCATCATCATCAGCTC ... TAGCATAA...

```

Figure 1.2 an Example of a Cloning Region within a Vector

If the DNA sequence was cloned into the plasmid at Xba I (N-terminus) and Eac I (C-terminus), the sequence would only express the protein inserted. If the protein is hard to purify alone, a Histidine tag can be attached to the N-terminus of the protein –using Xba I and Xho I restriction sites- or to the C-terminus using the Eac I restriction site.^{1,5,10} It is important to consider what the protein may need to effectively purify it after expression within a bacterial cell.

Another point of consideration during cloning is watching the open reading frames. Since amino acids are created in sequences of three nucleotides (a codon) at a time, it is important to consider where transcription starts, and what codons are being created as a result.¹

The Effect of Open Reading Frames

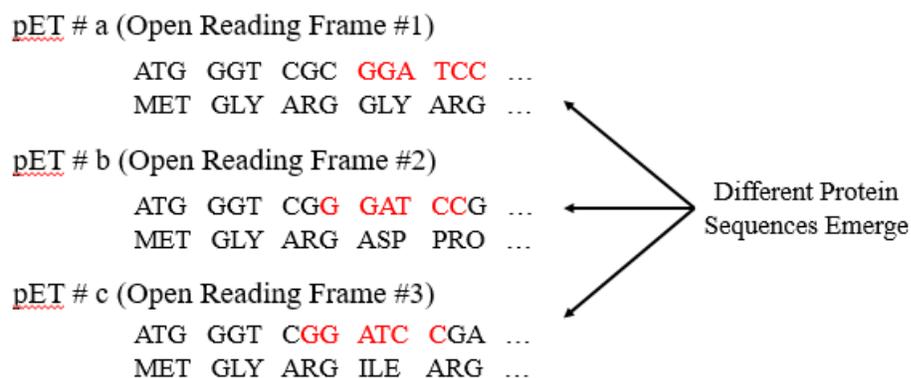


Figure 1.3 The Effect of Open Reading Frames on a DNA Sequence

As shown above, depending upon how a sequence is shifted, the DNA sequence inserted can result in entirely different protein sequences.^{1,10} Luckily, the pET vector family (as well as a few other expression vectors on the market) accounts for this by offering three different vectors (a,b, or c) to allow placement of your protein sequence into any reading frame of your choice. Always consider what resulting sequences within the coding region of the vector may emerge, and which reading frame the intended bacterial colony transcribes in.¹⁰

The process of cloning a DNA sequence into a vector is mentioned in literature, but this paper will not describe the procedure; only the considerations required to plan construction of a plasmid. Plasmids used for research within this thesis were constructed by external commercial companies.

1.2.3 Choosing an *E. coli* Strain for Bacterial Growth

Escherichia coli is the most commonly used bacterial host for protein expression due to its ability for rapid growth, simplicity of the cellular structure, and ease of protein expression.^{2,4,5} For most labs, the question of which bacteria should be used for expression is not asked; rather the question is, “which strain of *E. coli* should be used”?

The ‘gold standard’ strain of *E. coli* for expression is BL21. This cell is equipped for lac promoter driven gene expression, which means if the plasmid used contains a pUC family vector the protein of interest will be expressed effectively after the addition of IPTG. Unfortunately, any plasmid using a T7 promoter is unable to be effectively translated in this strain. Furthermore, transmembrane and other toxic proteins would not be able to be effectively expressed in BL21.¹⁰

Table 1.1 Short list of *E. coli* Strains and Their Differences

<i>E. Coli</i> Strain	Promoter	Pros	Cons
BL21	Lac	Doesn't overexpress toxic proteins Can express simple proteins	Cannot express transmembrane proteins well Cannot use T7 promoter plasmids Relatively lower yields of protein produced
BL21(DE3)	Lac & T7	Can use T7 promoter plasmids Produces higher yields of protein	Once induced, toxic proteins are expressed Cannot express transmembrane proteins well
ROSETTA(DE3)	Lac & T7	Express proteins with rare codons BL21 derivative	Once induced, toxic proteins are expressed Cannot express transmembrane proteins well
C43(DE3)	Lac & T7	Can overexpress toxic proteins Express transmembrane proteins BL21 derivative	Very expensive

BL21(DE3) is a superior variant, since it can express using the lac promoter –in the same way as BL21-, and can express with the T7 promoter. These cells are engineered to rapidly overexpress the protein of interest -as well as toxic proteins- when IPTG is added. This may produce a higher yield of protein, but also it is more important to watch cellular growth since it is easy to cause overgrowth and lose the batch.^{10,11}

ROSETTA(DE3) cells are a variant of BL21(DE3). The strain is engineered to recognize mammalian codons that are rarely found in bacterial cells. This allows certain mammalian proteins to be expressed in *E. coli* without the potential translation errors that commonly occur. Beyond this very helpful trait, the cells are almost identical to BL21.¹⁰

C43(DE3) cells are a variant of BL21(DE3) as well. They are engineered to proliferate within a toxic environment. This means proteins can be overexpressed more effectively within this cell, and toxic proteins can be effectively expressed in this strain.

Furthermore, hydrophobic or transmembrane proteins can be expressed in C43 with a relatively high yield. This strain of *E. coli* is considered to be one of the best strains for protein expression, but the cost may deter some from choosing C43 over other BL21 variants.¹¹

1.2.4 Expressing a Protein in Bacterial Cells

Once a plasmid is prepared and appropriate bacterial cells are chosen, expression of the intended protein can commence. To accomplish this, the plasmid must be first shuttled into the bacterial cell. A small volume –usually 1 or 2 μL – of the plasmid is chilled with a small volume –usually 10 μL – of bacterial cells within a suspension of Luria broth (LB) media. The solution is then heat shocked to encourage the bacterial cells to break open and shuttle the plasmid in.^{1,11,12}

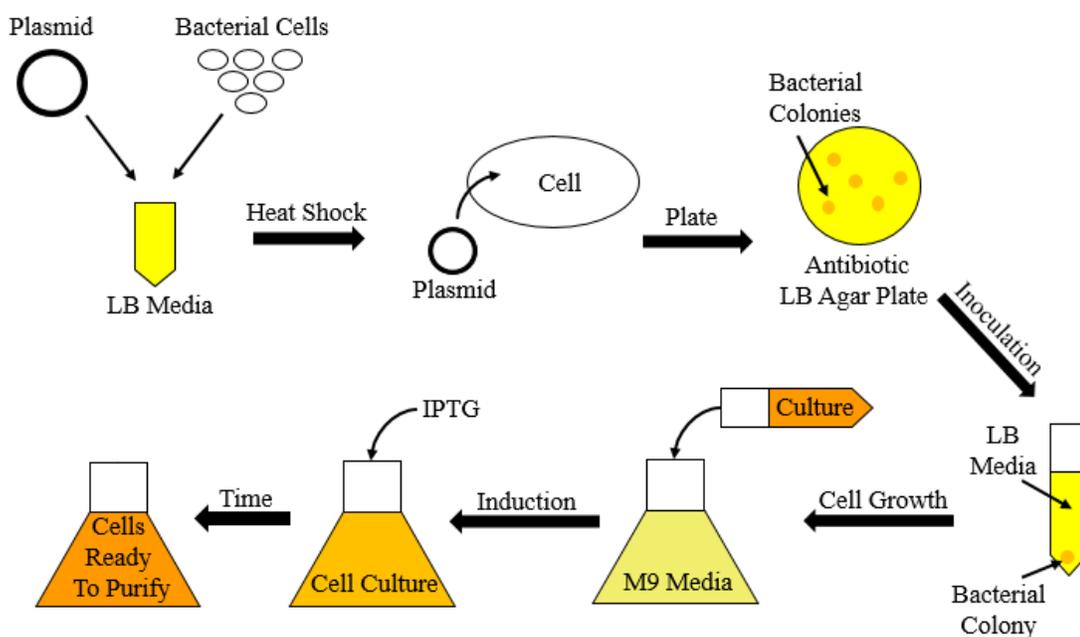


Figure 1.4 General Cell Culture Protocol

Once the bacteria have enveloped the plasmid, the solution is spread over an antibiotic LB agar plate. The antibiotic should match the antibiotic resistance of the plasmid introduced. This should encourage colony growth of cells exclusively carrying the plasmid. A colony can then be transferred to a vial containing a few milliliters of LB media, and incubated to encourage bacterial growth.^{1,11,12}

The culture can then be transferred to a half liter or liter of purified M9 media containing antibiotic. The media can contain isotopically labeled N^{15} ammonium salts or C^{13} glucose if desired for future research. The container should be either twice as large as the volume of the solution, or equipped with an aerator to ensure sufficient oxygen is present for cell growth. The solution is then incubated and agitated to encourage bacterial growth. As the optical density (A_{600}) of the solution reaches 0.7, IPTG is added to induce overexpression of the protein. After a period of incubation, the cells will be ready for purification. This incubation time may vary depending upon the cells used (see chapter 1.2.3), and the plasmid constructed.^{1,11,12}

1.3 Purifying a Cell Culture

1.3.1 Purification of a Water Soluble Protein

Hydrophilic proteins are relatively easy to purify, since they dissolve in most buffers safe for FPLC, and retain their native structural conformation without special detergents or lipids. The first step is to spin down the cell culture, and collect the pellet. The bacterial cells can then be re-suspended in a buffer viable for subsequent purification steps.^{1,12}

Buffers may vary depending upon the protein of interest and intended pH of the solution. For physiological pH (~7.4) 50 mM Tris-HCl is a useful buffer. Adding salt can mimic the environment of a cell, and adding a small concentration of Ethylenediaminetetraacetic acid (EDTA) can help prevent aggregation or protein fragmentation during purification.^{1,12}

Once a buffer is chosen, and the cells are re-suspended in it, the solution is ready for cell lysis. There are two common methods used for cellular lysis: First, using a French press and exert over 1000 PSI of force on the solution. Second, sonicate the solution with rapid, high frequency pulses. This ensures the proteins within the cell are extracted and suspended in the buffer. The solution is then re-centrifuged, and the supernatant - containing water soluble, cellular proteins- is collected.^{1,11,12}

Once the solution is collected, the proteins can be separated in a variety of ways. If the protein is a size quite different from other naturally occurring proteins within *E. coli*, the sample would be run through a gel filtration (size exclusion) column. At a slow flow rate, the larger proteins elute at an earlier rate than the smaller proteins, and elution fractions are collected over time. These fractions should contain a purified sample ready for further study.^{1,11,12,13}

Alternatively, if a sample cannot be effectively separated via size, the protein can be separated by charge. Either the protein may naturally have a high charge, or a histidine tag is attached to the end of itself. When introduced into the column, a buffer (buffer A) is washed through the column allowing the charged protein to stick to the resin, and all other proteins elute through the column. After the column is washed with at least 3 or 4 column volumes, a second buffer (buffer B) is introduced into the column as a gradient between buffers A and B. Buffer B should contain a chemical with a high affinity to the resin (such as Imidazole for a Ni affinity column) that displaces the protein of interest, and elutes it. The fractions are collected, and a pure sample should be extracted.¹

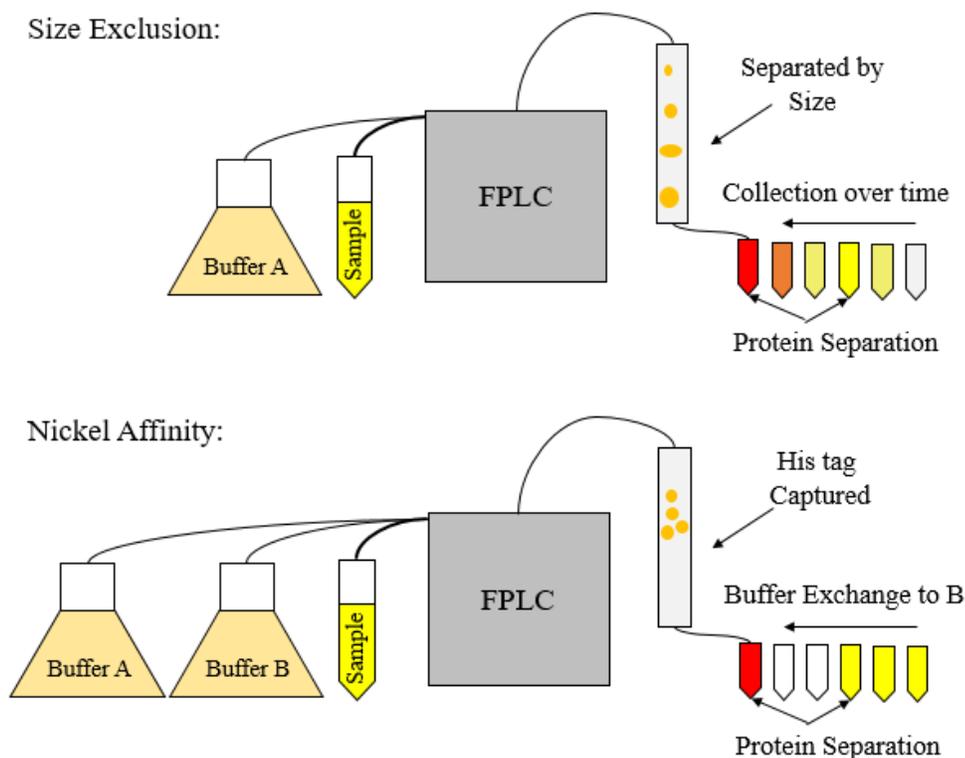


Figure 1.5 Diagram of Two Common Separation Columns

If a His tag is attached to a protein sample, it is imperative to remove the tag. Usually a His tag is followed by a thrombin cleavage site. After a sample is separated via Ni affinity column, the eluted fractions can be prepared for cleavage. The sample must first be buffer exchanged to a low salt buffer that doesn't contain imidazole –since imidazole inhibits activity of thrombin-. Furthermore, if the sample buffer contains detergents or lipids (See section 1.3.2), thrombin activity will also be inhibited. Buffer exchange can be accomplished via dialysis or filter centrifugation.^{1,12,13}

Once the protein is in an imidazole free buffer, 3-10 units of thrombin is added per 1mg of protein at room temperature overnight. This value can vary, and is highly dependent on temperature, time, and the protein being used. Much of literature suggests testing multiple conditions for optimal thrombin cleavage, since too much cleavage can result in degradation of the protein of interest.¹⁴

Once the protein is cleaved, the sample will need to be re-purified. This can be accomplished via dialysis (if the protein is large enough for cellulose membranes), or via size exclusion. Even if the protein doesn't require cleavage, it may require multiple purification steps to attain an NMR grade sample. It is important to balance the iterations of purification with the half-life of the protein. If the concentration halves every 6 hours and the procedure calls for 3 column purifications, there may not be a sample after purification ends. Given any of the above methods, once a sample is purified, it is ready for buffer exchange to a concentrated sample ideal for NMR analysis.

1.3.2 Purification of a Hydrophobic Protein

Hydrophobic or amphiphilic (partially hydrophobic) proteins are hard to purify since they usually degrade fast in a non-native environment, and require detergents or lipids to encapsulate the protein through purification. The first step for purification is deciding which detergent or lipid would be best for suspending the protein of interest in the buffer. Table 1.2 below describes a few of the common dissolving chemicals used for suspending insoluble proteins.^{11,13}

There are many factors to consider when choosing the right detergent/chemical for suspending a protein: Does the protein retain its native conformation? If it doesn't it may be much easier for the protein to degrade or become cleaved by other enzymes in solution. Does it work well with the method of purification intended? Urea is highly corrosive, and some detergent micelles may be too large for some filtration methods. Does the protein require a mild or harsh detergent? Mild detergents aren't as positively charged and have long hydrocarbon tails. They are harder to suspend some proteins, but are gentle on easily degradable/reactive proteins. Harsh detergents are the opposite. How easy is it to exchange the protein to another buffer? Some detergents like SDS adhere to proteins quite strongly, and require buffer exchange to urea, then to another detergent; Valuable time would be spent exchanging the protein to a buffer viable for NMR while the protein is degrading.^{15,16}

Table 1.2 Common Dissolving Chemicals for Suspending Insoluble Proteins

Dissolving chemical	Pros	Cons	Purification Steps
SDS	Highly efficient at dissolving hydrophobic proteins. Low concentrations of SDS is needed.	Adheres strongly to proteins. Reduces His-tag affinity to Ni columns. Unfolds proteins.	Buffer exchange to Urea, then Buffer exchange to detergent/lipids.
Urea	Very efficient at dissolving proteins. Doesn't interfere with Ni columns.	Can cause problems with HPLC's. Unable to prevent proteins from aggregating. Unfolds proteins.	Buffer exchange to detergent/lipids are required to allow proper folding.
LDAO	Allows proper folding of protein.	Very harsh detergent. A fifth of membrane proteins are soluble in it.	Can be characterized as is. Exchange into a nanodisc is quite simple.
DPC	Less harsh than LDAO. Forms bicelles. Mimics the behavior of a cell surface. Has been used in protein NMR experiments.	Still fairly harsh detergent. Doesn't always allow proper folding of a TM protein. Expensive detergent.	Can be characterized as is. Exchange into a nanodisc is quite simple.
DDM	Very mild detergent. Allows proper folding of proteins. Has been used in protein NMR experiments. Doesn't interfere with Ni columns.	Cannot suspend a lot of proteins.	Can be characterized as is. Exchange into a nanodisc is quite simple.
Triton X	Allows folding of proteins. Prevents oligomerization of proteins. Cheap detergent.	Cannot suspend a lot of proteins. Uncommonly used alone for characterizing proteins.	Can be characterized as is. Exchange into a nanodisc is quite simple.

Once a detergent is chosen, the buffer can be prepared in a similar manner as described in section 1.3.1. Tris-HCl is a useful physiological buffer, and salt and EDTA can be useful for any buffer. On top of the basic solution, the intended detergent should be added well above the critical micelle concentration to ensure micelles form to trap the protein of interest. If the protein of interest has a high propensity to aggregate, turning the buffer into a 10-30% solution of glycerol or adding DTT remedies the issue.^{15,16,17}

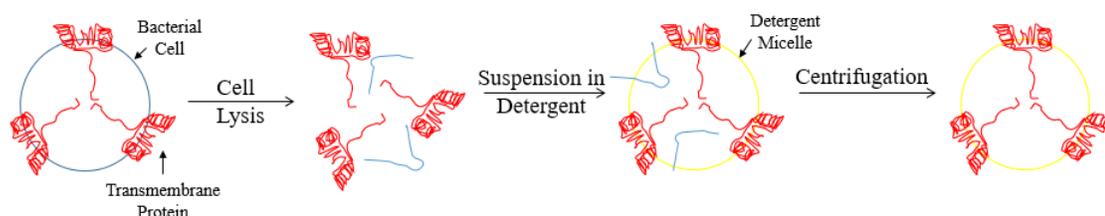


Figure 1.6 Process of Suspending a Protein in Detergent

Purification begins in a similar manner as described in section 1.3.1. The bacterial cells are centrifuged down to a pellet, and the pellet is re-dissolved in the premade buffer containing detergent. The cells are then lysed, and the proteins should dynamically re-suspend in a detergent micelle. Then the solution is re-centrifuged to discard the inclusion bodies. The supernatant is then prepared for purification. Ion and size exclusion columns are useful for purifying, although it is important to consider the effects of the detergent on the machines and columns being used (see table 1.2 and section 1.3.1 for more detail). Once purified, the samples are ready to be prepped for NMR analysis.^{15,16,17}

1.3.3 Preparing a Sample for NMR Analysis

For water soluble proteins, the NMR buffer is ideal when there is a low salt concentration. Citrate, Tris-d11, or Na/K phosphate buffers are useful for NMR analysis (using 95% deuterated water as a solvent). Buffers should be prepped for physiological pH as well as below pH 6, since NH exchange is slowed. 5mM DTT or 10mM CHAPS are useful to prevent dimer/oligomerization of the protein, and 5mM EDTA can be useful to prevent denaturation of the protein. It is generally advisable to first examine literature to mimic a buffer previously used for the protein of interest (or a similar protein), or modify the buffer based on literature.^{15,16}

If the sample is being prepped at a future time, the sample can be buffer exchanged to water via dialysis or filter centrifugation, and then lyophilized. The protein can then be directly rehydrated in an NMR buffer of choice at the proper volume. Alternatively, the NMR buffer can be added to the purified sample, and buffer exchanged via filter centrifugation, then concentrated to the proper volume.

For Hydrophobic proteins, the same methodology can be followed. Detergents such as DDM (n-Dodecyl β -D-maltoside) and DPC (diphosphocholine) can remain in the NMR buffer since it is ideal for characterizing proteins in their native conformation.^{15,16} Others –like urea- require buffer exchange to another detergent for NMR analysis. In many cases these detergent micelles can mimic a native lipid environment, but for certain proteins, detergent is not sufficient to encourage native behavior or conformational

structure. For these class of proteins, nanodiscs are ideal for mimicking the cell wall while being small enough for NMR characterization.^{18,19}

Nanodiscs are lipids (such as DMPC) encapsulated by a flexible α -helical amphiphilic protein (commonly MSP). The size of the disc is determined by the size of the amphiphilic protein, and the formation of a nanodisc is spontaneous and very favorable. To form the discs, a solution is prepared with MSP (or another protein) and lipids at an 80:1 lipid to amphiphilic protein ratio (suggested ratios may differ depending upon the protein). Nanodiscs spontaneously form. The Protein solution of interest is added to the nanodiscs, and agitated. The final mixture can be dialyzed in water to elute the detergent.^{18,19,21}

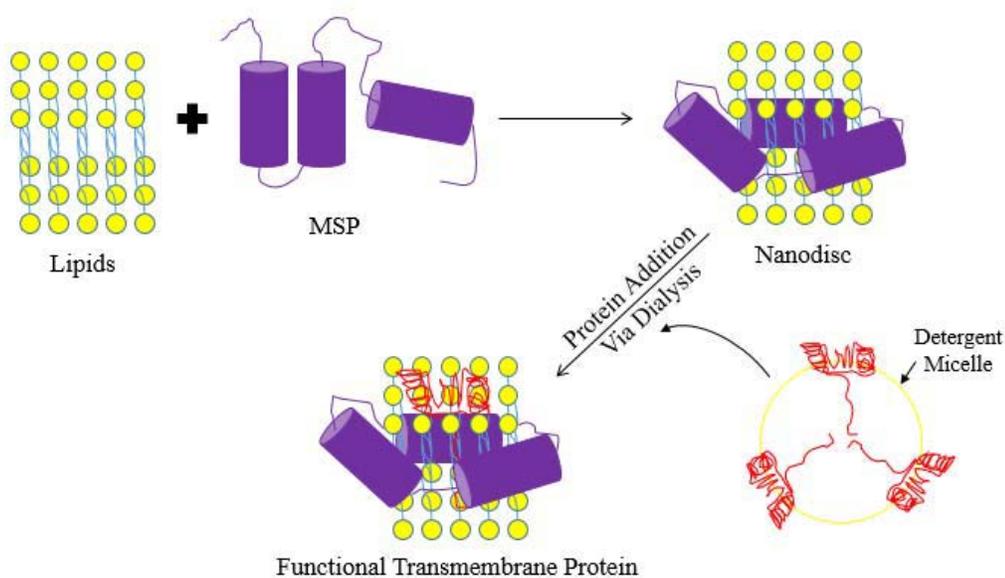


Figure 1.7 Diagram of Nanodisc Formation and Protein Encapsulation

Nanodiscs are incredibly useful since they generally increase the half-life of a protein dramatically, and even encourage transmembrane proteins to behave normally;

As if it were embedded in its native membrane within the cell. These attributes are ideal for 3D NMR experiments that may take days or even weeks to run.^{18,19,21}

1.4 Advantages and Disadvantages to Studying Proteins in This Manner

Concerning the use of bacteria as a medium for expression: When a eukaryotic protein is able to be expressed in *E. coli*, the yield is far higher than being able to extract it from its naturally occurring source, and with greater ease. Unfortunately, not all eukaryotic proteins can be effectively expressed in this manner. Some bacteria may disassemble proteins during transcription, others are simply unable to transcribe the protein, and some transcribe it improperly and the protein cannot fold in its native conformation.¹¹ There are currently cell-free methods of protein production that use the same molecular machinery for transcription, but floating in a solution instead of within a cell. This produces a pure product right away (which is always beneficial), but this method is relatively new, and has its own issues.²⁰

Preparing cell cultures for NMR analysis is quite challenging, relative to preparing a protein for mass spectroscopy, x-ray crystallography, isothermal calorimetry studies, or other methods of protein characterization. This is because NMR requires a relatively high concentration of protein, and a very pure sample of that protein. For some proteins that have a low half-life, this can prove to be more than challenging. Why study via NMR when many structural, binding, and dynamic studies can be performed through the means of various experiments?^{16,21}

The reason is NMR is able to do the work of many studies, and obtain more relevant data. NMR is able to not only characterize the structure of the protein, but it is able to do so in its native conformation, and see its dynamic structure (even for disordered proteins). NMR can perform binding studies, and see how the structure changes as a protein binds. The process of obtaining a sample for NMR analysis may be challenging, but information obtained is much faster and more illuminating once the sample is prepared.^{16,21}

1.4 Proteins Examined Within Thesis

1.5.1 Histone H4 Tail

The Histone H4 Tail is one of 5 proteins that comprise the structure of chromatin in eukaryotic cells. Chromatin helps regulate DNA replication, and packages DNA after replication. It also helps reinforce and protect the DNA during mitosis. For better or worse, the H4 tail is easily subject to acetylation and methylation, which can affect DNA replication -and ultimately alter gene expression- because it interferes with inter-nucleosomal packaging. The actual process of these interactions are not fully known, and is subject to much study.²²

The H4 Tail is useful for NMR in two ways: first, further studies are needed to fully understand its interaction with other proteins/chemicals and its ultimate effect on DNA replication. NMR would be useful for observing some of these interactions. Second, H4 is a small disordered protein that is hard to characterize given modern pulse sequences. It is a great protein for testing certain new pulse sequences, and furthering the scientific field of NMR as a whole.

1.5.2 Mucin 1

Mucin 1 is a transmembrane protein commonly located in the apical border of a mucosal cell. The protein contains a sea urchin sperm protein-enterokinase-agrin (SEA) domain that undergoes autocleavage, producing two subunits. Under cellular stress the N-terminal subunit of MUC1 is released outside the cell to form a physical barrier around the cell surface, while the C-terminal subunit remains partially embedded in the cell wall and extends into the cytoplasm. When this occurs across a series of cells on the surface of tissue, a thin mucus layer forms (e.g. within the colon, intestines, and breasts).²³

The cytoplasmic portion of MUC1 is able to interact with many transcription factors, epidermal growth factor receptors (EGFRs), tyrosine kinases, and other signaling

proteins throughout the cell that promotes cellular resistance and transcription of itself.^{23,24,25} It is believed MUC1 usually remains a homodimer when clustered in low concentration around the apical border of the cell, but when polarity of the cell is compromised, MUC1-C disperses across the entire cell surface, and is able to interact with the aforementioned proteins.^{23,25}

Polarity of the cell is affected when the cell undergoes extracellular stress, or when the cell becomes cancerous. In the prior condition, MUC1-C will temporarily encourage cellular signaling to bolster cellular defense, and then retreat back to the apical border as cellular stress is reduced. It is theorized that in cancerous cells, MUC1-C will spread across the cell long enough to proliferate, and bind with tyrosine kinases and ERBB2 to further disrupt the polarity of the cell.^{23,25} Once this occurs, MUC1-C will interact with JAK1 and STAT3 to promote transcription of itself, and ultimately establish an autoinductive loop that will encourage over expression of MUC1 and STAT3. As the concentration of MUC1 increases, MUC1-C domain will oligomerize and spill into the cytoplasm and further dimerize with Importin- β to cross the nuclear barrier, or with HSP70 and HSP90 to enter the mitochondria, and promote expression of genes involved in proliferation and survival of the cell.²³ This cascade of events leads to an inhibition of apoptosis, increased defense of the cellular membrane, and accelerated cell growth/division; many of the destructive hallmark signs of cancer. Although this proposed process is currently debated, it is known mucin 1 has a direct influence on the proliferation of cancerous cells.^{23,24,25}

It is believed MUC1-C dimerizes with many of these signaling proteins by forming a disulfide bond within its CQC motif found in the cytoplasmic tail of the protein. By blocking the CQC motif, or mutating it to AQA, MUC1-C loses its activity almost entirely, and renders a cancerous cell non-tumorigenic; Sometimes killing the cell entirely.^{23,24,25} Surprisingly, MUC1 does not bind to STAT3 with its CQC motif, but the inhibition of CQC with a cell penetrating peptide like GO-201 disrupted MUC1-STAT3 interaction completely; Indicating the CQC motif was still important in maintaining stability of the complex.²⁴

Beyond simply tracking MUC1-C signaling, and noting the significance of the protein's inhibition within the cell, animal and human trials have shown CQC inhibiting peptides - such as GO-201 - actually cause recession of mucosal cancer within a biological system.^{23,24} This makes MUC1-C a very significant protein, and one that should be studied thoroughly to devise the most effective means of inhibition for cancer treatment.

Unfortunately, besides the alpha helical transmembrane (~23 amino acids long) segment, the cytoplasmic subunit is intrinsically disordered. Because of the disordered nature of MUC1-C, it is hard to predict its dynamic structure, and ultimately, its method of interaction with important molecules of interest.²⁶ Furthermore, characterizing its structure via crystallography or other solid state methods alone will not provide any useful data, since it is improper to assume the packing structure of an intrinsically disordered protein is in any way indicative of its dynamic structure in solution. Therefore, Solution NMR is necessary to characterize the structure of MUC1-C alone, as well as characterizing it as a homodimer and heterodimer with many cytoplasmic proteins such as STAT3 to understand MUC1's true role in cancer growth and function.

1.6 Objectives

The objective of this research was to develop and optimize protocols for expressing and purifying the histone H4 tail and MUC1. For MUC1, once the C subunit was characterized, focus was directed to creating a plasmid to express the transmembrane domain –as well as the C subunit- of MUC1. Once accomplished, plans were made to purify the protein and prepare samples containing DDM detergent and nanodiscs (separately) for NMR analysis and characterization.

NMR studies have been performed on both proteins studied within this thesis, but the assignments will not be mentioned in detail since other members of Dr. Skrynnikov's lab completed this task. The work embodied here focuses almost exclusively on developing, expressing, and preparing protein samples for NMR analysis.

CHAPTER 2: EXPERIMENTAL PROCEDURES

2.1 Preparing Histone H4 Tail for NMR Analysis

2.1.1 Expressing and Purifying Histone H4 Tail

The plasmid for the histone H4 tail was developed by Adam Groves of Dr. Skrynnikov's lab, and created commercially by GenScript. The genetic sequence was cloned into a pET-32b vector containing ampicillin resistance. Thioredoxin and a histidine tag was attached to the N-terminal end of the protein. BL-21(DE3) ultracompetent cells were used as the bacterial medium for protein expression. The following is the optimized purification protocol developed for histone H4 tail.

1 μ L of 0.01mg/mL Plasmid was added to 10 μ L of BL-21(DE3) cells, and allowed to mix for 30 minutes on ice. The sample was then heat shocked at 37° C for 30 seconds, then cooled on ice for 2 minutes. 900 μ L of LB media was added to the mixture, and incubated for a half hour at 37° C. The cells were then spun down, and 200 μ L were transferred to an ampicillin resistant plate.

After 16 hours the plates were checked, and a colony was chosen for cell cultures. The colony was transferred to two vials containing ~7mL LB media and incubated for 16 hours at 33° C while agitated at 250 rpm. After that time, 0.5L of M9 media was prepared at pH 7.4 containing ampicillin in a 2 liter flask and autoclaved. For initial runs to test expression, nothing was isotopically labeled. After expression conditions were optimized, N¹⁵ ammonium salts and C¹³ glucose were added to the media.

5 mL of the LB media cultures were transferred to the M9 media, and incubated at 37° C while agitated at 300rpm. When the optical density (at 600 nm) reached ~0.7

(roughly 8 hours), 1mL of IPTG and more ampicillin was added to induce overexpression of our protein. The cells were ready for purification after about 15 hours.

The cellular media was centrifuged down at 8000 rpm for 15 minutes at 4°C. The supernatant was discarded, and the pellet was suspended in Buffer A (50 mM Tris-HCl, 200 mM NaCl, and 10 mM EDTA). The cells were lysed via French press at 1100 psi three times. The final product was centrifuged again at 28000 rpm for 30 minutes at 4° C, and the supernatant was collected.

The solution was prepared for FPLC by passing through a .22 micron filter. Since H4 contained a His tag, an Atka Start FPLC was equipped with a 10mL Ni affinity column, and equilibrated with buffer A. The sample was injected into the FPLC at 2mL/min, and washed with buffer A for 5 column volumes. A gradient between buffer A and buffer B (50 mM Tris-HCl, 200 mM NaCl, 10 mM EDTA, and 250 mM imidazole) was created, and fractions of the elution were collected. An example of the elution results are shown in figure 2.1 below.

As one can see, there was sufficient separation displayed on the chromatogram as imidazole (buffer B) saturated the column. The strong elution peak was confirmed to contain the Trx+H4+His-tag fusion protein via SDS PAGE gels (not shown). The peak fractions were collected, concentrated to 5 mL by centrifuging the sample at 1300 rpm using an amicon 5 kDa MW cut off filter tube.

1 mg of thrombin was added to the solution at 37° C for 4 hours to cleave the His-tag and thioredoxin (Trx). Normally a concentration of the protein –and the related amount of thrombin needed- would be determined using UV/VIS, but H4 has an extinction coefficient that is approximately zero. Therefore it is quite hard to determine a concentration. After a series of thrombin cleavage tests, the aforementioned method of thrombin activity was sufficient for protein cleavage (as seen in figure 2.2 below).

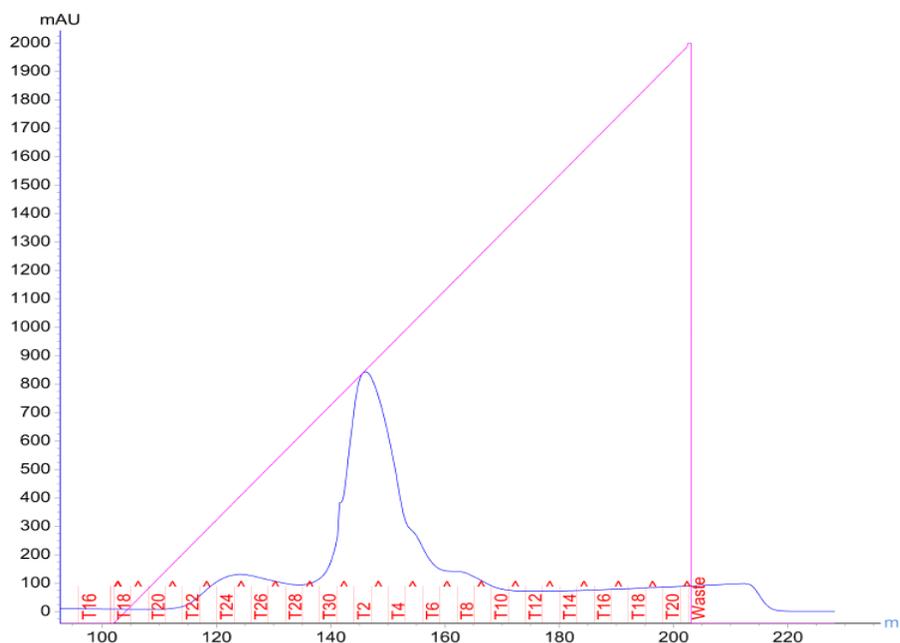


Figure 2.1 Absorbance Reading of H4 Tail Eluting from the Ni Column

Once the thrombin completed cleaving the protein, the sample was filtered through a .22 filter, and prepared for size exclusion. An Atka Start FPLC was equipped with a 120 mL HiLoad 16/60 superdex 75 prep grade size exclusion column, and the system was equilibrated with nanopure water. The sample was injected at a flow rate of 0.5 mL/min, and 5mL fractions were collected. After extensive sets of SDS PAGE gels, H4 was discovered purified (see figure 2.2).

The pure histone H4 tail was then lyophilized, and ~13mg samples were obtained per preparation. H4 was confirmed using mass spectrometry. The theoretical molecular weight of H4 with an N¹⁵ label should be 2693.4 Da, and experimental results (see figure 2.3) show the molecular weight of the protein is 2693 Da. Further confirmation was made via NMR analysis using a 2D HSQC sequence (not shown).

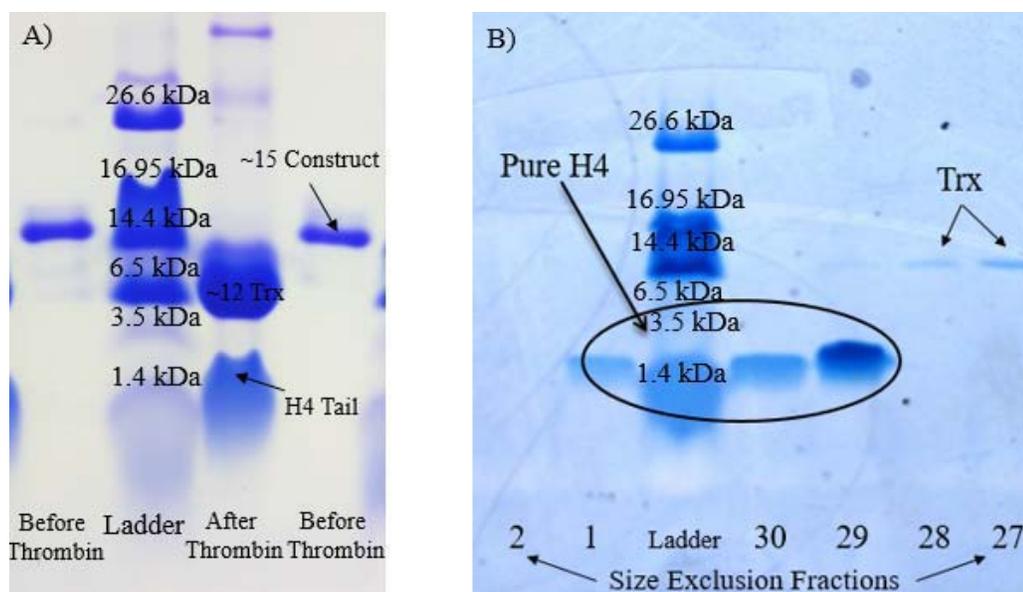


Figure 2.2 SDS PAGE Gel Results of A.) Thrombin Cleavage of H4 and B.) Size Exclusion Fractions Containing H4

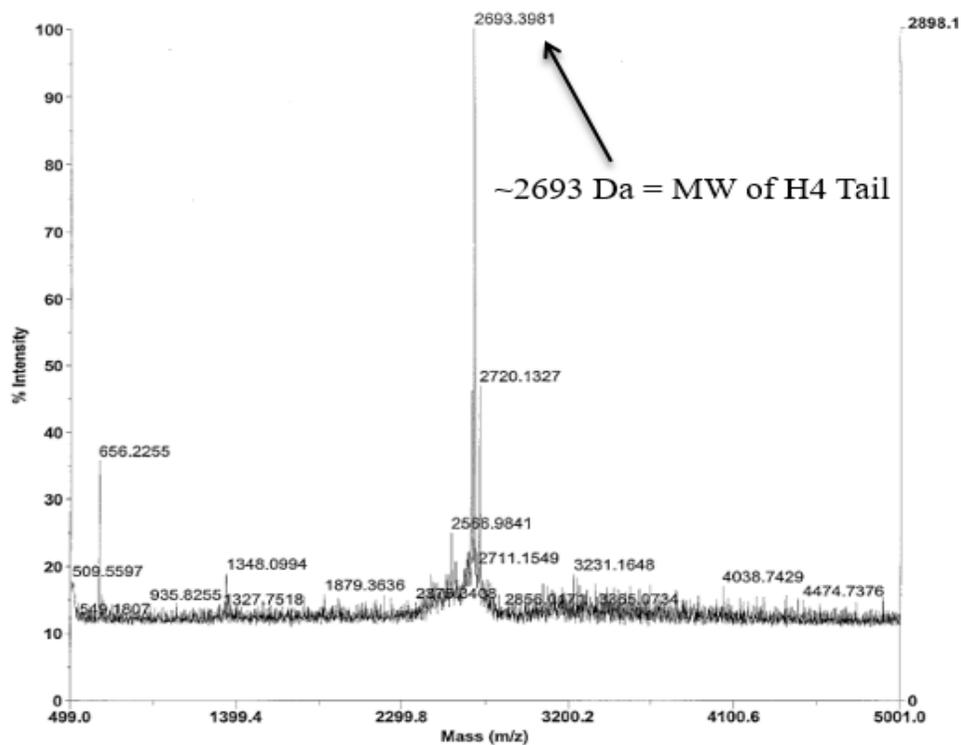


Figure 2.3 Mass Spectrometry Results for Purified H4 Tail

2.1.2 Preparing an NMR Sample of Histone H4 Tail

Since the H4 tail was lyophilized, and the protein is water soluble, preparing the sample for NMR analysis is relatively simple. A 50 mM citrate buffer containing 95% deuterated water, 5 mM DTT, and 5 mM EDTA was freshly prepared. 250 μ L of buffer dissolved 5mg of H4 tail, and the sample was transferred to a shigemi tube for analysis. Further NMR experiments were conducted on the sample by Adam Groves and Oleg Omikhail in Dr. Nikolai Skrynnikov's lab.

2.2 Preparing Mucin 1 for NMR Analysis

2.2.1 Developing a Mucin 1 C+TM Subunit Plasmid

Since it is known the mucin 1 C-subunit contains the CQC motif –which is known to dimerize with many cytoplasmic proteins- it is important to focus expression on a portion of Mucin 1 containing the cytoplasmic tail. Many of these interactions occur near the cell wall, so to fully attain proper folding and behavior of the C-subunit, the transmembrane segment must be included in this expression (See figure 2.4 below).^{23,24}

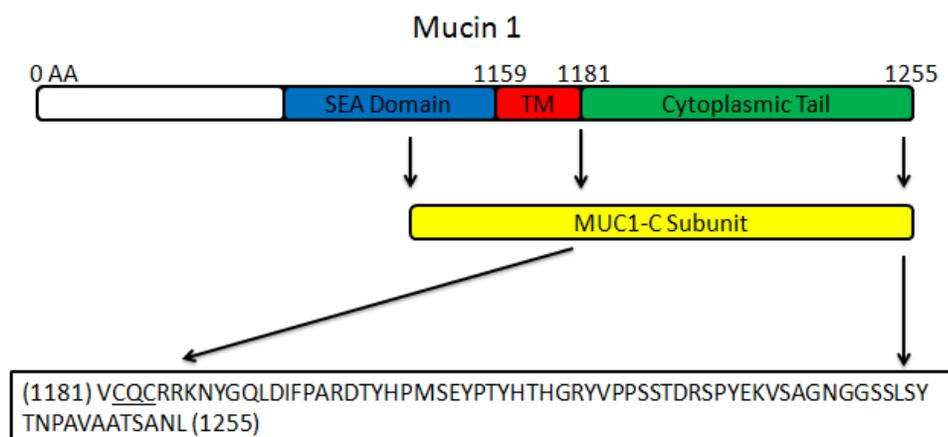


Figure 2.4 A General Map of Mucin 1 Sequence

The cytoplasmic portion of Mucin 1 (MUC1) is displayed in figure 2.4. On top of that genetic sequence, the transmembrane element (literature indicates the α -helix is ~23 amino acids) and an additional 9 amino acids (to ensure proper α -helical folding) on the N-terminal end were added to the intended protein sequence for expression. This provides the total sequence for expression:

QSGAGVPGW **GIALLVLCVLAIAIVYLIALA** **V** **CQC**RRKNYGQLDIFPARDTYH
PMSEYPTYHTHGRYVPPSSTDRSPYEKVSAGNGGSSLSYTNPAVAATSANL

Where the red region indicates the transmembrane segment, and the yellow region highlights the CQC motif. pET-28a was chosen as the vector, and the following restriction sites were chosen for inserting the MUC1-C+TM (cytoplasmic and transmembrane segment) sequence: Nde1 and Xho1. The intention behind this decision can be clearly seen in figure 2.5 below.

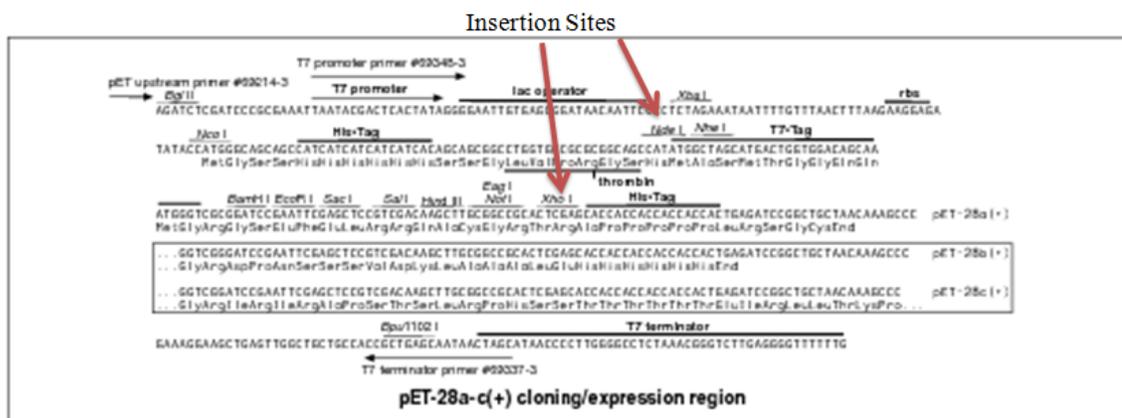


Figure 2.5 pET-28a Cloning Region and Intended Insertion Sites

This insertion allowed a His-tag and a thrombin cleavage site to attach to the N-terminal end of the protein. Adding a stop codon at the end of the sequence would prevent the addition of another His-tag at the C-terminal end. These specifications were given to GenScript, and their company constructed a plasmid for research. Once the

plasmid was obtained, and sequenced (to ensure the plasmid contains what was asked), it was prepped for expression of MUC1.

2.2.2 Expressing Mucin 1 C+TM

The plasmid was initially expressed in BI-21(DE3) ultracompetent cells. 1 μ L of a 0.001 mg/mL solution of plasmid was added to 10 μ L of BI-21(DE3) cells, and allowed to chill on ice for 30 mins. The sample was then heat shocked at 37° C for 30 seconds, then cooled on ice for 2 minutes. 900 μ L of LB media was added to the mixture, and incubated for a half hour at 37° C. The cells were then spun down, and 200 μ L were transferred to an ampicillin resistant plate. The rest of the expression protocol is exactly as described for the histone H4 tail in section 2.1.1.

After the media had been spun down, the pellet was re-suspended in a detergent buffer, since MUC1-C+TM is partially hydrophobic. The buffer contained 50mM Tris-HCl, 0.1M n-Dodecyl β -D-maltoside (DDM detergent), 200mM NaCl, and 5 mM EDTA. The cells were lysed via French press at 1100 psi in triplicate, then centrifuged down again. The supernatant was then injected into an Atka start FPLC with a 120 mL HiLoad 16/60 superdex 75 prep grade size exclusion column attached. The column was equilibrated with the same suspending buffer, and the sample was run through at 0.5 mL/min. The peaks were collected in 5mL fractions, and run on an SDS gel to determine the results.

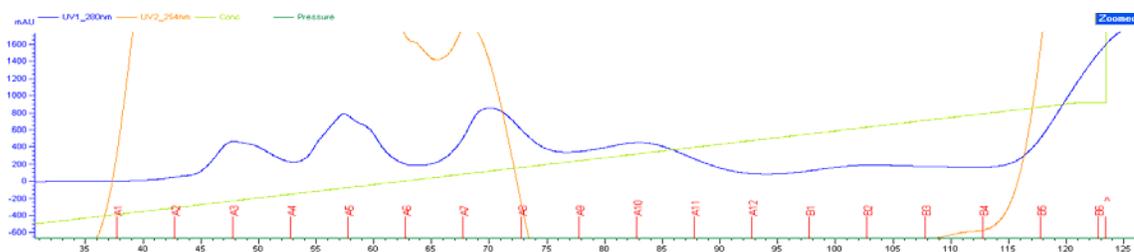


Figure 2.6 Size Exclusion Results for MUC1-C+TM

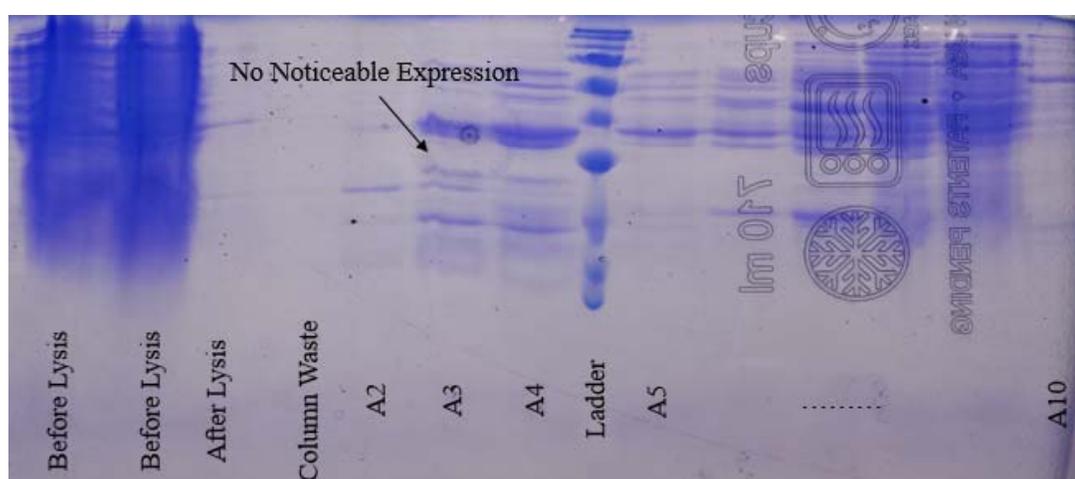


Figure 2.7 SDS PAGE gel of Size Exclusion Fractions Collected

These results show there was little to no significant expression of MUC1-C+TM. No band before or after lysis shows a strong band of protein around the molecular weight of Mucin (~26 kDa including the His-tag and thrombin cleavage site), and all elution absorbance peaks are relatively weak. This questions if MUC1-C+TM can really be expressed under these conditions.

Assuming MUC1-C crashed out of solution (DDM detergent may have been too weak), a new buffer was prepared containing 12M urea (as well as 50mM Tris-HCl, 200mM NaCl, and 5 mM EDTA) to encourage dissolution of all possible proteins within *E. Coli*. Two batches of cellular media were grown; one batch induced with IPTG

(mimicking the previous protocol) and one without IPTG induction. Before lysis, the both batches were dissolved in the urea buffer. When the solutions were spun down to remove any inclusion bodies, there was no pellet formation, indicating the entire cell had dissolved in solution. An SDS gel was run on the solutions (4x diluted to prevent urea interfering with SDS results). The results are noted in figure 2.8 below:

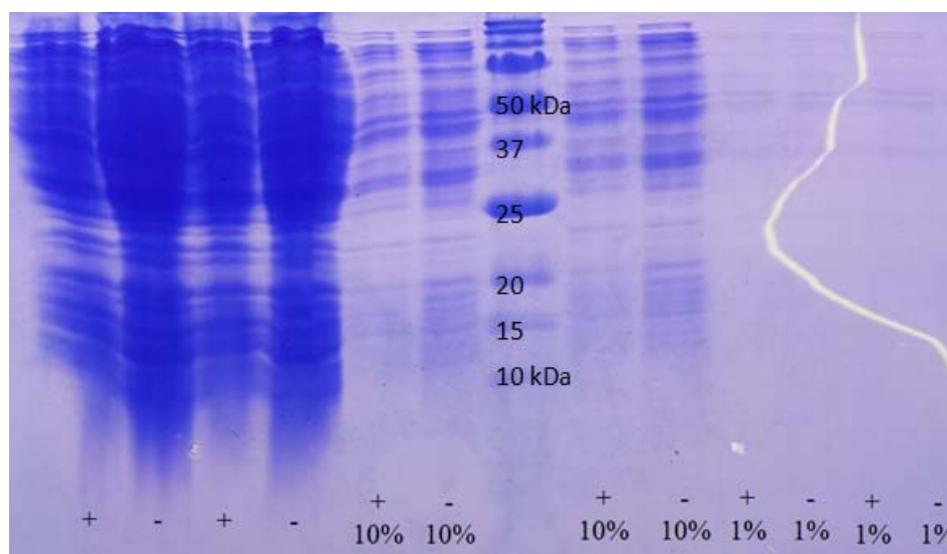


Figure 2.8 Mucin1 IPTG (+/-) SDS PAGE Results

Figure 2.8 shows the results for the two batches - IPTG (noted as +) and no IPTG (noted as -) - in various dilutions from their native concentration (described in percents). This IPTG +/- test revealed there is no expression of MUC1-C in the cell whatsoever. Both IPTG induced and non-induced bands are exactly the same, where if there was any expression at all, there would be a presence of a new strong band in the IPTG+ columns.

Since the protein cannot be expressed using BL21 cells, the same tests were conducted using ROSETTA(DE3) and C43(DE3) cells. During these experiments, careful attention was paid to ensuring optimal expression conditions. As seen in Figure 2.9 below, both strains failed to produce any Mucin1 protein. Since C43(DE3) is the most effective strain of *E. coli* in its ability to express transmembrane proteins, it was concluded the plasmid would not effectively express in a cell culture.

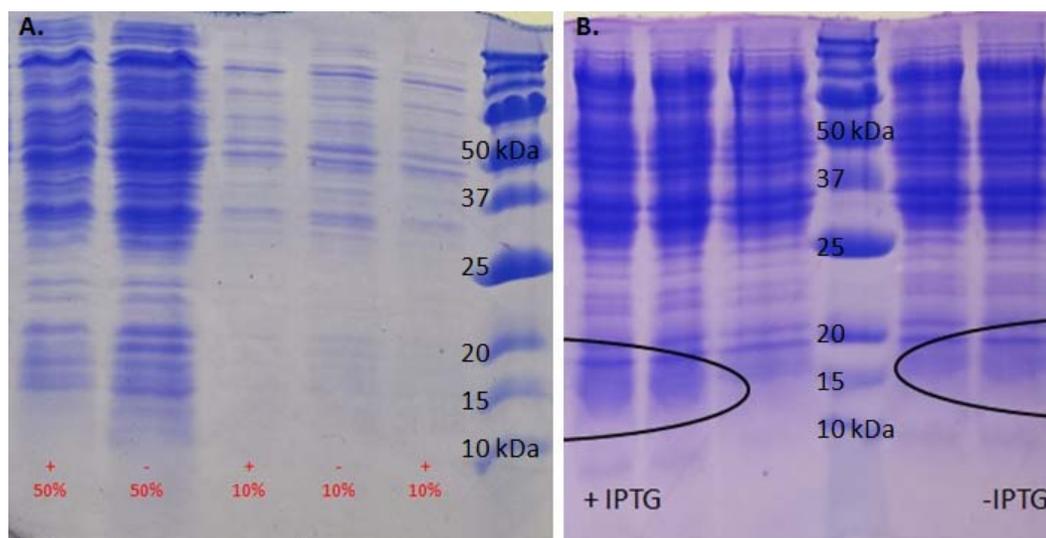


Figure 2.9 SDS PAGE results of A.) ROSETTA cells and B.) C43(DE3) Cells tested for Mucin1 Protein Expression

2.2.3 Revising the Mucin 1 Plasmid and Attempting Expression

A fellow colleague from St. Petersburg Russia, Yulia Pivovarova, suggested changing the vector used for expression. If the same protein sequence were used in a pET-32b vector, mucin1 could be fused to a thioredoxin protein found within the cloning region of the vector. Thioredoxin (Trx) is known to –when fused with another protein– encourage expression of normally inexpressible proteins.¹⁷ The decision was made to create another plasmid using this concept, and Trx was added to the N-terminus of the protein along with a thrombin cleavage site linking the two.

The plasmid was expressed in C43(DE3) cells using the same procedure discussed in section 2.1.1. The sample was then centrifuged, the pellet was collected, and re-suspended in an 8M Urea buffer. The cells were lysed, and re-centrifuged. After 30 minutes of centrifugation at 18000 rpm, no pellet formed; indicating all proteins were dissolved in the urea buffer. The same procedure was conducted without the addition of IPTG during expression, and both results were compared.

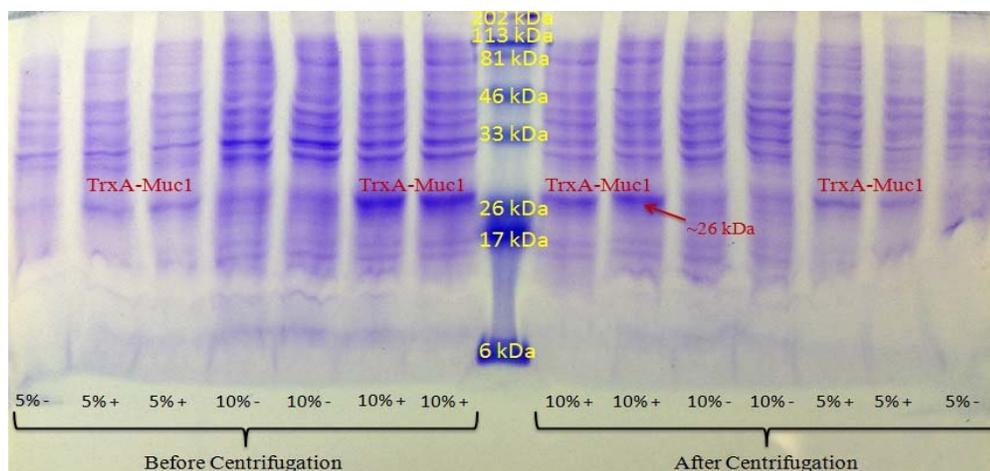


Figure 2.10 IPTG +/- Results for Trx-Mucin1 protein in 8M Urea

The results of the IPTG +/- test indicate absolute expression of the Trx-MUC1 fusion protein. The addition of thioredoxin actually encouraged expression of MUC1-C+TM, a normally inexpressible protein! Since High levels of urea can dissolve almost anything, it is important to test various detergents to see what the mucin construct can suspend in for future purifications and preparations for NMR study. Five detergents were added to the following buffer separately: 50mM Tris-HCl, 200 mM NaCl, 10 mM 2-mercaptoethanol, and 2mM EDTA. The five detergents were as follows: 8M Urea, 4% Triton X-100, 1% SDS, 2mM DPC, and 1.5% DDM. After Lysis, each cell culture suspended in the different detergent buffers were collected for an SDS-PAGE gel for comparison. The results are shown below. As a note, the DPC results were obtained on a separate gel, so the bands will not ideally align with the other bands. This does not affect the results though.

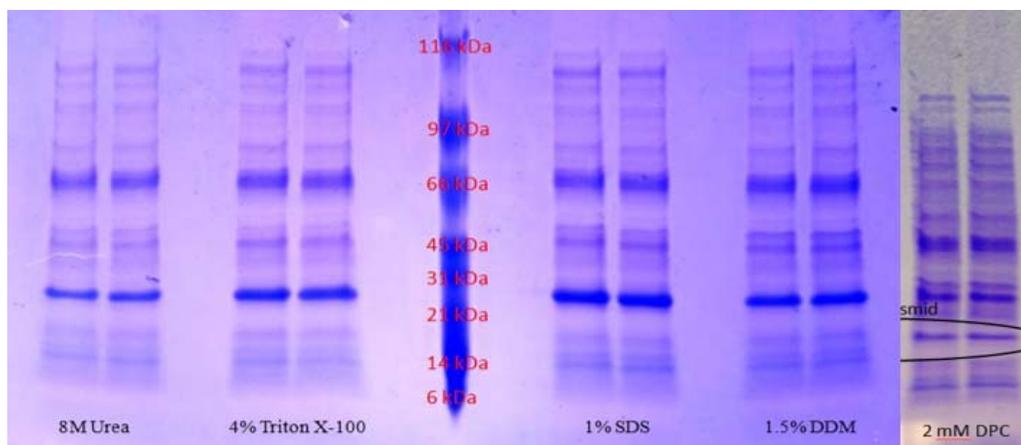


Figure 2.11 SDS-PAGE of Various Detergents Suspending Trx-MUC1 Protein

The results show Trx-MUC1 fusion protein is soluble in all detergents tested. This is great news, since options are available for future NMR studies.

Since Urea can be buffer exchanged to any detergent fairly easily, and 8M urea doesn't affect (to too much of an extent) Ni affinity or size exclusion columns at that concentration, the buffer of choice for purification was a urea buffer. After expressing Trx-MUC1 fusion protein, the sample was dissolved in 20 mL of buffer containing 50mM Tris-HCl, 8M urea, 200 mM NaCl, 10 mM 2-mercaptoethanol, and 2mM EDTA. The cells were lysed via French press at 1100 psi, and re-centrifuged for 30 minutes at 18000 rpm at 4° C. The supernatant was collected and introduced into the Atka start FPLC containing a 10 mL Ni affinity column. The flow rate was set to 3mL/min. After washing the column with 10 CV's of loading buffer, buffer B (containing 250mM imidazole) was introduced at a gradient. The elution fractions were collected in 5mL intervals, and associated UV peaks were run on an SDS-PAGE gel to find the fusion plasmid. The gel results are shown in figure 2.12 below. Note: Since 8M urea is too strong for SDS gels, the samples are diluted 4X. This is why the bands seem faint.

The results showed there was effective separation. The fractions showing presence of Trx-MUC1 were collected, buffer exchanged back to the initial suspending buffer (basically removing the imidazole), and concentrated to 5 milliliters. 300µL of

thrombin (5mg/mL stock) was added to cleave the thioredoxin, and incubated at 37°C for 3 hours.

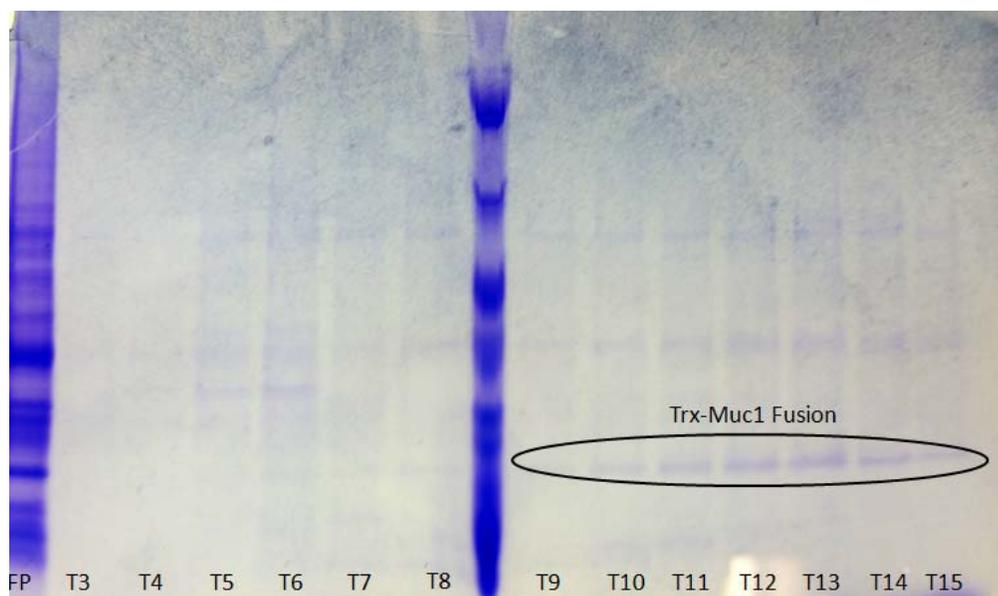


Figure 2.12 SDS-PAGE Gel of Ni Column Fractions Containing Trx-MUC1

After thrombin cleavage, the sample was injected back into the FPLC with a 120 mL HiLoad 16/60 superdex 75 prep grade size exclusion column attached. The column was equilibrated with the same urea buffer, and the flow rate was set to 0.5 mL/min. 5mL fractions were collected, and the results were examined via SDS-PAGE gels.

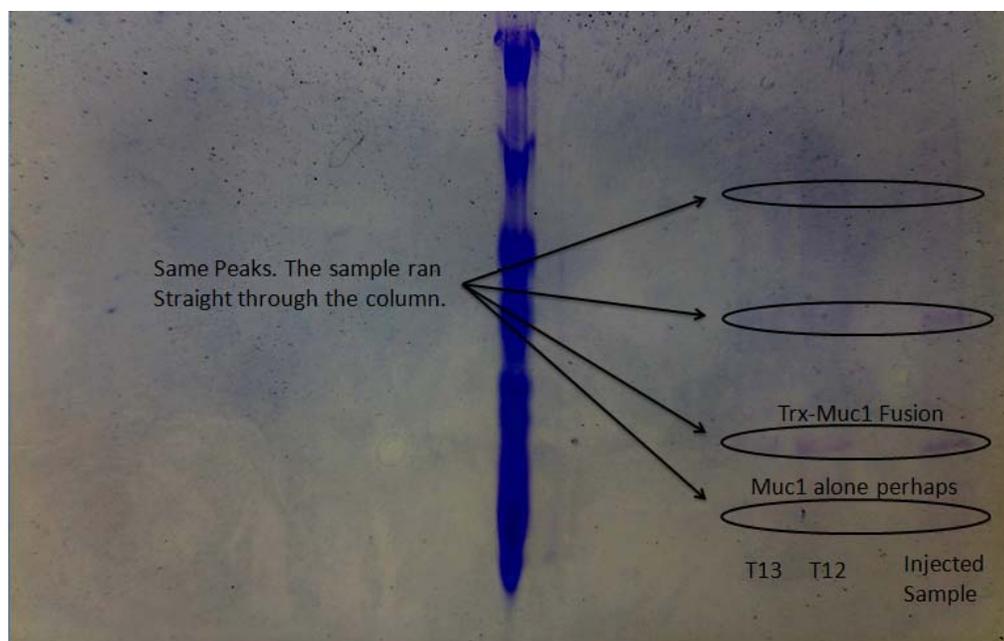


Figure 2.13 Size Exclusion Fractions Containing Trx-MUC1

The results are poor. The injected sample indicates some thioredoxin was cleaved, but not enough. The size exclusion column seemed to not effectively separate the proteins in the sample. Furthermore, the bands are quite faint, indicating the protein concentration is quite low. Overall, the protocol needs to be optimized to obtain a higher yield of MUC1-C+TM, and devise a way to obtain a more pure sample as well. On the positive side, it seems expression of mucin 1 was successful. For such a challenging protein, simply getting a product is quite an accomplishment.

CHAPTER 3: DISCUSSION AND FUTURE WORK

1.1 Histone H4 Tail

The purification of Histone H4 Tail was a success thanks to the addition of thioredoxin and a His-tag to the protein. The protocol was optimized to produce a large enough protein sample for NMR spectroscopy. NMR studies using existing and novel NMR pulse sequences were –and still are being- conducted in Dr. Skrynnikov's lab on the H4 tail. Future work on the histone H4 tail is being followed through by fellow colleague, Adam Groves.

1.2 Mucin 1

After much work, mucin 1 -a hydrophobic and disordered protein- was able to be expressed in a cell culture with the help of the fusion protein thioredoxin. The fusion protein allowed mucin 1 to be so soluble that even the most gentle detergent was able to suspend it in solution. Unfortunately, thrombin conditions and column purification have yet to be optimized, and therefore, a sample concentrated enough for NMR analysis has yet to be made.

The future work for this protein would first require optimizing thrombin conditions. Once accomplished, focus would need to be made on column purification, and ensuring effective separation is made on the sample. Once purification conditions are optimal, the sample can be buffer exchanged to either DPC or DDM for NMR analysis. Once mucin 1 is characterized via NMR in a detergent micelle, it can be buffer exchanged to a nanodisc (see section 1.3.2 for more information). These nanodiscs should mimic a cellular environment well enough that mucin 1 should fold

and behave as if it were in its natural environment. Once characterized in a nanodisc, binding studies can be conducted to observe interaction of mucin 1 with other cellular proteins.

1.3 Final Remarks

This research illuminates the versatility of protein expression via bacterial growth. Despite many obstacles, many eukaryotic proteins can be expressed in bacterial cells, and purified with ease for characterization. These methods are effective enough to produce a yield high enough for NMR studies. This recent methodology available to the scientific community has made protein NMR analysis more viable, and far more controlled than before. It has been a pleasure learning these methods, and building skills for future work in the field of protein NMR analysis. I hope this research helps readers with their own advances in research, and I am honored to publish this thesis for that very reason.

LIST OF REFERENCES

LIST OF REFERENCES

1. Nelson, D; Cox, M. Lehninger: Principles of Biochemistry Fifth Ed. New York, NY. W.H. Freeman and co. p. 708-720.
2. Leung, D. *et. al.* Growth Hormone Receptor and Serum Binding Protein: Purification, Cloning, and Expression. *Nature* 330:537-543. 1987.
3. Yu, T; Raschle, T; Hiller, S; Wagner, G. Solution NMR Spectroscopic Characterization of Human VDAC-2 in Detergent Micelles and Lipid Bilayer Nanodiscs. *Biochimia et Biophysica* 1818:1562-1569. 2012.
4. Newstead, S. *et. al.* High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in *Saccharomyces cerevisiae*. *PNAS* Vol. 104:35. 2007.
5. Terpe, K. Overview of Tag Protein Fusions: From Molecular and Biochemical Fundamentals to Commercial Systems. *Appl. Microbiol. Biotechnol.* 60:523-533. 2002.
6. Davis, G.; Elisee, C. *et. al.* New Fusion Protein Systems Designed to Give Soluble Expression in *Escherichia Coli*. *Biotech.* Vol. 65: 382-388. 1999.
7. Bok, J.; Keller, N. Fast and Easy Method for Construction of Plasmid Vectors Using Modified Quick-Change Mutagenesis. *Mol. Bio.* 944:163-174. 2012.
8. Yadav, M. Biotechnology. Second Ed. New Dehli, IN. Discovery Publishing House. P. 143.
9. Gupta, P. Elements of Biotechnology. First Ed. Meerut. Rastogi Publications. P.15-18.
10. Information on pET-28a-c, pET-32a-c, BL21, and ROSETTA Retrieved from 2016 Novagen Catalog.

11. Miroux, B.; Walker, J. Over-Production of Proteins in *Escherichia coli*: Mutant Hosts that allow Synthesis of Some Membrane Proteins and Globular Proteins at High Levels. *J. Mol. Biol.* 260:289-298. 1996.
12. Leung, D. *et al.* Growth Hormone Receptor and Serum Binding Protein: Purification, Cloning, and Expression. *Nature* 330:537-543. 1987.
13. Yu, T; Raschle, T; Hiller, S; Wagner, G. Solution NMR Spectroscopic Characterization of Human VDAC-2 in Detergent Micelles and Lipid Bilayer Nanodiscs. *Biochimica et Biophysica* 1818:1562-1569. 2012.
14. Lebediker, M. Thrombin Cleavage of GST-Fusion Protein. The Protein Purification Facility. wolfson.huji.ac.il Accessed 3/12/16.
15. Page, R. *et al.* Comprehensive Evaluation of Solution Nuclear Magnetic Resonance Spectroscopy Sample Preparation for Helical Integral Membrane Proteins. *J. Struct. Func. Genom.* 7:51-64. 2006.
16. Tian, C. *et al.* Membrane Protein Separation for TROSY NMR Screening. *Methods In Enzymology* 394:321-334. 2005.
17. LaVallie, E. *et al.* A Thoredoxin Gene Fusion Expression System that Circumvents Inclusion Body Formation in the *E. coli* Cytoplasm. *Nature: Biotechnology* 11:187-193. 1993.
18. Borch, J; Hamann, T. The Nanodisc: A Novel Tool for Membrane Protein Studies. *Biol. Chem.* 390:805-814. 2009.
19. Ritchie, T; Silgar, S; *et al.* Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. *Methods in Enzymology* Vol. 464. 2009.
20. Klammt, C. *et al.* High Level Cell-Free Expression and Specific Labeling of Integral Membrane Proteins. *FEBS Journal* 271:568-580. 2004.
21. Tsyr-Yan Yu, *et al.* Solution NMR spectroscopic characterization of human VDAC-2 in detergent micelles and lipid bilayer nanodiscs *Biochimica et Biophysica Acta.* 1818:1562–1569. 2012.
22. Pepenella, S.; Murphy, K.; Hayes, J. A Distinct Switch in Interactions of the Histone H4 Tail Domain upon Salt-Dependent Folding of Nucleosome Arrays. *J. Biol. Chem.* 114.595140. 2014.

23. Kufe, D. Mucins in Cancer: function, prognosis, and therapy. *Nature* vol. 9, 874-885, Dec. 2009.
24. Ahmad, R.; Rajabi, H.; *et.al.* MUC1-C Oncoprotein Promotes STAT3 Activation in an Autoinductive Regulatory Loop. *Science Signaling* vol. 4, 160, Feb. 2011.
25. Raina, D.; Kufe, D.; *et. al.* Targeting Cysteine-mediated dimerization of the MUC1-C oncoprotein in human cancer cells. *Journal of Oncology* 40: 1643-1649, 2012.
26. Radivojac, P.; Iakoucheva, L.; *et.al.* Intrinsic Disorder and Functional Proteomics *Biophysical J.* vol. 92, 1439-1456, Mar. 2007.