Membrane Protein Expression in *Mycobacterium tuberculosis*

Jason Crowe
National High Magnetic Field Laboratory

Abstract

Mycobacterium tuberculosis is a bacterial disease that kills millions worldwide every year. Recently, many strains have developed a resistance against one or all of the drugs being used to treat the disease. Now, research is being turned to find new ways to combat the disease. One major target has become the membrane proteins. By studying and analyzing the membrane proteins of *Mycobacterium tuberculosis*, it is possible that a new treatment may be found. At the very least, the Protein Data Bank will receive many new structures of membrane proteins to add to its database. This new information will be able to be applied throughout the biological and medical sciences, and will help further the study of cellular processes.

Introduction

Mtb, more commonly known as Tuberculosis, is the leading cause of infectious mortality throughout the world. Three million people annually die from the infection. Due to the fact that tuberculosis is extremely resistant to many clinical forms of treatment, the sequencing of its genome was done to hopefully provide important biological information about the bacterium, which could be used to combat the disease. Upon completion of the sequencing in 1998 (*Nature*, 1998), it was almost immediately noticed that a large part of the genome was devoted to coding for enzymes. Many of these enzymes are involved in lipogenesis and lipolysis, and therefore play a vital role in the interaction of the bacterium with the cellular membrane of the target cell. Also, since the cellular membranes are involved, membrane proteins are critical in the infection of the target cell. Studying and analyzing the membrane proteins of *Mycobacterium tuberculosis* can lead to the discovery of important information regarding its methods and mechanisms of infection. This information could provide new information about how to fight the infection. (NIH proposal)

Membrane proteins have many different structures and functions, such as ion and nutrient channels and transporters, intracellular signaling, cellular receptors, enzymes, and a variety of other functions. All of these functions mean that membrane proteins are essential for communication between the cell and the extracellular space. Membrane proteins differ from other proteins in that there are very few structures known, and very little in general is known about them. While close to a third of each genome is devoted to coding membrane proteins, less than 1 percent of known protein structures are membrane proteins. And since membrane proteins carry out these critical functions, drugs aimed at

affecting membrane proteins could have a profound effect on the way a cell functions. This research can lead to a breakthrough in the way diseases such as tuberculosis are treated. (NHMFL website)

Heading up this research is the *Mycobacterium tuberculosis* membrane protein pilot project. The goals of the project are to observe, classify, and analyze the membrane proteins of tuberculosis. Many of these proteins contain or are a part of virulence factors of tuberculosis. The characterization of these proteins will done using a variety of methods. Cloning of the tuberculosis genes, transformation of cloned inserts into E. Coli vectors, Polymerase Chain Reactions, and purification is being used to express the membrane proteins in a vector. Mass spectroscopy, solid-state and solution Nuclear Magnetic Resonance with a focus on ion cyclotron resonance, and X-ray diffraction are being used to determine the structure of membrane proteins after they are expressed. Successful completion of the membrane protein pilot project will triple the number of membrane protein structures in the Protein Data Bank. (NIH proposal)

Protein Expression

When the appropriate gene is selected to be expressed, it is introduced into the *E. Coli* cells (BL(DE3) Codon Plus-RP strain) via transformation. The pET-29b plasmid vector system is used for protein expression. (Novagen, Inc.) The copy of the gene was cloned into the plasmid DNA in place of the Nde-1 -Xho-1 region of the plasmid through a ligation reaction. The gene is under the control of the T7 promoter. Cells do not produce protein without induction, because the promoter is normally blocked by a repressor protein. Induction with IPTG (isopropylthio-β- D-galactoside) removes the repressor protein from the promoter region. The cells provide the necessary polymerase,

T7 RNA polymerase, for expression of the protein, but the polymerase is not active until induction with IPTG. After induction, the T7 RNA polymerase becomes extremely effective, and extremely selective for the inserted gene. The majority of the cell's resources become devoted to expressing the targeted gene. Also, within a few hours, the desired protein can compromise over half of the total cell protein content.

There is another important feature of the expression system. The vector carries a 6-histidine residue tag at the C-terminus of the gene, making it easy to track and purify the protein.

Materials and Methods

LB media was used for growing the bacterial cells containing the DNA insert. LB media is made by adding 10 grams of sodium chloride, 10 grams of bacto-peptone, and 5 grams of yeast extract to 800 mLs of deionized water. The solution is mixed, and the pH is adjusted to 8.0 by adding a sodium hydroxide solution. After adjustment of the pH, more deionized water is added until the total volume is one liter. The media is then autoclaved to remove any possible contaminants.

To check for the presence of the gene insert inside the plasmid, a DNA agarose gel electrophoresis is run. To make the agarose gel, 0.5 grams of high melting agarose is added to 50 mLs of deionized water. This forms a 1% agarose gel. The solution is then heated for 1.5 minutes to fully dissolve the agarose. When the agarose liquid starts to cool, but before it solidifies, 0.5 microgram/mL of Ethidium Bromide (EtBr) was added to the solution. The gel was then poured; the comb set, and allowed to solidify. Finally, after the gel has solidified, buffer is added and the comb is removed. The sample is

loaded into the gel, and the DNA is identified under UV light. This is due to the addition of EtBr, which binds to DNA and fluoresces.

To check for the presence of protein in a cell culture, an SDS-PAGE polyacrylamide gel is run. The gel is run in a tricine buffer solution. To make the gel, two separate gel solutions are necessary, a stacking gel to line up the protein fragments, and a resolving gel to separate the different proteins/fragments. To make a resolving gel, 1.2 mLs of 49.5% AA/Bis (Acrylamide/N1N methylbisacrylamide) (Novagen, Inc.) buffer, 2 mLs of pH 8.3 Tris buffer, 2 mLs of 56:44 glycerol:H2O, and 0.8 mLs of water were added to a small beaker. Just prior to casting the gel, 60 uL(10 uL/mL) of APS (ammonium persulfate) and 12 uL (2 uL/mLs) of TEMED are added to the solution. The gel is then cast immediately and allowed to solidify. To make the stacking gel, add 250 uL of 49.5% AA/Bis buffer, 250 uL of pH 8.3 Tris buffer, and 2 mLs of water to a small beaker. Again, add the appropriate amount of APS (10 uL/mL, so 25 uL) and (2 uL/mLs, so 5 uL) of TEMED. Then immediately cast the stacking gel on top of the solidified resolving gel.

For the protein purification, many buffers are needed. The chart below contains the instructions on how to create one liter of each required buffer. Tris pH 7.9 buffer is made by combining Tris base with Tris-HCl, in varying amounts depending on the pH desired and the desired temperature of the solution. 0.5 molar EDTA is made by adding 186.1 grams of EDTA to water, and adjusting the pH to 8.0. The pH must be adjusted as all of the EDTA will not go into solution until the appropriate pH is reached.

Buffer	Tris pH 7.9 (mL)	NaCl (g)	Imidazole (g)	0.5 M EDTA (mL)
8X Binding	160	233.76	2.75	-
8X Wash	160	233.76	33	-
4X Elution	80	116.88	275.2	-
2X Strip	40	58.44	-	500

After adding each of the ingredients, deionized water is added until the final volume of the solution is one liter. The exact concentrations of each chemical added can be found in the Novagen Catalog. Finally, to make the 1X charge buffer to charge the resin column, a 0.1 M solution of NiSO4 was made. This is done by adding 5.26 grams of NiSO4 to 200 mLs of water.

Methods

Propagation

When the desired protein to be analyzed is selected, it first must be expressed. To start off, the selected gene is transformed into a cloning vector plasmid, where it will be propagated. The cells must be cultured. They are grown on LB agar plates overnight. The following day, they are removed from the plates, and placed in LB media which had been inoculated with kanamycin, and are allowed to culture overnight. After the vectors containing the plasmids are grown, the plasmids must be check to see if the insert is actually in the plasmid. This is done by running a DNA agarose gel. If the plasmid is present, it must be purified.

Plasmid purification is a multi step process. A QIAprep Spin Miniprep kit made by QIAgen is used. The kit contains all of the appropriate buffers. First, the bacterial cells are spun down for 10 minutes at 10,000 rpm. They are resuspended in 250 uL of buffer P1 with RNAse. Then add 250 uL of buffer P2, and invert the tube gently to mix. Add 350 uL of buffer P3 and once again invert the tube to mix. Centrifuge the tubes for 10 minutes at 12,000 rpm. Take off the resulting supernatant and add it to the filter tubes given in the kit. Centrifuge for 1 minute. Discard the flow through and add 500 uL of

buffer PB, and centrifuge for 1 minute. Discard the flow through again. Add 750 uL of buffer PE and centrifuge for 1 minute. Discard the flow through and then centrifuge again for 1 minute. Finally, place the filter tube in a clean microcentrifuge tube, and add 50 uL of elution buffer. Let it stand for 1 minute and centrifuge for 2 minutes to recover purified DNA.

Transformation

Once the plasmids are purified, it is wise to check again to see if there is DNA present by running an agarose gel. If the plasmid is present and purified, PCR can be used to amplify the amount of DNA in the sample. One a pure plasmid is recovered, it is then transformed into the expression vector, *E. Coli*. To transform the plasmids, first thaw out the expression cells in an ice bath, and make sure they are resuspended in the solution before continuing. Add thawed cells to a microcentrifuge tube (25-50 uL of cells). Add 2 uL of ligation reaction (plasmid DNA) directly to the cells. Place the tubes on ice for 5 minutes. Removes the tubes and heat shock them by placing them in a 42 degree Celsius water bath for 30 seconds. Replace the tubes on ice for 2 minutes. Then add 250 uL of SOC medium to each tube. Allow the tubes to incubate on a 37-degree shaker at 250 rpm for 30 minutes. After incubation, spread the reaction on LB agar plates (with kanamycin), and allow the plates to incubate overnight at 37 degrees. The colonies can then be removed from the plates and placed in LB media for storage.

Protein Expression

Once the transformation reaction is completed, expression screening is done. A small-scale expression is done first to see if the particular protein is expressed, and if it is, a large-scale expression can be done to provide a stock of cells with the expressed

protein. To start, LB media is inoculated with kanamycin (1 uL/mL). The expression cells are then added to the LB media (500 uL/10 mLs). The cells are then placed on a 37-degree shaker at 250 rpm and allowed to grow. The cells are monitored until the observable density at 600 nanometers was between 0.4 and 0.6. When that point is reached, usually after 2-3 hours, the solutions are inoculated with IPTG. This is done to induce the cells to express the protein, as explained above. The cells are then allowed to grow for 3 hours on the 37-degree shaker. After 3 hours, the cells are harvested. They are collected and centrifuged for 10 minutes at 10,000 rpm. The supernatant is discarded and the cells are resuspended in water. The cells are then sonicated to break them apart. Sonication is done 5 times for 20 seconds per run, with a waiting period of 20 seconds between each run. After sonication, the cells are centrifuged for 10 minutes at 10,000 rpm. The supernatant is collected and placed in a new tube. This contains the soluble portion of the cells contents. The pellet, which contains the insoluble fraction, is resuspended in 100 uL of 2% SDS. When this is complete, the samples are loaded and an SDS-Page electrophoresis gel is run. The gel will show whether not the protein is expressed, and in what fraction the protein is located. If the gel confirms that a protein is present, the protein is ready to be purified. Also, a large-scale expression can be done to create a stock of cells. If follows the same procedure as above, except a liter of cells are grown as opposed to only 10 mLs.

Protein Purification

Before any protein purification can be done, all of the above-described buffers must be prepared. Protein purification can be done through either an affinity binding column, or an ion exchange column. To prepare the affinity column, resin is poured into

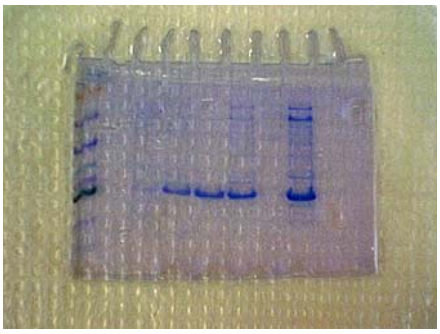
the column and allowed to settle. The water is also allowed to drip out of the column. As soon as the water has dripped, charge buffer is run through the column to charge it. This must be done immediately so the column does not dry out. While the charge buffer is passing through the column, the sonicated cells are frozen, thawed, re-frozen, thawed and sonicated again. The cells are then centrifuged for 10 minutes at 10,000 rpm to separate soluble from insoluble. The supernatant, containing the soluble fraction, is then run through the column as soon as the entire amount of charge buffer has passed into the column. A clean tube is used to collect whatever solution passes through the column. Upon completion, binding buffer is passed through the column. This is collected in another clean tube. After binding buffer, wash buffer is passed through the column and collected in a clean tube. When the wash buffer is complete, the elution buffer is passed through and collected in a clean tube. Finally, strip buffer is passed through the column and collected in a clean tube. All of the collected fractions are then run on an SDS-page electrophoresis gel to determine which fraction contains the protein. When the protein is located, it then must be run through dialysis to remove the imidazole. The proper fraction is placed in a dialysis tube and allowed to sit for several hours in 500 mLs of 1X binding buffer. To further remove imidazole, the dialysis tube can be transferred to a beaker of water, and allowed to sit for another few hours, or overnight. The above purification process is then repeated with the exception that the dialysis sample is run through the column instead of the supernatant. Another SDS-page gel is run to determine the purity of the protein. If the protein is not completely pure, then an ion exchange column can be run to further purify the protein. If the protein is pure, then it is ready for analysis.

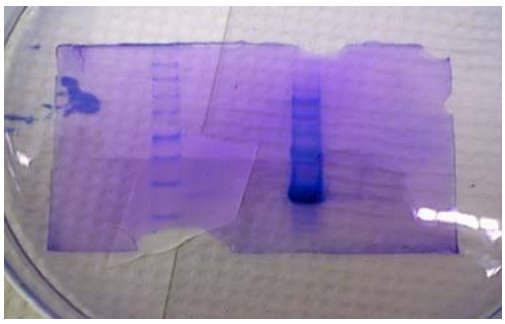
To run the ion exchange column, the fraction with the protein is once again dialyzed in water. The ion exchange column is set up in a similar way to the affinity column. The sample is run through the column, followed by a solution of 20 mM Tris-HCl and 500 mM NaCl, and all of the fractions are collected in separate tubes. The fractions are once again run on an SDS-page gel to confirm purity of the protein. The protein is now pure, and ready for analysis.

Results

Gene Name	DNA size (kb)	Protein M.W. (Da)	pET 29b plasmid	Transformation	Tried	Expressed	Large Scale
Rv1171	438	15168.6	+	+	+	+	+
Rv1342c	360	13369.8	+	+	+	+	+
Rv0011c	279	10418.2	+	+	+	+	+
Rv0514	297	10273.6	+	+	+	-	-
Rv0870c	387	13673.6	+	+	+	-	-
Rv3857	195	6748.47	+	+	-	-	-
Rv3789	363	13361.8	+	-	-	-	-
Rv2169c	402	14556.7	+	+	+	+	+
Rv2076	249	9029.31	+	+	-	-	-
Rv0460c	237	8063.45	+	+	-	-	-
Rv1305	243	8039	+	-	-	-	-
Rv0882	282	9612.04	+	-	-	-	-

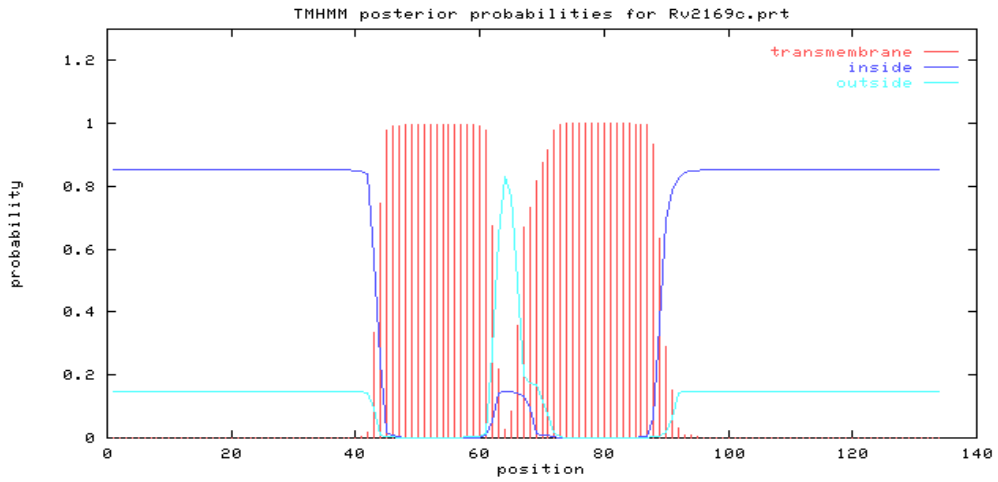
The chart on the preceding page shows a list of all the genes that where expression was tried. Out of the 9 proteins that were successfully transformed into expression cells, only 4 of them were actually expressed. Genes Rv1171, Rv1342c, Rv0011c, and Rv2169c were expressed and subsequently produced on a large scale for storage as stock. The remaining genes were not expressed in the expression cells.





The above SDS-PAGE gels show the results of the protein purification of the protein encoded by gene Rv2169c using a nickel-binding column. The bands show the presence of the protein on gel 1, but also the presence of other proteins, meaning that the purification procedure was partially successful. Also seen on gel 1 are bands in multiple lanes. The protein was supposed to elute in only a single band, but it was eluted from the column before it was expected to. Gel 2 shows the protein after ion exchange column protein purification. The band is light, meaning that the concentration was too low, but no other bands appear to be there meaning that the protein is relatively pure in that fraction.

Rv2169c is an unnamed membrane protein that is composed of 134 amino acids. The function of the protein is also unknown. Its molecular weight is14556.7 Daltons, and it has an isoelectric point of 11.7109. Based on computational analysis by TMHMM, as shown in the figure below, there appears to be two transmembrane regions within the protein. (TMHMM website)



Discussion

As seen in the results, it is difficult to express membrane proteins. There are many reasons why these genes could not be transformed or expressed. The systems used, pET29b vector and *E. Coli* cells were good systems to use for the genes that did express, but they may not have been optimal for the other genes of *Mycobacterium tuberculosis*. Other genes may require different cloning and expression vectors other than pET29b, or a different host cell to properly transform and express.

There are also many other biochemical reason of why the membrane proteins were not expressed. Membrane proteins are extremely hydrophobic, so expressing and producing then in any system would prove to be difficult. Since they do not like to be in contact with water, membrane proteins can only migrate to specific locations within a cell. They require a lipid environment, and there are not many cellular places with high lipid concentrations. This is why the production of membranes proteins is difficult for any cell. Simply put, cells do not like to produce membrane proteins normally, and in expression they are being forced to produce them. Also, another reason why expression may not have worked is aggregation of the proteins. When forcing a cell to produce foreign proteins, the proteins will not necessarily just go into the membrane where they are supposed to. The cell will try to rid itself of the foreign protein. It does this through the process of aggregation. The cell clumps the hydrophobic proteins together into a big lump called an inclusion body. Instead of nicely expressed proteins, there is a chunk of protein floating within the cell, which makes for tough expression.

Finally, the results from the experiments always have something in common. Whenever an SDS-PAGE gel was run with soluble and insoluble fractions, there are almost always protein bands in the insoluble fraction. This shows that there is some part

of protein that is not soluble and does not remain in solution. This may be an inclusion body, but it means that the protein will not be able to purified very well, as many other cellular parts are contained the insoluble fragments.

Acknowledgements

I would like to thank Dr. Philip Gao, my mentor, Dr. Alla Korepanova, who served as an extremely great resource in my intellectual journey this summer. I would also like to thank Yuanzi Hua, Yiming Mo, my laboratory partner April Jue, and the rest of the Cross lab staff for their help with the various experiments.

References

Fritsch, E F., T Maniatis, and J Sambrook. Novagen Molecular Cloning. 2nd ed. Plainview, NY: Cold Spring Harbor Laboratory P, 1989.

http://magnet.fsu.edu/~changlin/MBPweb/ (NHMFL NMR lab website)

http://magnet.fsu.edu/~changlin/MBPweb/proposal/proposalcover.htm (NIH grant proposal)

http://genolist.pasteur.fr/TubercuList/tmhmm/Rv2169c.prt.html (TMHMM website)