Development and Synthesis of Utrophin Actin Binding Domain 1 (ABD1)

Kate McMahon
Advised by Anne Hinderliter
University of Minnesota Duluth
April 2016

Abstract:
Duchenne Muscular Dystrophy (DMD) is an X-linked genetic disease containing point mutations in the muscle protein Dystrophin causing the protein to lose its function. Specifically, Dystrophin is critical for dissipating the mechanical stress placed on muscles during physical activity. Although Dystrophin is nonfunctional in DMD patients, its fetal homolog, Utrophin, is often present in higher amounts than common to adult cells. Because Utrophin and Dystrophin share 85% homology in their first actin binding domains (ABD1), the interrelatedness of structure and function validate Utrophin as a proposed therapeutic tool for combating DMD. To test this hypothesis, the thermodynamic character of Utrophin ABD1 and Dystrophin ABD1 will be compared. As Utrophin is not regularly studied, the gene for Utrophin ABD1 was designed, synthesized, and expressed in *E.coli* cells. Prokaryotic cells were utilized to express a eukaryotic protein because of rapid growth rate and the presence of an extra, self-replicating, circular DNA called a plasmid. A plasmid is evolutionarily advantageous because it can be passed quickly from prokaryotic cell to prokaryotic cell without the entire genome replicating, thus increasing variability. This unique attribute was utilized to express Utrophin ABD1 in *E. coli* cells. Although eukaryotic systems often have posttranslational modifications, this did not pose a threat for the prokaryotic cell amplification. The gene encoding the protein was designed using specific amino acid residues, not nucleotide sequences; the splicing of nucleotide sequences was irrelevant as posttranslational modification occurs before the amino acids are assembled into their primary structure. Specifically, Utrophin ABD1 was designed with BamHI and XhoI restriction enzymes flanking the 246 amino acid Utrophin ABD1 construct which was synthesized in a pUC57 *E. coli* plasmid. Using BamHI and XhoI, the amino acid sequence was restriction digested and subcloned into an expression vector containing components critical for nickel column chromatography like a histidine tag, TEV protease cut site, and maltose binding protein. The expression vector also contains a selective marker to find the correct ligated species such as the antibiotic Kanamycin. These plasmids were transformed into competent *E. coli* cells so the *E. coli* cells would replicate the inserted DNA the same way it replicates a plasmid. During the rapid growth, inclusion bodies, protein aggregates of overexpressed protein, are accounted for by the addition of the maltose binding protein which maintains solubility. The transformed cells were stored in a glycerol stock. Synthesis of this gene then allows growth and purification of the Utrophin ABD1 protein in a similar manner to those already classified for histidine tagged proteins. Purification is carried out at a pH of 8 so that the six histidines will be deprotonated and bind to the nickel column, thus washing out all other protein expect for the Utrophin bound to the column. Purification is important in that it insures pure protein by cleaving off the maltose binding protein using the tobacco etch virus (TEV) protease that recognizes a specific nucleotide sequence rarely found in the eukaryotic genome. Finally, thermodynamic analysis of this protein will give insight into the structure and function of Utrophin ABD1 and its potential capabilities as a therapeutic agent for patients with DMD.

Introduction

Muscular Dystrophy as a debilitative disease. Duchene Muscular Dystrophy (DMD) is an X-linked genetic disease that causes mutations to the muscle protein Dystrophin. Dystrophin is critical for dissipating the mechanical stress placed on muscles during physical activity. It has the ability to do this because of its large size- approximately 427kDa. The size of the protein allows it to take the brunt of the mechanical force the muscle cells undergo and the high number of repeating subunits allows a higher number of conformational changes. It is this transformation in conformations that dissipates the force. DMD is characterized by the complete lack of function or presence of Dystrophin. Although, Dystrophin is absent in DMD patients, its fetal homolog Utrophin is present in higher amounts than common to adult cells (Guiraud et al., 2015). As a result, we hypothesize that Utrophin is extremely important to people with DMD. In the early stages of life, DMD is characterized muscle fatigue that eventually leads to deterioration. As the patients age into their teens, death can be the result of either the lack of cardiac muscle contractions or damage to the diaphragm, which results in respiratory distress. Of the effected population, most are male as they only have one X chromosome and nothing to compensate for the disease. At this time, there is no way to treat this disease that continues to affect many.

Utrophin as a therapeutic agent. Utrophin and Dystrophin share 85% homology in their first actin binding domains (ABD1). As a result, it is proposed that they will have similar functions because oftentimes, structure and function are related. Utrophin is also considered a large protein at 394kDa. Additionally, both Dystrophin

and Utrophin have distinct energetic profiles that are a result of their chemical makeup. However, because of the relative low energy barriers, these subunits can be accessed easily, which is important to dissipating mechanical energy (Prochniewicz, et. al., 2009). Currently, the target of gene therapy is to create "mini-Dystrophin" subunits which can be used where there is no Dystrophin in DMD patients as Dystrophin itself is too large to utilize for gene therapy (Lin, et. Al., 2012). My project is to purify one of the ABD1 domains of Utrophin and characterize it thermodynamically as we have already done that with Dystrophin to test the hypothesis that is the basis of a NIH proposal in February. In order to test the validity of this hypothesis, the thermodynamics of Utrophin and Dystrophin must be compared as it will give insight into their function.

Proposed Experiment

This experiment will utilize molecular biology techniques to develop and synthesize the first Utrophin ABD1 gene in the Hinderliter lab. This functions to provide a Utrophin protein that can be parameterized macroscopically and microscopically. This information will then be available for further insight on why Utrophin does not function as an effective force dissipater in the muscle cell. Changes in energetics will indicate changes in structure that is inherently related to function.

The objectives for this experiment are as follows:

1. Develop the Utrophin ABD1 gene.

2. Insert the gene into an expression vector.

3. Transform the gene into competent E. coli cells.

4. Purify Utrophin ABD1 using nickel affinity chromatography.

Based on the thermodynamic and simulated data gathered on Dystrophin ABD1, the same aspects of Utrophin ABD1 can be hypothesized. Specifically, it is hypothesized that the purification process of Utrophin ABD1 will be similar to Dytrophin ABD1 because of their 85% homology, and that there will be difficulties with maintaining the solubility of the protein after the TEV protease is utilized. The proper purification of this protein will be critical for thermodynamic analysis of the actin binding domain of Utrophin so it can be compared with Dystrophin. It is hypothesized that Utrophin will display distinct thermodynamics from Dystrophin, but that the structural configuration of Utrophin can be manipulated in order to better mimic Dystrophin. This work is critical for the collaboration between DD Thomas at UMTC who focuses on structural dynamics and A. Cembran at UMD who focuses on molecular dynamic simulations. This work is the crucial link between the two labs as it provides the energetic basis for the differences between Utrophin and Dystrophin which is required for both groups.

Methods

Utrophin ABD1 will be designed using its already classified amino acid structure provided by Universal Protein Resource that contains a database of known proteins from every species. This sequence will be codon optimized for the E. coli cell and inserted into the pUC57 plasmid using restriction enzymes. Polymerase chain

reaction (PCR) will be conducted by an outside source to amplify the Utrophin ABD1 gene segment. Using the same restriction enzymes, the Utrophin ABD1 gene will be inserted into an expression vector so that it contains components necessary for nickel column chromatography. The expression vector will be transformed into E. coli cells so it will grow and proliferate in the lab. The development and synthesis of this gene allows further thermodynamic analysis that will give insight into the small molecule energetics of Utrophin ABD1 in comparison to the energetics of Dystrophin ABD1.

Results

Use of Prokaryotic Model Organism for Plasmid Expression

The development of the Utrophin gene began with investigating various hosts to carry the eukaryotic gene. It is important to chose the correct cells for the expression of the gene otherwise the gene will not proliferate in the cell. Prokaryotic cells, like bacteria, are the ideal expression systems because they contain a separate, self-replicating circular piece of DNA called a plasmid. Plasmids have been evolutionarily advantageous for prokaryotic cells because they can be transferred from cell to cell without the entire genome replicating, thus increasing cellular variability. Specifically, E. coli was the prokaryote chosen, as it is a model organism signifying that its genome has been sequenced and highly studied. This knowledge allows easier manipulation of the transcription within the cell and allows any form of desired gene to be expressed. Additionally, *E. coli* can be easily obtained and grown because they need normal physiological temperature (37

degrees Celsius) to survive, they have rapid growth rates, and can replicate every 20 minutes providing an extensive number of colonies in a day, and an inexpensive nutrient broth for food. Furthermore, *E. coli* is desirable in this lab because this lab has never purified Utrophin ABD1, so the purification will be a form of experimentation. As a result, *E. coli* cells provide efficient way of mass producing cells for this purification.

Development of Utrophin ABD1 for Insert into Plasmid

After deciding *E. coli* cells are the most efficient way to proceed, it must be decided if a eukaryotic protein can actually be expressed in a prokaryotic cell because eukaryotic and prokaryotic cells differ in their cellular modifications. Prokaryotes have one location for transcription and translation because they do not have membrane bound organelles. This allows them to have simultaneous transcription and translation. Conversely, eukaryotic transcription is unique because it takes place in two different locations, the nucleus and the cytoplasm. As a result of the different locations of eukaryotic transcription and translation, eukaryotic systems often have posttranslational modifications added to its mRNA (the directions for the assembly of the protein) after translation has occurred. This functions to protect the mRNA as it leaves the nucleus. Posttranslational modification could include the splicing out of introns, which are the coding regions of genes, and are important for adding variability to the genome; the addition of a polyAAA tail and a 5'cap that protect the mRNA as it leaves the cell's nucleus; or the addition of chemical markers to the DNA to tell if a protein should be expressed or not. However, posttranslational modification will not pose a threat because the

Utrophin ABD1 gene was composed of the amino acid sequence from the functional protein from *homo sapien* Utrophin. This means that any posttranslational modification that would occur in the eukaryotic cell would have already had to occur before the amino acids are assembled into their primary structure to create the protein. Therefore, the amino acid/protein sequence was legitimate to use for the creation of Utrophin ABD1 without any concerns.

Although the 20 codons and their nucleotides sequences are highly conserved between species, some codons for different amino acids are favored in different species. Because a eukaryotic gene was utilized in a prokaryotic organism, the amino acid sequence from humans needed to be codon optimized for *E. coli*. For example, if a human contains a gene that reads TTA and encodes the amino acid leucine, bacterial hosts may prefer the leucine amino acid in the terms of CTG. As a result, it was necessary to make sure the gene was codon optimized so that the codon for leucine will contain the three nucleotides most common in *E. coli* so Utrophin will be transcribed in the most efficient way possible (Codon Optimization Tool, 2016).

After the sequence of Utrophin ABD1 was optimized for *E. coli*, the development of the gene could occur. The gene was flanked with two restriction enzymes so that Utrophin ABD1 could be excised from the plasmid it will be inserted into and more easily placed into an expression vector (this decision will be discussed more in depth in the next section).

Figure 1: Representation of the designed gene for insertion into an E. coli plasmid.

Utrophin ABD1 is flanked with restriction enzymes BamHI and XhoI for excision of this segment from a plasmid into a vector using these two restriction enzymes.

The two restriction enzymes utilized were BamHI and XhoI and were chosen based on their position in the expression vector. Normally, when designing a plasmid, just the sequence of the desired gene is inserted. However, in order to skip the step of designing primers for PCR amplification, the restriction endonucleases will recognize these specific nucleotide sequences and cleave the DNA at this point and will be amplified. This was only possible because of the position of the restriction enzymes in expression vector. Additionally, it was necessary to add a series of three stop codons onto the gene of interest after the Utrophin ABD1 segment. This was necessary because ABD1 is just a fragment of the entire Utrophin gene so it does not have naturally occurring termination sequences. When this plasmid is finally grown in its cell, the protein will stop replication because of the addition of the stop codons.

The next step is to determine where the construct should be inserted in the plasmid. However, based on the properties of the pUC57 plasmid utilized, this has already been selected. The construct was inserted into the lac Z gene on the pUC57 plasmid (a plasmid that mimics *E. coli* plasmid pGEM). The lac Z gene codes for an enzyme called B galactosidase that when functional, cleaves lactose into glucose and

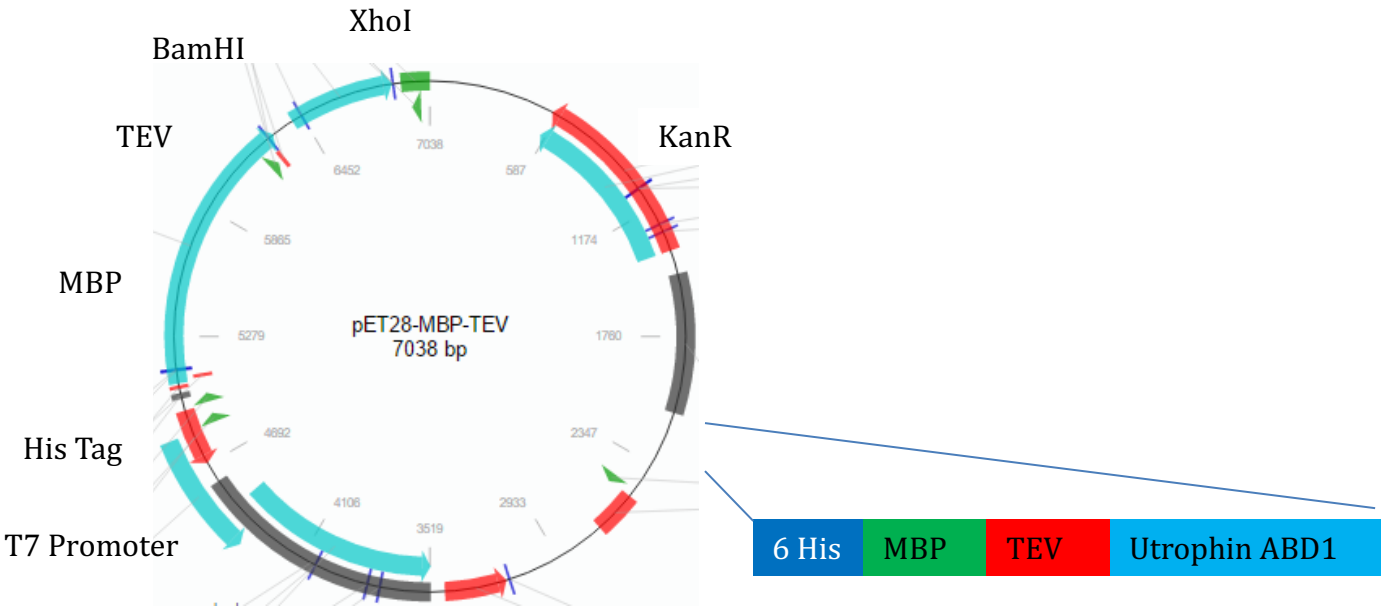
galactose. The gene is inserted in this segment because lac Z can be used as a screenable marker indicating it can give visual indication of the insert of the gene. When a gene is incorporated into this region, the lac Z gene function will be lost as a result of the gene disruption. When these bacterial are streaked on a nutrient agar plate, they will appear white because a special ingredient called Xgal does not detect the presence of glucose or galactose. However, when the B galactosidase gene is left functional, implying no inserted gene, the colonies will appear blue because Xgal detects the production of glucose and galactose. As a result, the addition of the gene into the lac Z gene gives visual representation about the success of the insertion of the gene (Elliott, Kilpatrick, Krebs, & Lewin, 2013). Additionally, the antibiotic Kanamycin provides another selectable marker because it selects for the survival of cells with only Kanamycin resistance while the cells without this resistance will die.

Ligation into Expression Vector for Ideal Chromatography Traits

After it was insured that our gene was synthesized, designed and inserted into an *E. coli* plasmid, it was subcloned into an expression vector. A vector is similar to a plasmid; however it is engineered to have components that can be specific to any process. In this case, the Utrophin construct needs to contain components necessary for nickel column purification and needs to be grown in *E. coli* cells. Synthesis begins with the T7 promoter. The promoter is part of the genome that initiates transcription, or the change of the sequence from its nucleotides into the amino acids that compose the protein. This promoter is critical to the synthesis of the gene because it is a viral promoter and is widely used because of this fact. The T7 promoter can be controlled outside of the parameters of the bacterial vector,

which is advantageous because the transcription of the vector can be initiated independently becoming faster and more efficient. A map of this vector is present in figure 3 below depicts the set up of the vector.

Figure 3: Expression vector for the insert of the Utrophin ABD1 gene with traits for nickel affinity chromatography.



This vector showcases the necessary components for the Utrophin ABD1 segment that were selected for high affinity purification including the final construct on the right that arises after insertion into this plasmid.

As stated before, the restriction enzymes BamHI and XhoI were chosen based on their position in this vector. After transcription begins, the synthesis of the gene will begin with the six histidine tag followed by a maltose binding protein, a TEV protease, and Utrophin ABD1 as displayed in Figure 3.

Transformation of vector into competent *E. coli* cells

The vectors containing components for purification will be transformed into *E. coli* cells for rapid growth of the desired construct. The process of transformation is extremely harsh on the cells receiving the insert and as a result, the majority of these cells die. However, only one *E. coli* needs to survive and proliferate for these methods to be successful. Transformation occurs in one of two ways. First, calcium shock can be used as the *E. coli* cells are put in a solution of cold calcium chloride. This high concentration of calcium binds to the negatively charged cell membrane with its positive charge. The addition of a high heat destabilizes the cell membrane creating lesions, allowing the vector to then enter the cell through these holes. Usually, the creation of lesions in the cell membrane is detrimental to the cell as the cell membrane cannot always rebuild the membrane before it loses all its contents. However, an extremely small percentage can survive and this is the percentage sought after. Additionally, another form of transformation occurs by electroporation. Electroporation fires an electric current at the *E. coli* cells by creating pores in the cellular membrane for the inserts to get in. This method also proves to be very hard on the cell. When transforming cells, it was important to have the correct vector to cell ratio in solution because if there was too high of a volume of *E. coli* cells, it will be very hard to recognize if any vector has been taken up. Conversely, if there is too high a concentration of the vector insert, the *E. coli* will be overwhelmed and harmed. For this reason, different ratios of vector to cell are utilized to see what is the most successful.

High Yield Purity is Ensured by Nickel Affinity Chromatography

In order to utilize nickel column chromatography, a six histidine tag, maltose binding protein, and a TEV protease were optimal for nickel chromatography. Nickel column chromatography is desirable because of its high yield in purification of protein. This is due to the strong covalent interaction between nickel and histidine. Specifically, nickel is the metal of choice because it is a readily available transition metal that chelates, or tightly binds, the resin beads (Bornhorst & Falke, 2000). In order to take advantage of this unique binding, six histidines are added to the construct as previously stated. Histidine is important because the lone pair on the nitrogen in the imidazole ring attaches to the positively charged nickel by donating the electrons from the lone pair of electrons in nickel. This whole process is carried out at the pH of 8 because this is where the histidine amino acid is deprotonated and the lone pair of electrons are most readily exposed. Specifically, six histidines are used because it has been found to have the best probability of binding to the nickel and staying on the column. Next, the maltose binding protein was desirable because it maintains the solubility of the protein in solution as it is being purified. It does so by preventing inclusion bodies, or aggregates of overexpressed protein, from forming which often occurs when there is such a high rate of expression of a certain protein. Finally the TEV protease comes from the tobacco etch virus in the tobacco plant. This virus has an extremely specific amino acid sequence not common to any eukaryotic or prokaryotic organism. As a result, the TEV protease has the ability to recognize this specific sequence and cleaves the protein at this point (Waugh, 2010).

After the TEV protease cleaves the construct, the purified Utrophin ABD1 complex is available.

In order to fully understand the whole process, figure 4 displays the overall scheme.

Figure 4 Development of Utrophin ABD1 From its Amino Acid Form to the Purified Product

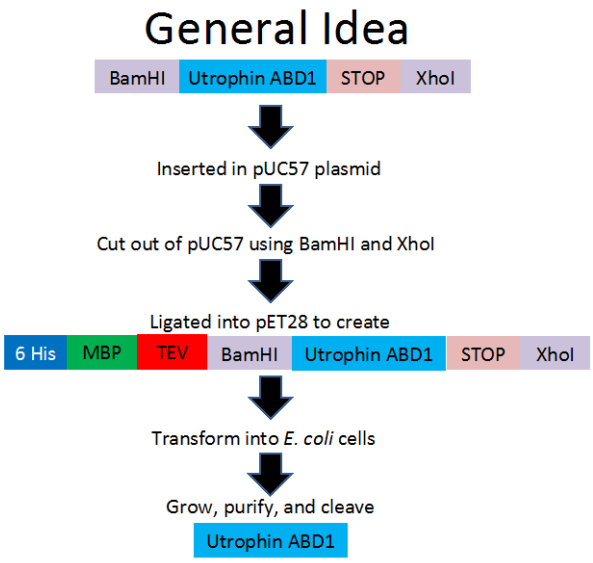


Figure 4 displays the full scheme of Utrophin ABD1 from the design of the Utrophin ABD1 segment, to the integration into the E. Coli pUC57 plasmid, to the ligation into the expression vector, and finally the purification of the desired gene.

Conclusion:

This research is critical for the future of Muscular Dystrophy patients in that the development and synthesis of this gene can provide insight into possible therapeutic cures for this crippling disease. Utrophin, the protein homologous to Dystrophin, is already up-regulated in patients with DMD. However, the exact

reason why Utrophin does not dissipate mechanical stress the same way as Dystrophin is the basis of this research. However, the gene itself must be developed before the thermodynamics of the unfolding of Utrophin can be parameterized. As already known, structure determines function and this will be necessary to determine how the function of Utrophin correlates with Dystrophin. Discovery in this area of science will lead to a better quality of life for patients with DMD in addition to an extended life expectancy.

Works Cited

Bornhorst, J. A., & Falke, J. J. (2000). [16] Purification of Proteins Using Polyhistidine Affinity Tags. *Methods in Enzymology*, *326*, 245–254.

Codon Optimization Tool. (2016). Retrieved 2016, from Integrated DNA Technologies: https://www.idtdna.com/CodonOpt

Elliott, S., Kilpatrick, T., Krebs, J., & Lewin, B. (2013). *Lewin's Essential Genes* (Third Edition ed.). Burlington, MA, USA: Jones & Bartlett Learning.

Guiraud, S., Squire, S. E., Edwards, B., Chen, H., Burns, D. T., Shah, N., … Davies, K. E. (2015). Second-generation compound for the modulation of utrophin in the therapy of DMD. *Human Molecular Genetics*, *24*(15), 4212–4224. http://doi.org/10.1093/hmg/ddv154

Lin, A.Y., Prochniewicz, E., Henderson, D.M., Li, B., Ervasti, J.M. and Thomas, D.D. (2012) Impacts of dystrophin and utrophin domains on actin structural dynamics: implications for therapeutic design. J. Mol. Biol. 420:87-98.

Prochniewicz, E., Henderson, D.M., Ervasti, J.M. and Thomas, D.D. (2009) Dystrophin and utrophin have distinct effects on the structural dynamics of actin Proc. National Acad. Sci. USA

Waugh, D. (2010, September). *TEV Protease FAQ*. (N. C. Institute, Producer) Retrieved April 2016, from Macromolecular Cyrstallography Laboratory: http://mcl1.ncifcrf.gov/waugh_tech/faq/tev.pdf