Myotonic Dystrophy: The Structure of Cug Repeats in Solution

by Brian Truong
Biochemistry / Honors College

Faculty Mentor: Andy Berglund
Chemisty/Institute of Molecular Biology
Faculty Mentor: Jill Murray: Molecular Biology

Abstract

Myotonic dystrophy (DM), a genetic and neuromuscular disorder, is the most common form of adult-onset muscular dystrophy and results in symptoms such as proximal muscle weakness, myotonia, iridescent cataracts and cardiac arrhythmia. Myotonic dystrophy type 1 (DM1), the most prevalent type of DM, is caused by a $(CTG)_n$ expansion in the 3' untranslated region of the dystrophin myotonin protein kinase gene. At the RNA level, the CUG repeats form a stem-loop that is thought to be the pathogenic element. We determined that the crystal structure of CUG repeats and found that the repeats form a structure similar to standard double-stranded A-form nucleic acid. Since crystal structures may be misleading, however, we will verify A-form conformation using a solution-based assay. A-form double-stranded RNA (dsRNA) is known to be cleaved into small RNA molecules in cells. Therefore, we predict that the cellular machinery recognizing and cleaving these dsRNA should degrade expanded CUG repeats. To test this hypothesis, we will monitor the cleavage of CUG repeats in solution using an enzyme that recognizes and cleaves A-form dsRNA.

Introduction

Myotonic dystrophy (DM), the most common form of adult-onset muscular dystrophy, is a neuromuscular disease that affects approximately 1 in every 8,000 individuals and is characterized by a defective gene passed from one generation to the next (Mooers). The disease is progressive so that the characteristic symptoms of muscle wasting and weakness develop gradually over the lifetime of an affected individual. In addition to general muscle weakness, people with DM also have myotonia, the delayed relaxation of voluntary muscles after contractions. While this is a mild problem in early onset patients, affected individuals eventually lose the ability to release the hand from a grip due to the myotonia (Kurihara). Besides the muscles in the hands, the first muscles to be affected are usually those of the face, neck, forearms and feet. In addition to targeting muscles, DM affects the nerves that control muscles and the interactions between muscles and nerves. The clinical symptoms vary in severity from person to person but typically include the following: myotonia, cardiac disease, cataracts, mental disorders (including mental retardation), testicular atrophy, respiratory impairment, adverse reactions to anesthesia, difficulty in swallowing (dysphagia) and other gastrointestinal tract involvement, excessive output of insulin, abnormal carbohydrate metabolism, and excessive sleeping (Ranum). DM is thus appropriately termed a multisystematic disease since it can affect the tissues and organs of many body systems in addition to the voluntary muscle system.

The symptoms of DM are so diverse that they must be examined carefully in order to provide a proper diagnosis. While fifty percent of those with the disease show visible signs by approximately twenty years of age, many do not develop distinct symptoms until

after age fifty. As a rule, the severity of the disease increases with successive generations while the age of onset decreases. In the most severe case, an infant can inherit a rare condition of congenital DM (maternally transmitted) leading to profound bodily weakness as well as difficulty in sucking, swallowing and breathing. Although myotonia is not a feature of the congenital DM, motor development is usually delayed and some children may even show some signs of mental retardation (Ranum). Since DM is progressive in children and adults, most symptoms usually do not lead to disability until fifteen to twenty years after the onset of symptoms. At an early stage of the disease process, most affected people demonstrate weakness and wasting of the facial, jaw and neck muscles as well as frontal balding in men. A simple physical examination can be performed to detect myotonia, and special laboratory tests (i.e. muscle biopsy) can be conducted to confirm the diagnosis. Because DM is a heredity disease, an examination of DM history in the family can also serve as a good indicator that a person has the disease. DM runs in families because it is transmitted from generation to generation by individuals who have inherited the defective gene. Because the defective gene is autosomal (non-sex linked) dominant, it only takes one DM gene from either the father or mother to produce the disease in an offspring (Voet). In other words, if one parent has the disease, there is a 50% probability that a child will inherit it.

There are two types of myotonic dystrophy: type 1 (DM1) and type 2 (DM2). Both types are caused by a repeating unit within the non-coding region of a gene on a chromosome, a large molecular structure that packages genetic information in the form of DNA in the cell. To picture the location of the disease in the body of an affected person, you can image yourself in a library. The library

represents 23 chromosomes of genetic material and contains shelves that represent individual chromosomes. When you pick one of the books on a shelf, you look at a particular gene on a single chromosome. The pages in the chosen book, in turn, represent DNA sequences (or the genetic alphabet). If there are multiple copies of the same page, a repeating unit exists. Because multiple copies of the page are unnecessary, they do not fit in the context of the book, the shelf and the library. This repeating unit illustrates the problem faced in DM1: a CTG repeat expansion in 3' untranslated (non-coding) region of the dystrophin myotonin protein kinase (DMPK) gene on chromosome 19.

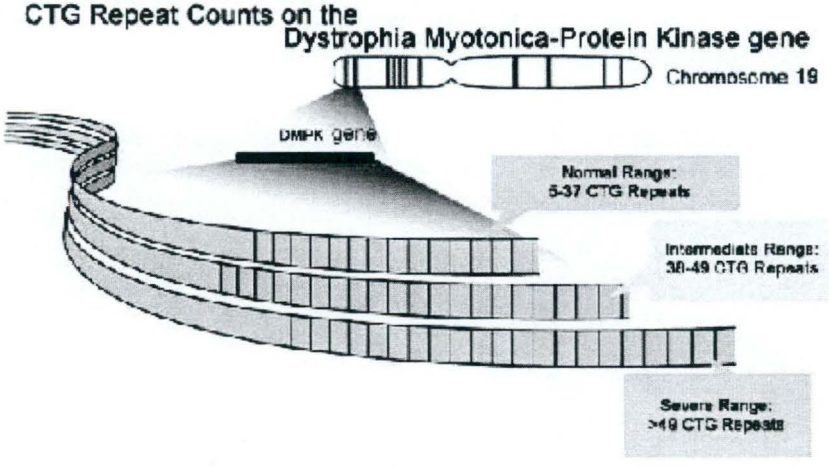


Figure 1: CTG Repeat Counts on the DMPK Gene

The number of repeats is directly proportional to the severity of the disease. Therefore, increasing the number of CTG repeats in the non-coding region of the DMPK gene on the long arm of chromosome 19 in DM1 amplifies the severity of the disease. In terms of myotonic

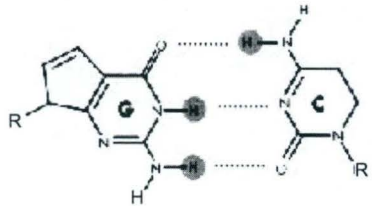
dystrophy, there is a significantly greater number of repeats in the congenital (the most severe) form than the adult form of DM1. Normal individuals typically have between 5 and 37 CTG repeats, whereas individuals with DM have more than 50 or sometimes even thousands of CTG repeats (Ranum). Since the number of repeats increases from generation to generation, the disease theoretically worsens as it is transmitted in the family lineage.

Myotonic dystrophy is one of many repeat expansion disorders. Aside from DM1 and DM2, repeat expansion disorders include Huntington's disease (CAG repeat), Fragile X syndrome (GCC repeat), Friedreich's ataxia (GAA repeat) and Spinocerebellar ataxia (CAG repeat).

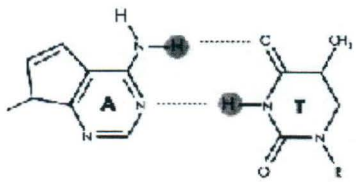
The number of CTG repeats leads to DM but is not the direct cause of the disease. The CTG repeats, which occur in DNA, are converted to RNA forming CUG repeats. The CUG repeats are what become the pathogenic element in DM.

To fully understand the conversion of CTG repeats to CUG repeats in DM1, we refer to nucleic acids, the molecules that enable living organisms to reproduce their complex components from one generation to the next (Campbell). There are two types of nucleic acids: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), both of which are composed of three components: a nitrogenous base, a pentose and a phosphate group. The two forms of nitrogenous bases are pyrimidines (six-membered ring of carbon and nitrogen atoms) and purines (six-membered ring fused to a five-membered ring). The pyrimidines include cytosine (C), thymine (T) and uracil (U). The purines include adenine (A) and guanine (G). Following Watson-Crick geometry, only a pyrimidine can pair with a purine through molecular interactions called hydrogen bonding.

However, there are only two possible pairings since one forms three hydrogen bonds and another forms two hydrogen bonds. In DNA, cytosine pairs with guanine and thymine pairs with adenine. In RNA, thymine is replaced with uracil. DNA and RNA then form helical structures that involve stacking base pairs that twist in several different conformations.



Formation of hydrogen bonds between guanine and cytosine

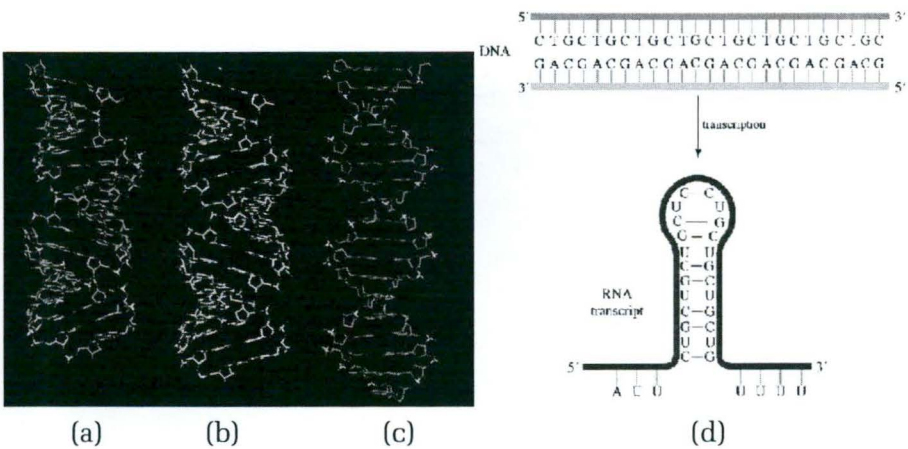


Formation of hydrogen bonds between adenine and thymine

Figure 2: Watson-Crick Base-Pairing

When two strands come together to make double-stranded nucleic acid, they can take many different structural forms. B-form DNA is the biologically predominant form of DNA, but DNA can also be A-form or Z-form, depending on a number of factors that include the quantity and orientation of base pairs per helical turn. RNA, on the other hand, is usually single-stranded, but will typically adopt the A-form conformation when double-stranded. In the case of DM, the CUG repeats fold back on themselves forming a long, double-stranded stem-loop structure held together by hydrogen bonds. While Cs pair with Gs, there is an unfavorable pairing between two Us at the RNA level. This odd pairing is called U-U mismatches. The U-U mismatches are tolerated as the energy

from the C-G pairing compensates for the unfavorable mismatching to form a thermodynamically favorable stem-loop structure of expanded CUG repeats. Crystallography studies have shown that expanded CUG repeats are similar to A-form in the crystal structure despite the presence of U-U mismatches (Mooers). This information provides insight on the structure of CUG repeats that may lead to the design of new drugs to treat DM.



(a) (b) (c) (d)

Figure 3: RNA Structure
(a) CUG repeats, (b) A-form RNA, (c) B-form DNA, (d) CUG Repeat Stem-Loop Structure

To connect the DNA and RNA worlds, the central dogma of biology states that DNA is transcribed (copied) to RNA, and the resulting RNA is translated (converted) to protein. Thus, this molecular process transforms DNA to functional proteins, complex biopolymers that are essential to the structure and function of all

living cells in the human body.

DNA → RNA → Protein

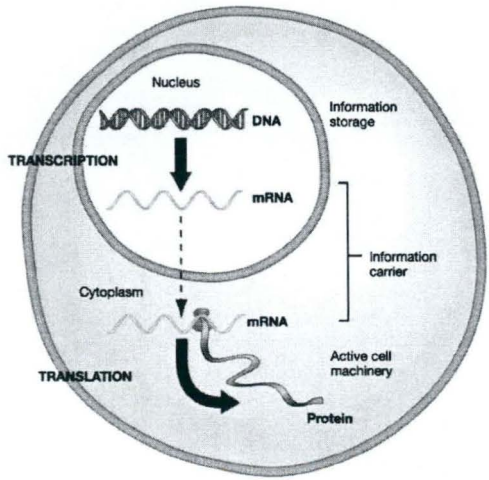


Figure 4: The Central Dogma

DNA is copied into translatable RNA by transcription and RNA processing. To initiate transcription, and thus regulate gene expression, a stimulus from the external or internal environment signals the local unpacking of DNA in the chromosome of the nucleus. Transcription produces unmodified RNA in the form of pre-mRNA with the assistance of proteins such as RNA polymerase, transcription factors and control elements including enhancers, activators and repressors. Pre-mRNA contains exons (coding regions) and introns (noncoding regions) that determine which regions code for proteins. To remove the noncoding regions, a splicing complex assembly called the splicesosome excises the introns and joins adjacent exons to form mRNA, which is then translated by the ribosomes in protein synthesis.

In DM1, however, improper splicing occurs resulting in incorrect proteins. At the RNA level, the CUG repeats form an extended stem-loop structure that sequesters a group of proteins collectively known as muscleblind (MBNL) proteins (Dansithong). This sequestration of MBNL causes the improper splicing to occur. The MBNL-sequestering CUG stem-loop structure thus leads to inappropriate gene expression programs that result in disease.

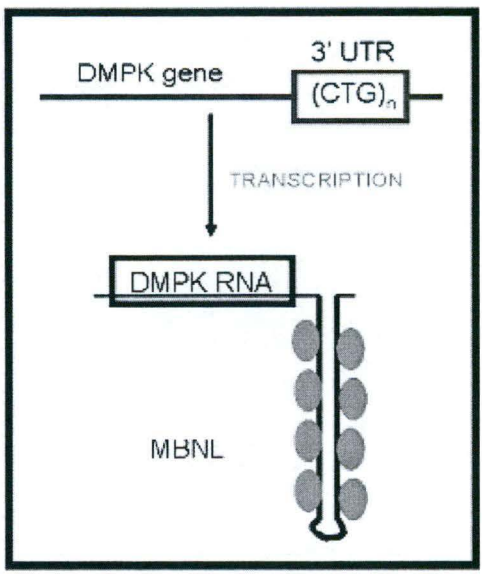


Figure 5: CUG Repeat Stem-Loop Structure with Sequestered MBNL

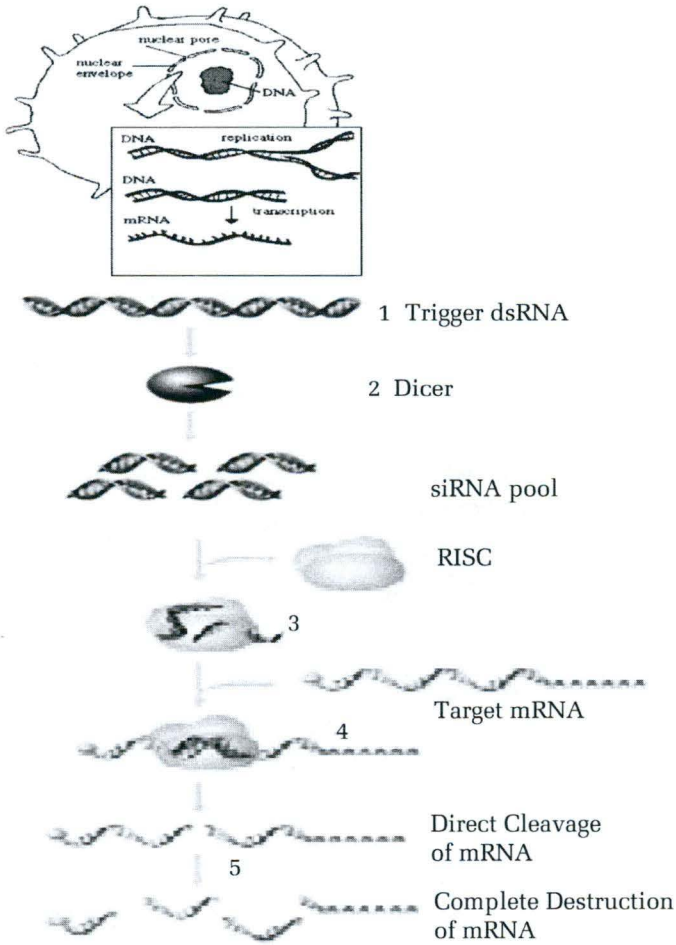
The DMPK protein cannot be made from the repeat-containing mRNA because a so-called toxic stem-loop structure forms and inhibits translation (Mooers). As a result, this turns off gene expression. To study this gene regulating mechanism, scientists have recently studied how specific double-stranded

RNA molecules silence certain genes using a phenomenon known as RNA interference (RNAi). In RNAi a cleaving enzyme, appropriately called Dicer, cuts double-stranded RNA molecules into short fragments (Vermeulen). While one strand degrades, the other strand (miRNA) associates with a large protein complex. This miRNA-protein complex then binds to mRNA molecules that have the complementary sequence and degrades them or blocks translation (Hutvágner). Since protein sequestration and RNAi both prevent the process of proper translation, understanding how Dicer cleaves double-stranded RNA becomes important in learning more about the structure of expanded CUG repeats in solution.

Dicer plays an important role in preventing protein synthesis, and recent research has shown that Dicer cleaves A-form RNA. Since expanded CUG repeats in *crystal* are similar to A-form and have been cut successfully by Dicer, cleaving expanded CUG repeats in *solution* by the same enzyme should confirm that the structure of DM is A-form in solution as well. Studying the effect of Dicer on double-stranded RNA is important in finding a possible treatment for DM. If we confirm the CUG repeats are A-form, then we can target this structure with small molecules that could function as a drug. This would eliminate protein sequestration that causes improper splicing in DM.

DM is a mysterious health problem because there are still many unanswered questions about the molecular basis or mechanism of the disease. Currently, no specific treatments have been found for the muscle weakness and wasting in DM. To help affected DM individuals, physicians prescribe ankle and leg braces that help support muscles as weakness progresses and medications that can

Figure 6: Dicer Activity



relieve myotonia. Some people choose to test for the defective gene in DM by obtaining a genetic examination that searches for the expanded CUG repeats. While this procedure is helpful in diagnosing the disease, it remains a highly controversial medical issue. In some cases, children (and not the parents) are interested in

knowing if they have the defective gene that causes DM. Whether this potentially life-changing decision is up to the children or the parents remains an ethical concern.

Overview of Methods

Myotonic dystrophy type 1 (DM1) is caused by expanded CUG repeats that form a stem-loop thought to be the pathogenic element at the RNA level. The crystal structure of CUG repeats has been found to form a structure similar to standard double-stranded A-form RNA. Dicer is a specific cleaving enzyme that cuts double-stranded A-form RNA. Since Dicer cuts CUG repeats in crystal, we predict that it should also degrade expanded CUG repeats in solution. To test this hypothesis, we will monitor the cleavage of CUG repeats in solution.

We begin with two circular DNA molecules called plasmids: one containing expanded CTG repeats and the other not containing CTG repeats. The plasmid containing CTG repeats is cut out, and using gluing enzyme ligase, is inserted into the one without CTG repeats. This procedure is done by a process known as molecular cloning, a technique that extracts a gene or piece of DNA from one organism and inserts it into a second organism. This cloning method allows genes to be used and studied.

We have a CTG-containing plasmid. We also have a CAG-containing plasmid in the laboratory. Using a restriction enzyme called BAM HI, we cut both plasmids to form two linearized double-stranded DNA molecules: one molecule contains expanded CTG repeats and the other contains CAG repeats. The two linearized DNA molecules are transcribed into RNA transcripts using RNA polymerase, a RNA-synthesizing enzyme. The RNA transcripts

contain expanded single-stranded CAG and CUG repeats, respectively. These RNA transcripts are annealed, forming a stem-loop or double-stranded structure.

Then we perform a Dicer assay by adding Dicer enzyme to (1) a stem-loop with CAG repeats, (2) a stem-loop with CUG repeats, and (3) a double-stranded structure with CAG and CUG repeats. Sets (1) and (3) are our controls. Set (1) is our negative control, and we expect to see no Dicer cleavage because expanded CAG repeats are not A-form. Set (3) is our positive control, and we expect to see Dicer cleavage because expanded CAG/CUG repeats are A-form. Set (2) is our test experiment, and we hope to see Dicer cleavage. This would confirm our hypothesis that the expanded CUG repeats of DM are A-form in solution.

Materials and Methods

3.1 Molecular Cloning

The $(CTG)_{54}$ repeat expansion is excised from a plasmid called PSP72 using BAM HI, a restriction enzyme. Using ligase, this DNA insert is put in an empty vector (host plasmid) called pcDNA3. The $(CTG)_{54}$-containing plasmid is grown overnight in E. coli cells to produce multiple copies of the defective gene-containing plasmid. Once grown to capacity, these plasmids are purified for enzymatic digestion. A $(CAG)_{54}$-containing plasmid is cloned in the same manner.

3.2 Enzymatic Digestion

Five micrograms of $(CTG)_{54}$- and $(CAG)_{54}$-containing plasmids are individually digested using BAM HI. Buffering solution and deionized water are added to complete the digestion. This reaction

is incubated for 1 hour at 37ºC. The linearized DNA is purified and extracted using a phenol/chloroform/isoamyl alcohol solution. The DNA concentration is determined using an ultraviolet (UV) spectrometer.

3.3 Transcription

One microgram of both $(CTG)_{54}$- and $(CAG)_{54}$-containing plasmids are individually transcribed using T7 RNA polymerase. Dithiothreitol (DTT), NTP Mix, radiolabeled CTP, Tris (pH 9), yeast pyrophosphate, RNAsin and deionized water are added to complete transcription. This reaction is incubated overnight at room temperature. The mixture is loaded with loading dye into a denaturing polyacrylamide gel to purify the CUG repeats from abortive transcription products. This procedure runs for approximately 2 hours at 300 volts. The radioactive RNA bands are visualized using a Storm 860 phosphoimager and IMAGEQUANT 1.2 software. The RNA transcripts are excised and eluted from the gel in sodium acetate (pH 5) overnight at -4ºC. The RNA transcripts are then purified by being precipitated in cold ethanol, spun, air-dried and washed.

3.4 Annealing

The transcripts are single-stranded $(CAG)_{54}$- and $(CUG)_{54}$-containing RNA. To form double-stranded RNA, two desired RNA strands are annealed using an annealing buffer solution and deionized water. The reaction is then heated for 5 minutes at 95ºC. To form the stem-loop structures for each of the repeat expansions, the reaction mixture is cooled for 20 minutes in ice. To form double-stranded RNA containing both complementary strands, the reaction

mixture is cooled for 20 minutes at room temperature.

3.5 Dicer Assay

Three double-stranded RNA (dsRNAs) are prepared: ds$(CAG)_{54}$, ds$(CUG)_{54}$ and ds$(CAG/CUG)_{54}$. The dsRNAs are digested using recombinant Dicer enzyme. Buffering solution and deionized water are added to complete digestion. This reaction is incubated for 18-20 hours at 37ºC. Once completed, the Dicer-cleaved dsRNAs are stored at -20ºC.

3.6 Visualization

The Dicer-cleaved dsRNAs are loaded with loading dye into a denaturing gel. This procedure runs for approximately 2 hours at 300 volts. The radioactive dsRNA bands are visualized using a Storm 860 phosphoimager and IMAGEQUANT 1.2 software.

Results

Figure 7 shows various RNA molecules running on a denaturing gel. A denaturing gel is a device used to separate molecules based on molecular size. The ladder is loaded to provide a standardized "ruler stick" that measures the distance molecules travel on the gel. In this case, single-stranded CUG repeats (ssRNA) and double-stranded CUG repeats (RNA stem-loop) migrate through the gel more quickly due to their relatively small molecular size. However, double-stranded CAG/CUG repeats (dsRNA) do not flow through the gel as quickly since they have a relatively large molecular size. Based on the distances that the RNA migrated through the gel relative to their molecular size, this annealing gel demonstrates a somewhat successful annealing of ssRNA to form dsRNA and

RNA stem-loop. However, the results are not reliable because there is a lot of smearing that could have altered the outcome. Further experimentation is needed before conclusions can be made.

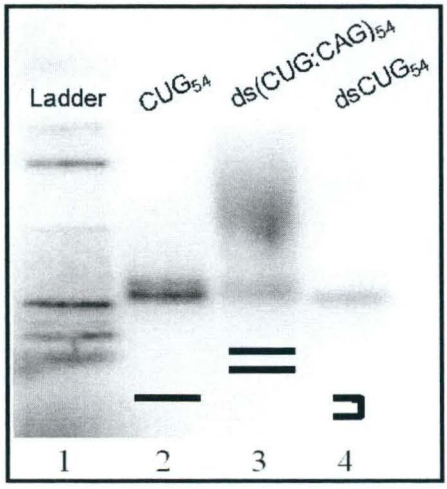


Figure 7: Annealing Assay

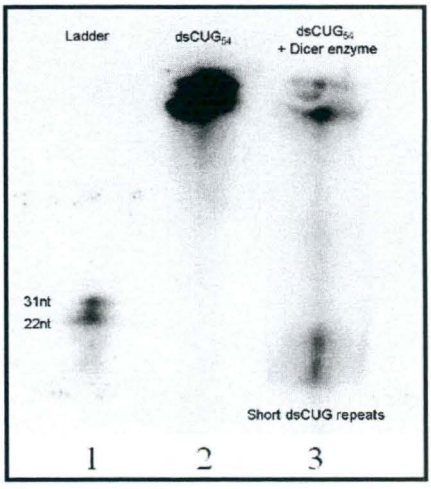


Figure 8: Dicer Assay

Figure 8 shows various RNA molecules running on a denaturing gel. The ladder is loaded to provide two points close to the size of the cleaved product. In this case, long double-stranded CUG repeats (long dsRNA) are visualized in the absence and presence of the Dicer enzyme. Dicer, an enzyme that specifically cleaves double-stranded molecules into fragments that are approximately 22 nucleotides (22 nt), cuts long double-stranded CUG repeats into short double-stranded CUG repeats (like miRNA in RNAi). This Dicer assay demonstrates a successful cleavage of long dsRNA to short dsRNA. The result needs to be reproduced, and the rate of RNA cleavage needs to be determined by varying the length of time the reaction is allowed to proceed.

Discussion and Future Direction

Since the summer of 2005, when work on this project begun, I have managed to demonstrate with some success that the expanded CUG repeats of DM are A-form in solution because they are cleaved by the A-form specific Dicer enzyme. This result confirms the previous study that showed expanded CUG repeats in DM are A-form in crystal. Continued research will strive to produce clearer and more conclusive images of the annealing and Dicer assays. To accomplish that goal, I will need to optimize the molecular cloning, transcription, annealing and cleaving procedures in this experiment. If successful with the Dicer assays, I will move onto circular dichroism, a laboratory method that I anticipate will support the hypothesis that the structure of CUG repeats are A-form in solution as they are in crystal.

References

1. Apirion, David. Processing of RNA. Boca Raton, Florida: CRC Press, Inc., 1984.

2. Broda, Magdalena, Elzbieta Kierzek, et al. (2005) Thermodynamic Stability of RNA Structures Formed by CNG Trinucleotide Repeats: Implication for Prediction of RNA Structure. *Biochemistry* 44: 10873-10882.

3. Campbell, Neil A., and Jane B. Reece. Biology. (7e) New York: Pearson Education, Inc., 2005.

4. Dansithong, Warunee, Sharan Paul, et al. (2005) MBNL1 is the Primary Determinant of Focus Formation and Aberrant Insulin Receptor Splicing in DM1. *The Journal of Biological Chemistry* 280: 5773-5780.

5. Krainer, Adrian R, ed. Eukaryotic mRNA Processing. New York: Oxford University Press, 1997.

6. Krieg, Paul A, ed. A Laboratory Guide to RNA: Isolation, Analysis, and Synthesis. New York: John Wiley & Sons, Inc., 1981.

7. Gesteland, Raymond F., Thomas R. Cech, and John F. Atkins, eds. The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA. (2e) New York: Cold Spring Harbor Laboratory Press, 1999.

8. Haley, Benjamin, Guiliang Tang, and Phillip D. Zamore. (2003) In vitro analysis of RNA interference in Drosophilia melanogaster. *Methods* 30: 330-336.

9. Handa, Vaishali, Tapas Saha, and Karen Usdin. (2003) The fragile X syndrome repeats form RNA hairpins that do not activate the interferon-inducible protein kinase, PKR, but are cut by Dicer.

Nucleic Acids Research 31: 6243-6248.

10. Hutvágner, Gyorgy, and Phillip D. Zamore. (2002) A microRNA in a Multiple-Turnover RNAi Enzyme Complex. *Science* 297: 2056-2060.

11. Kurihara, Teruyuki. (2005) New Classification and Treatment for Myotonic Disorders. *Internal Medicine* 44: 1027-1032.

12. Mooers, Blaine H.M., Jeremy S. Logue, and J. Andrew Berglund. (2005) The structural basis of myotonic dystrophy from the crystal structure of CUG Repeats. *PNAS* 102: 16626-16631.

13. Myers, Jason, Joshua T. Jones, et al. (2003) Recombinant Dicer efficiently converts large dsRNAs into siRNAs suitable for gene silencing. *Nature Publishing Group* 21: 324-328.

14. Neidle, Stephan. Topics in Nucleic Acid Structure. New York: John Wiley & Sons, Inc., 1981.

15. Nykamp, Keith R., and Maurice S. Swanson. (2004) Toxic RNA in the Nucleus: Unstable Microsatellite Expression in Neuromuscular Disease. *Progress in Molecular Subcellular Biology* 35: 57-77.

16. Ranum, Laura P.W., and John W. Day. (2004) Myotonic Dystrophy: RNA Pathogenesis Comes into Focus. *American Journal of Human Genetics* 74: 793-804.

17. Rickwood, D., and B. D. Hames, eds. Gel Electrophoresis of Nucleic Acids: A Practical Approach. (2e) New York: Oxford University Press, 1990.

18. Richter, Joel D. mRNA: Formation and Function. New York: Academic Press, 1997.

19. Rodger, Alison, and Bengt Nordén. Circular Dichroism & Linear

Dichroism. New York: Oxford University Press, 1997.

20. Vermeulen, Annaleen, and Linda Behlen, et al. (2005) The contributions of dsRNA structure to Dicer specificity and efficiency. *RNA* 11: 674-682.

21. Voet, Donald, and Judith G. Voet. Biochemistry. (3e) New York: John Wiley & Sons, Inc., 2004.

22. Zhang, Haidi, Fabrice A. Kolb, et al. (2002) Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *The EMBO Journal* 21: 5875-5885.