METHYLATION OF THE G-C RICH *DMPK* GENE:

CAUSATION FOR THE SEVERITY DIFFERENCE BETWEEN ADULT-ONSET AND CONGENITAL MYOTONIC DYSTROPHY?

by

Brian Barnette

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Honors Bachelor of Science in Biological Sciences.

Spring 2010

Copyright 2010 Brian Barnette
All Rights Reserved

METHYLATION OF THE G-C RICH *DMPK* GENE:

CAUSATION FOR THE SEVERITY DIFFERENCE BETWEEN ADULT-ONSET AND CONGENITAL MYOTONIC DYSTROPHY?

by

Brian Barnette

Approved: _____

Vicky Funanage, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____

Grace Hobson, Ph.D.
Committee member from the Department of Biological Sciences

Approved: _____

Marlene Emara, Ph.D.
Committee member from the Board of Senior Thesis Readers

Approved: _____

Alan Fox, Ph.D.
Director, University Honors Program

ACKNOWLEDGMENTS

This study could not have been possible without the help of several individuals. I would first like to thank Dr. Funanage for agreeing to sponsor not only my undergraduate research at the University of Delaware, but also for being my thesis advisor. Dr. Funanage had accepted me into her lab after a brief interview the winter of my sophomore year. Since then, I have greatly enjoyed my time in the Musculoskeletal Inherited Disease (MID) Laboratory, it was one of the greatest experiences I have had at the University of Delaware.

I would also like to thank two individuals I met the same day as my interview with Dr. Funanage, Sarah Swain and Susan Kirwin. Sarah Swain was instrumental in not only in the writing of my first grant proposal, but also in my laboratory training. Susan Kirwin, also immensely helpful in both writing the first grant proposal and training, was invaluable in her help with trouble-shooting, as well as assistance in completing certain aspects of experiments that my schedule prevented me from doing. In addition, it was her not so gentle nudges that kept the research moving forward every time a "brick-wall" arose. Without their help, this study would never have begun.

The other members of the MID Lab as well as member of the Neurogenetics Research Lab (NRL) deserve thanks as well for their help and patience. I would like to thank Priscilla Moses and Kathy Vinette for their assistance. Similarly, Karen Sperle and Dr. Jennifer Taube were also helpful, especially in their suggestions of bacterial cloning for sequencing and high fidelity PCR.

I would also like to thank the members of this thesis committee, Drs. Hobson and Emara. Your time and effort have been greatly responsible for the writing of this thesis. Finally, I would like to thank the University of Delaware, the Undergraduate Research Office, Charles Peter White Scholarship, and Howard Hughes Medical Institute, without their funding I would not have been able to conduct this research. Again thank you Dr. Funanage, Dr. Hobson, Dr. Emara, members of the MID lab, and members of the NRL lab, for all of your help with my research and thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Myotonic dystrophy type I is an autosomal dominant genetic disease. The disease arises from an expansion of a trinucleotide repeat, (CTG)n located in the 3' untranslated (3' UTR) region of the dystrophica myotonin protein kinase (*DMPK*) gene on chromosome 19q13.2-q13.3. Located proximal to this region is a CpG island, a region of DNA with a high concentration of CpG sequences that often associate with gene promoters in humans. Changes in these promoter regions, including the methylation status, modify expression of genes that they regulate. The purpose of this study was to test the hypothesis that there are allelic differences in the methylation status of this CpG island in normal, adult-onset and congenital myotonic dystrophy patients by using DNA prepared from somatic cell hybrids that carry either the normal or expanded *DMPK* allele. In addition, once the methylation status had been ascertained, it was the goal to determine if hypermethylation of the normal allele correlated with a decreased expression of the *DMPK* gene. This correlation could suggest the reason for severity differences between adult-onset and congenital DM1. The method of bisulfite-conversion was used to ascertain the methylation status of the CpG island. This method converts all unmethylated cytosine residues into uracil residues. The converted DNA of congenital DM1 patients was then amplified by means of High-Fidelity PCR where the converted uracil residues are amplified as thymine residues. This amplified DNA is then cloned and sequenced to compare the methylation status of the CpG island among individuals. After examining the

methylation status of this CpG island in one CDM1 patient and a patient presenting with signs and symptoms associated with DM1, it was determined that the non-mutant allele of these patients is not methylated. Since the non-mutant allele was determined to be unmethylated, hypermethylation could not be the cause of its decreased expression.

Chapter 1

INTRODUCTION

1.1 Myotonic Dystrophy Type I and *DMPK*.

Myotonic dystrophy type I (DM1; OMIM 160900) is a genetic multi-systemic disease with an incidence rate of 1 in 8,000, which can present at any age from infancy into late adulthood. The disease is inherited as an autosomal dominant trait and arises from an expansion of the triplet repeat, Cytosine-Thymine-Guanine (CTG), in the 3' untranslated region (UTR) of the dystrophica myotonin protein kinase (*DMPK*) gene (Figure 1.1.1). *DMPK* is a serine-threonine kinase that is found to be active in cardiac and muscle tissues; its particular targets and function are not known. In unaffected individuals, the CTG tract is between 5 and 35 repeats. In patients suffering from DM1, the triplet repeat ranges from greater than 50 up to 2000 (Ansved et al. 1997). A positive correlation exists between the number of repeats and the severity and time of onset of the disease. This correlation suggests that mildly affected patients with late adulthood onset have an expansion between 50 and 100, and that classical DM1 arises with expansions of 200 to 500. Finally, congenital DM1 (CDM1), the most severe form, presents in patients with greater than 1000 repeats. It is also noteworthy to mention the dynamic nature of DM1; the trait is passed to successive generations with increased repeat size and with increasing severity (Chi and Lam 2005; De Temmerman et al. 2008; Frisch et al. 2001; Longman 2006).

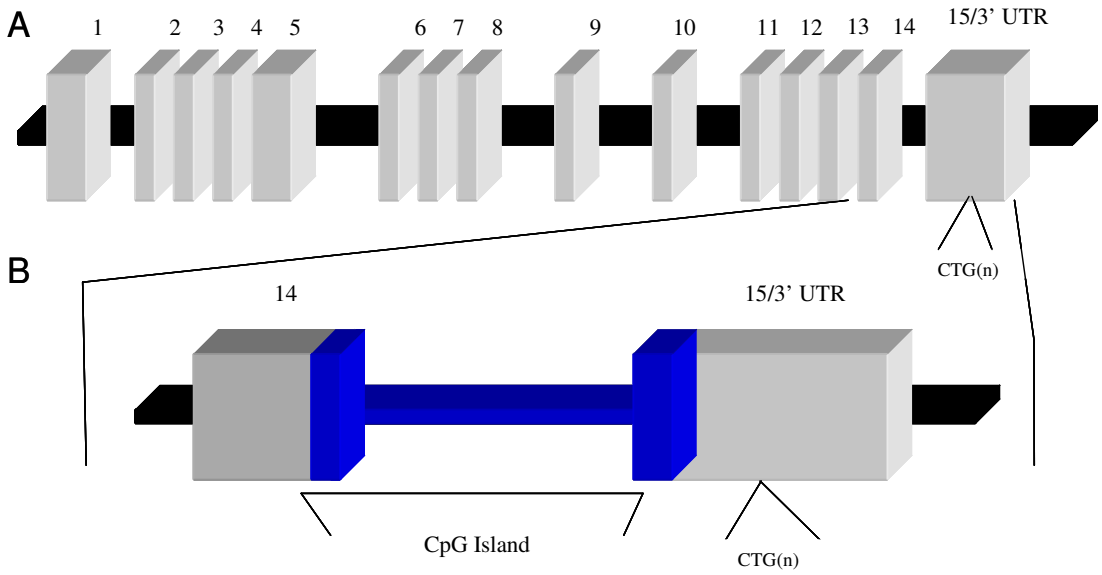


Figure 1.1.1 Diagram of the *DMPK* gene. A. The *DMPK* gene consists of 15 exons spread over approximately 13 kb on the long arm of chromosome 19. Alternative splicing of the *DMPK* transcript occurs, and removal of either exon 13 or 14 or both exons 13 and 14 occurs. Translation of transcripts lacking either exon 13 or exons 13 and 14 results in a frameshift and causes early termination of translation in exon 15, whereas transcripts lacking exon 14 do not cause a frameshift. The full-length transcript is the predominant mRNA species in skeletal muscle and brain. B. Located proximal to the CTG repeat is the CpG island, which is approximately 1.2 kb long (DMPK 2006; Frisch et al. 2001).

Clinical characteristics of myotonic dystrophy include skeletal, gastrointestinal, endocrine, respiratory, cardiac and cognitive problems. The most common symptoms across the varying degrees of disease severity are myotonia or slow muscle relaxation, distal muscle weakness (hands, lower legs, and face), and cognitive problems ranging from mild learning disorders to severe retardation in the congenital form. CDM1 patients also present at birth with hypotonia and respiratory distress, and half die in infancy. Those individuals that survive go on to develop the

classical features of DM1 in adulthood (Myotonic Dystrophy 2006; Huang and Kuo 2005; Orengo et al. 2008).

Located proximal to the CTG triplet repeat is a region of DNA with a high concentration of cytosine-guanine pairs, or a CpG island (Steinbach et al. 1998). Roughly 60-70% of CpG islands are associated with promoter regions in the human genome (England and Petterson 2005; Tooke et al. 2004). This particular island is within the promoter region for the downstream gene, DM locus-associated homeodomain protein/sine oculis homeobox homolog 5 (*DMAHP/SIX5*). Changes in these promoter regions modify expression of genes they regulate (Frisch et al. 2001). In the proposed project, we hypothesize that methylation (addition of a methyl group to Carbon5 of the cytosine nucleotide) of this CpG island in the normal allele of CDM1 patients is the basis for the increased clinical severity of the disease. Malumbres et al. (1999) have found that hypermethylation of the 3' UTR can interfere with transcription and that hypermethylation of CpG islands causes decreased gene expression. Changes in gene expression have been linked to the pathogenesis of many diseases, cancer being one common example (Rusk. 2005). In relation to DM1 patients, the hypermethylation of the CpG island proximal to the triplet repeat may cause expressional changes of either the *DMPK* gene or a number of genes downstream, such as *DMAHP/SIX5,* a transcription factor (Li et al. 2001; Umetani et al. 2006; Wang et al. 2008). These potential changes in gene expression may play a role in the pathogenesis of DM1, in particular the congenital form since the size of the disease-causing expansion overlaps between DM1 and CDM1. The expanded allele in CDM1 patients was found to be hypermethylated; the methylation status of the region under investigation was determined for the mutant allele as well in order to confirm or

refute previous findings (Steinbach et al. 1998). Another interesting finding is the decreased amount of full-length *DMPK* transcript expressed by the normal allele in CDM1 patients, which may be the result of hypermethylation of that allele (Frisch et al. 2001).

Although hypermethylation of the CpG island in the 3' UTR of the *DMPK* gene is suspected in the decreased expression of the gene, other theories regarding this decreased expression exist. Most notable among them is the binding of the CUG repeat tracts in the mRNA by CUG binding proteins (Cho and Tapscott 2007; Ranum and Day 2004; Ranum and Cooper 2006). Two CUG binding proteins have been implicated in the pathology of DM1, muscle-blind-like 1 (*MBNL1*) and muscle-blind-like 2 (*MBNL2*). The binding of these proteins to the *DMPK* mRNA not only forms aggregates in the nucleus, but also prevents the proteins from performing their normal functions within the nucleus. *MBNL1* is involved in RNA splicing; it functions to direct the splicesome's removal of introns and the ligation of exons. *MBNL2* is believed to be involved in gene expression, since its loss in mice results in aberrant gene transcription (Ho et al. 2004; Paul et al. 2006; Warf and Berglund 2007).

1.2 Bisulfite-Conversion and Bacterial Cloning for Sequencing.

To study the methylation pattern, the methods of bisulfite conversion and bacterial cloning were used. Bisulfite conversion is a technique in which DNA is treated with sodium bisulfite prior to PCR amplification (Figure 1.2.1). The bisulfite treatment converts unmethylated cytosine (C) to uracil, while methylated cytosine (^{m}C) is unreactive. The conversion is the result of the deamination of unmethylated cytosine residues. This conversion allows one to distinguish ^{m}C from C. The treated DNA sequence is then amplified using PCR, where the converted unmethylated

cytosine/uracil is amplified as thymine, and the ^{m}C is amplified as unchanged cytosine (Boyd and Zon 2004; Carless 2009).

Bisulfite conversion does present a difficulty when using direct sequencing to determine methylation status. The problem arises because of partial conversion. Incomplete bisulfite-conversion occurs when not all unmethylated cytosines are converted into uracil. This partial conversion then produces significant background noise in direct sequencing, demonstrated by the appearance of multiple peaks from 2 different sequencing reactions which can not be distinguished from one another. In addition, treating genomic DNA with sodium bisulfite also results in severe stress on the DNA. The DNA becomes photo- and temperature-sensitive as a result of oxidative damage done during the conversion (Ehrich et al. 2007; Harrison et al. 1998) For these reasons, direct sequencing of PCR products of bisulfite-treated DNA was abandoned; rather the technique of bacterial cloning was employed. Cloning involves the integration of the PCR product into a bacterial plasmid, in this case, the bacterium, *Escherichia coli*. Once integrated into the bacterium, the plasmid replicates along with the bacterial genome every time the bacterium undergoes division. Since bacterial cells are virtual clones of the parent cell, enough plasmids will be replicated in order to isolate the DNA and sequence by use of the TOPO TA Cloning Kit for Sequencing. This technique is better suited to bisulfite-converted sequencing than direct sequencing because of incomplete conversion. Since only one PCR product is incorporated into a plasmid, and since a colony of bacterial cells is founded by a single parent, only one PCR product will be sequenced at a time. This will greatly reduce the background found in sequences and identify if partial conversion occurred, for example if the sequence is rich in the cytosine residues (the vast majority of these

should have been eliminated in the bisulfite-conversion). Comparison of the sequences generated by bacterial cloning allowed for the identification of partial conversion, since incomplete conversion would result in differing appearances of cytosine and thymine residues at the same locations.

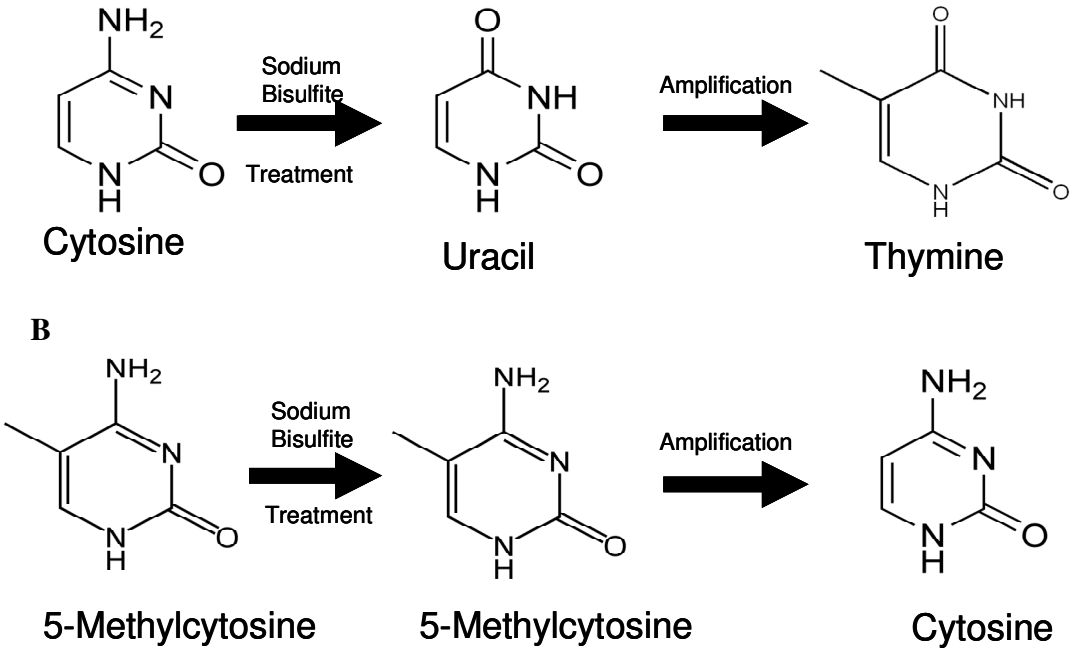


Figure 1.2.1 Explanation of Bisulfite Conversion. A: Unmethylated cytosine residues are deaminated and converted to uracil as a result of sodium bisulfite treatment. The uracil residues are then amplified as thymine residues in PCR. B: 5-methylcytosine residues are not deaminated to uracil by treatment with sodium bisulfite. As a consequence, 5-methylcytosine residues are amplified as cytosine in PCR.

1.3 Methylation as a Cause of Down-regulation of Transcription.

Since previous work had shown hypermethylation in the *DMPK* gene (Steinbach et al. 1998) and a decreased expression of both the normal and expanded *DMPK* alleles in CDM1 patients (Frisch et al. 2001), it was the scope of this study to demonstrate a correlation between methylation status of the CpG island located proximal to the CTG repeat and decreased expression of full-length *DMPK* transcripts from CDM1 patients. It was our hypothesis that the non-mutant alleles of CDM1 patients would be hypermethylated when compared to the non-mutant alleles of adult-onset DM1 patients. However, due to the evolution of the methods used in this study and high failure rates of the individual experiments, the methylation status of the CpG island was only examined in several CDM1 patients. Additional research is required to examine the methylation status of adult-onset DM1 patients, and *DMPK* expressional analysis of both CDM1 and adult-onset patients.

Chapter 2

MATERIALS AND METHODS

2.1 Cell Lines and Expansion Diagnosis.

In order to ascertain the methylation status of the CpG island located in the 3' UTR of *DMPK*, the methods of bisulfite conversion, PCR, bacterial cloning, and sequencing were used. These series of experiments were begun with the isolation of genomic DNA from three human-hamster hybrid cell lines from either DM1 or CDM1 patients. It is noteworthy to mention that the hybrid cell lines were established by this laboratory in a previous investigation (Funanage et al. 1997), and were constructed by fusion of skin fibroblasts with the Chinese hamster ovary cell line DXB11 (Carango et al. 1993). The cell lines used include: DM13 and DM1500, which carry (CTG)13 (normal) and (CTG)1500 repeats, respectively, from an adult-onset DM1 patient (19). The CDM14 andCDM2300 cell lines contain (CTG)14 (normal) and (CTG)2300 DMPK alleles, respectively, from a CDM1 patient. Hybrid cell lines NL5 and NL25 carry (CTG)5 (normal) and (CTG)25 alleles, respectively, from a patient who presented with hypotonia and proximal muscle weakness. A high fidelity PCR system was used to amplify regions of approximately 250 base pairs in length, and bacterial cloning for sequencing was used to analyze the PCR products. The single allele nature of the cells allowed for easy clarification of the methylation status of the mutant or normal allele. Diagnostic tests were performed on all isolated DNA samples to confirm the CTG repeat number by performing PCR and examination of the product

sizes on a 2% agarose gel stained with ethidium bromide. The resulting band sizes were estimated in comparison to a patient with known expansion sizes and a molecular weight standard via the Kodak Molecular Imaging software (Figure 2.1.1).

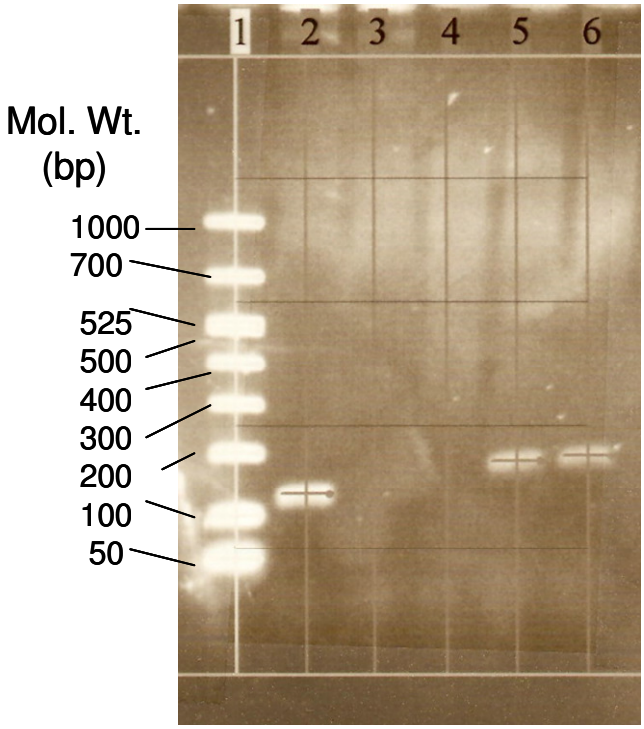


Figure 2.1.1 Example of the Diagnosis of Expansion Size in Cell Lines. The above picture is the diagnosis of the CTG expansion size within cells used. Lane 1 is the molecular weight marker and Lane 2 is the control, possessing five CTG repeats. Lanes 2 and 3 are CDM2300 and lanes 4 and 5 are CDM14. Since the expansion size of the control was known, its molecular weight could be used to ascertain the expansion size of the cell lines used.

2.2 Bisulfite-Conversion.

For the bisulfite conversion, several manufacturers' kits were used, including Qiagen (Germantown, MD) EpiTect Bisulfite, Zymo Research (Orange, CA) EZ-DNA Methylation kit, and Zymo Research EZ-DNA Methylation-Gold kit, in addition, to a "Homebrew" method as described by Warnecke et al. (2002). All of these procedures rely upon the treatment of genomic DNA with sodium bisulfite. The sodium bisulfite deaminates unmethylated cytosine residues, resulting in the conversion to a uracil residue. Methylated cytosine residues are not deaminated. All the methods used resulted in approximately equal quantities and qualities of DNA; however, the EZ-DNA Methylation-Gold kit (Figure 2.2.1) was chosen because of its comparatively shorter conversion time and similar quality of final DNA. Approximately 500 ng of genomic DNA were used per conversion reaction.

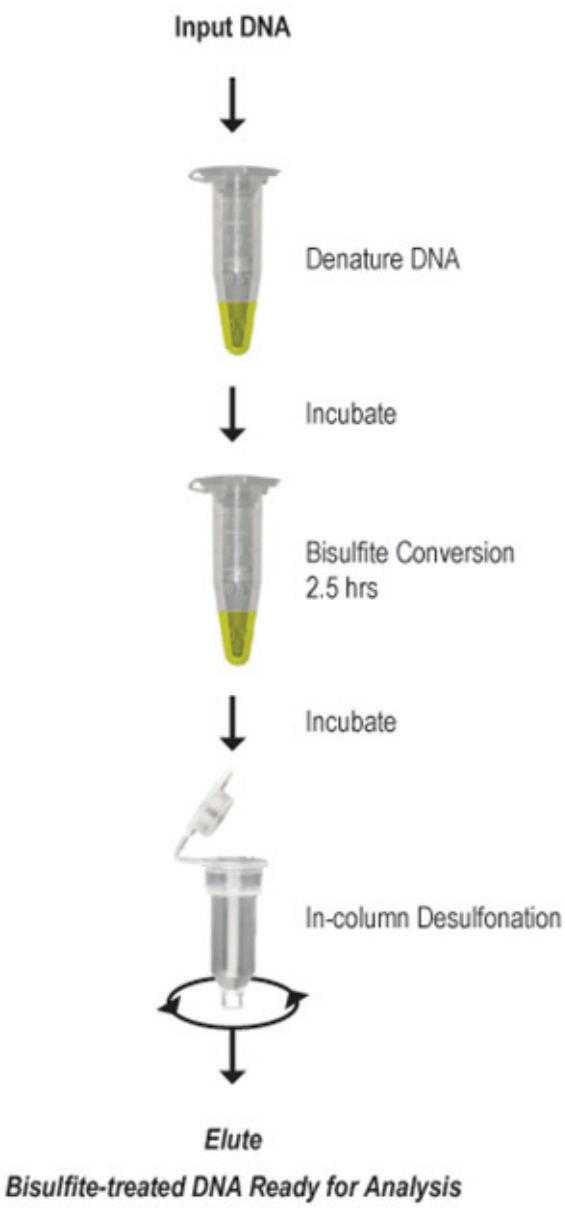


Figure 2.2.1 EZ-DNA Methylation-Gold. This flow diagram depicts the steps in the kit's conversion process, and the length of incubation for conversion. The total time needed to complete assay was 4 hours.

2.3 PCR Amplification.

Following the bisulfite conversion, the converted DNA is used to PCR amplify the region surrounding the area of concentrated CpG islands. Since the region is approximately 1.2 kilobases in length, it was reduced into smaller regions to ease investigation. At first, only a 450 base pair region was examined (Figure 2.3.1), but later, the entire CpG island was divided into 4 regions of approximately 250 base pairs in length. Primer pairs were created for each of the regions, and PCR conditions were optimized (Tables 2.3.1, 2.3.2, and 2.3.3). With certain pairs, touchdown PCR was used to amplify the desired region. Touchdown PCR follows the same pattern of normal PCR, initial denautration stage followed by cycles of denaturation, annealing, and extension, and a final extension time, however, it involves an initial cycling period in which the annealing temperature is dropped incrementally to the optimal annealing temperature of the primer pair being used. In the cases of the primer pairs used in this study, the annealing temperature was dropped in 1 ^{o}C increments for a total of five degrees. Once the optimal annealing temperature was reached, a second set of cycles is begun, in which the extension period is increased by 0:15 seconds for each subsequent cycle. The kit chosen for the amplification was the Roche (Indianapolis, IN) High Fidelity PCR system; this system utilizes not only a *Taq* polymerase, but also an enzyme that possesses a 3' to 5' exonuclease, which allows for proofreading of the PCR product. Once amplified, the product sizes were confirmed by examination on a 2% agarose gel stained with ethidium bromide. Depending upon whether or not multiple bands were present per PCR reaction, the bands were either excised directly from the gel and cleaned using the Qiaquik PCR cleaning kit (Qiagen), or the PCR products were cleaned directly using the same kit.

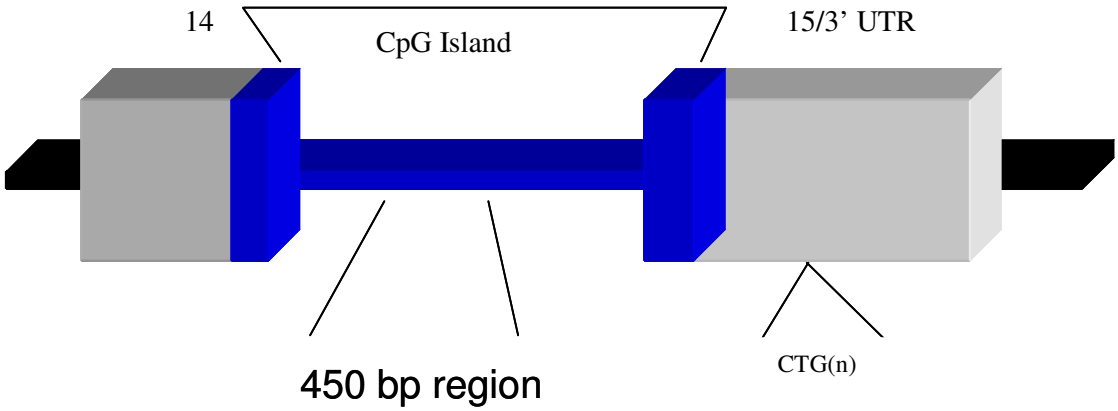


Figure 2.3.1 Initial Amplification Region of 450 Base Pairs. The initial amplification region of 450 bp was located approximately in the center of the CpG island. This would later be abandoned in favor of four smaller, approximately 250 bp, regions spread throughout the entire CpG island.

Table 2.3.1 Primers used for PCR Amplififcation.

Region/Pair	Primer Name	Primer Sequence	Annealing Temp. (^{o}C)
1	F167	CTTYGTTGGTGGGGTTAGGTTTTATGTAT	55
	R437	CAACAACCTACAACTATTATTAATCCACT	
2	F404	GTGYGAGTGGATTAATAATAGTTGTAGG	59
	R653	ACRTAAACCACCAACCCAATACAACCCA	
3	F619	GTYGTTTTGGGTTGTATTGGGTTGGTGG	61
	R848	AACCCRATACACAAAACTAAAACTAAAC	
3A	F626	TGGGTTGTATTGGGTTGGTGGTTTAYG	52
	R838	ACAAAACTAAAACTAAACRAAAACCCAC	
4	F811	GTGGGTTTTYGTTTAGTTTTAGTTTTGT	63
	R1035	CCCAACCTACAATTTACCCATCCACRT	

Table 2.3.2 PCR Conditions for Primer Pair/Regions 1,2, and 4.

			Touchdown (5 Cycles: decrease 1 °C of annealing per cycle)						Extension (additional 15 s of extension per cycle: 25 cycles)								
Primer Pair/Region	**Initial Denaturation**		**Denaturation**		**Annealing**		**Extension**		**Denaturation**		**Annealing**		**Extension**		**Final Extension**		**Hold**
	Time	°C	Time	°C	Time	°C	Time	°C	Time	°C	Time	°C	Time	°C	Time	°C	°C
1	5:00	95	:30	95	:30	60	:45	72	:30	95	:30	55	:45	72	6:00	72	4
2	5:00	95	:30	95	:30	64	:45	72	:30	95	:30	59	:45	72	6:00	72	4
4	5:00	95	:30	95	:30	65	:45	72	:30	95	:30	61	:45	72	6:00	72	4

Table 2.3.3 PCR Conditions for Primer Pair/Region 3 and 3A.

			30 Cycles								
Primer Pair/Region	**Initial Dentauration**		**Denaturation**		**Annealing**		**Extension**		**Final Extension**		**Hold**
	Time	^{o}C	Time	^{o}C	Time	^{o}C	Time	^{o}C	Time	^{o}C	^{o}C
3	5:00	95	:30	95	:30	61	:45	72	6:00	72	4
3A	5:00	95	:30	95	:30	52	:45	72	6:00	72	4

2.4 Bacterial Cloning and Sequencing.

Following the purification of the PCR products, the TOPO TA Cloning Kit for Sequencing by Invitrogen (San Diego, CA) was used (Figure 2.4.1). Once the PCR products had been inserted into bacterial plasmids, the bacteria were grown on LB agar plates with carbomycin (50ng/ml) for 24 hours at 37^{o}C. All colonies present were screened for correct insert size using primers M13F and M13R supplied with the TOPO TA Cloning Kit (Table 2.4.1). Only recombinant bacterial clones grew due to a kill gene that is inactivated by the inserted PCR product. Colonies that had the correct insert size were expanded in LB broth with carbomycin at 60 ng/ml for a period of 14-16 hours with aeration at 37^{o}C. The plasmids of the expanded colonies were then isolated using the Qiagen QiaQuik Mini Prep Kit. A minimum of 30 plasmids demonstrating the correct insert size were isolated per cloning reaction or PCR product; a range of 30-36 was usually isolated and analyzed.

Table 2.4.1 Primers used for Bacterial Colony Screening.

Primer Name	Primer Sequence	Annealing Temp. (^{o}C)
M13F	GTAAAACCACGGCCAG	55
M13R	CAGGAAACAGCTATGA	55

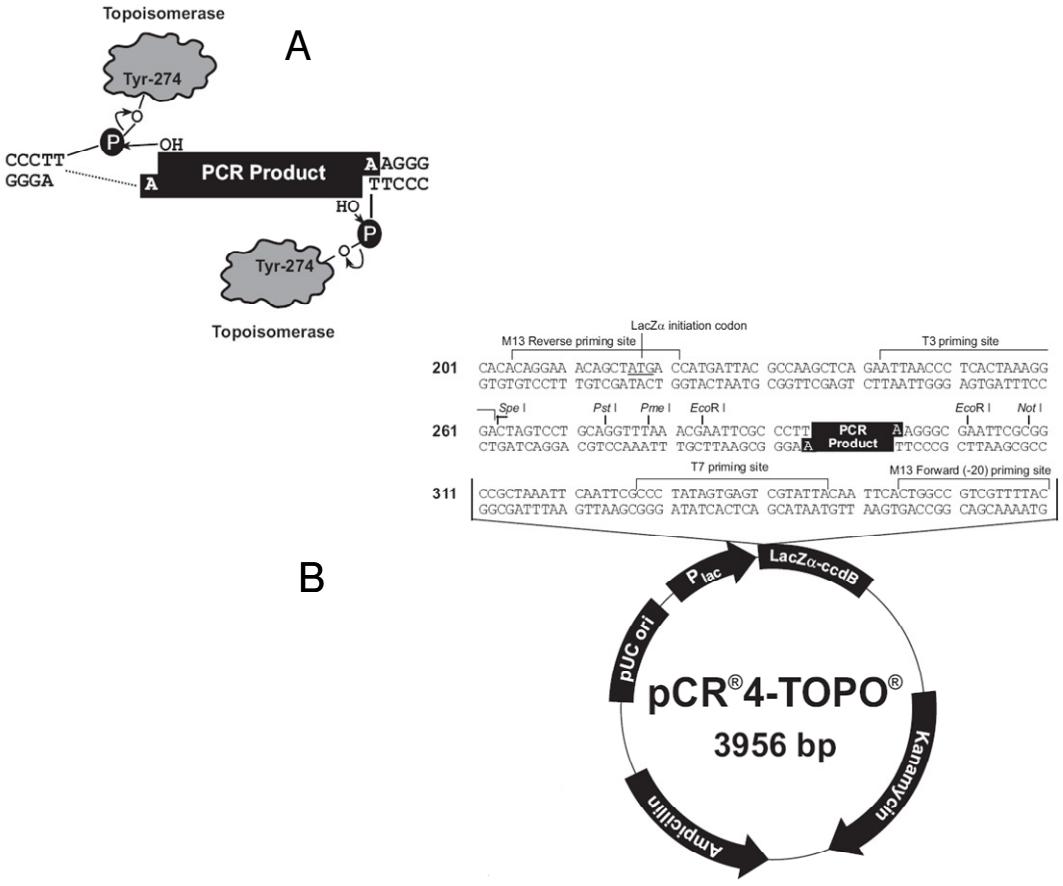


Figure 2.4.1 PCR Product Insertion. A. The PCR product possesses non-template based adenine residues on its 3' ends. These residues were added by the polymerase. Topoisomerase uses these adenine residues to insert the PCR product into the bacterial plasmid. B. Once inserted, it disrupts the kill gene present, thereby ensuring the survival of bacteria that have successfully undergone insertion.

The plasmids were sequenced at a concentration of 400 ng/ul. Sequencing was performed in only the reverse orientation using Big Dye Version 3.1 (Applied Biosystems, Foster City, CA). Sequencing reactions were analyzed using an Applied Biosystems 3130 Analyzer. All chromatograms were visually inspected for retention of cytosine residues at CpG locations prior to alignment to the converted *DMPK* reference sequence (L08835) using MacVector Software (Cary, NC). This reference sequence represented a CpG island in which all cytosines except those in a CpG dinucleotide, were assumed unmethylated, and therefore converted to thymines; cytosines in a CpG dinucleotide were left unchanged. Most commonly, cytosine residues within the dinucleotide CG are the only cytosine residues methylated. As is the case, if these cytosine residues are methylated they will appear as both a CG in the forward and reverse sequencing orientation. However, if the cytosine residues in a CG are unmethylated they will appear as a TG in the forward oriented sequence and a CA in the reverse oriented sequence.

Chapter 3

RESULTS

3.1 DNA Stability Following Conversion

To test the hypothesis that methylation status of the CpG island in the 3' UTR of the *DMPK* gene was responsible for the expression difference between normal and mutant alleles in DM1 and CDM1 patients, several bisulfite conversion kits and one "homebrew" method were first investigated. Starting with approximately equal quantities of DNA (0.5-1.0 μg), all methods resulted in relatively equal quantities of converted DNA (<0.5 μg). It became apparent that the converted DNA samples could tolerate only two to three freeze-thaws before the DNA was no longer able to be amplified (Figure 3.1.1). Therefore, the Zymo Research Methylation Gold Kit was chosen on the basis of its shorter conversion time (Table 3.1.1.).

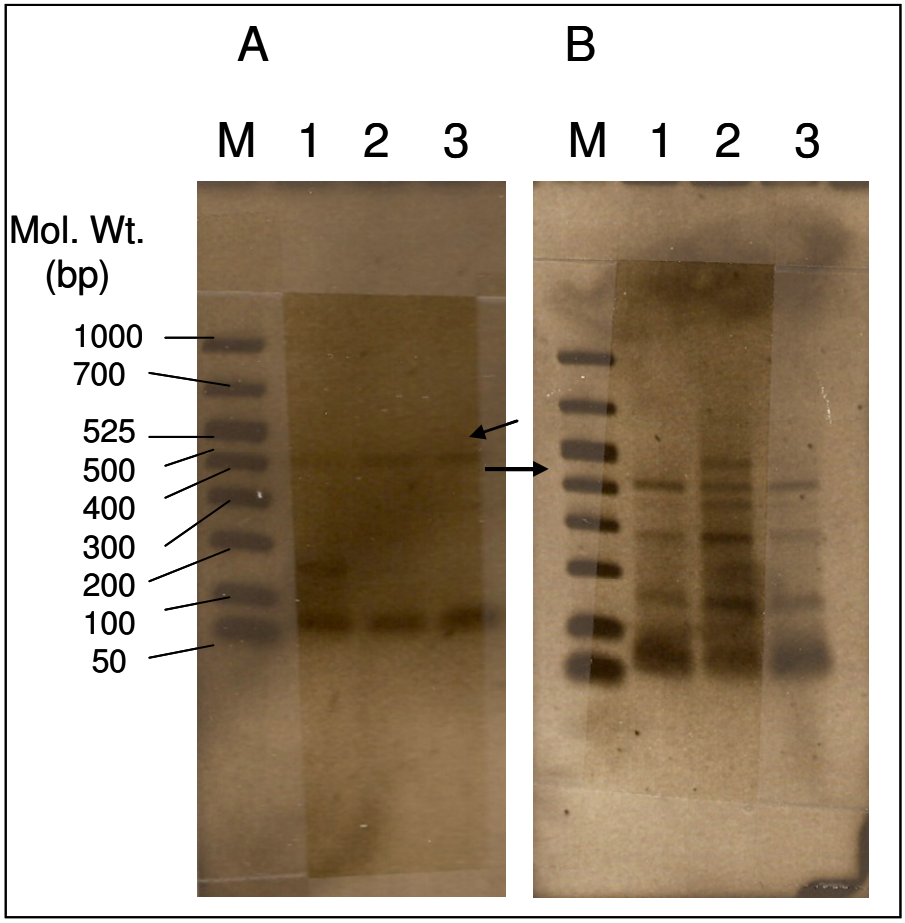


Figure 3.1.1 Comparison of PCR Quality Following Different Number of Freeze Thaws. The gel picture shows DNA that was converted on 5/11/2009. The arrows indicate the desired 450 bp product. A. The DNA in this gel picture has undergone a single freeze thaw. The desired bands are easily seen with minimal smearing beneath. Lanes 1 is CDM14, lane 2 is CDM2300, lane 3 is NL5, and lane M is the molecular weight marker. B. The image demonstrates three freeze thaw cycles. There is considerably more smearing and the appearance of undesired bands prevented easily read sequences from being produced by direct sequencing Lanes 1 is CDM14, lane 2 is CDM2300, lane 3 is NL5, and lane M is the molecular weight marker.

Table 3.1.1 Comparison of Incubation and Total Completion Times for Four Different Methods of Bisulfite Conversion.

Conversion Method	Incubation Time	Total Time
Qiagen's EpiTect Bisulfite	5 hours	6 hours
Zymo Research's EZ-DNA Methylation-Direct	5 hours	6 hours
Zymo Research's EZ-DNA Methylation-Gold	3 hours	4 hours
Homebrew	4 hours	36 hours

3.2 Difficulties Encountered during Amplification and Direct Sequencing.

Initial experiments, using the bisulfite-converted DNA, attempted to amplify a 450 base pair region of the CpG island. However, PCRs of this region often resulted in either lack of product or smearing of PCR product on 2% agarose gels, or the generation of unintended products, as demonstrated by the appearance of bands not 450 base pairs in length on 2% agarose gels (Figure 3.2.1).

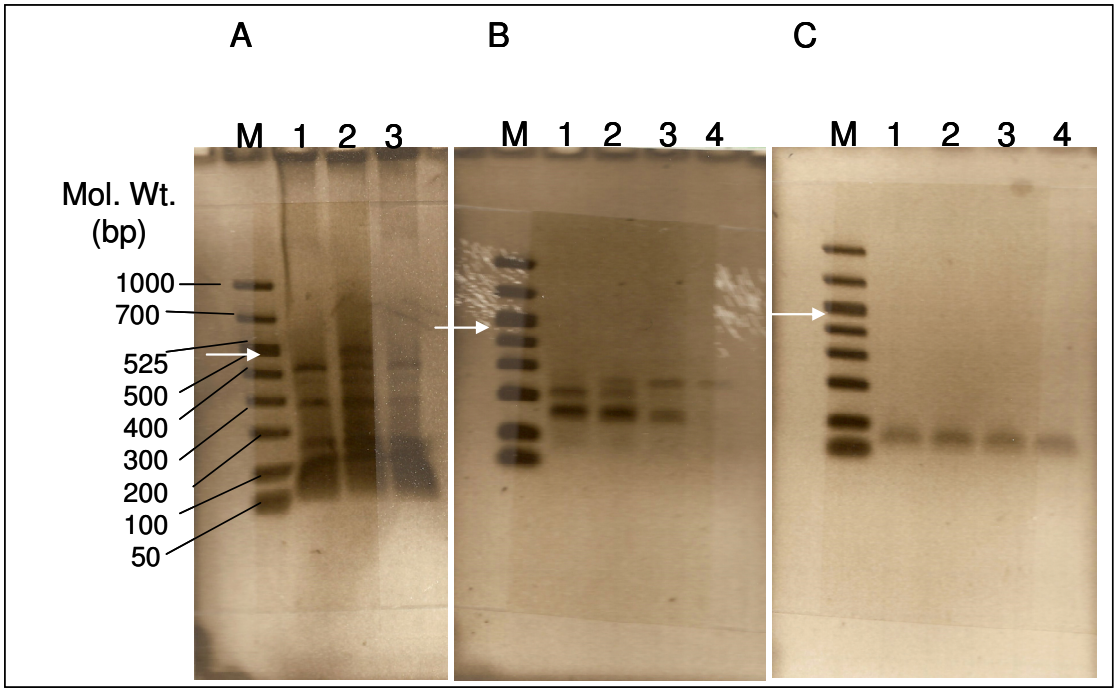


Figure 3.2.1. Gel Photographs Demonstrating Difficulties Encountered with Amplification of 450 Base Pair Region. Desired product sizes indicated with white arrows. A. Smearing of PCR product. Lane 1 is cell line CDM14, lane 2 is cell line CDM2300, lane 3 is cell line NL25, and M is the molecular weight marker. B. Appearance of spurious bands. Lane 1 is cell line CDM14, lane 2 is cell line CDM2300, lane 3 is cell line NL25, lane 4 is cell line NL5 and M is the molecular weight marker. C. No desired bands. Lane 1 is cell line CDM14, lane 2 is cell line CDM2300, lane 3 is cell line NL25, lane 4 is cell line NL5 and M is the molecular weight marker.

3.3 Primer Redesign and Bacterial Cloning for Sequencing.

Since numerous difficulties were encountered with trying to amplify a 450 base pair region of the CpG island a primer redesign was undertaken. The CpG island was divided into four regions of approximately 250 base pairs each; the largest region was 270 base pairs and the smallest was 225 base pairs (Fig. 3.3.1). The use of the

new primers and amplification of smaller regions resulted in a dramatic decrease in the amount of product smearing and generation of spurious bands (Figure 3.3.2). All PCR amplifications were performed using the Gold-Taq polymerase. Direct sequencing, again, resulted in significant background noise and co-sequencing of multiple PCR products.

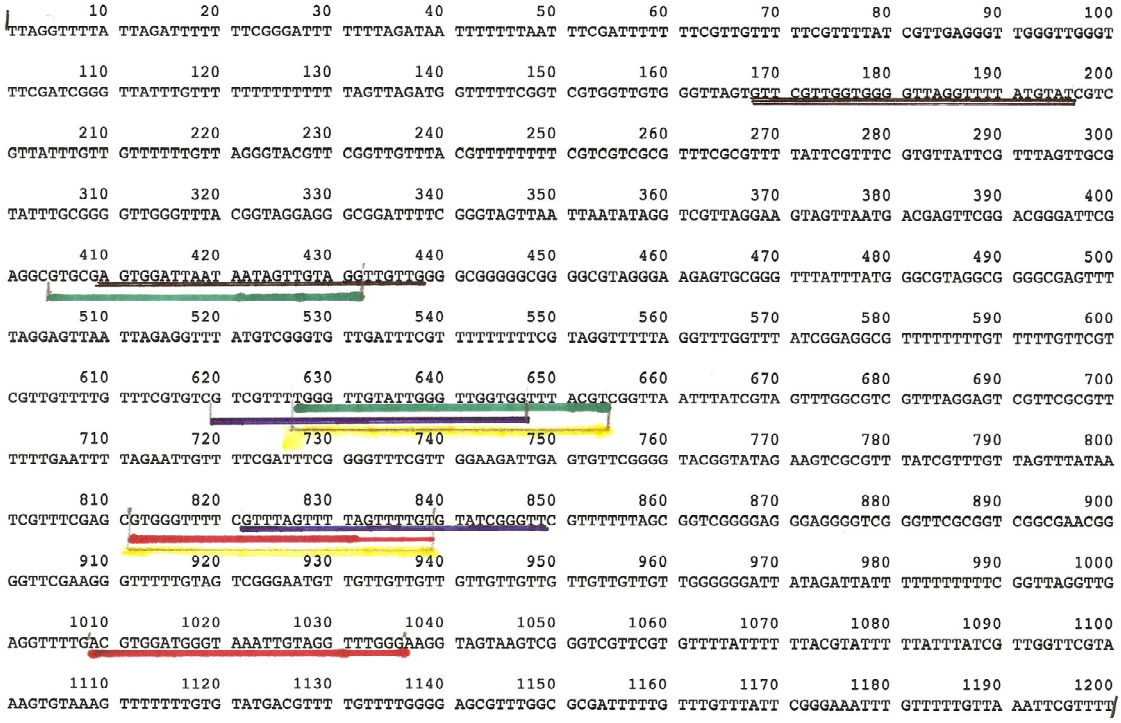
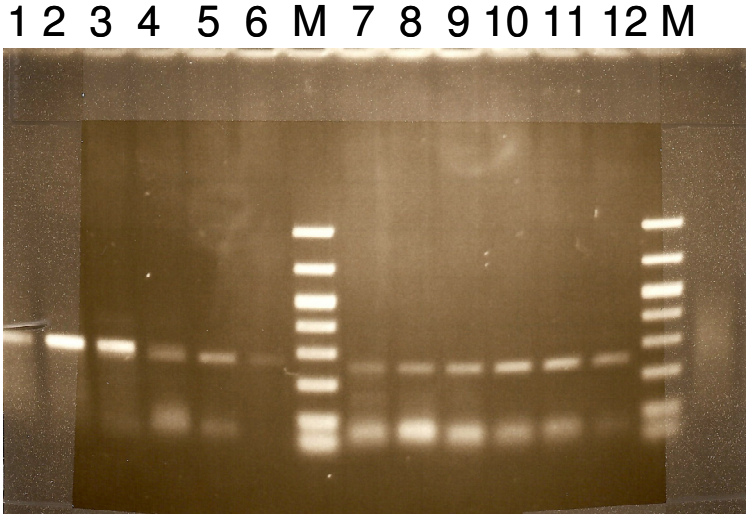


Figure 3.3.1. ***DMPK*** **CpG Island Broken Down into Four Regions of Approximately 250 Base Pairs. Five pairs of primers were generated for the smaller regions to be amplified. The first, underlined in black, amplified the first region at a product length of approximately 270 base pairs. The second pair, underlined in green, amplified the second region at a product length of 250 base pairs. The third region possesses two pairs. The initial third region pair, underlined in blue, amplified a region of 230 base pairs; however this pair was unable to produce a product. So a second pair for the region, designated Pair-3A and underlined in yellow, was generated and amplified a region of 213 base pairs. The final and fourth region, underlined in red, amplified a region of 225 base pairs.**

A



B

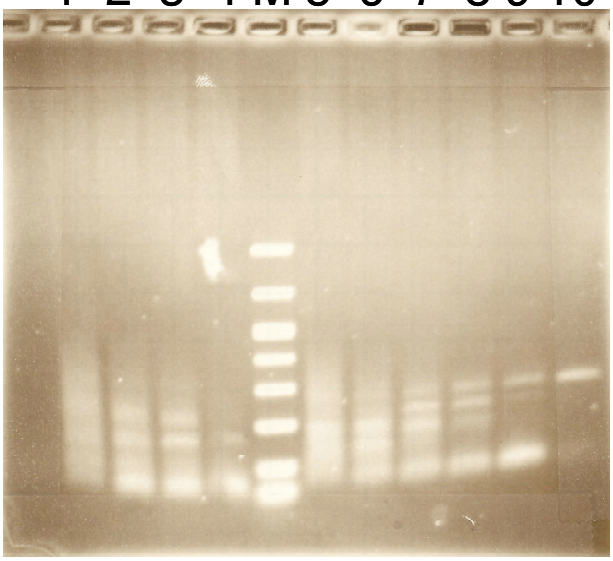


Figure 3.3.2 Amplification of the CpG Island using the Primer Pairs that Generated Products of 213 to 270 Base Pairs in Size. Although there is significant smearing and spurious band production, these products represent the variation of annealing temperatures to optimize PCR conditions for each primer pair. In comparison to Fig. 3.2.1, there is less smearing and spurious band production. The sole cell line used in these optimizations was CDM14. A. Lanes 1-6 are primer pair 1. Lanes 7-12 are primer pair 2. M is the molecular weight marker. B. Lanes 1-4 are primer pair 3. Lanes 5-10 are primer pair 4.

Bacterial cloning of PCR products for sequencing was used in order to obtain clearer sequences, sequences without the abundance of co-sequencing, which indicated the presence of converted and unconverted DNA. The use of cloning resulted in sequences showing little to no background noise and eliminated the issue of co-sequencing (Figure 3.3.3).

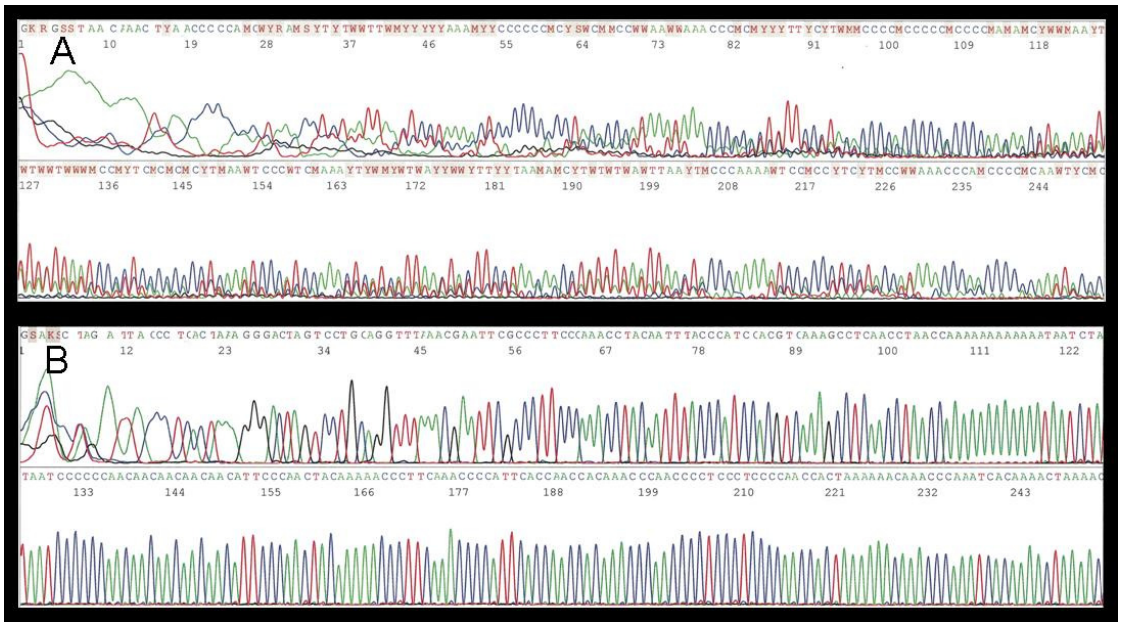


Figure 3.3.3 Comparison of Sequencing Results between Direct Sequencing of PCR Products and Bacterial Cloning for Sequencing. A. Chromatogram showing direct sequencing of amplified bisulfite-converted DNA, cell line CDM14; there is significant co-sequencing present. B. Sequence after bacterial cloning of amplified bisulfite converted DNA, CDM14; there is little to no co-sequencing or background.

3.4 No Detected Methylation.

Though the use of bacterial cloning for sequencing has resulted in significantly cleaner sequences, no methylation was detected in any of the six cell lines used. The lack of methylation was demonstrated by the failure of CGs to be seen in any converted, amplified, and sequenced cell lines; instead only CAs were visible at locations where CpG dinucleotides were known to be located (Figure 3.4.1).

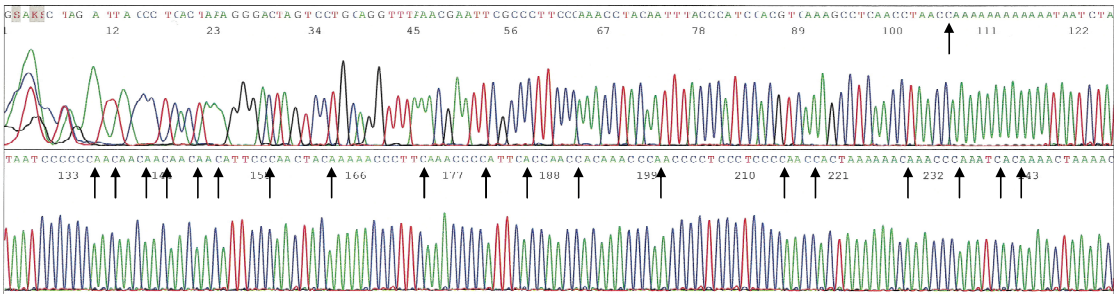


Figure 3.4.1 Sample Chromatogram Showing Lack of Methylation. With the use of bacterial cloning for sequencing, sequences were significantly clearer, but no methylation was detected in any of the cell lines used. The above sequence is represents the normal allele of the CDM1 patient, CDM14. The arrows indicate locations of CpGs, however the lack of methylation is indicated by the presence of CAs since the sequence is in the reverse orientation.

Chapter 4

DISCUSSION

Methylation of the CpG island located proximal to the triplet repeat of the *DMPK* gene was not detected in either the normal or mutant alleles of one CDM1 patient who was screened. This information therefore does not support the hypothesis that methylation of this CpG island is related to the lower expression levels of the non-mutant allele of CDM1 patients as was previously reported by Frisch et al. (2001). Interestingly, the lack of methylation of this region in CDM1 patients is in contrast to a report by Steinbach et al. (1998). A possible explanation for this difference is we examined the methylation status in proliferating cells (somatic cell hybrids), whereas Steinbach et al. examined muscle tissue from CDM1 patients. Funanage and Frisch (unpublished data) have shown that the expression of the *DMAHP/SIX5* gene is not decreased in proliferating hybrid cell lines, but is decreased once the cells terminally differentiate by transfection of MyoD into cells and conversion into muscle cells. Further work is needed to determine if methylation of the CpG island in CDM1 patients differs in proliferating vs. non-proliferating cells.

As was reported in the results, this investigation was marked by the difficulty in obtaining the desired full-length PCR products, whether they were the initial 450 base pair product or the smaller 250 base pair ones. The difficulty in obtaining such products is most likely the result of the use of bisulfite conversion. The treatment of DNA with sodium bisulfite causes a relatively high degree of oxidative stress. The stress imposed upon the DNA then makes it photosensitive, as well as

sensitive to freeze-thaws. Furthermore, the conversion process also results in the degradation of the DNA; this degradation can reach as high as 90% loss of the starting material. Much of the degradation is a consequence of depurination of the DNA, which causes breaks in the strand. Lastly, partial conversion was also a concern with the use of this technique. Partial conversion, again, is the failure of the sodium bisulfite treatment to deaminate all non-methylated cytosine residues (Ehrich et al. 2007; Harrison et al. 1998). Failure to do so not only results in detecting falsely positive methylated cytosine residues in the sequences, but also difficulties in the annealing of primers to the converted strands of DNA. Partial conversion is often the result of either too much initial DNA or insufficient incubation times. However, these two issues should have been avoided when kits were used for the conversion; the amount of DNA used was less than the maximum amount of DNA suggested by the kit's manufacturer and the incubation times given by the manufacturer were followed.

With these shortcomings of bisulfite conversion in mind, it is understandable as to why early attempts to amplify and directly sequence the converted DNA failed. The damaged DNA would have offered numerous partial strands for the polymerases to amplify, thereby causing the appearance of spurious bands on the agarose gels. This problem becomes more apparent when the length of the desired product is increased, which then increases the chance of a break in that strand. Similarly, the partially converted DNA may have also provided numerous sequences for the primers to bind and amplify, again resulting in un-wanted PCR products. The issue with photosensitivity and thermo-sensitivity were easily addressed, however, by decreasing the exposure to light and the number of usages of the converted DNA to 2-3 freeze-thaws. On a similar note, the presence of partial

products and products of partial conversion resulted in the high background noise and co-sequencing when the PCR products were directly sequenced. Sequencing primers would bind these undesired products and sequence them along with the desired products.

Switching to bacterial cloning for sequencing avoided the aforementioned problems. The method allowed not only for selection of products displaying the correct size, but also sequencing of individual PCR products of the correct size, therefore eliminating the presence of background noise and co-sequencing in the generated sequences. Occasionally, methylated CpGs were detected in the sequences, but were not considered significant for several reasons: the methylation was detected in only 1 of 30 or more sequences of different clones derived from the same PCR product, other mismatches occurred in the sequence, and different CpGs appeared methylated in different sequences. These errors could be attributed to either partial conversion or an error in the replication of the plasmid by the bacteria.

Since methylation of the normal *DMPK* allele was not detected in CDM1 patients, it cannot explain the expression difference of normal *DMPK* alleles in DM1 and CDM1 patients as was noted by Funanage et al (1997). Therefore, future research should first compare the methylation status of this region in proliferating cells with that of cells that have reached senescence. In addition, the methylation status of the 5' region of the *DMPK* gene, as well as the promoter or any enhancer regions, should be examined. These investigations would rely on nearly identical methods employed by this project. Additionally, the chromatin structure in the region of *DMPK* could also be investigated. Such studies would be concerned with methylation and/or acetylation of histone proteins and the resulting chromatin compaction.

Though this study did not result in the detection of methylation in the 3' UTR of the *DMPK* gene, it did allow for the optimization of methods to detect such methylation, particularly the use of a high fidelity PCR system to amplify regions of approximately 250 base pairs in length and bacterial cloning for sequencing, which will enable future investigation into the intriguing possibility that methylation of the CpG island in CDM1 depends on the developmental stage of tissues, leading a new therapeutic avenue into treatment of this disease.

REFERENCES

Myotonic Dystrophy [Internet]; c2006 [cited 2009 11/13].

DMPK [Internet]: U.S. National Library of Medicine; c2006. Available from: http://ghr.nlm.nih.gov/gene=dmpk .

TOPO TA Cloning Kit for Sequencing [Internet]: Invitrogen. Available from: http://www.invitrogen.com .

Ansved T, Edstrom L, Grandell U, Hedberg B, Anvret M. 1997. Variation of CTG-repeat number of the DMPK gene in muscle tissue. Neuromuscul Disord 7(3):152-5.

Boyd VL and Zon G. 2004. Bisulfite conversion of genomic DNA for methylation analysis: Protocol simplification with higher recovery applicable to limited samples and increased throughput. Anal Biochem 326(2):278-80.

Carango P, Noble JE, Marks HG, Funanage VL. 1993. Absence of myotonic dystrophy protein kinase (DMPK) mRNA as a result of a triplet repeat expansion in myotonic dystrophy. Genomics. 18(2):340-8.

Carless M. 2009. Investigation of genomic methylation status using methylation-specific and bisulfite-specific polymerase chain reaction. In: Chromatin protocols. Human Press. 217 p. Clifton, NJ.

Chi LM and Lam SL. 2005. Structural roles of CTG repeats in slippage expansion during DNA replication. Nucleic Acids Res 33(5):1604-17.

Cho DH and Tapscott SJ. 2007. Myotonic dystrophy: Emerging mechanisms for DM1 and DM2. Biochim Biophys Acta 1772(2):195-204.

De Temmerman N, Seneca S, Van Steirteghem A, Haentjens P, Van der Elst J, Liebaers I, Sermon KD. 2008. CTG repeat instability in a human embryonic stem cell line carrying the myotonic dystrophy type 1 mutation. Mol Hum Reprod 14(7):405-12.

Ehrich M, Zoll S, Sur S, van den Boom D. 2007. A new method for accurate assessment of DNA quality after bisulfite treatment. Nucleic Acids Res 35(5):e29.

England R, Petterson M. Pyro Q-Cp: quantitative analysis of the methylation in multiple CpG sites by pyrosequencing [Internet]: Nature Methods; c2005. Available from: http://www.nature.com/naturemethods .

Frisch R, Singleton KR, Moses PA, Gonzalez IL, Carango P, Marks HG, Funanage VL. 2001. Effect of triplet repeat expansion on chromatin structure and expression of DMPK and neighboring genes, SIX5 and DMWD, in myotonic dystrophy. Mol Genet Metab 74(1-2):281-91.

Funanage VL, Frisch R, Singleton KS, Carango P, Moses PA, Marks HG. 1997. Effect of CTG repeat expansion on chromatin structure and processing of DMPK mRNA in hybrid cell lines derived from myotonic dystrophy patients. Amer J H Gen 61:A308.

Harrison J, Stirzaker C, Clark SJ. 1998. Cytosines adjacent to methylated CpG sites can be partially resistant to conversion in genomic bisulfite sequencing leading to methylation artifacts. Anal Biochem 264:129-32.

Ho TH, Charlet-B N, Poulos MG, Singh G, Swanson MS, Cooper TA. 2004. Muscleblind proteins regulate alternative splicing. EMBO J 23(15):3103-12.

Huang CC and Kuo HC. 2005. Myotonic dystrophies. Chang Gung Med J 28(8):517-26.

Li B, Goyal J, Dhar S, Dimri G, Evron E, Sukumar S, Wazer DE, Band V. 2001. CpG methylation as a basis for breast tumor-specific loss of NES1/kallikrein 10 expression. Cancer Res 61(21):8014-21.

Longman C. 2006. Myotonic dystrophy. J R Physicians 36:51-55.

Malumbres M, Perez de Castro I, Santos J, Fernandez Piqueras J, Pellicer A. 1999. Hypermethylation of the cell cycle inhibitor p15INK4b 3'-untranslated region interferes with its transcriptional regulation in primary lymphomas. Oncogene 18(2):385-96.

Orengo JP, Chambon P, Metzger D, Mosier DR, Snipes GJ, Cooper TA. 2008. Expanded CTG repeats within the DMPK 3' UTR causes severe skeletal muscle wasting in an inducible mouse model for myotonic dystrophy. Proc Natl Acad Sci USA 105(7):2646-51.

Paul S, Dansithong W, Kim D, Rossi J, Webster NJ, Comai L, Reddy S. 2006. Interaction of muscleblind, CUG-BP1 and hnRNP H proteins in DM1-associated aberrant IR splicing. EMBO J 25(18):4271-83.

Ranum LP and Cooper TA. 2006. RNA-mediated neuromuscular disorders. Annu Rev Neurosci 29:259-77.

Ranum LP and Day JW. 2004. Myotonic dystrophy: RNA pathogenesis comes into focus. Am J Hum Genet 74(5):793-804.

Steinbach P, Glaser D, Vogel W, Wolf M, Schwemmle S. 1998. The DMPK gene of severely affected myotonic dystrophy patients is hypermethylated proximal to the largely expanded CTG repeat. Am J Hum Genet 62(2):278-85.

Tooke N, Petterson M. CpG methylation in clinical studies: utility, methods, and quality assurance [Internet]: IVD Technology; c2004. Available from: http://www.devicelink.com/ivdt .

Umetani N, de Maat MF, Sunami E, Hiramatsu S, Martinez S, Hoon DS. 2006. Methylation of p16 and ras association domain family protein 1a during colorectal malignant transformation. Mol Cancer Res 4(5):303-9.

Wang G, Hu X, Lu C, Su C, Luo S, Luo ZW. 2008. Promoter-hypermethylation associated defective expression of E-cadherin in primary non-small cell lung cancer. Lung Cancer 62(2):162-72.

Warf MB and Berglund JA. 2007. MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. RNA 13(12):2238-51.

Warnecke PM, Stirzaker C, Song J, Grunua C, Melki JR, Clark SJ. 2002. Identification and resolution of artifacts in bisulfite sequencing. Methods 27:101-107.