5-2016

Fluorescence Analysis of Expanded Trinucleotide Repeat Heterogeneity after in vitro PCR Amplification

Kidane M. Tewolde
University at Albany, State University of New York

Fluorescence Analysis of Expanded Trinucleotide Repeat Heterogeneity after *in vitro* PCR Amplification

An honors thesis presented to the
Department of Biological Sciences
University at Albany, State University Of New York
in partial fulfillment of the requirements
for graduation with Honors in Biological Sciences and
graduation from The Honors College.

Kidane M. Tewolde
Research Mentor: Pan T.X. Li, Ph.D.
Research Advisor: Sho-Ya Wang, Ph.D.
April 2016

Abstract:

The expansion of trinucleotide repeats in human genomic DNA manifests into multiple neurodegenerative diseases (Amrane et al., 2005). At least nine human diseases stem from the expression of expanded trinucleotide repeats (Box, 2007). Simple sequences such as, 5'CAG/CTG'3 repeats, contain a potential pathogenicity once expanded past their original lengths; which is exemplified in Huntington's Diseases and Myotonic Dystrophy. Once expanded CAG repeats are transcribed into RNA, these transcripts are translated, and the mal functioning proteins can lead to severe cell damage. Furthermore, the repeats can be passed down through generations, and in each germline, continuously expanding, gaining more repeated units (Orr & Zoghbi, 2007)-(Hartenstine, Goodman, & Petruska, 2000). Moreover, an increasing length of CAG repeats is directly correlated with an increase in the severity of the aforementioned diseases (Orr & Zoghbi, 2007)-(Hartenstine et al., 2000). In addition, the mechanisms by which elongated CAG repeats operate *in vivo* remains unknown; however, past literature denotes structural and functional aspects of CAG tracts (Zhang, Huang, Gu, & Li, 2012)-(Broda, Kierzek, Gdaniec, Kulinski, & Kierzek, 2005). These past studies were performed on shorter repeat lengths of 10-80, while *in vivo*, CAG sequences reach over 100 repeats in length (Petruska, Arnheim, & Goodman, 1996)-(Hartenstine, Goodman, & Petruska, 2002). The heterogeneity of these repeats reflect the innate replication difficulties present in *in vitro* studies and physiological conditions in Huntingtins and Myotonic Dystrophy (Lin & Wilson, 2011)- (Cooper, 2009). Petruska hypothesizes that when CAG repeats are being replicated, the strand forms structures that facilitated Taq's ability to jump over sequences and leave shortened replicons (Petruska et al., 1996). Furthermore, we aim to detect the heterogeneity of $(CAG)_n$ products after PCR amplification. By initially amplifying repeats 10-150 with long primers, we are able to verify the change in CAG sequence length after cutting the primers off with an enzymatic double digestion. In addition, $(CAG)_{10\text{-}150}$ were amplified once more with smaller series of primers: this enabled us to view how increasing CAG length induces a more heterogeneous product after PCR.CAG 10 through 60 show strong bands on agarose gels; however, the bands become less intense, and eventually degrade into complete smears with increasing length. This is seen in repeats 100, to 120, and then 150. The small primers allow us to view this degradation. The smear of $(CAG)_{150}$ displays a variety of repeats at different lengths; and a way to detect this variety is through a binding assay. SYBR Green is a chemical agent that fluoresces once bound to double stranded DNA. Doxorubicin- also called Daunomycinone- a well-known anticancer drug, naturally fluoresces. Once bound to double stranded DNA, Doxorubicin's fluoresces is quenched (Das & Kumar, 2013). Utilizing these two fluorophores will allow us to detect heterogeneity after PCR amplifications (Ning et al., 2015)-(Chan et al., 2012). By titrating in select CAG repeats after PCR we can record changes in fluorescence spectra and determine whether the amplicons are the correct length (Ning et al., 2015)-(Gatchel & Zoghbi, 2005). the Fluromax-3 Fluorimeter, we can record the fluorescence quenching of Doxorubicin by titrating in select CAG sequences (Ning et al., 2015)-(Gatchel & Zoghbi, 2005).

Acknowledgements:

I would like to thank Dr. Li for giving me this opportunity to work in his laboratory and further this investigation on CAG repeats. It was a pleasure and an honor to have worked alongside him and his graduate student Botros Toro. Botros Toro provided excellent guidance and leadership in critical moments. I also want to acknowledge my fellow undergraduate researcher, Gennaro Dellicarpini: he is an outstanding partner, and we achieved multiple goals together. In addition, I must also acknowledge my parents, Michael and Meskelu Tewolde, who have supported me throughout my undergraduate career.

Table of Contents:

Introduction:

Trinucleotide repeats are a common feature of genomes across multiple organisms (Li & Bonini, 2010)-(Usdin, 2008). These repeats often display instability or inconsistency in DNA structural integrity (Li & Bonini, 2010)-(Usdin, 2008). The expansion of trinucleotide repeats in human genomic DNA manifests into multiple neurodegenerative diseases (Li & Bonini, 2010)-(Mirkin, 2007). At least nine human diseases stem from the expression of expanded non-coding trinucleotide repeats (Mirkin, 2007)-(Usdin, 2008). Simple sequences such as, 5'CAG/CTG'3 repeats, contain a potential pathogenicity once expanded past their original lengths, which is exemplified in Huntington's Diseases and Myotonic Dystrophy (Mirkin, 2007)-(Usdin, 2008). The variability in repeat size governs the wide range of phenotypes associated in these diseases (Mitas, 1997). When expanded CAG repeats are transcribed into RNA, these transcripts are translated, and these mal functioning proteins can lead to severe cell damage (Li & Bonini, 2010)-(Usdin, 2008). Furthermore, while in the genome, the repeats can be passed down through generations in the germline, continuously expanding, and gaining more repeated units (Orr & Zoghbi, 2007)-(Hartenstine et al., 2000). This germline instability can cause an earlier onset age of the disease, and a more rapid progression (Mirkin, 2007)-(Mitas, 1997)-(Sobczak & Krzyzosiak, 2005). Moreover, an increasing length of CAG repeats is directly correlated with an increase in the severity of the aforementioned diseases (Orr & Zoghbi, 2007)-(Hartenstine et al., 2000). In addition, the mechanisms by which elongated CAG repeats operate *in vivo* remains unknown; however, past literature denotes structural and functional aspects of CAG tracts (Zhang et al., 2012)-(Broda et al., 2005). These past studies were performed on shorter repeat lengths of 10-80, while *in vivo*, CAG sequences reach over 100 repeats in length (Petruska et al., 1996)-(Hartenstine et al., 2002). Normal individuals have only 11 to 34 copies of 5'CAG/CTG'3 repeats; whereas those afflicted with Huntington's have upwards of 37 to 86 copies (Mariappan et al., 1998).

Studies have tried elucidated the expansion of CAG repeats in certain model organisms (Callahan, Andrews, Zakian, & Freudenreich, 2003), but the dynamic structure of these repeats is exemplified in Certain difficulties arise when studying such long CAG sequences *in vitro* as well (Sobczak et al., 2010)-(Sobczak & Krzyzosiak, 2005). The amplification of long repeats via Polymerase Chain Reaction (PCR) often leads to an unwanted stochastic shortening of DNA product- which is seen in CAG_{150} (Hartenstine et al., 2000)-(Stephenson, Keller, Tenenbaum, Zuker, & Li, 2014) . Petruska hypothesizes that when CAG repeats are being replicated, the strand forms structures that facilitated Taq's ability to jump over sequences and leave shortened replicons (Petruska et al., 1996). The PCR product of $(CAG)_{150}$ also appears as a smear after being ran on an agarose gel; this smear represents a population of CAG repeats at different lengths. $(CAG)_{150}$ presents with an inherent heterogeneity after amplification. This heterogeneity is also seen in lengthier repeats like 100 and 120, and these lanes also present with shortened products of after PCR, which laid above weak bands of correct CAG length. We aim to further study the heterogeneity of CAG repeats by displaying how the trinucleotide repeats sequence lengthens, and detecting these different lengths after PCR amplification. Initially a store of amplicons will be made with long oligonucleotide primers. Then using restriction enzymes, we will cut off the primers and visualize the change in $(CAG)_n$ sequences on an agarose gel. Afterwards, we will amplify repeats 10 through 150 with shorter primers. This will enable us to view the eventual shift of strong bands from shorter CAG lengths, to complete smears with longer repeats. This degradation is displayed with the amplicons of CAG 100, 120 and 150. Furthermore, with this varied store of CAG repeats, we can perform experiments that quantify the DNA product through utilization of fluorescence assays. Through fluorescence, we can quantify the CAG product and detect changes in length after PCR. An established anticancer drug, Doxorubicin- also known as Dyanomicin- will be used to detect the biding of CAG repeats through

fluorescence quenching (Karukstis, Thompson, Whiles, & Rosenfeld, 1998). Doxorubici$_n$ is a drug often prescribed after a diagnosis of malignant tumours (Das, Bhadra, Achari, Chakraborty, & Kumar, 2011). Doxorubicin binds to double stranded DNA of tumour cells and disrupts essential cellular functions (Das et al., 2011). Doxorubicin continuously fluoresces; however, once bound to 5'-GC-3' or 5'-GC-3' sequences of CAG tracts, a sharp decrease in fluoresces intensity is recorded (Ning et al., 2015)-(Hovorka et al., 2010). $(CAG)_n$ repeats that are rich in guanine (G) and cytosine (C). These repeats are susceptible to forming stable intramolecular structures (Ning et al., 2015). The chemical structure of Doxorubicin contains aromatic rings that are able to intercalate into the double strands of CAG repeats, and successfully bind to Watson-Crick base pairs (Box, 2007)-(Tran et al., 2014). Through spectroscopy techniques, we plan on titrating in varying concentrations of CAG repeats- and different lengths- into solution of Doxorubicin and associated buffer (Ning et al., 2015)-(Hovorka et al., 2010). This will allow us to observe how well Doxorubicin binds to longer repeats like CAG_{150}, and test how its natural heterogeneity after amplification affects these interactions (Changenet-Barret, Gustavsson, Markovitsi, Manet, & Monti, 2013). Later on, we can also perform experiments on how well Doxorubicin binds to single stranded CAG repeats(Das & Kumar, 2013)

Goals:

Our. We aim to investigate the heterogeneity of $(CAG)_n$ products after PCR amplification. The amplicons of CAG repeats are not all uniform in length: the CAG sequences varies in length after replication. We plan on developing techniques to detect changes in sequence length in order to better depict this inherent heterogeneity of CAG repeats. Initially, we plan on performing PCR experiments of $(CAG)_{10\text{-}150}$ with long oligonucleotide primers. With this initial amplification, we will then perform a double digestion reaction with enzymes BamHI and SallHF. These enzymes will cut off the primers leaving only the CAG sequences intact. Running the cut sequences on an agarose gel will reveal whether the repeats changed in length after the PCR amplification. Afterwards, we will then amplify $(CAG)_{10\text{-}150}$ with shorter primers. Utilizing shorter primers will allow us to monitor the effect of replication on each repeat length. Repeats of lengths 10 through 60 display solid, intense bands on an agarose gel; however, the bands begin to become fainter, and degrade into smears at lengths of 100 through 150. Using shorter primers provides a more accurate representation of how the CAG sequences are altered after amplification. We aim to view the degradation of strong bands, into fainter bands, and then smears. This is exemplified in the transition of CAG 100, 120, and 150. The next step in detecting the heterogeneity in these repeats is to develop a binding assay. We plan on using fluorophores to bind to CAG products after PCR. Tracking the intensities of the fluorescence that each repeat length emits will allow us to detect changes in sequence length. Doxorubicin (Dox) is a fluorophore that emits light in solution, except once bound to double stranded DNA: this intercalation into double stranded CAG repeats causes fluorescence quenching. The decrease in fluorescence allows us to track which CAG length is binding to Dox. Also, doxorubicin binds specifically to 5’GC’3 and 5’CG’3base pairs. Therefore, utilizing Dox in a binding assay with $(CAG)_n$ amplicons will allow us to detect changes in the original length of after PCR.

Method and Materials:

PCR with 3.0kb Primers:

We began with amplifying repeats of $CAG_{10\text{-}100}$ with long primers-as displayed in Figure 1- in PCR using Taq Polymerase (New England Bio Labs) to replicate the DNA, as seen in the PCR Protocols on Tables 1 and 2, and Thermal Cylcer Procedure on Table 3. We used Taq in order to display the heterogeneity in product after amplification. We used a 2720 Thermal Cycler from Bio Rad. A thermo buffer was used during these procedures from New England Bio Labs. In order to amplify CAG_{150} correctly, we used Q5 High Fidelity Polymerase (New England Bio Labs); and Q5 Reaction Buffer was used for this reaction. We use Q5 Polymerase instead of Taq because of CAG_{150}'s structural polymorphism; thusly, producing amplicons of the correct length. For both reactions, we used pUCTf –the forward primer- and pUCBr- the reverse primer. We then ran the PCR product on 0.8% agarose gels to correctly identify the size of the DNA amplified. Directly after PCR, purifying the product with a GeneJet Brand DNA Purification kit was the next step. With this pure

Table 1: PCR at .5mL Protocol

	uL
cag repeats	10-100
vector (10ng/uL)	10
Primer 1 (100uM)	5
Primer 2 (100uM)	5
dNTPs (25mM)	4
Buffer	50
H2O	420
Tap poly.	4
	498

Separate each 500uL into five 100uL samples.

Then perform PCR at p53 handle.

Table 2: PCR at .5mL Protocol

	uL
cag repeats	150
vector (10ng/uL)	10
Primer 1 (100uM)	5
Primer 2 (100uM)	5
dNTPs (25mM)	4
Q5 RXN Buffer	50
H2O	420
Q5 poly.	4
	498

Separate each 500uL into five 100uL samples.

Then perform PCR at p53 handle.

CAG DNA, we will then analyse it once more on a 0.8% agarose gel. This gel was made from 0.24g of agarose powder, along with 30mL of 10% Tris Buffer. The DNA on the gel was seen from a scanned image from a Typhoon Trio, on 526nm and 450700pmt emission settings for Figures 3 through 8. We also confirmed the correct replication of the PCR product by measuring their concentrations with a Nano Drop instrument - as seen on Table 5. The Nano Drop apparatus allows us to determine the concentrations of our amplified CAG repeats by only using 1µL of solution, with ng/µL precision. These measurements were compared with Dr. Li's Nano Drop concentrations, providing reassurance of the results.

Table 3: Thermal Cylcer Procedure

1. Choose p53 handle Procedure

A: PCR Amplification Steps

1. Heat to 95°C: Denaturation of Strands

2. Cool to 55°C: Anealing of Primers to Target Sequence

3. Heat to 75°C: Taq Begins Amplification

4: Repeat Steps 1, 2, and 3 for 30 Cycles

5. Cool to 4°C: Ensuring the Stop of Amplification and Mimic Refrigeration.

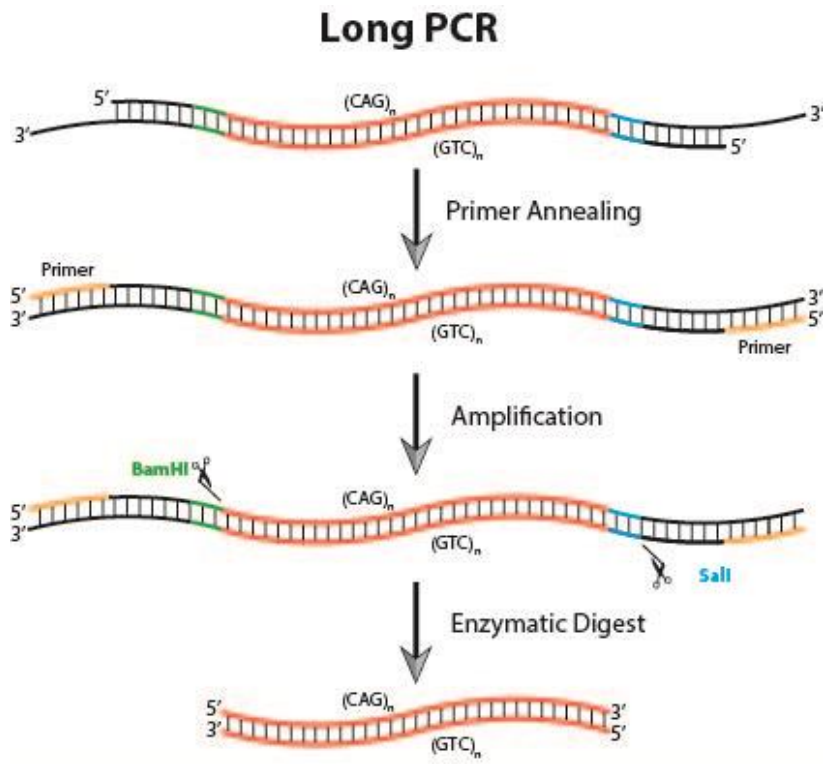

Figure 1: The methods for purifying CAG/DNA samples- in either linear templates or circular plasmids- are depicted in the figure above. After amplifying the repeats, the handles used in PCR would be cut off by Restriction Enzymes. Then the size of the purified samples will be analyzed on an agarose gel.

PCR Conditions:

The p53 handle program consisted of a series of –commonly 30- rounds of precise temperature changes termed cycles. Within each cycle, minute temperature fluctuations govern the replication of DNA. Initially, heating of the reaction vessel to 95 °C for 20 to thirty seconds is performed in order to denature the strands of template. The high temperatures elicit DNA melting through breaking of the hydrogen bonds between complementary bases, thereby producing two single stranded DNA. Next, the reaction temperature is lowered to 55 °C for 20-40 seconds, admitting the primers to anneal onto the single stands of DNA. This annealing step has to be extremely precise because of the hybridization of the primers to the template. The temperature needs to remain low enough to permit hybridization, while maintaining a temperature high enough allowing for specificity of binding. This delicate balance is essential for the primers properly binding to the complimentary sequences on the template strand. An excessively high temperature will completely inhibit binding, and a temperature too low will induce incorrect binding. In addition, the annealing temperature is usually 3 to 5 °C below the melting temperature (Tm) of the primers used. Incorrect annealing temperature inevitably cause replication errors as the reaction progresses. The next step consist of synthesizing a new DNA strand: the elongation phase. The temperature at this step is dependent on the DNA polymerase in the reaction. Taq polymerase's optimum activity is placed at 75 °C. The thermal cycler then heats up from 55 °C to 75°C, which enables Taq to attach to the primer/template strand, and begin synthesizing a complimentary DNA sequence. Taq creates this new daughter strand by adding dNTP's that are complementary to the template in a 5' to 3' direction. The time of this elongation phase is dictated by the enzymatic activity of the polymerase and the length of the desired DNA fragment amplified to be amplified. DNA polymerases at their respective optimized temperatures, are able to synthesize a thousand base pairs per minute. Therefore, during

optimal conditions, the only limiting aspect of the PCR is the amount of substrate. DNA polymerases are able to double the amount of targeted DNA each extension phase, thereby, causing an exponential increase of amplified DNA product. After elongation, the temperatures revert to the previous steps, and continues this process for 30 cycles. Once these cycles run to fruition, the final holding phase begins: the thermal cycler cools to 4°C, mimicking a temporary storage unit for the newly amplified product.

Enzymatic Double Digestion:

Furthermore, restriction enzymes will then be used to cut the long primers from the CAG repeats, thereby, indicating –with more specificity– how PCR replication affects the repeat lengths. As seen in Figure 1, the pair of the restriction enzymes that will be used are BamHI and Sall. Bam HI is a restriction endonuclease that binds at the recognition sequence 5'-GGATCC-3', and cleaves these sequences just after the 5'-guanine on each strand (New England Bio Labs). This cleavage results in sticky ends which are 4 base pairs long. Sal III is also a restriction endonuclease, however, it binds at the palindromic 5'-GTCGAC-'3 and cleaves in between the first guanine and thymine form the 5' end (New England Bio Labs). Table 4 depicts the protocol of how we tested each enzymes ability. Sample A is the control with no restriction enzymes; B contains only BamHI; C contains only Sall HF; and D both. After verifying how effective both enzymes are with sample D, we then replicated the double digest reactions with repeats 10 through 150.

Table 4: Double Enzymatic Digestion Protocol

Vol (µL)	**A**	**B**	**C**	**D**
DNA	3	3	3	3
Buffer 2	1	1	1	1
BSA	1	1	1	1
BamHI	-	1	-	1
Sal HF	-	-	1	1
H_2O	4	3	3	2
Total Volume	10	10	10	10

PCR with 1kb & 40bp Primers:

The short PCR experiments were performed in the same conditions as the long PCR's, as depicted in Table 1. The first set of short primers we used were the G series primers (1kb). The second set utilized was the M series (40bp). The M series are much shorter than the G primers. In addition, Figure 2 displays process of how short PCR will be run. The fluorescence assay was conducted with the purified PCR product made with M series primers. In order to correctly identify the repeats that we will be working with, we replenished our stock of CAG repeats by creating PCR volumes of 0.5mL. Dealing with the short PCR, we then purify the PCR products made from both the G and M series primers with the same GeneJet Brand DNA Purification kit. We also recorded the concentrations of each purified product from both sets of primers, with the same conditions as the long PCR products as seen in Tables 6, and 7.

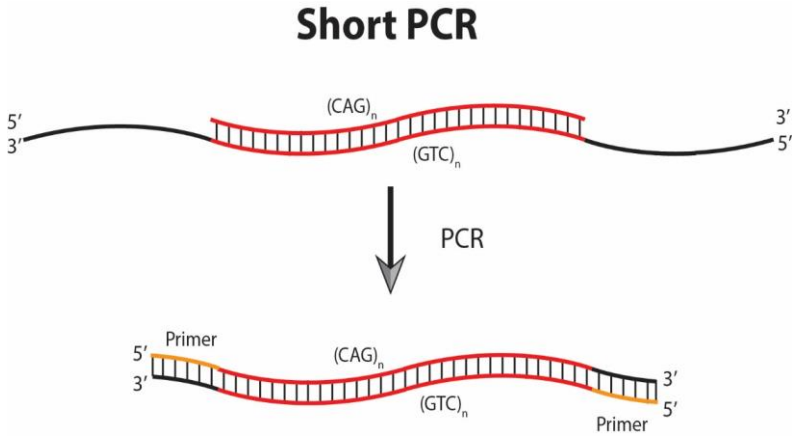

Figure 2: Representing the methodical steps in our plan, $(CAG)_n$ repeats will be amplified via PCR with various small primers. The length of these Primer handles will be heterogeneous at different

Binding Assay Procedure:

The fluorescence binding assay will consist of SYBR Green, Doxorubicin (dissolved in DMSO) and a 10mM Tris Buffer at a 7.4 pH. The Tris Buffer also contains 50mM NaCl. Utilizing fluorescence to tract the varying lengths of repeats in the heterogeneous population of the PCR product. Changes in fluorescence recordings will alert to alterations in (CAG)n length compared to the First we will excite a solution of SYBR Green and 10mM Tris Buffer under UV light. We will use a UV 1800 Shimadzu to emit UV light. This will ensure the emission spectra of SYBR Green. In addition, a solution of the buffer and the fluorescent chemical will have M-series $(CAG)_{10\text{-}150}$ products titrated in, and change in intensity recorded. Furthermore, the Fluorescence Intensities will be recorded with Fluromax-3 Fluorimeter. In order to test whether the fluorescent dye bound DNA to all of the binding sites we conducted experiments with multiple concentrations of CAG 60. We collected the fluoresce intensities with the Nano Drop Fluorimeter 3300.

Results:

PCR with 3.0kb Primers:

The difficulties of amplifying long CAG repeats are seen in our gel images. The PCR experiments for the repeats were all successfully produced, as seen in Figures 3 and 4; with a DNA Ladder as a control, the amplification of the DNA is shown as consecutive clear bands at 3 kbp. On Figure 3, CAG_{10} did no amplify well. However, the difficulties of replicating CAG_{10} were quickly rectified by simply redoing the PCR once more. Error may have originated from pipetting error while assembling the PCR experiment. We also added an extra 100uL of amplified CAG_{10} from previous test PCR's, and pooled the total CAG10 together, then ran it on the gel in Figure 4. This pooled product displayed a strong dark band, and the fact that the extra DNA was added is shown in the high concentration of 407.9 ng/µL, as seen in Table 5. CAG sequences between 20 and 100 repeats have concentrations between 200 ng/µL and 300 ng/µL. CAG_{150} also had a high concentration compared to the other repeats, at 394 ng/µL; however, this is expected due to its longer sequence. Furthermore, the way the CAG bands rise in accordance to increase repeat length verifies that the DNA product was correctly amplified, and purified- as seen in Figures 3 and 4. Moreover, the band of CAG_{150} being lower than 100 or 60, only exemplifies the intricate and complexity of CAG_{150} structure.

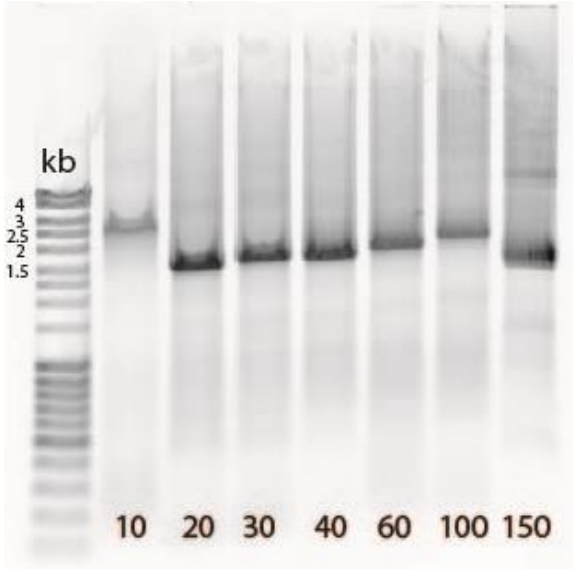

Figure 3: This gel image displays the amplification of $(CAG)_{10\text{-}150}$. Bands of each repeat are clearly shown expect for CAG_{10}. The reagents for CAG_{10} must not have properly reacted. However, for all the other the repeats, an increase in length, is verified with an increase in size, as seen in the rising bands in each lane. $(CAG)_{150}$ depicts a lower weight than 100 or 60 despite its shorter length; thus, signifying its natural unstable and complex nature.

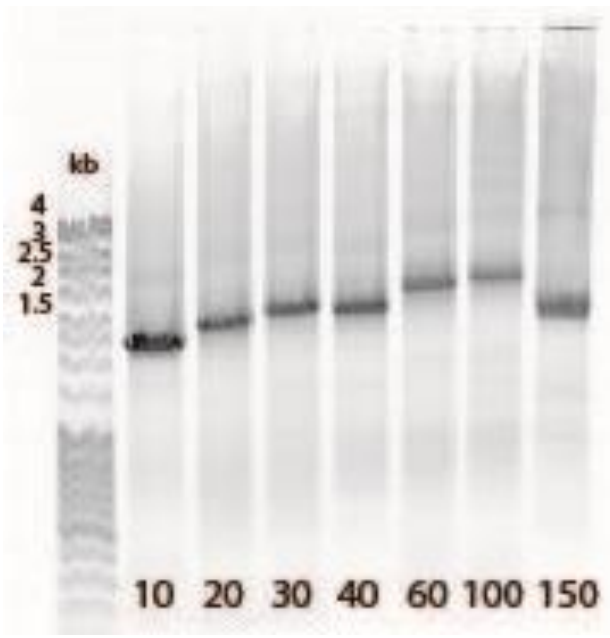

Figure 4: This gel image clearly displays $(CAG)_{10\text{-}150}$ repeats after PCR amplification, and purification. CAG10 was re-amplified, and purified once more. As the repeats increase in length, an increase in size is also shown for $(CAG)_{10\text{-}100}$. $(CAG)_{150}$ depicts a lower weight than 100 or 60 despite their shorter length.

Table 5: $(CAG)_{10\text{-}150}$ 3.0kb Primer Concentrations

$(CAG)_n$	(ng/µL)	A_{260nm}
10	407.9	8.159
20	248	4.96
30	227.1	4.541
40	272.1	5.442
60	227.1	4.541
100	196.3	3.926
150	394	7.879

Enzymatic Double Digestion:

The double digestion of the long PCR products highlights the heterogeneity that manifest in CAG repeats. Faint bands are present in the 25bp to 50bp for CAG's 10-30, as seen on Figure 5. Faint bands are also present on the 100bp range for CAG 40. CAG 60 shows a band in the 100-150bp region; and CAG 150 contains a smear at that weight. Ideally the image should present the CAG lengths at their correct sizes; however, during amplification the lengths of the repeats are altered. This gel displays the innate heterogeneity of CAG sequences after amplification. In the case of CAG 150, lengths of differing sizes are created. When these multiple lengths are ran on a gel, they produce a smear. Furthermore, Figure 5 displays the correlation between increased length, and heterogeneity manifestation. From 10 to 150, the bands become less and less defined, and eventually become a smear.

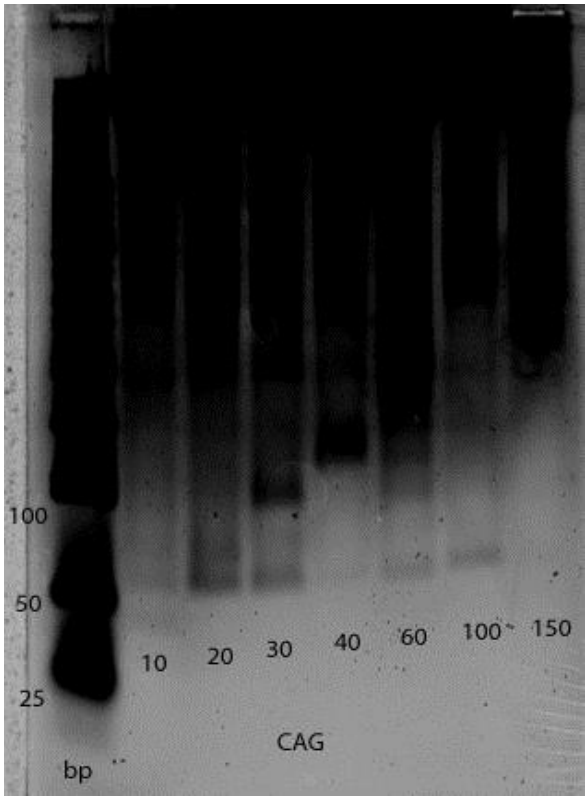

Figure 5: This gel image displays the purified PCR products of $(CAG)_{10\text{-}150}$ after a Double Enzymatic Digest. The enzymes, BamHI and Sal HF have cut off the long primers leaving only the CAG repeat base pairs, represented by the faint bands ranging from 25 to 50bp. Faint bands are seen in repeats 10-100 and 150 shows nothing. This images displays how amplification by PCR with long primers induces a heterogeneity seen in the CAG sequences alone. This is a 3% agarose gel, scanned at 700pmt.

PCR with 1.0kb Primers:

The second part of the heterogeneity investigation is the amplification of CAG repeats with smaller primers. As seen in Figure 6, we performed a 0.5mL PCR with smaller G series primers. In addition, we replicated a new length of CAG 120. More specifically, Figure 6 presents the crude amplified product- not the purified. The image displays a step wise increase of band weight along with increasing CAG length, which is similar to the long PCR images. Repeats 10 through 100 all display strong bands at appropriate weights. Moreover, Figure 6 shows a strong band for CAG 100, and then a smear with an unusual light weight band for 150. However, CAG 120 depicts fainter band that is slightly heavier than 100. CAG 120 serves as the link between 100 and 150: it displays the degradation of a strong band to a smear, which is all due to the increase of repeat length. Figure 7 displays the purified G-primer PCR product. The same features found in the last gel image, are also found in Figure 7. The bands weight increases along with longer CAG lengths, and the CAG 120 serves as the missing link once more. CAG 100 is a strong band, and 120 is much fainter and slightly heavier. Then 150 displays a smear with an oddly light

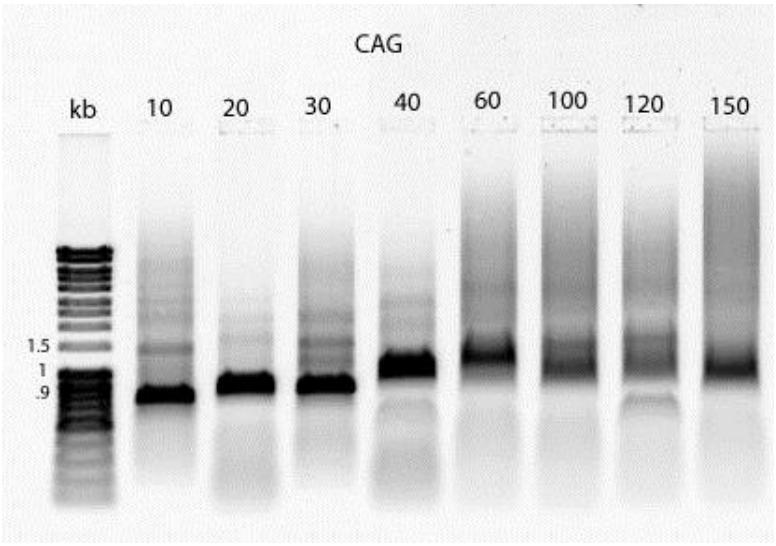

Figure 6: This image displays the crude PCR amplification of $(CAG)_{10\text{-}150}$ with shortened G series primers. High intensity bands are present for lengths 10 through 60. Meanwhile, the intensity of faint bands decreases from 100 to 120; and a smear – with a low weight band-is present for 150. From 100 to 120 the bands become fainter until a smear is shown on 150. This gel displays the eventual degradation of the bands because of the heterogeneity of lengthy CAG repeats. This is a 0.8% agarose gel, scanned at 500pmt.

band. Figure 7 –like Figure 6- also presents the effects of heterogeneity with increasing repeat length during PCR. Table 6 reviews the recordings of the G Primer product concentrations. The same procedure and materials were performed with the G primers as with the Long PCR products. We used the Nano Drop 1000 apparatus and deionized water to blank our measurements. As seen on Table 6, the concertation values are range between 350ng/µL and 588ng/µL. In this case, the longest repeat, 150 had the highest concentration of 587.2ng/µL. This is a considerable finding because ideally 150 should always have a highest concentration, since it is original started with the longest repeat length. The absorbance ratio of all the repeats ranges between 1.8 and 1.9; thusly, showing that the products are quite pure.

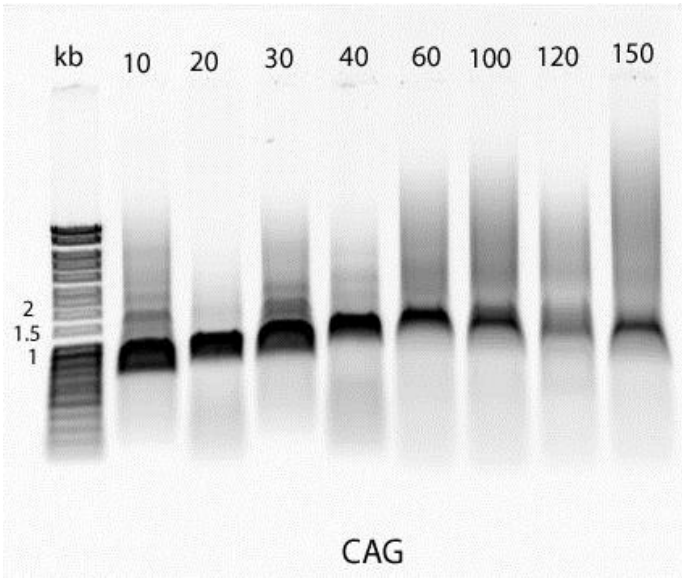

Figure 7: This image displays the purified PCR amplification of $(CAG)_{10\text{-}150}$ with shortened G series primers. These products are the purified samples from Figure 6; however, high intensity bands are present for lengths 10 through 100. Furthermore, the intensity of faint bands decreases from 100 to 120; and a smear –with a low weight band-is present for 150. Similar as to Figure 6, this gel displays the eventual degradation of the bands because of the heterogeneity of lengthy CAG repeats. This is a 0.5% agarose gel, scanned at 500pmt.

Table 6: $(CAG)_{10\text{-}150}$ G Primer Concentrations

$(CAG)_n$	ng/uL	A_{260}	A_{280}	260/280
10	458.3	9.166	4.831	1.9
20	355.5	7.109	3.778	1.88
30	470.2	9.404	5.133	1.83
40	379.7	7.593	4.009	1.89
60	486.8	9.736	5.114	1.9
100	577.2	11.543	6.21	1.86
120	478	9.561	5.055	1.89
150	587.2	11.745	6.191	1.9

PCR with 40bp Primers:

The next step in our investigation is the replication $(CAG)_n$ repeats with even shorter primers. We performed 0.5mL PCR experiments with M series primers. This PCR was kept at the same conditions as the previous experiments with G series, and Long primers. Figure 8 represents the image of the purified M series PCR product. The bands in Figure 8 are less intense than the bands in all other previous gel images; and this is because of the short length of the M series primers. As seen in the figure, CAG 10 through 60 show distinct bands at about 0.5 kb. CAG 100 appears as a faint band but is still distinguishable at 0.5 kb; also, CAG 120 is extremely faint and slightly heavier than CAG 100. Moreover, CAG 150 shows a smear as expected. In addition to the gel image, we recorded the concentrations of each sample with the Nano Drop 1000 apparatus. As seen on Table 7, the repeats ranged between 300ng/µL and 400ng/µL. The

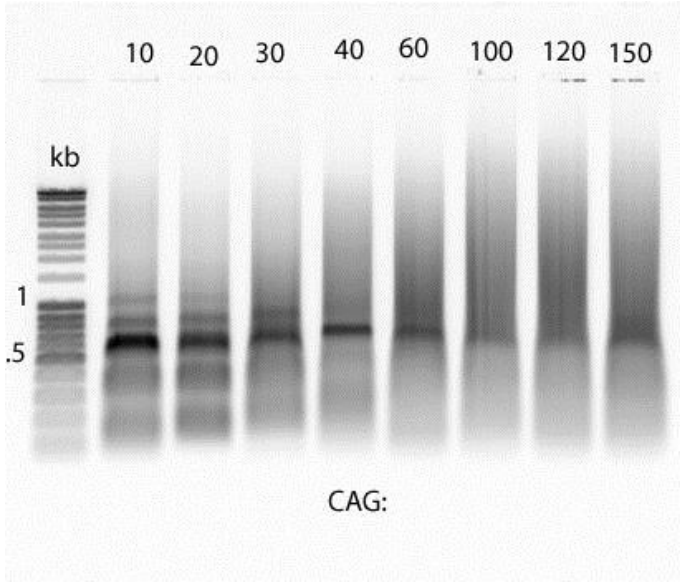

Figure 8: This gel displays the purified PCR product of $(CAG)_{10\text{-}150}$ with shortened M series primers. These M primers are shorter than the previous G series primers. Weak bands are shown in repeats 20 to 120, and 150 produced a smear. The shorter repeats amplified much better than the longer ones: a pattern shown with previous primers. These samples were run on a .5% agarose gel and scanned at 450pmt.

Table 7: $(CAG)_{10\text{-}150}$ M Primer Concentrations

$(CAG)_n$	ng/uL	A_{260}	A_{280}	260/280
10	306.58	6.132	3.296	1.86
20	315.86	6.317	3.384	1.87
30	346.48	6.930	3.668	1.89
40	243.13	4.863	2.565	1.90
60	382.42	7.648	4.049	1.89
100	314.73	6.295	3.356	1.88
120	406.10	8.122	4.274	1.90
150	373.84	7.477	3.987	1.88

absorbance ratio of 260/280 indicates the purity of the samples. All the repeats ratio's range from 1.86 to 1.90. These values indicate that our samples are quite pure.

Binding Assay:

-UV Illumination:

The genesis of the binding assay consisted of figuring the correct amount of DNA to bind with SYBR Green. Along with this, deciphering whether our fluorescent dye was emitting light at the correct wavelength. We then illuminated samples of the fluorescent probe with UV light- one with CAG 10, and another without. As seen in Figure 9, the SYBR Green is not excited by the short wavelength of UV light. No emission is seen. Once bound to CAG 10, SYBR Green then fluoresces-a bright green color- under UV light. This is shows that under UV, the fluorescent probe only emits light once bound to DNA.

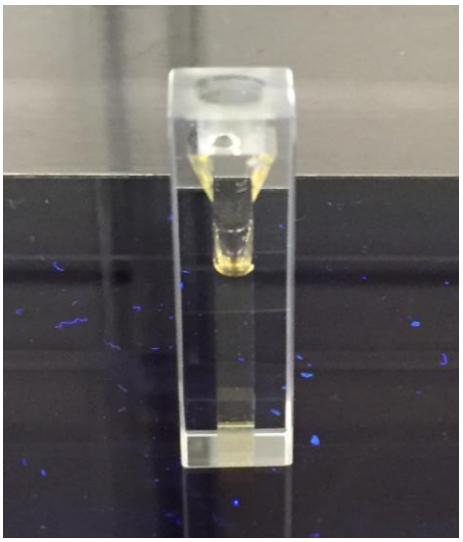

Figure 9: A quartz cuvette holding 10mM Tris Buffer at a 7.4 pH, and SYBR Green. As seen, when illuminated with UV light, no fluorescence emissions are present.

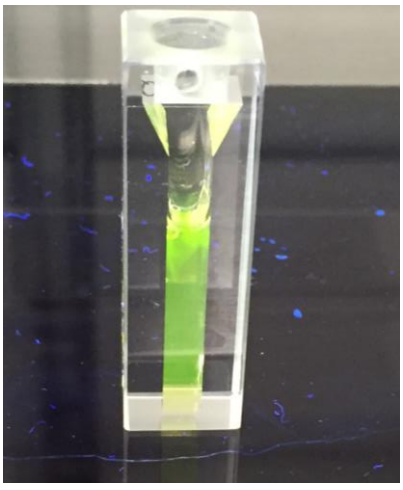

Figure 10: A quartz cuvette holding 10mM Tris Buffer at a 7.4 pH, and SYBR Green and CAG 10. As seen, when illuminated with UV light, fluorescence emissions are present. The binding of CAG10 with SYBR initiates this fluorescence.

-SYBR Green Titration

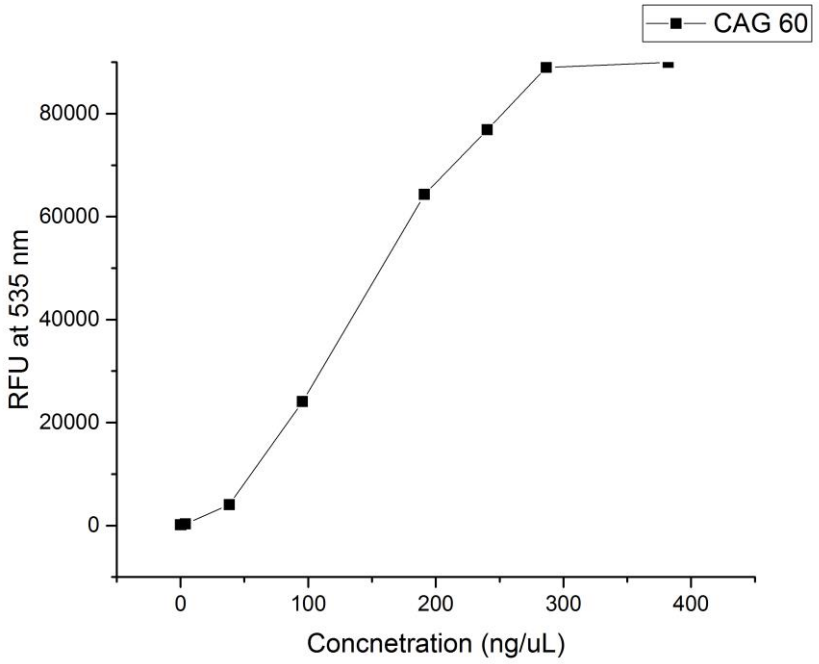

Figure 11: This graph depicts the concertation at which CAG 60 fills all of the binding sites of SYBR Green. The Relative Fluorescence Units (RFU) are the peak intensities of each concentration taken at 535nm. A plateau forms at a concentration of 300ng/µL. The peak intensities were recorded on the Nano Drop Fluorimeter 3300.

Next we titrated increasing concentrations of CAG 60 into constant amount of SYBR Green. This allowed us to view the concentration of at which this fluorescent probe has all of its binding sites bound with DNA. Figure 11 displays the fluorescence intensities at the peak wavelength of 535nm. The plateau at 300ng/µL depicts the concentration at which all the binding sites are filled due to the constant intensity.

-Doxorubicin Titration

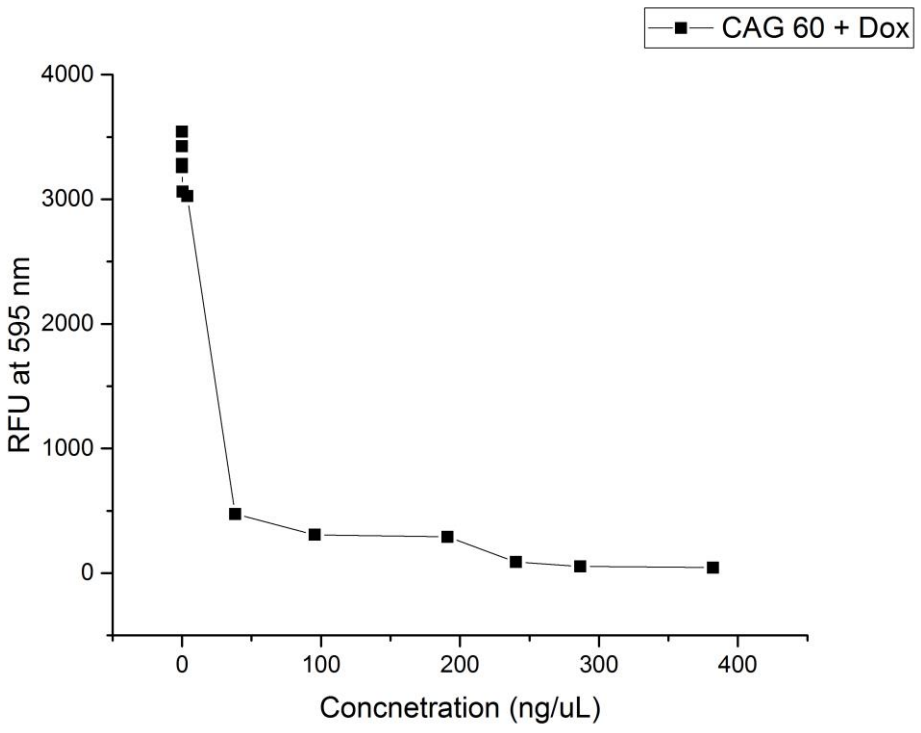

Figure 12: This graph depicts the concertation at which CAG 60 fills all of the binding sites of Doxorubicin. The Relative Fluorescence Units are the peak intensities of each concentration taken at 595 nm. A plateau forms at a concentration of 300ng/µL. The peak intensities were recorded on the Nano Drop Fluorimeter

A titration curve was also made for the interaction between Doxorubicin and CAG 60. A serial dilution of CAG 60 was made and titrated into constant concentrations of Doxorubicin. From these samples, the peak intensity- at 595nm- was recorded with the Nano Drop Fluorimeter 3300. The intercalation of Doxorubicin is present in Figure 12. With concentration plotted against fluoresce intensity, a higher concentration of CAG 60 correlates to a lower intensity. This is due to intercalation of Doxorubicin into the double stranded helix of CAG 60. Similar to Figure 11, a 300ng/uL concentration of DNA binds all the open sites of Doxorubicin.

-Doxorubicin Fluorescence Spectra

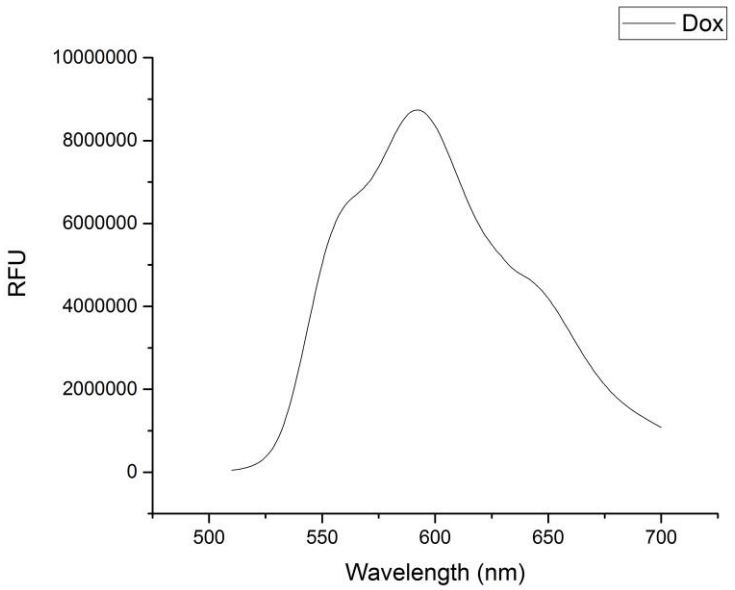

Figure 13: This graph shows the fluorescence spectra of Doxorubicin after excited at 450nm. The peak intensity is shown to be $8.73x10^6$, at wavelength 590nm. The Doxorubicin sample was recorded in a 10mM Tris Buffer at pH 7.4- the same conditions as the SYBR binding experiments were used with Doxorubicin.

As seen in Figure 13, the intensity spectra of Doxorubicin is quite different from the one of SYBR Green. Next, different repeats will be used to show how the increase of length will allow for more binding of Doxorubicin, and display fluorescence quenching.

Discussion:

CAG/CTG trinucleotide repeats are at the origin and proliferation of multiple neurodegenerative diseases through mechanisms that are not fully understood (Li & Bonini, 2010)-(Kashi & King, 2006). Insight into how these repeats affect DNA polymerase *in vivo* have shown that hairpin structures arise (Wheeler et al., 2012)- (Michlewski & Krzyzosiak, 2004); however, extended CAG/CTG repeats into the hundreds continue to be troublesome even in simple PCR amplifications, and in novel pulling experiments (Stephenson et al., 2014). Past literature denotes that expanded CAG/CTG repeats in genomic DNA have structural polymorphisms in thermodynamically stable forms that are predictable, and non-canonical states (Lin & Wilson, 2011)- (Amrane et al., 2005). The heterogeneity of these repeats the innate replication difficulties present in *in vitro* studies and physiological conditions in Huntingtins and Myotonic Dystrophy (Lin & Wilson, 2011)- (Cooper, 2009). When CAG repeats are being replicated, the strand forms structures that facilitated Taq's ability to jump over sequences and leave shortened replicons (Liu, Chen, Bissler, Sinden, & Leffak, 2010) . The agarose gels that represent the amplified CAG sequences merely prove how difficult extended repeats are to amplify correctly. Figures 5's presentation of the enzymatic digestion of the repeats displays the trend of heterogeneity of CAG tracts. The correlation between increased length, and heterogeneity manifestation. From 10 to 150, the bands become less and less defined, and eventually become a smear. Figures 6 through 8 show the degradation of intense bands to smears, as CAG repeats lengthen. This trend is shown in the amplification of repeats 100, 120, and 150. This is seen when using both G and M series primers. We were also able to identify that SYBR Green fluoresced only when bound to DNA with the images of Figures 9 and 10. In addition, the binding of SYBR Green and the CAG lengths will be tested with varying lengths at the correct concentration at which all of the SYBR binding sites will be

bound to DNA. This will help us probe whether longer lengths will induce higher emission intensities. Furthermore, future experiments that will enable us to continue our study of these repeats can be performed with the development of a fluorescence assay (Petruska, Hartenstine, & Goodman, 1998)- (Ning et al., 2015). Doxorubicin- also called Daunomycinone- a well-known anticancer drug, will be used in future fluorescence assays to further probe the heterogeneity of CAG_{150}. Doxorubicin intercalates between double stranded DNA that is rich in 5'-GC-3' or 5'-GC-3' sequences (Ning et al., 2015)-(Box, 2007). Furthermore, Doxorubicin acts a fluorophore that emits light at 560nm, once excited at 480nm (Karukstis et al., 1998)-(Das et al., 2011). When Doxorubicin is in solution alone, its full fluorescence is observed; when CAG sequences are added to the solution, Doxorubicin will intercalate between the double strands (Hovorka et al., 2010)-(Changenet-Barret et al., 2013). This intercalation will cause fluorescence quenching. A sudden decrease in fluoresce signifies the binding of Doxorubicin (Changenet-Barret et al., 2013),(Galka-Marciniak, Urbanek, & Krzyzosiak, 2012),(Yildirim, Park, Disney, & Schatz, 2013). By using the Fluromax-3 Fluorimeter, we can record the fluorescence quenching of Doxorubicin by titrating in select CAG sequences (Ning et al., 2015)-(Gatchel & Zoghbi, 2005). As seen in Figure 12, it seems like 300ng/µL is a concentration at which DNA can fill all of Doxorubicin and SYBR Green binding sites. Later on, we will aim to observe an increase of fluorescence quenching by titrating longer CAG tracts- starting from CAG_{10}, then ending at CAG_{150} (Ning et al., 2015)-(Chan et al., 2012). Along with the fluorescence assay, we can then move towards asymmetrical PCR. In this reaction, fluorescent probes on the primers can be used to track the amplification of only one strand of DNA. By replication one strand of $(CAG)_n$ we can test how well Doxorubicin binds to single strands of sequences that are rich with 5'-GC-3' and 5'-GC-3'. Possibly, a strong fluorescence quenching will be observed when performing titration tests with CAG_{150}. The dynamic structure of CAG_{150} would allow

for certain rearrangements that mimic the same GC accessibility as in double stranded sequences. As the single stranded repeats increase in length, the fluorescence would continue to drop. This pattern between double and single strand $(CAG)_n$ may be quite similar.

References:

Amrane, S., Saccà, B., Mills, M., Chauhan, M., Klump, H. H., & Mergny, J. L. (2005). Length-dependent energetics of (CTG)n and (CAG)n trinucleotide repeats. *Nucleic Acids Research*, *33*(13), 4065–4077. http://doi.org/10.1093/nar/gki716

Box, V. G. S. (2007). The intercalation of DNA double helices with doxorubicin and nogalamycin. *Journal of Molecular Graphics & Modelling*, *26*(1), 14–9. http://doi.org/10.1016/j.jmgm.2006.09.005

Broda, M., Kierzek, E., Gdaniec, Z., Kulinski, T., & Kierzek, R. (2005). Thermodynamic stability of RNA structures formed by CNG trinucleotide repeats. Implication for prediction of RNA structure. *Biochemistry*, *44*(32), 10873–10882. http://doi.org/10.1021/bi0502339

Callahan, J. L., Andrews, K. J., Zakian, V. a, & Freudenreich, C. H. (2003). Mutations in yeast replication proteins that increase CAG/CTG expansions also increase repeat fragility. *Molecular and Cellular Biology*, *23*(21), 7849–7860. http://doi.org/10.1128/MCB.23.21.7849-7860.2003

Chan, N. L. S., Hou, C., Zhang, T., Yuan, F., Machwe, A., Huang, J., … Li, G. M. (2012). The Werner syndrome protein promotes CAG/CTG repeat stability by resolving large (CAG)n/(CTG)n hairpins. *Journal of Biological Chemistry*, *287*(36), 30151–30156. http://doi.org/10.1074/jbc.M112.389791

Changenet-Barret, P., Gustavsson, T., Markovitsi, D., Manet, I., & Monti, S. (2013). Unravelling molecular mechanisms in the fluorescence spectra of doxorubicin in aqueous solution by femtosecond fluorescence spectroscopy. *Physical Chemistry Chemical Physics : PCCP*, *15*(8), 2937–44. http://doi.org/10.1039/c2cp44056c

Cooper, T. a. (2009). Chemical reversal of the RNA gain of function in myotonic dystrophy. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(44), 18433–18434. http://doi.org/10.1073/pnas.0910643106

Das, A., Bhadra, K., Achari, B., Chakraborty, P., & Kumar, G. S. (2011). Interaction of aristololactam-β-D-glucoside and daunomycin with poly(A): spectroscopic and calorimetric studies. *Biophysical Chemistry*, *155*(1), 10–9. http://doi.org/10.1016/j.bpc.2011.01.011

Das, A., & Kumar, G. S. (2013). Binding of the plant alkaloid aristololactam-??-d-glucoside and antitumor antibiotic daunomycin to single stranded polyribonucleotides. *Biochimica et Biophysica Acta - General Subjects*, *1830*(10), 4708–4718. http://doi.org/10.1016/j.bbagen.2013.06.001

Galka-Marciniak, P., Urbanek, M. O., & Krzyzosiak, W. J. (2012). Triplet repeats in transcripts: Structural insights into RNA toxicity. *Biological Chemistry*. http://doi.org/10.1515/hsz-2012-0218

Gatchel, J. R., & Zoghbi, H. Y. (2005). Diseases of unstable repeat expansion: mechanisms and common principles. *Nature Reviews. Genetics*, *6*(10), 743–755. http://doi.org/10.1038/nrg1691

Hartenstine, M. J., Goodman, M. F., & Petruska, J. (2000). Base stacking and even/odd behavior of hairpin loops in DNA triplet repeat slippage and expansion with DNA

polymerase. *Journal of Biological Chemistry*, *275*(24), 18382–18390. http://doi.org/10.1074/jbc.275.24.18382

Hartenstine, M. J., Goodman, M. F., & Petruska, J. (2002). Weak strand displacement activity enables human DNA polymerase ?? to expand CAG/CTG triplet repeats at strand breaks. *Journal of Biological Chemistry*, *277*(44), 41379–41389. http://doi.org/10.1074/jbc.M207013200

Hovorka, O., Šubr, V., Větvička, D., Kovář, L., Strohalm, J., Strohalm, M., … Říhová, B. (2010). Spectral analysis of doxorubicin accumulation and the indirect quantification of its DNA intercalation. *European Journal of Pharmaceutics and Biopharmaceutics*, *76*(3), 514–524. http://doi.org/10.1016/j.ejpb.2010.07.008

Karukstis, K. K., Thompson, E. H. Z., Whiles, J. a., & Rosenfeld, R. J. (1998). Deciphering the fluorescence signature of daunomycin and doxorubicin. *Biophysical Chemistry*, *73*(3), 249–263. http://doi.org/10.1016/S0301-4622(98)00150-1

Kashi, Y., & King, D. G. (2006). Simple sequence repeats as advantageous mutators in evolution. *Trends in Genetics*, *22*(5), 253–259. http://doi.org/10.1016/j.tig.2006.03.005

Li, L. B., & Bonini, N. M. (2010). Roles of trinucleotide-repeat RNA in neurological disease and degeneration. *Trends in Neurosciences*, *33*(6), 292–298. http://doi.org/10.1016/j.tins.2010.03.004

Lin, Y., & Wilson, J. H. (2011). Transcription-induced DNA toxicity at trinucleotide repeats Double bubble is trouble. *Cell Cycle*. http://doi.org/10.4161/cc.10.4.14729

Liu, G., Chen, X., Bissler, J. J., Sinden, R. R., & Leffak, M. (2010). Replication-dependent instability at (CTG) x (CAG) repeat hairpins in human cells. *Nature Chemical Biology*, *6*(9), 652–659. http://doi.org/10.1038/nchembio.416

Mariappan, S. V, Silks, L. a, Chen, X., Springer, P. a, Wu, R., Moyzis, R. K., … Gupta, G. (1998). Solution structures of the Huntington's disease DNA triplets, (CAG)n. *Journal of Biomolecular Structure & Dynamics*, *15*(4), 723–744. http://doi.org/10.1080/07391102.1998.10508988

Michlewski, G., & Krzyzosiak, W. J. (2004). Molecular architecture of CAG repeats in human disease related transcripts. *Journal of Molecular Biology*, *340*(4), 665–679. http://doi.org/10.1016/j.jmb.2004.05.021

Mirkin, S. M. (2007). Expandable DNA repeats and human disease. *Nature*, *447*(7147), 932–940. http://doi.org/10.1038/nature05977

Mitas, M. (1997). Trinucleotide repeats associated with human disease. *Nucleic Acids Research*, *25*(12), 2245–2253. http://doi.org/10.1093/nar/25.12.2245

Ning, L., Yang, D., Gao, T., Lu, S., Yin, Y., & Li, G. (2015). A new method to evaluate trinucleotide repeats length polymorphism. *Talanta*, *143*, 414–418. http://doi.org/10.1016/j.talanta.2015.05.016

Orr, H. T., & Zoghbi, H. Y. (2007). Trinucleotide Repeat Disorders. *Annual Review of Neuroscience*, *30*(1), 575–621. http://doi.org/10.1146/annurev.neuro.29.051605.113042

Petruska, J., Arnheim, N., & Goodman, M. F. (1996). Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases. *Nucleic Acids Research*, *24*(11), 1992–1998.

http://doi.org/10.1093/nar/24.11.1992

Petruska, J., Hartenstine, M. J., & Goodman, M. F. (1998). NUCLEIC ACIDS , PROTEIN SYNTHESIS , AND MOLECULAR GENETICS : Analysis of Strand Slippage in DNA Polymerase Expansions of CAG / CTG Triplet Repeats Associated with Neurodegenerative Disease Analysis of Strand Slippage in DNA Polymerase Expansions of CAG / , *273*(9), 1–8. http://doi.org/10.1074/jbc.273.9.5204

Sobczak, K., & Krzyzosiak, W. J. (2005). CAG repeats containing CAA interruptions form branched hairpin structures in spinocerebellar ataxia type 2 transcripts. *Journal of Biological Chemistry*, *280*(5), 3898–3910. http://doi.org/10.1074/jbc.M409984200

Sobczak, K., Michlewski, G., De Mezer, M., Kierzek, E., Krol, J., Olejniczak, M., … Krzyzosiak, W. J. (2010). Structural diversity of triplet repeat RNAs. *Journal of Biological Chemistry*, *285*(17), 12755–12764. http://doi.org/10.1074/jbc.M109.078790

Stephenson, W. T., Keller, S., Tenenbaum, S. a., Zuker, M., & Li, P. T. X. (2014). Structural Polymorphism of (Cag)N Repeat RNA Elucidated using Single Molecule Nanomanipulation. *Biophysical Journal*, *106*(2), 282a. http://doi.org/10.1016/j.bpj.2013.11.1653

Tran, T., Childs-Disney, J. L., Liu, B., Guan, L., Rzuczek, S., & Disney, M. D. (2014). Targeting the r(CGG) repeats that cause FXTAS with modularly assembled small molecules and oligonucleotides. *ACS Chemical Biology*, *9*(4), 904–912. http://doi.org/10.1021/cb400875u

Usdin, K. (2008). The biological effects of simple tandem repeats: Lessons from the repeat expansion diseases. *Genome Research*. http://doi.org/10.1101/gr.070409.107

Wheeler, T. M., Leger, A. J., Pandey, S. K., MacLeod, a. R., Nakamori, M., Cheng, S. H., … Thornton, C. a. (2012). Targeting nuclear RNA for in vivo correction of myotonic dystrophy. *Nature*, *488*(7409), 111–115. http://doi.org/10.1038/nature11362

Yildirim, I., Park, H., Disney, M. D., & Schatz, G. C. (2013). A dynamic structural model of expanded RNA CAG repeats: A refined X-ray structure and computational investigations using molecular dynamics and umbrella sampling simulations. *Journal of the American Chemical Society*, *135*(9), 3528–3538. http://doi.org/10.1021/ja3108627

Zhang, T., Huang, J., Gu, L., & Li, G.-M. (2012). In vitro repair of DNA hairpins containing various numbers of CAG/CTG trinucleotide repeats. *DNA Repair*, *11*(2), 201–209. http://doi.org/10.1016/j.dnarep.2011.10.020