2024-03-08

Tractor Beams and Disease: Probing Repeat RNA Structure through Single Molecule Force Spectroscopy

Joshua Webster-Ford

Honors Thesis

TRACTOR BEAMS AND DISEASE: PROBING REPEAT RNA STRUCTURE THROUGH SINGLE MOLECULE FORCE SPECTROSCOPY

by

Joshua Webster-Ford

Submitted to Brigham Young University in partial fulfillment of graduation requirements for University Honors

Department of Chemistry and Biochemistry

Brigham Young University

April 2024

Advisor: Kenneth Christensen

Reader: Sarah Woodson

Honors Coordinator: Walter Paxton

ABSTRACT

TRACTOR BEAMS AND DISEASE: PROBING REPEAT RNA STRUCTURE THROUGH SINGLE MOLECULE FORCE SPECTROSCOPY

Joshua Webster-Ford

Department of Chemistry and Biochemistry

Bachelor of Science

Huntington's Disease is characterized by an extended $(CAG)^n$ repeat found in the *huntingtin (Htt)* gene. While the typical human contains 6-20 CAG repeats, people with over 36 repeats manifest symptoms of the disease. The extended CAG repeats allow the *Htt* pre-mRNA transcript to form a hairpin loop structure which can sequester certain vital proteins such as Muscleblind like protein 1 (MBNL1), hindering their functional role. Further understanding the structure of the MBNL1-RNA complex is possible through force spectroscopy, or "optical tweezers", applied on the single molecule level. Force spectroscopy generates force versus distance plots for individual RNA molecules, revealing the dynamic "slipping" action of the hairpin, in which the hairpin shifts by 3 base pairs. Force spectroscopy assays have not been performed on the MBNL1-RNA complex, or on MBNL1-RNA treated with D6, a DAPI derivative that can intercalate into the RNA hairpin thought to be a potential small molecule treatment for HD. We aim to understand the structures of these complexes and the kinetics of their formation, in the hopes of unveiling potential avenues for developing novel treatments for Huntington's disease and other trinucleotide repeat expansion diseases.

ACKNOWLEDGEMENTS

I am deeply grateful for Dr. Sarah Woodson for hosting me in her lab over the summer and providing me with the chance to learn about her research. I am also grateful for her assistance and supervision in writing this manuscript and helping me to become interested in biochemical and biophysical research.

I would also like to thank my advisor, Dr. Ken Christensen, for his support and willingness to help me write up my research. His expertise and experience in biochemistry have inspired me since I first met him in Chem 584, and I consider myself very lucky to have worked with him.

I am grateful for Dr. Walter Paxton for his meeting with me throughout my undergraduate career to help me stay on track to graduate with honors, and his help in reading this transcript.

I am also grateful to the members of the Woodson lab, particularly Brett O'Brien, for mentoring me throughout my research over the summer. I also thank Dr. Christian Kaiser for allowing me to use his C-trap optical tweezers while at Johns Hopkins University.

Finally, I am especially grateful to my friends and family for their love and support during this process, especially Emma Webster, who moved with me across the country while I performed this research and supported me at every step of the journey.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

TRACTOR BEAMS AND DISEASE: PROBING REPEAT RNA STRUCTURE THROUGH SINGLE MOLECULE FORCE SPECTROSCOPY

INTRODUCTION

1.1 Genetic Basis

Huntington's Disease (HD) is a trinucleotide repeat expansion disorder (TRED) characterized by late-onset neurodegeneration. Roughly 40,000 people in the USA have the disease while another 200,000 are at risk of inheriting it.[1] HD is characterized by a $(CAG)^n$ trinucleotide repeat expansion in the first exon of the huntingtin (*Htt*) gene found on chromosome 4.[2] *Htt* is a widely expressed cytoplasmic protein thought to be primarily involved in microtubule-mediated transport and vesicle function.[3] Although the function of *Htt* is still being studied, the knockout of wildtype *Htt* is embryonically lethal in the early developmental stages of both Drosophila melanogaster and mouse models, proving its importance to cellular function.[4]

The *Htt* gene of a typical genome contains roughly 17 to 20 CAG repeats. Symptoms of HD start to appear once the trinucleotide repeat (TNR) number has reached roughly 36 repeats, which is when *Htt* has been shown to aggregate in vitro.[5] There is an inverse correlation between the length of the repeating segment and the age of disease onset, a well-established feature of genetic disease known as "anticipation."[6] The genetic mechanism of TNR expansion has not been completely elucidated, although it is hypothesized that repeats expand by forming a stable secondary structure (a hairpin loop) after the DNA polymerase "slipped" on the repeat section.[7]

Although visualizing a hairpin structure in vivo is not possible with our current structural resolution limits, this hypothesis is supported by most genetic data.

The expanded CAG region, often referred to as the polyglutamine or polyQ section, allows the HTT protein to aggregate with itself. In the N-terminus domain, repeat-expanded regions contain HEAT (huntingtin, elongation factor 3, protein phosphatase 2A, and TOR1) consensus sequences involved in aggregation, as they often fold into helix-turn-helix motifs that act as a scaffold for protein-protein interactions.[8]

HD pathogenesis can also be explained by the structure of repeat-containing RNA molecules. The cytosine residue of one CAG repeat can form a Watson-Crick base pair with the guanine residue of a different CAG repeat, leading to the formation of a hairpin loop (see Figure S1). The importance of RNA structure in pathogenesis was shown by the insertion of a single CAA trinucleotide into a sequence of 40 CAG repeats, which severely reduced the intensity of HD symptoms, solidifying the cytosine-guanine Watson-Crick pair formation theory, as CAA also codes for glutamine and as such the protein would possess the same affinity for aggregation.[9] These RNA hairpin structures can sequester proteins that naturally bind to double-stranded nucleic acids, particularly proteins that contain zinc finger motifs that recognize AU-rich elements.[10] One such protein is Muscleblind-like protein 1 (MBNL1), named because of its ability to cause blindness in mice whose protein is no longer functional (see Figure S2). MBNL1 acts to control the alternative splicing of exons during postnatal development by binding to A-A mismatches in double stranded sequences of nucleic acid (see Figure 2).[11] Sequestration of

MBNL1 leads to improper splicing in genes that are crucial to muscle development, leading to the progression of degenerative symptoms.[12]

1.2 Current Treatment

Although there are no current ways to prevent or slow down the progression of HD, several potential treatments are undergoing clinical trials. One such potential treatment is the use of antisense oligonucleotides (ASOs) which bind to the mRNA to both prevent hairpin loop formation and induce degradation before polyQ-protein expression. ASOs typically contain a sugar modification that enhances metabolic stability and improves binding to the target sequence, as well as a sequence of 8-10 nucleotides that promotes the binding of RNase H1 which can quickly cleave the mRNA sequence.[13]

Certain small molecules with high affinity for binding to 1x1 nucleotide RNA internal loops have also been proposed as a treatment method for HD. Such a molecule is D6, an analogue of the 4',6-diamidino-2-phenylindole (DAPI) stain that shows promising selectivity for A-A internal loops in computational screening studies.[14] The molecular structure of both DAPI and D6 are shown in Figure S3. In vitro studies on pre-mRNA of several genes that are affected by HD show that treatment of D6 improves splicing defects, evidence that D6 is competitively binding to the RNA hairpin, preventing the formation of the MBNL1-RNA complex. Such studies of D6 are novel and the mechanism by which D6 prevents MBNL1-RNA complex formation is still unclear, although it is likely that the planar shape of D6 allows it to intercalate into the RNA

hairpins, similar to the way in which DAPI intercalates into DNA when used as a fluorescent dye, disrupting protein binding.

1.3 Single Molecule Approaches

Understanding the structure of repeat-containing RNA hairpins is likely to provide insight into the way in which they sequester MBNL1 and other vital proteins and how D6 is able to prevent the formation of this complex. Here, we propose to probe the structure using single molecule approaches. Single molecule assays are useful because bulk studies produce ensemble averages drawn from an average of multiple samples, which often are not a possible conformation of the population. On the other hand, single molecule studies follow the folding and unfolding pathways of an individual molecule, allowing the structures of individual RNA conformations to be studied without losing specificity through ensemble averaging.

One such method of studying RNA hairpins at the single molecule level is force spectroscopy, colloquially referred to as 'optical tweezers'. The RNA of interest is connected to two beads of micron diameter using covalent biotin-streptavidin interactions. The beads are then suspended via a pair of high intensity lasers; the narrowest point of the focused laser beam contains a very strong electric field gradient, which attracts the dielectric beads and holds them at that point. Photon radiation pressure of the focused laser beams counters the downward force of gravity while also preventing lateral movement, effectively suspending the beads in the beam. One of the

lasers can then be adjusted, pulling the bead further from its counterpart bead and applying a force on the molecule (see Figure S4).

If the displacement from the trap centre is small enough, the force applied to the RNA can be assumed to be linear with respect to its displacement and the optical trap follows Hooke's law and can be treated as an ideal simple spring. Treating the two beads as a Hookean spring, a worm-like chain model can be used to extrapolate how the RNA will respond to a wide range of forces.[15] The worm-like chain model is a statistical mechanical method used to analyse flexible polymers in two dimensions, which is a good model for RNA. This model allows us to calculate the average extension of the polymer as a function of the external force applied. A worm-like chain can be generated for each conformation of the RNA molecule; the most obvious are the RNA in its most folded state, and the RNA in its most unfolded state. Perhaps of more interest are the intermediate conformations, where the RNA is in the process of unfolding and refolding.

Once the distance is determined at which the RNA constantly unfolds and refolds has been experimentally determined, the molecule can be left at this distance to calculate the proportion of times spent in different conformations. Such data allows us to determine kinetic data for the folding and unfolding of a single RNA model, providing insight into the stability of different conformations.[16]

1.4 Gap in Knowledge

Previous experiments done by the Woodson lab have measured the average extension distance at which certain unfolding events are prone to occur and compared that distance to the length of a

CAG trinucleotide. This showed that CAG repeats often "slip", meaning that Watson-Crick base pairs are broken and then reformed with a different CAG repeat as the molecule becomes more extended. This model explains why there is such conformational heterogeneity in RNA when probed with optical tweezers, as many different sizes of hairpin loops can form.

Although naked CAG-repeat RNA has been characterized using optical tweezers, neither the RNA-MBNL1 complex nor the RNA treated with D6 have ever been tested using these single molecule techniques. In this thesis, I aim to use force spectroscopy techniques to provide insight into the structure and binding stability of RNA-MBNL1 complexes. I hypothesize that the binding of MBNL1 will stabilize the RNA hairpin structure, decreasing the amount of intermediate conformational states that the RNA can adopt. Proving such a hypothesis to be true will not only increase our understanding of the structure of the MBNL1-RNA complex, but it may also provide insight into ways we can disrupt the binding stability as a treatment to prevent HD pathogenesis. In addition, I aim to study the effects of D6 on MBNL1 binding to RNA beyond what has been previously studied. Such treatment on a single molecule level may open up avenues of future treatment for people with HD on the macroscopic level.

METHODOLOGIES

2.1 MBNL1 translation and purification

The plasmid encoding GST-MBNL1-Hisx6, AmpR, and LacI (see Figure S5) was amplified via inverse PCR (50 μL) using primers (found in the appendix) to remove the final amino acid residues of the protein (260 – 388), as has been done previously in research on MBNL1. The PCR ran at a gradient temperature from 62-72°C. Following PCR amplification, all samples were run on a 1% agarose gel electrophoresis. Agarose gel electrophoresis was performed for 10 min at 60 V and then for 45 min at 120 V. It was determined by band intensity that the best amplification occurred at 65.5°C.

The linear PCR product was relinearized with a combination of kinase, ligase, and DpnI (KLD) in a 10 μL reaction. As the template had previously been isolated from E. coli, DpnI was used to remove the methylated template plasmid containing the full MBNL1 gene. Two separate KLD reactions were set up, one with 1 μL of PCR product and one with 2 μL of PCR product. Both reactions were incubated at room temperature for 10 min following the addition of KLD.

Competent C304OH cells were thawed on ice for 10 min, after which 1μL of each of the KLD reactions was added to two separate vessels of competent cells. Cell mixture was gently flicked 4-5 times to mix cells and DNA without vortexing and placed on ice for 30 min. Competent cells underwent heat shock at exactly 42°C for 30 sec, and were placed on ice for 5 min. After 5 min, 50 μL of competent cell mixture was added to 950 μL of stable outgrowth medium in separate incubation tubes and incubated at 37°C for 60 min with shaking. During the incubation, agar plates

were warmed via incubation at the same temperature. Following the hour incubation, 50 μL of competent cells were spread onto two separate plates and incubated overnight at 37°C.

Colonies were then chosen from the plate with the 1 μL KLD reaction (which was less overgrown) and inoculated in 5 mL LB with ampicillin to final concentration of 100 μg/mL for 18h with shaking.

Colonies were then miniprepped according to the following protocol: Bacterial overnight culture was pelleted by centrifugation at 8000 rpm for 3 minutes at room temperature. The supernatant was discarded, and the pellet was resuspended in 250 μL Buffer P1 and transferred to a microcentrifuge tube. 250 μL Buffer P2 was added and mixed thoroughly by inverting the tube until the solution turns blue, not allowing it to proceed past 5 min. 350 μL Buffer N3 was then added and mixed immediately by inverting the tube thoroughly. The suspension was centrifuged for 10 min at 13,000 rpm in a table-top microcentrifuge. 800 μL of supernatant was added to QIAprep 2.0 spin column and centrifuged for 30-60 s at 13,000 rpm and the flow-through was discarded. 0.75 mL Buffer PE was added to wash the column and spun down for 1 min at 13,000 rpm. The flow-through was discarded and the spin column was centrifuged for 1 min to remove residual wash buffer. The spin column was then transferred to a clean 1.5 mL microcentrifuge tube and eluted with 50 μL HPLC-grade water. Miniprepped samples were confirmed via Sanger sequencing.

Samples were then transformed into competent cells following the same procedure as above, but outgrowth medium (50 μL) was added directly into liquid culture (10 μL) without plating and grown for 18h at 37°C with shaking.

Following the incubation, OD_{600} was measured for the cultures, using LB as a blank. The cultures were then diluted with LB to a final OD_{600} of 0.06 in 500mL LB cultures. OD_{600} was measured roughly every 20 min until it had reached 0.6. IPTG was then added to a final concentration of 1 mM and samples were incubated for 2h at 37°C, shaking at 250 rpm. Cultures were then transferred into 500mL tubes and spun down at 5000 x *g* for 2 min. Supernatant was discarded and pellet was resuspended in 40 mL of 0.5 M NaCl, 0.5 mM Tris HCl (pH = 8.0), transferred to a 50 mL falcon tube, and centrifuged. Supernatant was discarded and cell pellet was stored at -80°C.

Cells were completely resuspended in 500 μL lysis buffer and agitated for 2 h at 4 °C and then centrifuged at 16000 x *g* for 20 min. The supernatant was collected, and pellet was discarded.

The Ni-NTA column was first recharged by washing the column two times with 8 mL stripping buffer (50 mM EDTA). The column was then washed two times with 8 mL distilled water and then recharged two times with 8 mL nickel solution. The column was then washed two more times with 8 mL distilled water.

10 mL binding buffer was put through the column and discarded. Then the cell lysate (10 mL) was pushed through the column, collecting the flowthrough for binding efficiency testing in a

later SDS-PAGE. 10 mL of Wash Buffer 1 and Wash Buffer 2 were pushed through the column, collecting the flowthrough for binding efficiency. Finally, protein was eluted with 15 mL Elution buffer and collected in a Falcon tube.

A 50% by volume slurry of glutathione Sepharose 4B beads and ethanol was created for GST purification. The slurry was sedimented by centrifugation at 500 x *g* for 5 minutes, and supernatant was discarded. The Glutathione Sepharose 4B beads were washed by adding 5 mL Binding buffer to each 1 mL of bead slurry and inverted to mix. The slurry was sedimented again by centrifugation at 500 x *g* for 5 minutes and supernatant was discarded. The wash step was repeated one more time.

The cell lysate was added to the glutathione Sepharose 4B beads and incubated overnight using end-over-end rotation. Following the incubation, a plastic pipette was used to transfer the mixture to a centrifuge tube and mixture was sedimented by centrifugation at 500 x *g* for 5 min. The supernatant was decanted and saved for measuring the binding efficiency in a later SDS-PAGE. Glutathione Sepharose 4B beads were washed by adding 5 mL Binding buffer to each 1 mL of bead slurry, inverting to mix, sedimenting by centrifugation at 500 x *g* for 5 minutes, and decanting the supernatant and saving it for binding efficiency. Wash step was repeated three times.

The bound protein was eluted by addition of 80 μL PreScission Protease and 920 μL PreScission cleavage buffer for each mL of Glutathione Sepharose 4B bed volume. The mixture was then incubated for 4 h using end-to-end rotation. Following incubation, the mixture was centrifuged at

500 x *g* for 5 minutes and the eluate, containing our protein of interest, was carefully transferred to a new tube. The PreScission Protease contains a GST tag and therefore remains bound to the Glutathione Sepharose 4B beads and will not be present in the eluate.

Protein purification was confirmed via SDS-PAGE. Loading buffer was created by combining 475 μL loading buffer with 25 μL β-mercaptoethanol. 20 μL of protein was combined with 20 μL loading buffer and denatured at 95°C for 2 min. 20 μL of each sample was loaded into the wells. Gel was run at 100 V for 5 min, then 150 V until the ladder was well spread out.

A Bradford microassay was then run to determine the concentration of our protein. 1 mL of Bradford's reagent, an acidic version of Coomassie blue G-250, was combined with 20 μL of BSA standards (1 mg/mL, 750 μg/mL, 500 μg/mL, 250 μg/mL, 125 μg/mL) as well as two 20 μL of our protein sample, each in a 1 mL cuvette. Solutions were allowed to incubate at room temperature for 10 min, and then the OD_{600} was measured for each cuvette, and a LSRL was generated, from which concentration of protein was calculated to be roughly 754.2 μg/mL.

Protein was diluted to 20 μL aliquots of 18 μM MBNL1, frozen using liquid nitrogen, and stored at -80°C.

2.2 RNA transcription and purification

The plasmid containing *Htt40* (the *huntingtin* gene containing 40 CAG repeats in the first exon) flanked by the T7 promoter and terminator was transformed into competent C304OH cells using heat shock, plated and then grown overnight at 37°C.

Colonies were then picked from the plate and inoculated in a liquid culture for 18h at 37°C with shaking. Six samples were miniprepped and sent out for Sanger sequencing. One sample showed repeat expansion (it had 64 repeats), but the others contained the proper 40 repeats.

Samples with the correct sequence were treated with SphI to linearize the DNA product, incubated at 37°C for 30 min, and then 60°C for 10 min, in a PCR thermocycler.

Linear DNA was then purified with five volumes of Buffer PB (250 μL) per one volume DNA (50 μL) in a microcentrifuge tube and mixed and then transferred to QIA quick column. Column was centrifuged for 13,000 rpm for 1 minute, and flow-through was discarded. 750 μL Buffer PE was added to the column and column was centrifuged for 1 min. The flow-through was discarded and the column was centrifuged again for 1 min. The column was transferred to a clean microcentrifuge tube and eluted with 50 μL HPLC-grade water. All five samples confirmed via Sanger sequencing were pooled and concentration was determined via NanoDrop.

Linearized DNA template was combined with 10 mM NTPs, T7 RNA polymerase reaction buffer, and 2 μL T7 RNA polymerase. Reaction was incubated at room temperature for 1h, after which treatment with DNase I removed the remaining linear DNA fragment. Reaction was incubated at 37°C for 10 minutes and then treated with EDTA to stop the DNase I reaction. Following the transcription of RNA, polyacrylamide gel electrophoresis (PAGE) was run using to confirm that we have the correct RNA product. 8 M urea was the denaturing agent, and an 8% gel was used. Gel was pre-run for 1 h before sample was loaded and run for an additional 2 h.

The gel was then imaged on an X-ray intensifying screen and a single band was seen. The band was cut out of the gel with a razor blade and placed in a microcentrifuge tube with 400 μL gel elution buffer, which incubated at 4 °C overnight with end-to-end rotation. After incubation, the tube was centrifuged for 10 minutes at 13000 rpm and supernatant containing RNA was drawn off and saved. The RNA was precipitated out with 2.5 volumes of ethanol, which was then vacuum dried off and RNA was resuspended in HPLC-grade water.

2.3 Electrophoretic mobility shift assays

Our *Htt40* RNA contains SA5 regions on both sides of the $(CAG)^{40}(CCG)^{12}$ repeat segment. Oligonucleotides antisense to the SA5 region of our RNA with a Cy5 fluorescent label attached were annealed to our *Htt40* RNA for visualization in the gel (50 μL reaction). Annealing occurred in the thermocycler at 85°C for 10 min, then cooled to 65°C over 90 min, cooled to 55°C over 90 min, and then cooled to 4°C, in order to limit conformational heterogeneity in the RNA.

RNA was incubated with differing concentrations of MBNL1 (2 μM to 25 nM) at room temperature for 30 minutes before insertion into the wells. For D6 trials, the RNA was incubated with D6 for 20 minutes at room temperature before MBNL1 was added. D6 was dissolved in DMSO, and so a control lane of DMSO without RNA was run on the gel to detect any effects DMSO had on the gel.

For time trial, RNA samples were mixed in one pot. First, RNA containing no MBNL1 were first loaded to act as a '0' time point. MBNL1 was then added to the pot, and 5 μL fractions were taken out at each respective time point.

All gels were run with 1X TBE, with loading buffer in one lane to monitor the progress of the bands. Gels were run at 15-18 V for 2-3 h, until the xylenes cyanol tracking dye was roughly halfway down the gel. The gel box often leaked, and extra care was taken to continually refill the apparatus with 1X TBE buffer. Gels were imaged on Amersham Typhoon gel imager.

2.4 Single Molecule Force Spectroscopy Assays

Htt40 RNA was annealed to DNA handles that were biotinylated at the end. After annealing reaction, streptavidin beads were incubated with RNA for 10 minutes prior to introduction into the fluidics chamber of the optical tweezers. For studying the MBNL1-RNA complex, MBNL1 was then added and incubated. Originally, samples were incubated with MBNL1 for 30 minutes before dilution and insertion into the fluidics chamber. However, EMSA data showed that MBNL1 binding occurred within roughly 15 seconds of insertion, and so subsequent MBNL1 samples were only incubated for 15 seconds before being diluted and inserted.

The fluid channel containing the beads was first opened until two beads had been caught by the 1026nm lasers, as monitored via microscopy. Next, the bead channel was closed, and the beads currently caught in the lasers were brought close enough to each other to determine a baseline force. Once the baseline had been calibrated, beads were slowly pulled apart, monitoring the force curve generated. Beads were then slowly brought back together, and this was done several

times per bead set to generate data. After roughly 10 pulls, the beads were released and a new set of beads was captured to repeat the experiment.

RESULTS AND DISCUSSION

To probe the effects of MBNL1 binding to RNA on the single molecule level, we first expressed and purified MBNL1 protein. Following the purification using a nickel column and glutathione beads, a polyacrylamide gel using Coomassie blue dye demonstrated that MBNL1 was isolated and pure enough to use for future assays (Figure 1).

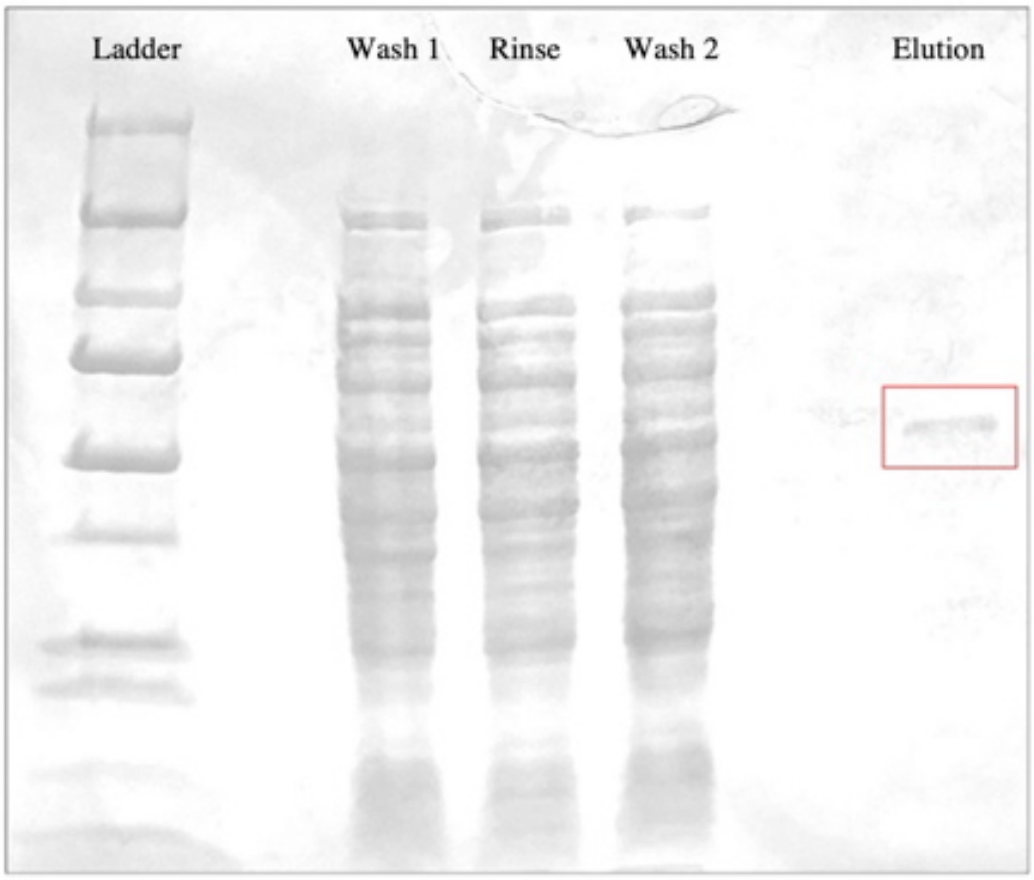

Figure 1. Elution of purified MBNL1 protein. A clear band occurs at our protein of interest, roughly 28.5 kDa. Other steps in elution process are shown to indicate that minimal protein was lost before elution.

The concentration of the expressed MBNL1 was necessary to prepare samples for the single molecule spectroscopy assays. A Bradford assay was run, indicating a concentration of 754.2 ug/mL which was high enough concentration for use in subsequent steps (Figure 3).

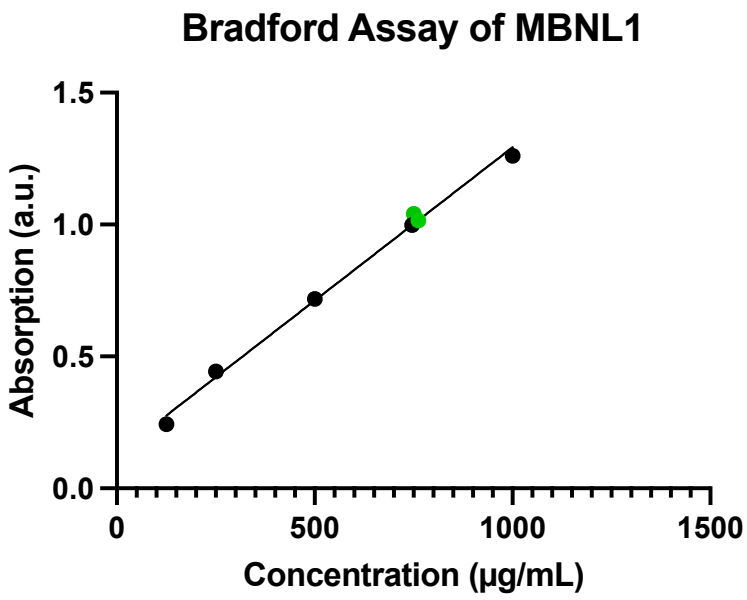

Figure 2. MBNL1 protein concentrated is 754 *µ*g/mL. MBNL1 absorption measurements shown in green. R^2 of LSRL is 99.40%.

In addition to MBNL1, pure RNA containing $(CAG)^{40}$ was synthesized and shown to be pure, as a single band appeared in the denaturing gel.

With the MBNL1 protein and RNA transcript synthesized, we next used electrophoretic mobility shift assays (EMSA) to determine a general kinetic understanding of the rate of binding of the MBNL1 to the RNA. A time-course gel was run, indicating that most of the MBNL1 had aggregated with the RNA within ten seconds, causing it to stay in the well (Figure 3).

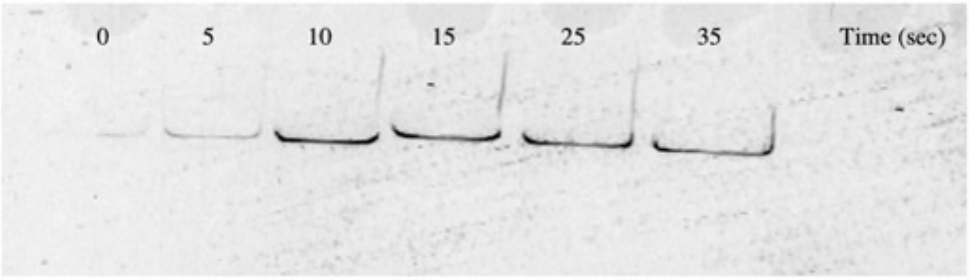

Figure 3. MBNL1 binds to RNA within roughly 10 seconds. Only the wells were imaged as no bands were visible.

We then ran a concentration-gradient gel to estimate the K_d of MBNL1 binding to RNA, as this information would allow us to determine an ideal concentration of protein to use for single molecule experiments. Our gel showed that concentrations at 64 nM and higher tended to aggregate enough to remain in the wells, indicating a K_d in this range. No second band above the free RNA band was seen, likely because the gel percentage (6%) and the acrylamide:bisacrylamide ratio (29:1) was too high to see the shifted band (Figure 4). Comparing the ratio of the K_d of MBNL1 and the concentration of RNA (64nM: 5nM), we determined that a stoichiometric ratio of roughly 13 molecules of MBNL1 per molecule of RNA caused the system to become sufficiently aggregated, such that it would no longer travel in the gel. Such a stoichiometric ratio is important in the calculations of concentrations for single molecule assays. If there is excess MBNL1 protein in the channel with RNA, it is likely that the optical tweezers will isolate MBNL1 molecules instead of RNA molecules, which will give a very different unfolding pattern that, for the purposes of this study, we are not interested in studying. In addition, concentration ratios are important to be sure that D6 can compete with MBNL1 for binding.

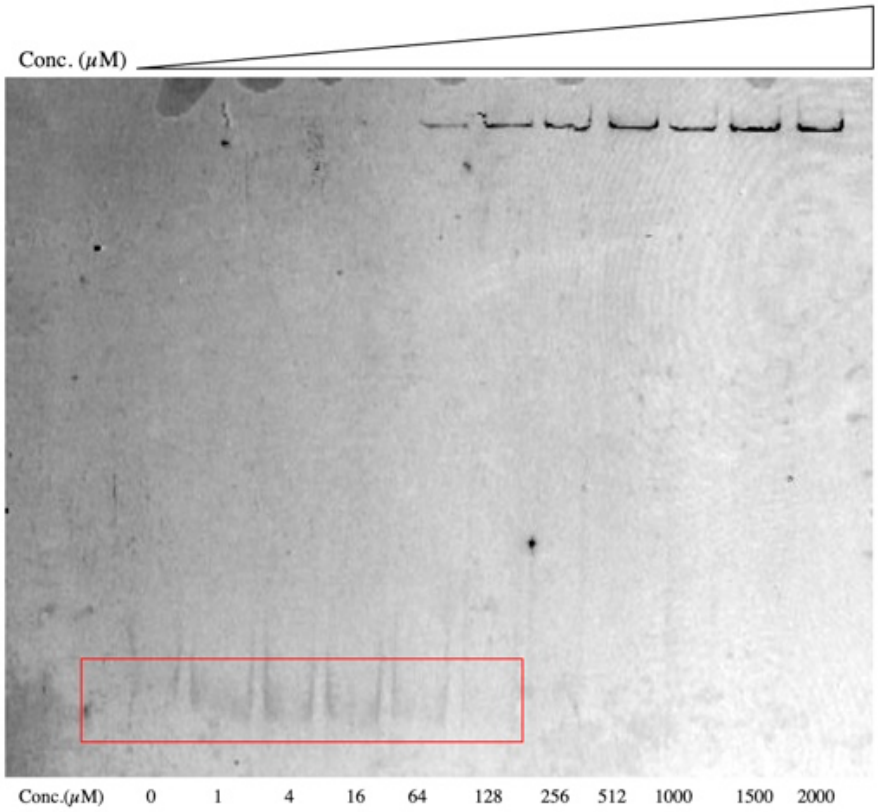

Figure 4. Gel shift of MBNL1 occurs at roughly 64 nM. Intermediate bands are likely not visible due to high percentage of gel (6%). Free RNA bands are indicated in red box.

With a rough estimate of the binding concentration and time required for MBNL1-RNA complexes to form, we then aimed to demonstrate the effect of D6 on MBNL1-RNA binding on a macroscopic level.

Precursor gel shift assays demonstrated that D6 has an effect on MBNL1-RNA complex formation, although it was minor (Figure 5). All of the wells showed heavy aggregation in the wells, regardless of treatment with D6, perhaps due to the high gel percentage, as well as acrylamide:bisacrylamide ratio. In addition, 3μM D6 was used as previous literature showed that

to be the IC_{50} of RNA binding, but in order to see a more pronounced effect, we did a new gel with a higher concentration of D6.

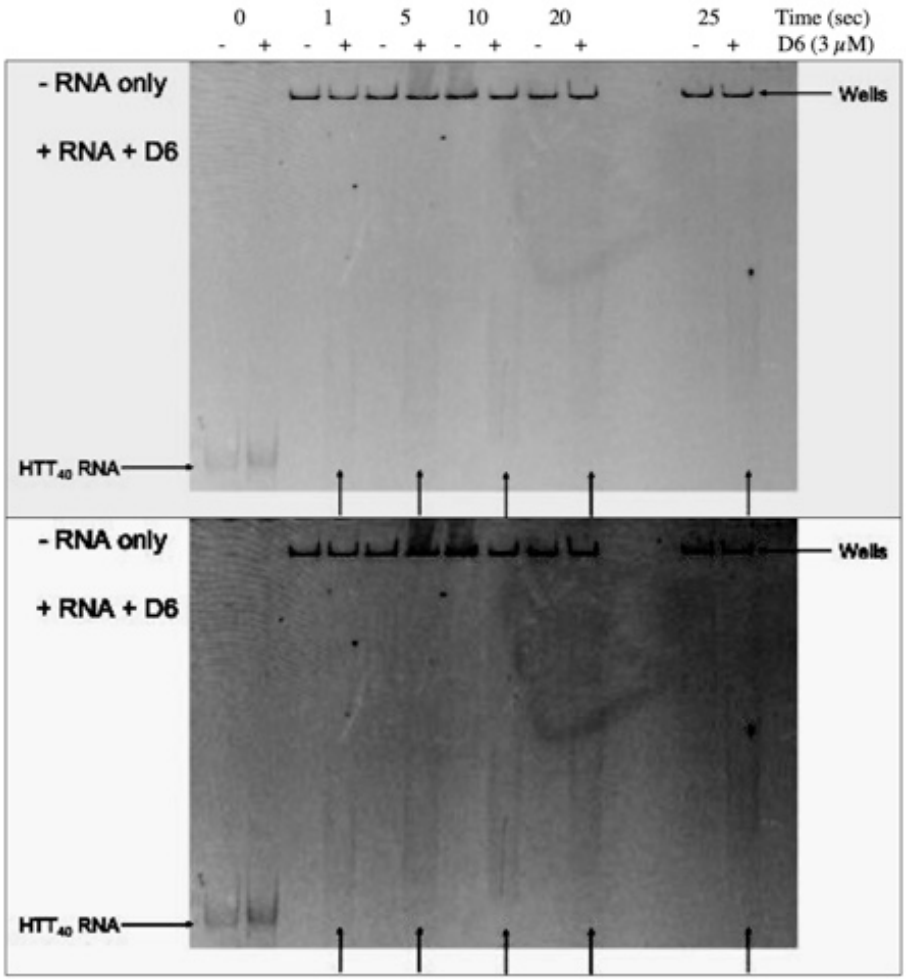

Figure 5. D6 shows slight destabilizing effect on MBNL1-RNA complex formation. IC_{50} concentration was used for the gel but it was too little to see a significant effect. Bottom picture has increased contrast.

Another gel was run (4%, 39:1 ratio) to allow the RNA to travel more easily through the gel, allowing better observation of shifted bands (Figure 6). A concentration gradient of MBNL1 was added to RNA with and without treatment. Although significant aggregation was still seen in the wells at higher concentrations, the intensity of the free RNA bands were quantified using ImageJ,

showing that D6 did inhibit MBNL1-RNA complex formation (see Table 1 in Supplementary Materials).

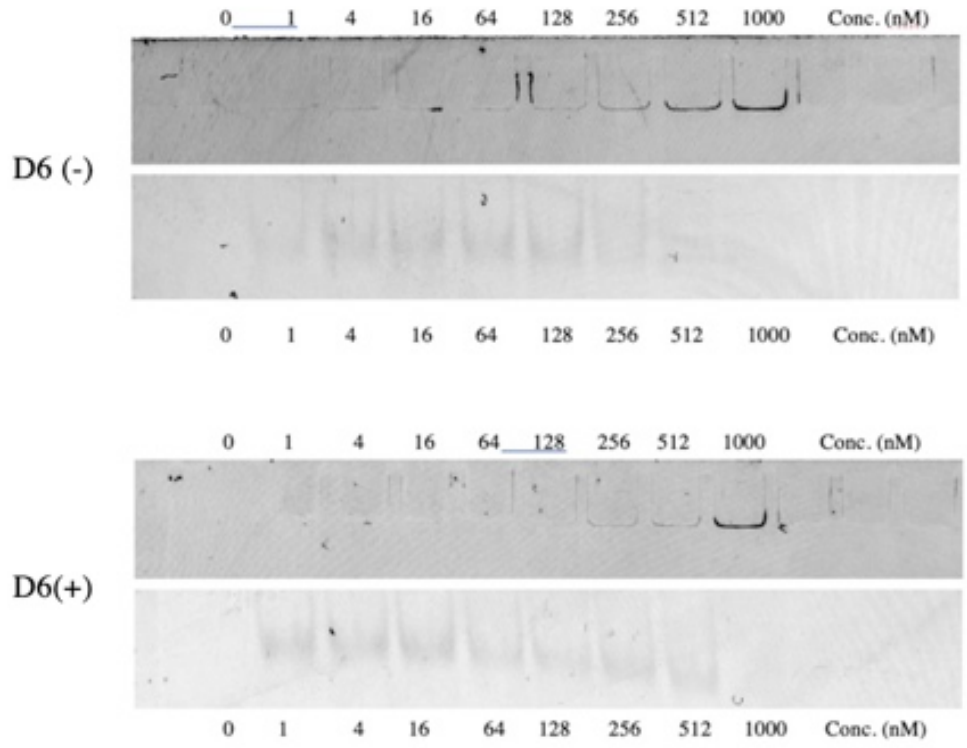

Figure 6. D6 decreases the intensity of free RNA band. An increase of intensity of free RNA with treatment of D6 indicates that it is interrupting the binding of MBNL1 to RNA, and therefore preventing aggregation in the wells.

With a macroscopic understanding of the interactions between MBNL1 and RNA, as well as the effect of D6, we began to probe the interactions at a single molecule level. First, free RNA was loaded into the C-trap and analyzed. Roughly 100 molecules of free RNA were analyzed, and the results were compared. A representative picture can be seen in Figure 7. A large increase in distance was seen at a force of roughly 13 pN, after which the plot appears quite messy until reaching a distance of 525 nm after which it increases exponentially.

We then incubated the RNA molecules with MBNL1 before loading it into the C-trap to test the responsiveness of the RNA-protein complex. From the results of the time course, we determined that within 15 seconds, the majority of the RNA had formed a complex with MBNL1, and as such we only let the RNA and MBNL1 incubate for 15 seconds before subsequent dilution and insertion. As opposed to the free RNA, several RNA-MBNL1 complexes reached a force of roughly 18 pN before a large distance increase event occurred, and the plot demonstrating this event appears much smoother than the free RNA.

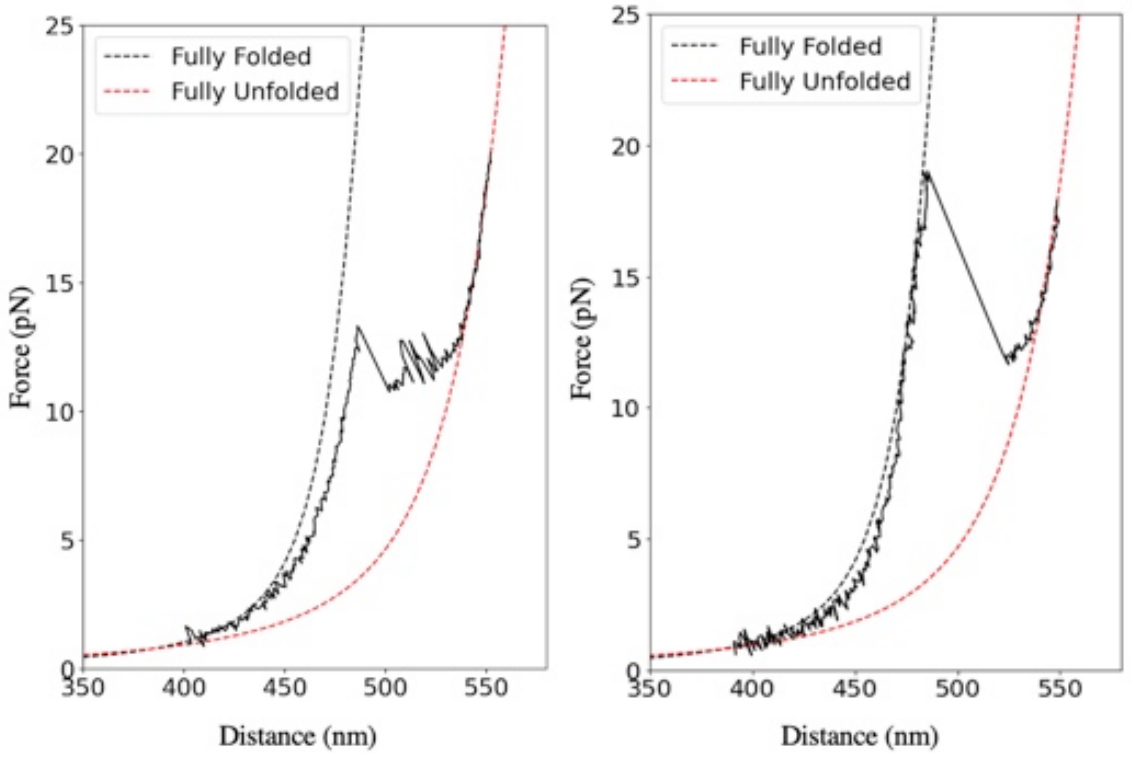

Figure 7. Increased stability of MBNL1-RNA complex compared to free RNA. MBNL1 bound to RNA reached higher forces (18 pN) before unfolding, compared to free RNA (13 pN).

CONCLUSION AND OUTLOOK

Our observations support a model in which D6 is able to compete with MBNL1 for binding to $(CAG)^n$-repeat containing RNA. Work done by Kumar et al. demonstrated that the IC_{50} of D6 was 3 μM; however, our gel shift experiments using that concentration of D6 had minimal effect on the RNA-MBNL1 binding. We were only able to see substantial changes at much higher concentration, roughly 50 μM. This would indicate that perhaps a significant portion of the D6 is unable to intercalate into the RNA, and as such future use of D6 as a small molecule treatment would require careful finetuning of D6 concentration to ensure the proper effect. It is also likely that D6 concentration would need to increase with the number of (CAG) repeats, showing the importance of advances in personalized genomic medicine.

Our results also point to a model in which the binding of MBNL1 to RNA stabilizes the folding complex. In the free RNA plot, the distance between the two beads fluctuated heavily around 13 pN of force, indicating that there were several folding and unfolding events, as folding would cause the distance to decrease while unfolding would cause the distance to increase. However, upon treatment with MBNL1, there are significantly less folding events; the plot appears to show that a single unfolding event occurs when a threshold force of 18pN is reached, at which the RNA molecule goes from folded to completely unfolded. Such a conclusion, although possible from our data, would need to be corroborated by more data gathering on the optical tweezers for a representative sample size.

As my research was collected during my time as an Amgen Scholar over the summer, my time was limited and I was only able to gather data to probe elementary insights into repeat RNA structure. Future work (currently being done by members of the Woodson lab) aims to further characterize the MBNL1-RNA complex through force spectroscopy, as well as study the effect of D6 on the MBNL1-RNA complex on a single molecule level.

In summary, we studied the effect of the small molecule D6 on the binding of MBNL1 to RNA containing (CAG) repeat structure. We observed a quantitative effect on the binding of MBNL1 in the presence of D6, indicating its use as a potential alleviation for the RNA-based pathogenesis of Huntington's Disease. In addition, we observed a distinct unfolding pattern in RNA-MBNL1 complex compared to free RNA on the single molecule level using force spectroscopy techniques, likely due to a stabilizing effect that results from the binding of the protein. Such discoveries provide more insight into the mechanism by which MBNL1-RNA associate with each other, allowing us to discover new ways to limit or disrupt the pathogenesis of HD. Such a treatment would provide a way to prevent the disease progression, improving the quality of life of thousands of people.

APPENDIX

MBNL1N_Forward: 5'-CAT CAT CAT CAT CAT CAT CAT TGA TAA GAA TTC CGG CCG – 3'

MBNL1N_Reverse: 5'-TTG TGC AGC TGC AGC CTG G – 3'

Buffer Concentrations:

Lysis Buffer (500 mL, pH 8.0): 25 mM Tris HCL**,** 500 mM NaCl**,** 10 mM Imidazole**,** 2 mM BME**,** 5% glycerol**,** 0.1% Triton-X100**,** 2 mg/mL lysozyme**,** Protease inhibitor tablet

Wash Buffer 1 (500 mL, pH 8.0): 25 mM Tris HCl**,** 500 mM NaCl**,** 20 mM Imidazole**,** 0.1% Triton-X100

Wash Buffer 2 (500 mL, pH 8.0): 25 mM Tris HCl**,** 300 mM NaCl**,** 5 mM BME**,** 0.1% Triton-X100

Elution Buffer (500 mL, pH 8.0): 25 mM Tris HCl**,** 500 mM NaCl**,** 250 mM Imidazole**,** 0.1% Triton-X100

PreScission Cleavage Buffer (pH 7.5): 50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM DTT

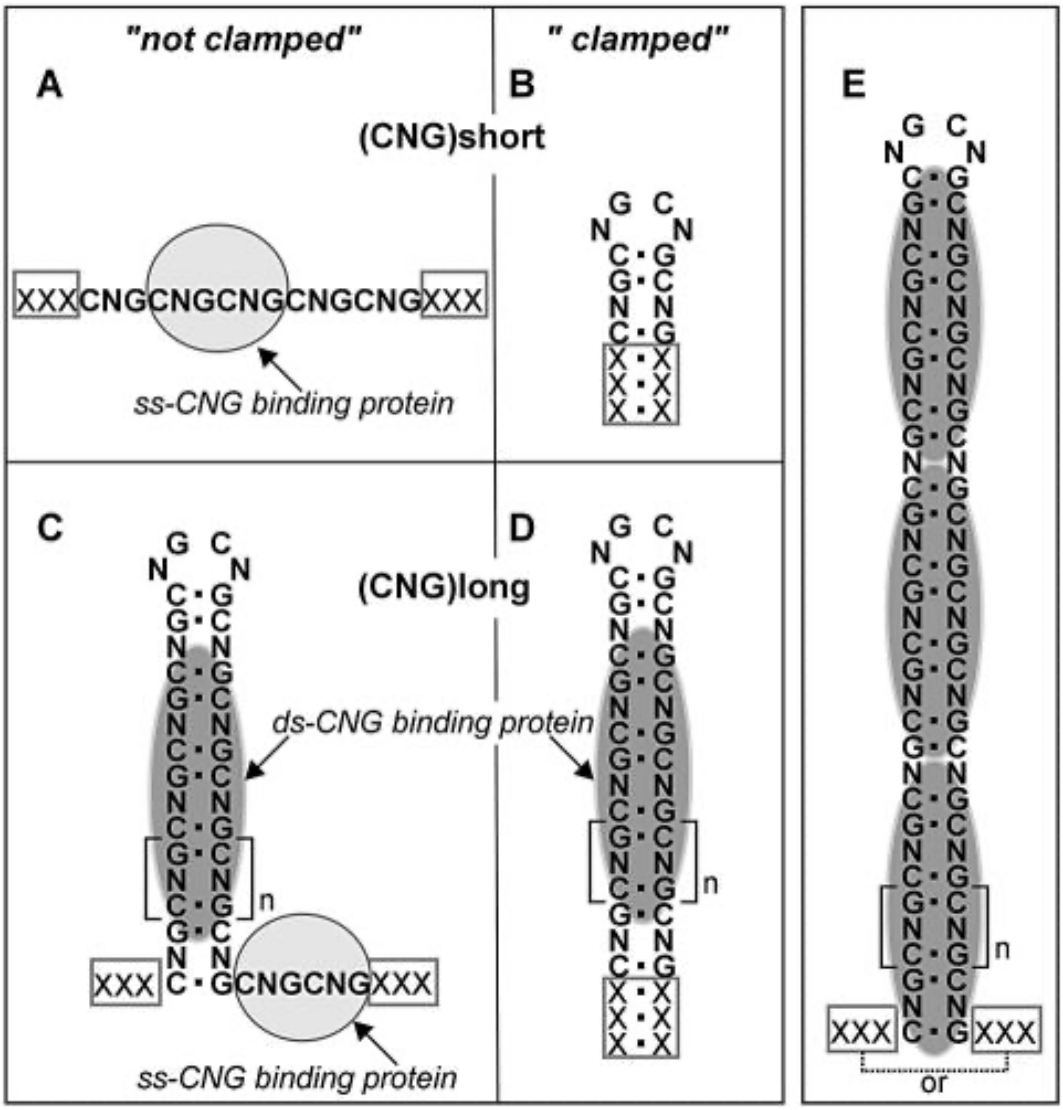

Figure S1. Structure of RNA Hairpin Loops. N represents either an adenine or uracil residue, both of which show proclivity to forming hairpin loops.

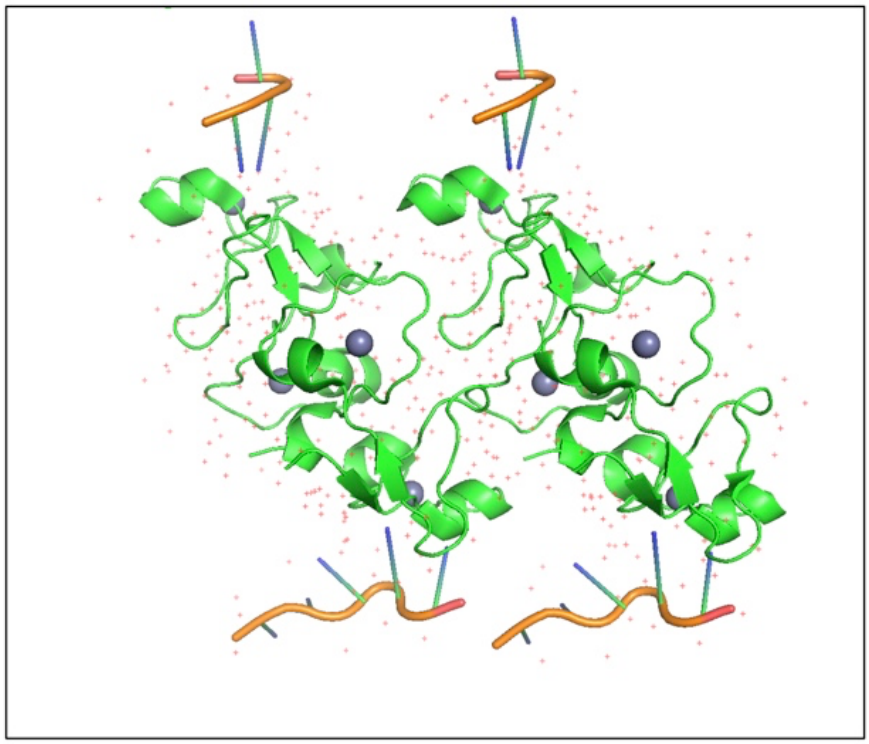

Figure S2. Structure of MBNL1. Structure was determined via protein x-ray crystallography with a resolution of 2.7 Å. Zinc fingers shown binding to nucleic acids shown in orange.

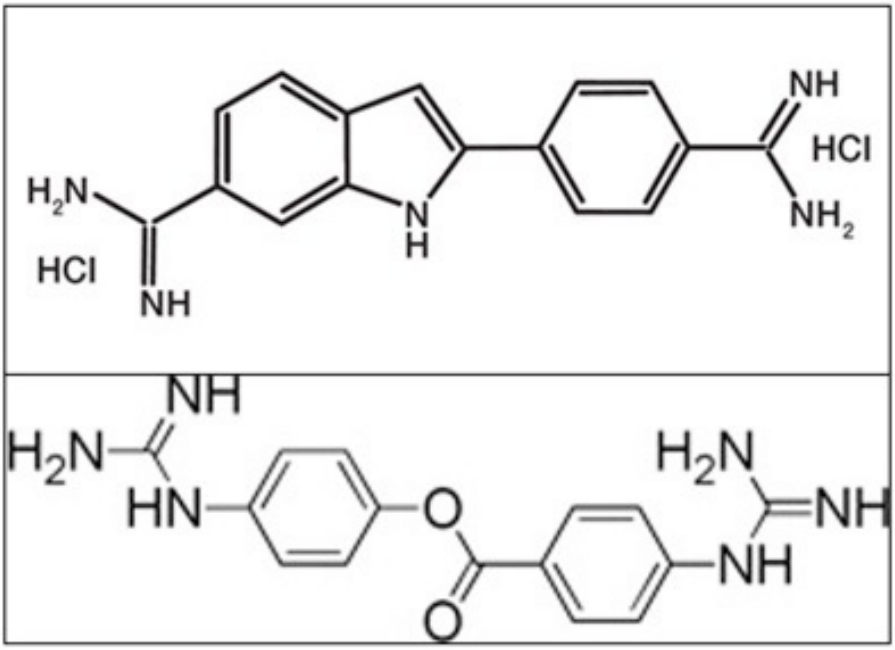

Figure S3. Molecular Structures of DAPI and D6. DAPI is seen above, D6 is shown below. The aromaticity of both molecules allow it to intercalate easily into the π stacking characteristic of nucleic acid polymers.

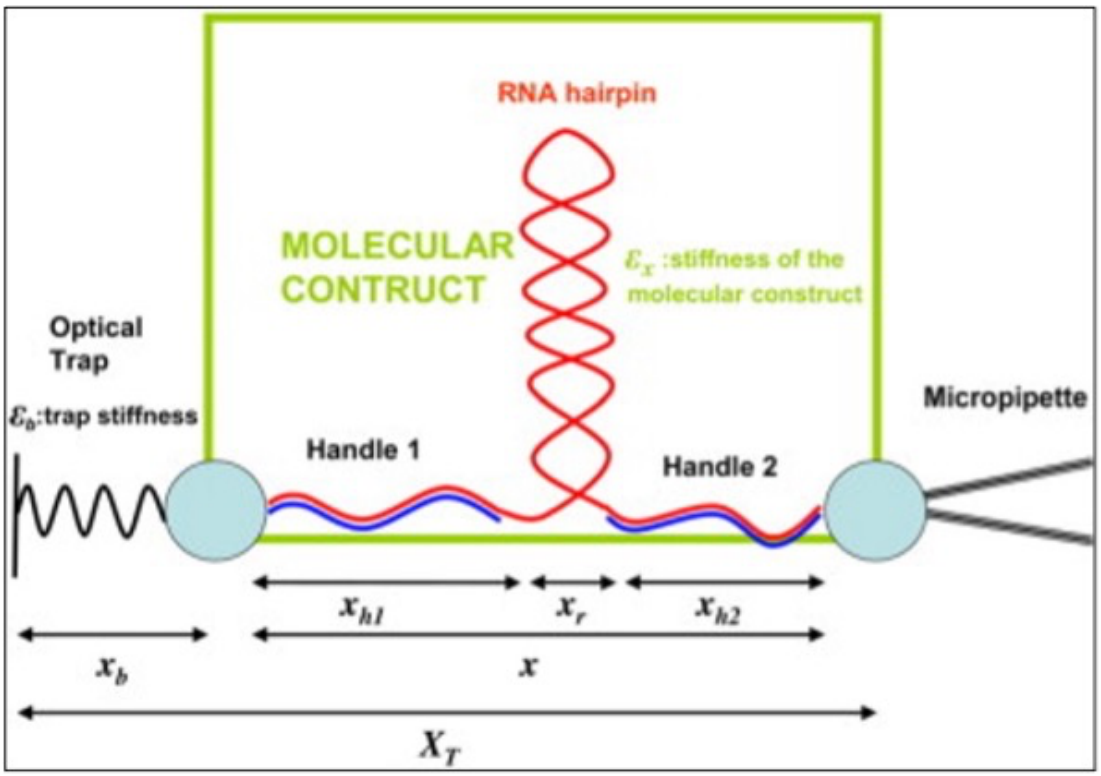

Figure S4. Optical Tweezers Setup with Nucleic Acid. Two beads are shown as cyan spheres attached to two handles which bind to the RNA construct of interest.

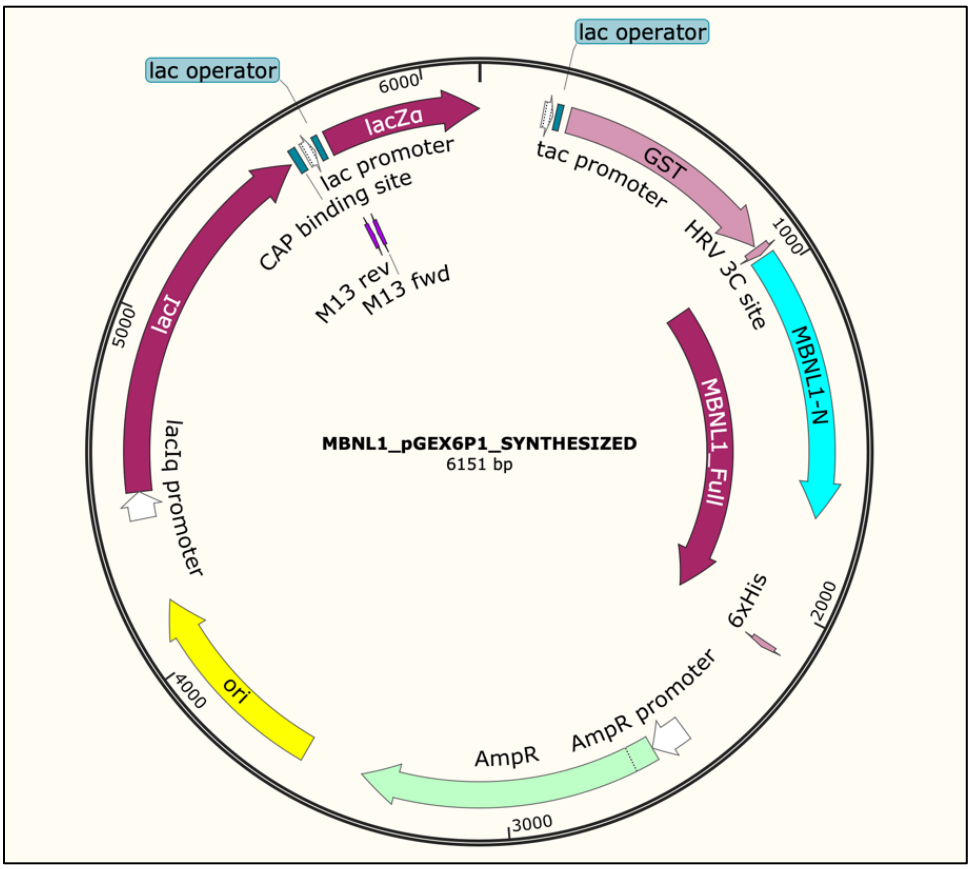

Figure S5. Plasmid map of MBNL1-containing DNA. GST tag and 6xHis tag are included, as well as AmpR used for E. coli selectivity. Red arrows show the primers used to remove the final amino acids of the *Htt* gene.

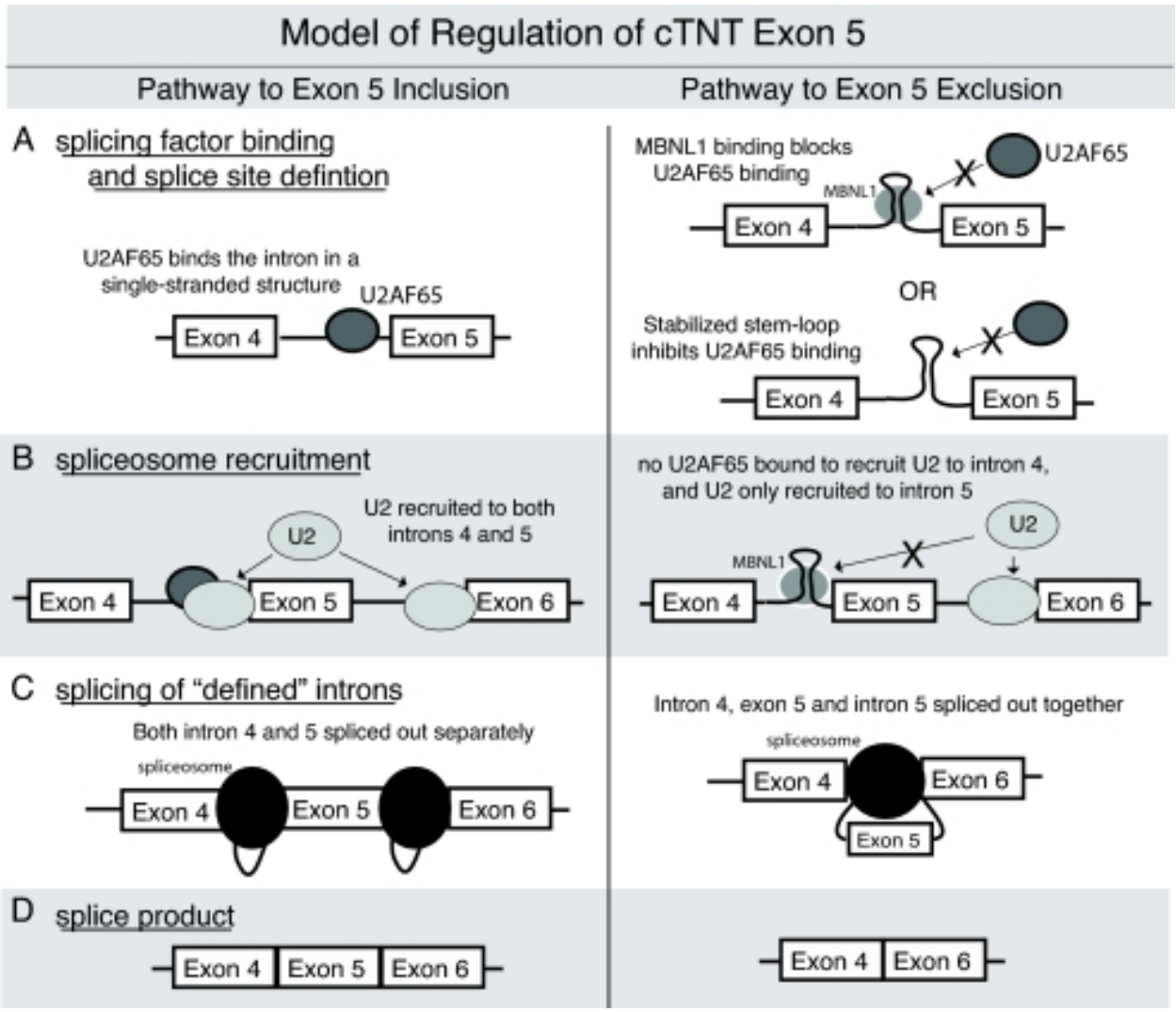

Figure S6. Pathway of MBNL1 Splicing Regulation Function. In the absence of a stem loop, U2AF65 can recruit U2 to all introns in the sequence, including them in the final splice product. If a stem loop is present, MBNL1 can bind and prevent U2AF65 from binding, causing the flanking exon to be excluded from the final splice product.

Table 1. LSRL of Bradford Assay with R^2 value of 0.9940.

Simple linear regression Tabular results	A Absorption (a.u.)
	Y
Best-fit values	
Slope	0.001162
Y-intercept	0.1307
X-intercept	-112.5
1/slope	860.3
Std. Error	
Slope	5.211e-005
Y-intercept	0.03204
95% Confidence Intervals	
Slope	0.0009966 to 0.001328
Y-intercept	0.02874 to 0.2327
X-intercept	-229.4 to -22.03
Goodness of Fit	
R squared	0.9940
Sy.x	0.03730
Is slope significantly non-zero?	
F	497.6
DFn, DFd	1, 3
P value	0.0002
Deviation from zero?	Significant
Equation	Y = 0.001162*X + 0.1307
Data	
Number of X values	15
Maximum number of Y replicates	1
Total number of values	5

Table 2. Higher quantity of free RNA band treated with D6. Quantification was done in ImageJ.

Conc. MBNL1 (μM)	D6(-) Band Intensity (a.u.)	D6(+) Band Intensity (a.u.)
1	3048	14250
4	10124	15196
16	8498	15670
64	7862	9411
128	6478	5026
256	3568	10975
512	1831	8066

REFERENCES

1. Medina, A.; Mahjoub, Y.; Shaver, L.; Pringsheim, T. Prevalence and Incidence of Huntington's Disease: An Updated Systematic Review and Meta-Analysis. *Mov. Disord.* **2022**, *37*, 2327-2335.

2. SUS Paulson, H. Repeat expansion diseases. *Handb. Clin. Neurol.* **2018**, *147*, 105-123.

3. Jimenez-Sanchez, M.; Licitra, F.; Underwood, B. R.; Rubinsztein, D. C. Huntington's Disease: Mechanisms of Pathogenesis and Therapeutic Strategies. *Cold Spring Harb Perspect. Med.* **2017**, *7*, a024240. doi: 10.1101/cshperspect.a024240.

4. Schulte, J.; Littleton, J. T. The biological function of the Huntingtin protein and its relevance to Huntington's Disease pathology. *Curr. Trends Neurol.* **2011**, *5*, 65-78.

5. Jain, A.; Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **2017**, *546*, 243-247.

6. Gusella, J. F.; Lee, J.; MacDonald, M. E. Huntington's disease: nearly four decades of human molecular genetics. *Hum. Mol. Genet.* **2021**, *30*, R254-R263.

7. Budworth, H.; McMurray, C. T. A brief history of triplet repeat diseases. *Methods Mol. Biol.* **2013**, *1010*, 3-17.

8. Schulte, J.; Littleton, J. T. The biological function of the Huntingtin protein and its relevance to Huntington's Disease pathology. *Curr. Trends Neurol.* **2011**, *5*, 65-78.

9. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. Electronic address: gusella@helix.mgh.harvard.edu; Genetic Modifiers of Huntington's Disease

(GeM-HD) Consortium CAG Repeat Not Polyglutamine Length Determines Timing of Huntington's Disease Onset. *Cell* **2019**, *178*, 887-900.e14.

10. Yuan, Y.; Compton, S. A.; Sobczak, K.; Stenberg, M. G.; Thornton, C. A.; Griffith, J. D.; Swanson, M. S. Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res.* **2007**, *35*, 5474-5486.

11. Ho, T. H.; Charlet-B, N.; Poulos, M. G.; Singh, G.; Swanson, M. S.; Cooper, T. A. Muscleblind proteins regulate alternative splicing. *EMBO J.* **2004**, *23*, 3103-3112.

12. Gusella, J. F.; Lee, J.; MacDonald, M. E. Huntington's disease: nearly four decades of human molecular genetics. *Hum. Mol. Genet.* **2021**, *30*, R254-R263.

13. Rook, M. E.; Southwell, A. L. Antisense Oligonucleotide Therapy: From Design to the Huntington Disease Clinic. *BioDrugs* **2022**, *36*, 105-119.

14. Kumar, A.; Parkesh, R.; Sznajder, L. J.; Childs-Disney, J.; Sobczak, K.; Disney, M. D. Chemical Correction of Pre-mRNA Splicing Defects Associated with Sequestration of Muscleblind-like 1 Protein by Expanded r(CAG)-Containing Transcripts. *ACS Chem. Biol.* **2012**, *7*, 496-505.

15. Marantan, A.; Mahadevan, L. Mechanics and statistics of the worm-like chain. *American Journal of Physics* **2018**, *86*, 86-94.

16. Bokinsky, G.; Zhuang, X. Single-molecule RNA folding. *Acc. Chem. Res.* **2005**, *38*, 566-573.