2013

# Diverse Mechanisms of Trinucleotide Repeat Disorders: An Exploration of Fragile X Syndrome and Huntington's Disease

Cara Strobel

# Diverse Mechanisms of Trinucleotide Repeat Disorders: An Exploration of Fragile X Syndrome and Huntington's Disease

CARA STROBEL

Cara Strobel authored this essay for the Cell Biology course in the spring semester of 2012. Given free reign with a cell biology related topic, she wanted to explore and contrast the specifics of several prevalent disorders. Cara plans to apply to medical school in Spring 2013.

Trinucleotide repeat disorders are an umbrella group of genetic diseases that have been well described clinically for a long time; however, the scientific community is only beginning to understand their molecular basis. They are classified in two basic groups depending on the location of the relevant triplet repeats in a coding or a non-coding region of the genome. Repeat expansion past a disease-specific threshold results in molecular and cellular abnormalities that manifest themselves as disease symptoms. Repeat expansion is postulated to occur via slippage during DNA replication and/or transcription-mediated DNA repair. Trinucleotide repeat disorders are characterized by genetic anticipation, which is defined by the increasing severity and earlier onset of a disease as it is inherited through consecutive generations. Through the analysis of Huntington's disease as an example of coding trinucleotide repeat disorder and Fragile X Syndrome as an example of non-coding trinucleotide repeat disorder, this work will explore the nature of this devastating group of diseases and the underlying basic molecular processes that construct them. Despite sharing key characteristics, these diseases differ significantly in the wide variety of havoc the repeat expansions can create depending on their location and the nature of the disrupted function. Understanding of the mechanism and specifics of each individual disease remains critical for development of proper therapies.

## Introduction

Trinucleotide repeat disorders are an umbrella group of genetic diseases that have been well described clinically for a long time; however, their molecular basis is still being elucidated. For example, Huntington's disease (HD), one of the most widely known diseases associated with trinucleotide repeats, was first described at the end of the 19th century, whereas the gene associated with this disorder was not discovered until the mid nineties of the 20th century (Chial, 2008). Currently, all known trinucleotide disorders are classified in two basic groups depending on the location of the relevant triplet repeats in a coding or non-coding region of the genome. The genome of a human cell could be viewed as a sum of the cellular DNA comprising its genetic blueprint. Coding

regions involve very small fraction of the human genome that codes for proteins. Protein molecules build most of the organism structures and function as working horses performing molecular work that underlies cellular physiology. The vast majority of human DNA does not code for proteins, thus comprising non-coding regions involved in orchestrating the regulation of protein expression. The execution of the genetic program of the human cell depends on the proper function of both coding and non-coding regions of the genome. As a general principle, mutations in coding regions inflict changes of protein sequence and function, whereas mutations in non-coding regions result in changes in the protein levels in the cell. Cells that are forced to work with altered or missing proteins exhibit abnormal physiology which on the organismal level is observed as disease symptoms.

Trinucleotide repeat disorders associated with coding regions include Huntington's disease, Kennedy disease (Spinobulbar muscular atrophy), Haw-River Syndrome (Dentatorubral-pallidoluysian atrophy) and five different types of Spinocerebellar ataxia. These diseases are also called polyglutamine diseases since all of them are associated with repeats of CAG triplets, which code for the amino acid glutamine. The pathology of all diseases in this group involves pathologies of the nervous system (Cummings and Zoghbi, 2000).

Non-coding trinucleotide repeat disorders include Fragile X syndrome (associated with CGG repeats), Fragile XE syndrome (associated with GCC repeats), Friedreich's ataxia (associated with GAA repeats), Myotonic Dystrophy (associated with CTG repeats), Spinocerebellar ataxia type 8 (associated with CTG repeats) and Spinocerebellar ataxia type 12 (associated with CAG repeats). Some of these disorders have been found to be neurodegenerative (Fragile X syndrome) while others, like myotonic dystrophy, affect muscle maintenance (Cummings and Zoghbi, 2000).

Both groups of trinucleotide repeat disorders share common molecular features and most likely common mechanisms of repeat extension that are directly connected to the flow of genetic information in the cell. The DNA as a genetic blueprint is securely stored in the nucleus of the human cell and constantly monitored for its integrity. DNA replication, or DNA copying, takes place only before cellular division, to ensure that exactly the same genetic information is passed to the resulting offspring. Since proteins are needed continuously in the cell and the protein synthesizing machinery resides outside the nucleus, the genetic code is first copied into messenger RNA (mRNA) via a process called transcription. mRNAs are easily transported outside the nucleus and then recognized by the protein-making machines, the ribosomes, which translate

the genetic information into proteins. Any changes that take place in the DNA blueprint such as mutations can be subsequently traced as changes in RNA and could potentially affect the integrity of the corresponding proteins.

All trinucleotide repeat disorders are associated with genes whose natural sequence includes multiple repeats. By their nature, repeats present a challenge for the cellular machinery handling the DNA or RNA molecules that contain them. One can rationalize the challenge by comparing the error rates of reading a page of repetitive text versus a page of non-repetitive text. The likelihood of skipping or repeating a line by accident is much greater for repetitive text than for the non-repetitive one. Similarly, when DNA repeats are replicated or transcribed, there is an increased chance of error that will result in change of the number of repeats. Trinucleotide repeat disorders are associated with repeat expansion. One proposed mechanism for the expansion has to do with strand slipping during DNA replication. Petruska, Hartenstine, and Goodman (1997) conducted experiments with synthetic DNA containing trinucleotide repeats and concluded that repeated sequences promote slippage via DNA looping. In a region of repeats, separation of the strands of DNA during replication can cause single-stranded loops of repeats. When the looping takes place in parental DNA strand, the net result is reduction of the number of the repeats, simply because the repeats in the loop are skipped and not copied. If the newly replicating DNA loops out, the net result is repeat expansion since the repeats in the loop are copied twice. Alternative mechanism involves transcription-mediated DNA repair pathways. Repeated sequences confuse the transcription machinery, which stalls and activates the DNA repair systems of the cell that occasionally make an error leading to repeat expansion. It is thought that the DNA slippage mechanism is prevalent in actively dividing cells, whereas the transcription-mediated DNA repair mechanism is prevalent in non-dividing cells, such as neurons, for example. The repeat expansion can take place in both somatic and germ cells, thus offering an explanation why the severity of the trinucleotide repeats disorders can sometimes change during the course of the disease in a particular individual and between generations. As a consequence of the described molecular events, trinucleotide repeat disorders share a common phenomenon named genetic anticipation, which is defined by the increasing severity and earlier onset of a disease as it is inherited through consecutive generations.

Through the analysis of Huntington's disease as an example of coding trinucleotide repeat disorder and Fragile X Syndrome as an example of non-coding trinucleotide repeat disorder, this work will explore the nature of this devastating group of diseases and the underlying basic molecular processes that construct

them. Despite sharing key characteristics, such as trinucleotide repeat expansion and genetic anticipation, these diseases differ significantly in the wide variety of havoc the repeat expansions can create depending on their location and the nature of the disrupted function.

## Molecular Mechanisms and Pathology of Huntington's Disease

Huntington's disease (HD) is one of the best known disorders within the polyglutamine diseases. It is an autosomal dominant disorder caused by the polyglutamine repeat expansion within the huntingtin protein (Htt) encoded by the IT15 (interesting transcript 15) gene located on chromosome 4. Individuals are affected by HD even if they inherit only a single abnormal copy of the gene. HD affects 1/10,000 individuals and is characterized by neurodegeneration. Specifically, it involves the depletion of neurons and an increased number of glial cells, primarily in the striatum region of the brain associated with movement planning, working memory, and decision-making. Other regions like the cortex, thalamus, and subthalamic nucleus are also affected (Sadri-Vakili and Cha, 2006). This atrophy of certain brain regions causes profound effects on the individual, but not until the fourth or fifth decade of life. An unfortunate truth of this, however, is that by the time the disease presents itself in symptoms, the mutated gene has likely been passed on to the next generation.

The three major types of symptoms associated with Huntington's are behavioral, motor, and cognitive. The disease presents itself as a progressive deterioration, usually taking the life of the affected individual 15-20 years after diagnosis (Longshore and Tarleton, 1996). Behavioral symptoms are the first to present themselves in the form of mood changes. Next, motor symptoms begin to occur in the extremities through involuntary twitching and loss of coordination. In conjunction with the motor symptoms, cognitive symptoms become more noticeable and complicated tasks become progressively more difficult to think through. As the disease advances, the severity of these symptoms becomes more intense to the point that the involuntary movements appear more like a frantic dancing as opposed to twitching. Changes in mood continue to develop though some people with HD become more apathetic in regards to their disease, losing interest in activities they used to enjoy. And finally, it becomes progressively harder for HD patients to think clearly as the disease worsens. In late stage HD, patients lose their involuntary movements and begin to experience rigidity and the lack of ability to move voluntarily at all.

It is well known in HD that the number of repeats is an important component of the disease. Huntingtin (Htt) is a natural protein found in numerous areas of the brain of all individuals. All of the Htt proteins have the polyglutamine repeats, but the number of repeats can affect the onset, progression, and overall severity of the disease (Figure 1). Normal individuals have between 7-34 CAG repeats. Individuals with more than 40 repeats develop HD. If the number of repeats is over 70, a juvenile onset of HD is very likely. It has been observed in clinical settings that the severity and percentage of early onset of HD increases as the mutation is passed through
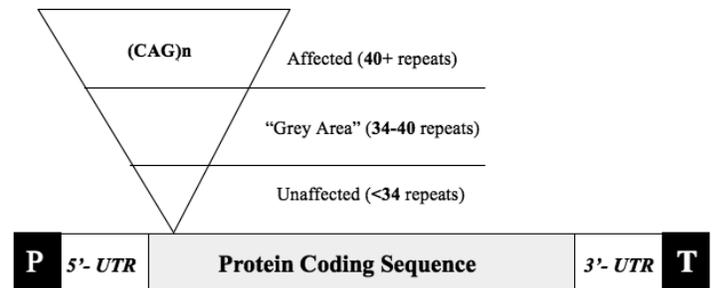


**Figure 1. *Htt* gene structure and relationship to HD pathology.**
CAG repeat expansions related to HD are located within the protein-coding sequence (in gray) of the *Htt* gene, thus affecting the sequence and the 3D shape of the protein. In contrast, the regulatory sequences of the *Htt* gene (Promoter (designated with P), Terminator (designated with T) and the UnTranslated Regions at the beginning of the gene (5'UTR) and the end of the gene (3'UTR)) are unaffected. The severity of the disease is directly correlated to the number of CAG repeats as depicted by the inverted triangle.

families (Longshore and Tarleton, 1996). There have also been differences in severity and onset depending on the parent from which the mutation was inherited. In HD, if the abnormal allele is inherited from the father, there is an increased risk of earlier onset, presumably from an expansion that had grown from the father's generation to his child's. It was found that the alleles involved are more unstable when transmitted through sperm as opposed to the egg, leading to changes in the number of repeats expanding through sperm transmission (Longshore and Tarleton, 1996).

There are several ideas of how glutamine expansion within the huntingtin protein causes HD. It is important to note, however, that neither mechanism is not fully understood. It has been proposed that the glutamine repeats cause a change in function of the Htt gene. For example, the mutant Htt protein binds to an intracellular protein HAP-1, to which the natural form of the protein would normally not bind. The HAP-1 protein has been noted to be involved in trafficking of vesicles and organelles, including mitochondria whose major function is energy production in the cell. The Htt protein, though found all throughout the body, is present at high concentrations in the brain. It is also thought that inefficient energy production

contributes to HD pathology since HD-related neurodegeneration could be reversed experimentally by treating with drugs interfering with energy-associated signaling (Varma, H., et al., 2007). Another idea for the effect the trinucleotide repeats have on Htt and subsequently the brain has to do with the way that the expansion affects protein folding, the process by which proteins adopt their 3D shape. When the Htt protein is misfolded, it gains a toxic function. The misfolded Htt protein has the potential to interact with nearby proteins that will cause selective degeneration and neuronal death (Cummings and Zoghbi, 2000). A second effect of the misfolded proteins is the potential for the formation of aggregates. The proteins attach to one another and form groups that can affect the nuclear architecture and the functions that it carries out (Cummings and Zoghbi, 2000).

Lastly, mutated Htt protein can alter transcription by interfering with histone modifications. Histone molecules are part of the efficient DNA packing in the cell. Strands of DNA wrap themselves around histone to become more compact when genes are not used (transcribed). When DNA needs to be transcribed, the strength of the hold the histone has on the DNA weakens and the DNA can be more easily accessed to allow the genetic information to be used to make proteins. When the mutant Htt interacts with histones, chromatin becomes more compact, therefore inhibiting necessary transcription (Sadri-Vakili and Cha, 2006). Changes in transcription regulation in the brain clearly have profound effects on brain function and degeneration, and they are found frequently in those individuals diagnosed with HD (Sadri-Vakili and Cha, 2006). It is quite possible that mutant Htt protein inflicts brain damage via multiple mechanisms since the Htt is a large multifunctional protein in its natural form. Logically, when it does not work properly, multiple physiological circuits of the cell could be disrupted.
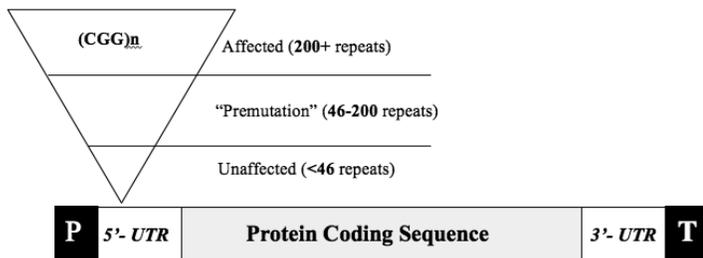


Figure 2. *FMR1* gene structure and relationship to Fragile X pathology. CGG repeat expansions related to **Fragile X syndrome** are located within the regulatory 5'UTR sequence (in white, on the left) of the *FMR1* gene, and do not affect the sequence and the 3D shape of the encoded protein. All regulatory sequences are labeled as described in the legend of Figure 1. The severity of the disease is directly correlated to the number of CGG repeats as depicted by the inverted triangle.

## Molecular Mechanisms and Pathology of Fragile X Syndrome

Just as the mechanism for HD is speculated upon and not fully understood, mechanisms for a fellow trinucleotide repeat disorder, Fragile X Syndrome (FXS), are also the subject of intensive research. Since FXS was one of the first trinucleotide disorders to be recognized and studied, the mechanisms are much better understood. The FXS relevant repeat is located in the 5' untranslated region (Figure 2) of the FMR1 (Fragile X Mental Retardation) gene on the X chromosome (Longshore and Tarleton, 1996), i.e. no changes in the protein itself are taking place, rather its amount is being changed. In contrast, the HD repeat is located in a translated region on chromosome 4, i.e. the defect alters the encoded protein directly introducing more glutamines. FXS is an X-linked dominant disorder that is one of the most commonly inherited forms of mental retardation, second only to Down syndrome. Since males have only one X-chromosome and females have two X-chromosomes, the disease manifests itself differently between males and females. Males usually display a significant intellectual impairment, but there is a range of severity. There are several distinctive physical features for FXS that include a long face, large ears, flat feet, and hyperflexible joints (National Fragile X Foundation, 2012). In terms of behavioral symptoms, increased aggression, ADHD-like attention issues, and social anxiety are all common in males affected by FXS. Males pass the disease only to their daughters since they are the ones inheriting an X chromosome from their father, whereas male offspring inherit their father's Y chromosome. In contrast, females can pass the disease to both sons and daughters and display less severe phenotypes as compared to their male counterparts. The differences in severity of the phonotype are at least in part associated with the X chromosome inactivation phenomenon in females. Each female cell randomly inactivates one X chromosome. If the inactivated chromosome contains a defective copy of the FRMP gene, then the cell will appear normal because the working X chromosome contains a functional copy of the gene. As a result, female bodies contain a variable mix of normal and defective cells, leading to a broad range of severity of the symptoms. Only one third of affected females have a significant intellectual disability and the behavioral symptoms are the same, but less severe than those affected males where all cells in the body have a defective X chromosome (National Fragile X Foundation, 2012).

FXS affects 1/1,250 males and 1/2,500 females and, like an HD, there is a range of repeats (Figure 2) that can cause the disorder (Timchenko and Caskey, 1996). Individuals with stable repeats that do not display the phenotype of FXS have repeats ranging from 6-46, and unaffected carriers have 50-200 repeats. The unaffected carriers, who have the coding for the

premutation, can pass along the expansion to their children. Those individuals with over 200 repeats in the FMR1 gene will exhibit the described phenotype. Just like HD, FXS displays unusual inheritance patterns of the nucleotide expansion. In addition to genetic anticipation, FXS is characterized with a parent of origin phenomenon or Sherman Paradox (Timchenko and Caskey, 1996). When a male with premutation has children, he passes on the premutation with minimal if any change in the number of the repeats. If a female passes on premutation alleles, the section is very likely to expand and develop into a full mutation resulting in Fragile X syndrome. This phenomenon is attributed to differences in the processes of sperm and egg formation (Timchenko and Caskey, 1996). Some individuals can have the FXS mutation but show no symptoms, a phenomenon known as incomplete penetrance.

Similar to HD, FXS is associated with a particular gene; FMR1 coding for FMRP, or fragile X mental retardation protein. In healthy individuals, FMRP is expressed in many tissues in the body with the highest levels in the neurons of the brain, testes and ovaries (Timchenko and Caskey, 1996). FMRP is an RNA-binding protein that shuttles between the nucleus and the cytoplasm presumably regulating the transport of specific mRNAs. In addition, FMRP has been found to associate with ribosomes in the dendritic protrusions of neurons, which suggests potential role in the translation of mRNAs specific for this area of the cell. FMRP interacts with several proteins; however the biological significance of the interactions is not understood (Cummings and Zoghbi, 2000: National Institutes of Health: FMR1, 2012).

 The accumulation of trinucleotide repeats associated with FXS results in alteration of the expression of the FMR1 gene leading to limited or absent fragile X protein production in the cell proposed to lead to disrupted neuronal communication. Two mechanisms have been experimentally supported: 1) decreased efficiency of translation initiation, most likely caused by the repeats impeding the proper assembly of the ribosome on the FMR1 mRNA and 2) transcription silencing via methylation of the DNA region governing FMR1 transcription resulting in lack of FMR1 mRNA and thus lack of Fragile X protein. The second mechanism is observed only in patients with more than 200 repeats (Timchenko and Caskey, 1996).

Collectively, both mechanisms result in insufficient levels of FMRP necessary for proper functioning, thus neuronal abnormalities and loss are observed and mental retardation occurs (Cummings and Zoghbi, 2000; Longshore and Tarleton, 1996).

**Do all trinucleotide diseases have a common mechanism?**
An interesting comparison has been made between HD and FXS in terms of the underlying mechanisms. There is a difference in the way in which the mutation affects the primary function of those pathways. Fragile X syndrome results from a reduced level of the FMRP or loss of function; whereas Huntington's disease results from changes in function of the associated protein (Cummings and Zoghbi, 2000; Longshore and Tarleton, 1996). In HD, the increase of the number of trinucleotide repeats causes the formation of aggregates and the acquisition of potentially toxic qualities that cause the neuronal death noted in the disorder. In FXS, the hypermethylation of the promoter region on the FMR1 gene causes a decrease in transcription, reducing the amount of FMRP which is vital for the proper development of neuronal tissue. It is the loss of function that causes the phenotype, not a gain in toxic function, as in the case of HD.

Despite their extensive differences, there have been attempts to link the trinucleotide repeat disorders together mechanistically. For example, Kaplan, Itzkovitz, and Shapiro (2007) proposed a universal mechanism for trinucleotide disorders specifically referring to the size of the repeat and how that affects the onset of the disorder. They explained that patients with large nucleotide repeats are born with all relevant cells having a copy of the mutation and as the individual grew older, the toxic effects of the repeat's misfolding or other behavior within the body grew progressively in numbers. There is a threshold for the amount of toxicity that is tolerated before phenotypic manifestations of the disorder begin to arise, and later in life with more and more developing due to these disorders, the phenotype begins to appear at these later points. Those with a greater number of repeats have a more severe display of the disorder and an earlier onset. According to their proposed mechanism, the trinucleotide repeat is more potent when it is longer and it therefore produces more drastic effects and quicker presentation. Though this mechanism does match the progression of HD, FXS is one trinucleotide disorder that does not manifest in this way. FXS presents itself from birth and does not progressively get worse. It very well may fit the mechanisms of other trinucleotide disorders, but one of the most well known of the group does not fit the proposed "all-encompassing" theory. The clinical observations could be very well explained with the underlying molecular events. Since Fragile X is caused by lack of functional FMRP, the increase of the number of repeats will not change the condition simply because all FMRP has been lost already.

As illustrated through the analysis of two of the most prevalent disorders, the trinucleotide repeats in both coding and non-coding regions of the genome do share several similarities, but

are very different mechanistically and phenotypically. They share the same basic type of molecular defect, but depending on the location of the trinucleotide repeat, the affected proteins can gain functionality in some cases, while in others can lose functionality, both of which can result in neuronal loss and subsequent symptoms. In addition, the natural functions of each affected protein are very different and their disruption results in widely varying abnormalities in cell physiology. Therefore, it seems unwise to try to offer a universal mechanism that would incorporate all aspects of the trinucleotide disorders and lead to development of common strategies for their management and treatment. Thus, it is highly unlikely that a universal mechanism would apply to all aspects of the various trinucleotide diseases solely based on the fact that they stem from the same type of mutation. Understanding of the mechanism and specifics of each individual disease remains critical for development of proper therapies.

## References

Chial, H. (2008). Huntington's disease: The discovery of the Huntington gene. *Nature Education* 1(1). <http://www.nature.com/scitable/topicpage/huntington-s-disease-the-discovery-of-the-851>. Date accessed February 18, 2013.

Cummings, C. J. and Zoghbi, H. Y. (2000). Fourteen and counting: Unraveling trinucleotide repeat diseases. *Human Molecular Genetics, 9*(6). 909-916.

"Htt."(2012). Genetics Home Reference. National Institutes of Health. <http://ghr.nlm.nih.gov/gene/HTT>. Date accessed March 25, 2012.

"FMR1." (2012). Genetics Home Reference. National Institutes of Health. <http://ghr.nlm.nih.gov/gene/FMR1> . Date accessed, March 25, 2012.

"Fragile X Syndrome."(2012). *National Fragile X Foundation.* <http://www.fragilex.org/fragile-x-associated-disorders/fragile-x-syndrome/>. Date accessed March 25, 2012.

Kaplan, S., Itzkovitz. S., and Shapiro, E. (2007). A universal mechanism ties genotype to phenotype in trinucleotide diseases. *PLoS Computational Biology, 3*(11). 2291-2298.

Longshore. J. W. and Tarleton. J. (1996). Dynamic mutations in human genes: A review of trinucleotide repeat diseases. *J. Genet., 75*(2). 193-217.

Petruska, J., Martenstine, M. J., and Goodman, M. R. (1998). Analysis of strand slippage in DNA polymerase expansions of CAG/CTG triplet repeats associated with neurodegenerative disease. *The Journal of Biological Chemistry, 273*(9). 5204-5210.

Sadri-Vakili, G. and Cha. J. J. (2006). Mechanisms of disease: Histone modification in Huntington's disease. *Nature Clinical Practice Neurology, 2*(6). 330-338.

Timchenko, L. T. and Caskey, C. T. (1996). Trinucleotide repeat disorders in humans: Discussions of mechanisms and medical issues. *The FASEB Journal, 10.* 1589-1597.

Varma, H., Cheng, R., Voisine, C., Hart. A. C., Stockwell, B. R. (2007). Inhibitors of metabolism rescue cell death in Huntington's disease models. PNAS, 104(36). 14525-14530.