Toward understanding the role of protein context in the polyglutamine disease, SCA3

Ginny Marie Harris
University of Iowa

Follow this and additional works at: https://ir.uiowa.edu/etd

© 2011 Ginny Marie Harris

This dissertation is available at Iowa Research Online: https://ir.uiowa.edu/etd/978

Recommended Citation
https://doi.org/10.17077/etd.wkmgd51q
TOWARD UNDERSTANDING THE ROLE OF PROTEIN CONTEXT IN THE POLYGLUTAMINE DISEASE, SCA3

by

Ginny Marie Harris

An Abstract

Of a thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Molecular and Cellular Biology in the Graduate College of The University of Iowa

May 2011

Thesis Supervisors:  Professor Henry L. Paulson
               Assistant Professor Pedro Gonzalez-Alegre
ABSTRACT

The polyglutamine diseases are a clinically heterogeneous group of inherited neurodegenerative disorders caused by expansion of polyglutamine-encoding (CAG)$_n$ trinucleotide repeats within the disease genes. It is increasingly clear that the amino acid sequences flanking the polyglutamine expansion in each disease protein, i.e. the specific protein context, contribute to selective neuronal toxicity by influencing the behavior of the disease protein within selectively vulnerable neuronal populations. In the studies described here, I explore the role that protein context plays in the polyglutamine disease, Spinocerebellar ataxia type 3 (SCA3). Toward this end, I utilize biochemical, cell-based, and animal models to gain a broader understanding of the SCA3 disease protein, ataxin-3, and generate tools for further exploration of the molecular properties of ataxin-3 that modulate its toxicity during disease.

In Chapter 1, I provide an overview of the recognized polyglutamine diseases, emphasizing the elements of protein context that are distinct among the polyglutamine disease proteins and may contribute to the neuropathological and clinical heterogeneity within this family of diseases.

Alternative splicing of the polyglutamine disease gene products adds an additional level of complexity to the tissue-specific protein context of expanded polyglutamine, yet this phenomenon has been underinvestigated. In Chapter 2, I examine the significance of ataxin-3 splice variation. Several minor 5’ variants and both known 3’ splice variants of ataxin-3, a deubiquitinating enzyme, are expressed at the mRNA level in brain. At the protein level, however, the C-terminal splice isoform with 3 ubiquitin interacting motifs (3UIM ataxin-3) is the predominant isoform in brain, independent of age or (CAG)$_n$ expansion. Although both C-terminal ataxin-3 splice isoforms display similar in vitro deubiquitinating activity, 2UIM ataxin-3 is more prone to aggregate and is more rapidly degraded by the proteasome. These observations demonstrate how alternative splicing of
sequences distinct from the polyglutamine-encoding \((CAG)_n\) repeat can alter disease-related components of protein context.

Knock-in models of polyglutamine diseases utilize pathogenic \((CAG)_n\) expansions within the endogenous genomic, transcript, and protein context to recreate key features of individual polyglutamine diseases. In chapter 3, I describe the creation of the first knock-in mouse model of SCA3. Hemizygous knock-in mice transmit the knock-in allele in Mendelian ratios and broadly express both the expanded \(\text{Atxn3(Q}_3\text{KQ}_82)\) protein and the wildtype murine \(\text{Atxn3(Q}_6)\) protein. In this chapter, I also compare the gene targeting efficiencies and rates of chromosomal instability of a novel C57BL/6J ES cell line (UMB6JD7) and two well established ES cell lines (W4 and Bruce4.G9). Of these, Bruce4.G9 ES cells proved superior based on lower rates of aneuploidy and the production of germline transmitting chimeras.

Finally, in Chapter 4 I discuss questions and concepts raised during the course of these studies, and suggest avenues of future research aimed at broadening our understanding of ataxin-3 physiology and of protein context-dependent elements in polyglutamine disease pathogenesis.

Abstract Approved: ____________________________________

Thesis Supervisor

____________________________________

Title and Department

____________________________________

Date

____________________________________

Thesis Supervisor

____________________________________

Title and Department

____________________________________

Date
TOWARD UNDERSTANDING THE ROLE OF PROTEIN CONTEXT IN THE
POLYGLUTAMINE DISEASE, SCA3

by
Ginny Marie Harris

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Molecular and Cellular Biology
in the Graduate College of
The University of Iowa

May 2011

Thesis Supervisors: Professor Henry L. Paulson
Assistant Professor Pedro Gonzalez-Alegre
CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Ginny Marie Harris

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Molecular and Cellular Biology at the May 2011 graduation.

Thesis Committee:

__________________________
Henry L. Paulson, Thesis Supervisor

__________________________
Pedro Gonzalez-Alegre, Thesis Supervisor

__________________________
Michael Welsh

__________________________
Stefan Strack

__________________________
Tomomi Kuwana

__________________________
Steven A. Moore
To Marci Powell (1981 – 2003),

who first introduced me to the Spinocerebellar ataxias,
whose struggle kept me motivated when nothing in lab was working,
and whose courage, joy, and guilelessness will remain an inspiration to me always.
ACKNOWLEDGMENTS

First of all, I would like to thank both of my wonderful mentors, Hank Paulson and Pedro Gonzalez-Alegre. Hank’s enthusiasm for his work is infectious, and I have enjoyed the many occasions upon which we have geeked out about science over the years. Pedro has always had an open door, and I have grown immensely from his pragmatic approaches to experimental design. Both have pushed me to become a better writer, a better scientist, and a better person. I also thank my thesis committee: Michael Welsh, Steven Moore, Stefan Strack, and Tomomi Kuwana. Their intellectual curiosity and fresh ideas have kept me excited about my work, their down to earth suggestions have kept me focused, and their encouragement has supported me every step of the way.

I owe an enormous debt of gratitude to all of my current and former lab mates. In particular, I would like to acknowledge Aislinn Williams, Sokol Todi, Edgar Rodriguez, Kevin Glenn, Sue Travis, Brett Winborn, Kara Gordon, and Janine Martin for their friendship, patience, and guidance over the years. I would also like to acknowledge Katerina Dodelzon for her preliminary work on the splicing project, Lijie Gong and Jim Dell’Orco for their assistance in long-distance establishment of the knock-in line, and Nicole Bode, Tina Knudson, Hsiang Wen, and Cynthia Gouvion for the assistance, organizational support, and training that they have provided. They have all been great co-workers and great friends, and every one of them has been a role model in some way.

I thank Baoli Yang for his technical expertise and for providing me with B6(Cg)-Tyr<2J>/J ES cell DNA, Thom Saunders, Elizabeth Hughes, Corey Ziebell, and Jessica Wiget at the University of Michigan Transgenic Animal Core for their work on ES cell culture and chimera generation, Veronica Colomer for providing the Q71B mice, and Olaf Riess, Mark Pook, and Ilya Bezprozvanny for providing the MJD15.4 and MJD84.2 mice, and Ullrich Wüllner for providing the Atxn3(-/-) knockout mice.
I thank the current and former chairs of the MSTP program, Pam Geyer, Steve Lentz, and Mike Knudsen, and of the Interdisciplinary Program in Molecular and Cellular Biology, Jackie Bickenbach and Minnetta Gardinier. They have been great resources for support and advice. I would also like to thank the program secretaries, including Elizabeth Snyder and Kathy Dautremont, without whom the trains would not run on time, and the tireless program administrators, Leslie Harrington and Paulette Villhauer, without whom the trains would not run at all.

Finally, I would like to thank my family and friends, including my sisters, Emily and Lauri, my brother-in-law John, my mother, Mary Helen, my father, Frederick, my nieces and nephews, Alexander, Colin, Kiera, and Rory, my partner David, my many medical school and MSTP stalwarts, as well as friends old and new who have kept me motivated, inspired, amused, and in possession of a great deal of perspective. Without their love and support I would not be where I am today.
TABLE OF CONTENTS

LIST OF TABLES ...........................................................................................................vii

LIST OF FIGURES ..........................................................................................................ix

LIST OF ABBREVIATIONS ...............................................................................................x

CHAPTER 1 PROTEIN CONTEXT IN POLYGLUTAMINE DISEASES ........................................1

Abstract .........................................................................................................................1
Polyglutamine Diseases are Trinucleotide Expansion Disorders ........................................1
A Common Molecular Mechanism of Pathogenesis in Polyglutamine Diseases? ......................4
Clinical Manifestations of the Polyglutamine Diseases ........................................................6
Influence of Protein Context on Polyglutamine Toxicity .......................................................9
Polyglutamine Disease Proteins ........................................................................................11
Protein Context as a Modifier of Protein Misfolding and Aggregation ...............................14
Polyglutamine Expansion-Dependent Alteration of Endogenous Function Depends on Protein Context .............................................................16
Protein Context as a Modifier of Subcellular Localization and Expression Patterns ...........17
Alternative Splicing as an Element of Protein Context ........................................................18
The Importance of Protein Context in Model Systems ........................................................21
Therapeutic Implications of Understanding the Role of Protein Context in Polyglutamine Diseases ..............................................................................................................23
Focus of this thesis: Protein Context in SCA3 ..................................................................25
Rationale and Aims of the Thesis .......................................................................................26

CHAPTER 2 ALTERNATIVE SPLICING OF ATAXIN-3 AS AN ELEMENT OF PROTEIN CONTEXT IN SCA3 ....................................................38

Abstract .........................................................................................................................38
Introduction .....................................................................................................................38
Materials and Methods ..................................................................................................40
Animal lines ....................................................................................................................40
Constructs and Primers .................................................................................................41
Cell culture and transfection .........................................................................................42
Collection of animal tissues ...........................................................................................43
Preparation of cDNA from animal tissues and pooled human RNAs ..............................43
Sequence-specific PCR-amplification of cDNA ...............................................................43
Preparation of protein from murine tissues and cultured cells ........................................44
2D-Western Blot Analysis ...............................................................................................45
In vitro deubiquitination assays .......................................................................................45
In vitro ubiquitin-AMC assays .......................................................................................46
Immunostaining of cultured cells ..................................................................................46
Cycloheximide treatment to analyze protein stability ......................................................48
Open Field Analysis .....................................................................................................48
Gait Analysis ...................................................................................................................49
Results.............................................................................................................49
Ataxin-3 is alternatively spliced in ATXN3 YAC transgenic and human brain .................................................................49
3UIM ataxin-3 is the predominant protein isoform in murine and human brain tissue .................................................................51
2UIM and 3UIM ataxin-3 display similar in vitro DUB activity against defined ubiquitin chains ...................................................53
2UIM ataxin-3 has a higher propensity to aggregate than 3UIM ataxin-3 ...........................................................................55
2UIM ataxin-3 is a less stable protein than 3UIM ataxin-3 and is subject to rapid proteasomal degradation ................................57
Discussion.......................................................................................................59

CHAPTER 3  GENERATION OF A (CAG)\text{n}-EXPANSION KNOCK-IN MURINE MODEL OF SCA3 ..................................................82

Abstract ...........................................................................................................82
Introduction .....................................................................................................83
Materials and Methods ...................................................................................87
Generation of the targeting vector ................................................................87
ES cell lines .............................................................................................89
Gene targeting by homologous recombination ............................................90
Screening ES cells for Homologous Recombination and Euploidy ............90
Chimera Generation and Screening for Germline Transmission ...............92
Excision of the Positive Selection Cassette .............................................93
Confirmation of Knock-in Protein Expression ..........................................94
Results.............................................................................................................94
Targeting vector construction: complications of (CAG)\text{n} repeat expansion instability .........................................................94
C57BL/6-derived ES cell lines demonstrated better gene-targeting efficiency, but significant chromosomal instability ..........97
Low germline transmission rates and premature mortality were observed in high percentage chimeras .........................................98
Comparison of germline transmission rates in chimeras and F1 progeny suggests nonclonality of ES cell colony 3C8 ..........99
F1 knock-in x FLPe transgenic matings resulted in mosaic excision of the positive selection cassette ...........................................101
Expanded Q\text{3}KQ\text{82} Ataxin-3 Protein is expressed in F2 mosaics and F3 founders ...............................................................102
Discussion.....................................................................................................102

CHAPTER 4  CONCLUSIONS AND FUTURE DIRECTIONS ................................132

Functional Significance of Tissue-specific Ataxin-3 Splice Variation ........132
Determination of UIM3 function .................................................................134
Cell-specific expression of minor ataxin-3 splice variants in affected and adjacent brain regions ..................................................138
Splice Isoforms and Protein Context in Other Polyglutamine Diseases ......140
Phenotypic Characterization of Atxn3(Q\text{3}KQ\text{82}) Knock-in Mice ..............142
Conclusions ..................................................................................................143

REFERENCES ........................................................................................................148
## LIST OF TABLES

Table 1. Clinical heterogeneity among the polyglutamine disorders.................................28
Table 2. Polyglutamine diseases and their gene products....................................................29
Table 3. Alternative splicing of polyglutamine disease gene products...............................30
Table 4. Predicted molecular weights and isoelectric points of ataxin-3 isoforms..............64
Table 5. Mass and exclusion criteria for open field analysis.............................................65
Table 6. Existing knock-in models of polyglutamine diseases.........................................108
Table 7. PCR-based methods for expansion of the (CAG)$_n$..............................................111
Table 8. ES cell lines utilized in gene targeting.................................................................113
Table 9. PCR conditions for ES and knock-in line screening...........................................114
Table 10. Assessment of euploidy and aneuploidy in homologously recombined ES cell populations. .................................................................................................116
Table 11. Summary of chimera production and breeding..................................................117
LIST OF FIGURES

Figure 1. The nucleotide repeat expansion disorders. .................................................32
Figure 2. Mechanisms of polyglutamine disease toxicity. .............................................33
Figure 3. Multi-domain aggregation mechanism for polyglutamine disease proteins with or without expansion of the polyglutamine repeat. ........................34
Figure 4. The genomic and protein context of Ataxin-3, which causes SCA3. ...............35
Figure 5. 3UIM ataxin-3 is well conserved among mammals and amphibians. ...........36
Figure 6. Minor Atxn3 5’ splice variants are present in murine brain tissue.................66
Figure 7. Both 3’ splice variants are expressed in ATXN3 YAC transgenic and human brain tissue. ........................................................................67
Figure 8. Distinct tissue-specific ataxin-3 isoforms suggest Atxn3 transcripts may be differentially spliced in mouse tissues. ..............................................................68
Figure 9. 3UIM ataxin-3 is the predominant transgenic isoform in YAC transgenic brain tissue. .........................................................................................69
Figure 10. 3UIM ataxin-3 is the major ataxin-3 isoform in human brain. .......................70
Figure 11. Unmodified 2UIM ataxin-3 is not detectable in ataxin-3 YAC transgenic brain. ..................................................................................................................71
Figure 12. 2UIM and 3UIM ataxin-3 display similar DUB activity against defined ubiquitin chains in vitro. .................................................................73
Figure 13. Recombinant 2UIM and 3UIM GST-ATXN3(Q22) cleave Ub-AMC at a similar rate. .................................................................74
Figure 14. 2UIM ataxin-3 has a higher propensity to aggregate than 3UIM ataxin-3 in mammalian cells. .................................................................75
Figure 15. 2UIM ataxin-3 has a higher propensity to aggregate than 3UIM ataxin-3 in brain tissue. .................................................................76
Figure 16. 18 – 23 month old MJD84.2 and Q71B hemizygotes show a mild ataxic phenotype.................................................................77
Figure 17. 2UIM ataxin-3 is much less stable than 3UIM ataxin-3. ..............................79
Figure 18. 2UIM ataxin-3 is subject to rapid proteasomal degradation, while 3UIM ataxin-3 is subject to autophagy..................................................80
Figure 19. Model for the differential aggregation properties and processing of 2UIM and 3UIM ataxin-3.................................................................81
Figure 20. Comparison of 129- and C57BL/6-derived ES cells for gene targeting by homologous recombination. .................................................................118

Figure 21. Overview of targeting vector construction. ...........................................119

Figure 22. Expansion of the trinucleotide repeat using the megaprimer method. .....120

Figure 23. Laccone method for expansion of the Q3KQ82-encoding repeat sequence. ........................................................................................................121

Figure 24. Targeting exon 10 of the murine Atxn3 gene for homologous recombination to insert a trinucleotide expansion. ........................................123

Figure 25. Five ES cell colonies showed evidence of targeted gene insertion. ..........124

Figure 26. Chimeras exhibited a very low rate of germline transmission. ..................125

Figure 27. Germline transmission of the Atxn3(Q3KQ82) knock-in allele with an intact positive selection cassette followed Mendelian ratios in F1 animals. ........................................................................................................126

Figure 28. Mosaic excision of the positive selection cassette in FLPe(tg/-) F2 Atxn3(Q3KQ82) knock-in animals.................................................................127

Figure 29. Atxn3(Q3KQ82) knock-in animals express expanded ataxin-3 protein. .......129

Figure 30. Summary of gene targeting and mating efficiencies.................................130

Figure 31. Mutation of UIM3 decreases tubulin levels..............................................145

Figure 32. Model for the role of UIM3 in aggresome formation and microtubule stability ...............................................................................................146
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2BP</td>
<td>ataxin-2 binding protein</td>
</tr>
<tr>
<td>AD</td>
<td>Alzheimer’s disease</td>
</tr>
<tr>
<td>AR</td>
<td>androgen receptor</td>
</tr>
<tr>
<td>ATN1</td>
<td>atrophin-1</td>
</tr>
<tr>
<td>ATXN1</td>
<td>ataxin-1</td>
</tr>
<tr>
<td>ATXN2</td>
<td>ataxin-2</td>
</tr>
<tr>
<td>ATXN3</td>
<td>ataxin-3</td>
</tr>
<tr>
<td>ATXN7</td>
<td>ataxin-7</td>
</tr>
<tr>
<td>BAC</td>
<td>bacterial artificial chromosome</td>
</tr>
<tr>
<td>CaV2.1</td>
<td>P/Q-type voltage-gated calcium channel, subunit α1</td>
</tr>
<tr>
<td>CACNA1A</td>
<td>gene encoding the P/Q-type voltage-gated calcium channel, subunit α1</td>
</tr>
<tr>
<td>cDNA</td>
<td>complimentary deoxyribonucleic acid</td>
</tr>
<tr>
<td>E4B</td>
<td>ubiquitination factor E4B protein</td>
</tr>
<tr>
<td>HD</td>
<td>Huntington disease</td>
</tr>
<tr>
<td>HEAT repeat</td>
<td>(Huntingtin, Ef3, regulatory subunit of PP2A, and mTor1) repeat</td>
</tr>
<tr>
<td>HHR23</td>
<td>human UV excision repair protein RAD23 homolog B</td>
</tr>
<tr>
<td>HTT</td>
<td>huntingtin protein</td>
</tr>
<tr>
<td>Lsm</td>
<td>(like Sm) motif – an RNA binding element</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>NI</td>
<td>nuclear inclusions</td>
</tr>
<tr>
<td>NMD</td>
<td>nonsense-mediated decay</td>
</tr>
<tr>
<td>PAM2</td>
<td>a motif that associates with PolyA mRNA Binding Protein</td>
</tr>
<tr>
<td>p300/CBP</td>
<td>CREB-binding protein – p300 histone acetyltransferase complex</td>
</tr>
<tr>
<td>PCAF</td>
<td>p300/CBP associated factor (a.k.a. lysine acetyltransferase 2B)</td>
</tr>
<tr>
<td>polyQ</td>
<td>polyglutamine</td>
</tr>
<tr>
<td>SBMA</td>
<td>spinobulbar muscular atrophy</td>
</tr>
<tr>
<td>SCA</td>
<td>spinocerebellar ataxia</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td>STAGA</td>
<td>SPT3/TAF GCN5 complex</td>
</tr>
<tr>
<td>TBP</td>
<td>TATA-box binding protein</td>
</tr>
<tr>
<td>TFTC</td>
<td>TATA-binding protein-free TBP-associated factor-containing complex</td>
</tr>
<tr>
<td>UPP</td>
<td>ubiquitin proteasome pathway</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
<tr>
<td>VCP</td>
<td>valsolin containing protein</td>
</tr>
<tr>
<td>YAC</td>
<td>yeast artificial chromosome</td>
</tr>
</tbody>
</table>
CHAPTER 1
PROTEIN CONTEXT IN POLYGLUTAMINE DISEASES

Abstract

At least nine neurodegenerative disorders are caused by the expansion of polymorphic polyglutamine-encoding (CAG)\textsubscript{n} trinucleotide repeats within specific genes. In all polyglutamine diseases, this shared molecular insult triggers a pathogenic cascade that ultimately leads to a restricted pattern of neurodegeneration, with increased expansion length strongly correlated to earlier disease onset and increased disease severity. Though all polyglutamine disease proteins are widely expressed, the patterns of neuronal loss and clinical features of these diseases vary significantly. The lack of a common “polyglutamine syndrome” suggests that the amino acid sequences flanking the polyglutamine expansion within each disease protein, i.e. the protein context of the polyglutamine repeat, must contribute to selective neuronal toxicity by influencing factors such as protein misfolding and oligomerization, subcellular localization, specific protein-protein interactions, and endogenous functions of the individual disease proteins. Alternative splicing of the disease gene products adds an additional level of complexity to the tissue-specific protein context of expanded polyglutamine proteins. This chapter provides an overview of the recognized polyglutamine diseases, with an emphasis on elements of protein context that are distinct among polyglutamine disease proteins and that may contribute to the neuropathological and clinical heterogeneity within this devastating family of diseases.

Polyglutamine Diseases are Trinucleotide Expansion Disorders

The polyglutamine diseases belong to a larger family of dynamic repeat expansion diseases (Figure 1). Clinically, this family encompasses a broad spectrum of diseases, ranging from mental retardation syndromes and muscular dystrophies to
movement disorders including ataxias and chorea. All are caused by the expansion of a polymorphic oligonucleotide repeat, but the pathophysiological manifestations of disease vary widely within this family of neuromuscular disorders (Cummings and Zoghbi, 2000, Gatchel and Zoghbi, 2005, Orr and Zoghbi, 2007). The trinucleotide repeat disorders can be subdivided broadly into four groups, based on what is known about their pathological mechanisms. Although individually many of the nucleotide repeat disorders are relatively rare, together they include the most common inherited ataxia, the most common dominantly inherited ataxia, the most common adult-onset muscular dystrophy, and one of the most common forms of inherited mental retardation (Gatchel and Zoghbi, 2005, Orr and Zoghbi, 2007). Collectively they impose a significant burden of disease worldwide.

Friedreich’s ataxia, the Fragile X mental retardation syndrome, Unverricht-Lundberg Disease (Joensuu et al., 2008), and possibly the Fragile Site FRA12A mental retardation syndrome (Winnepenninckx et al., 2007) are associated with pathogenic processes at the DNA level, such as epigenetic transcriptional silencing and prevention of productive transcriptional elongation (Kumari and Usdin, 2009). These diseases are caused by very large noncoding repeat expansions (> 200 trinucleotide repeats or > 30 dodecanucleotide repeats), and are considered to be “loss of function” disorders due to disrupted expression of the disease protein. The pathogenesis of the myotonic dystrophies, Fragile X-Associated Tremor Ataxia Syndrome, and Spinocerebellar ataxia type 8 (SCA8) is thought to occur at the RNA level. These diseases are considered to be RNA-mediated “gain of function” disorders because when expanded, these tri- or tetra-ribonucleotide repeats induce aberrant RNA-protein interactions, sequestering key proteins into ribonucleoprotein foci/inclusions and dysregulating critical cellular functions (O’Rourke and Swanson, 2009). At least three additional nucleotide repeat diseases have unknown pathological mechanisms, but are thought to have a nucleotide-mediated pathophysiology, because their nucleotide repeat expansions are located in
either constitutively noncoding (SCA10, SCA12) or alternatively spliced, variably coding sequences (HDL2) (Orr and Zoghbi, 2007).

Another subfamily of the nucleotide repeat disorders is made up of the polyglutamine diseases, the polyalanine diseases, and possibly a handful of diseases where very short (≤ 5) repeats vary by one or two aspartate- or glutamate-encoding codons (Amiel et al., 2004, Albrecht and Mundlos, 2005). These diseases are unique among the nucleotide repeat disorders, because the more modestly sized trinucleotide repeat expansions are located within translated exons of each disease gene. The polyglutamine and polyalanine diseases (Figure 1) occur with trinucleotide repeat expansions, and are associated with cytoplasmic or nucleoplasmic misfolded protein inclusions. Conversely, the diseases involving aspartate- or glutamate-encoding repeats occur with 1 – 2 codon expansions or contractions, and are associated with endoplasmic reticulum or nuclear envelope inclusions (Amiel et al., 2004); as such, they may not function by the same pathophysiological mechanism, and will not be discussed further here.

There are also a number of distinguishing features between the polyglutamine diseases and the polyalanine diseases. The polyglutamine diseases, including Spinobulbar muscular atrophy (SBMA), Huntington disease (HD), Dentatorubropallidoluysian atrophy (DRPLA), and the Spinocerebellar ataxias (SCAs) SCA1, SCA2, SCA3, SCA6, SCA7, and SCA17, are characterized by progressive degeneration of specific neuronal populations (Gatchel and Zoghbi, 2005). The polyalanine diseases, which include Sympolypdactyly type II, Cleidocranial dysplasia, Oculopharyngeal muscular dystrophy, Blepharophimosis-ptosis-epicanthus inversus syndrome, Congenital central hypoventilation syndrome, Hand-foot-genital syndrome, some X-linked mental retardation syndromes, X-linked hypopituitarism, and Holoprosencephaly, are all congenital malformation syndromes, with or without a degree of mental retardation (Amiel et al., 2004, Messaed and Rouleau, 2009). While
polyglutamine disease genes contain highly variable, generally perfect \((\text{CAG})_n\) repeat length polymorphisms even among healthy individuals, the imperfect \((\text{GCN})_n\) repeats seen in polyalanine diseases are generally shorter and have less variable repeat lengths among wildtype alleles (Amiel et al., 2004). Moreover, smaller repeat expansions are sufficient to cause disease in polyalanine diseases. Yet a toxic gain of protein function disease mechanism has been proposed in both, and it has even been suggested that frameshifts leading to aberrant polyalanine production may be involved in at least one polyglutamine disease (Gaspar et al., 2000).

Expression of expanded polyglutamine proteins leads to numerous cellular abnormalities (Figure 2), mediated at the protein level by a toxic gain in function conferred, at least in part, by misfolding of the repeat domain (Gatchel and Zoghbi, 2005). Yet, protein misfolding may not be the whole story in all polyglutamine diseases. Notably, the SCA8 locus is bidirectionally transcribed to encode a nearly pure polyglutamine stretch in addition to a noncoding \((\text{CTG})_n\) transcript. SCA8 is therefore considered to exhibit elements of both RNA- and protein-mediated gain of function families (Merienne and Trottier, 2009). Similarly, an element of additive RNA toxicity was demonstrated in a \textit{Drosophila} model of SCA3, a more prototypical polyglutamine disease, when a truncated transcript with a very large CAG expansion was expressed in the retina (Li et al., 2008). These examples illustrate the importance of understanding the molecular biology of nucleotide repeat disease loci across a broad spectrum: from genomic context through posttranscriptional processing to the behavior of the mature gene products within the cell.

A Common Molecular Mechanism of Pathogenesis in Polyglutamine Diseases?

At a cellular level, polyglutamine diseases share several pathological hallmarks, suggesting commonalities in the mechanisms of neurotoxicity. Perhaps, the best studied
of these is the behavior of the polyglutamine repeat itself. In the absence of any adjacent protein context, α-helical polyglutamine monomers are unstable and undergo a conformational change to a β-sheet rich conformation, which can self associate to form fibrillar oligomers as well as SDS-insoluble fibrillar aggregates. Formation of these amyloid-like fibrils is polyglutamine length- and concentration-dependent, with fibrillization kinetics consistent with a seeding and elongation model (Chen et al., 2002). Polyglutamine proteins that have been analyzed for their ability to do so (prior to the formation of mature SDS-insoluble high molecular weight species), to date, have also demonstrated a polyglutamine-independent tendency to self-associate, resulting in a two-stage aggregation process in the presence of an intact polyglutamine domain. This led to the proposal of a “multi-domain aggregation mechanism,” whereby conformational changes and early oligomerization are mediated by the protein context of the polyglutamine-flanking sequence, but the formation of SDS-insoluble fibrils is polyglutamine-dependent (Figure 3). Further, the observation that increasing the length of the polyglutamine repeat increases the kinetics of this first step suggests that polyglutamine expansion also has the ability to destabilize the native conformation of the flanking sequences (Saunders and Bottomley, 2009). These biochemical observations are consistent with the behavior of naked polyglutamine repeats and polyglutamine disease proteins in cells and tissues.

Yet within the cellular environment, polyglutamine toxicity involves dysfunction in many processes beyond the simple misfolding and aggregation of isolated proteins. As previously mentioned, expanded (CAG)\textsubscript{n} trinucleotide repeats may have some toxic features at the mRNA level, though most findings support a protein-based toxicity. At the protein level, the effects of expanded polyglutamine-induced conformational changes observed \textit{in vitro} become more complex within the intracellular milieu (Figure 2). When a polyglutamine protein has a non-expanded repeat length, it is able to maintain its native conformational state(s) and interact with its standard protein partners, leading to a non-
toxic, normally functioning protein. By inducing conformational changes, expanded polyglutamine repeats can lead to alterations in protein-protein interactions, disruption of endogenous polyglutamine disease protein function, proteolytic cleavage, sequestration of aggregated proteins into ubiquitinated inclusion bodies, and perturbations in the cellular mechanisms of protein quality control including the ubiquitin proteasome system, autophagy, and chaperone-mediated protein refolding. Polyglutamine toxicity is also associated with alterations in cellular processes such as transcriptional regulation, RNA metabolism, maintenance of appropriate cytoskeletal architecture, protein transport, mitochondrial function, and ion homeostasis (Cummings and Zoghbi, 2000, Gatchel and Zoghbi, 2005, Orr and Zoghbi, 2007, Williams and Paulson, 2008). Perplexingly, although the various polyglutamine disease proteins are widely expressed, and these pathological mechanisms can be demonstrated in a variety of cellular, yeast, invertebrate, and animal model systems, the patterns of neurodegeneration are restricted to a fairly limited subset of neuronal populations within each disease (Zoghbi and Orr, 2000). It will therefore be important to assess which of the polyglutamine-induced dysfunctional processes mediate clinically relevant selective neurotoxicity, and which represent ancillary symptoms of intracellular stress.

Clinical Manifestations of the Polyglutamine Diseases

Given their shared molecular insult, the polyglutamine diseases predictably share some clinical features. For example, all of the polyglutamine diseases are age-related neurodegenerative movement disorders. With the exception of SBMA, which is X-linked, polyglutamine diseases exhibit an autosomal dominant pattern of inheritance. Nearly all show a tendency to undergo anticipation – in other words, instability of the disease-causing repeat expansion during germline transmission favors greater repeat length in subsequent generations, particularly when transmitted from a male parent. There is also a strong correlation between increased repeat expansion length, earlier
disease onset, and increased disease severity. Individuals with juvenile-onset forms of polyglutamine diseases manifest different symptoms than do adult-onset disease forms, and exhibit a greater degree of phenotypic overlap among the individual polyglutamine diseases in juvenile-onset cases. This observation that the longest polyglutamine repeat expansions lose some specificity of neuronal toxicity suggests that there may be a broader range of cells undergoing compensated, subclinical dysfunction within patients suffering from adult-onset disorders (Zoghbi and Orr, 2000).

Despite these commonalities, there is clearly no single “polyglutamine syndrome,” even among the longest polyglutamine expansions (Table 1). SBMA is associated with degeneration of lower motor neurons in the anterior horn and medullary bulbar nuclei, as well as in primary sensory neurons in the dorsal root ganglia (Sobue et al., 1989). It manifests in males with symptoms of tremor, muscle cramping and weakness (progressive from proximal lower extremities), dysarthria, dysphagia, and features of androgen insensitivity, including gynecomastia, hypogonadism, and loss of libido (Arbizu et al., 1983, Atsuta et al., 2006). Symptoms are mild in homozygous females, and include tremor, mild muscle cramping, and rare periorbital fasciculations (Schmidt et al., 2002). In contrast to the motor neuron phenotype of SBMA, adult-onset HD is associated with presymptomatic striatal and cortical volume loss and progressive neurodegeneration of striatal medium spiny neurons, the globus pallidus, and certain cortical structures (Rosas et al., 2008); juvenile-onset HD involves a more generalized pattern of neurodegeneration, including loss of cerebellar Purkinje cells (Zoghbi and Orr, 2000). In both forms of HD choreiform movements, cognitive dysfunction, and psychiatric disturbances are prominent, while juvenile HD can also present with associated symptoms including dystonia, bradykinesia, oropharyngeal symptoms, and seizures (Gonzalez-Alegre and Afifi, 2006, Rosas et al., 2008). While both HD and SBMA are caused by expansion of translated trinucleotide repeats, these diseases present
such vastly different clinical pictures that they can be unequivocally distinguished from each other in the absence of molecular diagnostics.

Even among the seven polyglutamine diseases in which ataxia is a prominent symptom, there is significant clinical heterogeneity. The polyglutamine ataxias are SCAs 1, 2, 3, 6, 7, 17 and Dentatorubropallidoluysian Atrophy (DRPLA). In addition to the cerebellar signs (ataxia, dysarthria, nystagmus, and tremor) common to this subfamily, all of the polyglutamine ataxias except SCA6, which is considered a pure cerebellar ataxia, exhibit additional features due to degeneration of extracerebellar populations of neurons (Zoghbi and Orr, 2000, Orr and Zoghbi, 2007, Soong and Paulson, 2007). The extracerebellar degeneration seen in the cerebral cortex, striatum, basal ganglia, and subthalamic nucleus (Luys body) in DRPLA contribute to the seizures, psychosis, dementia, and choreoathetosis often seen in this disease. In SCA1, the extracerebellar degeneration of bulbar and inferior olivary nuclei and demyelination of ascending and descending spinal tracts contribute to the slurred speech, ophthalmoplegia, peripheral neuropathy, and extrapyramidal signs seen clinically, whereas the defects in verbal memory and executive function have been proposed to be due to disruption of cerebrocerebellar communication. Symptoms common in SCA2, which involves extracerebellar degeneration in the brainstem and frontotemporal lobes, include hyporeflexia and slow saccades; polyneuropathy, motor neuropathies, and dementia are also sometimes seen. SCA3 differs somewhat from other polyglutamine SCAs in that cerebellar degeneration is largely limited to the deep cerebellar nuclei, with relative sparing of Purkinje cells and the inferior olivary nuclei. Extracerebellar degeneration in SCA3 involves regions of the basal ganglia, brainstem, and spinal cord, which clinically correlates with spasticity, pseudoexophthalmos with impaired ocular movements, facial and lingual fasciculations, and parkinsonism that can be helpful diagnostic clues suggesting SCA3. SCA7 also features prominent degeneration in brainstem nuclei, but is unique among SCAs for consistent degeneration in photoreceptors and the visual cortex,
which results in most SCA7 patients experiencing profound visual impairment due to macular degeneration. Psychosis and seizures are also sometimes seen in SCA7, as is a very high degree of anticipation, leading to an infantile-onset disease involving deficits in cardiac function. Cerebral cortical degeneration plays a particularly predominant role in the extracerebellar symptoms of SCA17, which include behavioral changes, psychosis, cognitive impairment, and seizures in some individuals. Whereas much of the variability between SCAs is likely to result from toxicity-modulating effects of the disease protein context, much of the heterogeneity within each disorder has been attributed to variation in the length of the repeat expansions.

There is also heterogeneity of the toxic repeat length threshold among polyglutamine diseases. Most polyglutamine disease proteins can tolerate repeats of up to about 35 glutamine residues without causing disease symptoms. This threshold, however, is not fixed among the polyglutamine diseases. The one disease with a much shorter repeat threshold is SCA6, which manifests with repeats as small as 22 residues in length. At the other end of the spectrum, TATA-box binding protein (TBP), the disease protein in SCA17, normally has a repeat up to 42 glutamines long, and the smallest disease causing repeat length documented in the SCA3 protein, ataxin-3, is 52 glutamines (Orr and Zoghbi, 2007, Williams and Paulson, 2008). Combined with the observations that flanking protein sequences can influence the conformational stability and aggregation propensity of polyglutamine repeats (Thakur et al., 2009), these differences in repeat length threshold among polyglutamine diseases strongly support an important modulating role for protein context in defining the clinical spectrum of disease.

**Influence of Protein Context on Polyglutamine Toxicity**

Protein context could alter polyglutamine toxicity in numerous ways. Reduced stability in the amino acid sequence flanking the polyglutamine stretch could lead to increased rates of protein misfolding and oligomerization, resulting in enhanced toxic
gain of function effects, as previously discussed. Protein context can also influence subcellular localization. Multiple polyglutamine proteins have functional nuclear import and nuclear export signals, suggesting that nucleocytoplasmic trafficking of these proteins is a normal component of their endogenous function. Because nuclear localization signals outside of the polyglutamine domain and other post-transcriptional events that increase nuclear localization have been shown to increase polyglutamine toxicity, the nuclear accumulation of polyglutamine disease proteins has been suggested to be an important component of polyglutamine-induced toxicity in most, if not all, polyglutamine diseases (Truant et al., 2006b). Protein context also determines cellular interaction partners, which may protect against protein misfolding (Saunders and Bottomley, 2009). Normal protein-protein interactions of polyglutamine disease proteins may become dysregulated when polyglutamine expansions are expressed, leading to loss of endogenous function and/or to a toxic gain of function. A compelling example of this has been observed with the SCA1 disease protein, ataxin-1, and Capicua (CIC) containing protein complexes (Lim et al., 2008). Similarly, more structured polyglutamine proteins, including the P/Q-type voltage-gated calcium channel (Cav2.1) that harbors the polyglutamine expansion in SCA6, may be more sensitive to deleterious disruptions of endogenous function (Matsuyama et al., 2004). The influence of protein context on polyglutamine neuronal toxicity through modulation of polyglutamine-dependent gain of function, alteration of endogenous protein function, and loss of endogenous protein function are not mutually exclusive. These aberrant cellular processes may be mediated by either the disease protein monomer or larger protein complexes, and may work in concert to influence the pattern of selective neuronal toxicity in each polyglutamine disease.
Polyglutamine Disease Proteins

These complex influences of protein context rely on both structural features of the protein flanking sequence as well as functional aspects of the behavior of normal and expanded polyglutamine proteins within the cell. It is therefore important to understand the biochemical properties and physiological roles of polyglutamine disease proteins. This section will highlight what is known about the individual polyglutamine proteins, and provide a framework in which to discuss the exploitation or dysfunction of these properties in the context of polyglutamine expansion.

SBMA is caused by expansion of a polyglutamine repeat within the N-terminal ligand-dependent transactivator domain of the androgen receptor (AR), upstream of the well characterized DNA and ligand binding domains (McEwan, 2001). Expansions in the range of 40 – 62 glutamines are sufficient to cause disease. The androgen receptor is a 110 kDa member of the steroid hormone receptor superfamily. Binding of the AR to its ligand (testosterone or dihydrotestosterone) masks its nuclear export signal, causing it to shift from the cytoplasm to the nucleus, where it recruits TFIIF and p160 family transcriptional coactivators to androgen-responsive genes through its transactivator and DNA binding domains. AR loss of function mutations result in the Androgen Insensitivity Syndrome, and activating mutations are associated with various cancers; neither of these phenotypes, however, has been associated with neurodegeneration (McEwan, 2001, Hirata et al., 2003, Hu et al., 2009).

Huntington Disease is caused by expansions of 36 or more polyglutamine residues within the N-terminus of the huntingtin protein (HTT); repeat expansions containing from 70 – 121 glutamines result in juvenile-onset disease. HTT is a very large 350 kDa protein that contains several annotated motifs within its N-terminus, including a polyproline domain, several HEAT repeats, at least two caspase-6 cleavage sites, and multiple post-translational modification sites (SUMOylation, phosphorylation, and palmitoylation) in addition to the polyglutamine repeat. The C-terminus of HTT is less
well characterized, but is known to contain a nuclear export signal and a protein-protein interaction domain that binds the Rab5 effector, HAP40 (Truant et al., 2007). Although many aspects of HTT function remain to be elucidated, it is thought to be a scaffolding protein involved in energy-dependent endocytic, lysosomal, and synaptic vesicular trafficking along microtubules. HTT is also involved indirectly in transcriptional regulation, via cytoplasmic sequestration of a transcriptional silencing protein (Koshy and Zoghbi, 1997, Truant et al., 2006a, Truant et al., 2007).

DRPLA is caused by repeat expansions of 49-88 glutamine residues in atrophin-1 (ATN1). ATN1 is an 1185 amino acid protein that contains an N-terminal nuclear import signal, a putative nuclear export signal, and fixed, smaller polyserine and polyproline repeats near the polyglutamine domain. Although relatively little is known about ATN1 function, it is assumed to be involved in transcriptional regulation, because its expression causes transcriptional repression, albeit less so in the expanded atrophin-1 protein (Koshy and Zoghbi, 1997, Zoghbi and Orr, 2000, Orr and Zoghbi, 2007).

The disease protein in SCA1 is ataxin-1 (ATXN1). SCA1 occurs with pure glutamine repeat expansions of 39-84 residues; however some non-affected individuals have been documented to contain mixed polyglutamine-histidine repeats of up to 44 residues long. ATXN1 is an approximately 90 kDa AXH domain-containing RNA-binding protein that appears to be involved in transcriptional regulation and possibly also RNA processing. It is able to interact with numerous proteins including RORα, Gfi-1, Capicua (CIC), and RBM17, in both small and large functional protein complexes. (Truant et al., 2006b, Orr and Zoghbi, 2007, Lim et al., 2008).

Polyglutamine expansions between 32 – 200 repeats in the N-terminus of ataxin-2 (ATXN2) result in SCA2. Ataxin-2 is a 140 kDa protein that contains Lsm and PAM2 domains and that interacts with RNA, known RNA-binding proteins (PAM2 and A2BP), and possibly polyribosomes. It appears to have a role in RNA metabolism and

SCA3 is caused by 52-86 polyglutamine repeats in the C-terminal region of ataxin-3 (ATXN3). Ataxin-3 is a 42 kDa deubiquitinating enzyme composed of an N-terminal Josephin protease domain, two ubiquitin interacting motifs (UIMs) before the polyglutamine domain, and a third C-terminal UIM after the polyglutamine domain in some splice isoforms. It can be posttranslationally modified through ubiquitination and phosphorylation of multiple lysine and serine residues, respectively. It binds ubiquitin both through its Josephin domain and its UIMs, and interacts with several protein quality control-associated proteins including ubiquitination factor E4B protein (E4B), human UV excision repair protein RAD23 homolog B (HHR23), and valsolin containing protein (VCP), as well as with the chromatin remodeling complex proteins HDAC3 and NCoR, and the transcription factors p300/CBP and PCAF. Although its role in protein quality control is more clearly established, ataxin-3 may also play a role in transcriptional repression (Doss-Pepe et al., 2003, Evert et al., 2006, Zhong and Pittman, 2006, Orr and Zoghbi, 2007).

Relatively small polyglutamine expansions of 20 – 33 repeats in the P/Q-type voltage-gated calcium channel subunit αA1 (Ca\textsubscript{v}\textsubscript{2.1}) cause SCA6. Structurally, Ca\textsubscript{v}2.1 is a prototypical voltage-gated cation channel, composed of four domains, each containing six transmembrane segments, followed by a C-terminal cytoplasmic tail. Splice variation in Ca\textsubscript{v}2.1-encoding transcripts is extensive, and a subset of splice isoforms encode an elongated cytoplasmic tail containing the polyglutamine domain. Functionally, Ca\textsubscript{v}2.1 is responsible for P-type calcium currents in cerebellar Purkinje cells (Zhuchenko et al., 1997, Kanumilli et al., 2006, Orr and Zoghbi, 2007).

SCA7 exhibits the greatest degree of anticipation among the polyglutamine diseases, with a range of disease-causing polyglutamine expansions from 34 – 306 repeats in the N-terminus of the ataxin-7 (ATXN7) protein. Due to the marked
variability in the size of observed polyglutamine repeats, ATXN7 varies in size from 130 – 150 kDa in expanded alleles. Functionally, ATXN7 is a member of two GCN5 histone acetylase containing complexes: TFTC and STAGA. These complexes are particularly important for chromatin remodeling at genes involved in photoreceptor differentiation and function (Zoghbi and Orr, 2000, Helmlinger et al., 2004, Orr and Zoghbi, 2007).

In SCA17, polyglutamine expansions exceeding a particularly high threshold (> 50 repeats) in the TATA-box Binding Protein (TBP) result in disease. TBP is a small ~38 kDa protein, which together with numerous TBP-associated factors (TAFs) forms TFIID, a core component of the transcription initiation machinery utilized by RNA polymerase II (Truant et al., 2006b, Orr and Zoghbi, 2007).

Apart from homology in the polyglutamine domain, the polyglutamine disease proteins have very little in common. They vary widely in size, from the relatively small ATXN3 (42 kDa) and TBP (38 kDa) to the massive 350 kDa HTT protein. Functionally, the polyglutamine disease proteins are involved in diverse cellular processes including transcriptional regulation (AR, HD, ATN1, ATXN1, ATXN3, ATXN7, and TBP), translational regulation (ATXN2), ion homeostasis (Ca\textsubscript{V}2.1), intracellular signaling (HD), vesicular trafficking (HD), and protein quality control (ATXN3). Moreover, the location of the polyglutamine repeat within the various host proteins differs greatly, with N-terminally located repeats in AR, HTT, ATXN1, ATXN2, ATXN7, and TBP, C-terminally located repeats in ATXN3 and Ca\textsubscript{V}2.1, and a more centrally located repeat in ATN1 (Table 2).

### Protein Context as a Modifier of Protein Misfolding and Aggregation

Protein misfolding and aggregation are pathological hallmarks of many neurodegenerative diseases including Alzheimer’s disease, Parkinson’s disease, and prion-mediated spongiform encephalopathies in addition to the polyglutamine diseases.
While the final detergent-insoluble, β-rich amyloid fibrils likely do not represent the primary toxic species in most of these diseases, intermediates in the misfolding cascade and the significant burden that they place upon the cellular protein quality control systems both appear to be pathophysiologically significant. Similarly, the multi-domain aggregation mechanism (Figure 3) proposed to explain the involvement of protein context in polyglutamine-dependent protein misfolding is not unique to this set of disease proteins. For example, aggregation of the yeast prion protein Ure2p requires oligomerization mediated by its stable C-terminal domain before fibril formation is mediated by its amyloidogenic N-terminus (Thual et al., 2001). An emerging trend suggests that self-association of polyglutamine-flanking sequences may be a common feature among polyglutamine proteins. This hypothesis is supported by the observation that fusion of polyglutamine repeats to inherently stable proteins, such as S. aureus protein A and glutathione-S-transferase (GST), produces proteins with a greatly reduced ability to aggregate than either naked polyglutamine or polyglutamine within known disease proteins (Saunders and Bottomley, 2009). It is therefore not surprising that protein context is proving to be a significant modifier of polyglutamine protein conformational instability, misfolding, and aggregation.

The polyglutamine disease literature is rife with examples of alterations in protein context modulating protein aggregation. One common theme is the modulation of aggregation propensity by proteolytic cleavage of the disease protein at susceptible sites, often via caspases. Multiple polyglutamine proteins exhibit some degree of targeted proteolysis in vivo. In both HD and SCA7, caspase-mediated cleavage of the disease protein into truncated polyglutamine-containing fragments enhances aggregation (Truant et al., 2006b, Young et al., 2007). Young and colleagues also demonstrated that this enhanced aggregation phenotype correlated with transcriptional dysregulation and cell death (Young et al., 2007). The “toxic fragment hypothesis” of polyglutamine pathogenesis is based on these and many other examples.
Posttranslational modification of disease proteins is another element of protein context that can exert a significant effect on aggregation. Both SUMOylation (Mukherjee et al., 2009) and acetylation (Thomas et al., 2004) inhibit aggregation of expanded AR protein. Specific types of posttranslational modifications, however, do not always modulate aggregation in the same way. ATN1 intranuclear inclusions (NIs) are SUMOylated, and a defective SUMO1 mutant was able to significantly reduce the aggregation and toxicity exhibited by expanded ATN1 (Terashima et al., 2002). Similarly, phosphorylation of polyglutamine disease proteins by the kinase Akt has differential effects, leading to decreased aggregation and toxicity of HTT and the AR, while Akt-phosphorylation of ATXN1 reduces its degradation, leading to increases in inclusion formation and toxicity (Jorgensen et al., 2009, Pennuto et al., 2009).

Polyglutamine Expansion-Dependent Alteration of
Endogenous Function Depends on Protein Context

Polyglutamine length-dependent changes in endogenous protein functions can superimpose an additional deleterious loss of function or a polyglutamine-independent toxic effect on general polyglutamine-mediated toxicity. The best characterized example of superimposed loss of endogenous function is seen in SCA1. The RNA-binding protein ATXN1 interacts with various proteins in small (with RORα or Gfi-1) or large (with CIC or RBM17) functional protein complexes. Expansion of the polyglutamine domain of ATXN1 modulates these protein-protein interactions, leading to a decrease of ATXN1-CIC containing complexes, an increase in ATXN1-RBM17 containing complexes, and a disruption of the stabilizing effects of ATXN1 on RORα and Gfi-1 (Truant et al., 2006b, Orr and Zoghbi, 2007, Lim et al., 2008). Additionally, while nuclear export normally occurs with nonpathogenic (i.e. non-expanded) ATXN1, it is limited in the context of a polyglutamine expansion (Irwin et al., 2005). This combination of marked alterations in expanded ATXN1 behavior, compared to that of unexpanded ATXN1, led to the
hypothesis that disruption of endogenous ATXN1-interactions constitutes a loss of endogenous function, presumably transcriptional regulation or RNA processing. In support of this hypothesis, overexpression of the ATXN1 homologue, ataxin-1-like, partially rescues expanded ATXN1 toxicity (Bowman et al., 2007), whereas depletion of wild type ATXN1 on an expanded ATXN1 knock-in background exacerbates toxicity (Lim et al., 2008).

One of the best characterized examples of a superimposed aggregation-independent gain of function is the behavior of ATXN7 in the retina. ATXN7 functions as a member of the TFTC and STAGA complexes, which are involved in epigenetic chromatin remodeling through histone acetylation. Expanded and wild type ATXN7 are similarly incorporated into both of these complexes within mammalian cells from an SCA7 patient (Helmlinger et al., 2006b). However, within rod photoreceptors the TFTC/STAGA complexes containing expanded polyglutamine demonstrated abnormally enhanced recruitment to the promoters of genes involved in rod differentiation and function. At these sites, the aberrant complexes hyperacetylated histones, but paradoxically caused transcriptional deactivation of these rod-specific gene promoters (Helmlinger et al., 2006a). The occurrence of this protein context-dependent transcriptional deactivation in a rod-specific manner helps to explain macular degeneration in SCA7 – a unique and defining feature of this polyglutamine disease.

Protein Context as a Modifier of Subcellular Localization and Expression Patterns

Many polyglutamine proteins contain nuclear localization signals and nuclear export signals outside of the polyglutamine domain. Because intranuclear neuronal inclusions are present in most polyglutamine diseases, many groups have explored the differential toxicity of polyglutamine disease proteins in the cytoplasmic versus nuclear compartments. HTT contains a C-terminal nuclear export signal, which is lost when
caspase cleavage releases the N-terminal HTT fragment discussed earlier; loss of this export signal is associated with enhancement of aggregation-specific phenotypes (Xia et al., 2003). Similarly, disruption of the nuclear export signal in ATXN7 increases toxicity (Taylor et al., 2006). A more physiological example is provided by ligand binding to the AR, which masks its nuclear export signal, resulting in a net increase of ligand-bound receptors in the nucleus where they can influence transcription of androgen-responsive genes. This process seems to be crucial for manifestation of neurotoxicity in SBMA. Females who are homozygous for the X-linked expansion exhibit a much milder clinical course, presumably due to decreased androgen-dependent signaling (Schmidt et al., 2002). Similarly in a Drosophila model of SBMA, administration of androgens is needed to induce a rough eye phenotype; however this effect was blocked by cytoplasmic trapping of the expanded protein (Takeyama et al., 2002). Castration drastically rescued the disease phenotype in symptomatic SBMA transgenic male mice (Katsuno et al., 2002). These studies illustrate how disease protein-specific effectors of subcellular localization can modulate polyglutamine toxicity. Additional cell-specific elements of protein context, including various posttranslational modifications and posttranscriptional processes such as alternative splicing, must also contribute to protein context-dependent effects on the clinical manifestations of specific polyglutamine diseases.

Alternative Splicing as an Element of Protein Context

Alternative splicing is an important mechanism by which functional proteomic diversity is achieved in eukaryotes. Between 70 – 95% of mammalian transcripts undergo alternative splicing (Johnson et al., 2003, Wang et al., 2008). Patterns of alternative splicing are cell type-specific, and can be regulated through physiological processes, such as development, during pathological processes, such as cancer progression, and through pharmacological manipulation. Moreover, associations of silent mutations and single nucleotide polymorphisms with aberrations in alternative splicing
are becoming increasingly recognized. Alternative splicing, however, has rarely been investigated as an element of protein context in polyglutamine diseases.

Table 3 highlights our current understanding of alternative splicing among polyglutamine disease gene products. It illustrates the breadth of transcript diversity and highlights the known phenotypic implications of alternative splicing of polyglutamine disease transcripts. Multiple alternatively spliced isoforms of the AR have been documented in normal and transformed human tissues. These variants are generated through utilization of alternative transcriptional start sites, exon skipping, intraexonic splicing, and inclusion of cryptic exons. Alteration of physiological AR patterns of splicing can occur through cis-acting coding and noncoding SNPs/mutations and in-trans via alterations in the splicing microenvironment. Enrichment of AR isoforms that abrogate ligand binding, DNA binding, and transactivation result in Androgen Insensitivity Syndrome, whereas constitutively active isoforms have been associated with at least two neoplasias including prostate cancers that acquire resistance to anti-androgen therapy (Hirata et al., 2003, Hu et al., 2009).

Evidence of HTT transcript diversity is limited in humans. Although at least two transcripts varying in the selection of 3’ polyadenylation signals are detectable in mammalian tissues, including humans (Ambrose et al., 1994, Matsuyama et al., 2000), frank splice variants have only been described in the mouse and pig. The phenotypic significance of this variation at the isoform level has not been explored (Lin et al., 1994, Matsuyama et al., 2000).

ATN1 is alternatively spliced in human, murine, and rat tissues. In each case, utilization of an alternate splice acceptor site leads to the skipping of a single non-repeat associated glutamine-encoding (CAG) codon; this minor isoform shifts the ratio of nuclear to cytoplasmic ATN1 towards an enhanced cytoplasmic localization, compared to the major (glutamine-included) isoform (Tadokoro et al., 2005).
The ATXN1 transcript is alternatively spliced in human tissues through exon skipping within the 5'UTR. Although several variants are detectable, they are not predicted to alter the coding sequence (Banfi et al., 1994). The functional significance of this variation, if any, has not yet been explored.

ATXN2 mRNA is alternatively spliced in human and murine tissues. Multiple variants are derived from exon skipping. Although minor ATXN2 splice isoforms are predicted to result in the loss of proline-rich domains, and one variant was found to be relatively enriched in cerebellar mRNA, the functional implications of this splice variation has not been demonstrated at the protein level (Nechiporuk et al., 1998, Sahba et al., 1998, Affaitati et al., 2001).

Ataxin-3 is alternatively spliced in human and rodent tissues. Multiple forms of alternative splicing have been implicated in this variation, including exon skipping, utilization of alternate splice donor sites and splice acceptor sites, and the inclusion of cryptic exons. Transcript diversity is further increased through the usage of alternative polyadenylation signals. The functional significance of this alternative splicing has not been evaluated aside from the present study (in Chapter 2). Some variants are predicted to encode protein isoforms that lack the Josephin protease domain and/or a functional UIM3 (Kawaguchi et al., 1994, Goto et al., 1997, Paulson et al., 1997, Schmitt et al., 1997, Ichikawa et al., 2001, Bettencourt et al., 2009).

The CaV2.1-encoding transcript CACNA1A mRNA exhibits a broad range of alternative splicing in human, rabbit, and rat tissues. Alternative transcripts are formed combinatorially through exon skipping, mutually exclusive exon selection, and alternative splice acceptor site usage. Of particular relevance to SCA6, some variants lack (CAG)$_n$ repeats because they do not utilize a weak splice acceptor site that leads to the inclusion of a pentanucleotide GGCAG sequence at the beginning of exon 47 (Zhuchenko et al., 1997, Kanumilli et al., 2006). Alternative splice acceptor site variants that encode polyglutamine-containing isoforms are enriched in cerebellar Purkinje cells.
of SCA6 patients, but not in controls (Tsunemi et al., 2008). Whether this intriguing phenomenon alters levels of mutant protein in SCA6 cerebella has yet to be determined.

ATXN7 is alternatively spliced to produce two distinct transcripts in human and murine tissues, through the process of exon skipping. The minor ATXN7b splice isoform exhibits a more CNS-enriched pattern of expression, corresponding more closely to the distribution of SCA7 susceptible cell types than the predominant, ubiquitously expressed ATXN7a isoform. Yet, this isoform exhibits a more cytoplasmic distribution and is not found in NI’s in human disease brain (Einum et al., 2003, Strom et al., 2005). Both isoforms similarly repress transcription in reporter assays (Strom et al., 2005).

Finally, the skipping of a 3’ splice donor site in exon 3 of TBP results in a minor splice variant that is predicted to encode a truncated isoform of the protein indistinguishable from a putative proteolytic fragment. This variant has been demonstrated in human, murine, and bovine tissues, and in silico is predicted to exist in several other species. This aggregation-prone splice isoform of TBP appears to be enriched in the middle temporal gyrus of Alzheimer’s disease patients and in the brains of Huntington disease (HD) patients (Reid et al., 2009). Whether this isoform plays a significant role in SCA17 pathogenesis or is merely a marker of dysregulated splicing in AD and HD has yet to be determined.

The Importance of Protein Context in Model Systems

Polyglutamine diseases have been modeled in many cellular, invertebrate, and vertebrate systems. The spectrum of complete protein context utilization in these systems spans the gamut, from systems expressing pure polyglutamine repeats or truncated polyglutamine-containing fragments to systems expressing multiple physiological splice isoforms from a precise genomic context. In vitro, pure polyglutamine repeats exhibit markedly different aggregation properties including lower order kinetics than polyglutamine disease proteins (Saunders and Bottomley, 2009). In a cell-based yeast
model system of HD, expanded polyglutamine exon 1 fragments that also contain the neighboring proline-rich domain result in either a single large round aggregate or diffuse expression only, whereas yeast expressing isoforms that lack the polyproline moiety contained numerous, small, irregular aggregates. The single aggregate phenotype conferred by the polyproline containing fragment was accompanied by a marked reduction in cytotoxicity comparatively (Dehay and Bertolotti, 2006). Numerous mouse models of HD also exist, including transgenic animals expressing HTT fragments, transgenic animals expressing full length HTT, and knock-in animals expressing a polyglutamine expansion within the murine homologue, Htt (Lin et al., 1999, Heng et al., 2008). Comparison of a limited subset of the existing murine models of HD highlights some of the advantages and disadvantages of using variable stringencies of genomic and protein context, as discussed below.

R2/6 transgenic mice express an exon 1 fragment of HTT containing ~130 repeats. Although these mice exhibit a robust, age-related neurological phenotype, including involuntary movements, weight loss, and early death, they lack the complete HTT protein context and do not recapitulate the selective neurodegenerative features of HD pathology. The HTT cDNA transgenic model designed by Reddy and colleagues to overexpress full length Q_{89} HTT likewise exhibit a rapidly progressive neurological syndrome, including early clasping and hyperactivity and later akinesia, but also exhibit pathological HD hallmarks of selective neurodegeneration, mirroring that seen in HD patients, with the addition of neuronal inclusions in regions generally spared in human disease. Full length yeast artificial chromosome (YAC) constructs expressing full length Q_{89} HTT under its endogenous promoter develop a much more attenuated phenotype, with electrophysiological changes preceding mild neuropathological changes that include enhanced nuclear translocation of HTT selectively in medium spiny neurons of the striatum (a population of neurons known to be selectively vulnerable in HD). Knock-in of a 72–80 CAG repeat expansion into the murine Htt gene produced an even less robust
phenotype, manifesting with modest electrophysiological differences in some of the same pathways that were altered in the YAC Q_{89} model. Much larger repeat expansions were required to produce a robust phenotype in genomic context-specific models, as evidenced by the Hdh^{Q140}, Hdh^{Q150}, and YAC Q_{128} models of HD (Lin et al., 2001, Menalled et al., 2003, Slow et al., 2003).

In summary, models that maintain near endogenous levels and patterns of polyglutamine expression (e.g., YAC, Bacterial Artificial Chromosome (BAC), and knock-in models), provide the greatest physiological specificity for identifying dysregulated processes in vulnerable subsets of neurons. They are also likely to be the most useful for studying DNA and RNA context-specific phenomena, including repeat instability and alternative splicing. They are, however, unlikely to produce a robust, early onset phenotype in the absence of a suprapathophysiological repeat length.

Therapeutic Implications of Understanding the Role of Protein Context in Polyglutamine Diseases

In order to develop rational therapeutics to intervene in disease progression, we first need to know which elements of polyglutamine-mediated toxicity are fundamental triggers of selective neuronal toxicity and degeneration. An ideal therapeutic, from a drug design and delivery standpoint, would be an orally bioavailable small molecule that disrupts a key step common to the degenerative processes in all polyglutamine diseases. Such a therapeutic might target cellular mechanisms of protein quality control, such as chaperones or the ubiquitin proteasome system, or general neuroprotective pathways, such as neurotrophic factor signaling cascades or the antioxidant capacity of the intracellular environment. The next most desirable approach, from a drug design perspective, would be an effective polyglutamine-targeted treatment, although current strategies in this category present challenges in delivery. An example might be a locked nucleic acid or peptide-linked nucleic acid that preferentially targets expanded (CAG)_n.
repeats, resulting in translational silencing or enhanced RNA degradation (Hu et al., 2009). Unfortunately, any of these approaches can potentially be complicated by protein context-specific factors that underlie the restriction of toxicity to a subset of neurons that are vulnerable in each disease.

An alternate approach to rational therapeutic design is to exploit features of protein context that contribute to disease-specific neuronal toxicity and degeneration. An exciting example of this type of approach is currently underway in the field of SBMA therapy. The androgen dependence of polyglutamine-induced nuclear localization, toxicity, and neurodegeneration seen in SBMA (Katsuno et al., 2002, Schmidt et al., 2002, Takeyama et al., 2002), prompted further studies that compared the efficacy of the long acting luteinizing hormone releasing hormone (LHRH) agonist leuprorelin with the androgen receptor antagonist flutamide. While leuprorelin suppresses androgen production, the direct AR antagonist flutamide prevents transcriptional activation of androgen-responsive genes, but does not prevent the AR from translocating into the nucleus. Consistent with the hypothesis that protein context-dependent alterations in subcellular localization, rather than loss of transcriptional regulatory functions, are required for expanded AR toxicity, only leuprorelin significantly ameliorated protein aggregation and inclusion formation in anterior horn cells and muscle fibers (Katsuno et al., 2004). In phase II clinical trials, leuprorelin improved functional scores and swallowing studies over the course of 144 weeks of consecutive randomized placebo controlled and open label clinical paradigms. The results of these Phase II clinical trials were deemed sufficient to support the initiation of a Phase III clinical trial (Banno et al., 2009). Ideally, a better understanding of the mechanisms by which protein context modulates polyglutamine toxicity will lead to additional advances in rational therapeutic design capable of slowing or halting the progression neurodegeneration in other polyglutamine diseases as well.
Focus of this thesis: Protein Context in SCA3

The ATXN3 gene contains 11 annotated exons spanning just over 48 kb (Ichikawa et al., 2001). It encodes ATXN3, a deubiquitinating enzyme that contains a Josephin protease domain, followed by two UIMs, the polyglutamine domain, and either a third UIM or a hydrophobic domain, depending on 3’ alternative splicing of the ATXN3 transcript (Figure 4). The 3UIM splice isoform, which is encoded by all 11 exons, is well conserved among mammals and amphibians, but diverges in fish and worms at its extreme C-terminus. In contrast, the C-terminal domain of the 10 exon-encoded 2UIM isoform is poorly conserved across taxa (Figure 5). Alternative splicing of 5’ exons has also been reported, and is predicted to produce minor splice isoforms with disrupted catalytic activity (Ichikawa et al., 2001, Bettencourt et al., 2009). Functionally, ataxin-3 has been implicated both in protein quality control and in transcriptional repression. The complexity of ataxin-3 structure and function suggests that protein context-specific features of ataxin-3 could serve as potential modulators of polyglutamine toxicity.

Ataxin-3 illustrates several of the various mechanisms of protein context-mediated alterations in polyglutamine toxicity mentioned earlier. Ataxin-3 undergoes a multi-domain mechanism of aggregation in vitro, which implicates regions outside of the polyglutamine domain in the kinetics of protein conformational change, misfolding, and aggregation (Ellisdon et al., 2006). One mouse model of SCA3 implicates proteolytic cleavage outside of the polyglutamine domain of ATXN3 in the degenerative phenotype, consistent with the toxic fragment hypothesis. However, the exact nature of this cleavage site has not been elucidated (Goti et al., 2004, Colomer Gould et al., 2007). Posttranslational modification of ataxin-3 through glycogen synthase kinase 3β-dependent phosphorylation of ATXN3 serine 256 enhances polyglutamine-dependent aggregation (Fei et al., 2007). ATXN3 phosphorylation by casein kinase-2 at UIM-associated serine residues 340 and 352 enhances nuclear ATXN3 localization. This latter observation is particularly relevant to context-dependent polyglutamine toxicity, because
nuclear localization of expanded ATXN3 enhances its toxicity (Bichelmeier et al., 2007). Finally, involvement of ATXN3 in protein quality control mechanisms and transcriptional repression (a common feature of cellular responses to an abundance of misfolded proteins), suggests that any polyglutamine-dependent loss of endogenous ataxin-3 function may be particularly detrimental to selectively vulnerable neuronal populations.

Rationale and Aims of the Thesis

In my studies of polyglutamine disease, I have been particularly interested in the pathophysiological mechanism leading to selective neuronal toxicity. Because the clinical manifestations of individual polyglutamine diseases vary in a protein context-dependent manner, it is important to understand both the endogenous functions of the various disease proteins and how these properties are altered by polyglutamine expansion. Based on recent advances in our understanding of ataxin-3 structure and function, I chose to study elements of ataxin-3 protein context in the polyglutamine disease SCA3, the most common dominantly inherited ataxia. Ataxin-3 is alternatively spliced in multiple tissues. Therefore, before we can understand the role that protein context plays in SCA3, we need to define the range of variability in ataxin-3 protein context in brain and other tissues. To address the functional implications of ATXN3 splice variation, we must determine whether isoforms present within the brain affect known determinants of polyglutamine neurotoxicity, including the preservation or loss of endogenous function and protein context-dependent gain of function mechanisms. Chapter 2 investigates this issue in SCA3. Finally, in order to understand the role of protein context in vivo, we need to expand the range of protein context-specific murine models appropriate for the study of SCA3. In particular, knock-in of (CAG)_n expansions into the murine homologues of conserved disease genes is a desirable approach to modeling dominantly inherited ataxias, which can recapitulate the tissue-specific features
of an individual polyglutamine disease. Highly conserved orthologs of ATXN3 exist in many species, including *Mus musculus Atxn3*, making the prospect of a knock-in model of SCA3 highly feasible. Chapter 3 seeks to develop such a knock-in model of SCA3.

Overall, the studies described in this thesis help to refine our understanding of the molecular biology of human and murine ataxin-3, how these molecular biological processes change in the context of polyglutamine expansion, and how protein context-dependent alterations in these processes may influence selective neuronal toxicity in SCA3. In Chapter 2, I examine the extent of ataxin-3 alternative splicing in brain tissue and assess the impact of this splice variation on elements of ataxin-3 protein context that could exacerbate either alterations of ataxin-3 function or known polyglutamine toxicity associated phenotypes. In Chapter 3, I describe the creation of a polyglutamine-expanded knock-in model of SCA3, a genetically precise model of disease that will be ideal for future *in vivo* studies of SCA3 pathogenesis, including endogenous splice variant analysis, electrophysiological analysis of vulnerable neuronal populations, and studies of genomic instability. The knock-in model may also prove useful for in vivo testing of possible therapeutic agents. Finally, in Chapter 4 I discuss the significance of our findings and highlight several future directions for research addressing key questions raised by the current studies. Insights into how the tissue-specific molecular biology of ataxin-3 relates to SCA3 pathology will enhance our understanding of the complex interactions between generalized features of polyglutamine toxicity and protein context in polyglutamine disease pathophysiology.
Table 1. Clinical heterogeneity among the polyglutamine disorders.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Inheritance</th>
<th>OMIM number</th>
<th>Main clinical features</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCA1</td>
<td>AD</td>
<td>164400</td>
<td>Ataxia, slurred speech, spasticity, cognitive impairment</td>
</tr>
<tr>
<td>SCA2</td>
<td>AD</td>
<td>183090</td>
<td>Ataxia, slow saccades, decreased reflexes, polyneuropathy, motor neuropathy, infantile variant</td>
</tr>
<tr>
<td>SCA3</td>
<td>AD</td>
<td>109150</td>
<td>Ataxia, parkinsonism, severe spasticity</td>
</tr>
<tr>
<td>SCA6</td>
<td>AD</td>
<td>183086</td>
<td>Ataxia, dysarthria, nystagmus, tremor</td>
</tr>
<tr>
<td>SCA7</td>
<td>AD</td>
<td>164500</td>
<td>Ataxia, retinal degeneration, cardiac involvement in infantile variant</td>
</tr>
<tr>
<td>SCA17</td>
<td>AD</td>
<td>607136</td>
<td>Ataxia, behavioural changes or psychosis, intellectual deterioration, seizures</td>
</tr>
<tr>
<td>DRPLA</td>
<td>AD</td>
<td>125370</td>
<td>Ataxia, epilepsy, choreoathetosis, dementia</td>
</tr>
<tr>
<td>SBMA</td>
<td>XL</td>
<td>313200</td>
<td>Motor weakness, swallowing difficulty, gynecomastia, hypogonadism</td>
</tr>
<tr>
<td>HD</td>
<td>AD</td>
<td>143100</td>
<td>Severe movement abnormalities, chorea, dystonia, cognitive decline, psychiatric features</td>
</tr>
</tbody>
</table>


Note: Despite the common molecular insult shared by polyglutamine diseases, the clinical manifestations vary greatly among these disorders. This table reviews some of the commonalities and differences among the polyglutamine diseases, clinically.
Table 2. Polyglutamine diseases and their gene products.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Protein name</th>
<th>Protein function, size, polyQ position, and disease repeat range</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBMA</td>
<td>Androgen receptor</td>
<td>Testosterone-activated steroid receptor</td>
</tr>
<tr>
<td>HD</td>
<td>Huntingtin</td>
<td>Possible scaffolding protein linked to diverse cellular pathways</td>
</tr>
<tr>
<td>DRPLA</td>
<td>Atrophin-1</td>
<td>Possible transcriptional corepressor</td>
</tr>
<tr>
<td>SCA1</td>
<td>Ataxin-1</td>
<td>Transcriptional corepressor involved in transcription regulation, cell specification and synaptic activity</td>
</tr>
<tr>
<td>SCA2</td>
<td>Ataxin-2</td>
<td>Component of RNA processing and translational regulation pathways</td>
</tr>
<tr>
<td>SCA3</td>
<td>Ataxin-3</td>
<td>Deubiquitinating enzyme involved in protein quality control</td>
</tr>
<tr>
<td>SCA6</td>
<td>P/Q-type calcium-channel subunit α1A</td>
<td>Voltage-sensitive calcium-channel subunit</td>
</tr>
<tr>
<td>SCA7</td>
<td>Ataxin-7</td>
<td>Component of histone acetyltransferase complex (TFTC/STAGA) and transcriptional regulation pathways</td>
</tr>
<tr>
<td>SCA17</td>
<td>TATA-box-binding protein</td>
<td>Component of core transcriptional complex TFIIID</td>
</tr>
</tbody>
</table>


Note: Although all polyglutamine disease proteins contain a polymorphic \((CAG)_n\) trinucleotide repeat expansion, they vary widely in their protein context, including size, function, and location of the polyglutamine repeat. This table illustrates the diversity of protein context among polyglutamine disease proteins.
Table 3. Alternative splicing of polyglutamine disease gene products.

<table>
<thead>
<tr>
<th>Disease Protein</th>
<th>Species with Documented Splice Variation</th>
<th>Type of Splicing Observed</th>
<th>Associated Phenotype</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Androgen Receptor (AR)</td>
<td>Human</td>
<td>Alternative TSS, exon skipping, intraexonic splicing, inclusion of cryptic exons</td>
<td>Androgen insensitivity, constitutive AR activation, neoplasia</td>
<td>Hirata, 2003; Hu, 2009</td>
</tr>
<tr>
<td>Huntingtin (HTT)</td>
<td>Mouse</td>
<td>Exon skipping; alternative 3’ polyadenylation signals in mouse, rat, pig, and human</td>
<td>Not determined</td>
<td>Lin, 1994; Matsuyama, 2000</td>
</tr>
<tr>
<td>Atrophin-1 (ATN1)</td>
<td>Human</td>
<td>Alternate splice acceptor site usage, leading to skipping of a single non-repeat (CAG) trinucleotide</td>
<td>Enhanced cytoplasmic localization in the minor (CAG)-skipped isoform</td>
<td>Tadokoro, 2005</td>
</tr>
<tr>
<td>Ataxin-1 (ATXN1)</td>
<td>Human</td>
<td>5'UTR exon skipping</td>
<td>Not determined</td>
<td>Banfi, 1994</td>
</tr>
<tr>
<td>Ataxin-2 (ATXN2)</td>
<td>Human</td>
<td>Exon skipping</td>
<td>Not determined; predicted loss of proline-rich domains; one variant enriched in cerebellar mRNA</td>
<td>Sahba, 1998; Nechiporuk 1998; Affaitati, 2001</td>
</tr>
<tr>
<td>Ataxin-3 (ATXN3)</td>
<td>Human</td>
<td>Exon skipping, alternate splice donor site usage, alternate splice acceptor site usage, inclusion of cryptic exons; alternative polyadenylation signals</td>
<td>Not determined; predicted loss of protease domain and UIM3 function in some isoforms</td>
<td>Bettencourt, 2009; Goto, 1997; Ichikawa, 2001; Kawaguchi, 1994; Paulson, 1997; Schmitt, 1997</td>
</tr>
</tbody>
</table>
Table 3. Continued.

<table>
<thead>
<tr>
<th>P/Q-type voltage-gated calcium channel, subunit α1A1 (CaV2.1)</th>
<th>Human Rabbit Rat</th>
<th>Exon skipping, mutually exclusive exon selection, alternative splice acceptor site usage; some variants lack (CAG)$_n$ repeats</th>
<th>A subset of alternative splice acceptor site variants encoding polyglutamine containing isoforms are enriched in the Purkinje cells of SCA6 patients, but not in controls</th>
<th>Tsunemi, 2008; Zhuchenko, 1997 Kanumilli, 2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ataxin-7 (ATXN7)</td>
<td>Human Mouse</td>
<td>Exon skipping</td>
<td>A neurally-enriched isoform increases cytoplasmic distribution, is absent from nuclear inclusions, exhibits similar transcriptional activities</td>
<td>Einum, 2003; Ström, 2005</td>
</tr>
<tr>
<td>TATA-box binding Protein (TBP)</td>
<td>Human Mouse Cow</td>
<td>3’ splice donor site skipping</td>
<td>Putatively enriched in HD and AD brain, higher aggregation propensity in overexpression systems</td>
<td>Reid, 2009</td>
</tr>
</tbody>
</table>

Note: Of the nine recognized pure polyglutamine disease genes, eight are known to be alternatively spliced in humans, and all exhibit multiple mature mRNA transcripts. This table illustrates the breadth of transcriptional diversity and highlights the known phenotypic implications of the alternative splicing of polyglutamine disease gene products. Abbreviations: transcriptional start site (TSS), untranslated region (UTR), ubiquitin interacting motif (UIM).

Note: Human HTT transcripts utilizing alternative 3’ polyadenylation signals have been described; yet exon skipping events observed in other mammalian species have not been noted among human (or rat) transcripts.
Figure 1. The nucleotide repeat expansion disorders.
This superfamily of neuromuscular and neurodegenerative diseases is caused by the expansion of polymorphic tri-, tetra-, penta-, or dodecanucleotide repeats (dashed boxes) within the disease genes. This clinically heterogeneous group of disorders can be divided into subgroups based on pathological mechanism (ovals). The loss of function tri- and dodecanucleotide repeat disorders (green) result when very large untranslated repeat expansions (> 200 trinucleotide repeats or > 30 dodecanucleotide repeats) lead to epigenetic transcriptional silencing and/or disruption of productive transcriptional elongation. In the RNA-mediated gain of function diseases (blue), tri- or tetra-nucleotide expansions induce aberrant RNA-protein interactions, leading to ribonucleoprotein inclusions and dysregulation of critical cellular functions such as mRNA splicing. The polyglutamine and polyalanine diseases (pink) are mediated at the protein level by a toxic gain of function conferred at least in part by misfolding of the polyglutamine domain. SCA8 exhibits elements of both RNA- and protein-mediated gain of function. At least three additional diseases (gray) are caused by nucleotide repeat expansions in constitutively noncoding (SCA10, SCA12) or alternatively spliced, variably coding sequences (HDL2), through unknown pathological mechanisms.
Figure 2. Mechanisms of polyglutamine disease toxicity.
When a polyglutamine protein has a non-expanded repeat length, it maintains its native conformational state(s) and standard interaction partners, leading to a normally functioning protein that is not toxic. Expanded polyglutamine repeats induce conformational changes in polyglutamine disease proteins, which can lead to alterations in protein-protein interactions, disruption of endogenous function, proteolytic cleavage, protein aggregation, and perturbations in the cellular mechanisms of protein quality control, as well as alterations in cellular processes such as transcriptional regulation, RNA metabolism, cytoskeletal architecture, protein transport, mitochondrial function, and ion homeostasis.

Figure 3. Multi-domain aggregation mechanism for polyglutamine disease proteins with or without expansion of the polyglutamine repeat.

Some polyglutamine proteins exhibit intrinsic, polyglutamine-independent self-association. Saunders and Bottomley have proposed a multi-stage aggregation mechanism to explain the fibrillation kinetics observed in vitro. Initially, the monomer is in equilibrium between a native conformation and a more aggregation-prone conformation, which has the ability to self-associate. This initial protein context-dependent oligomerization increases the local concentration of polyglutamine domains within the oligomer, leading to a polyglutamine-dependent formation of β-rich, amyloid-like, SDS-insoluble fibrils. Compared to the unexpanded polyglutamine disease proteins, the sequence flanking an expanded polyglutamine repeat is more likely to be destabilized to an aggregation-prone conformation, leading to a higher oligomer concentration, and abundant seeding of fibrillation reactions, with a net effect of rapid aggregation kinetics in vitro, and detectable levels of aggregated protein in cells. Because unexpanded polyglutamine proteins have a lower rate of aggregation-prone conformer formation and smaller polyglutamine domains to facilitate fibrillogenesis, any oligomers formed would be more likely to be manageable by the cellular protein quality control machinery. These differences in aggregation kinetics may explain why non-expanded polyglutamine proteins do not form detectable intracellular aggregates alone, but can be recruited to aggregates pre-seeded by expanded polyglutamine proteins.

Figure 4. The genomic and protein context of Ataxin-3, which causes SCA3. Schematic representation of the ATXN3 gene showing exons that encode specific functional domains of the two predominant “full length” ataxin-3 isoforms. The splicing pattern of the originally identified 2UIM ataxin-3 encoding transcript is shown below (†); utilization of an alternative splice site linking exon 10 to exon 11 (shown above, ‡) generates 3UIM ataxin-3. Asterisks indicate exons encoding the three amino acids comprising the Josephin domain catalytic triad, polyQ denotes the polyglutamine repeat domain, and the arrowhead indicates a polymorphic Tyr/Stop-encoding residue within the hydrophobic domain (φ) found in the C-terminus of the 2UIM isoform. The product encoded by each 3’ splice variant is shown at the protein level.
Figure 5. 3UIM ataxin-3 is well conserved among mammals and amphibians. ClustalW alignment of ATXN3 isoforms in taxa from *Homo sapiens* to *C. elegans*. Red residues indicate identity, blue residues indicate similarity, and black residues indicate a lack of conservation. The N-terminal Josephin domain (green underscore) including the catalytic C-H-N triad (boxes) and the consensus residues (*) within each UIM (red underscore), are particularly well conserved. The 3UIM splice isoform (encoded by all 11 exons) is well conserved among mammals and amphibians, but diverges in fish and worms at its extreme C-terminus, whereas the C-terminal domain of the 10 exon-encoded 2UIM isoform is poorly conserved among all taxa included in this analysis. Note that this alignment shows the 2UIM-long C-terminal SNP variant produced by both MJD15.4 and MJD84.2 YAC transgenic lines; the polymorphic tyrosine/stop residue is underscored in this alignment (Y).
CHAPTER 2
ALTERNATIVE SPLICING OF ATAXIN-3 AS AN ELEMENT OF PROTEIN CONTEXT IN SCA3

Abstract
Protein context clearly influences neurotoxicity in polyglutamine diseases, but the contribution of alternative splicing to this phenomenon has rarely been investigated. Ataxin-3, a deubiquitinating enzyme and the disease protein in SCA3, is alternatively spliced to encode either a C-terminal hydrophobic stretch or a third UIM (2UIM and 3UIM isoforms, respectively). In light of emerging insights into ataxin-3 function, we examined the significance of this splice variation. We confirmed neural expression of several minor 5’ variants and both of the known 3’ ataxin-3 splice variants. Regardless of polyglutamine expansion, 3UIM ataxin-3 is the predominant isoform in brain lysates. Although 2UIM and 3UIM ataxin-3 display similar in vitro deubiquitinating activity, 2UIM ataxin-3 is more prone to aggregate and more rapidly degraded by the proteasome. Our data demonstrate how alternative splicing of sequences distinct from the trinucleotide repeat can alter properties of the encoded polyglutamine disease protein and thereby perhaps contribute to selective neurotoxicity.

Introduction
The polyglutamine neurodegenerative diseases are caused by the expansion of polyglutamine-encoding CAG trinucleotide repeats within specific genes. Polyglutamine expansion promotes disease protein misfolding, triggering a pathogenic cascade leading to neurodegeneration, with age of disease onset inversely correlated to expansion length. While all polyglutamine disease proteins are widely expressed, the patterns of neurodegeneration and clinical manifestations of disease vary significantly (Orr and
Zoghbi, 2007), suggesting that the protein context of each expansion contributes to selective neuronal toxicity by influencing factors such as subcellular localization, protein-protein interactions, endogenous function, and aggregation. Cell-specific elements of protein context are particularly attractive candidate determinants of selective toxicity. Because the precise protein context of a disease protein will vary between splice variants, alternative splicing may influence patterns of polyglutamine protein-induced neurodegeneration.

Alternative splicing is an important mechanism by which proteomic diversity is achieved in eukaryotes, with most mammalian transcripts undergoing alternative splicing (Johnson et al., 2003, Wang et al., 2008). Patterns of alternative splicing can be cell-specific and regulated through physiological or pathological processes. As reviewed in Chapter 1, many transcripts that encode polyglutamine proteins are alternatively spliced (Banfi et al., 1994, Goto et al., 1997, Paulson et al., 1997, Sahba et al., 1998, Affaitati et al., 2001, Ichikawa et al., 2001, Einum et al., 2003, Hirata et al., 2003, Tadokoro et al., 2005, Tsunemi et al., 2008, Bettencourt et al., 2009, Reid et al., 2009). For example in SCA6, splice variants encoding the polyglutamine domain of the Cav2.1 calcium channel are specifically enriched in Purkinje cells of SCA6 patients but not in controls (Tsunemi et al., 2008). This finding underscores the possibility that alternative splicing contributes to polyglutamine disease pathogenesis by regulating the expression of more “toxic” transcripts in specific cell populations.

Here we explore alternative splicing in the polyglutamine disorder Spinocerebellar ataxia type 3 (SCA3), the most common dominantly inherited ataxia. The disease protein in SCA3, ataxin-3, is a deubiquitinating enzyme. The original ataxin-3 transcript isolated from human brain encoded an isoform that contains a Josephin protease domain, two ubiquitin interacting motifs (UIMs), and the polyglutamine domain, followed by a C-terminal stretch of hydrophobic amino acids (Kawaguchi et al., 1994). Goto and colleagues subsequently isolated a variant that encodes a third UIM at its C-
terminus instead of this hydrophobic tail (Goto et al., 1997). In this report, we term the protein product of these alternative splice variants the 3UIM and 2UIM ataxin-3 isoforms, respectively (Figure 4). The 3UIM isoform is known to be widely expressed in multiple tissues (Schmidt et al., 1998). Additional N-terminal splice variants have also been observed in human and rodent tissues (Paulson et al., 1997, Ichikawa et al., 2001, Bettencourt et al., 2009). Although both 2UIM and 3UIM ataxin-3 are commonly considered “full length” ataxin-3 isoforms, and both have been used in mechanistic studies of SCA3, the impact of this 3’ splice variation on ataxin-3 function and disease pathogenesis has not been adequately examined.

In the current study, we investigate ataxin-3 alternative splicing. We characterize the range of splice variation in human and transgenic murine brain, establishing that while both 3’ splice variants are expressed, 3UIM ataxin-3 is the predominant isoform. We further show that although C-terminal splice isoform variation does not influence ataxin-3’s deubiquitinating activity, it significantly modifies both its tendency to aggregate and its intracellular stability. This observation highlights how splicing events that preserve the polyglutamine domain but alter protein context could influence selective neuronal toxicity.

Materials and Methods

Animal lines

Three murine models of SCA3 were used in this study. MJD15.4 and MJD84.2 (Cemal et al., 2002) are yeast artificial chromosome (YAC) transgenic lines that contain the full human ATXN3 gene with an unexpanded (Q15) or expanded (Q84) repeat. The presence of all genomic elements of the ATXN3 gene allows these mice to exhibit alternative splicing of both transgenic (human) and endogenous (murine) ataxin-3. Both lines possess the Tyr-encoding version of an A/C SNP within the extended portion of exon 10, at the position that encodes the final stop codon seen in the MJD1a isoform
(GenBank accession no. S75313.1), resulting in 2UIM-long isoforms that share an extreme C-terminus (though not all SNPs) with isoform MJD2-1 (Goto et al., 1997). These lines were maintained on a C57BL/6J x CBA/Ca background. Q71-B transgenic mice (Goti et al., 2004) express the human MJD1a splice isoform as a cDNA driven by the prion promoter, and thus only exhibit alternative splicing of endogenous ataxin-3. They were maintained on a C57BL/6J x C3H/HeJ background. Atxn-3 knockout mice (Schmitt et al., 2007) and their littermate controls were used to confirm that ataxin-3 isoforms detected in various tissues by 1H9 mAb were in fact derived from the Atxn3 gene. Knockout mice were maintained on a pure C67BL/6 background. All lines were maintained in accordance with the University of Michigan and University of Iowa AUCUC guidelines, including accepted measures to minimize pain or discomfort.

Constructs and Primers

3UIM (“full length, FL”) and catalytically inactive C14A GST-ataxin-3 fusion proteins were derived from pGEX-6P-1, as previously described (Todi et al., 2007). The 3UIM-encoding pGEX6P1-At3(Q22)FL expression vector has also been previously described as pGEX6P1-ATX3-WT (Winborn et al., 2008). pGEX6P1-At3(Q22)2UIM was derived from this construct by substituting the region downstream of the (CAG) repeat in the 3UIM-encoding construct for the 2UIM-encoding sequence in pEGFP-C1-ataxin-3(Q28) (Chai et al., 2002). Briefly, Q28 ataxin-3 was amplified using MJD.Nter.F#722 (see below) and the primer hMJD2UIM-R1N (5’gcggccgctttatgtcagataaagtg 3’), which creates a novel NotI restriction site, cloned into pCR2.1-TOPO (Invitrogen, Cat # K4500-01), and restriction digested from an endogenous PpuMI site to the novel NotI site to generate the donor 3’ sequence. The N-terminally Flag-tagged ataxin-3 eukaryotic expression vector pFlag-A22-FL-M1G contains the full 3UIM-encoding ataxin-3 sequence in a pFlag-CMV-6a backbone. pFlag-A22-2UIM-M1G was generated by inserting the 2UIM-encoding 3’ region, as
described above. The pFlag-A22-UIM3(SA/DG)-M1G mutant was similarly generated by exchanging regions of pFlag-A22-FL-M1G and pGEX6P1-At3(Q22)UIM3(SA/DG) using endogenous ataxin-3 MfeI and vector-derived NotI sites. All constructs were confirmed by restriction digestion and DNA sequencing.

MJD.Nter.F#722 (5’ ataaacatggagtccatcttc 3’) was the common forward primer used to amplify human and murine ataxin-3 cDNA. This primer targets the junction of the 5’UTR and exon1. The following reverse primers were used to amplify transgenic 5’ splice variants from YAC cDNA: HuMJD.Cter1.R#724 (5’ gtgtcatatcttgagatatg 3’) and HuMJD.Cter2.R#723 (5’ ttctgaagtaagatttgtac 3’) target exon 9 of human ataxin-3 to amplify 5’ variants independently of the documented 3’ variation. The following reverse primers were used to amplify 3’ splice variants from cDNA pools: 2HumExon10R#537 (5’ ctgctcttaatccagg 3’) was used to amplify transgenic and endogenous human 10-exon specific transcripts, 2HumExon11R#536 (5’ cacacggtatacagttgaagg 3’) was used to amplify transgenic and endogenous human 11-exon specific transcripts, MuMJDexon10R#598 (5’ cgagtaaagcatactg 3’) was used to amplify endogenous murine 10-exon specific transcripts, and MuMJDexon11R#597 (5’ ctgactgcctctttggc 3’) was used to amplify endogenous murine 11-exon specific transcripts.

Cell culture and transfection

Cos7 and 293T cells were maintained at 37°C, 5% CO₂ in DMEM, 10% FBS, 1% penicillin/streptomycin. Cells were transiently transfected with 0.2 – 2.0 µg DNA, using Lipofectamine PLUS Reagents (Invitrogen Cat #18324012, #11514-015), as per the manufacturer’s recommendations. Mock transfections were performed without DNA, and vector control transfections were performed with equivalent amounts of empty vector DNA.
Collection of animal tissues

Adult mice were either euthanized with Ketamine/Xylazine (1 mg / 0.1 mg per gram) prior to immediate dissection of brain tissue, or deeply anesthetized with 0.2 mg Ket / 20 µg Xyl per gram prior to transcardial perfusion (2 ml per minute) with 5-10 ml of ice cold sterile PBS followed by dissection of brain and other tissues. Neonatal mice were euthanized by decapitation. Harvested tissue was either homogenized immediately, snap frozen, or prepared for archival storage using RNAlater (Ambion, Cat #AM7020).

Preparation of cDNA from animal tissues and pooled human RNAs

Total RNA was isolated from murine tissues using TRIzol Reagent (Invitrogen, Cat #15596-026), and mature mRNA was extracted from total RNA using the Poly(A)Purist kit (Ambion, Cat #1916) as per the manufacturers’ standard protocols. Adult and fetal human total RNA were purchased from Clontech; the adult RNA (Cat # 636530, Lot 7120601A) was pooled from two neurologically normal male Caucasians, aged 47 – 55, the fetal RNA (Cat # 636526, Lot 7080344) was pooled from 21 third trimester spontaneous abortions, gestational age 26 – 40 weeks. Total cDNA was generated using random hexamer priming of mature mRNA with Superscript III Reverse Transcriptase (Invitrogen, Cat # 18080-085), prior to transcript-specific analyses.

Sequence-specific PCR-amplification of cDNA

All 5’ and 3’ splice variants were amplified from reverse-transcribed cDNA using the following conditions: 94°C, 2 min; (94°C, 30 s; 50°C, 30 s; 72°C, 1 min) x 35 cycles; 72°C, 1-10 min; 4°C, until further analysis. Fifty microliter reactions were carried out with Taq polymerase (Invitrogen, Cat # 10342-053) or Platinum Taq High Fidelity polymerase mixture (Invitrogen, Cat # 11304-011), following the manufacturer’s protocols in the recommended buffer systems, using 2 µl cDNA template.
Preparation of protein from murine tissues and cultured cells

For splice isoform analysis in tissue homogenates, protein was isolated in conjunction with total RNA, using the TRIzol Reagent, as per the manufacturer’s recommendations. For ataxin-3 aggregation analysis and 2D-PAGE, non-denaturing whole brain lysates were prepared by homogenization of 100 mg tissue per milliliter of ice cold RIPA buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) plus Complete Mini Protease Inhibitor Cocktail (Roche, Cat# 11836153001) in a Potter-Elvehjem homogenizer, and then centrifuged at 4000 rpm, 15 min, 4°C to separate supernatant and pellet fractions. Pellet fractions were re-homogenized in an equal volume of RIPA+PI. All RIPA lysates were stored at -80°C. Lysates were diluted in Laemmli buffer (50 mM Tris-Cl, pH 6.8, 2% SDS, 10% glycerol, 0.1% bromophenol blue) plus 100 mM DTT and sonicated 5-10 s for SDS-PAGE followed by Western blot analysis utilizing Western Lightening ECL (Perkin-Elmer Life Sciences, Cat # NEL102001EA) or dual color visualization of IRDye conjugated secondary antibodies (LI-COR Biosciences, Cat # 926-32210 & 926-32221) on an Odyssey IR imaging system. Densitometry for aggregate analysis was carried out using ImageJ software, after background subtraction with a rolling ball radius of 50; soluble and insoluble ataxin-3 levels were normalized to endogenous murine ataxin-3 signal to control for equal protein loading.

For transient transfection experiments, adherent cells were washed once with ice cold PBS, then directly lysed in Laemmli buffer plus 100 mM DTT (440 µl / well for 6-well plates; 200 µl / well for 12-well plates), sonicated for 5-10 s, heated for 3 minutes at 95°C, and centrifuged for 3 minutes at 14,000 rpm prior to SDS-PAGE and Western blot analysis, as above.
2D-Western Blot Analysis

Non-denaturing whole brain RIPA lysates (50 µg total protein) or purified GST-ataxin-3 proteins (5 µg for in gel Coomassie detection, 50 ng for Western analysis) were diluted into FOSB1 (7M urea, 2M thiourea, 1.25% CHAPS 32 mM DTT, 2.5 mM TCEP, 0.5% ASB-14, 0.5% Triton X-100, 0.5% Zwittergent 3-10, 0.3% carrier ampholytes, 0.001% Bromophenol Blue) using 100x BioLytes 3.9 – 5.1 as the carrier ampholytes (BioRad, Cat # 1632098) for optimal narrow range resolution. Diluted samples were used to passively rehydrate 11 cm narrow range pH 3.9 – 5.1 immobilized pH gradient (IPG) strips (BioRad, Cat # 163-2024). Isoelectric focusing was carried out in a PROTEAN IEF cell under the following conditions: Step 1 (0 – 250 V, 15 min, rapid ramp), Step2 (250 – 8000 V, 1 hr, slow ramp), Step 3 (8000 V constant, 30,000 V-hr, rapid ramp); all steps were set to a default temperature of 20°C, and wicks were changed periodically to remove unwanted salts and enhance actual running time. Focused IPG strips were equilibrated, run in the second dimension, and transferred to PVDF membrane using the Criterion Blotter system, as per the manufacturer’s recommendations.

In vitro deubiquitination assays

Recombinant 2UIM, 3UIM, and C14A ataxin-3 were expressed in BL21-A1 E. coli (Invitrogen Cat #C607003) as GST fusion proteins and purified as follows. Overnight cultures were subcultured at 1ml per 100 ml LB + 50 µg/ml Ampicillin until an OD<sub>600</sub> between 0.4 – 0.6. Recombinant protein expression was then induced with 400 µM isopropyl-1-thio-β-D-galactopyranoside for 3 hr at 30°C. Bacteria were lysed by sonication in 0.5x NPG buffer (150 mM NaCl, 25 mM NaH<sub>2</sub>PO<sub>4</sub>, 5 % glycerol, pH 8.0) plus protease inhibitors (1x Sigma Protease Inhibitor Cocktail, 1x Roche Complete mini Protease Inhibitor Cocktail, 2.25 mM PMSF, 0.5 mM Pefabloc SC). Lysates were pre-cleared for 15 minutes in 0.5xNPG-equilibrated PABA-agarose, and recombinant proteins were then bound to equilibrated GST-Sepharose, (GE Healthcare, Cat #27-4570-
03) for 30 minutes, on ice with periodic mixing, washed 5x with PBS plus Roche Complete Mini Protease Inhibitor Cocktail and 1x with PBS only, and then eluted with GSH elution buffer (3 mg/ml reduced glutathione, 10% glycerol, 1 mM DTT). Purified protein was quantified against BSA standards by in-gel Coomassie Brilliant Blue total protein stain. *In vitro* deubiquitination reactions were carried out at 37°C with 1 µM GST-ataxin-3 (3UIM, 2UIM, or C14A) and 250 nM ubiquitin chains in 50 mM HEPES pH 7.5, 500 µM EDTA, 100ng/ml ovalbumin, and 1mM DTT. Reactions were stopped by the addition of 1x Laemmli buffer plus 100mM DTT, and stored on ice until SDS-PAGE analysis. Twelve microliters of each reaction was fractionated on a 5-20% gradient gel and a 15% acrylamide gel, and analyzed by silver stain (Silver Stain Plus, BioRad) and Western blot analysis using P4D1 anti-ubiquitin mAb, respectively.

*In vitro* ubiquitin-AMC assays

Ubiquitin-7-amino-4-methylcoumarin (Ub-AMC, Boston Biochem Cat # U-550), GST-ATXN3(Q22)3UIM, GST-ATXN3(Q22)2UIM, and GST-FBXO2 were diluted to 2x stocks in 50 mM HEPES, 0.5 mM EDTA, 0.1 mg/ml ovalbumin, 1 mM DTT, pH 7.5 and pre-warmed to 37°C for 15 minutes. At time zero, 2x stocks of pre-warmed enzyme or buffer-only control were combined with Ub-AMC to yield 500 nM GST-tagged enzyme and 500 nM Ub-AMC in a final reaction volume of 100 µl. Ub-AMC cleavage at 37°C was detected using a Wallac 1420 multilabel fluorimeter using an excitation of 355 nm and emission at 460 nm, a lamp energy of 5160, and a counting time of 0.1 s using the Normal Aperture, Top Counter setting. Differences in initial reaction velocity were assessed by two-tailed heteroscedastic Student’s t-tests.

Immunostaining of cultured cells

Twenty-four hours before harvest, cells were replated on rat tail collagen-coated coverslips to ensure adequate spacing of transfected cells. Cells were rinsed with ice cold PBS, fixed in 4% paraformaldehyde/PBS, and blocked in 5% normal goat serum in
0.05% TX-100/PBS for at least 30 minutes. Primary antibodies (1H9 mAb 1:500 and Rb anti-Flag pAb (Sigma) 1:150 in 0.05% TX-100/PBS) were applied for one hour at room temperature. Cells were washed 3x in abundant PBST, before incubation for 1 hour in secondary antibodies (1:250 goat anti-mouse AlexaFluor 568 and goat anti-rabbit AlexaFluor 488 (Molecular Probes)). Cells were washed 5x in abundant 0.05% TX-100/PBS, 5 µg/ml DAPI (Sigma cat. D9564) counterstained for 10s, rinsed briefly, and prepared for mounting with the SlowFade Anti-Fade kit (Molecular Probes, Cat #S2828) using component A according to the manufacturer’s recommendations. Coverslips were sealed to glass slides with Permount (Fisher Scientific, Cat # SP15). All incubations and washes were carried out at room temperature, protected from light, unless otherwise noted.

Fixed and stained cells were gated into moderate and high levels of overexpression based on fluorescence intensity at defined exposure times using the Zeiss Axiocam MRGrab software. Moderate overexpression was defined as visibility of α-Flag staining of the cytoplasm and nucleus in the live image capture window with exposure times from 500 – 1000 ms; representative images for these cells were captured with the following exposures (DAPI 200 ms, 1H9 15,000 ms, α-Flag pAb 1000 ms), set to linear display (gamma = 1) with brightness and contrast levels adjusted to span the peaks of the pixel intensity histogram. High overexpression was defined as visibility of α-Flag staining of the cytoplasm and nucleus in the live image capture window at ≤ 500 ms; representative images for these cells were captured with the following exposure times (DAPI 200 ms, 1H9 4500 ms, α-Flag pAb 350 ms), and adjusted as above. All images were pseudocolored and overlayed in Adobe Photoshop without further adjustment. The number of puncta per cell that were brightly positive for both α-Flag and 1H9 immunostaining was counted for the first 25 cells identified in each replicate that met fluorescence gating criteria (or the maximum number of cells to meet criteria, if less than 25), and plotted in histogram form. Statistical differences between aggregation
phenotypes in 2UIM vs. 3UIM ataxin-3, 2UIM vs. UIM3-mutant ataxin-3, and 3UIM vs. UIM3-mutant ataxin-3 were assessed using the Chi-squared test of independence, with frequency data binned into populations containing 0, 1-5, 6-10, or >10 puncta per cell, giving 3 degrees of freedom (df) for each comparison.

Cycloheximide treatment to analyze protein stability

Cos7 cells were transiently transfected with 0.4 µg unexpanded Flag-ataxin-3 constructs one day before treatment with cycloheximide. At time zero cells were either harvested immediately or grown at 37°C, 5% CO₂ in DMEM, 10% FBS, 1% P/S plus or minus 10 µM cycloheximide, 10 µM epoxomycin, and/or 10 mM 3-methyladenine for 10 or 24 hours before harvest in 1X Laemmli buffer plus 100 mM DTT. Lysates were resolved by SDS-PAGE and analyzed by Western blot analysis using rabbit α-Flag antibody followed by Coomassie Brilliant Blue R250 total protein staining of the PVDF membrane. Flag-tagged and total protein were analyzed densitometrically using ImageJ software, after rolling ball background subtraction with a rolling ball radius of 50.

Differences in normalized abundance between constructs at each time point were assessed using a two-tailed Student’s t-test, assuming unequal variance. Differences in abundance of individual constructs under 24 hr degradation-rescue conditions compared to 0 hr and 24 hr + cycloheximide conditions were assessed using paired one-tailed Student’s t-tests (assuming a rescue value at 24 hours less than or equal to that at time zero, and a value greater than or equal to that at 24 hr without pharmacological rescue).

Open Field Analysis

Eighteen to 23 month old wild type or hemizygous MJD15.4, MJD84.2, and Q71B mice were acclimated to the behavior suite for at least 30 minutes prior to analysis in the open field apparatus. Each mouse was placed in a 25 cm x 25 cm enclosure. Spontaneous exploratory behavior was recorded for 1.5 hours with the VideoTrack recording system (View-point, Lyon, France), during which large and small movements
were recorded in centimeters traveled during each 5 minute interval of the recording. Simultaneously, the program plotted movement traces to document the area of the enclosure covered over this interval. Differences in the mean total distance traveled over 1.5 hours between hemizygous animals and wild type controls within each genotype were assessed by one-tailed heteroscedastic Student’s t-tests. Because we noticed a marked correlation between excessive body mass and hypoactivity during open field behavioral testing, any animals weighing more that two standard deviations above (or below) the Jackson Laboratory Mouse Phenome Database reference mean for 18 month old mice (Table 5) were excluded from the final analysis. Note that these means were calculated for 27 strains of male mice and 29 strains of female mice, including substrains of each of the parental strains of the YAC and Q71B lines (C57BL6, CBA, and C3H).

Gait Analysis

The gait analysis apparatus consists of a 90 cm paper-lined wall-enclosed course with an open topped entry chamber at one end and a covered escape chamber at the other. The forepaws and hindpaws of 18 - 23 month old wild type or hemizygous MJD84.2 and Q71B mice were painted with nontoxic water based paints (blue and black, respectively) immediately before each mouse was placed into the entry chamber. The mouse was then allowed to walk the length of the course to the escape chamber, and then allowed to remain in the escape chamber for at least one minute while the course was reset. Each animal was subjected to three trials.

Results

Ataxin-3 is alternatively spliced in ATXN3 YAC transgenic and human brain

Because most observations of ataxin-3 alternative splicing have been made in peripherally-derived, non-neuronal mRNA, we wanted to confirm the presence of
alternative splice variants in the brain. In addition to analyzing mRNA from pooled human brain tissue, we isolated and characterized mature mRNA from the brains of ataxin-3 YAC transgenic mice, which contain the full human ATXN3 gene and are an ideal model in which to examine alternative splicing of both transgenic human and endogenous murine ataxin-3 transcripts. To characterize 5’ splicing of ATXN3 mRNA independent of 3’ variation, we PCR amplified brain-derived YAC cDNA using primers targeting the 5’UTR/exon1 junction and exon 9 of the human ATXN3 transcript. We detected, cloned, and sequenced multiple minor splice variants from perinatal (data not shown) and adult and murine brain (Figure 6). Two of the 5’ variants we identified contain frameshift-induced stop codons upstream of multiple exon junction complexes, and thus are strong candidates for nonsense mediated decay (NMD). The remaining minor variants we identified are not likely to be NMD candidates but do excise at least one exon encoding a Josephin domain catalytic residue (indicated by asterisks). Accordingly, they are not predicted to encode functional deubiquitinating enzymes (DUBs).

We also evaluated the presence of 3’ splice variation among ATXN3 (human) and Atxn3 (murine) transcripts, using species-specific and sequence-specific reverse primers that selectively target either the 2UIM-encoding 10-exon or the 3UIM-encoding 11-exon variant (Figure 7A). Both endogenous 3’ Atxn3 variants were detected in nontransgenic and MJD15.4 hemizygous transgenic mice (Figure 7B). This mirrored the expression pattern of human ATXN3 variants in nonexpanded MJD15.4 transgenic mice (Figure 7B), CAG repeat-expanded MJD84.2 transgenic mice, and cDNA from pooled adult or fetal human brain tissue (Figure 7C). Although both 3’ variants were consistently detectable from whole brain-derived cDNA sources, non-quantitative amplification of 11-exon transcripts consistently produced a much more robust signal than did 10-exon transcripts, independent of CAG repeat length or species of origin.
3UIM ataxin-3 is the predominant protein isoform in murine and human brain tissue

Neuronal toxicity in polyglutamine diseases is thought to be mediated primarily by the disease protein. Therefore, it was important that we validate our observations of ATXN3 splice variation at the protein level. To assess the presence and relative abundance of 2UIM and 3UIM ataxin-3 isoforms, we carried out comparative standard and two-dimensional Western blot analyses. The monoclonal antibody 1H9 recognizes an epitope encoded by both 3' splice variants of human ATXN3 and murine Atxn3, whereas ataxin-3C polyclonal antibody specifically recognizes only human 3UIM ataxin-3 (Schmidt et al., 1998), as represented schematically in Figure 9A. First, to assess the range of tissue-specific isoform variation, we compared endogenous murine ataxin-3 expression in various tissues including forebrain, midbrain plus hindbrain, heart, kidney, liver, skeletal muscle, and spleen in wild type versus Atxn3 knockout animals (Figure 8). Endogenous ataxin-3 protein bands vary in apparent molecular weight between murine tissues, consistent with the expression of tissue-specific splice variants, posttranslational proteolysis, or a combination of both. The kidney and spleen in particular reveal prominent lower molecular weight isoforms distinct from the predicted 40.5 kDa 3UIM ataxin-3 protein. In contrast, a single predominant isoform consistent with full length ataxin-3 is present in brain tissue. The predominant ATXN3 bands recognized by 1H9 antibody in standard Western blot analysis of MJD15.4 (Figure 9B), MJD84.2 (Figure 9C), and human (Figure 10) brain lysates are also recognized by the 3UIM-specific antibody, ataxin-3C pAb. Although this result confirms that the 3UIM isoform of transgenic ataxin-3 is present in brain tissue from all of these sources, it is not conclusive evidence of relative abundance without further analysis, as included below.

Multiple synonymous, non-synonymous, and non-coding single nucleotide polymorphisms have been documented in the ATXN3 gene (Goto et al., 1997, do Carmo Costa et al., 2002). Both MJD15.4 and MJD84.2 lines were constructed using YACs that
possess a non-synonymous tyrosine-encoding SNP rather than the stop codon seen in the MJD1a ataxin-3 isoform, which results in production of a slightly longer 2UIM isoform from 10-exon transcripts (termed 2UIM-long; data not shown). As a result, unfortunately, the difference in predicted molecular weight between 2UIM-long and 3UIM ataxin-3 isoforms is only 0.5 kDa in both unexpanded (41.4 kDa vs. 41.9 kDa) and expanded (50.2 kDa vs. 50.7 kDa) transgenic lines (Table 4). To rule out the possibility that we might fail to detect a 2UIM ataxin-3 signal because it is obscured by its close proximity to the 3UIM isoform, we took advantage of the difference in isoelectric point (pI) between these isoforms to assess the relative abundance of 2UIM and 3UIM isoforms. We resolved whole brain lysates by 2D-PAGE followed by Western blot analysis with ataxin-3 specific antibodies (Figure 11A). A strong signal was detected in the expected pH range for unmodified 3UIM ataxin-3 (Table 4) with both ataxin-3C and 1H9 antibodies. Additional spots consistent with multiple phosphorylation events were also observed (slight increase in apparent molecular weight with an acidic shift). Additionally, 1H9 mAb exclusively detected spots consistent with endogenous murine 3UIM ataxin-3 (pI 4.69, MW of 40.5 kDa), mono-ubiquitinated 3UIM ataxin-3 (pI 4.81 MW ~ 8 kDa greater than the major transprotein and detectable only in the abundantly expressing MJD15.4 brain), and putative ataxin-3 splice isoforms or degradation products (apparent MW of < 37 kDa). Surprisingly, despite consistently detecting the 2UIM-encoding mRNA in brain, we were unable to detect any unmodified 2UIM transprotein (pI 4.81, MW 41.9) in whole brain lysates. We were similarly unable to detect 2UIM ataxin-3 in cerebellar lysates from MJD84.2 mice (data not shown), arguing against an enrichment of 2UIM ataxin-3 expression in this brain region, which is known to be disproportionately affected in human disease and in this mouse model. To rule out the possibility that the absence of 2UIM signal is an artifact (due to preferential insolubility of this isoform, to self-association, and precipitation during the IPG strip equilibration step, or to a lack of recognition by the 1H9 mAb), we also analyzed purified 2UIM and
3UIM GST-ataxin-3 by 2D-Western (Figure 11B). 2UIM and 3UIM ataxin-3 fusion proteins were similarly solubilized, focused, resolved in the second dimension, and detected by Western blot analysis, with 2UIM ataxin-3 demonstrating the anticipated basic shift compared to the 3UIM isoform. Similar results were obtained with GST-ataxin-3 isoforms resolved by 2D-PAGE followed by in-gel Coomassie staining (data not shown). Thus, the absence of detectable 2UIM ataxin-3 in brain lysates is not artifactual. We conclude that 3UIM ataxin-3 protein is the predominant isoform in the central nervous system, while 2UIM ataxin-3 is expressed at extremely low levels, is expressed only in a highly restricted subpopulation of cells within the CNS, or is posttranslationally modified so as to be undetectable at the anticipated pI/MW.

2UIM and 3UIM ataxin-3 display similar in vitro DUB activity against defined ubiquitin chains

Selective neuronal toxicity may result from an expanded polyglutamine-dependent, deleterious gain of function combined with a partial loss of function due to impaired activity of the endogenous protein (Matsuyama et al., 2004, Lim et al., 2008). Recent studies have shed light on the normal biological activity of ataxin-3. Ataxin-3 is a member of the Josephin family of deubiquitinating enzymes (DUBs). In a Drosophila model, full length ataxin-3 suppresses the toxicity of expanded ataxin-3 and other polyglutamine proteins. This protective ability requires the catalytic activity of the Josephin Domain and, to a lesser extent, intact UIMs 1 and 2 (Warrick et al., 2005). In vitro, 3UIM ataxin-3 binds K48-linked, K63-linked, or mixed linkage chains containing at least 4 ubiquitin molecules, and preferentially cleaves longer ubiquitin chains and K63 linkages within mixed linkage chains. These activities of ataxin-3 are UIM-dependent, as high affinity ubiquitin binding and cleavage specificity are lost when all three UIMs are mutated (Winborn et al., 2008). Intriguingly, while UIMs 1 and 2 are required for high affinity binding to ubiquitin chains, UIM3 is dispensable for this property and for the
ability of ataxin-3 to cleave ubiquitin-aldehyde (Burnett et al., 2003). Additionally, both
2UIM and 3UIM ataxin-3 can bind to polyubiquitinated proteins in cells (Berke et al.,
2005). The capacity of UIM3 to modulate the specificity of ataxin-3 DUB activity,
however, has not been adequately assessed.

To test whether replacing UIM3 with the hydrophobic tail of the 2UIM isoform
alters the specificity of ataxin-3 DUB activity, we incubated 1 µM purified recombinant
GST-ataxin-3(Q22)3UIM, GST-ataxin-3(Q22)2UIM, or catalytically inactive GST-
ataxin-3(Q22)C14A with 250 nM defined ubiquitin chains at 37°C in vitro. DUB activity
towards K48-hexaubiquitin, K63-tetraubiquitin, and K48-K63-K48 mixed linkage
tetraubiquitin was compared by Western blot (Figure 12); simultaneous silver stain
confirmed that equivalent levels of GST-ataxin-3 isoforms were used (data not shown).
Ataxin-3(Q22)2UIM showed the same cleavage activity as GST-ataxin-3(Q22)3UIM:
limited cleavage of K48-linked Ub chains, more robust cleavage of K63 linked Ub
residues and mixed linkage chains, and vigorous cleavage of the higher molecular weight
Ub chains that likely represent dimers of Ub4 or Ub6. As expected, catalytically inactive
ataxin-3(C14A) demonstrated no DUB activity. While these data do not support the
hypothesis that C-terminal splice variation alters the DUB activities of ataxin-3, they are
consistent with previous reports showing the greater importance of UIMs 1 and 2 for
ubiquitin-related activities of this protein, at least in a defined in vitro system (Burnett et
al., 2003, Berke et al., 2005).

To provide a more quantitative analysis of 2UIM and 3UIM ataxin-3 enzymatic
function, we utilized the fluorogenic substrate Ub-AMC, which emits fluorescence at 460
nm upon cleavage of a monoubiquitin from the 7-amino-4-methylcoumarin moiety. To
assess the relative ability of 2UIM and 3UIM ataxin-3 to cleave Ub-AMC, we incubated
500 nM of GST-ATXN3(Q22)3UIM or GST-ATXN3(Q22)2UIM with 500 nM of Ub-
AMC. GST-FBXO2 (a recombinant GST-fusion protein that is not a DUB) and a buffer
only control lacking any GST-fusion protein were used as negative controls. 2UIM and
3UIM ataxin-3 each cleaved Ub-AMC (Figure 13A). Moreover, there was no significant difference in initial reaction velocity between 2UIM and 3UIM DUB reactions (p > 0.4 by a 2 tailed heteroscedastic Student’s t-test). Ub-AMC was not cleaved by buffer alone or GST-FBXO2 (Figure 13B).

2UIM ataxin-3 has a higher propensity to aggregate than 3UIM ataxin-3

While visible aggregates, inclusions, and aggresomes may not directly cause polyglutamine-induced cytotoxicity, these pathophysiological hallmarks remain useful in identifying cells that are subject to a high burden of misfolded proteins. More subtle facets of intracellular polyglutamine protein behavior have been implicated in SCA3 toxicity, including alterations in subcellular localization (Bichelmeier et al., 2007), protein-protein interactions (Lim et al., 2008), and the formation of microaggregates (Williams et al., 2009). Thus, immunocytochemical analyses of expanded polyglutamine expressing cells remain useful as adjuncts to biochemical methods of assessing protein misfolding and cytotoxicity. To explore the differential behavior of 2UIM ataxin-3 and 3UIM ataxin-3 isoforms, we generated an N-terminal Flag-tagged ataxin-3(Q22)2UIM expression vector to complement our existing Flag-At3(Q22)3UIM construct. In addition, we generated a Flag-tagged ataxin-3 expression vector that encodes two mutations (A -> G, S -> D) in UIM3, which should abolish ubiquitin interactions with this UIM (Fisher et al., 2003, Todi et al., 2009). In our comparative studies of the 2UIM and 3UIM isoforms, this UIM3 mutant allows us to distinguish potential UIM3-specific effects from potential UIM-independent effects of substituting a hydrophobic domain (the C-terminus of 2UIM variant) for the largely hydrophilic sequence in UIM3. We transiently expressed these engineered forms of ataxin-3 in Cos7 cells (Figure 14) or HEK293T cells (data not shown) for 48 hours. Expression levels were confirmed by Western blot analysis with Flag and ataxin-3 antibodies (data not shown), and ataxin-3
subcellular localization was determined by immunofluorescence. 2UIM ataxin-3 expressing cells displayed moderately robust aggregation that was not observed in cells expressing 3UIM ataxin-3 or UIM3-mutant ataxin-3. To quantify this, cells were gated by fluorescence intensity into populations of moderate or high overexpressors and the number of immunopositive puncta per cell was counted for each ataxin-3 isoform (Figure 14B). 2UIM ataxin-3 expressing cells exhibited significantly higher aggregate formation than did 3UIM ataxin-3 and UIM3-mutant ataxin-3 expressing cells. This difference was present in moderately overexpressing cells ($\chi^2 = 52.6$ and 22.5, respectively, df = 3, p < 0.0001) and was even more pronounced in highly overexpressing cells ($\chi^2 = 75.4$ and 54.3, respectively, df = 3, p < $1 \times 10^{-11}$). There was no significant difference in aggregation behavior between 3UIM and UIM3-mutant ataxin-3, whether in moderately overexpressing or highly overexpressing cells ($\chi^2 = 5.56$, df = 3, p = 0.12 for moderate; $\chi^2 = 1.92$, df = 3, p = 0.59 for high overexpressors); however UIM3-mutant ataxin-3 expressing cells did exhibit some differences in cell morphology and abundance of cytoskeletal proteins (data not shown). Puncta in 2UIM ataxin-3 expressing cells generally exhibited a nuclear or nucleocytoplasmic distribution, whereas 3UIM and UIM3-mutant puncta were often exclusively in the cytoplasm. In addition, puncta formed by 2UIM ataxin-3 were qualitatively larger and more irregular. Thus in transiently transfected cells, 2UIM ataxin-3 confers a UIM3-independent aggregation phenotype, even with a non-expanded polyglutamine domain, which does not occur with the 3UIM ataxin-3 isoform.

To address the possibility that the increased aggregation of 2UIM ataxin-3 might simply reflect transient overexpression in nonneuronal cells, we compared levels of SDS-insoluble ataxin-3 aggregates in whole brain lysates from aged MJD84.2 mice (which express primarily expanded 3UIM ataxin-3 transprotein) and Q71B mice (which express expanded 2UIM ataxin-3 transprotein) (Figure 15). Despite expressing different predominant isoforms, hemizygous mice in both models exhibit a modest motor
phenotype at 18 – 23 months of age, demonstrating non-significant trends toward hypoactivity during open field analysis (Figure 16A and 16B), modest decreases in weight (Figure 16C), gait abnormalities (Figure 16D), and transient clasping behavior upon tail suspension (data not shown). During SDS-PAGE, insoluble ataxin-3 aggregates are retained at the base of wells and in the stacking gel, whereas soluble (nonaggregated) human ataxin-3 transprotein and endogenous murine ataxin-3 electrophorese within the resolving gel. Because polyglutamine protein misfolding and aggregation increase with longer repeats, ataxin-3(Q_{84}) in MJD84.2 mice would be expected to aggregate at least as readily as ataxin-3(Q_{71}) expressed in Q71 B mice, if 2UIM and 3UIM ataxin-3 behave similarly in vivo. Yet at 18 months of age, 3UIM-predominant MJD84.2 YAC mice show a significantly lower ratio of SDS-insoluble to soluble ataxin-3 than do age-matched 2UIM-only Q71B mice (p < 0.0005 by a 1 tailed heteroscedastic Student’s t-test) (Figure 15B). In summary, consistent with the enhanced aggregation we observe with 2UIM ataxin-3 in transiently expressing cells, we observe more ataxin-3 aggregation in brain tissue from Q71B mice, which express only 2UIM ataxin-3, than in brain tissue from MJD84.2 mice, which express primarily 3UIM ataxin-3, despite the smaller polyglutamine expansion in Q71B mice.

2UIM ataxin-3 is a less stable protein than 3UIM ataxin-3 and is subject to rapid proteasomal degradation

In cell culture experiments, we observed that transient transfection with equivalent amounts of 2UIM and 3UIM ataxin-3 expression vectors consistently yielded lower amounts of 2UIM protein. This observation together with the enhanced aggregation of 2UIM ataxin-3, prompted us to use cycloheximide treatment to explore the relative stabilities of 2UIM and 3UIM ataxin-3 (Figure 17A and 17B). Cells were transiently transfected 1 day before pharmacological treatment with one or more agents: 10 μM cycloheximide (CHX) to inhibit new protein synthesis, plus or minus 10 μM
epoxomycin (Epox) to inhibit proteasomal degradation, or 10 mM 3-methyladenine (3-MA) to inhibit macroautophagy. At 0, 10, and 24 hours after cycloheximide addition, cells were harvested for Western blot analysis with rabbit α-Flag antibody. Densitometry was used to quantify differences in the rate of protein degradation, and at each time point the Flag ataxin-3 signal, normalized to total protein, was expressed as a percentage of the normalized Flag ataxin-3 signal at time zero. In the presence of cycloheximide, 2UIM ataxin-3 levels decreased much more rapidly than did 3UIM ataxin-3, regardless of whether 3UIM ataxin-3 had a functionally intact UIM3 (*p < 0.02 by a two-tailed heteroscedastic Student’s t-test).

Epoxomycin and 3-MA were used to evaluate whether the proteasome, macroautophagy, or both contributed to the degradation of 2UIM versus 3UIM ataxin-3 isoforms (Figure 18). Rapid degradation of 2UIM ataxin-3 was nearly completely prevented by proteasomal inhibition (n.s. vs. t = 0 +CHX; p < 0.01 vs. t = 24 +CHX), and only slightly by inhibition of macroautophagy (p < 0.01 vs. t = 0 +CHX; p < 0.05 vs. t = 24 +CHX). In contrast, proteasomal inhibition did not significantly prevent the degradation of 3UIM or UIM3-mutant ataxin-3 (p < 0.05 vs. t = 0 +CHX; n.s. vs. t = 24 +CHX), whereas inhibition of macroautophagy appeared to have some effect. While the degradation of UIM3-mutant ataxin-3 was significantly prevented by 3-MA (n.s. vs. t = 0 +CHX; p < 0.05 vs. t = 24 +CHX), the prevention of 3UIM ataxin-3 degradation only approached statistical significance (t = 0 +CHX vs. t = 24 +CHX +3-MA cannot be assessed by a one-tailed analysis, as the mean trends higher than t =0, and is n.s. by a two-tailed analysis; t = 24 +CHX vs. t = 24 +CHX +3-MA approaches significance with p = 0.057). We were initially surprised to see consistently lower levels of 3UIM and UIM3-mutant ataxin-3 in the presence of cycloheximide plus epoxomycin compared to cycloheximide alone; however in the context of the results observed in the other treatment groups, a logical explanation exists. The combination of increased levels at t = 24 +CHX+3MA and decreased levels at t = 24 +CHX+Epox relative to t = 24 +CHX
only, is consistent with reports that epoxomycin treatment may also induce autophagy (Yang et al., 2009). In summary, the results of these measurements of ataxin-3 stability are consistent with the transient expression and aggregation data. While 3UIM and UIM3-mutant ataxin-3 appear to be relatively stable proteins with a low turnover rate (primarily via autophagy) 2UIM ataxin-3 is a highly unstable protein that is prone to misfolding and subsequent rapid degradation, primarily by the proteasome.

Discussion

In this study we confirmed the presence of multiple rare splice variants in murine brain and putative splice isoforms in other tissues. Recently, a study of 3’ ATXN3 alternative splicing in peripheral blood leukocytes (PBL) from SCA3 patients and normal controls identified 56 distinct alternatively spliced transcripts, several of which have been previously reported in the literature or predicted bioinformatically (Bettencourt et al., 2009). While we did not detect most of these transcripts, all five minor 5’ splice variants identified by us in MJD15.4 transgenic mouse brain were confirmed to be present in non-neuronal human PBLs. Moreover, at the protein level we found that the range of putative splice isoforms detectable in leukocyte-rich splenic tissues was greater than in brain, which appears to express a single full length isoform predominantly. Although it is possible that an ataxin-3 isoform encoded by one of these variants demonstrates clinically relevant enhancement of polyglutamine toxicity, which is masked due to the relatively short lives of white blood cells compared to terminally differentiated neuronal populations, the detection of these transcripts in peripheral tissues suggests a less pathologically significant, more ubiquitous function for these splicing events. Similar to the minor splice variants we identified in YAC transgenic brain, many PBL-derived variants are strong candidates for NMD because they contain frameshift-induced stop codons at least 50 bases upstream of exon junction complexes (Maquat, 2005, Silva and Romao, 2009). Of the remaining variants, Bettencourt et al. (2009) observed that most
lacked exons encoding one or more catalytic residues or structurally important features of the Josephin Domain. Thus, such isoforms cannot be active DUBs. Unlike our study, Bettencourt et al. also identified five poor candidates for NMD which encode severely truncated Josephin Domains followed by frameshift-induced polyalanine repeats. The apparent absence of these sequences in brain tissue is not consistent with a model of splicing-induced polyalanine toxicity. PBLs (Bettencourt et al., 2009) and lymphocyte-rich splenic tissues contain a greater number of splice variants and putative low molecular weight splice isoforms than does brain tissue; however, because all of the brain-derived 5’ ATXN3 splice variants we identified were also seen in these unaffected tissues, they are unlikely to explain the selective neuronal toxicity observed in SCA3.

While we provide evidence that both previously described 3’ splice variants (2UIM-encoding and 3UIM-encoding) are detectable in brain at the mRNA level, 3UIM ataxin-3 is clearly the predominant, physiologically relevant C-terminal splice isoform in brain tissue. Although an explicit function of the third UIM of ataxin-3 has been elusive, particularly in vitro, it is highly conserved in mammals and even in *Xenopus*. Recent studies have implicated UIM-3 in protein-protein interactions and in casein kinase 2-dependent ataxin-3 phosphorylation events (Mueller et al., 2009), a modification shown to influence the nucleocytoplasmic shuttling and intranuclear aggregation of ataxin-3. Consistent with these observations, on 2D-Western blots employing a UIM3-specific antibody we observed at least two spots consistent with phosphorylation events. Together, these data illustrate the importance of studying UIM3 modulation of ataxin-3 function within a cellular context. They also illustrate the need for researchers to be careful when choosing among murine models when attempting to identify polyglutamine-mediated effects on ataxin-3 or study SCA3 disease pathogenesis. YAC or BAC transgenic models would likely generate the full array of splice variants for the study of human ataxin-3, whereas existing cDNA transgenic mice only express single variants. A
knock-in model would also be suitable for the study of polyglutamine effects on endogenous murine ataxin-3.

The innate instability of 2UIM ataxin-3 and its propensity to aggregate suggest that while the 2UIM isoform is perhaps less physiologically relevant than the 3UIM isoform, 2UIM ataxin-3 may better facilitate in vitro studies of ataxin-3 fibrillation and aggregation or high throughput screening assays based on an aggregation phenotype. The fact that 2UIM ataxin-3’s propensity to aggregate is phenotypically distinct from a loss of UIM3 function favors a model in which hydrophobic protein-protein interactions flanking the polyglutamine domain significantly affect protein conformation and aggregate nucleation kinetics (Figure 19). This is not an unprecedented idea, given existing nucleation/fibrillization data both for ataxin-3 lacking a glutamine repeat and for the isolated Josephin domain (Ellisdon et al., 2006). In some circumstances, such as defined systems assessing the in vitro behavior of ataxin-3 as a DUB, 2UIM ataxin-3 and 3UIM ataxin-3 behave nearly identically and will produce similar results. Nevertheless, our results clearly demonstrate that the 2UIM and 3UIM isoforms of ataxin-3 are structurally distinct entities that behave differently within the cell, and should not be used interchangeably as “full length” ataxin-3 constructs.

Variations in alternative splicing are emerging as modifiers of numerous complex diseases. Previously ignored silent mutations and non-coding single nucleotide polymorphisms have been associated with aberrations in alternative splicing, some of which directly result in human disease (Tazi et al., 2009); no doubt scientists will continue to recognize more subtle effects of splice variation in modulating certain traits and cellular environments. Because most of the polyglutamine disease genes are alternatively spliced, and a subset of these splice variants have been shown to be enriched in affected neural tissues (Einum et al., 2003, Tsunemi et al., 2008), it is likely that select alternative splicing events also have an impact on polyglutamine-induced neurotoxicity.
Differences in protein context could alter polyglutamine toxicity in numerous ways. Increased conformational instability in the amino acid sequence flanking the polyglutamine stretch could lead to increased rates of protein misfolding and oligomerization, thereby enhancing toxic gain of function effects. Reid and colleagues have identified an aggregation-prone splice isoform of TBP, the disease protein in SCA17, which appears to be enriched in its soluble form in the brains of Alzheimer disease (AD) and Huntington disease (HD) patients (Reid et al., 2009). Whether this isoform plays a significant role in SCA17 pathology or is merely a marker of dysregulated splicing in AD and HD has yet to be determined. Changes in protein context can also alter subcellular localization, and functional nuclear localization signals outside of the polyglutamine domain have in fact been shown to increase polyglutamine toxicity (Perez et al., 1998, Bichelmeier et al., 2007). Protein context also determines interacting partners, which may protect against protein misfolding or become dysregulated in the presence of polyglutamine expansions. This type of polyglutamine-dependent alteration of physiological protein interactions is typified by the superimposed loss of endogenous function and toxic gain of function observed with ataxin-1 and Capicua-containing protein complexes (Lim et al., 2008). Similarly, more structured polyglutamine proteins, such as the voltage-gated calcium channel (Cav2.1) implicated in SCA6, may be more sensitive to deleterious disruptions of endogenous function (Matsuyama et al., 2004). Alternative splicing and other putative cell type-specific aspects of protein context are key candidate determinants of selective neuronal toxicity. Thus, in addition to identifying splice variants that include or exclude polyglutamine encoding exons, as in SCA6, it will be important to identify any splice variants that alter physiological functions and/or protein-protein interactions of polyglutamine disease proteins (leading to superimposed loss of function), or that affect polyglutamine expansion-induced properties, including protein misfolding and aggregation (leading to an increased gain of toxic function). In summary, we have demonstrated that alternative
splicing of sequences distinct from the polyglutamine-encoding trinucleotide repeat can produce protein isoforms with differences in disease-associated phenotypes. In one murine model of SCA3, the forced expression of a minor ataxin-3 variant results in the production of an unstable, aggregate-prone protein, which forms increased SDS-insoluble aggregates \textit{in vivo}. An overrepresentation of this type of transcript in a specific neuronal population could have a significant impact on the pattern of selective neuronal toxicity conferred by an expanded polyglutamine repeat.
Table 4. Predicted molecular weights and isoelectric points of ataxin-3 isoforms.

<table>
<thead>
<tr>
<th></th>
<th>MW (kDa)/pI</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Endogenous</td>
<td>MJD15.4</td>
<td>MJD84.2</td>
</tr>
<tr>
<td>3UIM</td>
<td>40.5 / 4.69</td>
<td>41.4 / 4.69</td>
<td>50.2 / 4.72</td>
</tr>
<tr>
<td>2UIM</td>
<td>37.9 / 4.68</td>
<td>41.9 / 4.81</td>
<td>50.7 / 4.86</td>
</tr>
</tbody>
</table>

Note: The predicted molecular weights (MW) in kilodaltons (kDa) and isoelectric points (pI) of various ataxin-3 isoforms are listed. While murine 2UIM and 3UIM ataxin-3 isoforms are sufficiently different in molecular weight to distinguish by standard SDS-PAGE, the transgenic isoforms of ataxin-3 expressed in the MJD15.4 and MJD84.2 lines are too similar to definitively resolve by one dimensional SDS-PAGE. 2D Westerns were used to rule out the possibility that a detectable 2UIM ataxin-3 signal is obscured by its close proximity to the 3UIM isoform, based on marked differences in pI between the splice isoforms in each model system.
Table 5. Mass and exclusion criteria for open field analysis.

<table>
<thead>
<tr>
<th></th>
<th>Values from Project Yuan1</th>
<th>Open Field Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (g)</td>
<td>s.d.</td>
</tr>
<tr>
<td>male</td>
<td>31.1</td>
<td>8.41</td>
</tr>
<tr>
<td>female</td>
<td>27.2</td>
<td>8.69</td>
</tr>
</tbody>
</table>


Note: Compared to reference body weights for age-matched mice available from the Jackson Laboratories Mouse Phenome Database, our cohort of 18 – 23 month old mice spanned the range of weights previously observed across a wide variety of strains (Low and High values from Project Yuan1 vs. our open field cohort). During our open field analysis, we noted a marked hypoactivity in obese mice within and between genotype cohorts, which was more pronounced than any genotype-specific effects. We therefore excluded any mice that weighed more than two standard deviations above (or below) the reference mean observed in the Project Yuan1 dataset from our post hoc open field quantitative analysis.
Figure 6. **Minor Atn3 5' splice variants are present in murine brain tissue.**
Diagram showing 5' ataxin-3 splice variants identified and confirmed by sequencing. Multiple variants are detectable in mature mRNA from adult murine brain by RT-PCR, using primers targeting the 5'UTR/exon 1 junction and exon 9 (arrows). All identified splice variants that maintain the open reading frame eliminate at least one Josephin domain catalytic triad residue, and thus are not likely to encode an active DUB. Darkly shaded areas in the bottom two variants are downstream of a frameshift-induced stop codon. Exon numbering is shown, with (*) indicating an exon encoding a catalytic triad residue. Josephin domain-encoding sequences (grey), and UIM-encoding sequences (light grey) are as indicated.
Figure 7. Both 3' splice variants are expressed in ATXN3 YAC transgenic and human brain tissue.

Endogenous Atxn3 and transgenic ATXN3 3 ‘full length’ splice variants were amplified by RT-PCR using species-specific (human, hum; murine, ms) and sequence-specific (10-exon 2UIM-encoding, 10; 11-exon 3UIM-encoding, 11) primers. 10-exon and 11-exon variants are both detectable in mature mRNA derived from all sources examined: (B) endogenous Atxn3 from all murine samples and unexpanded ATXN3 from MJD15.4(+/−) brain; and (C) expanded ATXN3 from MJD84.2(+/−) brain, and unexpanded ATXN3 from pooled human brain tissue (hum). Perinatal day 1-3 (P), adult (A), or fetal (F) sources were used, as indicated.
Figure 8. Distinct tissue-specific ataxin-3 isoforms suggest Atxn3 transcripts may be differentially spliced in mouse tissues.
Representative Western blot of wild type or Atxn3 knockout mouse tissue lysates, probed for endogenous murine ataxin-3 and GAPDH. Tissues include forebrain (FB), midbrain and hindbrain (M+H), heart (Ht), kidney (Kid), liver (Liv), skeletal muscle (SkM), and spleen (Spl). Although various putative tissue-specific splice isoforms exist, there is only one predominant isoform in brain tissue. These results were confirmed in three independent littermate pairs of wt and knockout mice.
Figure 9. 3UIM ataxin-3 is the predominant transgenic isoform in YAC transgenic brain tissue.

(A) Diagram of 2UIM (upper) and 3UIM (lower) ataxin-3 variants showing epitope specificity for 1H9 mAb, which recognizes both isoforms, and α-ataxin-3C, which recognizes only 3UIM ataxin-3. (B-C) The major isoform (arrow) of human ataxin-3 is recognized by the 3UIM-specific antibody (3C) and 1H9 in brain tissue from hemizygous transgenic (+/-) MJD15.4 (B) and MJD84.2 mice (C), whereas endogenous ataxin-3 (arrowhead) is recognized by 1H9 only in hemizygous transgenic mice and wild type (-/-) controls. Perinatal day 1-3 (P), adult (A), and non-specific 3C signal (*), as shown.
Figure 10. **3UIM ataxin-3 is the major ataxin-3 isoform in human brain.** Both 1H9 and 3C antibodies recognize the predominant non-expanded ataxin-3 isoforms (brackets) in healthy controls and SCA3 patients (S01-017 and LaLa), as well as the predominant expanded isoform in SCA3 patients (bold arrows). Lower molecular weight bands (bars) are preferentially recognized by 1H9; cortex (ctx), cerebellum (cb), putamen (p), caudate (cd) sources, as indicated.
Figure 11. Unmodified 2UIM ataxin-3 is not detectable in ataxin-3 YAC transgenic brain.

2D-Western blot analysis was used to distinguish 2UIM from 3UIM ataxin-3 protein. IPG range and predicted isoelectric points (pI) of each isoform are shown. (A,B) In brain lysates of (A) MJD15.4 or (B) MJD84.2 YAC transgenic mice, 1H9 recognizes multiple species including endogenous murine ataxin-3 (arrowhead) and 3UIM ataxin-3 transprotein (arrow), but does not detect any 2UIM ataxin-3, which would be 1H9-positive, 3C-negative, with a MW shift 0.5 kDa higher than 3UIM ataxin-3 and a basic shift in pI, as indicated. The prominent band detected by 3C (*) is nonspecific. (C) 2D-Western of 50 ng purified recombinant GST-tagged ataxin-3 isoforms shows that 2UIM GST-ataxin-3 is detected as readily as 3UIM GST-ataxin-3.
Figure 12. 2UIM and 3UIM ataxin-3 display similar DUB activity against defined ubiquitin chains \textit{in vitro}.

Recombinant GST-ataxin-3 (3UIM or 2UIM) can cleave K48-linked hexaubiquitin (A), K63-linked tetraubiquitin (B), and mixed linkage K48-K63-K48 tetraubiquitin (C) chains, whereas catalytically inactive GST-ataxin-3 (C14A mutant) cannot.
Figure 13. Recombinant 2UIM and 3UIM GST-ATXN3(Q22) cleave Ub-AMC at a similar rate.

(A) Ub-AMC reaction curves. Both 3UIM and 2UIM ataxin-3 are able to cleave Ub-AMC, while reactions with either an unrelated control protein (the non-DUB F-box protein FBXO2) or buffer only show no cleavage. Error bars show standard deviations.

(B) There is no significant difference between the initial reaction velocity of 2UIM and 3UIM ataxin-3(Q22) (p > 0.4 by a 2 tailed heteroscedastic Student’s t-test).
Figure 14. 2UIM ataxin-3 has a higher propensity to aggregate than 3UIM ataxin-3 in mammalian cells.

(A) Representative immunocytochemistry of Cos7 cells transiently expressing Flag-tagged ataxin-3 splice isoforms and the UIM3(SA/DG) mutant. Cells were gated by fluorescence intensity into populations of moderate and high expressors. (B) Quantification of the number of puncta per cell. Error bars represent the standard deviation within each bin. Frequency distributions are significantly different (*p < 0.0001, ** p < 1x10^{-11}) between 2UIM and 3UIM ataxin-3 and between 2UIM and UIM3-mutant ataxin-3, but not between 3UIM and UIM3-mutant ataxin-3, by a $\chi^2$ test for independence, df = 3.
Figure 15. 2UIM ataxin-3 has a higher propensity to aggregate than 3UIM ataxin-3 in brain tissue.

(A) Supernatant (sup) and pellet (pel) fractions of non-denaturing RIPA lysates from the brains of aged 3UIM-predominant MJD84.2 and 2UIM-expressing Q71B hemizygous ATXN3 transgenic mice were analyzed by Western blot with 1H9 anti-ataxin-3 antibody. Insoluble microaggregates were detected at the base of lane wells, whereas soluble transprotein and endogenous ataxin-3 were visualized within the resolving gel. (B) Quantification of the ratio of insoluble to soluble ataxin-3 transprotein seen in (A). 3UIM-predominant MJD84.2 mice show a significantly lower % normalized intensity ratio of insoluble:soluble transprotein than 2UIM-only Q71B mice (*p < 0.0005 by a 1 tailed heteroscedastic Student’s t-test).
Figure 16. 18 – 23 month old MJD84.2 and Q71B hemizygotes show a mild ataxic phenotype.

(A) Representative movement traces captured during open field analysis of 18 – 23 month old wild type or transgenic MJD15.4, MJD84.2, and Q71B mice. Long movements (red), short movements (green), stationary (black), and voxels never occupied (white) were traced over 1.5 hours of analysis. (B) Quantification of the distance traveled during 1.5 hours of open field analysis. Transgenic MJD84.2 and Q71B mice show a nonsignificant trend towards hypoactivity (error bars indicate SEM). (C) Average mass of 18 – 23 month old MJD15.4, MJD84.2, and Q71B animals. MJD84.2 hemizygotes weighed significantly less than nontransgenic controls (p < 0.05 by a heteroscedastic one tailed Student’s t-test; error bars indicate SEM). (D) Gait analysis of 18 – 19 month old MJD84.2 and Q71B mice, showing forepaw (blue) and hindpaw (black) prints. Compared to their nontransgenic controls, both MJD84.2 and Q71B hemizygotes showed a mildly ataxic gait.
A

(-/-)  (+/-)
MJD15.4
MJD84.2
Q71B

B

Distance Traveled (m)

MJD15.4  MJD84.2  Q71B
(-/-)  (+/-)  (-/-)  (+/-)  (-/-)  (+/-)

C

Mass (g)

MJD15.4  MJD84.2  Q71B
(-/-)  (+/-)  (-/-)  (+/-)  (-/-)  (+/-)

D

MJD84.2
(-/-)

MJD84.2
(+/-)

Q71B
(-/-)

Q71B
(+/-)
Figure 17. 2UIM ataxin-3 is much less stable than 3UIM ataxin-3.
(A) Representative cycloheximide treatment in Cos7 cells transiently transfected with Flag-tagged ataxin-3 constructs; data are visualized by anti-Flag Western blotting followed by Coomassie Brilliant Blue staining of the PVDF membrane. (B) Quantification of ataxin-3 levels during a 24 hour incubation with cycloheximide: 2UIM ataxin-3 is degraded significantly faster than 3UIM or UIM3(SA/DG) mutant ataxin-3 at 10 and 24 hours (*p < 0.02 by a two-tailed heteroscedastic Student’s t-test). Densitometry analysis is plotted as the percentage of signal at time zero, normalized to total protein signal.
Figure 18. 2UIM ataxin-3 is subject to rapid proteasomal degradation, while 3UIM ataxin-3 is subject to autophagy.

Quantification of ataxin-3 levels during a 24 hour cycloheximide incubation in the absence or presence of the proteasomal inhibitor epoxomycin or the macroautophagy inhibitor 3-methyladenine; loss of protein at 24 hours is rescued by proteasomal inhibition for 2UIM ataxin-3 and by inhibition of macroautophagy for 3UIM and UIM3-mutant ataxin-3 (†p < 0.05 or ‡p < 0.01 compared to time zero; *p < 0.05 or **p < 0.01 compared to 24 hour time point by paired one-tailed Student’s t-tests). Densitometry analysis is plotted as the percentage of signal at time zero, normalized to total protein signal.
Figure 19. Model for the differential aggregation properties and processing of 2UIM and 3UIM ataxin-3.

3UIM ataxin-3 follows the standard multi-domain aggregation mechanism for unexpanded polyglutamine proteins with self-association prone flanking regions (see Figure 3). 2UIM ataxin-3 exists in at least two monomeric states: the native conformation, in which the hydrophobic tail remains buried and protected from the aqueous environment, and an aggregation-prone monomeric conformation with an exposed hydrophobic tail. In order to shield the hydrophobic tail from the aqueous cellular environment, it can either revert to the native conformation or assemble into an oligomeric form, through the self-association propensity of the Josephin domain as well as via hydrophobic interactions of the 2UIM-specific domain. Within the oligomeric form, the hydrophobic C-termini will associate, increasing the local polyglutamine concentration beyond that seen in the 3UIM oligomer, favoring formation of a detergent-insoluble aggregate. Unstable forms of the monomer and the oligomers may be polyubiquitinated and degraded, predominantly through the UPP. The insoluble fibrils are more problematic for the protein quality control systems, and accumulate to form biochemically and microscopically detectable aggregates.

CHAPTER 3
GENERATION OF A (CAG)$_N$-EXPANSION KNOCK-IN MURINE MODEL OF SCA3

Abstract

Although knock-in models of polyglutamine diseases are genetically precise systems that recreate key features of individual polyglutamine diseases, a knock-in model of SCA3 does not currently exist. Toward the goal of developing a knock-in model, we generated a targeting vector designed to convert murine $\text{Atxn3}(Q_6)$ to $\text{Atxn3}(Q_3KQ_{82})$. We compared the gene targeting efficiencies and rates of chromosomal instability of three ES cell lines (W4, Bruce4.G9, and UMB6JD7), with the goal of creating an $\text{Atxn3}(Q_3KQ_{82})$ knock-in model directly on a C57BL/6 background, to avoid backcrossing before phenotypic analysis. We successfully targeted exon 10 of the $\text{Atxn3}$ gene by homologous recombination in both C57BL/6-derived ES lines, but UMB6JD7 exhibited higher chromosomal instability and did not produce chimeric animals upon blastocyst microinjection. Although we observed premature mortality in two out of the three chimeras that exhibited at least 40% coat chimerism, we achieved germline transmission, albeit at a low rate. Because predicted Mendelian ratios were observed in the progeny of hemizygous F1 knock-ins, however, the mortality and low germline transmission rate observed in chimeric animals are unlikely to reflect polyglutamine-mediated toxicity. Mosaic \textit{in vivo} FLPe-mediated excision of the FRT-PGK-Neo-FRT positive selection cassette was observed in our $F_2$ FLPe dual hemizygous animals. Importantly, $F_2$ FLPe hemizygous mice express the expanded (pathogenic) $\text{Atxn3}$ protein. Thus, we have created the first knock-in model of SCA3, which we anticipate will be an important resource with which to elucidate the effects of genomic, message, and protein context on polyglutamine toxicity in this disease.
Introduction

Animal models of disease are crucial tools for understanding disease mechanisms and for testing potential therapeutic strategies. In particular, murine models are particularly attractive due to the high fecundity, genetic malleability, and rapid life cycles of mice, as well as due to the extensive phenotypic, genomic, and proteomic characterization that is readily available for this species. While a variety of exogenous polyglutamine expression modalities, including cDNA, BAC, and YAC mouse models have provided useful insights into polyglutamine pathogenesis, there has been an increasing trend to create “knock-in” models through targeted insertion of expanded \((CAG)_n\) sequences into the appropriate gene regions of well conserved murine disease homologues. In addition to knock-in models of HD (discussed in Chapter 1), knock-in models of polyglutamine diseases have been created for SBMA (Yu et al., 2006a, Yu et al., 2006b), SCA1 (Lorenzetti et al., 2000, Watase et al., 2002, Watase et al., 2003), SCA6 (Watase et al., 2008), and SCA7 (Yoo et al., 2003). The details of these lines are summarized in Table 6. Several trends have emerged among these models that reinforce observations made in earlier comparisons limited to murine models of HD.

Knock-in models employing repeats in the pathophysiological range for human polyglutamine diseases have proven to be useful in numerous ways. At the DNA level, knock-in models display pathophysiological aspects of disease more faithfully than a number of transgenic models, particularly aspects of repeat instability and anticipation. Genomic context-dependent somatic and intergenerational repeat instability events that are exhibited when repeats reside within their appropriate genomic context are not observed in the corresponding cDNA transgenic models (Libby et al., 2003). Both somatic repeat instability (Wheeler et al., 1999) and intergenerational repeat instability (Lorenzetti et al., 2000) have been demonstrated in knock-in models with pathogenic repeat lengths. Knock-in models also faithfully recreate features of human disease that are not feasible to study in cDNA transgenic models. For example, enrichment of
polyglutamine-encoding splice variants within selectively vulnerable Purkinje cell populations, which was seen in human SCA6 brains compared to controls (Tsunemi et al., 2008), was also documented in disease-range \((CAG)_{30}\) knock-in brains compared to unexpanded \((CAG)_{14}\) knock-in controls (Watase et al., 2008). Changes in protein behavior within vulnerable neuronal populations have also been replicated by knock-in mouse lines expressing repeats within the pathophysiological range (Wheeler et al., 1999). While behavioral phenotypes are generally mild (Lorenzetti et al., 2000, Lin et al., 2001) to absent (White et al., 1997, Wheeler et al., 1999, Yu et al., 2006b) in knock-in models within the pathophysiological repeat range, the knock-in model for infantile-onset SCA7 (which expresses a massive expansion of 266 repeats) demonstrates a robust phenotype that mirrors many disease-specific features (Yoo et al., 2003).

Although knock-in models with repeat expansions longer than those seen in human disease arguably may be less faithful to human disease in terms of the precise inciting event, they often accurately recreate features of human disease. The intergenerational instability and alterations in mRNA splicing observed in pathophysiological repeat lines are even more pronounced when the \((CAG)_{n}\) repeats are expanded beyond the pathophysiological range (Watase et al., 2003, Watase et al., 2008). Perhaps more important for their utility in studying disease mechanisms and potential therapeutic approaches, very large repeats are more likely to exhibit disease-specific behavioral phenotypes and characteristic neuropathological findings. Knock-in of 113 repeats into the murine androgen receptor locus recreates the androgen-dependent early myotonia, late neurological changes, and androgen insensitivity observed in individuals with SBMA (Yu et al., 2006a, Yu et al., 2006b). Knock-in of a 150 glutamine-encoding repeat into the murine homologue \(Hdh\) induces motor deficits, weight loss, rare seizures, and neuropathological hallmarks of HD, including neuronal inclusions and reactive gliosis (Lin et al., 2001, Heng et al., 2007). The 154Q knock-in model of SCA1 recreates progressive motor defects, cognitive impairment, weight loss and premature mortality.
The Q84 knock-in model of SCA6 also recreates progressive motor dysfunction and pathological evidence of protein aggregation (Watase et al., 2008). The robust, measurable phenotypes seen in suprapathophysiological expansion knock-ins make them particularly useful for pre-clinical analysis of therapeutic interventions.

Comparison of the $Hdh^{Q80}$, $Hdh^{Q92}$, and $Hdh^{Q111}$ models of HD illustrates another important consideration in knock-in models. Despite the fact that the $Hdh^{Q92}$ and $Hdh^{Q111}$ models contain larger expansions, they do not have a behavioral phenotype, whereas the $Hdh^{Q80}$ model had mild but measurable motor defects. Although this discrepancy has not been definitively characterized, it is known that these strains are maintained on different genetic backgrounds (mixed 129/Ola with limited C57BL/6J backcross for $Hdh^{Q80}$, 129/CD1 for $Hdh^{Q92}$ and $Hdh^{Q111}$). Similarly, homozygous Sca1$^{Q78}$/Sca1$^{Q78}$ knock-in mice exhibited poorer motor coordination on a mixed 129/SvEv-C57BL/6 background than on a pure 129/SvEv background (Lorenzetti et al., 2000). Most of the existing polyglutamine disease knock-in models were developed in 129-derived embryonic stem cell (ES cell) lines. For many compelling reasons, however, these animals have not been maintained on a pure 129 substrain background. First and foremost, the polyglutamine diseases are neurodegenerative diseases, and 129 strains are neuroanatomically abnormal (Wahlsten et al., 1992). Compared to other commonly used strains of mice, they also exhibit differences in a number of behavioral tests (Crawley, 1996, Gerlai, 1996, Bothe et al., 2004). Therefore, existing polyglutamine knock-in models have typically been backcrossed to other strains, including CD1 and C57BL/6J. Although backcrossing solves some of the problems associated with the 129 substrains, it can cause its own problems, particularly in the context of subtle disease phenotypes (Gerlai, 1996).

Although there is no “ideal” mouse strain for all animal studies, C57BL/6J is the most widely used inbred mouse strain, and is considered the strain of choice in various neurological, immunological, and metabolic disease paradigms. C57BL/6J animals are
also favored for their reproductive capacity, as they are able to produce relatively large litters over a wide span of their adult life. Additionally, because they are genetically homogeneous and were used to create the reference mouse genome annotation, it is much easier to define and recreate specific genomic regions. For all of these reasons, many laboratories have created tools to perform targeted genetic manipulations directly on a C57BL/6 or mixed C57BL/6 background (Auerbach et al., 2000, Seong et al., 2004, Hughes et al., 2007, Keskin tepe et al., 2007).

Key advantages and disadvantages of gene targeting in 129 vs. C57BL/6 substrains are summarized in Figure 20. For instance, there is significant genetic diversity among 129 substrains and ES lines (Simpson et al., 1997). Although all characterized C57BL/6-derived ES cell lines have diverged somewhat from the reference C57BL/6J sequence, they show less variation than 129 substrains (Hughes et al., 2007). The exponential dependence of gene targeting efficiencies on the extent of flanking arm sequence identity (Deng and Capecchi, 1992) makes the ease of designing and creating a highly homologous targeting vector in C57BL/6 strains a clear advantage. Conversely, the ease of culture and maintenance of pluripotency in 129-derived ES cells compared to other historically “nonpermissive” strains including C57BL/6J has been a clear advantage of ES cells created from 129 substrains. C57BL/6-derived ES cell lines are also less efficient at blastocyst colonization, even in co-isogenic embryos, yielding relatively lower numbers of chimeras, with lower percentages of coat chimerism in those animals produced, than 129-derived ES cells. This phenomenon is less of a disadvantage than was initially predicted, however, because germline transmission has been reliably achieved from C57BL/6 chimeras in which low rates of embryonic sex conversion and low % coat chimerism would have predicted a failure to achieve germline transmission in 129-derived strains (Seong et al., 2004). Moreover, the higher fecundity of the C57BL/6 line greatly enhances the ability to rapidly and reliably generate F1 founders without the threat of line loss due to subfertility. The ability to begin expansion of the colony and
phenotypic analysis without time consuming and expensive backcrossing is also a distinct advantage of gene targeting on a C57BL/6 background.

Despite the multiple advantages of studying polyglutamine expansions in the most genetically precise context provided by the knock-in approach, to date there are no published reports of knock-in models for DRPLA, SCA2, SCA3, or SCA17. In this chapter, I describe the creation of a knock-in model of SCA3. I discuss methodological choices in the design of this model, including the choice of ES cell lines used for targeted insertion of the expansion into the Atxn3 gene, technical challenges of working with expanded repeats in their native DNA context, the successful establishment of knock-in mice that express the pathogenic gene product, and future directions for this project.

**Materials and Methods**

**Generation of the targeting vector**

Genomic murine Atxn3 DNA from a C57BL/6-tyr(c-2J) albino embryonic stem (ES) cell line (Millipore, Cat SCR011) was amplified using Pfu Ultra polymerase (Stratagene, Cat 600380). Specific primers were used to amplify: a 4 kb upstream flanking arm derived from intron 9-10 and a 2.6 kb downstream flanking arm that spans exon 10, from ~ 500 bp upstream of the intron 9-10 / exon 10 junction to a region of intron 10-11 downstream of an endogenous EcoRI site (Figure 21). NotI and SalI restriction sites were engineered to flank the upstream flanking arm using the forward primer, MJDNotA (5’ GCG GCC GCG CAG GCG GAT TTC TGA GTT TGC 3’), and the reverse primer, MJDSalA (5’ TTT TGT CGA CCT GAT GCT GCT TAC CTG TGT CC 3’). A novel KpnI restriction site was engineered onto the 5’ end of the downstream flanking arm with the forward primer MJDKpnA (5’ GGT ACC TGA GCT CTC ATT GCT TAC CTG TGT CC 3’), whereas an endogenous murine EcoRI site was used to define the 3’ limit of the downstream flanking arm by using a reverse primer, MJDRIB (5’ GTT TGC CGT GTT GAT GGG AAG G 3’) that targets a sequence beyond this site. Both
amplicons were initially cloned into pCR-Blunt-II-TOPO. The upstream flanking arm was then subcloned into pBY49a (provided by Baoli Yang) upstream of the FRT-PGK-Neo-FRT positive selection cassette to generate pBY/MJD4up without further modification.

We employed two methods to expand the (CAG)$_n$ repeat sequence within the targeting vector (Table 7). First, the downstream flanking arm was subcloned into pBluescript SK(-) to generate pBl/MJD10Q, and the (CAG)$_n$ repeat was expanded using a modified QuickChange approach (Figure 22A) (Geiser et al., 2001). Briefly, a human expanded (CAG)$_n$ template (either At3-Q129-GFP or At3-Q166-GFP) was amplified with partially complementary primers to generate an expanded (CAG)$_n$ “megaprimer” flanked by murine genomic sequence. This double-stranded megaprimer was used to insert an expanded (CAG)$_n$ repeat into the Atxn3 gene using the QuickChange Mutagenesis method (Stratagene). One clone generated by this method (#71) was chosen for additional repeat expansion through splicing by overlap extension (SOE, Figure 23A) (Laccone et al., 1999). Briefly, the 3’ flanking arm template was digested to create an upstream /KpnI/XhoI fragment containing the trinucleotide repeat domain and enough 5’ flanking sequence to bind the forward primer, MJDKpnA, and a downstream /Xmal/EcoRI fragment containing the trinucleotide repeat domain and enough 3’ flanking sequence to bind the reverse primer, MJDRIB. The /KpnI/XhoI fragment was amplified using MJDKpnA and a (CTG)$_n$ primer capable of annealing anywhere within the repeat domain. Similarly, the /Xmal/EcoRI fragment was amplified using MJDRIB and a (CAG)$_n$ primer. The products of these two reactions were then gel purified and spliced by overlap extension to generate a population of 3’ flanking arms with variable trinucleotide repeats. All clones produced by these methods were initially screened by colony PCR (Table 7), and promising expansion candidates were confirmed by restriction digestion with BglII and DNA sequencing.
In order to create a knock-in line with a repeat length at the upper limit of the human disease range, we inserted the 3’ flanking arm from clone #71 of the megaprimer expansion series, which contains a Q₃KQ₈₂-encoding expansion, into the targeting vector between the FRT-PGK-Neo-FRT positive selection cassette and the PGK-TK negative selection cassette, using KpnI and EcoRI restriction sites. In an attempt to create a targeting vector in the suprapathophysiological (hyperexpanded) repeat range, we also extensively subcloned clone #13 from a series of repeat expansion clones generated by the SOE method described by Laccone and colleagues (Laccone et al., 1999), prior to its insertion into the 3’ flanking arm. Notably, we carried out all cloning steps up to the point when repeat expansions were introduced in DH5α, One Shot Top10, or XL-1 Blue competent cells. Maintenance of plasmids containing significant expansions, however, was carried out in the recombinase-negative SURE2 and Stbl3 E. coli strains, in order to reduce the frequency of repeat contractions and recombination events involving the repeat-flanking regions.

ES cell lines

Gene targeting frequencies and the chromosomal stability of successfully targeted ES cell colonies were compared between three ES cell lines: W4, Bruce4-G9, and UMB6JD7 (Table 8). The W4 ES cell line (Auerbach et al., 2000) was generated on a 129S6/SvEvTac genetic background. W4 is a robust ES line, which is relatively easy to maintain in cell culture and exhibits good colonization of blastocysts upon microinjection, resulting in high levels of coat chimerism in the resulting animals. Based on these favorable qualities, they are often used for the production of gene-targeted mouse lines (Seong et al., 2004). The Bruce4 ES cell line was originally generated for immunology research from mice bred to contain the NZB-derived Thy1.1 allele on a C57BL/6 background (C57BL/6;NZB backcross at N3) (Kontgen et al., 1993). As of 2007, the majority of Jackson Laboratories-annotated genetically modified mice
generated on a C57BL/6 background were produced using Bruce4 ES cells. However, this line has documented genetic differences from the C57BL/6J reference sequence and displays a high rate of aneuploidy in electroporated ES cells (Hughes et al., 2007). Bruce4.G9 cells, which were sub-cloned from the Bruce4 line, have been selected for better rates of euploidy than the parent strain. Unfortunately, they have lower plating efficiencies after electroporation, and suffer from the same genetic differences from the reference sequence as the parental Bruce4 line. UMB6JD7 ES cells were generated in house at the University of Michigan Transgenic Animal Core on a pure C57BL/6J background. Although they have been used for successful gene targeting (personal communication with Thom Saunders), they were incompletely characterized at the time this work was carried out.

Gene targeting by homologous recombination

The complete targeting vector was purified with an endotoxin free plasmid megaprep kit (Qiagen Cat 12381), linearized with NdeI, and electroporated into W4, Bruce4.G9, and UMB6JD7 ES cells. G418 and ganciclovir selection were used to enrich for incorporation of the positive selection cassette and loss of the negative selection cassette, using the standard protocol of the University of Michigan Transgenic Animal Model Core (Soliman et al., 2007). Resistant colonies were picked and separated into triplicate 96-well plates for analysis and short term cryostorage. Homologously recombined clones were subsequently re-expanded for confirmatory testing, cytological and chromosomal analysis, blastocyst injection, and long-term cryostorage (Saunders, 2011).

Screening ES cells for Homologous Recombination and Euploidy

DNA was prepared from G418- and ganciclovir-resistant ES cell colonies in a 96-well format, as previously described (Ramirez-Solis et al., 1992). Each colony was
screened for homologous recombination at the 3’ flanking arm using PCR- and Southern blot-based strategies. Positive or indeterminate samples were assayed for homologous recombination at the 5’ flanking arm, using a PCR-based strategy, and the integrity of the \((CAG)_n\) trinucleotide repeat expansion was confirmed by a secondary Southern blot and DNA sequencing. The details of the PCR-based screening methods are listed in Table 9 and diagrammed schematically in Figure 25A. Briefly, both PCR-based strategies distinguish homologous recombinants from random insertions of the targeting vector by utilizing primer pairs that span from regions outside of the flanking arms into the FRT-PGK-Neo-FRT positive selection cassette. The 3’ flanking arm screen, which was used as a first-pass screening approach, utilizes an off-target 1 kb band as an endogenous control in addition to the 3 kb band that is specific for successfully targeted knock-in clones. The confirmatory 5’ flanking arm screen lacks an endogenous control band, and relies solely upon the presence or absence of the 4.7 kb knock-in specific band. Both PCR reactions were carried out using Platinum Taq High Fidelity polymerase (Invitrogen, Cat #11304-011), following the manufacturer’s protocols in the recommended buffer systems.

Southern blots were carried out using the DIG High Prime Starter Kit II (Roche, Cat #1 585 614), according to the manufacturer’s recommended instructions with the addition of a probe clean-up step using the High Pure PCR Product Purification Kit (Roche, Cat #11 732 668 001). DNA was transferred to Roche positively charged nylon membranes (Cat. #1 417 240) using the high salt transfer protocol, followed by crosslinking with the BioRad Genelinker, on program C3 (2xSSC-dampened membrane, 150 mJ). Pre-hybridization and probe hybridization were carried out at 43°C using pre-warmed (68°C) hybridization solutions. All post-hybridization washes were carried out in separate incubation trays, in order to reduce background signal.

Our Southern blot strategies are diagrammed schematically in Figure 25A. A first pass Southern blot used either probe DIG765 or probe DIG1178 to label AflIII-digested
genomic DNA, and homologous recombinants were recognized by the shift from 5.21 kb to 7.46 kb, which was conferred by the presence of both the knock-in allele and the positive selection cassette. A confirmatory SpeI Southern blot relied on probe DIG765 to detect the shift from 3.29 kb to 3.53 kb, which was conferred specifically by expansion of the trinucleotide repeat. Thus these approaches confirmed homologous recombination within the 3’ flanking arm and assured that this recombination event occurred downstream of the targeted (CAG)$_2$(CAAAAG)(CAG)$_{82}$ repeat expansion in Atxn3, exon 10.

The ploidy of each successfully targeted ES cell colony was determined as previously described (Hughes et al., 2007). Briefly, to generate metaphase spreads, overnight subcultures of the ES cell colonies were treated with colcemid, trypsinized to a single cell suspension, equilibrated in 75 mM KCl, fixed in 3:1 methanol:glacial acetic acid, and stained with Gurr’s Giemsa (Invitrogen, Cat #10092). Metaphase chromosome spreads from 10 – 25 individual ES cells were visualized at 1000x magnification to infer the ploidy of each Atxn3(Q$_3$KQ$_{82}$)-targeted colony. At least 20 individual cells were assessed from any colonies that were selected for blastocyst microinjection.

**Chimera Generation and Screening for Germline Transmission**

Knock-in positive ES cell clones confirmed to be sufficiently euploid were then microinjected into homozygous albino B6(Cg)-Tyr< c-2J>/J blastocysts, in order to allow black/white assessment of coat chimerism in the resulting progeny. Microinjected blastocysts were introduced into the uterine horns of pseudopregnant female mice. Chimeras were also crossed to albino B6(Cg)-Tyr< c-2J>/J animals, in order to facilitate black/white screening for pups carrying microinjected ES cell-derived alleles. All black pups were assayed for germline transmission of the knock-in allele using at least two separate reactions: one targeting the (CAG)$_n$ expanded exon 10 and one targeting the
positive selection cassette (see Table 9 for reaction details and Figure 26A for a schematic of primer locations). A limited number of F1 x C57BL/6J crosses were performed to compare germline transmission rates from F1 hemizygotes with an intact positive selection cassette with those observed in the chimera crosses. The F2 progeny from these crosses were genotyped using the same assays.

Excision of the Positive Selection Cassette

In order to facilitate in vivo excision of the FRT-site flanked positive selection cassette, F1 hemizygous knock-in animals were crossed with homozygous FLPe transgenic mice, which express an enhanced Flip Recombinase (FLPe) under the control of regulatory sequences from the human β-actin (ACTB) promoter (Jackson Laboratories, strain B6.Cg-Tg(ACTFLPe)9205Dym/J). Although this line can produce mosaic excision of FRT-flanked sequences in tissues of pups from FRT x FLPe crosses, the high expression of this transgene in male and female germ cells assures high rates of FLPe-recombined alleles in second generation progeny, including FLPe nontransgenic pups (Rodriguez et al., 2000). To confirm that all progeny were hemizygous FLPe transgenic, we utilized the PCR-based FLPe screen (Table 9). To detect complete or mosaic excision of the FRT-site flanked positive selection cassette in F2 mice, we utilized a novel XbaI restriction site introduced by incorporation of the positive selection cassette in homologous recombinants. This site remains intact after FLPe-mediated excision of the positive selection cassette, but is absent from intron 9-10 of the wild type Atxn3 allele (see schematic in Figure 28A). First, we amplified from MJD4F6 to MJD10R4, as in the Positive Selection Cassette Confirmation screen (see above and Table 9). As before, any bands running at 3.3 kb established the presence of an intact positive selection cassette. The amplicons from the wild type allele and any FLPe-recombined knock-in alleles, however, both run between 1.3 – 1.45 kb. Thus, this lower band was excised, gel purified, and digested with XbaI (NEB, Cat # R0145S) using the manufacturer’s
recommended buffer. While the wild type amplicon is resistant to XbaI digestion, the FRT-only knock-in allele is cleaved into two bands (926 bp and 514 bp). FLPe transgenic F2 mice with evidence of mosaic or complete excision of the positive selection cassette are currently being crossed with C57BL/6J mice. F3 pups that are Atxn3(Q3KQ82) knock-in positive, FRT-only (positive selection cassette deleted), and FLPe nontransgenic will be the founders of the final Atxn3(Q3KQ82) knock-in line.

Confirmation of Knock-in Protein Expression

Preparation of protein from murine tissues and 1D Western blot analyses was performed as described in Chapter 2. Tissue lysates were probed with 1H9 mAb (1:500) to detect endogenous unexpanded ataxin-3 and knock-in ataxin-3 expression. The polyglutamine expansion-specific 1C2 mAb (1:500) was used to detect the expanded polyglutamine expressed from the knock-in allele only. GAPDH mAb (1:1000, Chemicon MAB374) was used as a loading control.

Results

Targeting vector construction: complications of (CAG)$_n$ repeat expansion instability

In order to knock-in an expanded (CAG)$_n$ repeat into Atxn3, the murine homologue of the human ATXN3 gene, we constructed a targeting vector containing an expanded (CAG)$_n$ repeat within exon 10 of Atxn3 (Figure 21). Murine genomic DNA was amplified from a B6(Cg)-Tyr<+>-2J/J ES cell line to generate a 4kb 5’ flanking arm from intron 9-10 and a 2.6 kb 3’ flanking arm spanning exon 10 from approximately 500bp upstream of the intron 9-10 / exon 10 junction to a region in intron 10-11 downstream of an endogenous EcoRI site. While the upstream flanking arm was subcloned into the targeting vector template plasmid, pBY49a (Figure 21A), without further modification, the downstream flanking arm required expansion of the endogenous
murine (CAG)_6 repeat before it could be inserted into the targeting vector template. Two methods were used to expand the (CAG)_6 trinucleotide repeat within the murine ataxin-3 gene. The first method takes advantage of the QuickChange method of mutagenesis, utilizing a megaprimer containing a large (CAG)_n repeat. The second method takes advantage of the repetitive nature of the (CAG)_n sequence itself, allowing amplification and expansion of the CAG repeat size through the technique known as splicing by overlap extension.

In the first method (Figure 22), primers with 5’ homology to murine Atxn3 sequences and 3’ homology to the (CAG)_n repeat were used to amplify a “megaprimer” from plasmids containing expanded human ATXN3. In practice, this first step generated a population of megaprimers, ranging in size from ~ 100 – 500 bp (Figure 22B), which were binned into bands of ~100 bp, 100 – 250 bp, and 250 – 500 bp, gel purified and utilized as mutagenizing megaprimers for a small plasmid containing the unmodified 3’ flanking arm. Similarly, in practice, these double-stranded megaprimers generated a population of plasmids with variably expanded (CAG)_n repeat sequences within murine Atxn3 exon 10 (Figure 22C). Colony PCR was used to identify E. coli that contained expanded (CAG)_n repeats (Figure 22D). Promising candidate clones were subsequently analyzed directly by DNA sequencing.

This method not only generated a wide variety of (CAG)_n repeat lengths ranging from 6 to > 100 repeats, but also numerous minor, but significant, alterations of the trinucleotide repeat and flanking sequence (Figure 22 E). Because the trinucleotide expansion was amplified from a human ATXN3 cDNA template, we expected mutagenized clones to contain a Q_kKQ_n-encoding (CAG)_2CAAAAG(CAG)_n repeat domain rather than the pure (CAG)_n repeat that occurs in murine Atxn3. Our resultant population of plasmids included uninterrupted repeats and repeats interrupted by either one or two lysine-encoding (AAG) codons. At least one clone contained a small frameshift-inducing mutation. Additional clones contained in-frame repetitions of
variable lengths of 5’ and 3’ flanking sequences, producing clones with multiple splice acceptor sites or concatenated truncations of the repeat sequence. Of the clones sequenced, #71 was of particular interest because it contained a single Q

3

KQ

82

-encoding (CAG)

2

CAAAAG(CAG)

82

expansion without any duplications in its flanking sequence. It was ideal for insertion into the targeting vector template in order to create a knock-in model with a (CAG)\textsubscript{n} repeat expansion at the upper end of the SCA3 disease range in humans. Yet, because repeats that cause severe disease in humans have produced relatively mild phenotypes in knock-in models of other polyglutamine diseases, we also wished to use clone #71 as a template for further repeat expansion.

Because clone #71 provided us with a moderate expanded repeat flanked by endogenous murine Atxn3 sequence, we were able to utilize a method of trinucleotide expansion (Figure 23A) that others have shown to be effective for expanding trinucleotide repeats from the non-expanded range into the expanded range in human ataxin-3 cDNA (Laccone et al., 1999). First, fragments of clone #71 were amplified from nonrepetitive flanking sequences into the trinucleotide repeat. This amplification from specific primers into the repeat sequence was carried out in both directions separately. The 5’ population was made up of amplicons containing ~ 500 bp upstream of the repeat followed by variable trinucleotide repeat lengths. The 3’ population was made up of amplicons containing variable trinucleotide lengths followed by the remainder of Exon10 plus about 1.5 kb of intron 10 - 11 (Figure 23B). The products of these two reactions were spliced by overlap extension to generate a broad smear of products comprising the entire 3’ flanking arm with variable trinucleotide repeat lengths. This population of amplicons was then gel purified, cloned, and sequenced. Although some very large repeats were generated in this manner, every clone sent for sequencing contained two or more distinct repeat lengths, suggesting either repeat instability, recombination, and contraction during bacterial culture, or concatenation of the repeat-containing sequences during cloning (Figure 23C). Despite extensive subcloning of clone #13 in the
recombinase deficient *E. coli* strains SURE2 and Stbl3, we were unable to isolate a stable, supraphysiologically expanded repeat.

In summary, these methods of trinucleotide repeat expansion resulted in predominantly moderately-sized and longer, interrupted repeats, which were unstable in *E. coli*, and unsuitable for use in a targeting vector. Therefore, during these studies, we only generated a single targeting vector from clone #71, with a sequence-confirmed Q₃KQ₈₂ expansion inserted into exon 10 of the downstream flanking arm. The final linearized targeting vector consisted of a 5’ flanking arm containing ~ 4 kb from intron 9 – 10 of the murine *Atxn3* gene, the FRT-PGK-Neo-FRT positive selection cassette, the 3’ flanking arm spanning the expansion-containing exon 10, from 500 bp upstream of exon 10 to the endogenous EcoRI site in intron 10-11, and the PGK-TK negative selection cassette (Figure 24).

C57BL/6-derived ES cell lines demonstrated better gene-targeting efficiency, but significant chromosomal instability.

We compared three different ES cell lines (W4, Bruce4-G9, and UMB6JD7) with respect both to the efficiency of homologous recombination of murine *Atxn3* gene with the Atxin3(Q₃KQ₈₂) targeting vector and to the chromosomal stability of successfully targeted ES cell colonies. After electroporation of ES cells with the targeting vector, DNA from G418- and ganciclovir-resistant clones was screened for evidence of homologous recombination events resulting in the incorporation of the (Q₃KQ₈₂)-encoding trinucleotide expansion into the murine *Atxn3* gene (Figure 25). Homologous recombination within the 3’ flanking arm was confirmed by Southern blot of AflIII digested genomic ES cell DNA (Figure 25B) and by PCR amplification from the positive selection cassette to a primer downstream of the EcoRI site that defined the downstream end of the 3’ flanking arm. Positive or indeterminate clones were then assayed for homologous recombination within the 5’ flanking arm, utilizing a similar strategy (Figure
Because our targeted mutation lies about 500 bp downstream of the positive selection cassette, we employed a confirmatory Southern blot strategy on SpeI digested ES cell DNA and sequenced the exon 10 containing region of the 3’ flanking arm amplicon (Figure 25B and data not shown) to confirm the presence of an expanded repeat in positive clones.

Based on these screening strategies, we observed that 0 out of 192 W4 clones (0%), 2 out of 96 Bruce4-G9 clones (2.1%), and 3 out of 192 UMB6JD7 clones (1.6%) were positive for homologous recombination with an intact repeat expansion. Upon analysis of metaphase chromosomal spreads, however, two of the three positive UMB6JD7 clones exhibited unacceptable levels of aneuploidy (Table 10). While one of the two Bruce4.G9 clones exhibited chromosomal loss in a minority of cells analyzed, the UMB6JD7-derived clones exhibited higher degrees of aneuploidy, including chromosomal duplications, deletions, and translocations. Because no successful targeted gene mutation events were observed in the 129S6/SvEvTac-derived W4 cells, the incidence of aneuploidy among electroporated cells in this line was not assessed. Ultimately, of the five correct, homologously recombined clones, only three had acceptable levels of euploid cells (Bruce4.G9 clones 3C8, 3H11, and UMB6D7 clone 6C4). These clones were expanded, microinjected into B6(Cg)-Tyr< c-2J >/J blastocysts and introduced into pseudopregnant female mice.

Low germline transmission rates and premature mortality were observed in high percentage chimeras

Two rounds of blastocyst microinjections produced 11 chimeras with % coat chimerism ranging from 5 – 80% (Table 11). Notably, in the first round of blastocyst injections, only the 100% euploid clone (clone 3C8) produced an animal with appreciable coat chimerism. Neither 3H11 nor 6C4 successfully implanted into blastocysts to generate chimeric progeny. As a result, only clone 3C8 was thawed and expanded for a
second round of blastocyst injections. The animal displaying the highest degree of coat chimerism, Chimera 2.1, was observed to be small at weaning, and was found dead of unknown causes at one month of age. Chimeras 2.2 and 2.3, which displayed 60% and 40% coat chimerism, respectively, each achieved germline transmission to at least one hemizygous knock-in F1 pup, as determined by PCR-based screening assays that differentiate sequences containing the (CAG)$_2$CAAAAG(CAG)$_{82}$ expansion and the positive selection cassette from the wild type allele (Figure 26). Chimera 2.3 died of unknown causes at 5 months of age after a few productive matings. No germline transmission of the knock-in allele was observed in animals displaying less than 40% coat chimerism, and no Bruce4.G9 ES cell-derived pups were sired by animals displaying less than 20% chimerism. In total, of 647 pups sired by these chimeras, 488 (75%) were derived from donor blastocysts, 155 animals (24%) inherited the wild type allele from ES cell-derived germ cells, and only 4 animals (<1%) inherited the knock-in allele.

Comparison of germline transmission rates in chimeras and F1 progeny suggests nonclonality of ES cell colony 3C8

Based on the low rates of germline transmission and mortality of chimeras with high coat chimerism, we were concerned that our relatively small repeat expansion in \textit{Atxn3} might be causing unexpectedly severe toxicity, compared to phenotypes observed in previous polyglutamine disease knock-in models. It has been suggested that proximity to epigenetically modified CpG islands enhances instability of (CAG)$_n$ trinucleotide repeats (Brock et al., 1999). The human \textit{ATXN3} gene has a particularly low %GC surrounding its polymorphic repeat compared to other polyglutamine disease genes. This is also true for the murine sequence surrounding exon 10. Moreover, the difference in %CG content between the positive selection cassette and the flanking arm sequences are sufficient to predict creation of a CpG island \textit{in silico} (cpgplot analysis, http://www.ebi.ac.uk/Tools/emboss/cpgplot/index.html, data not shown). The majority
of polyglutamine knock-in models created using G418 resistance for positive selection have used electroporation to express recombinase in ES cells, causing positive selection cassette excision prior to blastocyst injection (Lorenzetti et al., 2000, Yoo et al., 2003, Yu et al., 2006b, Watase et al., 2008). One research group that retained the PGK-Neo cassette in living animals documented severe developmental defects in homozygous \( \text{Hdh}^{\text{neoQ50}}/\text{Hdh}^{\text{neoQ50}} \) mice, presumably due to a loss of murine huntingtin protein expression (White et al., 1997); however, the \( \text{Hdh}^{Q71} \), \( \text{Hdh}^{Q94} \), and \( \text{Hdh}^{Q140} \), lines were maintained in the presence of a 5’ PGK-Neo cassette in the reverse orientation, with no phenotype in the Q71 line (Levine et al., 1999, Menalled et al., 2003). In light of this limited data, we considered the hypothesis that inclusion of the positive selection cassette led to enhanced repeat instability in the germline of chimeric animals. Based on this hypothesis, we would expect that any germline transmission event occurring in the presence of the positive selection cassette would result in similar low rates of viable hemizygotes, including transmission from F1 animals.

However, in light of observations made during our ES cell screening, we also considered a second possibility. During our Southern blot analyses, we observed that the signal intensity of the knock-in band relative to the signal intensity of the wild type band was lower for ES colony 3C8 than for the other successfully targeted clones (See Figure 25B). Consequently, we formulated the alternate hypothesis that the low rate of germline transmission from chimeras might reflect non-clonality of the 3C8 colony – i.e. that wild type and successfully gene targeted ES cells grew together to form a single visible colony at the time of G418-resistant colony picking. Based on this hypothesis, we would expect that germline transmission from F1 hemizygous animals would result in roughly 50:50 Mendelian ratios of wild type to hemizygous offspring, despite the presence of an intact positive selection cassette.

In order to test these competing hypotheses, we set up a limited number of F1 hemizygous knock-in x C57BL/6J wild type crosses, and compared the number of wild
type progeny to the number of hemizygous pups containing the knock-in allele with an intact positive selection cassette. Genotyping of 30 F2_C57BL/6 pups revealed much better rates of germline transmission, with 17 wild type animals (57%) and 13 hemizygous knock-in animals (43%). Moreover, the knock-in specific bands produced during the expansion length-dependent genotyping assay did not vary appreciably in size between these progeny (Figure 27). Together, these observations refute the hypothesis that low rates of germline transmission from chimeric animals are due to positive selection cassette-dependent enhancement of repeat instability, and support the hypothesis that the 3C8 colony of ES cells was non-clonal.

F1 knock-in x FLPe transgenic matings resulted in mosaic excision of the positive selection cassette

Despite the fact that the presence of the positive selection cassette did not appear to cause a toxic or reproductive phenotype, we wanted to minimize any potential disruptions in the sequence of intron 9 – 10. To excise the positive selection cassette, we mated F1 hemizygous knock-in mice with homozygous FLPe(tg/tg) animals to promote enhanced Flip recombinase (FLPe)-mediated excision of the FRT-PGK-Neo-FRT cassette (leaving only a single FRT site). The anticipated Mendelian ratios were observed in these matings (~ 50:50 for knock-in:wildtype alleles and 100% FLPe(tg/-)). Next, to confirm that the FLPe recombinase worked properly in the F2_{FLPe} progeny, we took advantage of an XbaI restriction site within the FRT sequence that is absent in wild type intron 9-10 (Figure 28A). Combining Xbal digestion with PCR amplification from MJD4F6 to MJD10R4 (as in our standard positive selection cassette screening assay), we could distinguish animals containing an intact positive selection cassette (3.3 kb band) from animals containing the FLPe-recombined FRT-only sequence (2 smaller bands produced upon digestion of the 1.3 – 1.45 kb amplicon with Xbal) and from mosaic animals that exhibited both of these features. All F2_{FLPe} hemizygotes were mosaic for
FLPe-mediated recombination in genomic tail DNA samples at weaning (Figure 28B). These animals are currently being crossed to wild type C57BL/6J animals to produce founders for the final knock-in line that are FLPe non-transgenic and heterozygous for an expanded Atxn3(Q3KQ82) allele in which the selection cassette has been excised (FLPe(-/-), FRT-only Atxn3(Q3KQ82)(+/-) F3 pups).

Expanded Q3KQ82 Ataxin-3 Protein is expressed in F2 mosaics and F3 founders

To assess expression of the Atxn3(Q3KQ82) protein in knock-in mice, we performed Western blot analysis on lysates from various tissues of F2 FLPe Atxn3(Q3KQ82) knock-in mice and wild type littermate controls (Figure 29). Using ataxin-3 specific mAb 1H9, expanded Atxn3(Q3KQ82) knock-in protein is detectable in brain, heart, liver, skeletal muscle and splenic tissue, but not in renal tissue. In initial analyses of these positive selection cassette mosaic animals, Atxn3(Q3KQ82) protein is expressed at slightly lower levels than unexpanded murine ataxin-3. Expanded Atxn3(Q3KQ82) protein is also detectable with the conformation-specific 1C2 mAb, which specifically detects expanded polyglutamine domains in proteins. The lack of a 1C2-recognizable lower molecular weight ataxin-3 isoform in the kidney suggests that either (i) the 1H9 positive, kidney-specific ataxin-3 isoform does not contain a polyglutamine expansion or (ii) the presence of the trinucleotide expansion alters the expression and/or stability of this putative kidney-specific splice isoform. Both possibilities open avenues for further investigation, and highlight the importance of this knock-in model for the study of protein context-specific aspects of polyglutamine expansion within ataxin-3.

Discussion

In this chapter, I described the creation of a knock-in mouse model of the polyglutamine disease, SCA3. I produced a targeting vector capable of replacing the
murine (CAG)$_6$ in exon 10 of the $Atxn3$ gene with an expansion at the upper limit of the size range seen in humans with SCA3. I compared two published methods of increasing the length of (CAG)$_n$ expansions in vitro. I also compared gene targeting efficiency and chromosomal stability between a novel, pure C57BL/6J ES cell line and two existing, commonly used ES cell lines. Two chimeras achieved germline transmission of the targeted Q$_3$KQ$_{82}$ expansion, and multiple F2 mice exhibited mosaic excision of the positive selection cassette in vivo. These $F_{2FLPe}$ hemizygous transgenic heterozygous knock-ins, are being bred to wild type C57BL/6J animals to create FLPe(-/-), FRT-only $Atxn3(Q_3KQ_{82})$ knock-in mice that represent founders for the first murine knock-in model of SCA3 (summarized in Figure 30).

Despite a number of technical challenges prior to initial germline transmission of the $Atxn3(Q_3KQ_{82})$ allele, the subsequent Mendelian transmission rates, successful mosaic excision of the positive selection cassette, and expression of the $Atxn3(Q_3KQ_{82})$ protein in mosaic F2 animals all suggest that we should be able to rapidly generate an $Atxn3(Q_3KQ_{82})$ knock-in line suitable for further studies of polyglutamine expansion in a disease protein, expression level, tissue, and splice isoform-specific context. The expansion in this knock-in model is at the very high end of the range of repeat lengths seen in humans; thus, it is possible that the mouse will recapitulate motor and neuropathological features of SCA3. Yet, knock-in models of other polyglutamine diseases suggest that recapitulation of the severest motor and degenerative features of the human diseases requires the use of hyperexpanded repeats that are much longer than the normal disease range in humans. Given this, it is unlikely that our SCA3 knock-in mouse model will produce an early onset, robust behavioral phenotype suitable for therapeutic intervention studies. It should, however, be ideal for studies of somatic and intergenerational repeat instability, trinucleotide expansion-dependent alterations in alternative splicing, subtle dysfunction of vulnerable neuronal populations, genetic
interactions, and alterations in the macromolecular complexes formed by the disease protein.

Unfortunately, repeat instability events, including contractions of the (CAG)$_{\text{expanded}}$ repeat length and interruptions of long (CAG)$_{n}$ repeats with portions of normally repeat-flanking sequences, complicated targeting vector construction and precluded the creation of a hyperexpanded knock-in line during the course of these studies. Although disappointing, this result was not unprecedented. Expanded (CAG)$_{n}$ repeats within their normal genomic context are not only unstable within mammalian cells (Libby et al., 2003), but also in vitro and within E. coli. Contraction of expanded repeats during the cloning of a targeting vector was also encountered during the creation of SCA1 and SCA7 knock-in models (Lorenzetti et al., 2000, Watase et al., 2008). A large expansion of a (CAG)$_{78}$ repeat serendipitously generated the template for the Atxn1$^{154Q}$ knock-in, despite judicious use of the recombinase negative SURE strain during the bacterial propagation of these constructs (Watase et al., 2002). Similar cautious culture accompanied by rigorous screening for alteration in (CAG)$_{n}$ repeat size could be used in the future to generate larger expansions in the Atxn3 targeting vector, in order to create a hyperexpanded knock-in mouse model.

In light of the intergenerational repeat instability observed in other knock-in models of polyglutamine diseases, with a tendency towards anticipation in the male germline, it is conceivable that drift in repeat size within the Atxn3(Q$_{3KQ}_{82}$) line could result in rare hyperexpanded alleles leading to more severely affected animals. A similar phenomenon has been documented in Hdh$^{Q_{150}}$ knock-in mice, leading to mice with an expansion of (CAG)$_{200}$ or greater that recapitulate many disease features of human HD (Heng et al., 2009). Because it is likely that the Atxn3(Q$_{3KQ}_{82}$) line will exhibit similar intergenerational repeat instability, all progeny will be evaluated for large repeat expansions and contractions. Any pups exhibiting expansions $>100$ (CAG)$_{n}$ repeats in the Atxn3 locus will be separated as founders of a hyperexpanded line. We will use this
strategy, in addition to utilizing male knock-in animals in breeding pairs to encourage
anticipation, as a key strategy for generating animals with larger repeats in order to
achieve more severe phenotypes in subsequent generations of knock-in mice. If no large
expansions occur through targeted knock-in breeding, a second hyperexpanded line could
be generated by further expanding the (CAG)_n in a new targeting vector via the modified
QuickChange approach described above, through the splicing by overlap extension
approach starting with a smaller template, prior to a second round of ES cell
electroporation, blastocyst injection, and breeding. Ideally, the creation of both a knock-
in line with a repeat expansion at the upper level of the disease repeat range and a line
with a suprapathophysiological expansion will provide critical tools to the SCA3 field
with which to elucidate key pathogenic events leading to selective neuronal toxicity and
to conduct preclinical therapeutic trials.

Our comparison of gene targeting in 129- and C57BL/6-derived ES cell lines
yielded mixed results. Because they were the only ES cells to yield germline transmitting
chimeras, Bruce4.G9 ES cells proved superior to the other two ES lines used in this
study. It may appear surprising that we observed higher rates of gene targeting in the
C57BL/6-derived ES cell lines, when 129 substrain-derived ES cells are popular partly
because of the general consensus that they have favorable rates of homologous
recombination and blastocyst colonization. It is well documented, however, that
incomplete homology of the 5’ and 3’ flanking arms to the endogenous murine sequence
can greatly reduce targeting efficiency (Deng and Capecchi, 1992). The superior
performance of the Bruce4.G9 and UMB6JD7 cells likely reflects the fact that I
generated the 5’ and 3’ flanking arms of the targeting vector using genomic DNA from
the albino C57BL/6 line, B6(Cg)-Tyr<1-2J>/J. The high targeting efficiency, low plating
efficiency, and reduced rates of aneuploidy that we observed in Bruce4.G9 ES cells,
compared to the W4 and UMB6JD7 cells, are defining features of this subclone, relative
to the parent Bruce4 ES cell line. Unfortunately, the new in-house strain of pure
C57BL/6J-derived ES cells only outperformed the Bruce4.G9 line in plating efficiency, showing lower rates of successful homologous recombination (1.6% vs. 2.1%), and greater chromosomal instability, including chromosome duplications, deletions, and translocations. The only UMB6JD7 clone acceptable for blastocyst microinjection did not produce any chimeric animals, perhaps due to this tendency towards aneuploidy.

Although several of our observations while creating knock-in models on a C57BL/6 background were consistent with previous reports of the strengths and weaknesses of the Bruce4 line, a few were unanticipated. The percentage of blastocyst injections resulting in chimeric animals was approximately 10% (10 chimeras in 2 rounds of injections, with 50 blastocysts each), well within the range of 4 – 13% previously reported (Seong et al., 2004). Similarly, our average % coat chimerism was very near the reported average (30.5% versus 28%); however, we saw a much broader range of percentage coat chimerism than has been previously reported for Bruce4 cells (5-80% versus 10-40%). This is consistent with both the general trend to see lower coat chimerism in C57BL/6 lines in general and the reported variability among these ES cell lines ((Ware et al., 2003, Seong et al., 2004, Keskintepe et al., 2007). Previous reports of gene targeting in C57BL/6 ES cells have also suggested that there is less of a strict correlation of % coat chimerism with the likelihood of germline transmission, compared to 129-derived strains, with reports of germline transmission in chimeras with 10-30% coat chimerism (Seong et al., 2004). In our studies, however, we did not see germline transmission of the knock-in allele in any chimera exhibiting < 40% coat chimerism. It is not clear whether this is an anomaly or a feature of the Bruce4.G9 subclone. While our observation that ~25% of F1 offspring produced were of ES cell origin is within the spectrum of previously reported germline ES cell transmission, both our percentage of germline transmitting chimeras and our percentage of ES-cell derived knock-ins were surprisingly low.
Because 2 of 3 chimeras with ≥40% coat chimerism died of unknown causes before 6 months of age, we were initially concerned that our targeted insertion of a (CAG)_n expansion adjacent to an intact FRT-PGK-Neo-FRT cassette might be causing toxicity, leading to reduced viability in high percentage chimeras and either reduced fertility in the high percentage chimeras or reduced viability of F1 hemizygous knock-ins. Based on the Mendelian ratios observed in F1 x C57BL/6J and F1 x FLPe(tg/tg) crosses, however, we concluded that the cause of the low germline transmission rates observed in chimeras were not due to enhancement of repeat instability induced by the CG content of the positive selection cassette. While we cannot rule out the possibility that some combination of our knock-in allele with non-heritable Bruce4.G9 ES cell-specific factors resulted in this phenomenon (e.g. epigenetic dysregulation or chromosomal instability during embryogenesis), the low germline transmission rate is likely explained by non-clonality of the targeted knock-in containing ES cell colony 3C8. Importantly, because these issues have not been encountered after the initial germline transmission events, they should not impede the establishment or maintenance of the Atxn3(Q3KQ82) knock-in line.
Table 6. Existing knock-in models of polyglutamine diseases.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Human Disease Repeat Length</th>
<th>Repeat Length Knock-in Lines</th>
<th>Strain</th>
<th>Phenotype</th>
<th>Selection</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBMA</td>
<td>40 - 62</td>
<td>Control: 21</td>
<td>129S1 /SvImJ - F2 C57BL/6J backcross</td>
<td>In Q113 knock-in line: Decreased fertility, androgen insensitivity, testicular atrophy, hormone-dependent neuromuscular pathology, pelvic floor myotonia with functional urinary obstruction, azotemia and hyperkalemia, death by 2-4 months</td>
<td>Floxed PGK-Neo cassette, removed by Cre in ES cells</td>
<td>Yu, 2006a; Yu 2006b</td>
</tr>
<tr>
<td>HD</td>
<td>36 - 121</td>
<td>Control: Q20, Disease Model: Q50 Q92, Q111 Other: NeoQ50</td>
<td>129/SvEv - CD1 backcross =&gt; mixed background</td>
<td>In Hdh^{Q92} and Hdh^{Q111} knock-in lines: Repeat instability, in striatal MSNs – shift to nuclear protein localization, N-terminal inclusion/aggregate formation In homozygous Hdh^{NeoQ50}: Reduced Hdh expression, defects in neurogenesis Note – all lines contain a hybrid Hdh:HTT exon1</td>
<td>Floxed PGK-Neo cassette, removed by zygotic Cre injection or mating of hemizygotes to Cre(tg) mice</td>
<td>White, 1997; Wheeler, 1999; Wheeler, 2000</td>
</tr>
<tr>
<td>Disease Model: Q72 Q80</td>
<td>129/Sv/ter - &amp; 129/SvJ - C57BL/6 backcross =&gt; mixed background</td>
<td>In Hdh^{Q72,80} Germline repeat instability, aggressive behavior, reduced brain weight without characteristic neuropathology</td>
<td>Floxed PGK-Neo cassette, removed by Cre in ES cells and by mating of hemizygotes to Cre(tg) mice for line creation</td>
<td>Shelbourne, 1999; Kennedy, 2005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disease Model: Q71 Q94 Q140</td>
<td>129/Sv C57BL/6 =&gt; mixed background</td>
<td>In Hdh^{Q94} Motor phenotype: hyperactive by 2 mo, hypoactivity by 4-6mo, striatal microaggregates and inclusions In Hdh^{Q140} Motor Phenotype: early hyperactivity, late hypoactivity by 4mo, reduced stride length, broad nuclear localization, neuropil aggregates or intranuclear inclusions in striatum, olfactory system, nucleus accumbens, cortex, and globus pallidus Note – all lines contain a hybrid Hdh:HTT exon1</td>
<td>PGK-Neo cassette intact in the disease model (reverse orientation)</td>
<td>Levine, 1999; Menalled, 2002; Menalled, 2003</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 6.  Continued.

<table>
<thead>
<tr>
<th>Disease Model:</th>
<th>HD (cont’d)</th>
<th>HD</th>
<th>DRPLA</th>
<th>SCA1</th>
<th>SCA2</th>
<th>SCA3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q80, Q150</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>129/Ola - C57BL/J6 backcross =&gt; mixed background</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Hdh^{Q150} knock-in line:</td>
<td>Motor task deficits, gait abnormalities, striatal neuronal intranuclear inclusions and reactive gliosis rare seizures incompletely penetrant weight loss</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Hdh^{Q80} knock-in line:</td>
<td>Milder but detectable motor phenotype, rare striatal intranuclear inclusions, incompletely penetrant weight loss</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-phase in HPRT(-/-) ES cells</td>
<td>(i) insertion of HPRT minigene (ii) replacement of HPRT minigene with Hdh^{Q80} or Hdh^{Q150}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lin, 2001; Heng, 2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49 - 88</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Two populations: (i) 129/SvEv (ii) Mixed 129/SvEv - C57BL/6J</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minor repeat instability in hemizygotes Rotorod defects in mixed background homozygotes only no neuropathological changes at 18 mo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Floxed PGK-Neo cassette, removed by Cre in ES cells</td>
<td>Lorenzetti, 2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disease Model: Q78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>129/SvEv - F1, F2 C57BL/6J backcross</td>
<td>Progressive motor incoordination, cognitive deficits, wasting, premature death, Purkinje cell loss, age-related hippocampal synaptic dysfunction, somatic repeat instability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Floxed PGK-Neo cassette, removed by Cre in ES cells</td>
<td>Watase, 2002; Watase, 2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lorenzetti, 2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 - 207</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>SCA2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52 - 86</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>
Table 6. Continued.

| SCA6 | 20 - 33 | Control: Q14 | 129/SvEv variably backcrossed => 129/SvEv - C57BL/6J mixed background for some assays | In Sca6\textsuperscript{Q14} knockout line: Progressive motor impairment, aggregation of mutant channels, changes in alternative splicing patterns. Note: Inclusion of any polyglutamine repeat changes channel properties, but there are no differences in channel properties between Sca6\textsuperscript{Q14} – Sca\textsuperscript{Q84}. Note: Mouse Cacna1a has alternate splice acceptor sites like human; but no (CAG)\textsubscript{n} => they put the (CAG)\textsubscript{n} into the variant (humanized); | Floxed PGK-Neo cassette, removed by Cre in ES cells | Watase, 2008 |
| SCA7 | 34 - 306 | Disease Model: Q266 | 129/SvEv - F2 C57BL/6J backcross | Weight loss, ptosis, visual impairment, ataxia, muscle wasting, kyphosis, tremor, premature death dysregulation of the ubiquitin proteasome system late in disease | Floxed PGK-Neo cassette, removed by Cre in ES cells | Yoo, 2003; Bowman, 2005 |
| SCA17 | 47 - 63 | n/a | n/a | n/a | n/a | n/a |

Note: Knock-in mouse models of several polyglutamine diseases have been created. This table summarizes the construction and initial phenotypic characterization of these lines.
Table 7. PCR-based methods for expansion of the (CAG)$_n$.

| MegaPrimer PCR                  | **Primers:** | MjdMH-F - 5’ ATA TTC ACG TTT GAA TGT TTC AGG CAG CAG CAA AAG CAG CAA CAG C 3’  
|                                | MjdHM-R - 5’ GTC TAC CTC CTG CTG CTG CTG CTG CTG CTG CTG C 3’  
| **(megaprimer) Template:**     | At3-Q129-GFP or At3-Q166-GFP  
| **Cycling Conditions:**        | 1. 95°C for 2min  
|                                | 2. (95°C for 45s, 63°C for 45s, 72°C for 30s) x 35  
|                                | 3. 72°C for 10 min  
|                                | 4. Hold at 4°C  
| **Products:**                  | A smear of megaprimers, sized from ~ 100 – 500 bp  
| Reaction 2                     | **Primers:** | Reaction 1 megaprimers, binned by size into  
|                                | High ~ 250 – 500 bp  
|                                | Med ~ 100 – 250 bp  
|                                | Low ~ 100 bp  
| **(QuickChange) Template:**    | pBl/MJD10Q  
| **Cycling Conditions:**        | 1. 95°C for 30s  
|                                | 2. (95°C for 30s, 55°C for 30s, 68°C for 10 min) x 18  
|                                | 3. Hold at 4°C until DpnI digestion  
| **Products:**                  | A smear of products, sized from < 500 bp to > 6 kb  
| Laccone Method                 | **Primers:** | MJDKpnA - 5’ GGT ACC TGA GCT CTC ATT GCT TAC AGA CC 3’  
|                                | (CTG)$_n$ – 5’ CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG 3’  
| **Template:**                  | pBl/MJD10Q #71/KpnI/XhoI  
| **Cycling Conditions:**        | 1. 95°C for 4min  
|                                | 2. (95°C for 20s, 58°C for 30s, 72°C for 2 min 20s) x 35  
|                                | 3. 72°C for 10 min  
|                                | 4. Hold at 4°C  
| **Products:**                  | ≥ 500 bp upstream flanking arm + trinucleotide repeat (f)  

Table 7. Continued.

<table>
<thead>
<tr>
<th>Reaction 1(r):</th>
<th>Primers: MJDRIB – 5’ GTT TGC CGT GTT GAT GGG AAG G 3’ (CAG)$_n$ – 5’ CAG CAG CAG CAG CAG CAG CAG CAG CAG 3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template:</td>
<td>pBl/MJD10Q #71/XmaI/EcoRI</td>
</tr>
<tr>
<td>Cycling</td>
<td>1. 95ºC for 4min</td>
</tr>
<tr>
<td>Conditions:</td>
<td>2. (95ºC for 20s, 58ºC for 30s, 72ºC for 2 min 20s) x 35</td>
</tr>
<tr>
<td></td>
<td>3. 72ºC for 10 min</td>
</tr>
<tr>
<td></td>
<td>4. Hold at 4ºC</td>
</tr>
<tr>
<td>Products:</td>
<td>≥ 1.5 kb upstream flanking arm + trinucleotide repeat (r)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reaction 2: (SOE)</th>
<th>Primers/Template: f + r f + r (SOE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycling</td>
<td>1. (95ºC for 45s, 60ºC for 45s, 72ºC for 4 min) x 35</td>
</tr>
<tr>
<td>Conditions:</td>
<td>2. 72ºC for 10 min</td>
</tr>
<tr>
<td></td>
<td>3. Hold at 4ºC</td>
</tr>
<tr>
<td>Products:</td>
<td>≥ 2.25 kb</td>
</tr>
</tbody>
</table>

Colony PCR (CAG)$_n$ Expansion Screen

| Primers: MJD10F1 - 5’ CTA CGT GAT TGC CTG CTC TCC 3’ MJD10R3 - 5’ CAG GCA TAC ATA CAG ACA GAG C 3’ |
| Cycling       | 1. 95ºC for 5 min                                                                                    |
| Conditions:   | 2. (95ºC for 45s, 52ºC for 45s, 72ºC for 45s) x 30                                                   |
|               | 3. 72ºC for 10 min                                                                                   |
|               | 4. Hold at 4ºC                                                                                        |
| Products:     | 560 bp for wild type (CAG)$_b$ > 560 bp for expanded (CAG)$_n$                                      |

Note: Two methods were used to expand the (CAG)$_b$ trinucleotide repeat within the murine ataxin-3 gene. In the first method, primers with 5’ homology to murine Atxn3 and 3’ homology to the trinucleotide repeat were used to amplify a “megaprimer” from plasmids containing expanded human ATXN3 cDNA. QuickChange mutagenesis with this double stranded megaprimer was used to expand the (CAG)$_b$ repeat sequence within murine Atxn3 exon 10. In the second method, a modestly expanded murine ataxin-3 template was amplified from flanking sequences into the repeat, in both directions, before these flanking sequences were spliced by overlap extension. All clones resulting from either of these methods were screened by colony PCR, using primers flanking the repeat expansion. The details of these PCR reactions are listed.
# Table 8. ES cell lines utilized in gene targeting.

<table>
<thead>
<tr>
<th>ES cell line</th>
<th>Strain Derivation</th>
<th>Description</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>W4</td>
<td>129S6/SvEvTac</td>
<td>Commonly used strain for ES cell targeting.</td>
<td>Auerbach, 2000</td>
</tr>
<tr>
<td>Bruce4.G9</td>
<td>C57BL/6; NZB backcross at N3</td>
<td>Commonly used “C57BL/6” ES cell line; documented issues with chromosomal instability.</td>
<td>Hughes, 2007</td>
</tr>
<tr>
<td>UMB6JD7</td>
<td>C57BL/6J</td>
<td>In house derived experimental ES cell line; pure C57BL/6J background.</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Note: We compared the gene targeting efficacy and chromosomal instability of three ES cell lines after electroporation with the Atxn3(Q$_3$KQ$_{82}$) targeting vector. The well established W4 and Bruce4.G9 ES cells were used as references in order to test a pure C57BL/6J ES cell line created at the University of Michigan Transgenic Animal Core. Details of these strains are presented.
Table 9. PCR conditions for ES and knock-in line screening.

<table>
<thead>
<tr>
<th>5' ES Screen</th>
<th>3' ES Screen</th>
<th>F1 Screen: Knock-in Confirmation</th>
<th>F1 Screen: Positive Selection Cassette Confirmation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primers:</strong></td>
<td><strong>Primers:</strong></td>
<td><strong>Primers:</strong></td>
<td><strong>Primers:</strong></td>
</tr>
<tr>
<td>MJDint9-10F2 - 5’ CTT-GCT-CCA-AGT-CAG-ATG-CAG-C 3’</td>
<td>3’NeoF-2 - 5’ AAG-TAT-AGG-AAT-CTC-TTA-GGA-TCC 3’</td>
<td>MJD10F1 - 5’ CTA CGT GAT TGC TGC CTG CTC TCC TCT 3’</td>
<td>MJD4F6 - 5’ GCC CCT CGC AAT ACA CTT TA 3’</td>
</tr>
<tr>
<td>5’NeoR1 - 5’ TGG-ATG-AGG-AAT-GTG-TGC-GAG-G 3’</td>
<td>MJDRI-B - 5’ GTT-TGC-CGT-GTT-GAT-GGG-AAG-G 3’</td>
<td>MJD10R3 - 5’ CAG GCA TAC ATA CAG ACA GAG C 3’</td>
<td>MJD10R4 - 5’ GGA GAG CAG GCA ATC ACG TAG 3’</td>
</tr>
<tr>
<td><strong>Cycling Conditions:</strong></td>
<td><strong>Cycling Conditions:</strong></td>
<td><strong>Cycling Conditions:</strong></td>
<td><strong>Cycling Conditions:</strong></td>
</tr>
<tr>
<td>1. 94ºC for 30s</td>
<td>1. 94ºC for 30s</td>
<td>1. 95ºC for 2min</td>
<td>1. 95ºC for 2min</td>
</tr>
<tr>
<td>2. (94ºC for 15s, 62ºC for 30s, 68ºC for 6.5 min) x 35</td>
<td>2. (94ºC for 15s, 63ºC for 30s, 68ºC for 3 min) x 35</td>
<td>2. (95ºC for 30s, 51ºC for 30s, 72ºC for 1 min) x 30</td>
<td>2. (95ºC for 30s, 60ºC for 30s, 72ºC for 4 min) x 30</td>
</tr>
<tr>
<td>3. 72ºC for 10 min</td>
<td>3. 72ºC for 10 min</td>
<td>3. 72ºC for 10 min</td>
<td>3. 72ºC for 10 min</td>
</tr>
<tr>
<td>Hold at 4ºC</td>
<td>Hold at 4ºC</td>
<td>Hold at 4ºC</td>
<td>Hold at 4ºC</td>
</tr>
<tr>
<td><strong>Products:</strong></td>
<td><strong>Products:</strong></td>
<td><strong>Products:</strong></td>
<td><strong>Products:</strong></td>
</tr>
<tr>
<td>4.7 kb = intact 5’ flanking arm</td>
<td>~ 1 kb = wild type (nonspecific endogenous control band)</td>
<td>555 bp = wild type Atxn3(Q6)</td>
<td>1.33 kb = wild type</td>
</tr>
<tr>
<td>No band in wild type animals</td>
<td>2.98 kb = knock-in Atxn3(Q3KQ82)</td>
<td>795 bp = knock-in Atxn3(Q3KQ82)</td>
<td>3.30 kb = intact selection cassette (FRT-PGK-Neo-FRT)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.43 kb = excised selection cassette (FRT only)</td>
</tr>
</tbody>
</table>
Table 9. Continued.

<table>
<thead>
<tr>
<th>FLPe Screen</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primers:</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Cycling Conditions:</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Products:</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Note: The methodological details of key PCR reactions used to screen for homologous recombination in ES cells, germline transmission to F1 – F3 pups, and FLP-recombinase mediated excision of the positive selection cassette are listed.
Table 10. Assessment of euploidy and aneuploidy in homologously recombined ES cell populations.

<table>
<thead>
<tr>
<th>Clone</th>
<th>Euploid</th>
<th>Aneuploid</th>
<th>Injected?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% 40</td>
<td>% 39</td>
<td>% 41</td>
</tr>
<tr>
<td>W4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Bruce4.G9</td>
<td>3C8</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3H11</td>
<td>70</td>
<td>15</td>
</tr>
<tr>
<td>UMB6JD7</td>
<td>6C4</td>
<td>65</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>6H9</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>5F1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: Chromosome spreads were used to assess the ploidy of 10 – 25 individual ES cells positive for Atxn3(Q3KQ82) gene targeting. At least 20 individual cells were assessed for all clones used in downstream blastocyst injections. Because no homologous recombinants were identified in the electroporated W4 ES cells, no colonies were assessed. Among the C57BL/6 derived strains, UMB6JD7 exhibited higher degrees of aneuploidy, including chromosomal duplications, deletions, and translocations (trans). All colonies with detectable euploid cells were chosen for blastocyst injection.
Table 11. Summary of chimera production and breeding.

<table>
<thead>
<tr>
<th>Blastocyst Injection</th>
<th>Chimera</th>
<th>% Chimerism</th>
<th>White pups</th>
<th>Black pups</th>
<th>Knock-ins</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round 1</td>
<td>1.1</td>
<td>30</td>
<td>102</td>
<td>61</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Round 2</td>
<td>2.1</td>
<td>80</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>DIP, 1 mo</td>
</tr>
<tr>
<td></td>
<td>2.2</td>
<td>60</td>
<td>138</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.3</td>
<td>40</td>
<td>7</td>
<td>9</td>
<td>3</td>
<td>DIP, 5 mo</td>
</tr>
<tr>
<td></td>
<td>2.4</td>
<td>20</td>
<td>61</td>
<td>64</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>20</td>
<td>46</td>
<td>20</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.6</td>
<td>20</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.7</td>
<td>10</td>
<td>63</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.8</td>
<td>20</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.9</td>
<td>5</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Note: Eleven chimeras were produced from two rounds of blastocyst injections with clone 3C8. These animals ranged from 5 – 80% coat chimerism. Notably, only chimeras exhibiting 20% or higher coat chimerism produced ES cell-derived pups, and only animals exhibiting 40% or higher coat chimerism achieved germline transmission of the knock-in allele. Two of the three high percentage chimeras were found dead in pan (DIP) before 6 months of age.
Figure 20. Comparison of 129- and C57BL/6-derived ES cells for gene targeting by homologous recombination.

Although 129-derived ES cells are by far the most commonly used lines for transgenic and targeted gene mutation technologies, C57BL/6-derived strains are being increasingly utilized. This figure summarizes some of the key advantages and disadvantages of each line at each stage in the creation of a novel knock-in model.

Figure 21. Overview of targeting vector construction.
(A) Schematic of the parental targeting vector, pBY49a. (B) 5’ upstream and 3’ downstream flanking arms were amplified from B6(Cg)-Tyr<−c-2J>/J ES cells. The 5’ flanking arm was subcloned into pBY49a without further modification. The murine (CAG)$_n$ was expanded before subcloning into the final targeting vector.
Figure 22. Expansion of the trinucleotide repeat using the megaprimer method. (A) Schematic of the megaprimer method for expansion of the murine Atxn3(CAG)\textsubscript{6} repeat. (B) Amplification of a population of megaprimers from Q129 and Q166 human ataxin-3 cDNA templates using 5' murine specific / 3' human specific primers. (C) Quick-change incorporation of increasingly large megaprimer populations (determined by band size) into murine Atxn3, exon10. (D) Representative colony PCR reactions. Primers flanking the trinucleotide expansion were used to detect expanded clones for further analysis. (E) A summary of sequenced clones illustrates the variety of repeat lengths and sequence modifications generated using this method.
Figure 23. Laccone method for expansion of the $Q_3KQ_{82}$-encoding repeat sequence.
A) Schematic of the Laccone method for polyglutamine expansion, showing amplification from non-repetitive flanking sequences into the trinucleotide repeat separately in both directions, prior to splicing by overlap extension (SOE). (B) Reactions “f” and “r” generate populations of amplicons greater or equal to the length of the repeat-flanking sequences. (C) SOE amplification of these products generates a broad smear of products comprising the entire 3’ flanking arm of the proposed targeting vector, with
Figure 23. Continued.
variable numbers of trinucleotide repeats, which were gel purified, cloned, and sequenced. (D) Representative DNA sequence analysis demonstrating that while this method generated some very large repeats, every clone sequenced contained multiple distinct repeat lengths, suggesting either repeat instability, recombination, and contraction during bacterial culture, or concatenation of the repeat-containing sequences during cloning.
Figure 24. Targeting exon 10 of the murine Atxn3 gene for homologous recombination to insert a trinucleotide expansion.

Schematic representation of exons 8 through 11 of the murine Atxn3 gene, and the modifications they undergo during targeted gene mutation through homologous recombination with the Atxn3(Q3KQ82) targeting vector, and during subsequent FLPe recombinase-mediated excision of the positive selection cassette.
Figure 25. Five ES cell colonies showed evidence of targeted gene insertion.

(A) Schematic of screening strategies. PCR-based strategies utilized primers within the positive selection cassette with primers outside of the 5' and 3' flanking arms to assess homologous recombination at each site. Southern blot-based strategies were designed to confirm homologous recombination at the 3' flanking arm (AflII) and the presence of the trinucleotide expansion (SpeI). Five clones were confirmed to be positive for homologous recombination and inclusion of the Q3KQ82 repeat by (B) Southern blot and (C) PCR. Dashed arrows indicate the wild type control bands, and solid arrows indicate bands corresponding to the expanded knock-in allele.
Figure 26. Chimeras exhibited a very low rate of germline transmission.

(A) Schematic representation of genotyping strategies used to detect germline transmission of the Atxn3(Q₃KQ₈₂) knock-in allele to F1 progeny. (B) PCR-amplification of mouse genomic tail DNA from F1 pups with the primer pairs MJD10F1 – MJD10R3 and MJD4F6 – MJD10R4. Evidence of the repeat expansion and the positive selection cassette is visible in a small subset of F1 pups (*). Knock-in specific bands are indicated with arrows and endogenous control bands are indicated with brackets. (C) Only 2.5% of black F1 pups (derived from gametes of cultured ES cell origin) were hemizygous knock-ins containing an intact positive selection cassette.
Figure 27. Germline transmission of the Atxn3(Q3KQ82) knock-in allele with an intact positive selection cassette followed Mendelian ratios in F1 animals. (A) Schematic representation of genotyping strategies used to detect germline transmission of the Atxn3(Q3KQ82) knock-in allele to F2 progeny from F1 x C57BL/6J crosses. (B) PCR-amplification of mouse genomic tail DNA from F2 pups with the primer pairs MJD10F1 – MJD10R3 and MJD4F6 – MJD10R4. Evidence of the repeat expansion and the positive selection cassette is visible in approximately half of these animals (*), suggesting that the presence of the repeat expansion in the context of an intact positive selection cassette does not reduce viability of knock-in embryos or pups. Knock-in specific bands are indicated with arrows and endogenous control bands are indicated with brackets. (C) Nearly half (43%) of F2 progeny from F1 x C57BL/6J crosses were hemizygous knock-ins containing an intact positive selection cassette.
Figure 28. Mosaic excision of the positive selection cassette in FLPe(tg/-) F2 Atxn3(Q3KQ82) knock-in animals.
(A) Schematic representation of genotyping strategies used to detect the Atxn3(Q3KQ82) knock-in allele and excision of the positive selection cassette in F2 progeny from F1 x FLPe(tg/tg) crosses. (B) PCR-amplification of mouse genomic tail DNA from F2 pups with the primer pair MJD10F1 – MJD10R3 identifies animals containing the
Figure 28. Continued.
trinucleotide repeat expansion (*). PCR-amplification with the primers FLPe1 and FLPe2 confirm that all F2 animals from these crosses are FLPe(tg/-) (tg). The presence of a 3.3 kb band in the MJD4F6 – MJD10R4 reactions indicates the presence of an intact positive selection cassette in knock-in animals; however, digestion of the lower molecular weight MJD4F6 – MJD10R4 amplicon with XbaI confirms the presence of FRT-only alleles, suggesting that these animals are mosaic for FLPe recombination.
Figure 29. Atxn3(Q3KQ82) knock-in animals express expanded ataxin-3 protein. Western blot analysis of tissues from an Atxn3(Q3KQ82) hemizygous knock-in animal (mosaic for the positive selection cassette) and its wild type littermate control. Selected tissues include brain (B), heart (H), kidney (K), liver (L), and spleen (S). Left: 1H9 mAb recognizes both endogenous wild type murine ataxn-3 (brackets) and expanded ataxin-3 (large arrow) in select tissues. Right: 1C2-positive Atxn3(Q3KQ82) protein expression (arrow) is detectable in all tissues analyzed except the kidney. Note that expression of Atxn3 protein from the (CAG)6 wild type allele (bracket) does not appear to be greatly reduced by expression of the expanded protein.
Figure 30. Summary of gene targeting and mating efficiencies.
Schematic summary of the strategy for and efficiencies of gene targeting, blastocyst
colonization, germline transmission of the knock-in allele, FLPe-mediated excision of the
positive selection cassette, and generation of F3 colony founders. The Atxn3(Q3KQ82)-
encoding expansion is represented by a red asterisk, and (⊥⊥) markers indicate cells or
animals which are unsuitable for further analysis or utilization.

CHAPTER 4
CONCLUSIONS AND FUTURE DIRECTIONS

The overarching goal of my thesis work was to explore the role of protein context in the polyglutamine disease, SCA3. Toward this end, I have characterized the two most commonly used splice isoforms of ataxin-3 and shown that they are not interchangeable entities, are expressed at very different levels in brain tissue, have different in vivo aggregation properties, and are degraded by different components of the cellular protein quality control system. I have also generated the first knock-in mouse model of SCA3, which demonstrates Atxn3(Q3KQ82) protein expression in hemizygous knock-in animals. This genetically precise model of disease, while still requiring further characterization, should be ideal for studies exploring the effects of genomic, message, and protein context on polyglutamine toxicity in SCA3, and has a reasonable probability of leading to the generation of a hyperexpanded line through intergenerational repeat instability. In this closing chapter, I will discuss some of the questions and concepts raised during the course of these studies, and suggest avenues of future research to address these questions.

The work contained in this dissertation has laid the foundation for a broader understanding of ataxin-3 function and has provided tools for further elucidation of the ways in which the molecular properties of ataxin-3 protein modulate the toxicity of its polyglutamine expansion during disease. The future directions presented in this chapter will build upon these foundations to broaden our understanding of ataxin-3 physiology and polyglutamine disease pathogenesis.

Functional Significance of Tissue-specific Ataxin-3 Splice Variation

Numerous minor ataxin-3 splice variants are expressed at the mRNA level in brain and other tissues (Figures 6 and 7) (Bettencourt et al., 2009). While full length ataxin-3 is clearly the predominant, physiologically relevant isoform expressed in the
brain (Figures 9-11), lower molecular weight Atxn3 protein species are prominent in the heart, kidney, skeletal muscle, and spleen. While it is possible that these species result from posttranslational processing of the full length protein, they also are consistent with the translation of several minor 5’ splice variants that have been demonstrated at the mRNA level. Based on apparent molecular weight and reduced expression levels, the predominant band seen in cardiac and skeletal muscle is consistent with 2UIM Atxn3. Based on apparent molecular weight alone, the predominant isoform in the kidney is consistent with either Atxn3Δ2,3 (i.e. the splice isoform lacking exons two and three), Atxn3Δ3,4,5, or Atxn3Δ2.4.5. Similarly, the spleen appears to express full length 3UIM ataxin-3 as the major isoform with at least two lower molecular weight isoforms consistent with Atxn3Δ2 or Atxn3Δ4,5 and Atxn3Δ2,3, Atxn3Δ3,4,5, or Atxn3Δ2.4.5. Because all of these predicted Atxn3 splice isoforms involve disruption of functional domains within ataxin-3, it will be critically important to determine whether these lower molecular weight 1H9 positive bands represent bona fide splice isoforms that correlate with the previously described variants listed above. If so, it will be important to determine what function these catalytically impaired Atxn3 isoforms may be playing in specific tissues.

To determine whether the lower molecular weight Atxn3 species represent documented or novel splice isoforms, the first step would be to purify, amplify, and sequence mature Atxn3 mRNA transcripts from these tissues, as was done for brain tissue in Chapter 2. If any transcripts are identified which are consistent with the protein isoforms observed, the immunopurified isoforms could be resolved by 1D or 2D PAGE and characterized by MALDI-TOF analysis. Functionally, any isoforms confirmed to be expressed at the protein level could be expressed recombinantly as GST-fusion proteins, purified, and analyzed for ubiquitin binding and in vitro DUB activity, to confirm that disruption of the Josephin domain interferes with these activities, as predicted. Additionally, these isoforms could be transiently overexpressed in mammalian cell
culture, to determine whether they alter the subcellular localization of full length ataxin-3, the levels or distribution of polyubiquitinated proteins within the cell, or levels and HAT activities of ataxin-3:p300/CBP and ataxin-3:PCAF complexes. Transcriptional activities would be particularly important to assess in any DUB-deficient isoforms, as binding of ataxin-3 to these transcriptional activator complexes is dependent on the C-terminal portion of ataxin-3, while the N-terminal portion of ataxin-3 interacts with histones H3 and H4 (Li et al., 2002). Currently it is not known whether histone ubiquitination is required for Josephin domain:histone interactions, whether the DUB activity of ataxin-3 is important for the transcriptional co-repressor activity of ataxin-3, or whether ataxin-3 functionally interacts with ubiquitinated histones H3 and H4 during DNA repair. Naturally occurring DUB-deficient isoforms of ataxin-3 could provide important clues to these questions about the role of ataxin-3 in regulating transcription.

**Determination of UIM3 function**

Although it is a highly conserved motif, the function of ataxin-3 UIM3 has been elusive. It is not required for ataxin-3 binding to free ubiquitin chains or polyubiquitinated proteins, *in vitro* cleavage of the model substrate ubiquitin-aldehyde (Burnett et al., 2003, Berke et al., 2005), or *in vitro* cleavage of defined ubiquitin chains (Figure 12). Moreover, it does not appear to alter the specificity of ataxin-3 cleavage for preferred substrates (Figure 12). Loss of UIM3 function does, however, induce a phenotype when highly overexpressed in mammalian cells. Morphologically, cells transiently overexpressing high levels of the UIM3(SA/DG) mutant of ataxin-3 are more likely to extend cellular protrusions into multiple planes of focus upon inspection at 1000x magnification under oil emersion, compared to the intact 3UIM isoform. Perplexingly, this phenotype is often accompanied by a loss of nuclear and perinuclear Flag-epitope signal, without a loss of 1H9 mAb (ataxin-3) staining (Figure 14). Biochemically, these cells have greatly reduced levels of alpha-tubulin by Western blot
(Figure 31), compared to cells expressing either full length 3UIM or 2UIM ataxin-3. These observations implicate ataxin-3 in cytoskeletal regulation and tubulin stability, and provide the first clues to the function of this highly conserved but poorly characterized domain.

Several lines of evidence already suggest direct and indirect links between tubulin and ataxin-3. First, ataxin-3 interacts directly with both α-tubulin and β-tubulin with a high affinity ($K_d$ 50 – 70 nM), and partially colocalizes with α-tubulin in Cos7 cells (Mazzucchelli et al., 2009). 2UIM Ataxin-3 also interacts with dynein and the microtubule-associated deacetylase HDAC6 to facilitate sequestration of misfolded proteins into centrosomal aggresomes (Burnett and Pittman, 2005). Intriguingly, both ataxin-3 and HDAC6 interact with the microtubule-associated E3 ubiquitin ligase, parkin. Within neurons, parkin localizes in a punctate pattern along microtubules, where it can bind both tubulin dimers and microtubules. Parkin has been shown to ubiquitinate tubulin, leading to enhanced proteasomal degradation, yet it also stabilizes established microtubules (Ren et al., 2003, Yang et al., 2005). Parkin is also recruited to the centrosome during inhibition of the 26S proteasome, in a process that requires parkin:HDAC6 interaction and HDAC6-dependent microtubule deacetylation (Jiang et al., 2008). Ataxin-3 and parkin also directly interact, with the Josephin domain interacting with the same regions required for parkin binding to tubulin and HDAC6, and at least one of the ataxin-3 UIMs interacting with the ubiquitin-like (UBL) domain of parkin. Functionally, ataxin-3 preferentially cleaves K63 linkages within chains formed during parkin autoubiquitination, enriching the K48-only nature of these chains, and enhancing degradation of parkin (Durcan et al., 2008); it may also act to edit ubiquitin chains on parkin substrates. Although centrosomal γ-tubulin, microtubule associated proteins, and cytoskeletal regulatory GTPases can also be regulated by ubiquitination, ataxin-3 binding or DUB activities have not been implicated specifically in these processes. A model of functional interactions between ataxin-3, HDAC6,
dynein/dynactin, parkin, and microtubules, however, may explain the phenotype observed in cells highly overexpressing the UIM3(SA/DG) mutant ataxin-3.

Combining these observations, I propose the following model (Figure 32A). In the presence of wild type ataxin-3, HDAC6 senses the presence of proteasomal stress through a signal, such as the accumulation of appropriately edited K48-linked chains on polyubiquitinated protein substrates, which is dependent on ataxin-3 catalytic activity and UIMs 1 and 2. In response to this signal, HDAC6 forms a complex with dynein/dynactin, parkin, and likely polyubiquitinated substrates, and translocates in a microtubule deacetylation-dependent fashion to the pericentrosomal area to form a compartment enriched in 20S proteasomal components, polyubiquitinated proteins, and ataxin-3 (Kawaguchi et al., 2003, Burnett and Pittman, 2005). This concentration of protein quality control machinery may allow protein:protein interactions between ubiquitin ligases, such as parkin, and chain-editing DUBs, such as ataxin-3, to enhance the efficiency of proteasomal degradation. This sequestration of misfolded proteins and protein degradation machinery can be reversed, if the proteasomal stress is mitigated; however, in the presence of persistent proteasomal stress, this compartment remains compact and is walled off by a vimentin “cage” to form a cytoprotective aggresome.

I hypothesize that the UIM3(SA/DG) mutant acts in at least one of two ways: (1) by increasing the ability of ataxin-3 to activate HDAC6, and (2) by altering the activity of the ataxin-3:parkin protein complex, leading to inefficient autoubiquitination and degradation of parkin (Figure 32B). According to this model, when the UIM3(SA/DG) mutant is overexpressed, HDAC6 is more efficiently activated to recruit parkin to a pericentrosomal domain that is still enriched in protein quality control machinery, but which is not as heavily enriched with misfolded, polyubiquitinated proteins. Within this pericentrosomal compartment, I hypothesize that ataxin-3:parkin complexes are still formed (mediated by the Josephin domain of ataxin-3 and the RIR domain of parkin). It is possible that the ataxin-3:parkin complex does not specifically require UIM3 for
efficient, selective chain editing, but that the increase of HDAC6-dependent microtubule destabilization and accumulation of protein quality control machinery in the absence of ample misfolded protein substrates leads to efficient ubiquitination and degradation of the most abundant parkin substrates available – tubulin dimers. Alternately, it is possible that in the absence of UIM3:parkin-UBL domain interactions, the efficient, selective K63 chain editing activity of the ataxin-3:parkin complex is impaired, leading to a reduction of K48-only autoubiquitinated parkin and reduced parkin degradation. The resulting increase in parkin levels may then increase the degradation of parkin substrates, including tubulin, and increase the stabilization of intact microtubules by parkin. Either of these two mechanisms could lead to the observed loss of perinuclear Flag epitope signal, drastic decreases in total tubulin levels, and persistence of long cellular protrusions in the presence of overexpressed UIM3(SA/DG). It will be important to test this model to determine whether UIM3 has an in vivo function involved in either the sensing of misfolded proteins by HDAC6 or the substrate specificity of ataxin-3 when in physiological protein-protein complexes.

To test the effect of UIM3 mutation on HDAC6-dependent formation of a pericentrosomal protein degradation compartment/aggresome, it would not only be important to observe the subcellular distribution of HDAC6, parkin, 20S proteasomal subunits, and vimentin during transient UIM3(SA/DG) overexpression, compared to 3UIM ataxin-3 overexpression, but also to observe this process in cells depleted of endogenous ataxin-3 and reconstituted as previously described (Burnett and Pittman, 2005) with wild type 3UIM and UIM3-mutant ataxin-3. To test the effect of UIM3 mutation on the activity and linkage specificity of deubiquitination of ataxin-3:parkin complexes, wild type or UIM3-mutant ataxin-3 could be used in modified in vitro DUB reactions using autoubiquitinated parkin (Durcan et al., 2008) or other defined substrates, such as immunopurified ubiquitinated tubulin (Ren et al., 2003), in the presence of parkin. Similarly, reconstituted in vitro systems could also be used to examine the effect
of wild type or UIM3-mutant ataxin-3 on active parkin autoubiquitination and substrate ubiquitination. To directly examine the effects of wild type versus UIM3-mutant ataxin-3 on the tubulin cytoskeleton, immunofluorescence could be used to compare microtubule distribution, acetylation, and stability in the presence or absence of microtubule destabilizing agents, such as colchicine and nocodazole. Note that because serine phosphorylation occurs within ataxin-3 UIMs (Fei et al., 2007, Mueller et al., 2009), an important control to include in all of these experimental systems is a UIM3 mutant, such as UIM3(S/A) or UIM3(SA/AG), that is not capable of serving as a UIM3 serine phosphorylation mimic, in order to demonstrate that these alterations are truly due to disruption of UIM:ubiquitin or UIM:UBL interactions. These studies would aid our understanding of both the functional importance of UIM3 and the role of ataxin-3 in protein quality control systems, including aggresome formation.

**Cell-specific expression of minor ataxin-3 splice variants in affected and adjacent brain regions**

While 3UIM ataxin-3 is clearly the predominant, physiologically relevant splice isoform in brain tissue, and the isoform of choice for any experimental attempts to understand “full length” ataxin-3 function, the aggregation propensity and proteasomal route of degradation of the 2UIM isoform warrant additional characterization. Although we were unable to detect unmodified 2UIM ATXN3 in the brains of MJD15.4 or MJD84.2 animals, it should be noted that a rapidly degraded minor isoform expressed in a limited subpopulation of cells would be difficult, if not impossible, to detect in whole brain or even whole cerebellar lysates. If the disease-associated properties of 2UIM ataxin-3 have any pathophysiological significance, this isoform must be present and enriched in affected populations of cells. To address the hypothesis that enhanced expression of 2UIM ataxin-3 contributes to selective neuronal toxicity, it would be important to determine whether this isoform is enriched at the mRNA level or detectable
at the protein level in affected neuronal populations. *In situ* hybridization would be a reliable method to determine whether selectively vulnerable neuronal populations are enriched for 2UIM ataxin-3 at the message level. Because there are no existing antibodies to specifically detect the 2UIM isoform, and because the hydrophobicity of the domain unique to this isoform makes the successful generation of such a tool unlikely, immunohistochemistry is not a viable approach to confirm results at the protein level. Instead, microdissection followed by 2D-Western analyses that compare affected (pontine and dentate nuclear) and unaffected (cerebral cortical and hippocampal) neuronal populations would be a viable alternative approach.

While it would be ideal to perform these studies in brain tissue from SCA patients and controls, this approach is complicated both by the difficulty of obtaining post-mortem brain tissue with acceptable RNA quality for *in situ* hybridization, and by the technical challenge of 2D-based protein analysis due to non-synonymous ataxin-3 SNP variation within human populations. These studies could also reliably be conducted in MJD15.4 and MJD84.2 mice (to study splicing of the human transgene), and possibly in the Atxn3(Q₃KQ₈₂) homozygous knock-in mice generated here. If the knock-in mice exhibit selective neurodegeneration, MALDI-TOF analysis could supplement a 2D-Western approach, as the murine putative 2UIM isoform has a shorter hydrophobic domain (similar to the 2UIM, short SNP variant), and would have a low enough mass to facilitate a time of flight within the range of detection. Humans with SCA3 and MJD84.2 mice exhibit neurodegeneration in cerebellar dentate and pontine nuclei. In the MJD84.2 mice, 30 – 40% of neurons in these nuclei have been lost by 12 months of age, so it would be important to look both at and before this time point. Therefore, analysis of RNA and protein from selected neuronal populations in 6 month and 12 month old mice would be ideal for these studies. An absence of detectable enrichment of 2UIM-encoding transcript variants and detectable 2UIM protein in these neuronal populations would refute the hypothesis that this minor splice variant contributes significantly to the
selective neuronal toxicity observed in SCA3. In such a scenario, any future research proposals focusing on the 2UIM isoform of ataxin-3 would be considered to have a much lower impact and significance to the field, compared to studies utilizing the 3UIM isoform.

**Splice Isoforms and Protein Context in Other Polyglutamine Diseases**

The prevalence of alternative splicing among polyglutamine disease proteins provides many opportunities to explore the role of alternative splicing in the other polyglutamine diseases. Some of the most straightforward projects would involve those variants that have already been partially characterized. For example, the observation that CACNA1A transcripts containing polyglutamine-encoding splice variants are enriched in the Purkinje cells of SCA6 patients, but not in controls (Tsunemi et al., 2008), needs to be followed up with studies at the protein level. While it is possible that this enrichment at the transcript level is mirrored at the protein level, this transcript enrichment could also be either unrelated to protein levels or compensatory for the selective degradation or sequestration of polyglutamine-containing CaV2.1 isoforms. Although higher numbers of Purkinje cells would need to be captured by microdissection (Ball and Hunt, 2004), a similar comparison of Purkinje cell versus granule cell CaV2.1 protein levels could be performed on existing snap-frozen human tissues (Tsunemi et al., 2008), as well as on tissues from SCA6\(^{Q84}\) versus SCA6\(^{Q14}\) knock-in mice (Watase et al., 2008). Because the isoforms of interest either contain or lack the polyglutamine domain, they exhibit a shift in molecular weight corresponding to the documented repeat length, which can be confirmed by utilization of the polyglutamine-specific monoclonal antibody 1C2 (Ishikawa et al., 2001). Thus, identification of these variants does not require the development of a novel, isoform-specific antibody. If the enrichment of \((\text{CAG})_n\) containing CACNA1A transcripts in diseased Purkinje cells is accompanied by an increase
in polyglutamine containing isoforms, it would provide the strongest evidence to date that alternative splicing results in cell type-specific, pathologically relevant alterations in protein context.

Due to the androgen-dependence of polyglutamine toxicity in SBMA, it is generally thought of as a disease that only affects men; some heterozygous and homozygous females, however, exhibit mild symptoms (Ishihara et al., 2001, Schmidt et al., 2002, Tomik et al., 2006, Karaer et al., 2008, Soraru et al., 2008). Moreover, symptoms in female carriers are not necessarily correlated with X-inactivation (Ishihara et al., 2001, Paradas et al., 2008). Similarly, while there is a strong inverse correlation between age of symptom onset and repeat length in SBMA, there is still a wide variation in the age of onset for any particular symptom between individual patients with a given repeat length. Splicing of the androgen receptor is dysregulated in some breast (Zhu et al., 1997) and prostate cancers (Jagla et al., 2007, Dehm et al., 2008, Hu et al., 2009, Marcias et al., 2010). This alternative splicing produces some isoforms that exhibit enhanced retention in the cytoplasm (Jagla et al., 2007) and others that exhibit constitutive activation with androgen-independent nuclear translocation and transcriptional regulatory activities (Dehm et al., 2008, Hu et al., 2009, Marcias et al., 2010). It would be interesting and clinically relevant to determine whether any splice variants belonging to this latter category are observed or enriched in females who exhibit some SBMA symptoms (compared to asymptomatic carriers) and SBMA patients who are resistant to clinical improvement during leuprolelin therapy (compared to responders). If constitutively activated isoforms are overrepresented in either of these groups of individuals, AR alternative splicing might be a useful prognostic indicator of disease penetrance and therapeutic response in SBMA.

Finally, examination of splice variation in affected neuronal populations compared to unaffected cell types may provide clues to the question of which elements of protein context are important to disease pathophysiology. The use of tissue
microdissection coupled with splice variant analysis, similar to that discussed earlier for CACNA1A mRNA, may reveal variants that are selectively enriched in vulnerable neuronal populations (either constitutively or only in the context of a (CAG)$_n$ expansion). If this type of splice variation exists among other polyglutamine disease gene products, careful examination of the corresponding protein isoforms may highlight peptide sequences important for conformational stability, subcellular protein localization, post-translational modifications, and specific protein:protein interactions that could modulate polyglutamine toxicity.

**Phenotypic Characterization of Atxn3(Q$_3$KQ$_{82}$) Knock-in Mice**

While we know that the F2 FLPe(tg/-), positive selection cassette mosaic, Atxn3(Q$_3$KQ$_{82}$)(+/-) knock-in animals are viable and express detectable polyglutamine-expanded ataxin-3 protein, significant work remains for the establishment of this line. These F2 animals are currently being mated to wild type C57BL/6J animals to generate FLPe(-/-), FRT-only Atxn3(Q$_3$KQ$_{82}$)(+/-) F3 founders. These animals will be crossed to C57BL/6J mice to establish the knock-in line, and crossed to each other to produce homozygous knock-in mice. These animals and their progeny will also be used for the initial phenotypic characterization of this mouse model of SCA3. The colony will be monitored to assess whether transmission of the knock-in allele follows the predicted Mendelian ratios in Atxn3(Q$_3$KQ$_{82}$)(+/-) x wild type C57BL/6 and Atxn3(Q$_3$KQ$_{82}$)(+/-) x Atxn3(Q$_3$KQ$_{82}$)(+/-) crosses. Genotyping will also routinely be conducted with the primers MJD10F1 and MJD10R3, to monitor progeny for large intergenerational changes in repeat length (Figure 25 – 27). Selected animals will also be assessed for somatic and intergenerational repeat instability analysis by PCR-amplification using primers flanking the (CAG)$_n$ repeats followed by DNA sequencing, as previously described (Boy et al.,
Both male and female knock-in animals will be used as breeders, with in order to increase the likelihood of anticipation and repeat stability, respectively.

Wild type, hemizygous knock-in, and homozygous knock-in mice will be longitudinally followed for rigorous phenotypic analysis. An initial cohort of animals will be followed to assess monthly weights and survival. At 3, 6, 9, 12, and 18 months of age, cohorts of animals will undergo anatomic, biochemical, and behavioral assessment, based on the SmithKline Beecham, Harwell, Imperial College School of Medicine, Royal London Hospital, Phenotype, Assessment (SHIRPA) protocol (Rogers et al., 1997). While a comprehensive evaluation will include a complete SHIRPA protocol, priority studies will be dedicated to assessing the protein aggregation and neurodegenerative phenotypes in these animals. Protein misfolding and aggregation will be assessed by non-denaturing protein fractionation of brain lysates and immunolabeling of prepared brain tissue. Behavioral analysis focused on the anticipated neurodegenerative phenotype will include open field analysis to assess activity level, as well as beam walking and gait analysis to detect an ataxic phenotype. Additionally, cohorts of these animals will be assessed for proteasomal dysfunction (Bowman et al., 2005), transcriptional dysregulation (Li et al., 2002, Evert et al., 2006), and alterations of alternative splicing (Chapter 2).

Conclusions

The importance of protein context in the manifestation of specific polyglutamine diseases has become increasingly clear. While therapeutic strategies for the treatment of polyglutamine diseases have included approaches designed to reduce toxic protein expression or non-specifically neutralize the toxicity of the expanded polyglutamine repeats, the most promising therapeutic strategy in clinical trials to date was developed based on a protein context-specific understanding of the behavior of the AR within cells (Banno et al., 2009). Data from leuprorelin trials (Katsuno et al., 2004, Banno et al.,
and conditional mouse models (Orr and Zoghbi, 2000, Yamamoto et al., 2000) suggest that the progression of polyglutamine disease symptoms are reversible with interventions that reduce key disease-specific toxic events. Although there is ample room for exploration of approaches that decrease levels of all expanded polyglutamine proteins or generally enhance cellular protein quality control mechanisms, protection of vulnerable neuronal populations may require protein context-specific approaches. Future studies must seek to better understand polyglutamine-induced loss of protein function events, to characterize polyglutamine-induced aberrations in functional protein complexes and restore endogenous protein-protein interactions, and to understand the complexity that alternative splicing adds to disease-specific protein context. To meet these goals, the polyglutamine disease field requires additional model systems that preserve as much of the genomic, transcript, and protein context as possible, while manifesting quantifiable symptoms that are robust enough to test the efficacy of interventions.
Figure 31. Mutation of UIM3 decreases tubulin levels.  
Representative Western blotting of lysates from Cos7 cells transiently expressing Flag-tagged ataxin-3 constructs. Overexpression of the UIM3(SA/DG) mutant, but neither endogenous C-terminal splice variant, dramatically decreases β-tubulin levels. GAPDH is used as a loading control.
Figure 32. Model for the role of UIM3 in aggresome formation and microtubule stability.

(A) The normal role of ataxin-3, HDAC6, and parkin in pericentrosomal degradation of misfolded proteins. (A1) Ataxin-3 is required for HDAC6 to sense abundant misfolded proteins. (A2) HDAC6 forms a complex with dynein/dynactin to transport parkin and polyubiquitinated substrates towards the centrosome, in a microtubule deacetylation-dependent manner. (A3) This compartment is enriched in 20S proteasomal components, polyubiquitinated proteins, and ataxin-3. The concentration of protein quality control machinery and misfolded proteins facilitates an interaction between parkin and ataxin-3. (A4) The cooperative ubiquitin chain elongation and editing activities of ataxin-3:parkin complexes enhance the efficiency of proteasomal degradation. (B) The ataxin-3 UIM3(SA/DG) mutant dysregulates this process, leading to decreases in total tubulin, loss of IF signal from an N-terminal Flag epitope in the pericentrosomal area, and an increase in 3-dimensional cellular complexity. (B1) The ataxin-3 UIM3(SA/DG) mutant may be able to more efficiently activate HDAC6 to (B2) recruit parkin to the pericentrosomal domain in the absence of a large burden of misfolded proteins. (B3) The resulting domain is still enriched in protein quality control machinery, but not as heavily enriched with misfolded, polyubiquitinated proteins. In this environment, HDAC6-dependent accumulation of protein quality control machinery in the absence of ample misfolded proteins may lead increased ubiquitination and (B4) degradation of abundant tubulin dimers provided during the deacetylation-dependent HDAC6 destabilization of microtubules. (B3) Alternately, UIM3-mutant ataxin-3:parkin complexes may behave abnormally, reducing the autoubiquitination and degradation of parkin, and leading to (B4) increased degradation of parkin substrates, including tubulin. Decreased levels of soluble tubulin dimers will further destabilize the tubulin cytoskeleton; increased ratios of parkin to microtubules, however, may increase the peripheral stability of the remaining, intact microtubules. Note that in this model, formation of the ataxin-3:parkin complex masks the N-terminal Flag epitope.
REFERENCES


Slow EJ, van Raamsdonk J, Rogers D, Coleman SH, Graham RK, Deng Y, Oh R, 
Bissada N, Hossain SM, Yang YZ, Li XJ, Simpson EM, Gutekunst CA, Leavitt 

linked recessive bulbospinal neuronopathy. A clinicopathological study. Brain 

Soliman GA, Ishida-Takahashi R, Gong Y, Jones JC, Leshan RL, Saunders TL, Fingar 
DC, Myers MG, Jr. (2007) A simple qPCR-based method to detect correct 
insertion of homologous targeting vectors in murine ES cells. Transgenic Res 

20:438-446.

Soraru G, D'Ascenzo C, Polo A, Palmieri A, Baggio L, Vergani L, Gellera C, Moretto G, 
Pegoraro E, Angelini C (2008) Spinal and bulbar muscular atrophy: skeletal 
muscle pathology in male patients and heterozygous females. J Neurol Sci 
264:100-105.

Strom AL, Forsgren L, Holmberg M (2005) Identification and characterization of 
Spinocerebellar Ataxia Type 7 (SCA7) isoform SCA7b in mice. Biochim Biophys 

Tadokoro K, Yamazaki-Inoue M, Tachibana M, Fujishiro M, Nagao K, Toyoda M, Ozaki 
M, Ono M, Miki N, Miyashita T, Yamada M (2005) Frequent occurrence of 
protein isoforms with or without a single amino acid residue by subtle alternative 
splicing: the case of Gln in DRPLA affects subcellular localization of the 

T, Kato S (2002) Androgen-dependent neurodegeneration by polyglutamine- 

(2006) Ataxin-7 can export from the nucleus via a conserved exportin-dependent 


with mutant atrophin-1 with expanded polyglutamines accelerates intranuclear 

Thakur AK, Jayaraman M, Mishra R, Thakur M, Chellgren VM, L Byeon I-J, Anjum 
DH, Kodali R, Creamer TP, Conway JF, M Gronenborn A, Wetzel R (2009) 
Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex 


