Triplet Repeat Diseases

Stephan J. Guyenet and Albert R. La Spada
University of Washington, Seattle, WA 98195

1 A Novel Mechanism of Genetic Mutation Emerges
1.1 Repeat Sequences of All Types and Sizes
1.2 Trinucleotide Repeat Expansion as a Cause of Disease: Unique Features Explain Unusual Genetics

2 Repeat Diseases and Their Classification
2.1 Summary of Repeat Diseases
2.2 Differences in Repeat Sequence Composition and Location within Gene
2.3 Classification Based upon Mechanism of Pathogenesis and Nature of Mutation

3 Type 1: The CAG/Polyglutamine Repeat Diseases
3.1 Spinal and Bulbar Muscular Atrophy
3.2 Huntington’s Disease
3.3 Dentatorubral Pallidoluysian Atrophy
3.4 Spinocerebellar Ataxia Type 1
3.5 Spinocerebellar Ataxia Type 2
3.6 Spinocerebellar Ataxia Type 3/Machado–Joseph Disease
3.7 Spinocerebellar Ataxia Type 6
3.8 Spinocerebellar Ataxia Type 7
3.9 Spinocerebellar Ataxia Type 17
3.10 Role of Aggregation in Polyglutamine Disease Pathogenesis
3.11 Protein Context
3.12 Transcriptional Dysregulation
3.13 Proteolytic Cleavage

4 Type 2: the Loss-of-function Repeat Diseases
4.1 Fragile X Syndrome
Triplet Repeat Diseases

4.2 Fragile XE Mental Retardation 34
4.3 Friedreich’s Ataxia 36
4.4 Progressive Myoclonus Epilepsy Type 1 39

5 Type 3: the RNA Gain-of-function Repeat Diseases 40
5.1 Myotonic Dystrophy Type 1 40
5.2 Myotonic Dystrophy Type 2 42
5.3 Spinocerebellar Ataxia Type 8 43
5.4 The Fragile X Tremor – Ataxia Syndrome (FXTAS) 44

6 Type 4: The Polyalanine Diseases 46
6.1 Overview 46
6.2 Oculopharyngeal Muscular dystrophy 46
6.3 Synpolydactyly (Syndactyly Type II) 48
6.4 Cleidocranial Dysplasia 48
6.5 Holoprosencephaly 48
6.6 Hand-foot-genital Syndrome 49
6.7 Blepharophimosis-ptosis-epicanthus Inversus Syndrome 49
6.8 Syndromic and Nonsyndromic X-linked Mental Retardation 49
6.9 Congenital Central Hypoventilation Syndrome 50

7 Unclassified Repeat Diseases Lacking Mechanistic Explanations 50
7.1 Spinocerebellar Ataxia Type 10 50
7.2 Spinocerebellar Ataxia Type 12 51
7.3 Huntington’s Disease Like 2 (HDL2) 52

Bibliography 54
Books and Reviews 54
Primary Literature 54

Keywords

Aggregate
Accumulation of proteinaceous and/or ribonucleic acid material into a structure that is visible in a cell at the light microscope level.

Anticipation
Worsening severity of a disease phenotype as the causal (typically dominant) genetic mutation is transmitted from one generation to the next in a family segregating the disease of interest.
Gain of function
Refers to a type of mutation that imparts a novel activity or action to the gene product containing the mutation.

Loss of function
Refers to a type of mutation that eliminates the action of the gene product encoded by the gene within which the mutation resides.

Repeat Expansion
An elongation of a repeat to a larger size that no longer falls within the size distribution range typically seen in the normal population; this process is now recognized as a mechanism of human genetic mutation.

Trinucleotide
Three DNA base pairs of specific sequence composition (e.g. cytosine-adenine-guanine).

The repeat expansion disorders are a group of human diseases that are caused by the elongation of a DNA repeat sequence. In this chapter, we provide an overview of the discovery of repeat expansion as an important cause of human disease, and we summarize the molecular genetics and mechanistic basis of 27 microsatellite repeat disorders. Comparison of the many repeat expansion disorders reveals distinct categories of repeat diseases, allowing us to propose a classification of the repeat expansion disorders based upon mutation sequence and pathogenic mechanism. The four types of repeat expansion disorders defined by this approach are the CAG/polyglutamine repeat diseases; the loss-of-function repeat diseases; the RNA gain-of-function repeat diseases; and the polyalanine diseases. Although the genetic basis for most of these diseases was determined less than a decade or so ago, considerable advances have been made in our understanding of how “dynamic mutations” produce molecular pathology and human disease.

1
A Novel Mechanism of Genetic Mutation Emerges

1.1
Repeat Sequences of All Types and Sizes

Long before the sequencing of the human genome was undertaken, the discovery of bacterial enzymes that recognize specific DNA sequences (“recognition sites”) and cleaved them, yielded a new methodology for differentiating individuals (“molecular fingerprinting”). This methodology also found application in the mapping of inherited genetic diseases, taking advantage of human variation in the form of so-called restriction fragment length polymorphisms (RFLP’s). In the search for even more informative genetic markers, investigators uncovered a variety of very
short repeat sequences ("short tandem repeats" – STR’s; "simple sequence length polymorphisms" – SSLP’s) that were highly dispersed throughout the human genome. These latter repeat sequences came to be known as "microsatellites" to differentiate them from the "minisatellites" that were being used for molecular fingerprinting. Minisatellites were originally defined as tandem arrays of 14–100 bp repeating sequences. In the case of minisatellites, the repeat sequence was in essence a "consensus" as deviations from the exact repeat sequence were common. Microsatellites, however, were typically pure perfect repeats of less than 13 bp, the most common microsatellites being dinucleotide repeats, trinucleotide repeats, or tetranucleotide repeats. The discovery of minisatellites and microsatellites yielded a virtual bonanza of reagents for molecular fingerprinting and genetic linkage mapping, while also providing evolutionary biologists and population geneticists with intriguing material to attempt to reconstruct evolutionary relationships and inter- or intraspecies relationships. Although human geneticists and evolutionary biologists were applying microsatellites in different ways for their own studies, they shared the commonly held belief that such repeats were neutral and therefore unlikely to be of much functional consequence, let alone play a role in causing human disease. Of course, all that would soon change in the last decade of the twentieth century.

1.2 Trinucleotide Repeat Expansion as a Cause of Disease: Unique Features Explain Unusual Genetics

In 1991, two groups working on seemingly unrelated inherited genetic diseases independently made paradigm-shifting discoveries. In one case, a CAG trinucleotide repeat expansion within the first coding exon of the androgen receptor (AR) gene was found to be the cause of an X-linked neuromuscular disorder known as spinal and bulbar muscular atrophy (SBMA or Kennedy’s disease). CAG encodes the amino acid glutamine; thus, elongation of a polyglutamine tract within the AR protein was hypothesized to be the molecular basis for the motor neuron degeneration in SBMA. In the other case, a disorder known as the fragile X syndrome of mental retardation (FRAXA), also X-linked but much more common, was reported to result from expansion of a CGG repeat. In the latter case, although originally envisioned to encode a polyarginine tract, the CGG repeat turned out to be in the 5’ untranslated region of a novel gene, the so-called FMR-1 gene (for "fragile X mental retardation-1").

The identification of triplet repeat expansions as the cause of two inherited diseases was an exciting turning point in the field of human molecular genetics not only because of the novel nature of these findings, but also because of the unusual genetic characteristics of this new type of mutation. Analysis of families segregating FRAXA revealed the existence of three distinct allele categories defined by the length of the pure CGG repeat: a normal size range; a disease size range; and an intermediate, "premutation" size range. As discussed below, individuals carrying premutation-sized CGG repeats never develop the mental retardation phenotype, but instead are at risk for passing on even larger CGG repeats to their children or grandchildren who then display the mental retardation phenotype. An important tenet that emerged from these early studies was that expanded CGG repeats (larger
than the normal size range) displayed an exceptionally high rate of further mutation. This observation reversed a commonly held view of the genetic material— the notion that any single nucleotide in the human genome displayed a mutation rate of $\sim 1 \times 10^{-5}$. In the case of expanded CGG repeats in premutation carriers, the rate approached unity ($10^0$)! Besides displaying this high mutation rate, there were other unusual features: (1) the CGG repeats showed a marked tendency to further expansion, suggesting that repeat mutation was a polar process; and (2) the sex of the individual transmitting the premutation-sized CGG allele determined whether a large expansion into the disease range would be possible. For FRAXA, expansion into the disease range could only occur if the premutation allele was transmitted by a female carrier. All of these unusual aspects of FRAXA CGG repeat genetics thoroughly accounted for the bizarre non-Mendelian inheritance patterns described in FRAXA families as the “Sherman paradox.”

The recognition of repeat expansion as the cause of the neurodegenerative disorder SBMA and as an explanation for the puzzling genetics of FRAXA set the stage for further discoveries in the field of neurogenetics (the study of inherited neurological disorders). One disorder known as myotonic dystrophy (DM), the most common of all muscular dystrophies, had been the center of a genetic controversy that had gone on for nearly a century. The controversy involved the debate over whether the clinical phenomenon of anticipation truly existed or was simply an artifact of clinical study (Anticipation may be defined as a progressively earlier age of disease onset with increasing disease severity in successive generations of a family segregating an inherited disorder). Although anticipation was initially proposed as a defining feature of DM in 1918, a number of leading geneticists, among them Penrose, L.S., dismissed its authenticity, claiming that it was a product of ascertainment bias due to better clinically defining the profound variable expressivity in this disorder. This view persisted from its promulgation in the 1950s, reinforced by the concept that the genetic material is seldom subject to alteration or modification that could be heritable. However, with the discovery that expanded trinucleotide repeats could further expand, and that indeed the expansion process was a prerequisite for the FRAXA disease phenotype—accounting for maternal inheritance and greater percentages of affected individuals in more recent generations—a role for repeats in diseases displaying anticipation was entertained. Consequently, the third repeat mutation disorder to be identified—just one year after the SBMA and FRAXA discoveries—was DM. Studies of the causal CTG repeat expansion in DM demonstrated a strong correlation between disease severity (i.e. age of onset and rate of progression) and the length of the CTG repeat. In this way, it became clear that anticipation was a genuine phenomenon and that expanded repeat mutational instability was its long awaited molecular explanation. The mutational instability characteristics of expanded repeats have led to their designation as so-called “dynamic mutations.”

2 Repeat Diseases and Their Classification

2.1 Summary of Repeat Diseases

The list of diseases caused by microsatellite repeat expansions (involving repeats of
Triplet Repeat Diseases

3–12 bp) now includes more than 25 disorders (Table 1). Certain aspects of the list of repeat diseases deserve emphasis. Twenty of the repeat diseases either principally or exclusively affect the neuraxis – that is, anywhere from the brain to the cerebellum/brainstem, to the spinal cord, to the peripheral nerve or muscle – and are degenerative disorders. Almost all of these inherited neurological disorders are caused by large expanded repeats that display the property of pronounced genetic instability (dynamic mutation). On the other hand, the developmental malformation syndromes for which repeat expansion mutations have been implicated all involve modestly sized disease repeats by comparison, and these repeats do not exhibit such dynamic mutation genetic instability.

2.2 Differences in Repeat Sequence Composition and Location within Gene

When faced with the task of categorizing the various repeat expansion diseases into different classes, a number of approaches are possible. We have found that consideration of the sequence of the repeat and its location within the gene are the most useful characteristics to apply for grouping the different repeat diseases. As shown in Table 1, there are many different types of repeats varying in length and sequence composition. However, among the recurrent sequence types are CAG trinucleotide repeats, CTG trinucleotide repeats, and GCG trinucleotide repeats. The rest of the repeat sequences are unique in composition and differ widely in size, ranging from 3 to 12 bp as noted above. Comparison of the location of a repeat within the gene it is affecting also yields different types of repeats. The largest single-repeat location category is within the coding region of a gene, which applies to both CAG (glutamine) and GCG (alanine) repeats, and has been proposed but not yet demonstrated for CTG (leucine) repeats. The rest of the locations defined for repeats vary widely, ranging from the gene’s promoter to its 5’ untranslated region to an intron to the 3’ untranslated region, and no more than two to three repeat expansions can be placed in each of these categories at this time.

2.3 Classification Based upon Mechanism of Pathogenesis and Nature of Mutation

To allow us to reconstruct how the different repeat expansion mutations cause molecular pathology in the various disorders that they cause, we have chosen to categorize the 25 repeat disorders into four classes (Table 2). The first class of disorders, the Type 1 repeat diseases, are the “CAG-polyglutamine disorders.” This class of repeat diseases includes nine inherited neurodegenerative disorders (SBMA, Huntington’s disease, dentatorubral pallidoluysian atrophy, and six forms of spinocerebellar ataxia) that all share the common feature of being caused by a CAG repeat located within the coding region of a gene. Upon CAG repeat expansion, a mutant protein with an extended polyglutamine tract is produced, making the protein then adopt an abnormal conformation and misfold to initiate the pathogenic cascade. The resultant pathology is believed to primarily stem from a gain of function of the mutant protein imparted by the expanded polyglutamine tract. As discussed below, much effort has gone into trying to define what the gain-of-function effect is for each polyglutamine disease protein and into determining if
# Tab. 1 Compilation of repeat expansion diseases in humans.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Gene</th>
<th>Repeat</th>
<th>Normal</th>
<th>Premutation</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blepharophimosis-ptosis-epicanthus inversus</td>
<td>FOXL2</td>
<td>GCG</td>
<td>14</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>Cleidocranial dysplasia</td>
<td>RUNX2</td>
<td>GCG</td>
<td>11–17</td>
<td></td>
<td>27</td>
</tr>
<tr>
<td>Congenital central hypoventilation</td>
<td>PHOX-2B</td>
<td>GCG</td>
<td>20</td>
<td></td>
<td>25–29</td>
</tr>
<tr>
<td>Dentatorubralpallidoluysian atrophy (Haw River syndrome)</td>
<td>Atrophin-1</td>
<td>CAG</td>
<td>7–35</td>
<td></td>
<td>49–88</td>
</tr>
<tr>
<td>Fragile X syndrome of mental retardation</td>
<td>Fmr-1</td>
<td>CCG</td>
<td>6–53</td>
<td>45–200&lt;sup&gt;a&lt;/sup&gt;</td>
<td>200–&gt;2000</td>
</tr>
<tr>
<td>FRAXE mental retardation</td>
<td>Fmr-2</td>
<td>GCC</td>
<td>6–40</td>
<td>61–200</td>
<td>200–1000</td>
</tr>
<tr>
<td>Friedreich’s ataxia</td>
<td>Frataxin</td>
<td>GAA</td>
<td>7–38</td>
<td>33–60</td>
<td>66–1700</td>
</tr>
<tr>
<td>Hand-foot-genital</td>
<td>HOX-A13</td>
<td>GGC</td>
<td>18</td>
<td></td>
<td>24–27</td>
</tr>
<tr>
<td>Huntington’s disease</td>
<td>Htt/IT15</td>
<td>CAG</td>
<td>6–35</td>
<td>27–36</td>
<td>39–250</td>
</tr>
<tr>
<td>Huntington’s disease like 2</td>
<td>JPH3</td>
<td>CAG</td>
<td>7–27</td>
<td></td>
<td>50–60</td>
</tr>
<tr>
<td>Holoprosencephaly</td>
<td>ZIC2</td>
<td>GCG</td>
<td>15</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>Myoclonus epilepsy type 1</td>
<td>Cystatin B</td>
<td>CCCC</td>
<td>2–3</td>
<td>12–17</td>
<td>30–150</td>
</tr>
<tr>
<td>Myotonic dystrophy type 1</td>
<td>DMPK</td>
<td>CTG</td>
<td>5–37</td>
<td>50–80</td>
<td>50–4000</td>
</tr>
<tr>
<td>Myotonic dystrophy type 2</td>
<td>ZNF9</td>
<td>CCTG</td>
<td>104–176</td>
<td></td>
<td>175–11,000</td>
</tr>
<tr>
<td>Oculopharyngeal muscular dystrophy</td>
<td>PABP2</td>
<td>GCG</td>
<td>6–7</td>
<td></td>
<td>8–17</td>
</tr>
<tr>
<td>Spinocerebellar ataxia type 1</td>
<td>Ataxin-1</td>
<td>CAG</td>
<td>6–44</td>
<td></td>
<td>39–&gt;10</td>
</tr>
<tr>
<td>Spinocerebellar ataxia type 2</td>
<td>Ataxin-2</td>
<td>CAG</td>
<td>13–33</td>
<td></td>
<td>31–&gt;10</td>
</tr>
<tr>
<td>Spinocerebellar ataxia type 3 (Machado-Joseph disease)</td>
<td>Ataxin-3</td>
<td>CAG</td>
<td>12–40</td>
<td></td>
<td>55–84</td>
</tr>
<tr>
<td>Spinocerebellar ataxia type 6</td>
<td>CACNA1A</td>
<td>CAG</td>
<td>4–18</td>
<td></td>
<td>19–33</td>
</tr>
<tr>
<td>Spinocerebellar ataxia type 7</td>
<td>Ataxin-7</td>
<td>CAG</td>
<td>4–35</td>
<td></td>
<td>37–306</td>
</tr>
<tr>
<td>Spinocerebellar ataxia type 8</td>
<td>?</td>
<td>CTG</td>
<td>16–&gt;800</td>
<td></td>
<td>71–&gt;1000</td>
</tr>
<tr>
<td>Spinocerebellar ataxia type 10</td>
<td>Ataxin-10</td>
<td>ATTC</td>
<td>10–22</td>
<td></td>
<td>800–4500</td>
</tr>
<tr>
<td>Spinocerebellar ataxia type 12</td>
<td>PPP2R2B</td>
<td>CAG</td>
<td>7–32</td>
<td></td>
<td>55–78</td>
</tr>
<tr>
<td>Spinocerebellar ataxia type 17</td>
<td>TBP</td>
<td>CAG</td>
<td>25–48</td>
<td></td>
<td>43–66</td>
</tr>
<tr>
<td>Syndromic and nonsyndromic X-linked mental retardation</td>
<td>ARX</td>
<td>GCG</td>
<td>10–16</td>
<td></td>
<td>17–23</td>
</tr>
<tr>
<td>Sympolodylacty</td>
<td>HOX-D13</td>
<td>GCG</td>
<td>9–15</td>
<td></td>
<td>16–29</td>
</tr>
<tr>
<td>X-linked spinal &amp; bulbar muscular atrophy</td>
<td>AR</td>
<td>CAG</td>
<td>5–34</td>
<td></td>
<td>37–70</td>
</tr>
</tbody>
</table>

<sup>a</sup> Male premutation carriers of the FMR-1 CGG repeat develop the Fragile X tremor-ataxia syndrome.

<sup>–</sup>, no premutation alleles identified, so they may not exist.

<sup>?</sup>, existence of premutation alleles is unknown.
<table>
<thead>
<tr>
<th>CAG/polyglutamine disorders</th>
<th>Loss-of-function disorders</th>
<th>RNA gain-of-function disorders</th>
<th>Polyalanine disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dentatorubral pallidoluysian atrophy</td>
<td>Fragile X syndrome of MR</td>
<td>Myotonic dystrophy type 1</td>
<td>Blepharophimosis-ptosis-epicanthus inversus</td>
</tr>
<tr>
<td>Huntington’s disease</td>
<td>FRAXE MR</td>
<td>Myotonic dystrophy type 2</td>
<td></td>
</tr>
<tr>
<td>Spinal and bulbar muscular atrophy</td>
<td>Friedreich’s ataxia</td>
<td>Spinocerebellar ataxia type 8</td>
<td>Cleidocranial dysplasia</td>
</tr>
<tr>
<td>Spinocerebellar ataxia type 1</td>
<td>Myoclonus epilepsy type 1</td>
<td>Fragile X tremor-ataxia syndrome</td>
<td>Congenital central hypoventilation</td>
</tr>
<tr>
<td>Spinocerebellar ataxia type 2</td>
<td></td>
<td></td>
<td>Hand-foot-genital syndrome</td>
</tr>
<tr>
<td>Spinocerebellar ataxia type 3</td>
<td></td>
<td></td>
<td>Holoprosencephaly</td>
</tr>
<tr>
<td>Spinocerebellar ataxia type 6</td>
<td></td>
<td></td>
<td>Oculopharyngeal muscular dystrophy</td>
</tr>
<tr>
<td>Spinocerebellar ataxia type 7</td>
<td></td>
<td></td>
<td>Synpolydactyly</td>
</tr>
<tr>
<td>Spinocerebellar ataxia type 17</td>
<td></td>
<td></td>
<td>X-linked MR (syndromic &amp; nonsyndromic)</td>
</tr>
</tbody>
</table>

MR = mental retardation.
shared pathways of toxicity are initiated in the different diseases. At least one, and perhaps a greater number of these disorders may principally involve a simultaneous dominant-negative partial loss of the normal function of the disease protein. The next class of disorders, the Type 2 repeat diseases, are a much more disparate group of repeat disorders. The Type 2 repeat diseases are the “Loss-of-function repeat disorders.” These disorders include different repeats that vary in sequence composition and gene location, but share a final common pathway of disease pathogenesis—a loss of function of the disease gene within which they occur. This group includes various classic trinucleotide repeat disorders such as the two fragile X syndromes of mental retardation (FRAXA and FRAXE) and Friedreich’s ataxia—but also encompasses the dodecamer repeat expansion in progressive myoclonic epilepsy type 1, and possibly the CAG repeat expansion in Huntington’s disease like-2 (HDL2) gene. Strong evidence for a loss-of-function pathway in the form of nonrepeat loss-of-function mutations supports many of these classifications. The third group of repeat diseases, the Type 3 disorders, comprise a shared class because all of them have been proposed to involve the production of a toxic RNA species. This category of repeat diseases is thus called the RNA gain-of-function disorders. Included among these disorders are two closely related forms of DM, the common and classic myotonic dystrophy type 1 (DM1) and its uncommon phenocopy, myotonic dystrophy type 2 (DM2). Another member of this group is the recently described fragile X tremor-ataxia syndrome (FXTAS) in male premutation carriers—a fascinating example of two different disease pathways operating upon the same expanded repeat mutation based upon size range differences. One form of spinocerebellar ataxia (SCA8) with an unclear mechanism of pathogenesis has also been provisionally placed into this category, based upon current working models of how its repeat causes disease. The last class of repeat disease, the Type 4 disorders, are the “GCG-polyalanine disorders” that are grouped together because all involve short GCG repeat tracts falling within the coding regions of unrelated genes that become expanded to moderately sized GCG repeats. With the exception of oculopharyngeal muscular dystrophy, all are developmental malformation syndromes, and while gain-of-function polyalanine toxicity has been proposed for a number of these disorders, loss of function due to the polyalanine expansion seems more likely for others. Finally, a number of repeat disorders, spinocerebellar ataxia type 10 (SCA10), spinocerebellar ataxia type 12 (SCA12), and Huntington’s disease like 2 (HDL2), currently defy classification because very little is known about their molecular basis. These diseases will be considered in the final section of this chapter.

3 Type 1: The CAG/Polyglutamine Repeat Diseases

3.1 Spinal and Bulbar Muscular Atrophy

Spinal and bulbar muscular atrophy (SBMA; Kennedy’s disease) is a late-onset neurodegenerative disease with an inheritance pattern resembling X-linked recessive. It has a prevalence of about 1 in 50,000 males. Patients suffer a late-onset, progressive degeneration of primarily lower motor neurons in the anterior horn of the spinal cord and in the bulbar region of the
Triplet Repeat Diseases

brainstem; however, variable involvement of sensory neurons in the dorsal root ganglia also occurs. SBMA typically presents with cramps, followed by proximal muscle weakness and atrophy. Patients often exhibit dysarthria, dysphagia, and fasciculations of the tongue and lips. Affected individuals have symptoms of mild androgen insensitivity, such as gynecomastia, reduced fertility, and testicular atrophy.

SBMA is caused by a polymorphic (CAG)n repeat in the first exon of the AR gene, which is expressed as a glutamine tract. Unaffected individuals carry 5 to 34 triplet repeats, while affected individuals carry 37 to 70 repeats. SBMA exhibits a paternal expansion bias. The disease does not appear to involve a simple loss-of-function mechanism, as complete loss of AR does not result in motor neuron degeneration.

AR is widely expressed in males and females, and is a member of the steroid receptor–thyroid receptor superfamily with a highly conserved DNA binding domain, ligand binding domain and transactivation domain (Fig. 1). In its inactive state, it forms an apo-receptor complex with heat-shock proteins (HSPs) 70 and 90, and resides in the cytoplasm. Upon binding androgen, it dissociates from these HSP chaperone proteins and translocates to the nucleus. Once in the nucleus, AR dimers transactivate certain genes, many of which are responsible for generating and maintaining male characteristics. Although the glutamine expansion does not affect the binding of its ligand, androgen (testosterone), the glutamine tract is in the major transactivation domain, and may affect transactivation competence. However, the effect of the polyglutamine expansion upon AR transactivation competence remains controversial.

Many lines of evidence suggest that AR must translocate to the nucleus to exert its toxicity. Nuclear inclusions (NIs) are present in motor neurons of the spinal cord and brainstem in SBMA patients. In a Drosophila model of SBMA, retinal expression of mutant AR only yielded a degenerative phenotype in the presence of ligand. As in humans, androgen binding causes the nuclear translocation of AR in mice. Transgenic male SBMA mice produce testosterone and will only

Normal = 5–34
SBMA = 37–70

\[\ldots\text{CAGCAGCAG}\ldots\]

![Fig. 1](image) Diagram of the androgen receptor. The androgen receptor is a member of the steroid receptor-thyroid hormone receptor superfamily, and consequently displays a stereotypical architecture. The CAG repeat – polyglutamine tract (striped box) resides within the amino-terminal domain, which mediates transcription activation through an "activation function" domain (charcoal gray box). Additional conserved domains include the DNA binding domain (light gray box), nuclear localization signal (black box), and the ligand binding domain (checkered box). Expansion of the CAG repeat to alleles of \( \geq 37 \) triplets is the cause of spinal and bulbar muscular atrophy (SBMA).
Triplet Repeat Diseases

develop motor neuron disease if expressing polyglutamine-expanded full-length AR protein, yet when castrated, such SBMA transgenic mice do not develop a phenotype. In the same study, female transgenic mice hemizygous for expanded AR develop motor neuron degeneration when exposed to exogenous androgen. Importantly, even rare human females who are homozygous for an AR CAG repeat expansion mutation do not develop SBMA, despite widespread expression of mutant AR throughout the CNS (central nervous system). Thus, SBMA is not a true X-linked recessive disorder, but rather is classified as a sex-limited disorder, since expression of the disease phenotype is dependent upon male levels of androgen.

A number of studies on SBMA patients and mouse models have characterized the NIs seen in this disease. While NIs are widely distributed in lower motor neurons of the spinal cord and brainstem, NIs also occur in a variety of nonneuronal tissues that appear to function normally. NIs colocalize with components of the proteasome and with molecular chaperones. How this contributes to the phenotype is unknown, although HSP70 overexpression attenuates toxicity in cell culture and in transgenic mouse. The presence of proteasome components and HSPs in NIs may thus be revealing a protective intervention by the cell, or alternatively may be deleterious due to depletion of these important cellular proteins. Interestingly, only antibodies raised to amino-terminal fragments of AR detect NIs, indicating that proteolysis may play a role in the disorder. Caspase-3 has been shown to cleave AR in a polyglutamine tract length-dependent manner in vitro, and this may have pathogenic significance, as truncated AR is more toxic than full-length protein in cell culture studies and transgenic mouse models. The phosphorylation of AR is modulated by androgen, and this posttranslational modification appears to enhance caspase-3 cleavage.

One possible mechanism of expanded AR toxicity is through transcription interference. Polyglutamine-expanded AR interacts with the transcription coactivator CREB-binding protein (CBP) in a polyglutamine tract length-dependent manner, colocalizes with CBP in spinal cord NIs from patients, and can interfere with CBP-dependent transcription. CBP is a transcription cofactor that regulates the expression of vascular endothelial growth factor (VEGF), among other genes. VEGF is important in motor neuron health, as deletion of a portion of its promoter called the hypoxia response element (HRE) causes motor neuron degeneration even in normoxic mice. Pathologically expanded AR reduces VEGF transcript expression in males, with VEGF165 isoform expression reduced at both the RNA and protein levels. Adding VEGF165 to a motor neuron-like cell line (MN-1) expressing polyglutamine-expanded AR significantly rescues its cell death, again supporting the role of transcription interference in SBMA and suggesting that VEGF165 may serve as a neurotrophic factor for motor neurons.

While no effective treatment for SBMA has yet been validated in human patients, the role of testosterone in SBMA disease progression has received considerable attention. Interestingly, testosterone supplementation was initially used as a treatment for SBMA. Rather than aggravating the disease as one might fear, it was reported to attenuate the phenotype slightly, but did not significantly retard disease progression. Such a beneficial effect of testosterone may be due
to anabolic effects on muscle strength or to a downregulation of AR in response to elevated androgen levels. Still another possibility is that AR-mediated transcription in motor neurons somehow performs a trophic function in the face of damage or injury. Very recent work, however, strongly indicates that elimination of ligand by surgical or pharmacological castration is a very effective treatment in SBMA transgenic mouse models, even showing efficacy when SBMA mice display an advanced phenotype. As ligand binding is associated with nuclear translocation of mutant AR and aberrant effects upon nuclear transcription appear crucial for SBMA disease progression, abrogation of nuclear localization may account for the success of castration. Consistent with this hypothesis are results of studies with flutamide, a drug that appears to block AR-dependent transactivation without preventing its nuclear translocation. While pharmacological castration with leuprolide is highly effective therapeutically, flutamide does not ameliorate symptoms or disease progression in an SBMA mouse model. Attempted translation of these preclinical trial results to human SBMA patients is currently underway.

3.2 Huntington’s Disease

Huntington’s disease (HD) is an autosomal dominant disorder with a prevalence of 1 per 15 000 persons worldwide. It is a debilitating disease that often presents clinically in the fourth or fifth decade of life with chorea (i.e. spontaneous, involuntary dancelike movements). Personality change and cognitive impairment may precede the clinical onset by years, and ultimately culminate in dementia after onset of the movement disorder. Chorea gives way to bradykinesia and rigidity late in the disease. CNS atrophy occurs most prominently in the striatum, which is reduced to a fraction of its original size (Fig. 2). However, significant neurodegeneration and neuron loss in the cortex is also typical, while cerebellum, brainstem, and spinal cord are relatively spared – except in juvenile-onset cases.

HD is caused by a \((\text{CAG})_n\) repeat expansion in the \textit{huntingtin} gene (\textit{htt}), which encodes a 350-kDa protein containing 67 exons. Unaffected individuals carry 6 to 35 repeats, while affected individuals carry 39 to 250 repeats. The largest repeats cause juvenile-onset HD and display a paternal transmission bias. The \textit{htt} protein is ubiquitously expressed in brain tissue.

![Huntington's disease (HD) neuropathology](image)

Fig. 2  Huntington’s disease (HD) neuropathology. Hemi-coronal sections of postmortem brains from (a) a classic, adult-onset HD patient and (b) a normal control reveal marked degeneration of the striatum (midmedial region) and considerable atrophy of cortical regions. (From Robataille et al. (1997) \textit{Brain pathol.} 7, 901. Used with permission).
with highest levels in striatal interneurons and cortical pyramidal cells. Two htt mRNA transcripts have been detected: one 10 kb and the other 13 kb in length, expressed most highly in CNS neurons. Ultrastructurally, the wild-type full-length protein is predominantly cytoplasmic and is located in the pre- and postsynaptic regions of dendrites and axons. Htt protein associates with microtubules, vesicles, and organelles.

HD appears to have a predominantly dominant mechanism due to its inheritance pattern, evidence that homozygotes are no more severely affected than heterozygotes, and the observation that heterozygous deletion of huntingtin does not cause HD. The fact that no HD-causing loss-of-function mutations have been documented in the huge \textit{htt} gene further supports a gain-of-function mechanism. However, postnatal elimination of htt expression in regions of the cortex can cause striatal degeneration in mice, so the notion that gain of function fully accounts for the HD phenotype is being reexamined. As htt is a regulator of transcription activation and/or mediator of vesicular brain-derived neurotrophic factor (BDNF) transport up and down axons, many investigators now envision the effects of expanded htt as twofold: simultaneously causing protein misfolding leading to gain-of-function toxicity, and inducing partial loss of an ill-defined normal function.

In 1997, it was first reported that mutant htt forms dense amyloid-type aggregates in the nucleus, perikarya, and neuropil of neurons from a transgenic mouse model and in human patients. The role of aggregation in the polyglutamine diseases and in a wide range of neurodegenerative disorders including Alzheimer’s disease, Parkinson’s disease, amyotrophic lateral sclerosis, and the prion diseases, thus emerged as an important theme at the end of the last decade. While initially it was thought that the formation of large aggregates is the basis of polyglutamine neurotoxicity, the weight of evidence now suggests little correlation between visible aggregate formation and disease pathology. However, the occurrence of aggregates along with neuropathology suggests that aggregation is inextricably linked to the pathogenicity of polyglutamine disease proteins. We will address the role of polyglutamine aggregation in neurotoxicity in detail in a separate section.

Many theories have been proposed to account for how polyglutamine-expanded htt causes neurotoxicity. A thorough discussion of this literature goes beyond the scope of this chapter, so the reader is referred to the relevant books and reviews in the Bibliography for a more intensive treatment of this topic. Major theories of htt neurotoxicity that will be considered herein include transcription dysregulation, proteasome inhibition, mitochondrial dysfunction/excitotoxicity, and proteolytic cleavage. As multiple independent toxicity events may be occurring concomitantly, these disease pathways are not mutually exclusive.

Expanded htt protein may interfere with transcriptional processes. Many transcription factors, such as CBP, contain glutamine tracts that mediate important protein–protein interactions. CBP interacts directly with htt, and mutant htt toxicity is ameliorated in striatal neurons in \textit{vitro} when CBP lacking the htt interaction domain is overexpressed. CBP mediates the transcription of a number of neuronal survival factors, such as BDNF, and functions by acetylating histones, one aspect of chromatin remodeling that permits transcription to occur. Studies of HD fruit fly and mouse models have highlighted the
potential importance of histone acetylation by demonstrating that histone deacetylase inhibitors (HDAC Is) can successfully rescue degenerative phenotypes in these model organisms. Htt can affect transcription mediated by p53, Sp1, and REST interaction, thereby potentially altering the expression of a large number of genes. Differential expression of survival factors, neurotransmitter receptors, and a number of other genes may thus contribute to HD pathogenesis.

The ubiquitin-proteasome protein degradation pathway has also been implicated in HD. Nuclear inclusions of htt colocalize with ubiquitin, which indicates that the expanded protein has been identified as misfolded and thus targeted for proteasomal degradation. However, polyQ sequences longer than 9 glutamines are impossible for eukaryotic proteasomes to cleave. The proteasomes of htt-transfected cells are consequently less capable of degrading proteins other than htt, as demonstrated by the reduced degradation of GFP-tagged proteins in culture. If proteasome components are clogged with mutant protein and/or sequestered into inclusions, they may be unable to degrade other misfolded or damaged proteins that carry out important functions. Alternatively, accumulation of improperly degraded proteins may interfere with other normal cellular processes, such as autophagy or mitochondrial oxidative phosphorylation.

Perhaps the longest and most thoroughly studied htt toxicity pathway involves metabolic disturbances and excitotoxicity. Glucose metabolism and oxygen consumption are reduced in HD brains as measured by PET. Severe deficits in the activity of complexes II/III and IV of the mitochondrial electron transport chain (ETC) are evident in HD brains. Inhibitors of complex II of the mitochondrial electron transport chain, such as 3-nitropropionic acid (3-NPA), can cause a selective degeneration of the striatum in rat and primate models when injected systemically, since the striatum has among the highest energy demands of all neuronal regions. 3-NPA reduces levels of ATP, resulting in mitochondrial and cellular depolarization with activation of voltage-dependent NMDA receptors. Excitotoxic damage may act in concert with increased production of free radicals to cause selective striatal degeneration in HD. Some of the earliest animal models of HD were thus generated by exposing the striatum or entire brain of rats or primates to metabolic or excitotoxic insults. In 1976, the first such model of HD was created by injection of the glutamate analog and excitotoxin kainic acid into the striatum of rats. Such lesioned rats exhibited a selective degeneration of neurons in the striatum reminiscent of HD. Recent studies have suggested that impaired Ca++ flux due to aberrant interaction of htt with the inositol phosphate-3 receptor (IP3-R) may underlie excitotoxic pathology. Studies of mitochondria from HD patients reveal abnormal mitochondrial Ca++ handling and decreased mitochondrial depolarization thresholds, in support of this hypothesis. One very recent study has found that deletion of peroxisome proliferator-activated receptor (PPAR) gamma coactivator 1 alpha (PGC-1α), a key mediator of mitochondrial biogenesis, yields HD-like striatal degeneration in mice. Thus, metabolic insults and excitotoxicity may be key steps in HD disease pathogenesis.

Another important theory of htt neurotoxicity posits that proteolytic cleavage of htt is a required step in neuronal dysfunction and the degeneration process. As noted above, the htt protein is enormous, with full-length product consisting of >3100 amino acids. Analysis of human
HD material has indicated an absence of midprotein and C-terminal epitopes in htt aggregates. Careful biochemical studies of htt have characterized a variety of putative caspase and calpain cleavage sites. Various studies suggest that the more truncated the polyglutamine-expanded htt protein, the more toxic it is in cell culture and in animal models, and the more likely it is to enter the nucleus and produce toxicity there. In the case of htt, a series of proteolytic cleavage steps culminating with cleavage to an ~100 amino acid peptide fragment by an aspartyl protease has been proposed to yield a final “toxic fragment.” As it turns out, the “toxic fragment hypothesis” (as it has also been called) may be applicable to a number of polyQ diseases, which will be reviewed later in this section.

3.3 Dentatorubral Pallidoluysian Atrophy

Dentatorubral pallidoluysian atrophy (DRPLA) is a rare, autosomal dominant neurodegenerative disorder most prevalent in Japan. Adult-onset DRPLA typically involves progressive cerebellar ataxia, choreoathetosis, epilepsy, and dementia, while juvenile-onset cases also display myoclonus, epilepsy, and mental retardation. Neuropathological abnormalities include degeneration of the dentate nucleus of the cerebellum, rubral nucleus, and globus pallidus, as well as a more generalized degeneration and gliosis involving the brainstem, cerebellum, cortex, and pons. DRPLA is caused by a polymorphic (CAG)n repeat in the carboxy-terminal coding region of the atrophin-1 gene on chromosome 12. Normal individuals carry 6–35 repeats, while affected individuals carry 49–88 repeats. Anticipation is prominent in DRPLA, as very large repeat expansions can occur in a single generation, typically via paternal transmission.

The DRPLA disease protein, atrophin-1, is widely expressed and appears to be predominantly cytoplasmic. Atrophin-1 contains both a putative nuclear localization signal (NLS) and nuclear export signal (NES), and nuclear localization of normal atrophin-1 is observed. Nuclear localization of polyglutamine-expanded atrophin-1 has been linked to increased toxicity in cell culture models. Ubiquitinated NIs are present in neurons and glia from patient brains, and are also immunoreactive for small ubiquitinlike modifier (SUMO) protein. Atrophin-1 is a substrate for c-Jun N-terminal kinase (JNK), with JNK’s affinity for atrophin-1 inversely proportional to the size of the polyglutamine tract expansion.

Although the relevance of JNK phosphorylation to atrophin-1 function is unknown, other insights into the function of atrophin-1 have been reported. Using Drosophila melanogaster as a model system, Zhang et al. took advantage of the existence of a fly ortholog of atrophin-1 (Atro) and created lines of flies carrying mutations in the Atro gene. These flies demonstrated severe developmental abnormalities due to complex pattern defects, and subsequent experiments revealed that Atro is a transcription corepressor whose activity diminishes with increasing polyglutamine tract length. Independent studies of human atrophin-1 in cell culture studies found evidence for transcription interference of polyglutamine-expanded atrophin-1 with CREB-mediated gene activation. Transcription dysregulation may thus play a prominent role in DRPLA. Another study suggests that disturbed carbohydrate metabolism may contribute to the DRPLA neurodegenerative phenotype.
Spinocerebellar ataxia type 1 (SCA1) is an autosomal dominant disorder characterized primarily by a progressive cerebellar ataxia. It accounts for about 6% of all autosomal dominant cerebellar ataxias (ADCAs) worldwide. SCA1 patients suffer from coordination difficulties including dysarthria, dysphagia, and ophthalmoplegia. The cerebellum undergoes atrophy, gliosis, and severe loss of Purkinje cells, accompanied by degeneration of the dentate nucleus, inferior olive, and some brainstem nuclei. Disease onset typically occurs in the third or fourth decade of life, but presentation in childhood or adolescence to late life may be seen. SCA1 is caused by the expansion of a coding (CAG)$_n$ repeat in the amino-terminal coding region of the ataxin-1 gene. The ataxin-1 CAG repeat is highly polymorphic, ranging in size from 6 to 44 triplets in unaffected individuals. The repeat length associated with disease ranges from 39 to more than 100 CAGs. Unaffected individuals with more than 20 repeats have CAT triplet interruptions within their (CAG)$_n$ repeat tracts. Such interruptions stabilize the repeat expansion, while absence of the CAT repeat interspersion is noted in SCA1 patient alleles. As in many other polyglutamine disorders, there is strong evidence supporting a gain-of-function mechanism in SCA1.

Ataxin-1 is widely expressed in the CNS and throughout the periphery, although expression levels are several-fold higher in nervous system tissues. The protein is predominantly nuclear in the CNS, however, some cytoplasmic staining is apparent in Purkinje cells of the cerebellum and in brainstem nuclei. Ataxin-1 knockout mice do not develop SCA1, although they do exhibit impairments in motor and spatial learning. These mice also have decreased paired-pulse facilitation in the CA1 region of the hippocampus, suggesting that ataxin-1 may normally function in synaptic plasticity and learning.

Large ataxin-1 containing NI’s occur in the brainstem of affected individuals, and are immunoreactive for ubiquitin, the 20 S proteasome subunit, and HSPs HDJ2 and Hsc70. Work done on SCA1 in transgenic mice by the Orr and Zoghbi laboratories has been crucial in formulating models of not only SCA1 disease pathogenesis but also for influencing views of the molecular basis of all polyglutamine diseases. Indeed, the first mouse model for a polyglutamine disease was generated by transgenic overexpression of polyglutamine-expanded ataxin-1 in Purkinje cells (Fig. 3). This SCA1 transgenic mouse model has laid the foundation for

---

**Fig. 3** The original spinocerebellar ataxia type 1 (SCA1) mouse model. (a) Diagram of the Pcp2-SCA1 transgene construct. A Purkinje cell-specific expression cassette based upon inclusion of the promoter (straight line), first two noncoding exons (black boxes), and first intron (bent line) of the Purkinje cell protein 2 (Pcp2) gene and a SV40 polyadenylation sequence (open box) was the basis for this landmark work. Ataxin-1 cDNAs (gray box) containing either 30 CAGs (control) or 82 CAGs (expanded) were inserted into the Pcp2 expression cassette. Sites of various PCR primer sets are also shown. (b) Pcp2-SCA1 CAG-82 mice display ataxia. Still photographs of a 30-week-old Pcp2-SCA1 CAG-82 mouse from a transgenic line that greatly overexpresses the mutant ataxin-1 transgene illustrates the inability of this mouse to maintain its balance when ambulating. Loss of balance when walking is consistent with the gait ataxia seen in human SCA1 patients. (From Burright et al. (1995) *Cell* 82, 937, used with permission).
numerous follow-up studies of polyglutamine disease pathogenesis. For example, expression of mutant ataxin-1 lacking an intact self-association domain precluded aggregate formation, but permitted neurotoxicity, demonstrating that NIs are not required for SCA1 in mice. In a later study, crossing of SCA1 transgenic mice with mice lacking a ubiquitin ligase enzyme yielded SCA1 mice incapable of aggregate formation. These mice were more severely affected than their transgenic counterparts.
due to absence of the ubiquitin ligase. This work supported the view that visible aggregate formation may represent a protective cellular response for neutralizing misfolded polyglutamine-containing peptides. In another study, mice expressing ataxin-1 with a mutated NLS were found not to develop SCA1, suggesting that nuclear localization is absolutely required for SCA1 molecular pathology. Although targeting ataxin-1 to Purkinje cells appears sufficient to recapitulate a dramatic SCA1-like disease in mice, a subsequent knockin mouse model of SCA1 indicated that expression in other regions of the CNS yields a more representative disease phenotype.

Over the past few years, numerous leads have emerged in the search for the pathogenic basis of SCA1. In one line of investigation, phosphorylation of serine 776 of the ataxin-1 protein was shown to affect pathogenesis, highlighting the importance of this posttranslational modification. SCA1 transgenic mice in which serine 776 had been mutated to an alanine, exhibited a dramatically attenuated phenotype and a complete lack of NI’s. In an accompanying study, interaction of polyglutamine-expanded ataxin-1, but not wild-type ataxin-1, with several isoforms of the phosphoserine/threonine binding protein 14-3-3 was reported. This extremely abundant peptide is thought to serve a regulatory function by binding proteins and determining their subcellular localization, among other things. The interaction between ataxin-1 and 14-3-3 was shown to be dependent upon the Akt-mediated phosphorylation of serine 776. A novel mechanism for ataxin-1 toxicity was thus proposed: upon Akt phosphorylation of polyglutamine-expanded ataxin-1, ataxin-1 binds 14-3-3, is stabilized, and ultimately accumulates in the nucleus. The downstream effects of the nuclear accumulation of mutant ataxin-1 on neuronal function, however, remain undefined.

Transcriptional dysregulation is also a reasonable hypothesis for SCA1 pathogenesis. One study has demonstrated that ataxin-1 interacts with polyglutamine binding protein 1 (PQBP-1), and that this interaction results in interference with RNA polymerase-dependent transcription. Independent studies have supported a role for ataxin-1 as a transcription co-repressor. In pull-down assays and in Drosophila, ataxin-1 interacts with the proteins SMRT (silencing mediator of retinoid and thyroid hormone receptors) and HDAC3 (histone deacetylase 3), both transcriptional repressors. Aggregates of polyglutamine-expanded ataxin-1 sequester SMRT, the Drosophila ortholog of SMRT. Transcription repressors and histone deacetylases also appeared in an earlier screen for modulators of the Ataxin-1 phenotype in the fly. In addition, two separate studies of gene expression alterations in presymptomatic SCA1 transgenic mice have uncovered changes in the levels of transcripts encoding proteins involved in Ca++ flux and metabolism. It thus appears that transcription dysregulation may be a key step in SCA1 disease pathogenesis.

3.5 Spinocerebellar Ataxia Type 2

Spinocerebellar ataxia type 2 (SCA2) is an autosomal dominant, progressive cerebellar ataxia that accounts for about 13% of all ADCA cases. Although patients display problems with voluntary coordinated movements as in the rest of the SCAs, its main distinguishing clinical feature is extremely slow saccadic eye movements. Other symptoms may include hyporeflexia, myoclonus, and action tremor. Patients suffer a gradual degeneration of
the cerebellum, inferior olive, pons, and spinal cord. Both Purkinje and granule cells degenerate in the cerebellum, and CNS atrophy in some patients can be widespread. Interestingly, in certain cases, there is involvement of the substantia nigra, with such patients displaying a prominent degree of parkinsonism. The disorder is caused by a coding (CAG)n expansion in a novel gene of unknown function, named ataxin-2. The SCA2 CAG repeat displays little polymorphism in the normal population, with 95% of the population possessing 22 or 23 repeats in each allele. The remaining 5% of alleles in unaffected individuals do nonetheless range from 15–31 CAG repeats. Such unaffected individuals typically have two CAA interruptions in their CAG repeat tract. Affected SCA2 patients typically display CAG repeats numbering from 32–63 triplets, and such disease alleles are always uninterrupted CAG tracts. There is a nonlinear inverse correlation between expansion size and age of onset in SCA2, with some studies documenting a paternal transmission bias for larger expansions.

Both wild-type and expanded ataxin-2 mRNA is widely expressed, with highest levels in the substantia nigra and Purkinje cells of the cerebellum. The 140-kDa protein is cytoplasmic and its function remains uncertain. Since it contains Sm1 and Sm2 motifs common in proteins involved in RNA splicing and protein–protein interactions, involvement in RNA processing has been proposed. A yeast two-hybrid screen indicated that ataxin-2 has a binding partner, ataxin-2 binding protein-1 (A2BP1). A2BP1 is highly conserved throughout the animal kingdom and its expression pattern corresponds well with that of ataxin-2. A2BP1 contains a domain that is also conserved in RNA-binding proteins, the RNP motif. Thus, a complex including ataxin-2 and A2BP1 may be involved in RNA processing or metabolism.

The ataxin-2 gene is evolutionarily conserved. The murine ortholog of ataxin-2 does not contain a polyglutamine repeat tract, however, but instead possesses a single glutamine residue at the analogous location. (Absence of a substantial glutamine repeat tract is typical for mouse orthologs of polyglutamine disease proteins.) In the case of ataxin-2, study of the Drosophila ortholog (Datx2), which does show two regions of marked amino acid similarity and does contain polyglutamine repeat regions, has yielded some potentially important insights into ataxin-2s normal function. Modulation of Datx2 dosage resulted in mutant phenotypes whose cause could be traced to aberrant actin filament formation (Fig. 4). This work suggests that alteration of ataxin-2-mediated regulation of cytoskeletal structure could affect dendrite formation or other aspects of neuronal function in SCA2. As SCA2 is one of the few polyglutamine disorders to display prominent cytosolic aggregates, such a model of SCA2 pathogenesis seems plausible. In other experiments, eliminating the C. elegans ortholog of ataxin-2 (ATX-2) yielded a lethal phenotype, indicating that ATX-2 is required for early embryonic development of this nematode worm. It remains to be seen how these observations in worms and flies will apply to ataxin-2 function in mammals, especially since no simple or conditional knockout of the mouse ataxin-2 gene has been performed.

3.6 Spinocerebellar Ataxia Type 3/Machado–Joseph Disease

Spinocerebellar ataxia type 3 (SCA3) is the most common inherited dominant
Studies of the ataxin-2 ortholog in Drosophila melanogaster reveals a role for Drosophila ataxin-2 (Datx2) in actin filament formation during oogenesis. Egg chambers from normal (Wild-Type (WT)) and mutant flies with reduced expression (Datx2) of Drosophila ataxin-2 were stained with DAPI (blue) and phalloidin (red) to indicate nuclei and filamentous actin respectively (a–d). The egg chambers of WT flies prior to cytoplasmic transport (a) display well demarcated, separated but interconnected cells (blue) as expected, while the egg chambers of Datx2 flies (b) contain irregularly arranged cells. After cytoplasmic transport, the egg chambers of WT flies (c) show one greatly enlarged oocyte (dashed line) with only a small section of compressed cells, while the egg chambers of Datx2 flies (d) have failed to yield an enlarged oocyte, instead retaining dispersed and large adjacent cells. Confocal images of egg chambers prior to cytoplasmic transport stage reveal a prominent actin filament network in WT flies (e), but a remarkably transparent actin filament network in Datx2 flies (f). The decreased density of the actin filament network underlies the cytoplasmic “dumping” defect in the Datx2 flies. (From Satterfield, T.F., Jackson, S.M., Pallanck, L.J. (2002) A Drosophila homolog of the polyglutamine disease gene SCA2 is a dosage-sensitive regulator of actin filament formation, Genetics 162, 1687–1702, used with permission of Genetics.) (See color plate p. xxiii).

Spinocerebellar ataxia worldwide. SCA3 is also known as Machado–Joseph disease (MJD) because of its initial description in a group of Portuguese residents of the Azores islands. It is a progressive, autosomal dominant cerebellar ataxia whose clinical features include ophthalmoplegia, dystonia, dysarthria, and signs of lower motor neuron disease, such as tongue and facial fasciculations. Degeneration occurs in the spinocerebellar tracts, dentate nuclei, red nuclei, substantia nigra, and spinal cord. This disease is unique among the SCAs because the cerebellar cortex and inferior olive are largely spared. SCA3 is caused by a (CAG)n repeat near the 3’ end of the coding region of a novel gene (ataxin-3). Normal alleles range from 12–40 CAG repeats while affected individuals carry 55–84 repeats. Unlike many other triplet repeat disorders, there is a substantial gap between the largest normal repeat allele and the smallest disease-causing repeat allele. Some researchers have proposed that this could be due to a SCA3 founder effect. The presence of SCA3 in every major racial population worldwide, however, would require the founder to be truly ancient. There is the typical inverse correlation between repeat number and age of onset in SCA3, and in this disease, paternal expansion bias is characteristic, as documented by repeat sizing of sperm.

Ataxin-3 is a ubiquitously expressed 42-kDa protein, making it the smallest of the polyglutamine proteins. Ataxin-3 is highly conserved in eukaryotes, with homology to ENTH and VHS domain proteins involved in regulatory adaptor functions and membrane trafficking. Ataxin-3 has several splice variants which reside in the nucleus and the cytoplasm, and it appears developmentally regulated. The existence of NIs was first documented in the brains of patients with SCA3, and this feature of polyglutamine-expanded
ataxin-3 remains evident in SCA3 cell culture and mouse models. Ataxin-3 NIs are ubiquitinated and contain numerous transcription factors.

A number of recent discoveries regarding the domain structure and normal function of ataxin-3 suggest pathways by which polyglutamine repeat expansion could result in disease pathogenesis. Comparison of ataxin-3 amino acid sequences across a wide range of eukaryotic species revealed the presence of an extremely highly conserved amino-terminal sequence that was named the josephin domain. Study of this region indicates that it may play a role in aggregate formation in concert with the expanded polyglutamine tract. A rather intriguing feature of ataxin-3 that was discovered independently is the presence of multiple ubiquitin interaction motifs (UIMs), and the demonstration that ataxin-3 is a polyubiquitin binding protein. This suggests a role for ataxin-3 in mediating protein refolding and degradation.

In addition to a possible normal role in protein surveillance, other studies have found ataxin-3 directly interacts with the histone acetyltransferases CBP and p300, and can block histone acetyltransferase activity by inhibiting access of such coactivators to their histone substrates. This appears to be mediated by the interaction of ataxin-3 with histones. In vitro and in vivo studies of ataxin-3 further revealed an interaction with histones and the chromatin remodeling machinery that led to a repression of transcription activation. Thus, polyglutamine-expanded ataxin-3 may also affect transcriptional processes once it begins to accumulate in the nuclear compartments of the cell types where it is expressed.

3.7 Spinocerebellar Ataxia Type 6

Spinocerebellar ataxia type 6 (SCA6) is an autosomal dominant, slowly progressing cerebellar ataxia that accounts for \( \sim 10\% - 20\% \) of ADCA worldwide. It is characterized predominantly by cerebellar dysfunction that may have an episodic component. Other common features may include dysarthria, nystagmus, loss of vibration sense and proprioception, and imbalance. Histopathological changes include loss of Purkinje cells, cerebellar granule neurons, and neurons in the dentate nucleus and inferior olive. SCA6 is caused by a coding (CAG)\( n \) expansion in exon 47 of the gene CACNA1A, which encodes the \( \alpha_{1A} \) subunit of the P/Q-type voltage-gated calcium channel. This gene is located on chromosome 19 at band p13. The SCA6 CAG repeat is small compared to other polyglutamine disorders, with a pathogenic range of only 19–33 triplets. Unaffected individuals carry alleles of 4–18 repeats. There is an inverse correlation between repeat length and age of disease onset, and minimal intergenerational and somatic instability has been reported for the SCA6 repeat expansion in affected patients.

The 9.8 kb CACNA1A transcript implicated in SCA6 is expressed throughout the CNS, and most highly in cerebellar Purkinje cells and granule neurons. The \( \alpha_{1A} \) subunit is the pore-forming component of the channel, which is important in neurotransmitter release at the synapse. Polyglutamine expansions in this subunit cause variable changes in calcium transmission rates, depending on the system and the \( \beta \) subunit coexpressed. Expansions do not cause a reduction in membrane channel density in HEK293 cells, suggesting that aggregation does not occur
at pathogenic repeat lengths. Ubiquitin-negative neuronal inclusions are visible in the cytoplasm of SCA6 patient Purkinje cells, however.

In addition to displaying a disease allele range that does not overlap with the other polyQ diseases, there are several other reasons to suspect that SCA6 is different from the rest of the polyglutamine repeat group. The first difference is the existence of two disorders, episodic ataxia type 2 (EA2) and familial hemiplegic migraine (FMH), that are both allelic to SCA6, as both are caused by point mutations in the CACNA1A gene. EA2 resembles SCA6 as affected EA2 individuals suffer from a slowly progressive form of episodic ataxia, accompanied by nystagmus, dysarthria, loss of balance and sometimes cerebellar atrophy. EA2 is usually caused by truncation mutations in CACNA1A, resulting in loss-of-function of the calcium channel. CACNA1A knockout mice similarly develop ataxia and late-onset cerebellar atrophy, characteristic of SCA6 and EA2 patient phenotypes. Another reason SCA6 is different from other polyQ disorders is that the CACNA1A protein probably does not undergo a structural change that converts it into an aggregate-prone, beta-sheet adopting, amyloid-like conformer. Indeed, even the largest SCA6 polyglutamine tract is below the threshold required for stable β-pleated sheet formation in other polyglutamine disorders. Consistent with this prediction, channel localization of mutant polyglutamine-expanded CACNA1A protein is not affected in cell culture models, and its channel function is not completely abolished. Cytoplasmic aggregates in patient material are ubiquitin-negative, suggesting that the protein may not be grossly misfolded. This evidence supports a model whereby a dominant-negative loss of function of the P/Q-type voltage-gated calcium channel due to association of the mutant α1A subunit with other subunits causes SCA6. The coincidence that this disorder is caused by a polyQ tract expansion and causes a progressive cerebellar ataxia cannot be ignored, however, more data will be required to soundly refute a possible concomitant toxic gain-of-function effect.

3.8 Spinocerebellar Ataxia Type 7

Spinocerebellar ataxia type 7 (SCA7) is an autosomal dominant, progressive cerebellar ataxia. It is unique among autosomal dominant SCAs, as patients typically develop visual impairment in addition to their cerebellar ataxia. The visual impairment is due to a cone-rod dystrophy that results in retinal degeneration. Patients first develop problems distinguishing colors, but ultimately go blind in this type of retinal degeneration. SCA7 patients may present either with cerebellar ataxia or visual impairment. The likelihood of their presentation is dictated by the size of their CAG repeat disease allele, with larger repeats typically favoring presentation with visual impairment. Affected individuals display prominent dysarthria, and can develop increased reflexes, decreased vibration sense, and oculomotor disturbances. Neuronal degeneration and reactive gliosis occur in the cerebellar cortex, dentate nucleus, inferior olive, pontine nuclei, and occasionally in the basal ganglia. NIs are widespread. Infantile-onset SCA7 has been documented, and in this fatal form of the disease, nonneuronal tissues such as the heart and the kidney are severely affected. SCA7 is caused by a highly polymorphic (CAG)n repeat in the 5’ coding region of the ataxin-7 gene. Unaffected individuals carry 4–35 repeats,
while affected individuals carry 37–306 repeats. The SCA7 trinucleotide repeat is one of the most unstable of all polyglutamine disease genes, sometimes expanding by as much as 250 repeats in a single generation. There is a pronounced paternal expansion bias, with large expansions occurring in male germ cells, frequently causing embryonic lethality that results in reduced transmission. Marked repeat instability can occur in the brain, and produce large somatic expansions on occasion.

Ataxin-7 is a ubiquitously expressed protein of 892 amino acids. It is expressed most highly in heart, skeletal muscle, and pancreas. Expression levels within the CNS are highest in the cerebellum and brainstem. One splice variant, ataxin-7b, is expressed predominantly in the CNS (32). Ataxin-7 contains a functional arrestin domain, a protein interaction domain that is highly selective for phosphorylated forms of its interacting protein(s). This suggests that it interacts with specific phospho-proteins, although none have been discovered to date. It also contains two SH3 domains, which are protein interaction domains that bind proline-rich sequences and mediate a number of cell signaling processes. Ataxin-7 also contains three putative NLSs and one NES.

Ataxin-7 is conserved throughout eukaryotes, and its yeast ortholog SGF73 is part of a multisubunit histone acetyltransferase complex called SAGA (Spt/Ada/Gcn5 acetyltransferase). The human orthologs of SAGA comprise the so-called STAGA (SPT3/TAF9/ADA2/GCN5 acetyltransferase) complex, and are essential transcription coactivators required for the transcription of certain genes. STAGA components immunoprecipitate with ataxin-7. Although pathogenic expansion of ataxin-7 does not alter its ability to be integrated into the STAGA complex, the presence of the polyglutamine-expanded ataxin-7 has a dominant-negative effect upon the GCN5 histone acetyltransferase activity of the STAGA complex, resulting in transcription dysregulation. The transcription dysregulation caused by polyglutamine-expanded ataxin-7 likely causes a disease phenotype by altering the ability of certain transcription factors to activate expression of their target genes.

The best characterized example of transcription dysregulation by polyglutamine-expanded ataxin-7 is its interference with the cone-rod homeobox protein (CRX), a glutamine domain containing transcription factor expressed only in the retina and the pineal gland. Ataxin-7 interacts directly and functionally with CRX, according to studies performed in vitro and in a mouse model of SCA7. Importantly, the interaction between ataxin-7 and CRX appears to involve the glutamine tract regions found in both proteins. Autosomal dominant mutations in CRX can cause a cone-rod dystrophy in humans, further supporting a model in which CRX’s diminished transactivation competence is central to the SCA7 retinal degeneration phenotype. Several transcription factors, including CBP, can be found in SCA7. CRX may be but one of a number of transcription factors whose function is diminished by polyQ-expanded ataxin-7 interaction and dysregulation of STAGA complex-mediated gene expression.

One intriguing feature of SCA7 was discovered upon generation of transgenic mice expressing ataxin-7 with the mouse prion protein promoter. This promoter drives expression in every tissue, with the occasional notable exception of the Purkinje cells of the cerebellum. Despite lack of Purkinje cell expression of ataxin-7, mice
developed a cerebellar ataxia accompanied by degeneration of the Purkinje cells (Fig. 5). This noncell-autonomous degeneration may point to a disease mechanism involving withdrawal of trophic support by communicating neurons (olivary, deep cerebellar, brainstem, or granule neurons) or dysfunction of glutamate transporters expressed by surrounding glia. Damage to inferior olivary neurons or Bergmann glia can indeed cause the degeneration of Purkinje cells, lending credence to this theory.

Another interesting feature of SCA7 that is common to other neurodegenerative disorders is the prominence of morphological and functional degeneration without pronounced apoptosis. Some neurons in SCA7 humans and mouse models exhibit indentations in the nuclear envelope, reduced arborization, ectopy, and increased autophagy. (Autophagy is the bulk degradation of cellular components by a membrane-bound autophagosome and is accelerated by a number of cellular insults.) Caspase-3, a proteolytic enzyme classically associated with apoptosis, is activated at abnormally high levels in SCA7 patient brains. This may suggest that caspase activation, rather than causing apoptosis, is contributing to a nonapoptotic degenerative process.

In normal human neurons, ataxin-7 is variably located in the nucleus or the cytoplasm. In SCA7 patients, ataxin-7 gradually undergoes a shift in localization into NIs. This process has been replicated in SCA7 transgenic mice, in which ataxin-7 immunoreactivity shifts from the cytoplasm to the nucleus, and ultimately forms foci there. In patient’s brains, these NIs colocalize with promyelocytic leukemia (PML) protein, which is an integral part of nuclear bodies (NBs). NBs are associated with transcription regulation and the
ubiquitin-proteasome pathway; thus, accumulation of ataxin-7 in NBs may represent an attempt by the cell to degrade misfolded protein.

3.9 **Spinocerebellar Ataxia Type 17**

Spinocerebellar ataxia type 17 (SCA17) is the most recently discovered polyglutamine repeat disease. It is an autosomal dominant, progressive disorder characterized by dementia as well as cerebellar ataxia and involuntary movement abnormalities. Its symptoms are diverse and heterogeneous, but typically begin with behavioral disturbances and cognitive impairment. This is followed by ataxia, rigidity, dystonia, hyperreflexia, and rarely parkinsonism. Neuropathologically, patients may suffer degeneration of the cortex, cerebellum (including Purkinje cells), inferior olive, caudate nucleus, and medial thalamic nuclei. SCA17 is caused by the expansion of a coding (CAG)n repeat in the TATA-binding protein (TBP) gene on chromosome 6q27. Unaffected individuals carry 25–48 repeats, while affected individuals carry 43–66 repeats. The area of overlap between affected and unaffected alleles indicates incomplete penetrance at intermediate repeat lengths. The pathogenic threshold is higher than for most other polyglutamine disorders, yet there is an inverse correlation between repeat length and age of onset.

TBP is a ubiquitously expressed transcription initiation factor that is a core component of the RNA polymerase II transcription factor D (TFIID) complex. TBP possesses DNA binding activity in the TFIID complex, and is therefore required for the transcription of numerous genes. SCA17 patients have NIs immunoreactive for TBP, polyglutamine, and ubiquitin. Given TBP’s central role in transcription regulation, the basis of SCA17 disease pathogenesis is proposed to involve transcription dysregulation, but this is yet to be proven.

3.10 **Role of Aggregation in Polyglutamine Disease Pathogenesis**

In most polyglutamine disorders and in many neurodegenerative diseases in general, protein aggregation is a prominent feature. It occurs in almost all polyglutamine diseases, despite the lack of domain or structural similarity between the different disease proteins. Aggregates have long been considered reliable indicators of disease, although their pattern and onset often do not correspond well with the cell-type specificity of disease pathology. As the role of aggregation in pathogenesis has been one of the most hotly debated issues in the field of neurodegeneration, it may have implications for Alzheimer’s disease, Parkinson’s disease, and amyotrophic lateral sclerosis. Why? There are several characteristics of polyglutamine-mediated aggregation that are presumably common to all of the disorders that show aggregation. First of all, aggregates are rich in β-pleated sheets and have many of the properties of amyloid. This is supported by Congo red birefringence and immunoreactivity with anti-amyloid antibodies. Various studies have shown that the conformational change is associated with the production of visible aggregates rather than in the soluble nonaggregated phase, where polyglutamine tracts are thought to remain in a random coil conformation.

*In vitro*, the kinetics of aggregation are consistent with nucleated-growth polymerization, in which the rate-limiting step is
the formation of misfolded peptide assemblies, often referred to as oligomers. A mechanism involving nucleated-growth polymerization would predict that only one (or a few) visible aggregates would appear in each affected cell, and this appears to be the case for most polyglutamine disorders. In the cellular milieu however, where membranes and intermolecular interactions compartmentalize proteins, this prediction may not be valid. The discovery of polyglutamine microaggregates confirms this suspicion. The threshold for aggregation of polyglutamine tracts is similar to the disease threshold for most polyglutamine diseases. In vitro, longer polyglutamine tracts have a lower concentration threshold for aggregation and nucleate more quickly than shorter tracts. This is one possible explanation for the correlation between tract length and disease onset and progression.

Some data support a role for aggregation in polyglutamine disease pathogenesis. Some proponents of this theory invoke the fact that aggregates sequester many proteins besides the disease protein. Intermolecular interactions between transcription factors often involve glutamine tracts or glutamine-rich regions. Some studies have shown that normal proteins with glutamine tracts or glutamine-rich regions are enriched in polyglutamine aggregates in cell culture and animal models. Several transcription factors colocalize with such aggregates, including CBP, TBP, and numerous TBP-associated factors (TAFs). CBP is responsible for the prosurvival effects of BDNF, and its soluble concentration is lowered in HD patient’s brains. Postnatal mice lacking CREB and its homolog CREM develop a progressive degeneration of the hippocampus and striatum.

Titration of enzymes and factors required for protein refolding and degradation away from the soluble phase and into aggregates has been proposed as a potential cause of cell toxicity in neurons with aggregates. Caspase activation is another way in which aggregation could be linked to pathology. Caspase recruitment into aggregates can lead to their activation, which could result in dysfunction or cell death in neurons. Further supporting the role of aggregates in polyglutamine protein toxicity, injection of preformed polyglutamine aggregates into the nuclei of cultured cells causes cell death, while the injection of nonpolyglutamine aggregates does not. At the same time, there is strong evidence that soluble polyglutamine protein, rather than aggregates, is the primary source of toxicity. Importantly, the pattern of aggregates observed in human patients often does not correlate with the pattern of neuronal dysfunction. For example, in the striatum of HD patients, large interneurons contain aggregates more frequently than medium spiny neurons, yet the former neurons are largely spared, while the latter are most vulnerable. HD transgenic mice expressing full-length mutant htt will develop inclusions in many brain regions many of which are neuropathologically unaffected, while few inclusions are detected in the striatum, the region of the brain displaying the most prominent pathology.

Aggregation and toxicity have been directly dissociated in other polyglutamine disease model systems. Studies of the SCA1 knockin mouse model revealed that neurons lacking aggregates were more susceptible to dysfunction and demise, while those neurons displaying prominent aggregates were protected. Similarly, SCA7 transgenic mice exhibit retinal pathology before the occurrence of visible aggregates. Finally, in a very provocative study, SCA1
transgenic mice lacking the ubiquitin ligase E6-AP were significantly less capable of forming large and numerous aggregates in neurons in comparison to SCA1 transgenic mice on a wild-type background, but displayed earlier onset of an ataxic phenotype and accelerated neurodegeneration. This suggests that the aggregates may be protective, although the toxicity of compromising the proteasome may have contributed to the accelerated phenotype. Thus ensued a contentious debate over the role of aggregate formation in polyglutamine disease – with some workers espousing the view that aggregates were responsible for disease pathology, others suggesting that the aggregation process was a protective coping mechanism of the cell and thereby beneficial, and still others insisting that aggregates were incidental and irrelevant. This debate was complicated by the existence of “microaggregates,” small clumps of aggregated protein visible only at the electron microscope level.

Over the last few years, studies deconstructing polyglutamine tract aggregation into a multistep process suggest a parsimonious explanation for the divergent views. Using a variety of biophysical approaches, including transmission electron microscopy (TEM), Fourier transform infrared spectroscopy (FTIR), and atomic force microscopy (AFM), one study dissected the process of huntingtin (htt) exon 1 peptide aggregation and found evidence for a number of sequential morphological and structural intermediates (Fig. 6). By showing that misfolded htt exon 1–44Q adopted intermediate structures, such work opened up the possibility that intermediates (not visible at the light microscope level) are the toxic species and that the ultimate visible aggregated forms of htt exon 1–44Q are neutralized versions of mutant protein. To examine the role of aggregate formation in polyglutamine toxicity, another group then tracked survival time versus diffuse, soluble htt protein expression levels in htt-transfected striatal neurons undergoing aggregate formation. Comparison of neurons that developed aggregates over time versus those that did not confirmed that the level of soluble nonaggregated mutant polyglutamine protein was the more reliable predictor of cellular toxicity. Such studies have shifted our attention to the role of the intermediates (oligomers, protofibrils, etc.) as the toxic species instead of the final, visible aggregates.

A reasonable model for polyglutamine toxicity predicts that the process starts with a protein that misfolds because of the presence of an expanded polyglutamine tract. The misfolded protein is initially detectable in the soluble phase due to the cell’s ability to maintain the protein in a properly folded state and direct it to the degradation machinery. However, the refolding capacity of the cell is ultimately exceeded, and since the degradation machinery cannot turnover the misfolded mutant polyglutamine protein, accumulation occurs. Misfolded polyglutamine proteins can spontaneously change their structural properties and adopt abnormal conformations. These abnormally folded proteins can then nucleate, forming oligomers. Oligomers then form protofibrils that grow into fibrils. The transition to the fibril stage is characterized by the attainment of a β-sheet structure, so at this point the structures are amyloid-like. Fibrils then grow into fibers (also known as microaggregates), which then assemble into aggregates visible under a light microscope. According to such a model, blocking an intermediate step could be therapeutically effective. Consistent with this, one
group has nicely shown that Congo red can bind polyglutamine peptides once they are amyloid-like and prevent their conversion into fibers, while an independent group has shown that Congo red delivery to an HD mouse model is a highly effective treatment intervention.

One satisfying aspect of this model is that it allows us to simultaneously view aggregates as harmful, protective, and innocuous. How? First, aggregates clearly must be toxic as their creation is predicated upon the production of earlier intermediate toxic forms. At the same time, the final visible aggregates are less toxic than the earlier intermediates, so anything that enhances their sequestration into visible aggregates is beneficial. Finally, since oligomers are difficult to detect, it is often not possible to know whether cells are successfully sequestering the toxic precursors into aggregates or if high levels of toxic intermediates are building up. So, aggregates are thus also incidental, since their presence does not provide us with insight into the steady state levels of the toxic precursors. In conclusion, advances in our understanding of the role of aggregate formation in polyglutamine disease processes suggest that aggregates may be simultaneously viewed as harmful, beneficial, and incidental.
3.11 Protein Context

The importance of protein context in the pathogenesis of polyglutamine disorders is a subject of great interest to many researchers, because it is directly related to the mechanisms of toxicity in each disease. It is evident in the polyglutamine disorders that protein context plays a role in cell-type specificity. For example, a polyglutamine expansion in the Htt protein causes a severe degeneration of the striatum and cortex, while the same size glutamine tract in ataxin-1 causes a degeneration of structures in the cerebellum and brainstem, while sparing the striatum and cortex. This cannot be attributed to gross differences in expression patterns, since both proteins are pan-neural (Fig. 7).

One effect that protein context may have on pathology is to affect subcellular localization. For example, the presence of a functional NLS or NES dictates preferential subcellular localization in the nucleus or cytoplasm. Mice expressing expanded ataxin-1 with a mutated NLS do not develop the SCA1 phenotype, while those without mutated NLSs do. Interaction domains also affect polyglutamine protein toxicity. The ability of expanded polyglutamine proteins to interact with other molecules is a promising avenue in the search for mechanisms of cell-type specificity. Yeast two-hybrid screens and other techniques have yielded interaction partners for a number of polyglutamine proteins. For example, Htt’s interactions with transcription factors may prove to be central to its pathological effects. At the same time, it is also evident that polyglutamine tracts are innately toxic. Pure polyglutamine tracts cause toxicity in cell culture. Mice that express a glutamine tract of 150 amino acids in the Hprt protein, which does not naturally contain any such tracts, exhibit progressive neurological deficits and NILs.

Thus, it appears that pathology in each polyglutamine disease is due to a gain of function of the glutamine tract that is then modulated by the protein context in which it resides.

Fig. 7 The conundrum of cell-type specificity in the polyglutamine diseases. Although the different polyglutamine disease proteins are expressed throughout the central nervous system, only select populations of neurons degenerate in the different disorders. The principal regions of selective vulnerability are shown for certain of the polyglutamine diseases.
3.12 Transcriptional Dysregulation

Transcriptional dysregulation is one theory of polyglutamine toxicity that has gained support from many lines of evidence in several disorders, and it appears increasingly likely that it is one of the fundamental causes of pathogenesis. Microarray technology and other genomic techniques are facilitating rapid advances in this area of the polyglutamine field. The majority of polyglutamine disorders are caused by proteins that are either transcription factors/cofactors or interact closely with transcription factors. TBP and AR are transcription factors, atrophin-1 is a transcriptional corepressor, ataxin-7 is part of a transcriptional coactivator complex, and ataxin-1, htt, and ataxin-3 all interact with various transcription factors. It is therefore reasonable to suspect that an abnormally long polyglutamine tract could interfere with the transcriptional activity of these proteins and/or their interactors.

CBP is one of the most studied transcription factors in the polyglutamine field. It is a transcription activator that mediates part of the cellular response to cAMP, and it can interact with numerous polyglutamine proteins in the soluble or insoluble phase, including huntingtin, ataxin-3, ataxin-7, and AR. Expanded htt represses CBP-regulated genes, and overexpression of CBP causes a considerable rescue of the cell death phenotype in HD and SBMA cell culture models. Postnatal mice lacking CBP’s upstream activator CREB and its homolog CREM display a profound degeneration of the striatum and hippocampus. Thus, interference with CBP appears to be a common theme in polyglutamine pathology.

Another common theme in the transcription dysregulation equation is the tendency for the polyglutamine disease proteins to alter transcription factor/coactivator-mediated histone acetyltransferase (HAT) activity. The ability of a gene to be transcribed depends upon its chromatin structure, and thus upon the degree of histone acetylation in its vicinity. This is because acetylated histones cause chromatin to be in an “open,” transcription-friendly conformation. Acetylation status depends upon the balance between the activity of HATs and their countervailing counterparts, the HDACs. CBP, p300, and p300/CBP-associated factor (PCAF) are all HATs that are inhibited by polyglutamine-expanded htt exon 1 peptide and ataxin-3. In cell culture, expression of a mutant htt fragment reduces global histone acetylation.

Finally, according to a very recent study, ataxin-7 directly interacts with GCN5 as part of the STAGA complex, and upon polyglutamine expansion, mutant ataxin-7 causes a dominant-negative effect upon GCN5 HAT activity. This results in CRX transcription interference that may contribute to the SCA7 retinal degeneration phenotype. Inhibiting HDACs with drugs known as HDAC inhibitors (HDAC Is) attenuates the phenotype of HD and SBMA in mouse and fly models, presumably by shifting the cells’ acetylation status. HDAC Is such as sodium butyrate and especially suberoylanilide hydroxamic acid (SAHA) have thus emerged as possible candidates for therapeutic trials in human polyglutamine disease patients.

3.13 Proteolytic Cleavage

The occurrence of proteolytic cleavage in polyglutamine diseases first became apparent when it was shown in HD that htt can be cleaved by extracts derived
Triplet Repeat Diseases

Fig. 8 Proteolytic cleavage in the polyglutamine diseases. Production of amino-terminal truncated proteins is observed in many of the polyglutamine diseases, as is shown here for SCA7. In this experiment, nuclear fractions of retinal lysates from age- and sex-matched nontransgenic (nt), ataxin-7 CAG-24 (24Q), and ataxin-7 CAG-92 (92Q) mice were prepared and immunoblotted with an antibody directed against the amino-terminal region of the protein ataxin-7. As shown here, in addition to soluble full-length ataxin-7 (which is typically reduced in cells where it is aggregating into insoluble inclusions), an ∼50–60 kD fragment is detected. (Adapted from Garden, G.A., Libby, R.T., Fu, Y.H., Kinoshita, Y., Huang, J., Possin, D.E., Smith, A.C., Martinez, R.A., Fine, G.C., Grote, S.K., et al. (2002) Polyglutamine-expanded ataxin-7 promotes noncell-autonomous Purkinje cell degeneration and displays proteolytic cleavage in ataxic transgenic mice, J. Neurosci. 22, 4897–4905. Used with permission of the Journal of Neuroscience.)

from apoptotic cells. Subsequent publications showed that only amino-terminal epitopes of htt protein are detectable in aggregates. SBMA, DRPLA, and SCA7 were then added to the list of diseases in which aggregates are only immunoreactive for a glutamine-containing fragment of the disease protein. Since then, evidence of proteolytic cleavage has been published for nearly all known polyglutamine disorders. Experimental studies revealed that inhibition of caspase cleavage reduces aggregate formation and toxicity in a cell culture model of HD, underscoring the relevance of proteolytic cleavage in polyglutamine disease pathogenesis. Cleavage promotes aggregation and/or toxicity in SCA3 and SCA7 (Fig. 8). Polyglutamine tract containing htt fragments are more toxic in cell culture than full-length htt, and the most widely used HD mouse model expresses only exon 1 of the htt gene.

There are several mechanisms by which cleavage may modulate polyglutamine neurotoxicity. Abnormal proteolysis of polyglutamine proteins may lead to a toxic species that can cause damage in the soluble or insoluble phase. Alternatively, proteolysis may be a normal event in protein turnover and the inability to clear cleaved protein may be the problem. Another possibility is that soluble polyglutamine proteins may cause the activation of proteases such as caspases, which then cleave the disease protein and send the cell on a path to degeneration and ultimately apoptosis. As we will see, each theory has support and not all are mutually exclusive.

Abnormal proteolysis of polyglutamine disease proteins is one possible mechanism of toxicity. Some studies have shown that polyglutamine repeat length modulates susceptibility to proteolytic cleavage, but results have been inconclusive. Few studies have been performed using human brain tissue, however. Considerable evidence exists to show that polyglutamine tracts themselves are resistant to degradation by mammalian proteasomes. The 20S and 26S eukaryotic proteasomes are incapable of cutting within polyglutamine tracts longer than 9 amino acids in vitro. The inability of the proteasome to digest glutamine tracts may contribute to the toxicity and aggregation of expanded polyglutamine peptide fragments.
Proteolysis may be a normal part of the turnover of wild-type and mutant htt. One group published evidence that caspase-3 cleaves both mutant and wild-type htt in the cytoplasm of HD brains, suggesting that this particular cleavage is a normal event in its turnover. Since this fragment was located exclusively in the cytoplasm, it also implies that further cleavage occurs before the nuclear translocation of htt. Another study suggested that pepstatin-sensitive aspartyl endopeptidases such as the presenilins and cathepsins D and E may be involved in the normal turnover of htt. This process, which may normally aid in maintaining the balance between htt synthesis and degradation, could generate a toxic, highly aggregation-prone (and hence intermediate oligomer/protofibril-prone) species when it cleaves mutant htt.

Another possible source of toxic cleavage products is the ubiquitin-proteasome system, which is responsible for labeling and digesting misfolded proteins, among other things. Polyglutamine tracts beyond a threshold of about 35 glutamines adopt an abnormal $\beta$-pleated sheet conformation, which may cause the host protein to be ubiquitinated and thus targeted for degradation. In support of this theory, a common feature of polyglutamine diseases is the presence of ubiquitin-positive inclusions. Arguing against this idea is an experiment showing that in a cell culture model of SCA1, increased proteasome degradation due to the introduction of a degradation signal reduces aggregates and toxicity. Another potential source of toxic fragments is the autophagy pathway, which is responsible for the bulk degradation of cellular components. According to one study, degradation of htt by autophagy and the autophagosome-associated cathepsin D creates toxic htt fragments.

There are many proteases that are suspected to play a role in the proteolytic cleavage of polyglutamine proteins, the most studied of which are the caspases. Caspases are cysteine-dependent aspartyl proteases that play a crucial role in apoptosis, but are increasingly being investigated for their nonapoptotic functions. In healthy cells, caspases are present predominantly as inactive proenzymes, which must be cleaved for full activity. Caspase-3 is the only protease that has been shown to cleave htt in HD patient’s material, although caspases 1 and 6 have also been implicated in cell culture and in vivo. Inhibiting the cleavage of htt by any of these three proteases is protective, and an HD mouse model expressing dominant-negative caspase-1 showed significantly delayed disease onset. Caspase-3 has also been implicated in the proteolytic cleavage of atrophin-1, ataxin-3 and AR. Caspases are activated in response to a number of cellular insults, and in postmitotic cells such as neurons where inhibitors of apoptosis proteins (IAPs) are highly expressed, this may cause cellular damage that does not result in classical apoptosis.

Calpains are another family of proteases implicated in the cleavage of polyglutamine proteins. They are calcium-activated cysteine proteases that exist predominantly as proenzymes. As mouse models of HD and material from HD patients exhibit abnormal mitochondrial calcium regulation, some investigators have proposed that altered calcium flux is the mechanism of calpain activation. Indeed, several cell culture studies support the role of calpain proteases I, II, and “m” in htt cleavage. As mentioned previously, other aspartyl proteases such as the Alzheimer’s disease-associated presenilins and autophagy-associated cathepsins have also been implicated in htt cleavage.
4 Type 2: the Loss-of-function Repeat Diseases

4.1 Fragile X Syndrome

Fragile X syndrome (FRAXA) is an X-linked disorder with a prevalence of 1 in 4000 in males and 1 in 8000 in females, making it the most common form of inherited mental retardation. Neurological presentation frequently includes mild to severe mental retardation, hyperactivity, poor eye contact, high-pitched speech, and flapping or biting hand movements. Physical signs in males include long, prominent ears, and jaws, macrocephaly, postpubescent macroorchidism, and occasionally, connective tissue abnormalities. The mutation responsible for FRAXA is typically an expansion of a polymorphic (CGG)n repeat found in the 5′-untranslated region of the fragile X mental retardation-1 (FMR1) gene. Expanded chromosomes have a folate-sensitive fragile site at Xq27.3 that can be viewed under a light microscope under special cell culture conditions. Normal alleles contain 6–53 triplets punctuated by one or more AGGs, which are considered to have a stabilizing influence on the repeat. Disease alleles contain expansions beyond 200 and up to 2000 repeats, with no AGG interruptions. Pathogenic expansions result exclusively from maternal transmission. FRAXA appears to be a loss-of-function disorder, since deletion of FMR1 and loss-of-function point mutations can also cause FRAXA. This view is further supported by the FMR1 knockout mouse, which reproduces certain aspects of the human disorder.

FRAXA patients have reduced levels of the FMR1 gene product, FMRP, and there is a linear correlation between reduced protein levels and IQ test scores. Expansion of the disease allele results in hypermethylation of the (CGG)n tract, which spreads to a nearby CpG island in the FMR1 promoter region. Some of the proposed secondary structures formed by the FRAXA repeat contain C–C mispairs, which are good targets for human DNA methyltransferase. Another theory invokes the RNAi protein Dicer’s ability to cleave CGG repeat RNA, postulating that the resulting siRNAs recruit DNA methyltransferases to the FMR1 locus. Affected alleles also display condensed chromatin, loss of histone acetylation and increased histone methylation. These data support a process whereby DNA methylation recruits transcription silencing machinery, which subsequently suppresses FMR1 transcription in the nearby promoter region. Interestingly, while premutation carriers had been viewed as perfectly normal for decades, recent work indicates that some premutation carriers (60–200 uninterrupted CGG repeats) display a phenotype distinct from FRAXA. Females have a predisposition to premature ovarian failure (POF), and males may develop late-onset ataxia and tremor with the presence of neuronal intranuclear inclusions (NIIs) that consist of RNA. (This new FXTAS is discussed in a separate section.) Expression studies have shown that premutation carriers may express up to seven times more FMR1 mRNA than normal individuals. This upregulation is probably not due to compensation for loss of function, because individuals with point mutations resulting in loss of function of FMRP do not have higher levels of the transcript. Abnormal transcript levels due to premutations or full mutations are thought to be as a result of the expansion’s effect.
on transcription initiation rather than on mRNA stability.

The FMR1 promoter lacks a functional TATA box and initiator sequence. Four functional transcription factor binding sites have been identified in normal individuals: two sites binding USF1/USF2 and Nrf-1, and two GC boxes that bind members of the “Sp” family of transcription factors. Sp1 and Nrf-1 binding is disrupted in cells from FRAXA patients. The FMR1 gene encodes the widely expressed FMRP, which is an RNA-binding protein most highly expressed in the brain and testes. It is 60% homologous to two other proteins, FXR1P and FXR2P, with which it interacts. The protein is thought to bind approximately 4% of all brain mRNA transcripts, through its RGG box domain. It selectively recognizes RNA-containing hairpin or tetraplex secondary structures (“G-quartet”), and shuttles into and out of the nucleus and associates with polyribosomes in messenger ribonucleoprotein complexes (mRNPs). The FMRP-containing mRNP complexes also contain Pur α and mStaufen, proteins involved in the transport of neuronal granules. These granules, which contain RNA and associated proteins, are transported to dendritic spines in a metabotropic glutamate receptor 5 (mGluR5)-dependent manner. FMRP can suppress the translation of certain transcripts in vitro and in vivo. FMRP may thus regulate the transport, localization, and translation of certain mRNAs in an activity-dependent manner (Fig. 9). In Drosophila, FMRP interacts with components of the RNAi machinery, which is involved in gene silencing and thus translational control. Although FMRP does not seem to affect the siRNA pathway, it may regulate microRNAs (miRNAs), which are noncoding RNAs thought to control the translation of mRNAs by binding to their 3’-untranslated region. FMRP associates with miRNAs, and proteins in miRNA-containing complexes in mammals, interactions that may be relevant to FMRPs regulation of translation.

FMRP is important for the development of dendritic spines and synaptic plasticity. It regulates the expression of MAP1B, an important regulator of microtubule stability. FMRP knockout mice have abnormally high levels of MAP1B, resulting in increased microtubule stability. This may affect the development of dendrites and/or dendritic spines. The absence of FMRP in hippocampal neurons results in immature dendritic spine morphology and delayed synaptic connections, perhaps contributing to the neurological phenotype observed in FRAXA. FMRP may also affect mGluR5-dependent long term depression (LTD). LTD is a process by which neuronal activity can cause a lasting desensitization of neurons to depolarization. LTD-associated protein synthesis at synapses may be enhanced by mGluR5 activation, while it appears to be suppressed by FMRP. This is supported by evidence of enhanced LTD in the FMR-1 knockout mouse. FMRP suppression and mGluR5 activation of LTD-dependent local protein synthesis may be opposing forces that are out of balance in FRAXA, perhaps accounting for part of the cognitive abnormalities.

4.2 Fragile XE Mental Retardation

Fragile XE mental retardation (FRAXE) is an X-linked disorder with a prevalence of about 1–4% that of FRAXA. It accounts for approximately two-thirds of families with nonspecific X-linked mental retardation (MRX), a classification in which mental retardation is the only consistent
Advances in the understanding of FMRP's normal function. FMRP binds to recently transcribed messenger RNAs in the nucleus to form a ribonucleoprotein complex (mRNP). The FMRP–mRNP complex moves out of the nucleus and then either directly associates with ribosomes or interacts with the RNA-induced silencing complex (RISC). The FMRP–mRNP complex is also transported to dendrites. Whether in the perinuclear cytosol or in dendrites, FMRP is believed to regulate protein translation. (From Jin, P., Alisch, R.S., Warren, S.T. (2004) RNA and microRNAs in fragile X mental retardation, Nat. Cell Biol. 6, 1048–1053. Used with permission of Nature Cell Biology).
clinical feature. Other characteristics are variable and may include hyperactivity, mild facial hypoplasia and nasal abnormalities. FRAXE is caused by the expansion of a (GCC)$^n$ repeat in the promoter region of $FMR2$, a gene 600 kb downstream of $FMR1$. It also overlaps with the putative promoter of a gene called $FMR3$, transcribed in the opposite direction. Both maternal and paternal transmission can result in expansion. Normal individuals bear 6–35 GCC repeats, premutation carriers bear 61–200 repeats, and full mutation carriers bear >200 repeats. Pathogenic expansion results in a folate-sensitive fragile site at the disease locus.

In mice, $FMR2$ mRNA is expressed in adult and fetal brain, kidney, lung, and placenta, with highest brain levels in the hippocampus and amygdala. Transcripts of both $FMR2$ and $FMR3$ are reduced beyond detection in FRAXE patients. This may be due to a methylation-dependent silencing process similar to FRAXA. The presence of the repeat within the preinitiation region of the $FMR2$ promoter may also suggest a more direct disruption of transcription. $FMR2$ mRNA is highly expressed in the hippocampus, an area critical for learning and memory. Its paralogs AF4 and LAF4 are both transcription transactivator proteins, and the $FMR2$ protein seems to be a potent transcription activator itself. Furthermore, the $FMR2$ protein is nuclear, consistent with its proposed role in transcription regulation. The Drosophila ortholog of $FMR2/AR4$, Lilliputian, is essential for proper organ development, and its loss of function is lethal. $FMR2$ knockout mice display impairment in the conditioned fear test and enhanced LTP (long-term potentiation) in the hippocampus. The role of $FMR3$ in FRAXE, if any, is unknown. The mechanism of FRAXE pathogenesis may therefore involve silencing of the human $FMR2$ gene resulting in altered transcription in many parts of the developing nervous system and mature brain. The neurological deficits characteristic of FRAXE may involve effects of altered transcription on the hippocampus, amygdala, and possibly other brain structures.

### 4.3 Friedreich’s Ataxia

With a prevalence of about one per 50,000 individuals in the Caucasian population, Friedreich’s ataxia (FRDA) is the most common inherited ataxia in this ethnic group. It is a multisystem degenerative disease that is unusual among the triplet repeat disorders due to its autosomal recessive inheritance. Neurological symptoms include gait, limb and truncal ataxia, loss of position and vibration senses, diminished tendon reflexes, and dysarthria. Neuropathology changes include degeneration of the posterior columns of the spinal cord, loss of large primary sensory neurons in the dorsal root ganglia (DRG), and mild, late-onset degeneration of the cerebellar cortex. Other common clinical features are cardiomyopathy, diabetes mellitus, scoliosis, and other skeletal abnormalities. Patients often present with symptoms in childhood, become wheelchair-bound by their late teens or early twenties, and have reduced lifespans due to cardiac failure. Adult presenting patients, however, can have nearly normal lifespans with more protracted progression and less severe nonneuronal involvement. Indeed, until the identification of the causal mutation, many of these adult-onset cases went undiagnosed.

FRDA results from the expansion of a polymorphic (GAA)$^n$ repeat in the first intron of the gene $X25$, now known as frataxin. FRDA patients have at least one
Triplet Repeat Diseases

expanded allele. For the disease to occur, the second allele must contain either an expanded repeat, or rarely, a copy of frataxin containing a loss-of-function mutation. No patients have been described with loss-of-function allele mutations in both FRDA genes. This is probably due to prenatal lethality, as frataxin-null mice die in utero.

Disease severity and onset age are determined by the size of the (GAA)_n repeat of the smaller expansion allele. Unaffected individuals carry at least one allele between 7–38 repeats, while affected individuals carry two alleles of 66–1700 repeats. Interruptions of the (GAA)_n repeat result in later onset and attenuated presentations. GAA expansion at the FRDA locus can result in enormous repeats in a single generation. Maternal transmission can result in expansion or contraction, while paternal transmission results primarily in contraction.

FRDA patients exhibit reduced frataxin RNA and protein levels, and evidence suggests a defect in transcription or RNA maturation. Current models propose that frataxin RNA elongation may be disturbed by triplex structure formation between expanded DNA strands during transcription. DNA triplex is formed when one strand of a double-stranded DNA molecule folds back upon itself, and interacts with two previously annealed strands. This creates a local structure comprising three DNA strands held together by hydrogen bonds. Certain sequences favor this process, among them extended (GAA)_n repeats. Triplex formation in the first intron would presumably affect transcript elongation but not initiation. In support of this model are in vitro transcription experiments demonstrating no effect of repeat length on transcript initiation.

Frataxin is a 210 amino acid protein that is well conserved from prokaryotes to mammals. It contains mitochondrial targeting signal sequences and localizes to the mitochondrial matrix. Frataxin expression occurs at the primary sites of pathology: dorsal root ganglia, spinal cord, sensory nerves, heart, and pancreas. These are tissues that rely on high levels of oxidative metabolism and consequently are rich in mitochondria. In addition, such cell types are often postmitotic, meaning that most dividing cell types are spared in FRDA patients.

Studies with FRDA patient material have demonstrated an increased heart iron content and a deficiency in iron–sulfur cluster-containing proteins, including aconitase, a protein involved in iron homeostasis. In addition, fibroblasts derived from FRDA patients are abnormally sensitive to iron and hydrogen peroxide induced stress. Ablation of the yeast frataxin homolog results in respiratory dysfunction, abnormal accumulation of mitochondrial iron, impaired biogenesis of iron–sulfur proteins, and increased sensitivity to oxidative stress. Conditional knockout of frataxin in striated muscle results in a heart-specific phenotype resembling the cardiac abnormalities in human FRDA (Fig. 10). This supports a genetic mechanism involving loss of function of the disease protein. Frataxin may be part of a complex that delivers iron to Iron–sulfur clusters (ISCs), which are cofactors essential to the activity of many important cellular proteins. The accumulation of iron outside the mitochondria, although probably not central to FRDA pathogenesis, supports a deficiency in iron delivery to ISCs. Among ISC-dependent cofactors are proteins involved in mitochondrial electron transport and thus respiration. Disturbances in oxidative metabolism are often associated with increases in the production of
Triplet Repeat Diseases

Fig. 10 Friedreich’s ataxia mice display cardiac muscle pathology. (a) Transmission electron microscopy of mice lacking expression of frataxin in only their muscle reveals the abnormal accumulation of lipid droplets (L) in cardiac muscle at 4 weeks of age. (b) By 7 weeks of age, mitochondria appear abnormal and iron deposits are visible (arrows). (c) With further progression, large vacuoles emerge (inset-top) as mitochondria undergo continued prominent degeneration. Ultimately, the mitochondria become engorged with electron-dense material consistent with iron deposits. (From Seznec, H., Simon, D., Monassier, L., Criqui-Filipe, P., Gansmuller, A., Rustin, P., Koenig, M., Puccio, H. (2004) Idebenone delays the onset of cardiac functional alteration without correction of Fe-S enzymes deficit in a mouse model for Friedreich ataxia, Hum. Mol. Genet. 13, 1017–1024, used with permission of Human Molecular Genetics).

toxic reactive oxygen species (ROS). FRDA therefore is thought to result from reduced frataxin levels, leading to abnormal iron–sulfur metabolism, mitochondrial dysfunction, oxidative stress, and tissue degeneration. FRDA is thus caused by the dynamic mutation of a nuclear-encoded mitochondrial protein. The disease shares features with classic mitochondrial disorders such as MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes syndrome) and MERRF (myoclonus epilepsy with ragged red fibers), which are caused by stable mutations to mitochondrial-encoded proteins. FRDA appears to be a classic mitochondrial disorder with an unusual genetic basis – dynamic mutation.

Given the data suggesting a role for oxidative metabolism and iron transport in FRDA pathogenesis, antioxidants and iron transport molecules are being evaluated as treatments for this disease. Several clinical trials have been conducted in FRDA
patients using idebenone, a synthetic analog of coenzyme Q10 and a potent antioxidant. Some studies have shown substantial improvements in the heart function of patients, although the lack of randomized placebo controls and the variability of the disorder render the results controversial. A recent conditional knockout mouse model of FRDA also supports the use of idebenone as a treatment for the disorder, making treatment of FRDA patients with this antioxidant a distinct possibility.

4.4 Progressive Myoclonus Epilepsy Type 1

Progressive myoclonus epilepsy type 1 (EPM1) is a rare, autosomal recessive neurological disorder most prevalent in Finland and parts of North Africa. Its most prominent symptoms are progressive photosensitive myoclonus and tonic-clonic epilepsy. Some patients also experience a progressive cerebellar ataxia and cognitive decline. Neurodegeneration occurs in the thalamus, spinal cord, and the Purkinje and granular neurons of the cerebellum. Typical age of onset is 6 to 16 years of age. EPM1 is sometimes caused by missense mutations in the cystatin B gene (CSTB); however, analysis of affected patients lacking such mutations revealed a dodecamer repeat expansion upstream of CSTB. The dodecamer repeat is located between 66 and 77 bp 5′ of two putative transcription start sites, and has the sequence CCCCGC-CCCGCG. Unaffected individuals have a repeat number of 2–3, premutation carriers have 12–17 repeats, and affected individuals carry 30–150 repeat alleles.

CSTB is a highly conserved gene in the cystatin family of cysteine protease inhibitors. CSTB mRNA is ubiquitous, with high transcript levels in the hippocampus. The protein binds to and inhibits lysosomal proteases such as cathepsins B, H, L, and S. Pathogenic expansion of the repeat causes a reduction in CSTB transcription in some cell types. Cstb knockout mice develop symptoms similar to those seen in EPM1 patients, and missense mutations causing the disease disrupt the ability of CSTB to bind cysteine proteases. Together, these data suggest a loss-of-function mechanism for EPM1. Lowered inhibition of cysteine proteases may result in neuronal damage, causing the phenotype observed in EPM1.

There are several ways in which the repeat might reduce CSTB expression. The first is that reduced gene expression may be due to an increase in the distance between promoter elements and the transcription start site. AP-1 binding sites have been located upstream of the repeat. One study demonstrated a two- to fourfold reduction in promoter activity when the repeat is expanded. An equivalent reduction also occurred when similarly sized heterologous DNA was used in place of the expanded repeat. Independent studies have confirmed a large reduction in promoter activity due to repeat expansion. Another way the repeat could affect CSTB expression is through alterations in chromatin structure. The repeat region is G-C rich and might exclude nucleosomes or promote the formation of secondary structures. Abnormal DNA secondary structures have been observed in the EPM1 repeat region. These alterations could affect the expression of CSTB or other genes in the vicinity. A third possible mechanism for expression changes lies in the sequence of the repeat region. Each dodecamer repeat contains a GC box, which could act as an Sp1 binding site in vivo. In cases of abnormally large repeat numbers,
increased Sp1 binding could contribute to the suppression of CSTB transcription.

5
Type 3: the RNA Gain-of-function Repeat Diseases

5.1 Myotonic Dystrophy Type 1

Myotonic dystrophy or DM is an autosomal dominant, multisystem disorder with a prevalence of 1 in 100,000 worldwide, and 1 in 8000 in European and North American Caucasian populations. Patients typically present with proximal or distal muscle dysfunction including weakness, pain, and myotonia (failure of muscle relaxation). DM exhibits a combination of characteristic symptoms: cardiac conduction abnormalities, subcapsular iridescence cataracts, and unusual endocrine changes. Other features include testicular atrophy, type II diabetes, and late-onset cognitive impairment. In its most severe form, congenital DM, mental retardation, craniofacial deformities, and other developmental abnormalities are present. It exhibits both paternal and maternal transmission, although there is an almost exclusive maternal transmission in congenital DM. Myotonic dystrophy type 1 (DM1) is caused by a (CTG)n expansion in the 3′-untranslated region of the gene, dystrophica myotonica protein kinase (DMPK), on chromosome 19. Unaffected individuals carry 5–37 repeats, while affected individuals carry 50–4000 repeats. DM exhibits pronounced anticipation and dramatic somatic instability. Interestingly, the clinical phenomenon of anticipation (worsening severity as a disease gene is transmitted from one generation to the next) was first described nearly a century ago in a family segregating DM1. Although the anticipation phenomenon was dismissed as an artifact of ascertainment by geneticists of the mid twentieth century, anticipation is now known to be a genuine feature of dynamic mutation diseases. Anticipation results from the tendency of disease repeats to expand and the inverse correlation between repeat length and age of disease onset.

Several theories have been advanced to explain the molecular basis of DM1, including haploinsufficiency of DMPK, local chromatin effects on neighboring genes, and gain of function exerted by expanded RNAs. DMPK is proposed to have many functions, some of which could relate to the disease, such as modulation of skeletal muscle sodium channels, RNA metabolism, calcium homeostasis, and the cell stress response. Initial expression studies reported a decrease in DMPK RNA and protein levels, but a DMPK knockout mouse designed to test the haploinsufficiency theory exhibited only mild myopathy that was inconsistent with the DM phenotype. The (CTG)n tract is a strong nucleosome assembly site, and it was therefore hypothesized to have trans-effects on the expression of neighboring genes. According to this model, the myriad symptoms of DM1 are caused by disturbances in the expression of multiple nearby genes due to the expansion – making DM1 a “contiguous gene syndrome.” There are several genes in close proximity to the DMPK gene (i.e. <5 kb) whose roles in DM1 were considered; the most studied of these genes has been SIX5. The SIX5 homolog in Drosophila is required for normal eye development and its mouse homolog is involved in regulating distal limb muscle development. Expression data on SIX5 in
DM1 patients has been inconsistent, however. A Six5 knockout mouse developed cataracts, although they were not of the type observed in DM1. Another gene implicated by this model is DMWD, which is expressed in the testis and suspected to be involved in male infertility. Studies indicated that the expression of DMWD is not altered.

The third theory for the molecular basis of DM proposes that an RNA gain-of-function mechanism is responsible. RNA foci, which are accumulations of expanded transcripts, accumulate in the nuclei of patient cells (Fig. 11). The RNA gain-of-function theory was buttressed by the discovery that DM2, a disorder with symptoms almost identical to DM1, is caused by a (CCTG)n expansion in intron 1 of the zinc finger 9 protein (ZNF9). The two genes responsible for DM1 and DM2 are unrelated and reside on different chromosomes. Genes near the two loci bear no obvious resemblances. The only striking parallels between the two expansions are: (1) they both contain CTG triplets; and (2) they both occur in transcribed but noncoding regions of the genome. More support for the RNA gain-of-function model came from a mouse model generated by Mankodi et al. in 2000, which contained a (CTG)n expansion in the 3′ untranslated region of the skeletal actin (HSA) gene. This mouse exhibited myopathy typical of DM1, although the skeletal muscle-restricted expression pattern of HSA precludes broader conclusions. Thus, there are several lines of evidence suggesting that CUG expansion-containing RNAs are capable of causing DM.

RNA-binding proteins and transcription factors colocalize with the RNA foci found in DM, potentially altering nuclear processes. Elevated levels of CUG-containing RNA has been shown to alter gene splicing in specific transcripts which could be relevant to DM: cardiac troponin T (cTNT), involved in cardiomyopathy; Insulin Receptor (IR), involved in diabetes; and Clc-1, the main chloride channel in muscle. Abnormal splicing of cTNT, IR, and Clc-1 are proposed to account for the cardiac abnormalities, insulin insensitivity and myotonia observed in DM.

![Fig. 11](image_url)

*Fig. 11*  *In situ* hybridization of muscle sections with fluorescently labeled antisense oligonucleotide probes reveals accumulation of mutant RNA in DM1 and DM2. (a) Probing of DM2 muscle with a CAGG probe indicates that multiple RNA foci are present. (b) Probing of normal muscle with a CAGG probe demonstrates absence of RNA foci. (c) Probing of DM1 muscle with a CAG probe yields prominent RNA foci. (From Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W., Ranum, L.P. (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9, *Science* 293, 864–867, used with permission of *Science*).
Splicing abnormalities cause a reduction in the membrane concentration of ClC-1 and reduce chloride conductance to levels consistent with myotonia. These splicing alterations are thought to be due to the repeat-expanded RNA’s effects on two families of RNA-binding proteins: “CUG-BP1 and ETR-3-like factors” (CELF) and “muscleblind-like proteins” (MBNL). CELF proteins regulate pre-mRNA splicing in cTNT, IR, and ClC-1. In patient tissue and cell culture, CUG-BP1 levels and activity are increased in response to elevated levels of CUG-containing RNA. This may be due to a lengthening of the protein’s half-life. MBNL proteins were named after their Drosophila ortholog muscleblind, which is required for photoreceptor and muscle differentiation in flies. The three known proteins in this family (MBNL1 (MBNL), MBNL2 (MBLL), and MBNL3 (MBXL)) are splicing regulators that are thought to act antagonistically to CELF family proteins. They colocalize with RNA foci in vivo, and a mouse model lacking specific isoforms of MBNL1 recapitulates the myotonia, cataracts, and splicing dysregulation observed in DM.

The splicing alterations seen in DM patients are consistent with loss of function of MBNL proteins or an increase in CELF protein activity in muscle and brain. Because of the colocalization of MBNL proteins with RNA foci, it has been proposed that sequestration and subsequent depletion of these proteins from the cellular milieu is responsible for the symptoms of DM. Although this process probably plays a critical role in the disease, recent evidence suggests it is not solely responsible. The most plausible current theory explaining DM pathogenesis proposes an imbalance between the antagonistic MBNL and CELF proteins, resulting in specific splicing abnormalities that cause DM’s diverse range of symptoms.

5.2 Myotonic Dystrophy Type 2

Myotonic dystrophy type 2 (DM2) is an autosomal dominant, multisystem disorder very similar to DM1. The majority of its symptoms resemble DM1: progressive weakness, myotonia, cardiac disturbances, iridescent cataracts, and insulin insensitivity. There are some notable differences, however. DM2, unlike DM1, predominantly affects proximal muscles at its onset, which is why many cases were originally classified as proximal myotonic myopathy (PROMM). Other interesting differences are that mental retardation is not observed in DM2, DM2 patients show increased sweating, and DM2 congenital forms have not been observed. DM2 is caused by a (CCTG)n tetranucleotide expansion in the first intron of the zinc finger protein 9 (ZNF9) gene on chromosome 3q21.3. The tetranucleotide repeat can expand to stunning lengths, with the longest cases comprising 44 kb of DNA, making them the longest tracts observed in the repeat expansion disorders. The DM2 locus exhibits marked somatic instability. Over the course of a patient’s lifetime, the average repeat length increases substantially, as judged by blood drawings.

The similarities between DM1 and DM2 are not restricted to the clinical presentation. The expansions are both large, CTG-containing tracts that are transcribed, but not translated. Both disorders cause nuclear RNA foci that sequester specific RNA-binding proteins (Fig. 11), including CELF and MBNL family members. Similar splicing abnormalities are also observed in DM2. Since the genes associated with the repeat expansions in DM1 and DM2 are
unrelated, it is likely that the cause of both disorders is a toxic gain of function of long tracts of CUG-containing RNA. A mouse model generated by Mankodi et al. in 2000, which contained a (CTG)n expansion in the 3’ untranslated region of the skeletal actin (HSA) gene, exhibited myopathy typical of DM. Thus, even outside the context of the DM1 and DM2 loci, CTG-containing expansions can cause DM-like symptoms. This supports the RNA gain-of-function theory, as transcribed CTG-containing expansions are inherently capable of causing disease. Differences between DM1 and DM2 may be due to regional or temporal expression patterns or differences in the affinity of RNA-binding proteins for CTG or CCTG tracts.

5.3 Spinocerebellar Ataxia Type 8

Spinocerebellar ataxia type 8 (SCA8) is a dominantly inherited cerebellar ataxia. Affected individuals suffer from late-onset, slowly progressing gait ataxia as well as dysarthria, oculomotor incoordination, spasticity, and decreased vibration sense. The cerebellar cortices and vermis undergo a slowly progressing but dramatic atrophy, while the brainstem exhibits little evidence of degeneration. Patients may become wheelchair-bound as early as their fourth decade of life. Using a direct method for cloning expanded triplet repeats (i.e. RAPID cloning), a gene containing an expanded CTG tract was identified in a large family (MN-A) segregating the SCA8 phenotype. Interestingly, this CTG is contained within a gene on the long arm of chromosome 13 that is transcribed, but apparently not translated into a protein product. Thus, it was proposed that the production of an RNA transcript containing this expanded CTG repeat tract is the cause of SCA8. Numerous subsequent studies, however, have indicated that possession of the expanded CTG repeat tract appears necessary, but not sufficient, for the production of the SCA8 phenotype. Thus, reduced penetrance is viewed as a key feature of SCA8 at this time. On the basis of the available genetic data, normal individuals always carry fewer than 70 CTG repeats, while affected SCA8 patients can carry anywhere from 71 to >1000 CTG repeats. As extremely large CTG repeat alleles (>800 triplets) were shown to not cause disease in early reports, this was initially attributed to lack of stability of the mutant RNA product. However, further work has revealed considerable overlap between disease-causing CTG repeat expansions and nonpenetrant CTG repeat alleles, indicating that a secondary factor – either a trans-acting genetic factor or an environmental factor – must be present to yield the SCA8 disease phenotype. In the case of the original MN-A family, evidence for a cis modifier appears responsible for the extremely high penetrance in this large pedigree. Thus, while the causality of the CTG repeat expansion in SCA8 has been somewhat controversial, review of the current literature suggests that the CTG repeat expansion is directly involved in SCA8, but may not alone be sufficient to produce the disorder.

Another unique feature of the SCA8 CTG repeat is its extreme and unusual genetic instability. Paternal transmissions generally result in contractions, while maternal transmissions generally result in expansions. Expansions of up to 600 repeats have occurred in one generation through maternal transmission. Large deletions in expanded alleles often occur in sperm cells, offering an explanation for the paternal contraction bias. Pathogenic expansions
often have 5’ triplet interruptions, the role of which is unclear.

The molecular basis of how the expanded SCA8 RNA causes disease remains unclear. The transcript has been detected exclusively in the brain, and appears to be transcribed in the CTG orientation. Translation of a polyglutamine tract from the CAG-containing transcript in the opposite direction is precluded by the existence of stop codons flanking the repeat. The longest transcript identified to date contains 6 exons, and many alternatively spliced forms of the gene have been detected, although all these isoforms are expressed at very low levels. None contain significant open reading frames. In 1999, Koob et al. reported that a gene partially overlaps the 5’ end of the SCA8 gene locus on the antisense strand. This gene, called Kelch-like 1 (KLHL1), for its homology to the Drosophila KELCH gene, is highly conserved and predicted to encode an actin-binding protein of 748 amino acids. Its expression overlaps with that of the SCA8 transcript, suggesting that the SCA8 gene may produce an antisense RNA whose function is to regulate KLHL1 expression. Whether or how this is occurring remains uncertain at this time.

With the discovery of the DM RNA toxic gain-of-function pathway, an emerging theory for SCA8 pathology holds that the production of a CUG-expanded transcript results in RNA gain-of-function toxicity within the restricted neuronal populations where the SCA8 gene is expressed. In support of this hypothesis, transgenic mice derived with a human bacterial artificial chromosome containing the entire SCA8 gene with a 118 CTG repeat expansion develop neurological disease. The severity of their phenotype depends on the expression level of the transgene. In the more moderate expressing lines, the SCA8 mice display a slowly progressive gait ataxia reminiscent of the human disease. While no protein product has been detected, 1C2 and ubiquitin antibody staining reveal intranuclear inclusions in cerebellar neurons, consistent with the RNA-containing inclusions observed in FXTAS brains. An alternative interpretation is that translation of CTG into a polyleucine tract is occurring. Further work will be needed to distinguish these two possibilities.

Studies of the SCA8 CTG repeat expansion in D. melanogaster have also been informative. Whether directing expression of the SCA8 gene to fly retina with a normal CTG repeat tract or an expanded CTG repeat tract, a neurodegenerative eye phenotype results. Modifier screens using fly stocks carrying mutations in genes encoding RNA-binding protein yielded a number of genes that could either enhance or suppress this retinal degeneration phenotype. Interestingly, Drosophila muscleblind, whose mammalian counterpart has been implicated in the DM RNA toxicity pathway, modified the retinal degeneration caused by expanded SCA8 CTG repeat expression more so than the retinal degeneration caused by normal SCA8 CTG repeat expression. Such data supports the hypothesis that the SCA8 CTG repeat expansion is producing neurotoxicity by altering the function of RNA-binding proteins as in DM. Although it would be premature to definitively conclude that SCA8 is an RNA gain-of-function repeat disease, provisional classification in this category is appropriate.

5.4 The Fragile X Tremor – Ataxia Syndrome (FXTAS)

Fragile X tremor-ataxia syndrome (FXTAS) is a rare and unusual disorder
that is associated with premutation of the FRAXA locus in males. Patients develop a progressive ataxia and intention tremor. This is sometimes accompanied by dementia, parkinsonism, and autonomic dysfunction. Neuropathological changes include degeneration of the cerebellum and ubiquitin-positive intranuclear inclusions in neurons and glia. FXTAS is associated with the expansion of a (CGG)\textsubscript{n} tract in the 5′-untranslated region of the \textit{FMR1} gene, found on chromosome Xq27.3. Expansion of this tract beyond 200 repeats causes FRAXA, a disorder resulting from a reduction in the expression of FMR1. Males with tracts between 55 and 200 CGG repeats are considered to be in the “premutation range,” and are now considered to be at risk for FXTAS beyond middle age.

In contrast to FRAXA, FXTAS patients show increased expression of the \textit{FMR1} transcript, with longer repeats corresponding to higher transcript levels. Premutation carriers with more than 100 repeats have an average of five times more \textit{FMR1} mRNA than individuals in the normal repeat range. Despite the increase in transcript level, the protein product of the \textit{FMR1} gene, FMRP, is slightly reduced in FXTAS patients. Since elevated transcript

\textbf{Fig. 12} Intranuclear inclusions are present in the brains of FXTAS patients. (a) Hematoxylin & eosin staining of cerebral neurons reveals a refractile, eosinophilic nuclear inclusion of \(~5\ \mu\text{M}\) in diameter (white arrowhead). (b) Hematoxylin & eosin staining of cerebral astrocytes reveals refractile, eosinophilic nuclear inclusions of \(~2\ \mu\text{M}\) in diameter (white arrowhead). (c) Antineurofilament antibody staining of cerebellum demonstrates presence of dystrophic neurites, consistent with ongoing Purkinje cell degeneration. (d) Antiubiquitin antibody staining of cerebral neurons labels intranuclear inclusions. (e) Antiubiquitin positive intranuclear inclusions form in both neurons (white arrowhead; larger cell) and astrocytes (white arrowhead; smaller cell). (From Greco et al. (2002) \textit{Brain} \textbf{125}, 1760, used with permission).
levels are present without great alteration in protein levels, this disorder may be due to an RNA gain of function of the FMR1 transcript. Indeed, immunostaining of FXTAS patients’ brain sections reveal intranuclear inclusions in neurons and glia, believed to be comprised of accumulated Fmr1 RNA transcripts and various proteins (Fig. 12). To determine the molecular basis of FXTAS, a knockin mouse model of FXTAS was generated using a human 98 CGG repeat, and was noted to produce elevated levels of the Fmr1 transcript and display intranuclear ribonucleoprotein inclusions, containing ubiquitin, Hsp40 and the 20S proteasome subunit. The inclusions may be a response to RNA toxicity or a result of the aggregation of CGG-binding proteins. While the mechanism of FXTAS may be due to a dominant gain of function of triplet-expanded RNA on RNA-binding proteins as in DM, there is currently no direct evidence supporting this theory.

6.2 Oculopharyngeal Muscular dystrophy

Oculopharyngeal muscular dystrophy (OPMD) is a predominantly autosomal dominant, late-onset disorder characterized by progressive drooping eyelids, dysphagia, and proximal limb weakness. Certain skeletal muscles in affected patients degenerate and contain nuclear inclusions and rimmed vacuoles. This effect is particularly striking in the levator palpebra and pharyngeal muscles, responsible for lifting the eyelids and swallowing. OPMD is caused by an alanine expansion in the gene polyadenine-binding protein 2 (PABP2). PABP2 normally contains a 10-alanine tract encoded by (GCG)6(GCA)3GCG. Affected individuals carry 12–17 alanines, probably resulting from unequal crossing-over of the two alleles. Repeat length appears to correlate with disease severity. PABP2 is an abundant, ubiquitously expressed pre-mRNA-binding protein that plays a role in controlling the formation and length of mRNA polyA tails. NIs immunoreactive for PABP2 are present in the skeletal muscle cells of OPMD patients and sequester polyA-containing transcripts. The aggregates are filamentous and contain PABP2, ubiquitin and
proteasome subunits, leading investigators to conclude that expansion of the alanine tract probably causes PABP2 to misfold and aggregate. Aggregation in OPMD has been linked to toxicity in several experimental systems. Polyalanine-expanded PABP2 aggregates and causes cell death in cultured cells and transgenic mice. Indeed, widespread expression of the human PABP2 gene in transgenic mice yielded polyalanine length-dependent muscle pathology, including rimmed vacuoles, central nuclei, and numerous dystrophic changes (Fig. 13). Deleting the C-terminal oligomerization domain, overexpressing chaperones, or exposing cells to aggregation inhibitors such as Congo red and doxycycline, reduces aggregate formation and toxicity in cell culture. A polyalanine-expanded peptide, similar to the amino-terminal region of PABP2 was shown to adopt a β-sheet conformation, whereas the same peptide with 7 alanines adopted an α-helical conformation. This is reminiscent of polyglutamine proteins, which also adopt a β-sheet conformation when the polyglutamine tract exceeds a threshold of about 35 glutamines. Expanded polyalanine proteins also activate caspase-3 and -8 in cultured cells, another feature reminiscent of polyglutamine toxicity.

In 2004, Wirtschafter et al. proposed a model for the selective vulnerability of

![Fig. 13](image)
extraocular muscles in OPMD. Unlike other skeletal muscles, extraocular muscles are not postmitotic and they continually undergo remodeling. This requires the frequent upregulation of genes involved in cell cycling and protein synthesis. Failure of correct mRNA polyadenylation or transport in these cells may have a cumulative toxic effect resulting in the progressive degeneration of these muscles. The mechanism of OPMD could be due to an interference with polyadenylation, disturbances in intracellular trafficking of mRNA, or toxicity due to the aggregation of misfolded and/or aggregated PABP2 species. Further research will be required to distinguish between these possibilities.

6.3 Synpolydactyly (Syndactyly Type II)

Synpolydactyly (SPD) is a rare, autosomal dominant developmental disorder characterized by fused and extra digits (syndactyly and polydactyly, respectively). It is caused by the expansion of a polyalanine tract in the amino-terminal region of the transcription factor HOX-D13. HOX genes act in concert to coordinate axial patterning in animals. Tracts of 7–14 alanines in HOX-D13 have been linked to SPD, and disease severity is proportional to repeat length. Multiple studies support a “dominant-negative” role for expanded HOX-D13 protein in SPD. Mice null for Hox-d13 have a phenotype less severe than SPD, while mice with alanine tract expansions have a form that more closely resembles the disorder. Mice lacking Hox11, Hox12, and Hox13 have a phenotype similar to SPD, suggesting that the alanine expansion in SPD antagonizes the function of other HOX genes. This is supported by a genetic complementation study. In further support of this theory, humans with suspected loss-of-function mutations in the HOX-D13 gene do not have a phenotype consistent with SPD.

6.4 Cleidocranial Dysplasia

Cleidocranial dysplasia (CCD) is a rare, autosomal dominant developmental disorder characterized by holes in the skull, dental malformations, absent or hypoplastic clavicles and maxillae, and other skeletal malformations. The primary cause of CCD is thought to be loss-of-function mutations in the gene RUNX2. This gene affects osteoblast differentiation and is a member of the Runt family of transcription factors. In one family, phenotypically distinct from classic CCD, an expansion from 17 to 27 alanines in RUNX2 has been detected. This family exhibited brachydactyly and a mild CCD phenotype. The difference in phenotype between the polyalanine expansion mutation family and typical presumed haploinsufficient, loss-of-function CCD patients supports a gain-of-function effect of the expanded alanine tract in this atypical family. Contraction of the tract in RUNX2 is common and does not cause a detectable phenotype.

6.5 Holoprosencephaly

Holoprosencephaly (HPE) is a common developmental malformation resulting in partial or full cyclopia, failure to develop midline structures in the ventral forebrain, and prenatal lethality. In rare cases, HPE is caused by the heterozygous expansion of a 15-amino acid alanine tract in the protein ZIC2. This protein is one member of a family of zinc finger proteins believed to regulate neurulation,
left–right axis formation and other developmental processes. Other individuals with heterozygous loss-of-function mutations have a phenotype indistinguishable from those with alanine tract expansions. Partial loss of function of Zic2 in mice similarly causes developmental abnormalities similar to HPE. Therefore, expansion of the alanine tract likely causes loss of function of ZIC2 in the case of HPE.

6.6 **Hand-foot-genital Syndrome**

Hand-foot-genital syndrome (HFGS) is a rare, dominantly inherited developmental abnormality characterized by malformation of the distal limbs and lower urogenital tract. Short thumbs, short great toes, and abnormal carpals and tarsals are some of its salient features. Some HFGS patients have alanine tract expansions in the protein HOX-A13, which is in the same family as the SPD-associated protein HOX-D13. HOX-A13 contains three alanine tracts, the most C-terminal of which is the most commonly mutated. Expansions enlarge the second or third tract by 6 to 9 alanines. In humans, deletion of HOX-A13 causes a phenotype that is mild in comparison to HFGS caused by alanine tract mutations. Hoxa13-null mice also have a milder phenotype than mice carrying a frameshift deletion suspected to confer gain of function. This evidence suggests a dominant-negative mechanism for HFGS caused by alanine tract expansions.

6.7 **Blepharophimosis-ptosis-epicanthus Inversus Syndrome**

Blepharophimosis-ptosis-epicanthus inversus syndrome (BPES) is a rare, autosomal dominant developmental disorder resulting in malformation of the upper eyelids and forehead, and occasionally premature ovarian failure in women. It is caused by a number of different mutations in the gene *Forkhead L2* (Foxl2), the most common of which is expansion of its carboxy-terminal 14-amino acid alanine tract. FOXL2 is a highly conserved transcription factor whose role in the ovary has been studied most thoroughly. It is expressed in the ovaries and eyelids during development and adulthood. Studies in which Foxl2 have been ablated in mice show that it is important for the differentiation of ovarian granulosa cells, and its absence causes accelerated follicle cell depletion leading to POF. Knockout mice also display craniofacial abnormalities consistent with BPES, suggesting that a dominant-negative mechanism may be responsible in humans.

6.8 **Syndromic and Nonsyndromic X-linked Mental Retardation**

Several loosely related disorders are associated with alanine tract expansions in the Aristaless related homeobox (ARX) protein on chromosome Xp22.13. One is nonsyndromic X-linked mental retardation (XLMR), a heterogeneous condition in which mental retardation is the main consistent feature. Several syndromic XLMR disorders linked to alanine expansion in ARX include West syndrome (WS) and Partington syndrome (PRTS). WS causes progressive mental retardation with abnormal EEG and infantile seizures, while PRTS causes mental retardation, dysarthria, and dystonic movements of the hands. Brain anatomy appears normal in these disorders. ARX is a paired-class homeodomain protein that is expressed in the ventricular
and marginal zones of the developing mouse brain and continues to be expressed in adult cortex. It is suspected to play a role in neuroepithelial cell differentiation and maintenance of neuronal subtypes in the adult cortex. ARX-null mice have small brains and neuronal migration deficits. Expansions of two different alanine tracts in ARX cause XLMR: a 12-alanine tract at amino acids 144–155, and a second tract at amino acids 100–115. Both sites cause highly variable forms of XLMR, with expansions in the former alanine tract resulting in nonsyndromic XLMR, WS or PRTS. Alanine tract expansions in ARX probably result in partial loss of function of the protein. Heterozygous female carriers of expanded ARX do not exhibit XLMR, suggesting that one normal copy of the protein is sufficient for normal cognitive ability. Also, humans with null mutations of ARX suffer from X-linked lissencephaly with abnormal genitalia (XLAG), a much more severe condition causing major developmental abnormalities in the brain and genitalia. Thus, XLMR due to alanine expansions in ARX may be because of partial loss of function leading to subtle developmental defects and/or failure to maintain specific neuronal populations.

6.9 Congenital Central Hypoventilation Syndrome

Congenital central hypoventilation syndrome (CCHS) is a rare, autosomal dominant disorder causing a failure of autonomic control of breathing. It attenuates or abolishes responses to hypercarbia and hypoxemia. In the majority of cases, it is caused by the expansion of one of two alanine tracts in the protein PHOX-2B. Mutations expand the 20-residue tract to 25–29 alanines. PHOX-2B is a paired-class transcription factor containing a homeodomain. A loss-of-function mutation in murine Phox2b is homozygous lethal, and specifically prevents the development of parasympathetic ganglia. Heterozygous mice show chronic pupil dilation but no parasympathetic or respiratory disturbances. Also, a patient hemizygous for a 5-Mb deletion including Phox2b does not have CCHS. No cases of CCHS have been reported in which PHOX2B is truncated before the homeobox domain. This evidence suggests that the expansion of alanine tracts in PHOX2B may cause a subtle, dominant-negative effect on the development of respiratory control pathways.

7 Unclassified Repeat Diseases Lacking Mechanistic Explanations

7.1 Spinocerebellar Ataxia Type 10

Spinocerebellar ataxia type 10 (SCA10) is an autosomal dominant, progressive ataxia that exhibits nearly pure cerebellar signs. It appears restricted to individuals of Mexican ethnicity. SCA10 patients typically present with gait ataxia, followed by dysarthria, dysphagia, and ocular dysmetria, and most patients also experience recurrent motor seizures. Cerebellar atrophy is the most prominent neuropathological change. SCA10 is caused by the expansion of a highly polymorphic pentanucleotide repeat, (ATTCT)n, on chromosome 22. Unaffected individuals carry 10–22 repeats. While expanded alleles have not been successfully PCR amplified, transcript sizes on Northern blots and fragment sizes on Southern blots indicate that disease alleles can range from ~800 to 4500 repeats. Anticipation occurs in SCA10, independently
supporting the causal nature of the repeat expansion. The ATTCT repeat is located in intron 9 of the SCA10 gene, a previously unrecognized 66-kb gene that encodes a novel putative 475 amino acid protein of unknown function with few recognizable motifs or domains. The 2-kb SCA10 transcript is ubiquitously expressed though the highest levels of expression occur in the brain, testis, and adrenal glands. Within the brain, it is most highly expressed in the cerebellum and associated structures. The carboxy-terminal portion of the protein appears to contain “armadillo repeats,” which are responsible for the membrane association of β-catenins. Ataxin-10 does not seem to associate with membranes, however. The entire protein is highly conserved between humans and rodents, and potential orthologs exist in Arabidopsis and Drosophila.

There are several proposed mechanisms for SCA10 pathogenesis. The most obvious hypothesis is that the large intronic expansion affects ataxin-10 expression, perhaps by altering local chromatin structure. siRNA knockdown of ataxin-10 in cell culture experiments yields higher rates of cell death in cerebellar neurons than in cortical neurons. However, the expression levels of the SCA10 transcript are not reduced in patient’s lymphoblast cells, arguing against simple haploinsufficiency. Consequently, RNA gain-of-function toxicity has been proposed as the potential mechanism; however, there are no further data at this time to support such a hypothesis. Additional work will need to be done to distinguish between these and other possibilities, and will need to account for the cell-type specific pattern of neurodegeneration that occurs in the face of apparently widespread expression of the SCA10 gene mutation.

7.2 Spinocerebellar Ataxia Type 12

Among the more recent additions to the unstable repeat disease group is spinocerebellar ataxia type 12 (SCA12), a rare, autosomal dominant disorder that may be most prevalent in Indian populations. Its symptoms are distinct from the other SCAs, typically beginning with an action tremor of the upper extremities and progressing to include hyperreflexia, mild cerebellar dysfunction, bradykinesia, increased muscle tone, psychiatric symptoms, and dementia. The brains of SCA12 patients likely undergo a slow, generalized atrophy that is most prominent in the cortex, but also results in loss of Purkinje cells in the cerebellum. Disease onset typically occurs in the third or fourth decade, and a gradually progressive disease course is typical.

SCA12 is caused by a \((\text{CAG})^n\) expansion in chromosome 5q31 – q33. The expansion is 5′ to the \(\text{PPP2R2B}\) gene, encoding a regulatory subunit of the protein phosphatase 2A enzyme (PP2A). This gene has many transcription start sites, some of which include the repeat (but many of which do not). GENSCAN predicts an exon including the expansion that would encode a polyserine tract, but this prediction is of low probability and considered unlikely. Other evidence suggests that the expansion is located in the promoter region, and this is currently the most widely accepted view. Unaffected individuals carry 7–32 CAG repeats while affected individuals carry 55–78 CAG triplets. The most common repeat size in unaffected individuals is 10 CAGs. The repeat is fairly stable, with only modest expansions and contractions resulting equally from maternal and paternal transmission. A significant correlation between repeat length and age of onset has not been documented.
The protein PP2A is an essential serine/threonine phosphatase expressed in all known eukaryotic cells. It is involved in diverse cellular functions, including cell growth, differentiation, DNA replication, neurotransmitter release, and apoptosis. PPP2R2B is a brain-specific regulatory subunit of PP2A. The class of regulatory subunits including PPP2R2B may affect PP2A's phosphatase activity for certain substrates, including histone-1, vimentin, and tau. It may also affect PP2A's subcellular localization.

There are several possible explanations for SCA12 pathogenesis. The first is that expanded PPP2R2B may generate a polyamino acid tract-containing protein, resulting in toxicity. Northern blots probing for sequence flanking the CAG repeat did not detect a PPP2R2B transcript, indicating that if a repeat-containing exon exists and is transcribed, it is not present at appreciable levels. Nevertheless, the possibility of polyglutamine, polyserine, or polyalanine toxicity, though unlikely, cannot be completely ruled out at this time. A second possible cause of SCA12 pathogenesis is RNA gain of function, as occurs in DM and FXTAS. The repeat expansions are smaller in SCA12 than in DM, however, and more importantly, CAG tract-containing transcripts appear to be rather scarce. A third possibility is altered splicing of the PPP2R2B transcript. Several different amino-termini are possible, the ratio of which may affect PP2A's subcellular localization. Yet another theory of SCA12 pathogenesis is that the expansion affects PPP2R2B transcript levels. Repeat expansion causes a substantial increase in PPP2R2B expression as measured in reporter assays using a neuroblastoma cell line. Altered levels of the protein could affect PP2A's specificity or subcellular localization. This has the potential to disturb a multitude of processes in the CNS. At this time, all of the above theories of SCA12 CAG repeat expansion neurotoxicity remain plausible.

7.3 Huntington’s Disease Like 2 (HDL2)

Perhaps the most exciting and enigmatic recent discovery in the repeat expansion field is that of the mutational basis of a disorder known as Huntington’s disease like 2 or HDL2. HDL2 is so named because it is in essence a genocopy of classical HD, as the original HDL2 pedigree was labeled with a diagnosis of HD until HD CAG repeat testing indicated that this family’s HD-like disease did not result from a CAG repeat expansion in the *htt* gene. HDL2 patients present with weight loss and diminished coordination, and then develop tremors, dysarthria, hyperreflexia, and rigidity. Patients display psychiatric involvement, chorea, and dystonia, and ultimately become demented. Death occurs 15–25 years after onset, when patients become bedridden as in typical HD cases. MRI findings reveal marked atrophy of the caudate and of the cerebral cortex, making the neuropathology indistinguishable from classic HD. HDL2 patients do not show cerebellar signs or neuropathology.

As soon as it was found that HDL2 patients do not have the HD CAG repeat expansion, direct cloning methods for triplet repeat expansions of the CAG/CTG type were applied to patient samples and an expanded CAG/CTG repeat (*n* = 55) was isolated. Sequence flanking this repeat indicated that HDL2 is caused by CTG repeat expansions in the *junctophilin-3 (JPH3)* gene, one of a family of structural proteins whose function is to link
Ca++ channels on the ER with voltage sensors on the plasma membrane. Analysis of JPH3 CTG repeat indicates that normal individuals typically carry 7–27 CTG repeats while affected HDL2 patients usually have expansions of 50–60 CTG repeats. Widespread screening of HD-like patients from around the globe suggests that the HDL2 JPH3 CTG repeat expansions are most common in individuals of African ethnicity.

The question of how the JPH3 CTG repeat expansion causes HDL2 remains unknown, but there are at least four possible explanations. These alternative (but not mutually exclusive) theories stem from the documented alternative processing of the JPH3 gene, which permits the prediction of the CTG repeat tract as: (1) part of intron 1; (2) as part of the 3′ untranslated region; or (3) as encoding either a polyleucine or polyalanine tract. One theory is that haploinsufficiency of the Ca++ regulating brain- and testes-specific junctophilin-3 protein is responsible for HDL2. While JPH3 knockout mice display motor incoordination, no histological abnormalities are found in their brains. Further work with these mice is ongoing to evaluate this hypothesis. Evidence against simple haploinsufficiency has come from study of HDL2 patient’s brain material, however. 1C2 and ubiquitin antibody immunostaining reveal intranuclear inclusions in neurons throughout the brains of these patients, with dramatic similarity in distribution to patients with classic HD (Fig. 14). At this time, the molecular basis of HDL2 is unknown; however, the incredible overlap between HD and HDL2 in terms of clinical phenotype and neuropathology strongly suggests that solving HDL2 should have profound implications for our mechanistic understanding of HD.

Fig. 14 HDL2 patients have 1C2-positive intranuclear inclusions. 1C2 staining of cerebral cortex (frontal lobe) from an HDL2 patient reveals a prominent intranuclear inclusion that resembles the intranuclear inclusions seen in classic HD patients. As the 1C2 antibody is directed against expanded, misfolded polyglutamine tracts, which are not predicted to be expressed from the causal HDL2 gene, the explanation for the presence of such nuclear inclusions in HDL2 patients remains unknown. (From Margolis, R.L., O’Hearn, E., Rosenblatt, A., Willour, V., Holmes, S.E., Franz, M.L., Callahan, C., Hwang, H.S., Troncoso, J.C., Ross, C.A. (2001) A disorder similar to Huntington’s disease is associated with a novel CAG repeat expansion, Ann. Neurol. 50, 373–380, used with permission of Annals of Neurology, and John Wiley & Sons, publisher).
See also Genetics, Molecular Basis of; Motor Neuron Diseases: Cellular and Animal Models; Motor Neuron Diseases: Molecular Mechanism, Pathophysiology, and Treatments; Noncoding Tandemly Repeated DNA Sequences.

Bibliography

Books and Reviews


Primary Literature


Bibliography


Chen, S., Peng, G.H., Wang, X., Smith, A.C., Grote, S.K., Sopher, B.L., La Spada, A.R. (2004) Interference of Crx-dependent transcription by ataxin-7 involves interaction between the glutamine regions and requires the ataxin-7 carboxy-terminal region for...


and the dodecamer: progressive myoclonus epilepsy, cystatin b and a 12-mer repeat expansion, Cytogenet. Genome Res. 100, 213–223. 
Triplet Repeat Diseases


Sutherland, G.R., Baker, E. (1992) Characterisation of a new rare fragile site easily confused for...


Triplet Repeat Diseases

Tumor Immunology: see Immunology of Cancer

Tumorigenesis, Epigenetic Mechanisms: see Epigenetic Mechanisms in Tumorigenesis