Investigation of the pathogenic agent in a *Drosophila* model of polyglutamine disease

A thesis submitted for the degree of Doctor of Philosophy, March 2006

Catherine McLeod, B. Sc. (Biomed.) (Hons.)

School of Molecular and Biomedical Sciences, Discipline of Genetics
Centre for the Molecular Genetics of Development,
Adelaide University

Chapter 1 – Background Literature

Repeat Expansion Diseases

Over a decade ago, it was discovered that the expansion of repeats in the human genome can lead to disease [1,2]. Since then, at least 17 diseases involving neurological or neuromuscular degeneration have been attributed to repeat expansion [3-5]. Despite the fact that the causative mutations and the genes in which they reside have been identified, the mechanisms by which the mutations lead to degeneration remain unclear.

Typically, the expanded repeat disease phenotypes manifest later in life; why this is so is unknown. The severity of each of the diseases is related to the length of the repeat expansion with longer repeats causing more severe phenotypes and an earlier age of onset. The diseases also display anticipation, where an increased severity and earlier age of onset are observed in successive generations. The cause of this phenomenon was initially unclear. However, once the nature of the mutations causing the diseases was identified, it became apparent that anticipation is due to the instability of the expanded repeats and their tendency to expand further upon transmission through the germline [6].

The expanded repeat diseases can be divided into two groups. Many are caused by the expansion of CAG repeats in the coding region of a gene, resulting in an expanded polyglutamine tract in the encoded protein, and are thus referred to as ‘polyglutamine’ diseases. The remaining diseases are caused by repeat expansions located in untranslated regions and, like the polyglutamine diseases, most are dominantly inherited, despite the sequence of the encoded proteins remaining unchanged [7]. There is also one disorder that does not fit either of these categories, oculopharyngeal muscular dystrophy, which is caused by expansion of a GCA repeat encoding polyalanine [8].

Polyglutamine diseases

To date, nine human diseases have been attributed to the expansion of polymorphic CAG tracts within coding regions, causing the resulting proteins to contain an expanded polyglutamine region. These include spinocerebellar ataxia (SCA) types 1, 2, 3, 6, 7 and 17, Huntington's disease, spinobulbar muscular atrophy (SBMA) and dentatorubral pallidoluysian atrophy (DRPLA) (Table 1.1).

Disease	Gene	Protein	CAG repeat size	
			Normal	Disease
Huntington's disease	*HD*	huntingtin	6-34	36-121
SBMA (Kennedy Disease)	*AR*	androgen receptor	9-36	38-62
DRPLA	*DRPLA*	atrophin-1	6-35	49-88
spinocerebellar ataxia type 1 (SCA1)	*SCA1*	ataxin-1	6-44*	39-82
SCA2	*SCA2*	ataxin-2	15-31[+]	36-63
SCA3 (Machado-Joseph disease)	*SCA3*	ataxin-3	12-40	55-84
SCA6	*SCA6/ CACNA1A*	α1A subunit of voltage-dependent calcium channel	4-18	21-33
SCA7	*SCA7*	ataxin-7	4-35	37-306
SCA17	*TBP*	TATA binding protein	27-42	47-55

**Normal SCA1 alleles with repeat copy numbers above 21 are interrupted by CAT codons, whereas disease alleles consist of pure CAG repeats. [+]Normal SCA2 alleles are frequently interrupted with CAA repeats, whereas diseases alleles consist of pure CAG repeats.*

Table 1.1 Polyglutamine diseases described to date, caused by expansion of a coding CAG repeat tract, enlarging it from the normal range into the disease-causing range. The gene containing the CAG repeat tract, affected protein, and normal and disease-causing ranges of repeat are shown for each disease. For references see [3], [5] and [9].

Huntington's Disease

The most frequently occurring polyglutamine disease is Huntington's Disease (HD), which is caused by expansion of a polymorphic CAG tract in the *HD* gene [10]. *HD* contains 67 exons, the first of which contains the CAG tract. The function of the encoded 350 kD protein, huntingtin, is unknown. Huntingtin is predominantly localised to the cytoplasm [11], where it shows partial colocalisation with vesicles and microtubules, leading to the suggestion that it plays a role in intracellular transport [12]. The characteristic clinical features seen in HD patients are progressive chorea, or uncontrollable involuntary movements, impairment of voluntary movement and dementia. This is associated with atrophy and degeneration of a subset of neurons in the brain, specifically the medium spiny neurons of the striatum [10]. The reason behind this selective neurodegeneration is unknown – it cannot simply be explained by analysing the expression pattern of huntingtin, as huntingtin from both wild-type

and expanded alleles is expressed ubiquitously in the brain, and also at lower levels throughout the body [13].

Spinocerebellar ataxia type 1 (SCA1)

Six of the polyglutamine diseases are spinocerebellar ataxias (SCAs), which are progressive diseases associated with degeneration of the cerebellum. They are characterised by late-onset ataxia, or lack of coordination of movement, as well as other symptoms, many of which are overlapping [9]. The best characterised of these is SCA1, caused by a CAG expansion in the *SCA1* gene. Normal alleles above 21 repeats contain 1-3 CAT interruptions, which interrupt the polyglutamine tract with histidines; expanded disease-causing alleles always consist of pure CAG tracts [14]. The normal function of the encoded protein, ataxin-1, is unknown, although there is evidence for a role in RNA processing [15]. Ataxin-1 shows predominantly nuclear localisation in neurons [16] and is expressed throughout the CNS and also at lower levels in peripheral tissues [17]. SCA1 is characterised by the death of Purkinje cells in the cerebellar cortex. As in HD, it is unclear why this specific subset of neurons is sensitive to the effects of the polyglutamine expansion in ataxin-1, as the protein is expressed in all neurons.

SCA2

SCA2 is characterised by degeneration of the cerebellum and brainstem, although again the protein containing the expanded polyglutamine tract, ataxin-2, is widely expressed throughout the brain [18]. However, in contrast to most of the other disease-related proteins, ataxin-2 shows no nuclear localisation and is found exclusively in the cytoplasm [18,19]. The function of ataxin-2 is unknown, although based on the role of its yeast ortholog, it may play a role in RNA metabolism [20]. Non-pathogenic alleles frequently contain CAA interruptions, while disease-causing alleles consist of pure CAG repeats [21]. However, CAA also encodes glutamine, so repeat tracts that are interrupted by this codon at the DNA/RNA level still encode a pure polyglutamine repeat.

SCA3

SCA3, also known as Machado Joseph Disease, is the most prevalent form of SCA. The phenotype is characterised by progressive ataxia in combination with various other non-cerebellar symptoms including peripheral neuropathy. Degeneration of neurons in the basal ganglia, brainstem and spinal cord is observed, with mild

neuronal loss in the cerebellum [22]. The pathogenic threshold for SCA3 is slightly higher than for the other polyglutamine diseases, although intermediate length repeat tracts (53-54 repeats) are associated with a milder neurological phenotype [23]. The CAG expansion occurs in the gene encoding ataxin-3, which is widely expressed in neurons and outside the CNS and shows predominantly cytoplasmic localisation [24]. Ataxin-3 plays a role in the ubiquitin-proteasome pathway [25], possibly acting as a deubiquitinating enzyme [26].

SCA6

SCA6 is rare among the polyglutamine diseases in that the role of the protein containing the repeat expansion was well characterised before the disease-causing mutation was mapped. The CAG expansion occurs in *CACNA1A*, the gene encoding the alpha (1A) subunit of the neuronal P/Q-type voltage-gated calcium channel [27]. This subunit is expressed throughout the brain but most strongly in the Purkinje cells of the cerebellum, and is localised to the cytoplasm [28]. In accordance with this expression pattern, neurodegeneration in SCA6 is primarily localised to the Purkinje cells [29]. SCA6 also differs from the other polyglutamine diseases in that the size of the polyglutamine tract in disease-causing alleles, 21-33 repeats, is much shorter than those found for the other diseases. Furthermore, there is evidence that SCA6 may be caused by alteration of the kinetic properties of the ion channel containing the polyglutamine repeats, leading to questions regarding its classification as a polyglutamine disease [30,31]. Supporting this is the observation that different mutations in *CACNA1A* that do not involve the polyglutamine tract give rise to disease phenotypes sharing similarities with SCA6 [32].

SCA7

SCA7 is characterised by neural loss in the cerebellum and regions of the brainstem and spinal cord. The retina is also a significant site of degeneration leading to blindness in SCA7 patients [33]. Ataxin-7 is expressed throughout the brain and in the retina and also in peripheral tissues [34]. The protein shows both nuclear and cytoplasmic localisation [34,35] and is thought to be involved in transcriptional regulation [36].

SCA17

Like SCA6, the gene mutated in SCA17 was well characterised before it was linked with the disease phenotype. In SCA17, the CAG expansion occurs in the gene

encoding TATA box binding protein (TBP), a general transcription initiation factor [5,37]. Larger polyglutamine tracts are tolerated in TBP: the pathogenic threshold for disease is 47 repeats, significantly higher than the other polyglutamine diseases. However, there is some evidence that repeats in the 43-48 range can induce SCA17, but with reduced penetrance [38]. Normal alleles contain CAA repeat interruptions, which are sometimes lost in disease alleles [39]. The major site of pathogenesis is the Purkinje cells [40], although TBP is ubiquitously expressed [39].

SBMA

SBMA, also known as Kennedy disease, is the only polyglutamine disease that is not dominant – instead it shows an X-linked mode of transmission. The CAG repeat expansion responsible for this disease occurs in the gene encoding the androgen receptor (AR), a ligand-activated transcription factor [2]. Testosterone activates the AR and causes its translocation to the nucleus, an event that occurs at an increased level in males and may explain why females are unaffected; alternatively this may be due to X-inactivation, resulting in females only expressing the expanded allele in approximately half of their cells. SBMA is primarily a motor-neuron disease; motor and sensory neurons of the spinal cord undergo degeneration, leading to muscle weakness and wasting [41]. Symptoms of androgen insensitivity are also present, suggesting that the polyglutamine expansion impairs the normal function of the AR [42].

DRPLA

DRPLA shares phenotypic overlap with HD, characterised by features including chorea, myoclonus, seizures and ataxia [43]. These symptoms are caused by widespread neurodegeneration in the cortex, globus pallidus, striatum and cerebellum [44,45]. The CAG expansion in DRPLA is located in the *atrophin-1* gene [46,47]. Atrophin-1 is widely expressed in the brain and peripheral tissues, and shows both nuclear and cytoplasmic localisation [48,49]. The function of atrophin-1 is unknown, although there is evidence that it associates with nuclear receptor co-repressor complexes and is involved in transcriptional regulation [48].

Similarities and differences

The proteins involved in the polyglutamine diseases generally show widespread expression throughout the brain, and many are also expressed in other tissues as well. Therefore it unclear why only a subset of neurons is sensitive to the

polyglutamine expansion and furthermore, why this subset varies between the diseases. It seems likely that gene context, the only significant difference between the disease-causing mutations, is responsible for the difference in the subset of neurons that are affected. How this occurs at the molecular level is unknown – possibilities include the normal function of the protein being particularly impaired in sensitive neurons, or perturbation of interactions with other proteins that are expressed only in the sensitive neurons.

However, there is significant overlap between the affected regions of the brain in each disease, and as each disease progresses, generalised atrophy of the cortex is frequently observed. Also, in severe juvenile-onset cases of disease caused by large polyglutamine expansions, there is loss of neuronal specificity and the disease phenotypes show significantly more overlap [9]. These observations suggest that the expanded polyglutamine-containing proteins are indeed toxic to a widespread range of neuronal cells and that their toxic effects are moderated by protein context when the repeat number is in the lower ranges. This notion is supported by evidence from mouse models, where expanded polyglutamine repeats are sufficient to cause neurological disease, but selective neuronal loss mimicking the human disease state occurs only when the full-length protein containing an expanded glutamine tract is expressed [50-52].

Despite the expanded CAG repeats occurring in otherwise unrelated genes, which show no homology outside of the polyglutamine repeat and appear to function in unrelated cellular processes, the resulting diseases are similar and share many overlapping symptoms. All of the diseases are progressive, typically beginning in midlife, and show dominant transmission (except for SBMA). Dysfunction, and eventually loss, of neurons in the cerebellum is observed in each disease, although the subset of affected neurons varies. The disease threshold length for repeat alleles is common between the diseases at 35-40 repeats, with the exceptions being SCA6, where it is slightly lower, and SCA17, where it is slightly higher. These similarities lead to the hypothesis that there is a pathogenic disease-causing mechanism common to all of the polyglutamine diseases, involving the expanded polyglutamine tract [53,54].

Toxicity of expanded polyglutamine

For HD, there is convincing genetic evidence that the disease phenotype is not simply due to partial loss of Huntingtin function, but instead is caused by a gain of function of the expanded allele. The disease is dominantly inherited, meaning that affected individuals have one normal *HD* allele. However, patients who have lost one copy of *HD* do not show the HD phenotype, suggesting that it is not caused by haploinsufficiency [55]. Similarly, deletion of one copy of the *HD* homologue in mice does not cause Huntington's disease [56]. There is also evidence that SBMA is caused by a gain of function, as the phenotype resulting from loss of AR function is testicular feminisation, rather than the motor neuronopathy of SBMA [57]. Mouse models suggest that this is also the case for SCA7: the SCA7 phenotype in mice with an expanded *ataxin-7* allele is no worse when the other allele is null than when it is wild-type [58]. Furthermore, overexpression of ataxin-7 does not modulate the SCA7 phenotype in mice [59]. Thus it appears that the polyglutamine diseases are not simply caused by loss of function resulting from expansion of one allele, but rather the expanded polyglutamine is exerting a dominant, toxic effect.

More direct evidence of expanded polyglutamine toxicity comes from studies of cultured cells. Proteins consisting of long polyglutamine tracts fused to GFP are toxic when expressed in COS cells, whereas shorter repeat lengths (less than 35) are not [60]. Expanded polyglutamine tracts are also toxic when expressed in *E. coli* [61]. Furthermore, when a $(CAG)_{146}$ repeat is engineered into a mouse gene that is unrelated to any of the genes involved in the human polyglutamine diseases, a late onset progressive neurological phenotype results [52]. This evidence supports the suggestion that expanded CAG repeats cause disease through a pathogenic mechanism that is shared by the polyglutamine diseases, and occurs regardless of gene context [62].

Mouse models of Polyglutamine Disease

In order to study the mechanisms of polyglutamine pathogenesis, many transgenic mouse models have been generated. One of the earliest successful models was a SCA1 mouse generated by driving expression of the full-length human *SCA1* gene in Purkinje cells, the major site of neuropathology in SCA1. The mice expressed high

levels of either wild-type *ataxin-1* containing a $(CAG)_{30}$ repeat, or an expanded *ataxin-1* allele, containing $(CAG)_{82}$ (referred to as B05 mice). While mice expressing ataxin-1 containing 30 glutamines (ataxin-1-Q30) remained identical to wild-type littermates, B05 mice developed cerebellar impairment at 5 weeks of age, which progressively worsened to severe ataxia. At 3.5 weeks of age, aggregates of ataxin-1-Q82 were seen in the Purkinje cells of the B05 mice, and continued to accumulate. By 6 weeks of age, Purkinje cell death was observed; however this is long after the first appearance of symptoms, suggesting that the neurological impairment in these mice is due to neuronal dysfunction, and not directly to neuronal loss [63].

The B05 mice appeared to faithfully reproduce some of the pathology of SCA1, namely neurological impairment, nuclear aggregation of the expanded protein, and neuronal cell death. However, some aspects of the human condition are not accurately represented. The Purkinje cell-specific promoter used was relatively strong, driving expression of the transgene in B05 mice at approximately 100 times the level of endogenous ataxin-1. In mice with lower expression of the transgene, comparable to endogenous levels, no phenotype was observed [63]. This requirement for overexpression to produce a phenotype has been observed in other polyglutamine mouse models, and suggests that either higher levels of protein are required in the mouse to cause disease, or that a longer time of exposure of neurons to the expanded polyglutamine is required, which falls outside the relatively short lifespan of the mouse. In a later study of the same mice it was observed that $(CAG)_{30}$ mice show mild Purkinje cell degeneration in old age, suggesting that even wild-type ataxin-1 is toxic if expressed at sufficiently high levels for a long enough period of time [64]. This toxicity induced by high levels of wild-type protein may imply that expansion of the polyglutamine tract enhances some property of the protein that occurs at a low level in the wild-type situation: perhaps increasing protein levels by decreasing clearance of the protein, or increasing the likelihood that the protein will misfold [64].

Another major discrepancy in this mouse model of SCA1 is that in B05 mice, ataxin-1 was only expressed in Purkinje cells, and not ubiquitously in the brain as it is in human SCA1. Thus these mice only show a phenotype related to Purkinje cell dysfunction, which does not represent the complexity of the human condition. To overcome this, knock-in mice were generated, containing an expanded $(CAG)_{78}$ repeat inserted into the endogenous *SCA1* locus, so that the expanded protein was

expressed in the correct spatial and temporal pattern [65]. Expanded ataxin-1 was expressed at endogenous levels in these mice and this was not sufficient to cause a SCA1 phenotype. Instead they showed only mild behavioural changes later in life, whereas in humans a repeat of this length is sufficient to cause juvenile onset SCA1 [65]. In a recent study, this approach has again been used to generate a line of knock-in mice, but this time an expanded $(CAG)_{154}$ repeat was targeted to the *SCA1* locus. The resulting mice demonstrate a SCA1 phenotype that more accurately represents the human SCA1 disease state, including Purkinje cell loss during the later stages, intranuclear inclusions in various brain regions, dysfunction of hippocampal and cortical neurons without significant death, and a shortened lifespan. These mice thus appear to represent a relatively accurate model of SCA1 [66].

Mouse models of HD also require overexpression or longer repeat lengths to produce a phenotype similar to the human disease state. The mouse homologue of *HD*, *Hdh*, is 86% identical to the human gene and shows a similar expression pattern [67], suggesting that it performs a homologous function in the mouse. Therefore the creation of transgenic mice containing an expanded CAG tract in *Hdh* seems as though it would be an accurate model of the human disease. However, as for *ataxin-1*, insertion of a CAG repeat tract large enough to cause HD in humans (50 or 80 repeats) fails to produce an HD phenotype in the mouse [68,69]. Insertion of a longer $(CAG)_{150}$ repeat does cause an HD-like phenotype, including disruption of movement. However, no neuronal death, even of the medium sized striatal neurons that are primarily affected in HD, has been detected in these mice [70].

Another transgenic mouse commonly used to study the molecular mechanisms of HD is the R6/2 line, which was initially created to observe intergenerational stability of expanded CAG repeats [71]. These mice carry a transgene containing the 5' end of *HD,* which consists of exon 1 including an expanded $(CAG)_{144}$ repeat and 1 kb of the promoter region. This promoter sequence is sufficient to drive ubiquitous expression of the huntingtin fragment in the brain, resulting in a progressive neurological phenotype, decrease in striatal volume and widespread neuronal nuclear and cytoplasmic inclusions [71]. Despite a lack of detectable neuronal death in these mice, they still recapitulate many of the features of HD, and have been utilised to test the effect of administration of transglutaminase inhibitor [72] or caspase inhibitor [73] on disease progression.

A particularly informative strain of transgenic mouse is the conditional HD mouse model generated by Yamamoto *et al* (2000). These mice contained exon 1 of *HD* with a $(CAG)_{94}$ repeat, the expression of which was controlled by a tetracycline-regulated transcription activator. Without tetracycline, huntingtin was expressed at high levels in the forebrain, including the striatum. This caused a progressive neurological phenotype in the mouse similar to that of the R6/2 mice, including reduction in striatal volume and clasping of the limbs when the mice were suspended by the tail, a behavioural abnormality common to mouse models of HD. All of these symptoms were apparent by 8 weeks of age, with motor behaviour disruption and neuropathology evident at 18 weeks. In mice given tetracycline, expression of the huntingtin fragment was turned off. Tetracycline administration for 16 weeks starting from 18 weeks of age improved the phenotype significantly. The progressive striatal changes were arrested and neurological function was partially restored (evident by a reduction in the clasping phenotype). These results indicate that continuous expression of the gene containing the repeat expansion is required to maintain the disease state, and that the mutant proteins can be cleared by the cell, even after neuronal dysfunction has occurred [74]. This conclusion was supported by results from a conditional SCA1 transgenic mouse model, where again cessation of mutant *ataxin-1* expression caused a complete reversal of pathology and motor dysfunction [75]. Recovery, albeit not complete, was observed even when expression was halted in the later stages of disease, suggesting that even in these later stages, cells maintain the ability to clear the mutant protein and repair the damage it has caused.

From analysis of these transgenic HD mouse models, it appears that shorter protein fragments containing expanded repeats are capable of producing a phenotype, whereas longer proteins are less toxic. This is also supported by evidence from a SCA3 transgenic mouse model: mice expressing full-length SCA3 containing Q79 show no phenotype, whereas mice expressing a truncated form of SCA3, consisting of only 43 amino acids and the Q79 repeat, develop ataxia and progressive degeneration of the cerebellum [76]. In the human disease context, short fragments of huntingtin also appear to be responsible for the HD disease phenotype [77]. Huntingtin acts *in vitro* as a substrate for caspase-3, a cysteine protease involved in apoptotic death, and the rate of this cleavage increases with the length of the polyglutamine tract [78]. The inhibition of caspase-3 in R6/2 mice slows progression of their HD-like disease phenotype [79], implicating caspase-3 cleavage of Huntingtin in disease progression. However, it is unclear whether this cleavage is a cause or an

effect of the neurodegenerative process in HD, and also whether a similar mechanism is involved in the other polyglutamine diseases.

Drosophila models of polyglutamine disease

In addition to mouse models of the polyglutamine diseases, *Drosophila melanogaster* has successfully been used to replicate the disease phenotype. Full-length ataxin-1 and truncated versions of ataxin-3 and huntingtin containing expanded polyglutamine tracts have been expressed in *Drosophila*, causing late onset, progressive neurodegeneration [64,80,81]. To create a *Drosophila* model of polyglutamine disease, the bipartite expression system UAS/GAL4 is frequently utilised, where various tissue-specific promoters drive the expression of GAL4 that in turn drives expression of a transgene located downstream of GAL4 upstream activating sequences (UAS). A *Drosophila* model of SCA3 used the GAL4-UAS system to allow comparison of the effects of expressing a fragment of ataxin-3 containing either Q27 or Q78 in various tissues of the fly. In all tissues, expression of ataxin-3-Q27 had no effect. In contrast, the ataxin-3-Q78 fragment caused early adult death or lethality when expressed in all developing cells of the peripheral and central nervous system (using *elav-GAL4* to drive expression). However, when this construct was expressed in epithelial cells of the imaginal discs (driven by *dpp-GAL4*), no deleterious effects were detected. Expression of expanded ataxin-3 using the eye-specific driver *GMR-GAL4* caused late onset progressive degeneration of photoreceptor neurons and pigment cells [80]. Expression of a fragment of huntingtin containing expanded polyglutamine repeats also caused a similar degenerative effect in a late onset, progressive manner. In the eye, huntingtin expression resulted in a decrease in the number of detectable rhabdomeres (subcellular light-gathering structures which are part of the photoreceptor neurons) from the normal seven to an average of less than five at 12 days after eclosion [64,82,83].

Drosophila has also been used to study the intrinsic toxicity of expanded polyglutamine repeats alone. Instead of expressing a protein implicated in polyglutamine disease, a Q108 repeat within a small peptide was expressed in various tissues of the fly, and the effects of this compared to the expression of the same small peptide containing Q22. While this shorter glutamine tract had no effect, the longer polyglutamine tract caused lethality or early adult death when expressed in

all neurons throughout development (using *elav-GAL4*). Lethality also occurred upon expression in the eye (driven by *GMR-GAL4*), suggesting that the *GMR* promoter may drive expression in tissues other than the eye. In non-lethal lines, degeneration of cells in the eye was observed, particularly of the photoreceptor neurons, whilst the mechanosensory neurons that produce the hairs of the eye were unaffected. Lethality was also observed when expression was driven by *dpp-GAL4* in the epithelial cells of late-stage imaginal discs. However, the addition of a 26 amino acid myc/flag epitope tag to the expanded polyglutamine peptide dramatically reduced its lethality when expressed in all of these tissue types [84]. These results suggest that polyglutamine tracts are intrinsically toxic in a cell type-dependent manner, and that this toxicity is modified by the inclusion of other amino acids.

One of the advantages of the *Drosophila* system is the relative ease of performing large scale screens to identify genes that modify a given phenotype when they are mutated or have their expression pattern altered [85]. This technique has been used successfully to identify genes that modify the readily visible eye phenotype resulting from expression of expanded polyglutamine protein in the eye [64,86]. Large-scale screens and also testing of candidate genes has been performed. These methods identified many groups of modifiers, including some already implicated in pathogenesis such as genes involved in protein folding and protein clearance, as well as suggesting novel mechanisms which may be involved, such as RNA processing, transcriptional regulation and cellular detoxification [64,86]. However, the large number of modifiers identified in these screens has made it difficult to separate those involved in the primary pathogenic pathway from those involved in the degenerative pathways downstream of the initial pathogenic insult.

Nuclear aggregates

Polyglutamine aggregates as pathogenic agent

Aggregates of expanded polyglutamine-containing proteins are a prominent pathological hallmark of polyglutamine disease. They are observed in each one of the diseases in neurons which undergo degeneration and also sometimes in those that do not. In post-mortem HD brains neuronal intranuclear and also cytoplasmic inclusions are observed in the cortex and the striatum [77]. Intranuclear inclusions consisting of aggregates of the protein containing the expanded repeat are also seen

in autopsy brains from patients with SCA1 [16], SCA3 [24], SCA6 [87], SCA7 [33], SCA17 [5], SBMA [88] and DRPLA [89]. However, in SCA2 and some cases of SCA6 only cytoplasmic inclusions have been reported [19,28]. In mouse and *Drosophila* models of polyglutamine disease, aggregates are a feature of the pathogenic process and aggregate formation in mice was shown to precede the onset of symptoms, consistent with aggregation playing a causative role in the disease process [90,91]. These observations led to the hypothesis that intranuclear aggregates of expanded polyglutamine proteins are central to the pathogenic mechanism [82,92,93].

The expanded polyglutamine aggregates in human disease tissue and in transgenic mice attract numerous additional cellular proteins such as transcription factors, caspases and normal interacting partners of the disease protein [94]. Ubiquitin is consistently found in polyglutamine aggregates [24,91,95], as are components of the proteasomal machinery and heat shock proteins [95,96]. The ubiquitin-proteasome pathway, whereby proteins are covalently modified with ubiquitin and then degraded by the proteasomal apparatus, is the major pathway by which cells remove misfolded proteins [97]. Therefore the detection of ubiquitin in aggregates raises the possibility that alterations in the ubiquitin-proteasome pathway may be involved in the pathogenesis of the polyglutamine diseases, resulting in the accumulation of misfolded proteins and a decreased ability to clear expanded polyglutamine [98,99]. This theory of pathogenesis has been supported by studies showing impairment of the ubiquitin-proteasome pathway by expanded polyglutamine in cell culture models [100] and also in mice [101].

In addition to ubiquitin and proteasome components, heat shock proteins (HSPs) have been identified in polyglutamine aggregates, again implying a role for protein misfolding and aggregation of misfolded proteins in pathogenesis. HSPs are involved in the folding of nascent polypeptides and also facilitate refolding and resolubilisation of misfolded proteins by acting as molecular chaperones which bind to the native or misfolded protein and allow it to adopt the correct conformation [102]. In this role, they are important during development and in response to cellular insult. Various HSPs co-localise with polyglutamine intranuclear aggregates [103,104], and the overexpression of HSPs suppresses neurodegeneration and improves motor function in B05 mice [96,105] and ameliorates toxicity of expanded polyglutamine proteins in cultured cells [106] and *Drosophila* [107]. A decrease in the formation of aggregates is also observed, leading to the conclusion that this decrease in aggregate formation

that is responsible for the beneficial effect of HSPs. However, others have reported that the protective effect may be mediated by inhibition of caspase activity, and is unrelated to the effect on aggregates [108].

Other cellular components that appear to be sequestered by polyglutamine aggregates are transcription factors. More than 20 transcription-related factors have been reported to interact with expanded polyglutamine proteins. These include TATA-binding protein [109], CREB-binding protein [110] and Sp1 [111], all of which contain a normal polyglutamine stretch that might mediate the interaction with expanded polyglutamine. The sequestration of these transcription factors and others is associated with repressed gene expression, which appears to be an early feature of pathogenesis and can be detected prior to the appearance of symptoms [109,111,112]. It is possible that this transcription factor depletion is an early step in the pathogenic pathway, causing misregulation of specific transcriptional programs, to which neurons are particularly sensitive.

The above evidence suggesting toxicity of polyglutamine aggregates has led to recent searches for substances that disrupt aggregate formation, with the aim of finding agents that may have therapeutic benefits. These studies have identified various substances that disrupt aggregate formation and also improve survival [113] [114], providing more evidence for a link between aggregation and pathogenesis.

Polyglutamine aggregates as innocent bystanders

In contrast to the findings described above, there is a significant body of evidence that does not support a role for aggregates in pathogenesis. Using the B05 line of SCA1 mice, which express human ataxin-1-Q82, Orr and colleagues investigated the importance of nuclear localisation and aggregation in disease pathogenesis. A line of mice carrying the same transgene as the B05 line, except with a mutation in the ataxin-1 nuclear localisation signal (NLS), was generated. In these mice, ataxin-1 showed almost exclusive cytoplasmic localisation, and no intranuclear aggregates were observed. Furthermore, these mice demonstrated no sign of ataxia or neurodegeneration, instead remaining disease free. These results suggest that nuclear localisation of expanded ataxin-1 is required for pathogenesis in this model of SCA1 [115]. In further experiments, a similar line of mice in which ataxin-1 contained a wild-type NLS but a mutation in the self-association domain was generated. This form of ataxin-1 was able to enter the nucleus but once there, did not form

aggregates. However, mice expressing this form of ataxin-1 still developed Purkinje cell pathology and a neurological disease nearly indistinguishable to that of B05 mice [115]. These findings demonstrate that while nuclear localisation is required for disease pathogenesis, the formation of nuclear aggregates is not, and therefore aggregates cannot be the pathogenic agent in this SCA1 mouse model.

The role of aggregates has also been investigated using an alternative model: cultured striatal and hippocampal neurons expressing an N-terminal fragment of huntingtin [116]. N-terminal huntingtin containing Q68 caused degeneration of the striatal neurons, but not the hippocampal neurons, mimicking the selectivity seen in HD. Addition of a nuclear export signal to the huntingtin fragment alleviated toxicity, suggesting again that the nucleus is the site of pathogenesis. However, intranuclear aggregates were seen in both striatal and hippocampal cells, even though only the striatal cells underwent huntingtin-induced death, separating aggregates from neurodegeneration. The presence of aggregates in neurons that do not undergo degeneration has also been noted in HD brains [77,117] and in mouse models [50,51].

The cultured striatal cell model was further used to directly investigate the link between aggregates and neuronal death. N-terminal huntingtin containing Q68 was co-expressed in the striatal cells with a dominant negative form of a ubiquitin-conjugating enzyme. This prevented ubiquitination, leading to a dramatic reduction in the presence of intranuclear aggregates, but surprisingly increased cell death. These results were interpreted to mean that aggregates form as the cell attempts to degrade mutant protein, or sequester it so that it is no longer free to cause toxicity [116]. Similar results have also been obtained in the mouse: inhibition of ubiquitin protein ligase in the Purkinje cells of SCA1 mice leads to a decrease in the number of intranuclear inclusions, but a significantly worse level of neurodegeneration [118].

A more recent investigation of the toxicity of aggregates utilised an HD cell model in which individual cells were followed throughout the study. This demonstrated that formation of aggregates in individual neurons decreased the level of mutant huntingtin elsewhere in the neuron and was a good predictor of improved survival of that neuron [119]. Another recent development is the combining of a SCA7 knock-in mouse model with a reporter for impairment of the ubiquitin-proteasome system. This was used to investigate the role of aggregation and impairment of the ubiquitin

proteasome system in pathogenesis, and demonstrated a striking inverse correlation between neuropathology and nuclear inclusions [120]. Furthermore, neuronal dysfunction was observed before impairment of the ubiquitin proteasome system. Thus there are many examples where aggregates and proteasome impairment have been dissociated from pathogenesis, suggesting that these processes may not play a central role in the pathogenic pathway.

Polyglutamine aggregates as protective agent

Rather than acting as a toxic agent, it has also been proposed that aggregates play a protective role by sequestering expanded polyglutamine-containing protein [94]. Many of the studies described above have found that a decrease in aggregation correlates with an increase in toxicity [116,118,119], lending weight to this theory. Recently a molecular mechanism for this protective role was proposed. Polyglutamine aggregates in HD patient brains, mouse and fly models and cell culture were shown to sequester mTOR, the mammalian target of rapamycin that acts to inhibit autophagy. Sequestration of mTOR increases autophagy, which causes an increased clearance of polyglutamine fragments and thereby protects against polyglutamine toxicity [121]. This mechanism would provide an explanation for the protective role that aggregation seems to play under some paradigms.

Thus, despite an ever-increasing amount of evidence regarding the role of aggregation in pathogenesis, the contribution it makes remains unclear. The question of whether polyglutamine aggregates are deleterious, harmless or protective remains unanswered, and continues to be an area of contention.

One pathogenic pathway or many?

Another unanswered question in the polyglutamine field is whether the polyglutamine diseases are manifestations of the same pathogenic process, or instead caused by different pathways of pathogenesis. There can be striking similarities in phenotype between the diseases. Overlap between the disease phenotypes leading to misdiagnosis has been reported, with cases including a SCA17 homozygote displaying a HD phenotype [122], SCA17 presenting like HD [123] and clinical overlap between the DRPLA and HD phenotypes [124]. In addition, common features such as aggregates and the involvement of the ubiquitin-proteasome pathway

suggest they share a common pathogenic pathway, and this seems to be the most parsimonious explanation. However, there is also evidence supporting disease-specific pathogenic mechanisms.

A compelling example comes from the B05 SCA1 transgenic mouse expressing ataxin-1-Q82. Using this model, it was demonstrated that a specific residue (serine 776) in ataxin-1 outside of the polyglutamine repeat tract is phosphorylated *in vitro* and *in vivo* [125]. Phosphorylation at this residue appears to affect the ability of ataxin-1 to aggregate, as mutation of the serine to an alanine (that can no longer be phosphorylated) prevents the protein from forming aggregates in cultured cells. To investigate the role of phosphorylation at this residue, a variation of the B05 mice expressing ataxin-1-Q82-S776A (with serine 776 mutated to alanine) was generated. These mice showed a substantial decrease in pathogenesis compared to B05 mice expressing ataxin-1-Q82, based on both Purkinje cell morphology and behavioural analysis [125]. Thus it seems that serine 776 is critical for ataxin-1 containing an expanded polyglutamine to induce pathogenesis, and is likely to be involved in SCA1 pathogenesis.

Further investigation using a biochemical approach to identify proteins that interact specifically with ataxin-1 phosphorylated at S776 identified 14-3-3, a ubiquitous regulatory molecule that binds phosphomotifs to regulate a wide variety of cellular processes. In addition, the protein that acts to phosphorylate S776 was identified as Akt kinase, an anti-apoptotic signalling molecule. Remarkably, in a *Drosophila* model of SCA1 both Akt kinase and 14-3-3 appear to be required for pathogenesis [126]. Although other proteins implicated in the polyglutamine diseases are phosphorylated, the involvement of Akt kinase and 14-3-3 in disease pathogenesis seems likely to be specific to SCA1. Therefore, it is unclear how these findings relate to the other disorders, and whether they are an indication that there may not be a unifying pathogenic mechanism.

In a different study looking for interactors of ataxin-1, an interaction was identified between the AXH domain of ataxin-1 and the transcription factor Senseless, which leads to the degradation of Senseless, in both mice and *Drosophila*. Loss of *Senseless* expression in the mouse phenocopies SCA1, causing Purkinje cell degeneration and ataxia. Furthermore, in mice the destabilization of Senseless caused by ataxin-1 overexpression leads to the loss of the specific neuronal types

(Purkinje cells) that degenerate in patients with SCA1. This lead to the proposal that polyglutamine expansion in ataxin-1 serves to stabilise the interaction between ataxin-1 and Senseless, leading to increased degradation of Senseless and Purkinje cell degeneration [127]. This model provides a mechanism for SCA1 pathogenesis, and the selective vulnerability of Purkinje cells in this disease. However, the mechanism proposed is specific for SCA1, based on interaction with Senseless that does not involve the polyglutamine domain of ataxin-1, again supporting the notion that the polyglutamine diseases may be caused by different pathogenic mechanisms.

Alternative hypothesis: pathogenic polyalanine

In addition to the polyglutamine disorders, there are another nine diseases caused by expansion of a homopolymeric amino acid repeat, but in these cases the repeated amino acid is alanine. Eight of these alanine expansions occur in transcription factors that play important roles during development, and the resulting phenotypes are congenital malformation syndromes [128]. In these cases, the same phenotypes can arise from different mutations in the same gene, suggesting that they are due to loss of function of the affected transcription factor [129]. However, this is not the case for one of the polyalanine diseases, oculopharyngeal muscular dystrophy (OPMD), which seems to instead share similarities with the polyglutamine diseases. OPMD is a late-onset, progressive disorder characterised by drooping eyelids, difficulty in swallowing and proximal limb-muscle weakness. This phenotype is caused by expansion of a $(GCG)_6$ repeat encoding a polyalanine tract in poly(A) binding protein 2 (PABP2), which expands to 8-13 copies [8]. PABP2 is ubiquitously expressed and functions to stimulate poly(A) polymerase to control the length of poly(A) tails after mRNA processing. In OPMD, aggregates of PABP2 containing the alanine expansion are observed in muscle cells, the major site of pathology. These aggregates also contain ubiquitin and components of the proteasome, as well as poly(A) RNA [130].

Based on the observation of polyalanine aggregates in OPMD, and the fact that much shorter alanine tracts can give rise to disease compared to polyglutamine, a novel pathogenic mechanism for the polyglutamine disorders involving aggregation of polyalanine tracts has been proposed. Peptides composed of polyalanine tracts appear to be more toxic and prone to aggregation than polyglutamine when they are expressed in cultured cells [131]. Therefore Gaspar *et al.* proposed that frameshifting

to the GCA frame occurring during the transcription of expanded CAG results in the production of polyalanine-containing proteins that are more toxic than polyglutamine, and thus contribute to pathogenesis [132,133]. Indeed, an antibody raised against the C-terminal sequence of ataxin-3 predicted to result from polyalanine-producing frameshifts showed accumulation of such frameshift products in the polyglutamine aggregates of a SCA3 patient brain [132]. Furthermore, in a cell culture model, frameshifts were shown to occur in a CAG length-dependent manner and to be harmful to cells, with the addition of a ribosome-interacting drug that reduces frameshifting able to decrease the toxicity of expanded polyglutamine [133]. This evidence supports the theory that polyalanine toxicity contributes to polyglutamine pathogenesis, and by a mechanism that could potentially occur in all of the polyglutamine diseases – however, more investigation into whether such frameshifting occurs in diseases other than SCA3 is required.

Thus, in spite of the considerable amount of data generated over the last decade in the search to explain the mechanism behind polyglutamine toxicity, many fundamental questions remain unanswered. It is not known how much overlap there is between pathogenesis in the different disorders and what the role of aggregation is in the pathogenic process. Even the nature of the major pathogenic pathway is unknown, although multiple processes that are perturbed in the disease state have been identified. Although there are many suggestions as to what might be the primary insult, it seems that these have raised more questions than provided answers.

Untranslated repeats

The toxic polyglutamine hypothesis appears to at least partially explain the phenotypes of those diseases involving translated CAG repeats. However, finding a unifying mechanism is complicated further by the existence of similar dominant neurodegenerative diseases caused by repeats that are not translated. Three of these, SCA8, SCA10 and SCA12, show pathology and clinical features typical of spinocerebellar ataxia, and show late onset, anticipation, and selective sensitivity of a subset of neurons like the polyglutamine diseases. However, the molecular basis of these SCAs is expansion of repeats in non-coding regions, either in the 5' untranslated region (UTR) of a transcript in SCA12 or in the case of SCA8, in a non-coding RNA [134,135]. Furthermore, in SCA10 the causative mutation is the expansion of a pentanucleotide repeat located in an intron [136,137]. How these repeat expansions can cause diseases so similar in phenotype to those resulting from polyglutamine expansion is unclear.

To date, nine diseases have been ascribed to the expansion of untranslated repeats. These diseases tend to be relatively mild in the lower repeat length range and exhibit a larger range of repeat lengths, with more severe pathology observed when repeat lengths are in the upper range. A broader range of repeats has been observed to undergo expansion in untranslated regions (Table 1.2). For some of these diseases, the mechanism by which the repeat expansion causes disease has been determined. These include fragile X and XE, where the expansion causes transcriptional silencing due to methylation, leading to loss of function of the affected genes, and Friedreich's ataxia, where the expansion of an intronic repeat causes loss of expression by interfering with transcription elongation [138]. These loss-of-function diseases are recessive and therefore unlikely to share a common pathogenic mechanism with the remaining untranslated repeat diseases, which are dominantly inherited. In these cases a mechanism whereby these silent mutations exert a dominant effect is, as yet, unclear.

Disease	Repeat	Gene, region	Inheritance, mechanism	Repeat length	
				Normal	Disease
Fragile X	CGG	*FMR1*, 5' UTR	X linked, Loss-of-function	6-53	200-1000
FXTAS	CGG	*FMR1, 5' UTR*	Dominant, unknown	6-53	60-200
Fragile XE	GCC	*FMR2*, 5' UTR	X linked, Loss-of-function	7-35	130-750*
Friedreich's ataxia	GAA	*X25*, intron	Autosomal recessive, loss-of-function	6-34	80-1700*
Myotonic dystrophy type 1 (DM1)	CTG	*DMPK*, 3' UTR	Dominant, RNA gain-of-function	5-37	50-3000
DM2	CCTG	*ZNF9*	Dominant, as for DM1	10-26[+]	75-11000
SCA8	CTG	*SCA8*, non coding	Dominant, unknown	16-37	107-127
SCA10	ATTCT	*SCA10,* intron	Dominant, unknown	10-29	800-4500
SCA12	CAG	*PPP2R2B*, 5' UTR	Dominant, unknown	9-28	55-78
HDL2	CTG	*junctophilin-3*, depends on splicing	Dominant, unknown	14-19	51-57

**Shorter expansions in this range are premutations, intermediate alleles that do not cause disease but are long enough to be unstable and can expand into the full mutation through germline transmission.*
[+]DM2 alleles in the normal range are interrupted by GCTG and TCTG motifs, whereas expanded alleles consist of uninterrupted CCTG repeats.

Table 1.2 Diseases caused by expansion of untranslated triplet repeats. The gene containing the repeat, region of the gene in which the repeat lies, mode of inheritance and mechanism by which the expansion causes disease are shown for each disease, as well as the number of repeats in normal and premutation/disease alleles. FXTAS=Fragile X cerebellar tremor/ataxia syndrome. HDL2=Huntington's disease-like 2. For references see [139], [134], [140], [4], [141] and [137].

SCA8

SCA8 is caused by a CUG repeat expansion in the 3' end of a non-coding RNA [134,142]. The *SCA8* transcript encodes a natural antisense transcript of the mRNA that encodes the Kelch-like 1 protein (KLHL1), and therefore it is likely to play a role in regulating expression of KLHL1 [142]. The *SCA8* transcript is expressed at low levels in brain and lung [134]. SCA8 is a slowly progressive, predominantly cerebellar ataxia with marked cerebellar atrophy, affecting gait, swallowing, speech, and limb and eye movement [143]

SCA10

SCA10 is the only known disease caused by expansion of a pentanucleotide repeat. It is characterised by degeneration of neurons in the cerebellum and to a lesser extent in the brainstem and cortex, leading to ataxia that is sometimes accompanied by seizures [144]. The molecular basis of SCA10 is expansion of an ATTCT repeat in the 9th intron of the *SCA10* gene [136]. This encodes a protein of unknown function that is highly expressed in the brain and also in the testis, heart, kidney and skeletal muscle with lower levels of expression in the liver and blood leukocytes [137]. The ATTCT repeat expansion is large in all SCA10 cases identified so far, ranging from

800 to 4,500 repeats, making it one of the biggest microsatellite repeats in the human genome.

SCA12

The only disorder caused by a CAG repeat that does not encode polyglutamine is SCA12. In this case the repeat is located in the 5' UTR of *PPP2R2B*, encoding a brain-specific regulatory subunit of the protein phosphatase PP2A [145]. *PPP2R2B* is widely expressed in neurons throughout the brain, including constitutive expression in Purkinje cells of the cerebellar cortex [146]. In SCA12, degeneration of neurons is widespread, with generalised atrophy of the central nervous system (CNS) dominantly affecting the cerebral cortex and cerebellum. The SCA12 phenotype presents as a tremor followed by ataxia, decreased movement and other variable features such as psychiatric symptoms [145,147].

FXTAS

Fragile X cerebellar tremor/ataxia syndrome (FXTAS) is the most recently described disorder caused by an untranslated repeat expansion. The expansion occurs at the Fragile X locus, and previously patients harbouring repeats in the FXTAS range (53 - 200 copies) were thought to be non-symptomatic carriers of the Fragile X premutation. Repeat lengths in this range are unstable and prone to expansion upon transmission, which can result in fragile X mental retardation in the offspring of permutation carriers. Fragile X is an X-linked, loss of function disorder caused by silencing of the *FMR1* gene. Repeat expansion above 200 copies of a CGG repeat in the promoter region causes hypermethylation of the promoter, leading to transcriptional silencing of *FMR1* [148]. In contrast, FXTAS is not caused by loss of function – repeat lengths in the premutation range appear to result in an increase in transcription of *FMR1* [149]. The FXTAS phenotype is quite distinct to that of Fragile X; instead of mental retardation from birth, it is a late onset neurodegenerative disorder characterised by intention tremor, ataxia and cognitive decline [141]. Degeneration of Purkinje cells in the cerebellum is observed, as well as generalised cortical atrophy [150]. The FXTAS phenotype is not observed in all premutation carriers – it is infrequently observed in female carriers (possibly depending on the proportion of cells in which the expanded allele undergoes X inactivation) [151], and is seen in 15 - 75% of males [152]. Recently it has been proposed that this incomplete penetrance may be caused by the frequent occurrence in the population of AGG interruptions in the CGG repeat [153].

A surprising feature of FXTAS is the presence of intranuclear aggregates in the brain of affected patients. These are detected throughout the cerebellum and brainstem, and are ubiquitinated but do not stain positively for FMR1 protein or polyglutamine [154]. The presence of aggregates in a disease that is not caused by a mutation affecting a coding sequence is unexpected. In the polyglutamine disorders, aggregation of proteins containing an expanded polyglutamine tract is not surprising, but in FXTAS there would seem to be no reason for aggregates to form. No such aggregates have been reported in any of the other untranslated repeat disorders; however this may be because prior to the description of the FXTAS aggregates, there was no reason to suspect that aggregation would occur in these disorders.

Huntington's disease phenocopies

In some patients who present with an HD phenotype and are diagnosed with HD, a CAG repeat in *HD* is not detected. In the majority of these patients, the HD phenotype is due to a HD phenocopy termed Huntington's disease-like 2 (HDL2). As this disease also displays anticipation, it was suspected that it is also caused by repeat expansion. Indeed this is the case, with an expanded repeat recently detected in these patients at a locus unrelated to *HD*. This expansion was mapped to the gene *junctophilin-3*, encoding a protein involved in the formation of junctional membrane structures. The repeated triplet is CTG, located in a variably spliced exon. Alternative transcripts contain the repeat located in the 3' UTR, or translated as a polyalanine or polyleucine tract [4]. In none of these detected transcripts does the repeat encode polyglutamine. How this mutation can cause a phenotype that is clinically indistinguishable from that caused by polyglutamine expansion in HD is not known.

Myotonic dystrophy (DM)

The most well characterised untranslated repeat disease is myotonic dystrophy type 1 (DM1). The molecular basis of DM1 is expansion of a CTG repeat in the 3' UTR of the *Dystrophia myotonia protein kinase* (*DMPK)* gene [155]. DM1 is a multisystemic disease characterised by hyperexcitability of skeletal muscle (myotonia), muscle degeneration (myopathy), cardiac conduction defects, cataracts and neuropsychiatric impairment. Diabetes, kidney failure, testicular atrophy, infertility and early frontal balding in males are also common [156]. In DM1, the expansion of the CTG repeat from the normal range (5 - 37 repeats) into the pathogenic range (50 to several thousand repeats) appears to have an effect on expression of multiple genes in the

region. The most obvious target is *DMPK*, as the repeat expansion is located in the 3' UTR of this transcript. Indeed, *DMPK* transcripts with CUG expansions show altered splicing and are retained in the nucleus instead of translocating to the cytoplasm for translation, reducing cellular levels of DMPK [157]. This led to the suggestion that DM1 results from haploinsufficiency of DMPK, a hypothesis tested by the generation of *DMPK* knockout mice. However, heterozygous mice fail to show a phenotype, and -/- mice have mild myopathy [158] and cardiac conduction defects [159], but do not show any other characteristic features of the disease. Directly upstream of *DMPK* is the S*ix5* gene, a member of the *Six* gene family, which are mouse homologs of the *Drosophila sine oculis* gene [160]. Like *DMPK*, reduced levels of *Six5* transcript are observed in DM1 muscle tissue [161,162]. However, the only phenotype of *Six5* -/- mice is early development of cataracts [163,164]. Downstream of *DMPK* is a WD-repeat gene (*DMWD)*, which is highly expressed in the testis and also shows decreased levels of transcription in DM1 [165]. Disruption of *DMWD* expression could therefore plausibly contribute to the infertility aspect of the phenotype in males. From these observations of the genes in the region, it was suggested that DM1 is a multigene disease, with decreased levels of *DMWD*, *Six5* and *DMPK* all making a contribution to the multisystemic phenotype [166]. However, the hallmark features of DM1, myotonia and myopathy, are still not accounted for in this multigene model. Furthermore, no *DMPK* point mutations have been associated with the DM1 phenotype, suggesting that DM1 is not caused by DMPK haploinsufficiency.

RNA Pathogenesis

A landmark in DM1 research came in the form of a transgenic mouse expressing the *human skeletal actin* gene containing an expanded CTG repeat in the 3' UTR. The resulting CUG-containing transcript was expressed in skeletal muscle and mimicked the mutation that causes DM1 except that the gene involved was unrelated to *DMPK*. The mice showed myotonia and myopathy, the characteristic features of DM1, with pathology dependent on expression of the actin transgene containing the CTG repeat and also on the length of the repeat [167]. These results suggest that the CUG repeat-containing RNA exerts a dominant toxic effect to cause the characteristic DM1 features in these mice, a process which is unrelated to *DMPK* [168].

Further evidence from cell culture studies and examination of human disease tissue supports an RNA gain-of-function mechanism in DM1. *DMPK* mRNA containing the expanded CUG repeat forms double-stranded hairpin structures *in vivo* [169,170]. These stable structures form foci in the nuclei of cells from DM1 patient tissue, and also in cultured cells [171], whereas normal transcripts are processed and transported to the cytoplasm. Evidence that this accumulation of mRNA can exert a dominant effect comes from the observation that expression of mutant *DMPK* 3' UTR mRNA in cultured myoblasts inhibits their differentiation [172]. A possible mechanism for this dominant effect is via an interaction with proteins that bind to CUG repeats in RNA.

Expanded *DMPK* transcripts interact with a number of proteins that bind to CUG repeat-containing RNA. These include CUG-binding protein (CUG-BP) and human muscleblind proteins. Expression of both of these proteins is altered in DM1 tissue and transgenic mouse models: muscleblind proteins show co-localisation with foci of DMPK transcript in the nucleus and appear to be sequestered by the mutant RNA [173-175], while CUG-BP levels are increased in DM1 tissue [176]. The normal role of both the muscleblind proteins and CUG-BP is in regulation of alternative splicing [177,178]. Indeed, they appear to play antagonistic roles, with CUG-BP generally promoting foetal splice forms involved in development and muscleblind proteins promoting adult splice forms [177]. Therefore an increase in CUG-BP levels and/or a decrease in the activity of muscleblind proteins could lead to the inappropriate persistence of foetal isoforms of target transcripts. This mechanism of pathogenesis has been supported by a number of observations. Importantly, a number of target transcripts whose splicing is regulated by CUG-BP and muscleblind proteins show altered splicing in DM1 tissues. These transcripts include:

- cardiac troponin T (cTNT), a component of cardiac muscle, which shows altered splicing in DM1, providing a hypothetical link between the RNA gain-of-function mechanism and cardiac conduction defects [178].

- Muscle chloride channel 1 (ClC-1), which shows aberrant splicing and decreased levels on the surface of in DM1 cells in a mouse model and in human disease tissue, accounting for the hyperexcitability of skeletal muscle seen in DM1 [179].

- Insulin receptor (IR), implicated in DM1 by the frequent occurrence of insulin resistance in DM1 patients. Aberrant regulation of alternative splicing of the insulin receptor pre mRNA occurs in DM1 skeletal muscle tissue, resulting in predominant expression of the lower-signalling non-muscle isoforms [180].

Additional support for the splicing alteration model of DM1 pathogenesis comes from a *muscleblind* knockout mouse, which displays myotonia, myopathy, cataracts and RNA splicing abnormalities characteristic of DM1 [181], further implicating the muscleblind family of proteins in DM1.

RNA toxicity in DM1 may also arise from mechanisms other than altered splicing. One of these may be the trapping and depletion of transcription factors by the double-stranded hairpin structures formed by expanded CUG repeats in the nucleus. In a cultured DM1 muscle cell model, expanded *DMPK* transcripts bind transcription factors involved in activation, cell maintenance and differentiation, including Sp1 and retinoic acid receptor gamma. This results in a depletion of these transcription factors from chromatin, associated with a decrease in expression of their target genes. One such target is *CLCN1*, the gene encoding ClC-1, which showed a suppression of transcript levels in cultured DM1 cells that can be restored to normal by increasing expression of *Sp1*. From this data the authors argue that Sp1 depletion leading to decreased ClC-1 levels may be the primary cause of hypotonia in DM1 patients, on which with splicing alterations are secondarily imposed [182]. However, this transcription factor-leaching hypothesis is yet to be confirmed in tissues from affected patients.

Myotonic dystrophy type 2

Analysis of a subset of patients presenting with the DM1 phenotype revealed that some do not have the CTG repeat expansion in *DMPK*, or indeed do not have any mutations in *DMPK.* However they present with the same clinical phenotype, and the disease shows anticipation, suggesting that it is caused by repeat expansion [183]. Linkage analysis connected the disease phenotype in these patients, termed myotonic dystrophy type 2 (DM2), to a gene unrelated to *DMPK* and located on a different chromosome. The nature of the disease-causing mutation is expansion of a CCTG tract in intron 1 of *zinc finger protein 9* (*ZNF9)*. The CCTG repeat is polymorphic, with normal alleles containing up to 27 repeats and disease alleles analysed in the study containing 75 – 11,000 uninterrupted repeats [184].

Remarkably DM2 shows the same multisystemic features as DM1: characterised by myotonia and myopathy with cataracts, cardiac conduction defects, insulin resistance and male hypogonadism [183]. This suggests that the two diseases share a common pathogenic mechanism. However, the genes in the *ZNF9* region have no relationship to those surrounding *DMPK* [184]. This provides compelling evidence that DM1 and DM2 are caused by RNA gain-of-function, rather than loss of function of genes such as *DMPK*, *Six5* or *DMWD*, as the only common entity in the two diseases is mRNA carrying an expanded repeat. Analysis of DM2 further suggests that this is the case: Expanded *ZNF9* transcripts form a hairpin structure [185], are retained in the nucleus and form foci in DM2 muscle tissue like those seen in DM1 [184]. These RNA foci colocalise with muscleblind proteins [173-175]. Furthermore, aberrant splicing of CUG-BP and muscleblind target transcripts is observed in DM2 [179,186]. Thus valuable evidence gained from DM2 confirms the identification of RNA as the pathogenic agent in DM1.

Animal models of untranslated repeat disorders

Aside from DM1 and DM2, relatively little is known about the pathogenic mechanism in the untranslated repeat diseases, and because their causative mutations have only recently been discovered, there are fewer tools such as mouse models available to study their pathogenesis. However, in the case of SCA8 and FXTAS, *Drosophila* models have been generated to investigate how repeat expansion in these diseases contributes to neurodegeneration.

To generate a SCA8 model, the *SCA8* transcript containing 9 or 112 CUG repeats was expressed in the fly eye using the UAS-GAL4 system and *GMR* as a driver. Expression of *SCA8-(CTG)$_{112}$* lead to disorganization of the ommatidia and mechanosensory bristles, resulting in a rough eye. However, this phenotype was also observed upon expression of *SCA8-(CTG)$_9$*, creating doubt as to whether the expanded CUG or expression of the *SCA8* transcript itself is responsible for the degeneration. A screen for genes that differentially modify the rough eye phenotype revealed two dominant enhancers of the *SCA8-(CTG)$_{112}$* phenotype. These modifiers are loss-of-function alleles of *muscleblind* and *split ends*, which have different effects on the *SCA8-(CTG)$_9$* phenotype: *muscleblind* shows milder enhancement and *split*

ends shows stronger enhancement [187]. This difference in interaction strengths suggests that there may be different molecular pathways involved in *SCA8-(CTG)*$_{112}$ and *SCA8-(CTG)*$_{9}$ pathogenesis. The involvement of muscleblind is interesting, as it points to commonalities in pathogenesis between SCA8 and DM1 and raises the possibility of a dominant RNA mechanism in SCA8 similar to that which has been characterised in DM1 and DM2.

Generation of a *Drosophila* model has also implicated RNA in the pathogenesis of FXTAS. The 5' UTR of *FMR1* containing a CGG repeat in the expanded or normal range was fused to the reporter gene GFP and expressed in the eye using *GMR-GAL4*. Moderate expression of the CGG repeat-containing transcripts in the normal range (60 copies) resulted in no phenotype, whereas neurodegeneration was observed when the repeat was expanded to 90 copies. Expression in all neurons using the driver *elav-GAL4* resulted in lethality, but expression in the epithelial cells of the developing imaginal discs using *dpp-GAL4* did not cause a phenotype, suggesting that the degenerative effect is cell type-specific. In the eye, expression of $(CGG)_{90}$ induced neurodegeneration, observed as cell death and ommatidial disruption. In addition, the effect was progressive and dosage sensitive. Interestingly, nuclear and cytoplasmic aggregates were observed, which stained positively for ubiquitin, Hsp70 and the proteasome but did not contain GFP or polyglutamine. Furthermore, the neurodegenerative phenotype could be suppressed by overexpression of Hsp70, or enhanced by expression of a dominant negative form of Hsp70 [188]. The reason behind this genetic interaction with Hsp70 is unclear, as in FXTAS only the RNA contains the repeat expansion, so there is no mutant protein to misfold and be a chaperone target. However, it raises the possibility that protein misfolding may be involved in this disorder; perhaps the proteins that misfold are those that have an altered or inappropriate interaction with the transcript containing the expanded repeat. These results also imply that the primary pathogenic agent in FXTAS is RNA.

Hypothesis:
RNA as a unifying pathogenic agent in expanded repeat diseases

While it is widely accepted that expanded polyglutamine-containing proteins are responsible for the disease phenotype in the polyglutamine diseases, examination of

the available evidence shows that this has not been proven. Many unanswered questions remain, in particular regarding how untranslated repeats can cause a similar phenotype if polyglutamine is the toxic agent. An alternative hypothesis is that the phenotype in the polyglutamine diseases is due to the presence of the expanded repeats in mRNA.

There is strong evidence implicating RNA gain-of-function in the pathogenesis of DM and FXTAS, demonstrating that expanded repeats in mRNA are capable of causing disease. In these cases, the RNA containing the expanded repeat forms a stable secondary structure, termed a hairpin, caused by the binding of the complementary C-G bases in the repeat. This has been shown to occur in the CUG repeats in the *DMPK* transcript [189], where the transcript is retained in the nucleus and induces pathogenesis by inappropriate interactions with RNA binding proteins. In FXTAS, the CGG repeats in the *FMR1* transcript can also form a stable hairpin secondary structure [153], and RNA pathogenesis has been implied using the *Drosophila* system. Furthermore, AGG interruptions in the *FMR1* repeat tract that occur frequently in the population can act to prevent formation of a long, stable hairpin, providing a convincing explanation for the incomplete penetrance of FXTAS if hairpin formation is a requisite step in pathogenesis [153].

In accordance with RNA being the pathogenic agent in polyglutamine disorders, it has been demonstrated that RNA containing expanded CAG repeats can also form a stable hairpin structure [185]. Furthermore, hairpin formation occurs in the context of the full length transcript for three genes investigated: *ataxin-3*, *CACNA1A* and *atrophin-1* [190]. In these studies, the hairpin secondary structures were predicted by the Mfold algorithm [191], and confirmed using nuclease digestion of mRNA transcripts generated *in vitro*. Thus hairpin formation of expanded CAG repeat-containing transcripts may also occur *in vivo*, providing support for the suggestion that it could act as the pathogenic agent.

In addition to hairpin formation by RNA, the presence of aggregates in FXTAS suggests that RNA pathogenesis in this disease may share parallels with the polyglutamine diseases. Polyglutamine aggregates are regarded by some as the pathogenic agent in the polyglutamine disorders, yet there is significant evidence suggesting otherwise. Furthermore, despite many years of research, the primary pathogenic agent in polyglutamine diseases remains to be determined. This suggests

that either the true pathogenic agent has not been uncovered, or that the polyglutamine disorders do not share a common pathogenic pathway, a notion that is supported by disease-specific mechanisms uncovered by some studies [125,192]. However, disease-specific pathways are not the most parsimonious explanation for the similar phenotypes seen in the polyglutamine diseases and many of the untranslated diseases. The RNA hypothesis instead predicts that an agent common to all of the expanded repeat diseases (an mRNA transcript carrying an expanded repeat) is the primary pathogenic agent in all of the disorders, perhaps involving formation of a hairpin secondary structure. Even the *SCA10* transcript, the odd one out in terms of the fact that it does not contain a CNG repeat but rather a pentanucleotide AUUCU repeat, forms an RNA hairpin structure [193]. Thus, in addition to a transcript containing a repeat expansion, the ability to form an RNA hairpin seems to be a common property of the dominant diseases caused by repeat expansions. In addition, the pathogenic threshold for the translated and untranslated repeat diseases is roughly similar (Figure 1.1), which is in accordance with them sharing a pathogenic mechanism.

How could RNA cause pathogenesis in the polyglutamine diseases?

The assumption that proteins containing expanded polyglutamine are pathogenic in the polyglutamine disorders has meant that other potential mechanisms, specifically those involving RNA, remain unexplored. If RNA is indeed the primary pathogenic agent, likely mechanisms involve inappropriate interactions between the RNA transcript and double-stranded RNA (dsRNA) binding proteins or proteins that bind specifically to rCAG repeats. A study aiming to identify such CAG binding proteins found two proteins isolated from human cortex and striatum that interact specifically with expanded CAG repeats in RNA in a length dependent manner [194]. The identity of these proteins has not been determined, but they would be prime candidates in a pathogenic pathway involving RNA.

Another protein that binds to RNA containing CAG expansions is protein kinase R (PKR), a dsRNA-dependent protein kinase. PKR is implicated in DM pathology: the PKR dsRNA binding domain binds directly to expanded CUG repeats in *DMPK* with length-dependent affinity [170]. Similarly, PKR also binds to *HD* transcripts, with a preference for those containing a CAG expansion. In addition, PKR shows increased activation in regions of the brain most affected in HD, in both HD autopsy brains and in HD transgenic mouse brains [195]. The normal role of PKR is in virally-induced

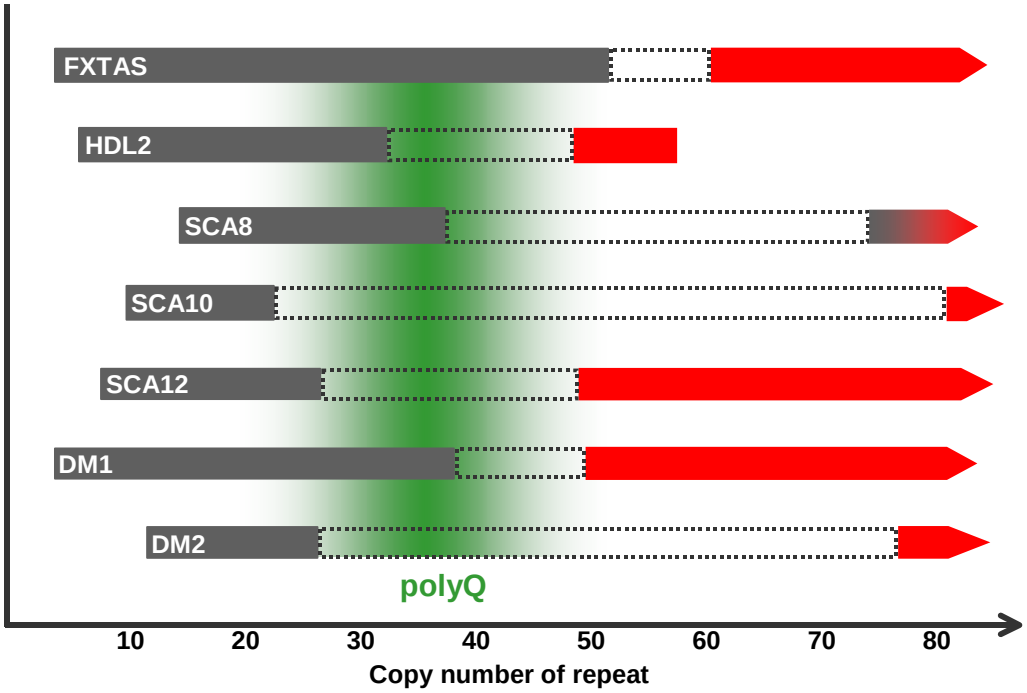


Figure 1.1 Pathogenic thresholds for the untranslated repeat diseases roughly overlap the pathogenic threshold for the polyglutamine diseases. Repeat copy numbers in the normal range are indicated in grey for each untranslated repeat disease, and copy numbers in the pathogenic range are in red. Dashed regions in the middle demonstrate copy numbers that have not been detected; therefore it is unknown precisely where the pathogenic threshold lies. The pathogenic threshold range for the polyglutamine disorders is shown in green. In the case of SCA8, the relationship between repeat length and disease is complex, with unaffected carriers of expanded alleles reported, suggesting that incomplete penetrance occurs in some genetic backgrounds.

and stress-mediated apoptosis [196], suggesting that inappropriate activation may lead to an increase of cellular sensitivity to apoptotic stimuli. Therefore, inappropriate PKR activation in both HD and DM by mRNA containing expanded repeats could provide a common mechanism leading to increased cell death in translated and untranslated repeat diseases. However, PKR is also activated by inflammation and stress signals, and is upregulated in Alzheimer's disease brains [197] and Parkinson's disease brains [198], suggesting that PKR activation may be a general feature of degeneration. This observation suggests that it is more likely to be a downstream effect of neurodegeneration in the expanded repeat disorders, rather than being involved in the primary pathogenic insult.

In DM1 and DM2, the muscleblind proteins are known to bind to mRNA transcripts containing expanded repeats, leading to perturbation of their normal function. Using a yeast three-hybrid assay, which detects the interaction between RNA and protein, it

was determined that muscleblind proteins bind to CHG and CHHG repeats, where H = A, U or C. This included binding to rCAG containing 16-70 repeats [199], raising the possibility that muscleblind could be involved in RNA pathogenesis in the polyglutamine disorders, as it is in DM. In support of this, expression of mRNA transcripts containing CUG or CAG repeats in cultured cells leads to formation of nuclear foci that colocalise with endogenous muscleblind proteins [200]. Fluorescent recovery after photobleaching (FRAP) analysis was used to investigate the strength of interaction between muscleblind and the repeat transcripts in these foci, and this demonstrated similar half-times of recovery and similar fractions of immobile molecules in foci formed by CUG and CAG repeats, which were also similar to recovery times seen from foci in DM cells. These results suggest that rCAG can form foci and relocalise muscleblind proteins in a similar manner to rCUG. However, upon examination of splicing patterns of *cTNT* and *IR*, two muscleblind targets that show altered splicing in DM, splicing alterations in these transcripts were observed in cells expressing rCUG but not rCAG [200]. This suggests that foci formation and muscleblind colocalisation are separable from splicing alterations in DM. The implications of this finding for RNA pathogenesis in CAG repeat disorders are unclear: it may suggest that muscleblind is not involved in rCAG pathogenesis, or that different targets of muscleblind are affected by the CAG repeat. In addition, the CUG and CAG repeat tracts used in this study consisted of 960 repeats, interrupted every 20th repeat with a TCGAG motif. It is unclear what effect these interruptions have on the secondary structure of the repeats and their ability to form a hairpin. Also, 960 copies of the CAG repeat is well in excess of the copy numbers observed in the polyglutamine disorders, so this may cast doubt over the relevance of this study to polyglutamine pathogenesis. Nonetheless, it does demonstrate that rCAG repeats are capable of forming nuclear foci and interacting with RNA binding proteins such as muscleblind.

Evidence against RNA-mediated pathogenesis

There is little evidence directly contradicting a role for RNA gain-of-function in the polyglutamine diseases, as the dominance of the toxic polyglutamine hypothesis has meant that research into the disease mechanism has focused on polyglutamine. In 1996, Goldberg *et al.* generated a transgenic mouse containing the full-length huntingtin cDNA containing 44 CAG repeats and including the 5’ UTR. However, a 120 bp region of the 5’ UTR acted as an inhibitor of translation, meaning that the gene was transcribed but not translated. These mice showed no HD-like phenotype,

even though they were expressing high levels of *HD* mRNA. This was interpreted to mean that a translated polyglutamine tract is essential for development of disease, and thus mRNA containing the expanded repeat is not the pathological agent [201]. However, repeats of this length have consistently failed to cause a phenotype when expressed in transgenic mice [202]. Therefore the lack of phenotype does not necessarily imply that mRNA is not the mediator of disease, but may instead be due to the fact that larger repeat expansions are necessary to generate a phenotype with an age of onset that falls within the lifetime of the mouse.

The most compelling evidence dispelling an RNA gain-of-function mechanism in the polyglutamine diseases comes from analysing the results of the previously described study investigating the requirement of nuclear localisation for pathogenesis in B05 mice. Mice expressing ataxin-1-Q82 developed neuropathology. However, when the NLS in ataxin-1 was mutated, causing ataxin-1 to remain in the cytoplasm, mice expressing this construct failed to develop the same phenotype, even though they were expressing both mRNA containing the expanded CAG repeat and an expanded polyglutamine-containing protein [115]. These results appear to rule out mRNA as a pathogenic agent, and also suggest that nuclear localisation is required for disease. However, this cannot apply to all polyglutamine diseases, as SCA2 and SCA6 are caused by CAG expansions in non-nuclear proteins. Furthermore, the effect of the introduced mutations on the structure of the mRNA transcript was not examined in this model. Similarly, mutation of one specific amino acid in ataxin-1 that lies outside the repeat region severely diminishes pathology in a mouse model of SCA1 [125], apparently due to its effect on the interaction between ataxin-1 and 14-3-3 [126]. This finding is again difficult to account for using an RNA-based model of pathogenesis, and can only be explained by the introduced mutation affecting the secondary structure of the mRNA transcript, a suggestion that seems unlikely.

In addition to these findings, the X linked mode of transmission of SBMA points to protein-mediated pathogenesis. The lack of an SBMA phenotype in females carrying the repeat expansion has been clarified by findings from a *Drosophila* model in which pathogenesis was dependent on the ingestion of androgen, which activates the AR and causes its translocation to the nucleus [203]. Similar results were obtained from a mouse model, where significant pathology was not observed in females or castrated males [204]. This requirement for androgen is difficult to reconcile with an RNA-based model of pathogenesis – it is also difficult to relate this finding to the

other polyglutamine disorders, particularly SCA2, where pathogenesis occurs despite ataxin-2 showing no nuclear localisation.

Another finding that does not support RNA-mediated pathogenesis is that the addition of polyglutamine aggregates generated *in vitro* to cultured cells induces toxicity. Aggregates of simple polyglutamine monomers were added to and passively taken up by the cells and showed cytoplasmic localisation; addition of an NLS to the polyglutamine sequence was required to induce toxicity [205]. However, this finding is contradicted by another study comparing the toxicity of nuclear and cytoplasmic aggregates of huntingtin. No significant difference in toxicity of nuclear compared with cytoplasmic aggregates was detected, supporting the notion that both the nucleus and the cytoplasm are sites of pathogenesis [206].

Thus, while there are many reports documenting the toxic effects of expanded polyglutamine tracts on cells, and the main focus of research into the polyglutamine diseases is on the polyglutamine-containing proteins, there is contradictory information, and no clear picture of pathogenesis emerges that can be applied to all of the disorders. This lack of clarity supports the proposal that the role of RNA in the disease process requires investigation, as RNA based pathogenesis can provide a unifying mechanism, not just for the polyglutamine diseases but also those caused by untranslated repeats. Alternatively, the data available can also be interpreted to suggest that there is not a common pathogenic pathway, for the polyglutamine diseases or the untranslated repeat diseases, and that different mechanisms are at work in each of the polyglutamine disorders and also in the untranslated repeat disorders.

Evidence supporting RNA hypothesis

As the suggestion that RNA is the pathogenic agent in polyglutamine disorders has not been addressed, there is little direct evidence to support such a hypothesis. However, indirect support comes from studies examining the structures of the mRNA transcripts in the polyglutamine diseases. In the case of SCA2, interruptions to the CAG repeat by CAA codons are observed frequently in the population. Expanded alleles in the pathogenic range containing interruptions are either not associated with disease or correlate with a significantly later age of onset than would be predicted by looking at the total length of the repeat, despite the fact that CAA also encodes glutamine and so the polyglutamine tract is not interrupted. Investigation of the effect

of these CAA interruptions on the secondary structures formed by the transcripts revealed that the repeat regions form shorter, branched hairpins, meaning that the length of the longest hairpin is decreased. In fact, the length of the longest hairpin shows good correlation with age of onset, whereas the length of the polyglutamine repeat does not, leading the authors to conclude that their data “may be considered supportive for the RNA contribution to SCA2 pathogenesis” [207]. Similar findings have also been described for the structure of the *ataxin-1* transcript and its correlation with age of onset of SCA1 [208].

In another study by the same group, sequences flanking the repeat and their effect on its structure were examined for the *CACNA1A*, *atrophin-1* and *ataxin-3* transcripts. These transcripts were chosen based on differing pathogenic thresholds for disease: as few as 20 CAG repeats in the *CACNA1A* gene cause SCA6, whereas as many as 53 and 55 repeats in *atrophin-1* and *ataxin-3* are required to confer the full symptoms of DRPLA and SCA3, respectively (Table 1). Analysis of the sequences flanking the repeat revealed that in the *CACNA1A* transcript, the sequences flanking the repeat are complementary and base-pair with each other, which acts to stabilise the hairpin structure formed by the repeat. In contrast, no binding of flanking sequences occurs in the *atrophin-1* and *ataxin-3* transcripts, meaning that their repeat tracts form several slipped hairpin variants [190]. Thus these findings demonstrate a rough correlation between stability of the RNA hairpin and the pathogenic threshold for disease, and provide an explanation for the variation in these pathogenic thresholds if the diseases are RNA-mediated. In contrast, the polyglutamine hypothesis offers no such explanation for the observed variation in pathogenic thresholds.

Therefore, whilst there is no direct evidence for RNA pathogenesis in the polyglutamine disorders, there is circumstantial evidence supporting it, and the possibility has not been conclusively ruled out. In addition, there are many observations that do not fit together based on our current understanding of polyglutamine pathogenesis. Based on this, RNA as the pathogenic agent warrants further investigation. Recently the possibility of a ‘mechanistic overlap’ between polyglutamine and RNA disorders was raised [209], and a review of repeat expansion diseases questioned whether “expansions that are mediated at the protein level also have an RNA component” [210]. The following experiments are designed to examine and address this possibility.

Approach: use of the *Drosophila* eye as a model

To investigate the RNA hypothesis, and other aspects of polyglutamine pathogenesis, the *Drosophila* eye was used as a model system. *Drosophila* is well established as a model for polyglutamine pathogenesis: expression of expanded CAG repeats, whether in the context of a human disease gene or not, induces neurodegeneration in a length dependent manner [80,81,86,203]. The *Drosophila* eye is a convenient system in which to study polyglutamine pathogenesis, as the effect of expression on the cells and overall structure of the eye can be readily observed.

The *Drosophila* eye is composed of a hexagonal array of approximately 800 ommatidia (Figure 1.2 A). Each ommatidium contains 8 photoreceptor cells and a mechanosensory bristle, which are neuronal, and pigment and cone cells, which are non-neuronal. The photoreceptor cells consist of a rhabdomere, or microvillar stack of membranes in which photopigment resides, and an axon projecting along the optic stalk to the brain. The rhabdomeres of photoreceptors R1 – R6 lie in an asymmetric trapezoid, with R7 and R8 in the middle (Figure 1.2 B and C). As R8 lies directly below R7, tangential sections across the eye reveal only 7 of the 8 photoreceptors: R1 to R6, and either R7 or R8 (Figure 1.2 B). Each cluster of 8 photoreceptors is surrounded by pigment cells that contain pigment granules and act to optically insulate the unit. Above the photoreceptors lie four cone cells, which secrete the central part of the lens. Each ommatidial unit also contains one small mechanosensory bristle that projects a sensory axon into a different region of the brain [211]. Thus the eye is a complex structure composed of various neuronal and non-neuronal cell types.

The *Drosophila* eye represents a useful model system that has been utilised to study polyglutamine toxicity. Many well-characterised driver lines exist which express GAL4 in a tissue-specific manner and can be used to drive polyglutamine expression in various patterns. In the case of the eye, *GMR-GAL4* drives expression in all cells following differentiation, allowing the effects on both neuronal and non-neuronal cells to be determined. In addition, expression in the eye can be restricted to neurons using *elav-GAL4*, allowing the effects on neurons to be examined in isolation, and compared to the effect of expression in all cells. However, *elav-GAL4* drives expression in all neurons of the central and peripheral nervous system, and effects on the whole organism (such as early death) can also arise.

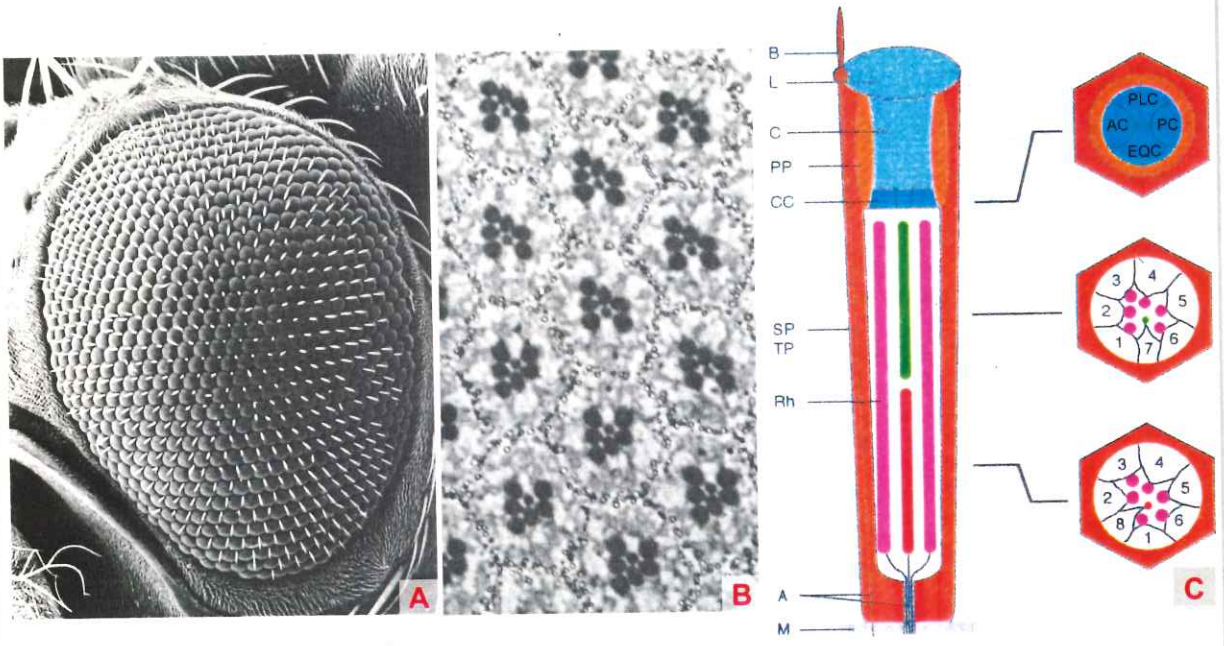


Figure 1.2 Appearance and structure of the *Drosophila* eye. (**A**) Scanning electron microscope image of a wild-type adult *Drosophila* eye, showing the regular array of ommatidia and bristles. (**B**) Tangential section across the eye. Each ommatidium contains seven visible photoreceptors surrounded by pigment granules. (**C**) Schematic representation of an ommatidial unit, showing the positions of the components of the ommatidium. A: photoreceptor cell axons, AC: anterior cone cell, B: bristle, C: liquid-filled pseudocone, CC: cone cells, EQC: equatorial cone cell, L: lens, M: basal membrane, PC: posterior cone cell, PLC: polar cone cell, PP: primary pigment cell, Rh: rhabdomere, SP: secondary pigment cells, TP: tertiary pigment cells, 1-8: photoreceptor cells R1-R8. Images reproduced from [212].

Previous investigations into the toxicity of polyglutamine have demonstrated that it causes degeneration of cells in the *Drosophila* eye. This is the case for both polyglutamine tracts alone [84,86] and when they are in the context of proteins involved in the human diseases such as huntingtin [81], ataxin-3 [80] or the androgen receptor [203]. The resulting eye phenotypes are characterised by degeneration and appear to show cell type-specificity, affecting some cells but not others. For example, expression of a polyglutamine peptide composed of 108 glutamines, separated from the context of a disease-related protein, in all cells of the eye driven by *GMR-GAL4* leads to degeneration (Figure 1.3 C). This was visible externally as loss of colour due to complete degeneration of the non-neuronal pigment cells, suggesting that they are sensitive to polyglutamine toxicity. In contrast, the mechanosensory cells that produce the bristles of the eye were unaffected. To examine the effect of polyglutamine expression specifically on neurons, *elav-GAL4* was used (Figure 1.3 D). This also induced degeneration, which was visible in tangential sections as a

disruption to the organisation of the cells, and holes in the tissue due to cell death [84].

These findings, and others from the disease-specific models of SCA3 and Huntington's disease described earlier, suggest that the fundamental mechanisms of polyglutamine-mediated neurodegeneration, namely cell-type specificity and age-dependent degeneration, are conserved in *Drosophila*. Therefore it seems that the *Drosophila* eye represents a suitable system for studying the mechanisms of polyglutamine toxicity, using these previously established genetic tools.

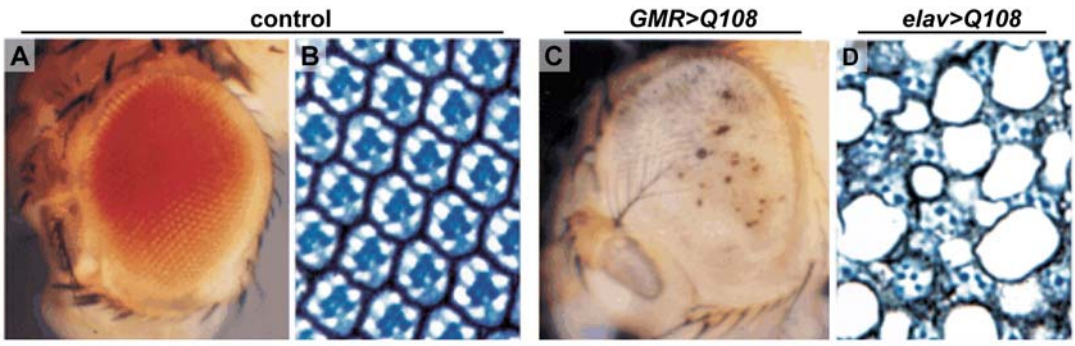


Figure 1.3 Effect of polyglutamine peptide expression in the *Drosophila* eye. (**A**) Light microscope image of a control *Drosophila* eye, carrying the *GMR-GAL4* driver alone. The eye shows red pigment and a regular array of ommatidia. (**B**) Tangential section of a wild-type eye stained with toluidine blue. The ommatidia are in an orderly array and surrounded by pigment granules. (**C**) External appearance of the eye of a fly expressing a polyglutamine peptide consisting of 108 polyglutamines, driven by *GMR-GAL4*. Pigment loss is apparent throughout the eye, and black patches are visible in the eye; these may be due to necrotic death. (**D**) Tangential section through the eye of a fly expressing the 108-glutamine peptide in all neurons, driven by *elav-GAL4*. The orderly array of ommatidia is disrupted, cell death has led to holes appearing in the tissue, and many ommatidia contain fewer than seven photoreceptors. Images reproduced from [84].

Chapter 2 – Materials and Methods

Materials

Enzymes

T4 polynucleotide kinase, 3' phosphatase free: Roche

T4 DNA ligase: Roche

Restriction endonucleases: New England Biolabs (NEB)

LR Clonase: Invitrogen

Pfu DNA polymerase: Stratagene

Taq DNA polymerase: Invitrogen

DNase I: Invitrogen

RNase H: Invitrogen

Superscript II reverse transcriptase: Invitrogen

Proteinase K: Sigma-Aldrich

SYBR green PCR master mix: Applied Biosystems

Kits

QIAquick gel extraction kit: Qiagen

QIAquick PCR clean-up kit: Qiagen

GenElute plasmid miniprep kit: Sigma-Aldrich

QIAGEN spin midiprep kit: Qiagen

QuickChange site directed mutagenesis kit: Stratagene

Enhanced Chemiluminescence (ECL) kit: Amersham Pharmacia

Plasmids

pBluescript KS+: Stratagene

pUAST: obtained from R. Saint

pDEST-UAST: pUAST modified by H. Dalton with Gateway cloning cassette, Invitrogen

pENTR/D-TOPO: Invitrogen

Antibiotics

Ampicillin: Sigma-Aldrich

Kanamycin: Sigma-Aldrich

Molecular weight markers

DNA:	1 Kb Plus DNA Ladder (Invitrogen) sizes (in bp) 100 200 300 400 500 650 850 1000 1650 2000 up to 12,000 in 1000 bp increments
Protein:	Benchmark pre-stained protein ladder (Invitrogen) sizes (in kDa) 8, 15, 20, 27, 38, 50, 65, 80, 115, 180 (sizes vary between batches; approximate sizes are given)

Oligonucleotides

Oligos were obtained from Geneworks. Oligos used for cloning and *in vitro* mutagenesis were purified to mutagenesis grade by reverse phase HPLC; Oligos for PCR and sequencing were standard PCR grade. All oligo sequences are given 5'-3'.

Cloning Oligos to generate repeats:

$(CAG)_{10}$-Fw CAGCAGCAGCAGCAGCAGCAGCAGCAG

$(CAG)_{10}$-Rv CTGCTGCTGCTGCTGCTGCTGCTGCTG

$(CAA)_{10}$-Fw CAACAACAACAACAACAACAACAACAA

$(CAA)_{10}$-Rv TTGTTGTTGTTGTTGTTGTTGTTGTTG

linker AAGCTTGGCCAAGCTT

Cloning oligos to generate coding sequence of short peptide with myc and flag tags:

part1-Fw CACCATGAGGAGCCGAAAGCTTAGGAGCCAAG

part1-Rv GGCCTTGGCTCCTAAGCTTTCGGCTCCTCATGGTG

part1 frameshift-Fw CACCATGAGGAGCCGAAAAGCTTGGAGCCAAG

part1 frameshift-Rv GGCCTTGGGTCCAAGGTTTTCGTCGGCTCCTCATGGTG

part2-Fw GCCCTGAGCAGAAACTCATCTCTGAAGAGGATCTGAATG

part2-Rv AATTCATTCAGATCCTCTTCAGAGATGAGTTTCTGCTCAG

part3-Fw AATTCCTGCAAGATTACAAGGATGACGATAAGTAGTCTAGA

part3-Rv TCTAGACTACTTATCGTCGTCATCCTTGTAATCTTGCAGG

In vitro mutagenesis oligos:

IVM-Fw CCACCATGAGGAGCAGCTGAAAGCTTAGGAGC

IVM-Rv GCTCCTAAGCTTTCAG TCCTCATGGTGG

Oligos for PCR and sequencing:

pUAST-F GAAGAGAACTCTGAATAGGG

PUAST-R GTCACACCACAGAAGTAAGG
UAS-F CCTTAGCATGTCCGTGG

Oligos for real-time PCR:
repeat-Fw TGTGGTGTGACATAATTGGACAA
repeat-Rv TGCTCCCATTCATCAGTTCC
GAL4-Fw CACTGACCCCGTCTGCTTTG
GAL4-Rv GGTTCGGACCGTTGCTACTG
Rp49-Fw ATCGATATGCTAAGCTGTCGCAC
Rp49-Rv TGTCGATACCCTTGGGCTTG

Antibodies

primary:
α myc 9E10, mouse monoclonal (D. Lawrence, IMVS) at 1/200 (western) or 1/250 (immuno-staining)
α Hsp70, mouse monoclonal (Stressgen) at 1/100 and 1/300 (immuno-staining)
α β-tubulin, mouse monclonal (Sigma-Aldrich) at 1/10,000 (western)

secondary:
α mouse-alexa488 (Rockland) at 1/300 (immuno-staining)
α mouse-HRP (Rockland) at 1/2000 (western)

Bacterial media

All media were prepared with distilled and deionised water and sterilised by autoclaving, except heat labile reagents, which were filter sterilised. Antibiotics were added from sterile stock solutions to the media after it had been autoclaved.

L-Broth (LB):	1% (w/v) amine A, 0.5% yeast extract, 1% NaCl, pH 7.0.
SOC:	2% bactotryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl,10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose.
Plates:	L-Broth with 1.5% (w/v) bactoagar supplemented with ampicillin (100 mg/L) or Kanamycin (50 mg/L) where appropriate.

Drosophila media

Fortified (F1) *Drosophila* medium: 1% (w/v) agar, 18.75% compressed yeast, 10% treacle, 10% polenta, 1.5% acid mix (47% propionic acid, 4.7% orthophosphoric acid), 2.5% tegosept (10% para-hydroxybenzoate in ethanol)

Grape juice agar plates: 0.3% agar, 25% grape juice, 0.3% sucrose, 0.03% tegosept

Buffers and solutions

TAE: 40 mM Tris-acetate, 20 mM sodium acetate, 1 mM EDTA, pH 8.2

PBS: 7.5 mM Na_2HPO_4, 2.5 mM NaH_2PO_4, 145 mM NaCl

Annealing buffer: 10 mM Tris pH 7.5, 50 mM NaCl, 1 mM EDTA

Agarose gel loading buffer: 50% (w/v) glycerol, 50 mM EDTA, 0.1% (w/v) bromophenol blue, 0.1% (w/v) xylene cyanol

Heat shock competent cell buffer 1: 30 mM potassium acetate, 100 mM $RbCl_2$, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 15% (w/v) glycerol

Heat shock competent cell buffer 2: 10 mM MOPS (or PIPES), 75 mM $CaCl_2$, 10 mM $RbCl_2$, 15% (w/v) glycerol

Embryo injecting buffer: 5 mM KCl, 0.1 mM $NaPO_4$ pH 6.8

Squishing buffer: 10 mM Tris-Cl pH 8.2, 1 mM EDTA, 25 mM NaCl, 200 µg/ml Proteinase K

SDS sample buffer (2 x): 125 mM Tris pH 6.8, 4% sodium dodecyl sulfate (SDS), 10% glycerol, 0.006% bromophenol blue, 2% β -mercaptoethanol

Transfer buffer (1 x): 3.0275 g/L Tris base, 14.413 g/L glycine, 20% methanol

Western blocking buffer: 10 mM Tris pH 7.5, 100 mM NaCl, 0.1% Tween-20

Nitrocellulose stripping buffer: 62.5 mM Tris-HCl pH 6.8, 2% SDS, 100 mM β-mercaptoethanol

Cytoskeletal buffer: 10 mM Hepes (pH 7.4), 200 mM sucrose, 3 mM $MgCl_2$, 50 mM NaCl, 0.5% Triton-X, 0.02% NaN_3

Methods

Standard molecular genetic techniques were performed as described in [212].

Annealing Oligonucleotides

To anneal two single-stranded oligonucleotides to generate a double-stranded DNA fragment for cloning, equimolar quantities of the two oligos were combined in annealing buffer to a concentration of 250 ng/µl. This was heated to 95 °C then cooled slowly to room temperature. The 5' ends were then phosphorylated using T4 polynucleotide kinase in the supplied kinase buffer (50 mM Tris-Hcl, 10 mM $MgCl_2$, 5 mM dithioretol) for 2 hours at 37 °C. The resulting double-stranded DNA fragments were then ready for ligation or cloning.

Generation of recombinant plasmids

Ligation: DNA fragments to be ligated were placed in a mix (total volume 20 µl) containing 1 U of T4 DNA ligase and 1 x ligation buffer and incubated at 15 °C overnight. For transformation by electroporation, the ligation was phenol/chloroform extracted, precipitated by adding 1 ml glycogen, 1/10 volume 3 M NaAcetate pH 5.2 and 2.5 volumes ethanol, then washed in 70% ethanol prior to resuspension in 10 µl Milli Q water (MQ H_2O). For transformation by heat shock, the ligation mix was transformed without purification.

Purification of DNA from agarose gels: a QIAquick gel extraction kit was used to purify DNA bands excised from agarose gels, using the manufacturers protocol.

Transformation of Bacteria

Electroporation:

500 ml of L-broth was inoculated with 5 ml of an overnight culture of *E. coli* DH5α cells or SURE cells (Invitrogen) and grown to an $OD_{A}600$ of 0.5-0.6. The culture was then chilled in an ice slurry for 15 to 30 minutes and the cells harvested by centrifugation at 4000 g for 15 minutes. The cells were then resuspended in 500 ml of ice-cold MQ H_2O, pelleted at 5000 g, resuspended in 250 ml of ice-cold MQ H_2O, pelleted at 4000 g, resuspended in 10 ml of ice-cold 10% glycerol, re-pelleted at 3000 g and finally resuspended in 1 ml of ice-cold 10% glycerol. The competent cells were then snap frozen in liquid nitrogen and stored as 45 µl aliquots at -80 °C.

For transformation, cells were thawed at room temperature, added to ligation reaction mixture and incubated on ice for at least 30 seconds. Cells were then transferred to an ice-cold 2 mm electroporation cuvette and electroporated in a Bio-Rad Gene Pulser set to 2500 V, 25 mFD capacitance and Capacitance Extender set to 500

mFD. The cuvette was immediately washed out with 1 ml of SOC, and the suspension incubated at 37 °C for 30 minutes. Cells were then pelleted for 8 seconds at 16,000 g, then 800 ml of the supernatant was removed, and the cells gently resuspended in the remaining SOC. The cell suspension was plated onto L-agar plates supplemented with ampicillin or kanamycin as appropriate, and incubated at 37 °C overnight. If selection for β-galactosidase activity (blue/white colour selection) was required, 56 µl of 100 mM IPTG and 16 µl of 50 mg/ml X-Gal were added before plating.

Heat shock:
250 ml of L-broth was inoculated with 1 ml of an overnight culture of *E. coli* DH5α cells or SURE cells (Invitrogen) and grown to an $OD_{A}600$ of 0.5-0.6. The cells were chilled on ice for 30 minutes, and harvested by centrifugation at 4 °C at 3000 g for 10 minutes. The cells were then resuspended in 2/5 volumes (100 ml) of ice-cold heat shock competent cell buffer 1, and incubated on ice for 10 minutes. They were then pelleted by centrifugation at 4 °C at 3000 g for 10 minutes and resuspended in 1/25 volumes (10 ml) of ice-cold heat shock competent cell buffer 2. They were then incubated on ice for 30 minutes and aliquoted into pre-chilled microfuge tubes in 200 µl aliquots, snap frozen in liquid nitrogen and stored at -80 °C.

For transformation, cells were thawed at room temperature, 50 µl added to each ligation reaction and the mixture incubated on ice for 20 minutes. They were then heat-shocked for 45-50 seconds at 42 °C, returned to ice for 2 minutes, then 450 µl LB was added and the mixture inverted gently. The cells were incubated for 1 hour at 37 °C, pelleted for 5 minutes at 600 g, 200 µl of the supernatant was removed, and the cells gently resuspended in the remaining LB, and plated as above.

Isolation of plasmid DNA
Small-scale preparation of plasmids was performed using the Sigma GenElute plasmid miniprep kit according to the manufactures instructions. Large-scale preparation was performed using the Qiagen Midi prep kit according to the manufactures instructions.

PCR amplification of DNA
Pfu polymerase was used according to the manufacturers instructions. PCR conditions were: 0.5 U Pfu polymerase, 1 ng template DNA, 0.1 ng primers and 0.2

mM dNTPs in 1 x Pfu PCR buffer (20 mM Tris-HCl, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$, 0.1% Triton X-100, 0.1 mg/ml BSA, pH 8.75). Reactions were performed using an MJ Research PTC-200 peltier thermal cycler, with the following conditions: 94 °C for 2 minutes, 9 cycles of: 94 °C for 30 seconds, 53 °C for 30 seconds and 72 °C for 2 minutes, then 29 cycles of 94 °C for 30 seconds, 51 °C for 30 seconds and 72 °C for 2 minutes.

The resulting PCR products were purified using a QIAquick PCR clean up kit, or analysed by agarose gel electrophoresis, excised and extracted using a QIAquick gel extraction kit according to the manufacturers instructions.

In vitro site directed mutagenesis

The design of primers for the site directed mutagenesis followed the instructions provided in the QuickChange Site-Directed Mutagenesis kit. The reaction was carried out in a 20 µl volume. To each reaction the following was added: 2 µl of 10 x reaction buffer, 50 ng double-stranded DNA template, 125 ng forward primer, 125 ng reverse primer, 1 µl of 10 mM dNTP mix (2.5 mM each dNTP) and MQ H_2O to a final volume of 19 µl. 1 µl Pfu Turbo DNA polymerase (2.5 U/µl) was added last. The reaction was cycled according to the manufacturers instructions in an MJ Research PTC-200 peltier thermal cycler. The reaction products were transferred to a new microfuge tube, 1 µl of *Dpn* I (10 U/µl) was added, mixed and incubated at 37 °C for 1 hour to digest the parental (nonmutated) DNA. The reaction was then phenol/$CHCl_3$ extracted, ethanol precipitated and resuspended in 10 µl of MQ H_2O. 5 µl was then transformed into *E. coli*. The introduced mutation was confirmed using sequencing analysis.

Automated Sequencing

DNA was sequenced using the ABI Prism™ Dye Terminator Cycle Sequencing Ready Reaction Mix (Perkin-Elmer), essentially as described in the manufacturer's protocol with the modification of using half the described amount of reaction mix. 400-800 ng of double-stranded DNA was used as a template, and approximately 100 ng or 18 pmol of primer was used. Reactions were performed using an MJ Research PTC-200 peltier thermal cycler, with the following conditions: 25 cycles of 96 °C for 30 seconds, 50 °C for 15 seconds and 60 °C for 4 minutes. Running and analysis of Dye Terminator gels was conducted by the Sequencing Centre at the IMVS.

Generation of repeat constructs

The construct encoding a short peptide with myc and flag tags was generated in three sections by annealing forward and reverse (complementary) oligos, creating three separate double stranded fragments that were then ligated together to form the whole coding region. An alternative version of the first section was also generated that is frameshifted relative to the first version by insertion of one base pair before the repeat and deletion of one base pair after the repeat. This has the effect of changing the reading frame of the repeat to the +2 frame, while leaving the myc and flag coding regions unaltered. A third version was also generated, in which a termination codon was inserted upstream of the repeat by *in vitro* mutagenesis. These various constructs were either ligated into pBluescript, or cloned into pENTR/D-TOPO by TOPO-mediated cloning according to the manufacturers instructions.

To generate the repeats, double stranded $(CAG)_{10}$ or $(CAA)_{10}$ fragments were generated by annealing complementary oligos. These were then blunt-end ligated, and the products of the ligation run on a 2% agarose gel. Ligation products of the desired size were excised from the gel, gel extracted, and then double stranded linkers containing *Hind* III sites were ligated onto the ends. The ligation products were digested with *Hind* III, and cloned into the short peptide coding sequence, in pBluescript or pENTR/D-TOPO, which had also been digested with *Hind* III. The longest repeats generated using this method consisted of 30 copies of the repeat; these were expanded using PCR-based methods [213,214]. The resulting repeat constructs were subcloned into pUAST by restriction digest and ligation, or into pDEST-UAST by LR clonase-mediated recombination. Constructs were then analysed by restriction digest and sequencing to confirm that the repeat copy number was unchanged and that no mutations had been introduced. These were then used for microinjection to generate transgenic *Drosophila*.

Drosophila cultures

Flies were raised at 18 °C or 25 °C with 70% humidity on F1 medium. Unless otherwise stated, crosses were performed at 25 °C, and newly emerged adults (0 - 24 hours old) used for analysis.

P-element mediated transformation of Drosophila

Micro-injection: DNA for injection was prepared using the GenElute plasmid miniprep kit or the QIAGEN spin midiprep kit according to the suppliers protocol. An injection

mix was prepared to a concentration of 0.5-1 mg/ml transformation vector plasmid DNA and 0.3 mg/ml of pp25.7wc (Δ2-3 transposase) plasmid in 1 x Embryo injecting buffer. The injection needle was back-filled using a drawn out capillary containing 2 ml of the injection mix that had been centrifuged briefly to remove any particulate matter. w^{1118} embryos to be injected were collected from 30 minute lays on grape juice agar plates at 25 °C, dechorionated in 50% bleach for 3 minutes, and rinsed thoroughly in MQ H_20. Embryos were then aligned along a strip of non-toxic rubber glue such that their posterior ends face the needle. A drop of liquid paraffin was placed over the embryos and the slide placed on the stage of an inverted microscope. A micromanipulator was used to position the needle, with injection carried out by moving the microscope stage to bring the embryos to the needle, such that a very small amount of DNA was injected into the posterior cytoplasm.

Identification of transformants: The slides of microinjected embryos were placed in a petri-dish containing moist tissue paper with a small amount of yeast paste surrounding the embryos and kept at 25 °C to allow the embryos to hatch. Larvae were collected after 2 days using strips of Whatman paper and placed in a fly food vial where they developed into adult flies. The flies were then crossed to w^{1118} flies to identify transformants on the basis of the *white*$^{+}$ eye pigmentation phenotype. Numerous independent transformants were mapped to determine the chromosome of insertion, using the dominantly marked balancer chromosomes CyO and TM6B in the stocks *w ; Gla / CyO* and *w ; Df(3R)roXB3 / TM6B*, and homozygous lines generated using these balancers.

Fly crosses and strains

To generate flies carrying two independent insertions on the same chromosome, two lines of flies carrying independent insertions were crossed, and trans-heterozygous virgins selected in which the chromosomes containing the two insertions can recombine. These were crossed to w^{1118} males, and in the progeny of this cross, flies carrying a recombinant chromosome containing both insertions were selected based on eye colour. This method was used to generate flies carrying two copies of term$(CTG)_{114}$ and term$(CAG)_{93}$, described in Chapter 5. To generate lines of flies carrying both GMR-GAL4 and a polyglutamine-encoding construct to allow screening of modifiers described in Chapter 4, eye colour could not be used to identify recombinants as the eye colour associated with the GMR-GAL4 insertion is too dark to allow visualization of the extra pigment resulting from expression of the second

copy of the *white* gene. Therefore, PCR was used to identify flies carrying the polyglutamine-encoding insertion. In the case of the term$(CAA)_{99}$ construct, the insertions used to generate the line of flies carrying two copies of the construct were on different chromosomes. Therefore in the resulting flies with two independent insertions, one insertion is on the II chromosome and one is on the III.

Genomic DNA preparation

To prepare genomic DNA, a single fly was crushed in 50 µl squishing buffer and incubated at 37 °C for 30 minutes. The proteinase K was then inactivated by heating to 95 °C for 2 minutes. 1 µl of the resulting solution was used as a template for PCR, allowing amplification and sequencing of the inserted transgene using the primers pUAST-F and pUAST-R, or UAS-F and pUAST-R.

Protein gel electrophoresis

Protein samples were collected from wandering third instar larvae, and consisted of ten pairs of eye discs with mouth hooks attached. These were boiled in 1 x SDS sample buffer for 2 minutes, then loaded into a 15% SDS-PAGE gel and electrophoresed at 120 V for 10 minutes then 160 V for 1 hour. A BioRad mini-Protean 3 gel electrophoresis system was used with 0.8 mm analytical gels prepared according to manufacturer's instructions.

Western blotting

Transfer of proteins onto nitrocellulose membrane was performed using a BioRad mini trans-blot electrophoretic transfer cell according to manufacturer's guidelines. Electrophoresis gels were equilibrated in transfer buffer for 30 minutes and then transferred at 100 V for 1 hour. Nitrocellulose blots were blocked for 1 hour at room temperature in 5% skim milk powder in western blocking buffer. Primary antibody incubations were carried out overnight at 4 °C and secondary antibody incubations for 2 hours at room temperature, with the appropriate dilutions of antibody in western blocking buffer with 5% skim milk powder. The secondary antibody was HRP-conjugated and detected by ECL. To strip the membrane for re-probing with the loading control, the membrane was incubated in nitrocellulose stripping buffer at 50 °C for 30 minutes with rotation/shaking. It was then rinsed and blocked in blocking buffer, and reprobed.

Preparation of whole RNA from Drosophila heads

Approximately 100 heads were collected from flies aged 0 – 24 hours. These were crushed with a micropestle in 100 µl Trizol (Invitrogen), then a further 900 µl Trizol was added and they were incubated at room temperature for 5 minutes. 200 µl chloroform was added, and the solution was spun at 12000 g for 10 minutes. The supernatant was discarded, and the pellet washed with 70% ethanol (DEPC treated) and spun at 7000 g for 5 minutes. This was then resuspended in MQ H_20 (DEPC treated), and further purified using an RNeasy spin column from Qiagen according to the manufacturers instructions.

Real time PCR

cDNA synthesis:1 µg of total RNA was treated with DNase I according to the manufacturers instructions. cDNA was then synthesised using oligo(dT)12-18 or gene-specific primers and Superscript II reverse transcriptase. This was then treated with RNase H to remove complementary RNA. The cDNA was diluted (20 µl in a total of 70 µl), and 11.1 µl of this dilution used as template for real-time PCR.

Real time PCR: SYBR green master mix was used for real time PCR, with 1.26 pmoles of each primer and template cDNA described above. Each reaction was performed in triplicate in a 96-well plate, and a no-reverse transcriptase control was used to determine whether genomic DNA contamination was present. The reactions were heated to 50 °C for 2 minutes, 95 °C for 10 minutes, then 40 cycles of 95 °C for 15 seconds and 60 °C for one minute were performed. A dissociation curve was performed for each 96 well plate. The real time PCRs were performed on an ABI 7000 sequence detection system (Applied Biosciences). Results were initially analysed using the ABI prism 7000 SDS program to determine CT values. These were then exported and analysed using Microsoft excel. Amplification of serial dilutions of cDNA was used to establish a standard curve, from which relative amounts of each transcript present in the samples were calculated. The amount of repeat transcript relative to the amount of *GAL4* transcript in each sample was then calculated, to take cell death into account (which could result in fewer cells expressing the repeat transcripts, but should also affect *GAL4* transcript levels to the same extent). These were then expressed for each sample relative to an arbitrarily chosen sample, which was set at 1.

Horizontal sectioning and antibody staining

Heads were dissected from flies aged 0 – 24 hours. These were positioned in optimal cutting temperature medium (OCT), frozen on dry ice and 10 µm horizontal sections cut using a Leitz Kryostat 1720 with the chamber temperature at –30 °C and the block at –15 °C. The sections were collected on a poly-lysine slide and fixed in 4% paraformaldehyde for 10 minutes. They were then washed in PBS three times, immersed in cytoskeletal buffer for 5 minutes, and washed in with 0.01% saponin in PBS three times. Primary antibody at the appropriate concentration was added and the slides were incubated overnight at 4 °C. They were then washed in 0.01% saponin, and secondary antibody was added. This was incubated at room temperature for 2 hours, washed in saponin three times, and Hoechst 33258 (Sigma-Aldrich) was used at 10 µg/ml to stain the DNA. The sections were then washed three times in PBS and mounted in 80% glycerol in PBS.

Tangential sectioning

Adult eyes were dissected, the proboscis removed and the heads cut in half. These were fixed overnight in 2.5% gluteraldehyde, 0.1 M sodium phosphate (pH7.2), post-fixed in 2% osmium tetroxide, washed in water, and dehydrated in acetone. Specimens were mounted in epoxy resin, sectioned at 1 µm using an RMC Mt7 ultramicrotome, mounted onto slides and stained with methylene blue. Tangential sectioning was performed by H. Rodgers.

Microscopy

Scanning electron micrographs:

Whole adult flies were dehydrated progressively through an acetone series (25%, 50%, 75%, 100%), dried at room temperature, and viewed with a field emission scanning electron microscope (Phillips, at CEMMSA, Adelaide University). Digital images were collected in Adobe Photoshop 6.0.

Light and fluorescence microscopy:

Light microscopy was performed on an Olympus SZH10 light microscope with a Polaroid digital microscope camera attached. Images were collected digitally using Adobe Photoshop 6.0. Fluorescence microscopy was performed on an Olympus AX70. Digital images were collected with a cooled CCD camera and V++ software. Photoshop 6.0 was used for image preparation. Anterior is to the left in all eye photos.

Introduction

Use of the Drosophila model system

The *Drosophila* eye was used to investigate the hypothesis that RNA containing expanded CAG repeats is responsible for, or contributes to, polyglutamine pathogenesis. For the purposes of this study, gene context was not being investigated; rather the aim was to determine the intrinsic toxicity of the CAG repeat. Therefore, constructs encoding peptides consisting of polyglutamine tracts alone were designed to be expressed in the *Drosophila* eye. These were based on those previously described by Marsh *et al*. [84], where a polyglutamine tract is located within a small coding region, flanked by short amino acid sequences (six on the N-terminal and four on the C-terminal side) and a myc/flag epitope tag of 26 amino acids at the C-terminus (Figure 3.1 A).

Figure 3.1 Schematic representation of repeat constructs. (**A**) All repeat constructs were inserted downstream of UAS sites to allow expression under the control of the driver *GMR-GAL4*. The repeats were flanked by six amino acids on the N-terminal side and four on the C-terminal, and a myc/flag epitope tag was placed downstream. (**B**) Constructs containing polyglutamines encoded by $(CAG)_{30}$, $(CAG)_{52}$ and $(CAG)_{99}$ were generated. (**C**) These were compared to constructs containing polyglutamine encoded by $(CAA)_{20}$ or $(CAA)_{94}$. (**D**) A variation of this was generated with a termination codon inserted immediately upstream of $(CAG)_{93}$, effectively moving the repeat into the 3' untranslated region. This is referred to as term$(CAG)_{93}$. (**E**) The reading frame of the repeat was altered to the +2 frame, changing the CAG repeat into a GCA repeat encoding polyalanine. This was accomplished by inserting one base pair before the repeat and deleting one following the repeat, resulting in a minor alteration to the encoded protein sequence. (**F**) $(CTG)_{86}$, encoding polyleucine, was also generated in the original construct.

Approach: comparison of CAG and CAA

In all of the human diseases caused by polyglutamine expansion identified to date, the expanded repeat is encoded by CAG. CAA also encodes glutamine; however it is unknown whether a CAA repeat expansion can induce pathogenesis, or whether there is a sequence requirement for the repeat to be involved in pathogenesis. One way in which the sequence of the repeat could be involved is via the RNA hypothesis. A feature of the RNA hypothesis is that it is supported by data showing that in all disease-inducing repeat expansions, the mRNA transcripts carrying the repeats are able to form hairpin secondary structures. In DM, this propensity to form a hairpin is related to the toxicity of the expanded CUG repeat, as is its retention in the nucleus where it forms foci. Therefore, to determine whether the ability of the RNA to form a hairpin, or some other sequence-specific attribute of CAG, contributes to pathogenesis in the polyglutamine diseases, the effect of expressing expanded CAA or CAG repeats was compared. As both CAG and CAA encode glutamine, the constructs encode proteins that are identical at the protein level, but the RNA differs in its ability to form a hairpin. Whilst it has been shown that rCAG repeats form hairpins *in vitro* and nuclear foci *in vivo*, CAA repeats are not predicted to form such a structure as the requisite C-G base pairs are lacking. To confirm this, the transcripts encoded by the constructs were analysed using Mfold, a software application that predicts the secondary structure of single stranded nucleic acids [191,215]. This supported the predicted hairpin formation of the CAG repeat, but demonstrated no such secondary structure for the CAA repeat (Figure 3.2). Thus comparison of the effect of expressing these different repeats in the *Drosophila* eye should allow dissection of the contribution made by RNA, and its ability to form a hairpin secondary structure, to the neurodegenerative phenotype. If hairpin-structured RNA is entirely responsible for polyglutamine pathogenesis, then expression of a CAA repeat should not cause a phenotype in the eye. If polyglutamine pathogenesis is partially mediated by RNA, then the CAA phenotype should be milder than the CAG phenotype, since the RNA component would be lacking.

Results

Generation of flies expressing expanded repeats

Repeat constructs were generated by annealing complementary single stranded oligonucleotides to form double stranded 30mers consisting of 10 CAA or CAG

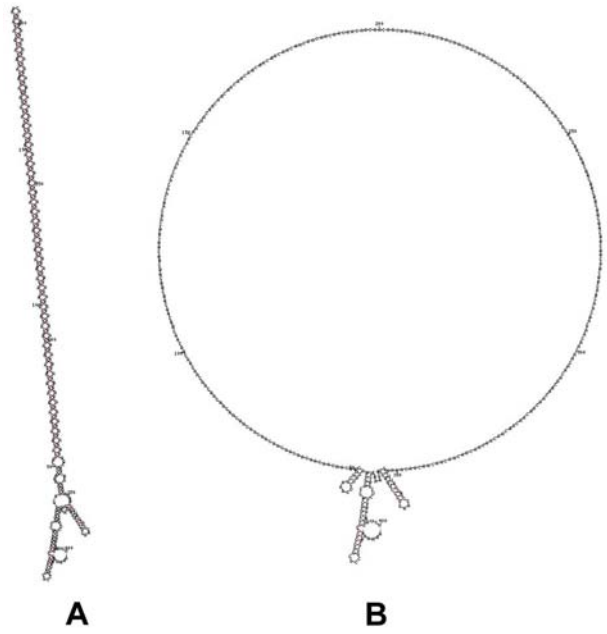


Figure 3.2 Secondary structure predictions for CAG and CAA repeats. The mRNA sequence encoded by the $(CAG)_{99}$ and $(CAA)_{94}$ repeats flanked by 50 base pairs of surrounding sequence on either side was analysed using the Mfold program. The structures represented here are the optimal predictions. (**A**) CAG repeats form a hairpin structure with complementary base pairing between the C-G pairs. (**B**) No such hairpin is formed by the CAA repeats, and indeed no structure is predicted, leading to the repeat tract being represented here as a circle.

repeats, then performing blunt ended ligation to increase their length. The longest copy numbers successfully generated using this method were 30 repeats; therefore these repeat tracts were expanded using a PCR-based methods [213,214]. They were then subcloned into the *Drosophila* transformation vector pUAST [216], which had been modified to contain the amino acid sequence surrounding the repeat, including the myc and flag epitope tags, that has been described previously [84]. These were then used to generated transgenic flies by microinjection and P-element mediated transformation [217]. A range of copy numbers above and below the pathogenic threshold for both CAG and CAA repeats was analysed (Figure 3.1 B and C); for each construct, multiple independent transgenic lines were generated. They were analysed by crossing to the driver *GMR-GAL4*, which drives expression of the repeat construct in all cells of the eye following differentiation [218].

CAG and CAA have the same effect on the Drosophila eye

Expression of GAL4 driven by GMR causes a very mild rough eye due to induction of apoptosis in the developing eye [219]; therefore flies of the genotype *GMR-GAL4/+* were used as a control. Flies expressing $(CAG)_{30}$ or $(CAA)_{20}$, repeat copy numbers below the pathogenic threshold, had the same appearance as that of control flies expressing only GAL4 (Figure 3.3 A, B and E). This is in accordance with what has been previously reported for CAG repeats in this range, with the exception of ataxin-1 containing $(CAG)_{30}$, which can cause a similar phenotype to that caused by repeat expansion when expressed at a high level in either the *Drosophila* eye or a mouse model of SCA1 [64]. This effect may be due to the effect of the *ataxin-1* transcript or

protein on the eye, rather than the repeat itself, as similar findings have not been described for any other polyglutamine disease models. The lack of phenotype reported here in any of the 13 independently generated transgenic lines containing $(CAG)_{29}$, despite the fact that some are likely to be expressing the repeat construct at higher levels than others, supports this conclusion.

Expression of the $(CAG)_{52}$ or $(CAG)_{99}$ repeat constructs, which are above the pathogenic threshold, induced degeneration in the *Drosophila* eye (Figure 3.3 C and 3.5). A range of phenotypes was observed upon expression of these constructs (Figure 3.4). Presumably, this is due to positional effects, where varying levels of expression between independent lines are caused by differences in the site of integration [220]. For $(CAG)_{99}$, this phenotype differed in independently generated lines ranging from no phenotype to lethality, with different severities of eye phenotype in between. Degeneration of cells of the eye occurred, causing the structure to collapse, and pigment loss was also observed. The phenotypes observed upon expression of $(CAG)_{52}$ were similar but less severe. For the $(CAG)_{52}$ phenotypes and the $(CAG)_{99}$ phenotypes at the milder end of the spectrum, the only obvious defect

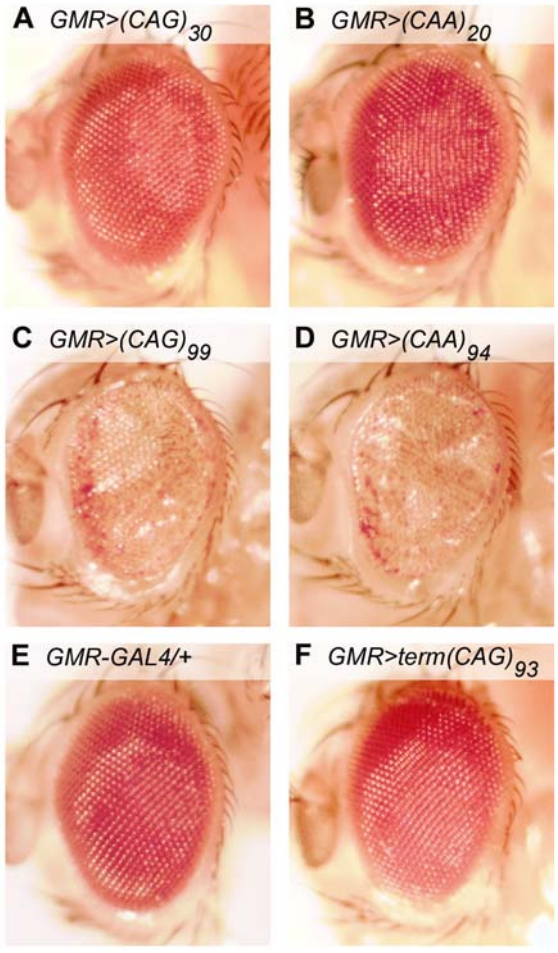


Figure 3.3 Effect of expressing various repeat constructs on the external appearance of the *Drosophila* eye. All repeats were downstream of UAS sites, and expression in the eye was driven by *GMR-GAL4*. (**A**) No alteration from the wild-type eye phenotype is observed in flies expressing polyglutamine encoded by $(CAG)_{30}$. (**B**) Expression of polyglutamine encoded by $(CAA)_{20}$ also has no effect on the external appearance of the eye. (**C**) Expanded polyglutamine encoded by $(CAG)_{99}$ causes degeneration of the eye, seen here as a lack of pigmentation and collapsed eye structure. (**D**) Expression of expanded polyglutamine encoded by $(CAA)_{94}$ repeats also causes degeneration, with a phenotype indistinguishable from that caused by $(CAG)_{99}$. (**E**) Control fly, expressing only GAL4 in the eye. (**F**) Expression of an untranslated CAG repeat encoded by term$(CAG)_{93}$ has no effect on the appearance of the eye.

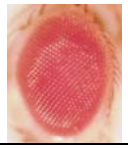
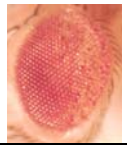
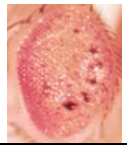
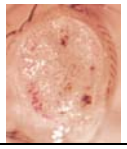
Eye Phenotype:	None 	Mild 	Moder. 	Severe 	Lethal	Total
$(CAG)_{29}$	13	0	0	0	0	13
$(CAA)_{20}$	1	0	0	0	0	1
$(CAG)_{52}$	7	3	0	0	0	10
$(CAG)_{99}$	2*	5	1*	2	1	11
$(CAA)_{94}$	1*	2	3	6	1	13
term$(CAG)_{93}$	16*	0	0	0	0	16

Figure 3.4 Effect of expressing expanded repeats in the *Drosophila* eye driven by *GMR-GAL4* varies between lines. For each repeat construct, the total number of independently generated lines is indicated, as well as the number of lines showing no eye phenotype, a mild, moderate or severe phenotype, or lethality, upon expression with *GMR-GAL4*. Examples of these phenotypic categories are shown, which do not correspond to any of the particular lines. *Lines analysed by quantitative RT-PCR.

was a loss of pigment in the posterior part of the eye; this was not associated with any disorganisation to the external appearance of the ommatidia as revealed by scanning electron microscope (SEM) images (Figure 3.5). However, tangential sections of these eyes revealed significant disorganisation compared to the *GMR-GAL4/+* control, with the normal regular trapezoidal arrangement of seven visible rhabdomeres reduced to only a few, and the ordered structure disrupted (Figure 3.5 E, F). For $(CAG)_{99}$ at the more severe end of the spectrum, the eye lost all pigment and collapsed completely; small dark patches of cells were frequently observed, which may be necrotic or dying cells. Lethality at the late pupal stage was observed in the most severe case; this is likely to be due to expression in tissues other than the eye, which have previously been reported using *GMR* [84]. The observed phenotypes appear to be in accordance with those previously reported for a similar construct containing $(CAG)_{108}$ by Marsh *et al.* Previous characterisation of this phenotype revealed it to be cell type-specific and degenerative [84].

Expression of the $(CAA)_{94}$ repeat construct had an indistinguishable effect from $(CAG)_{99}$. Again, eye phenotypes ranging from none to lethality were observed, with phenotypes in between characterised by degeneration of cells leading to collapse of the eye and pigment loss (Figure 3.3 D). Furthermore, the observed range of phenotypes was the same as that seen for the CAG repeat construct (Figure 3.4). No

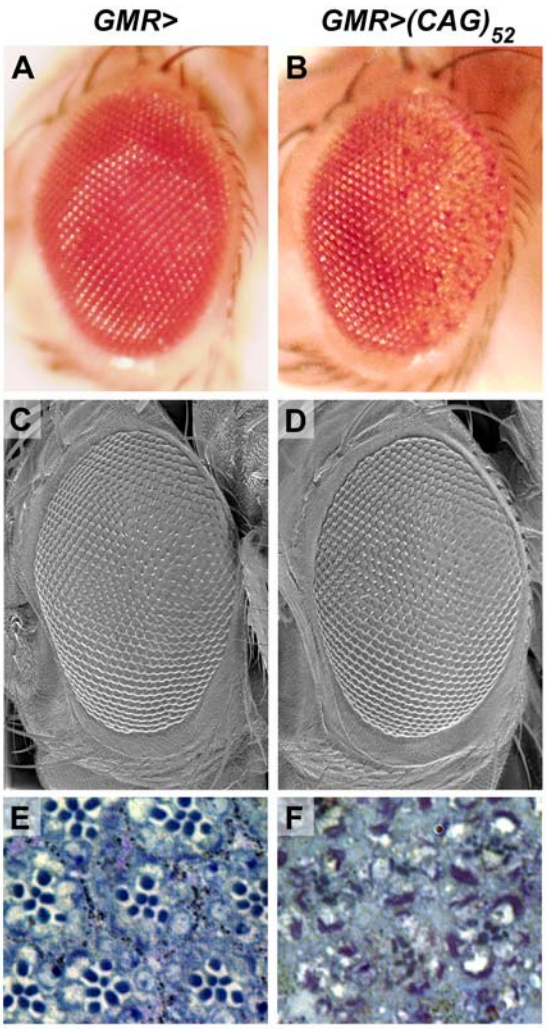


Figure 3.5 Expression of $(CAG)_{52}$ in the *Drosophila* eye causes loss of pigment and photoreceptor neurons. The CAG repeat was downstream of UAS sites, and expression in the eye was driven by *GMR-GAL4*. (**A**) Appearance of the eye of control flies, carrying *GMR-GAL4* alone, under the light microscope. (**B**) Expression of the $(CAG)_{52}$ repeat causes loss of pigment in the posterior half of the eye. (**C**) Scanning electron micrograph of the control fly reveals only a very mild disruption to the ommatidia. (**D**) At the external level, expression of $(CAG)_{52}$ does not disrupt this any further. (**E**) Tangential sections stained with methylene blue show the pattern of the control eye, with 7 visible photoreceptor neurons forming a hexagonal shape, and pigment granules surrounding each ommatidium. (**F**) This order is severely disrupted in the fly expressing $(CAG)_{52}$, with many photoreceptor neurons lost or smaller and pigment granules not visible.

increase in severity was seen for CAG compared to CAA as would be predicted by the RNA hypothesis; instead both types of repeat appear to have the same effect and exhibit equal toxicity in the *Drosophila* eye.

Expanded CAG and CAA-induced pathology involves protein aggregation

The observation that the expression of expanded CAA and CAG repeats in the *Drosophila* eye has the same effect suggests that pathogenesis is independent of RNA sequence and secondary structure and that it is indeed polyglutamine-containing protein which is the pathogenic agent. To determine the distribution of the polyglutamine protein in the eyes of these flies, horizontal sections of the head were stained with a 9E10 myc antibody, which recognises the myc epitope tag on the C-terminus of the polyglutamine protein. In flies expressing polyglutamine encoded by either CAG or CAA repeats that are below the pathogenic threshold, the polyglutamine protein appeared to be diffusely distributed throughout the eye. However, when the repeats are expanded to around 100 copies, the protein forms aggregates in the eye (Figure 3.6), a result that has been reported previously for

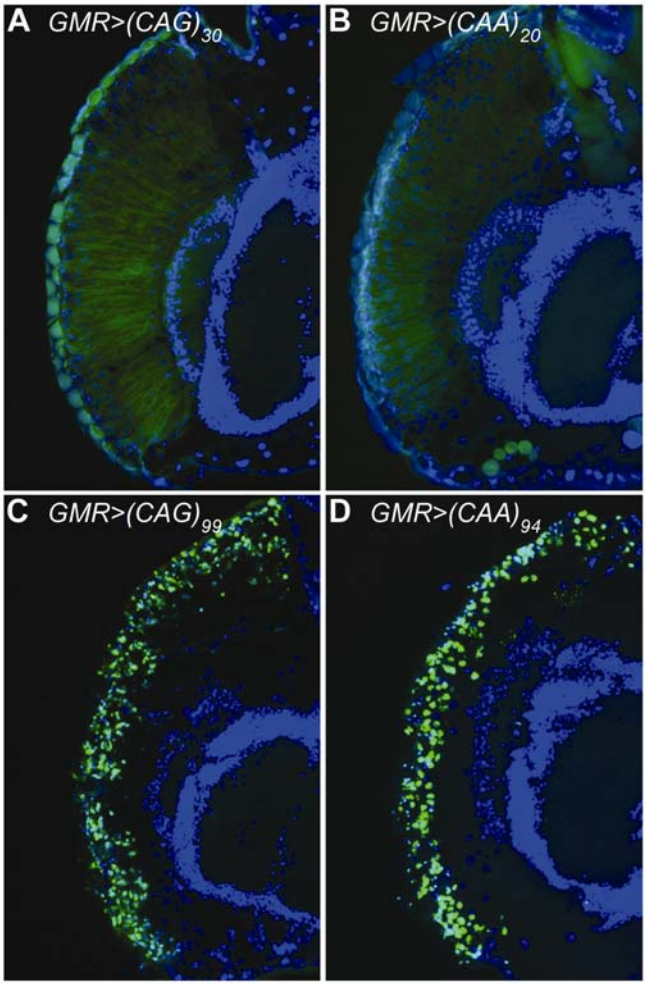


Figure 3.6 Distribution of polyglutamine proteins in the *Drosophila* eye. Expression of the UAS repeat constructs was driven by *GMR-GAL4.* Horizontal cryosections were stained with Hoechst (blue) to show DNA and myc antibody (green), which recognizes the myc epitope tag on the C-terminus of the polyglutamine protein. (**A**) Short polyglutamine proteins encoded by $(CAG)_{30}$ show a diffuse distribution. (**B**) short polyglutamine proteins encoded by $(CAA)_{20}$ are also diffusely distributed. (**C**) Polyglutamine proteins above the pathogenic threshold encoded by $(CAG)_{99}$ form tight aggregates. (**D**) This is also seen when polyglutamine is encoded by $(CAA)_{94}$.

polyglutamine encoded by CAG repeats [86]. Such aggregation was observed for expanded polyglutamine proteins encoded by both CAG repeats and CAA repeats. This suggests that the behaviour of the encoded proteins is a major determinant of the phenotypes observed in this *Drosophila* model of the CAG repeat expansion diseases.

Untranslated CAG repeats do not induce degeneration

Whilst the results obtained by comparing CAA and CAG repeats suggest that polyglutamine pathogenesis is not mediated by RNA, a contribution by both rCAA and rCAG to their respective phenotypes cannot be ruled out. It is possible that the pathogenesis observed in these cases involves multiple pathways and that both RNA and the encoded polyglutamine contribute to pathogenesis via independent pathways. To examine this possibility, an alternative version of the CAG repeat construct was generated, in which a termination codon is inserted upstream of the repeats. This has the effect of moving the CAG repeats into the 3'-untranslated region of the transcript so that it now encodes only a short peptide (Figure 3.1 D). Expression of this construct allowed the possibility of a contribution by RNA to be

examined, since the RNA transcript still contains the expanded repeat, but it no longer encodes a polyglutamine tract. Therefore if RNA can induce pathogenesis, a phenotype in the *Drosophila* eye would be expected. Multiple independent transgenic lines were generated carrying this construct, and crossed to the driver *GMR-GAL4*. However, no phenotype was observed when this construct was expressed in the eye, with the resulting flies having an eye phenotype indistinguishable from flies expressing GAL4 alone (Figure 3.2). These observations support the conclusion that RNA is not the pathogenic agent in the polyglutamine disorders.

The untranslated CAG repeat transcript is expressed

As expression of term$(CAG)_{93}$ did not induce a phenotype in the *Drosophila* eye, it was necessary to investigate whether the transcript is expressed and present in the eye. In particular, the introduced termination codon could alter the stability of the mRNA or cause the transcript to be recognised as nonsense and degraded via the nonsense mediated decay mechanism [221]. Thus the lack of phenotype may simply be due to the fact that the transcript is not present at a sufficient level. To address this, quantitative (real time) RT-PCR was performed on whole adult heads to compare the level of repeat transcript in two lines expressing the untranslated term$(CAG)_{93}$ transcript in the eye and two lines expressing the translated $(CAG)_{99}$ repeat. These were normalised to the level of the *GAL4* transcript, which should be constant in all lines. The rationale behind the use of the *GAL4* transcript was to avoid inaccuracy due to the death of cells expressing the repeat transcript. This could lead to an under estimation of the repeat transcript level in these lines if a housekeeping gene expressed in all cells was used, as a greater proportion of those cells not expressing the repeat transcript would remain.

Of the two translated $(CAG)_{99}$ lines analysed, one caused a moderate phenotype when expressed in the eye and one caused no phenotype, presumably owing to differences in expression level of the transgene caused by variation in the site of integration. Indeed, the quantitative RT-PCR demonstrated that the level of the $(CAG)_{99}$ transcript was higher in the line showing a phenotype, whereas the line showing no phenotype demonstrated a lower level of expression. In comparison, the two lines expressing the untranslated term$(CAG)_{93}$ transcript did so at a higher level than the translated $(CAG)_{99}$ lines, including the one showing a phenotype (Figure 3.7). This suggests that the introduction of the termination codon is not causing the transcript to be degraded, and that it is present at levels comparable to those

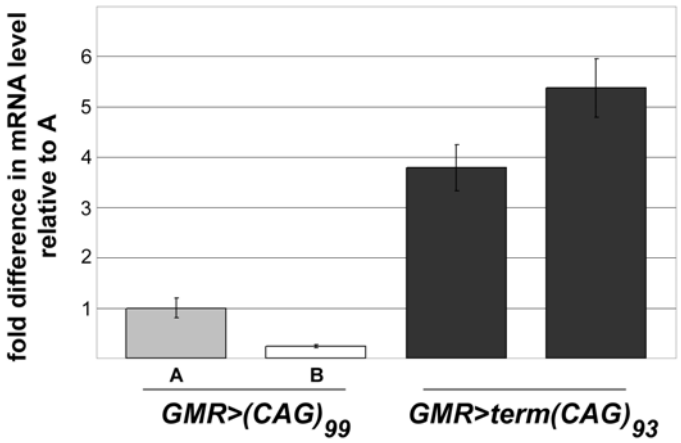


Figure 3.7 Expression levels of translated and untranslated CAG repeat transcripts. *GMR-GAL4* was used to drive expression of UAS repeat constructs. Quantitative (real time) PCR was performed to determine levels of repeat transcripts in the following flies: $(CAG)_{99}$ showing a phenotype (A) in light grey, $(CAG)_{99}$ showing no phenotype (B) in white and two independent lines of term$(CAG)_{93}$ shown in dark grey. The expression level of the repeat transcripts were normalized to the level of the *GAL4* transcript, and the expression level of each line is expressed as the fold difference relative to the $(CAG)_{99}$ line that shows a phenotype (A), which was arbitrarily set at one.

sufficient to induce a phenotype when translated. The lack of phenotype observed in these lines can therefore be interpreted to mean that untranslated CAG repeats are not pathogenic in the *Drosophila* eye.

Evidence against the polyalanine hypothesis

In addition to addressing the RNA hypothesis, the translated CAG/CAA experiments described here also address the question of whether frameshifting to the polyalanine frame contributes to polyglutamine pathogenesis. The CAA repeat construct cannot encode alanine in any alternative frames. Therefore any contribution to pathogenesis by the presence of alanine peptides in the CAG-expressing flies would be lacking in the CAA-expressing flies and a milder phenotype would be expected in these flies. However, this was not observed, suggesting that even if frameshifting is occurring in the case of CAG, it is not making a significant contribution to pathogenesis.

To further examine this issue, an attempt was made to generate lines of flies expressing polyalanine instead of polyglutamine, with the aim of determining whether polyalanine is more toxic than polyglutamine in the *Drosophila* system. To achieve this, the repeat construct was altered by inserting 1 base directly upstream of the CAG repeat tract and deleting 1 base directly downstream (Figure 3.1 E). This shifted the reading frame of the repeat into the GCA frame, encoding polyalanine. The construct was generated in pUAST and sequencing confirmed that it contained 99

copies of the GCA repeat. However, difficulty was encountered generating transgenic *Drosophila* containing this repeat sequence; despite repeated rounds of microinjection that resulted in the generation of 23 independent lines, none were obtained in which the $(GCA)_{99}$ repeat length was maintained. The longest GCA repeat detected contained 56 copies of the repeat, and a total of six lines with a copy number above 50 were used for analysis. The insertions in these lines were analysed by sequencing, which revealed that the repeats had contracted precisely, leaving surrounding sequences intact and the reading frame unaltered.

Expression of these polyalanine proteins in the *Drosophila* eye driven by *GMR-GAL4* did not cause a phenotype, with the resulting eyes having an appearance indistinguishable from flies expressing GAL4 alone (Figure 3.8). However, the difficulty experienced in generating lines of longer repeat numbers may suggest that expression of the GCA repeat is toxic to *Drosophila*, and thus the lines obtained may be those in which the repeat is not expressed, or below the pathogenic threshold in this system. To address this possibility, quantitative RT-PCR was used to determine the repeat transcript level in three of the lines with a copy number above 50. This demonstrated that the transcripts are being expressed, and at a higher level than the $(CAG)_{99}$ polyglutamine line showing a phenotype (Figure 3.9). Therefore they have presumably not inserted into heterochromatic regions of the genome.

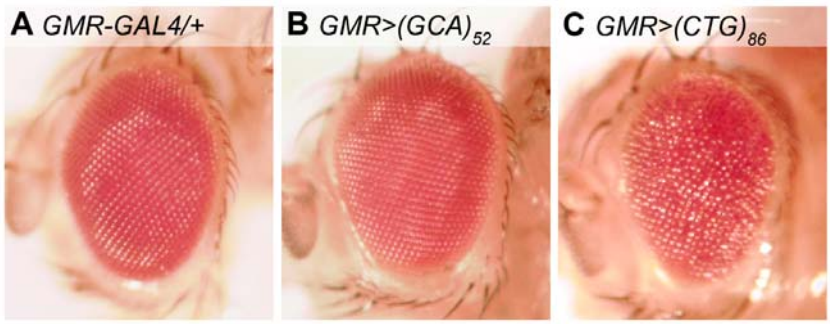


Figure 3.8 Effect of GCA and CTG repeat constructs on the external appearance of the *Drosophila* eye. All repeats were downstream of UAS sites, and expression in the eye was driven by *GMR-GAL4*. (**A**) Control fly, expressing GAL4 alone in the eye. (**B**) Expression of a GCA repeat encoding polyalanine has no effect on the eye, with the external appearance indistinguishable from the control. (**C**) Expression of a CTG repeat encoding polyleucine induces a rough eye phenotype due to disorganisation of the ommatidia.

To analyse expression of the polyalanine repeat at the protein level, western analysis using an antibody directed against myc was performed. This should detect the myc

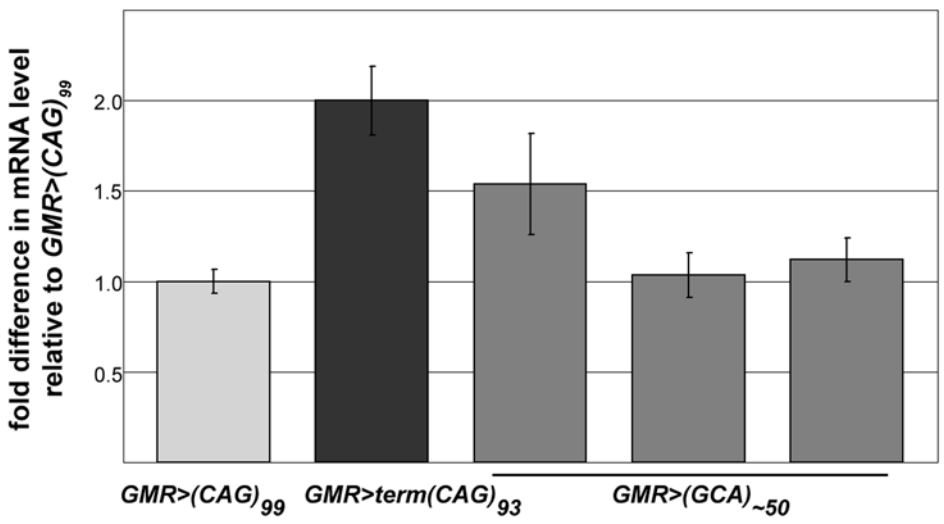


Figure 3.9 The GCA repeat transcript is expressed at similar levels to the translated and untranslated CAG repeat transcripts. *GMR-GAL4* was used to drive expression of UAS repeat constructs. Quantitative (real time) PCR was performed to determine levels of repeat transcripts in the following flies: $(CAG)_{99}$ showing a phenotype in light grey, term$(CAG)_{93}$ showing no phenotype in dark grey, and three independently generated lines of term$(GCA)_{\sim 50}$ shown in mid grey. The expression level of the repeat transcripts were normalized to the level of the *GAL4* transcript, and the expression level of each line is expressed as the fold difference relative to the level of the $(CAG)_{99}$ transcript, which was arbitrarily set at one.

epitope tag at the C terminus of the polyalanine protein, just as it detected myc-tagged polyglutamine proteins when used for immunofluorescent staining of eye sections. To perform the western analysis, eye discs were collected from third instar larvae expressing GAL4 alone (negative control), polyglutamine proteins containing repeat tracts of 99, 52 and 49 copies (positive controls), and the three polyalanine lines analysed by quantitative PCR. The eye discs were run on a polyacrylamide gel that was then probed with the myc antibody. Bands above background were detected in the lanes containing the polyglutamine proteins. They appeared to run higher than would be predicted based on their molecular weight – this is presumably due to some property/structure of the polyglutamine tract, and has previously been reported for polyglutamine proteins [222]. However, no significant bands were detected in the lanes containing the putative polyalanine proteins; only faint background bands were detected that are also present in the GAL4 control lane (Figure 3.10). This suggests that the myc antibody is unable to detect the polyalanine protein (perhaps due to some secondary structure formed by the repeat) or alternatively that the polyalanine proteins are not being expressed.

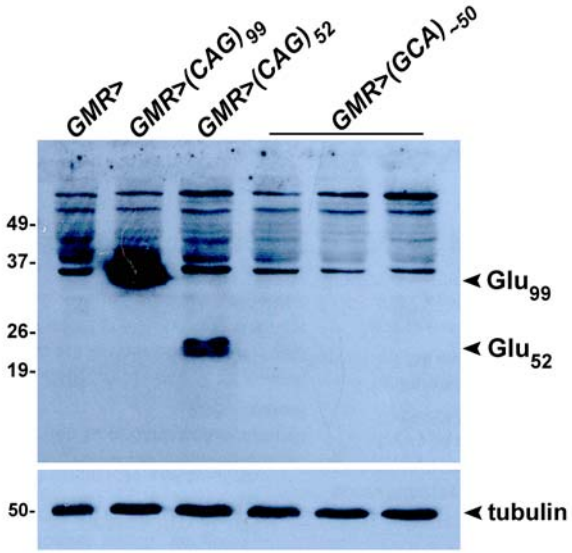


Figure 3.10 Polyalanine proteins cannot be detected by western blot. *GMR-GAL4* was used to drive expression of UAS repeat constructs. Eye discs from third instar larvae were run on a 15% polyacrylamide gel, which was transferred to nitrocellulose and probed with an antibody against the C-terminal myc tag present on each repeat protein (top panel). Lane one shows control larvae expressing GAL4 alone. Lane two shows flies expressing a protein containing a 99 glutamine tract (Glu_{99}), which is detected by the myc antibody at approximately 33 kDa. Lane 3 shows a smaller protein containing 52 glutamines (Glu_{52}), again detected by the myc antibody and running at approximately 22 kDa. The final three lanes contain samples from flies expressing GCA repeat tracts; no polyalanine proteins are visible in these lanes. The membrane was also probed with anti α-tubulin as a loading control (bottom panel).

Expression of a CTG repeat causes a phenotype in the Drosophila eye

It has been suggested that it is the hydrophobicity of alanine and its propensity to aggregate than makes polyalanine more toxic than polyglutamine [131]. Leucine is a similar amino acid to alanine, as it also has a non-polar side chain. Therefore it was proposed that polyleucine, like polyalanine, is more toxic than polyglutamine, and this has been demonstrated in cultured cells [131]. Polyleucine expansions may be the cause of HDL2; however, it has not yet been determined whether the CTG expansion in this disorder encodes leucine, alanine or is non-coding. To investigate the toxicity of polyleucine proteins in *Drosophila*, a construct containing a $(CTG)_{86}$ repeat was created by excising the CAG repeat and ligating it back into the construct to generate a clone containing the repeat sequence in the opposite orientation (Figure 3.1F). Unlike the polyalanine construct, no difficulty was encountered generating transgenic *Drosophila* carrying the $(CTG)_{86}$ repeat. Expression in the eye using *GMR-GAL4* caused a rough eye phenotype in 5 of the 13 independent lines generated. The eye

phenotype was distinct from the polyglutamine phenotype, showing no loss of pigment or collapse but rather a disorganisation of the ordered structure of the eye (Figure 3.8).

The presence of a phenotype in these flies suggests that the polyleucine protein is being expressed; however it is possible that the phenotype is due to expression of RNA containing the CUG repeat, rather than the polyleucine protein, as CTG expansions in non-coding regions are the cause of DM and SCA8. Western analysis was used to investigate whether polyleucine protein is present in the eyes of flies expressing the $(CTG)_{86}$ construct. As was the case for the putative alanine proteins, no bands were detected corresponding to the polyleucine protein (Figure 3.11). The bottom of the stacking gel was retained, included in the transfer and probed in case the formation of large aggregates by the polyleucine proteins prevented them from entering the resolving gel; however no protein was detected in this region.

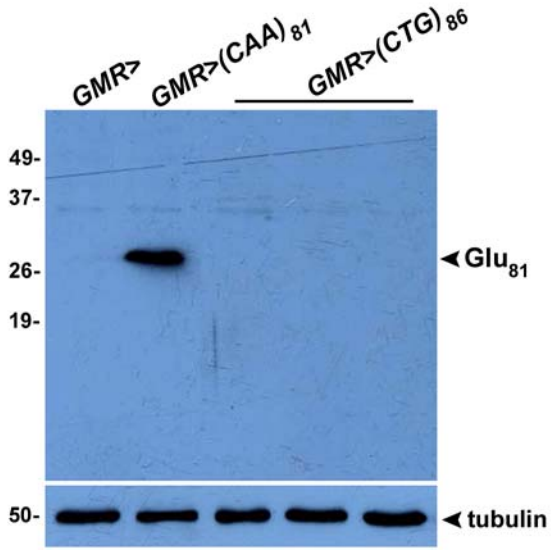


Figure 3.11 Polyleucine proteins cannot be detected by western blot. *GMR-GAL4* was used to drive expression of UAS repeat constructs. Eye discs from third instar larvae were run on a 12% polyacrylamide gel, which was transferred to nitrocellulose and probed with an antibody against the C-terminal myc tag present on each repeat protein (top panel). Lane one: control larvae expressing GAL4 alone. Lane two: larvae expressing an 81-repeat polyglutamine tract, from a line generated from the $(CAA)_{99}$ construct which underwent a spontaneous deletion, contracting the repeat cleanly to 81 copies. The 81-copy polyglutamine protein is detected by the myc antibody at approximately 30 kDa. Lanes three to five: larvae from independently generated lines expressing a $(CTG)_{86}$ repeat tract; no polyleucine proteins are detected in these lanes. The membrane was also probed with anti α-tubulin as a loading control (bottom panel).

Polyleucine proteins are expressed the Drosophila eye

The lack of detection of polyalanine and polyleucine proteins using western analysis may be due to their propensity to form aggregates, which prevents them from being extracted from the eye discs and entering the gel. Therefore, horizontal sections of the head of flies expressing GCA and CTG repeats were stained with the myc antibody, in an attempt to detect the proteins using immunofluorescence. In the control fly expressing GAL4 alone, staining with the anti-myc antibody resulted in a ubiquitous pattern of staining, which may be due to background or the endogenous myc protein (Figure 3.12). In flies expressing the $(GCA)_{52}$ repeat, no polyalanine aggregates were visible in sections of the eye. Diffuse staining with the anti-myc antibody was observed; however this was similar to that seen in control flies, suggesting it could be background rather than the polyalanine protein. There is conflicting data on the propensity of polyalanine proteins to form aggregates: two separate studies of the behaviour of polyalanine proteins in cultured cells have reported that a repeat tract consisting of 19 alanines is sufficient to induce aggregate formation [223,224], yet a third study found no aggregate formation by a pure alanine tract consisting of 29 copies, with the protein remaining diffuse in the cytoplasm [225]. Thus it is not clear whether the polyalanine proteins reported here would form aggregates, and a low level of cytoplasmic expression cannot be ruled out in the eye sections stained with the myc antibody. Therefore no conclusions regarding the toxicity of polyalanine can be drawn from these flies, as it is unknown whether or not the protein is being expressed.

In contrast, in flies expressing the CTG repeat, tiny aggregates or speckles were visible throughout the eye upon anti-myc staining (Figure 3.12 J and K, arrows point out two examples). These were not seen in the control or the polyalanine flies, suggesting that they correspond to the polyleucine protein. In support of this conclusion, polyleucine proteins also form small aggregates described as speckles when expressed in cultured cells [131]. Therefore, despite the lack of detectable protein on the western blot, it appears that polyleucine proteins are being expressed in the *Drosophila* eye; these experiments do not demonstrate whether it is this polyleucine protein or the CUG-containing RNA that is responsible for the phenotype observed.

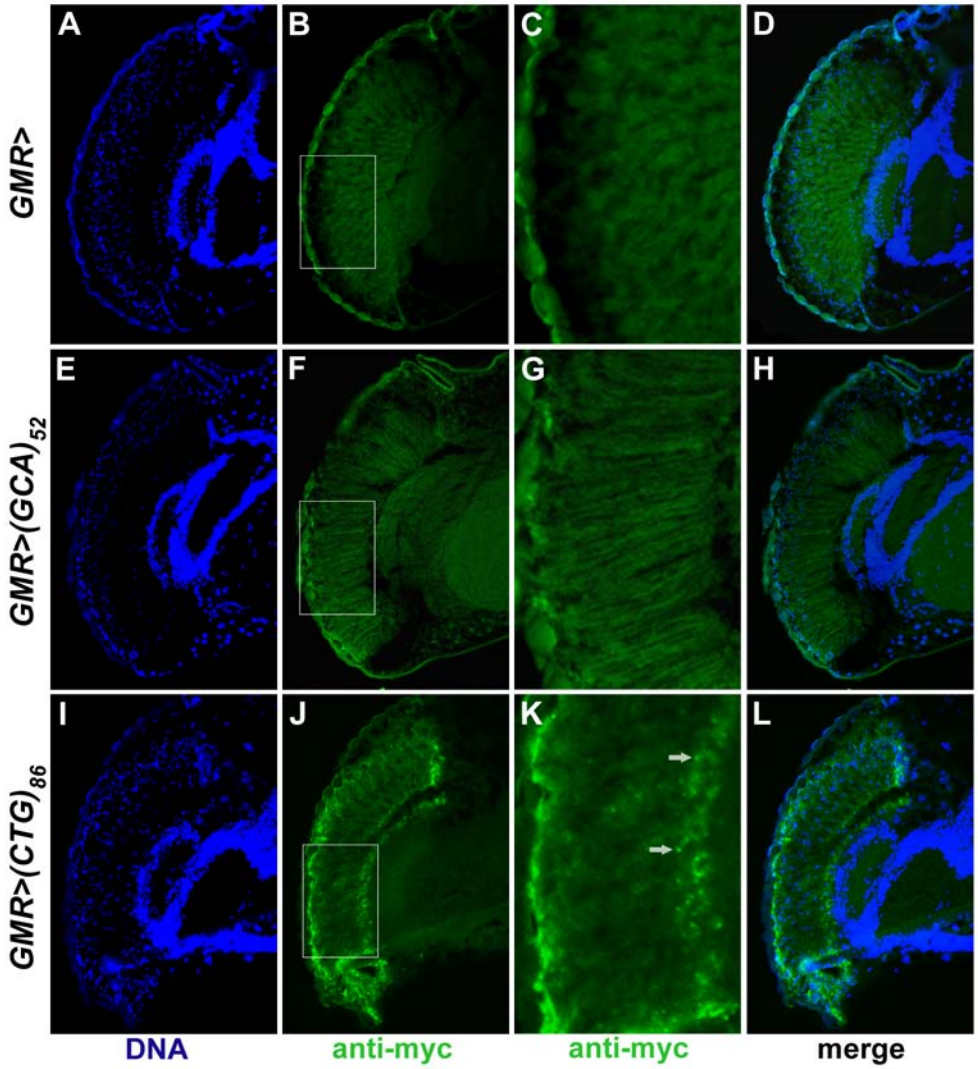


Figure 3.12 Detection of polyalanine and polyleucine proteins in the *Drosophila* eye. Expression of the UAS repeat constructs was driven by *GMR-GAL4.* Horizontal cryosections were stained with Hoechst (**A**,**E**,**I**) to show DNA and myc antibody (**B**,**F**,**J** with boxed regions enlarged in **C**,**G**,**K**), which recognizes the myc epitope tag on the C-terminus of the proteins. (**A** –**D**) Control fly expressing only GAL4 in the eye, to determine the level of background staining. (**E**-**H**) Fly expressing a $(GCA)_{52}$ repeat. No polyalanine protein is detected, with the staining pattern resembling that of the control fly. (**I**-**L**) Fly expressing polyleucine protein in the eye, encoded by $(CTG)_{86}$. The staining pattern is distinct from that of the control, with small speckles observable (**K** arrows), suggesting that polyleucine protein is being detected.

Discussion

No evidence supporting RNA as the pathogenic agent

The primary aim of this work was to investigate the possibility that RNA may be the pathogenic agent, or contribute to pathogenesis, in the polyglutamine diseases. The results obtained provide evidence against this hypothesis, showing that CAA repeats cause a phenotype in the *Drosophila* eye that is indistinguishable to that caused by

CAG repeats, and is no less severe. Furthermore, an untranslated CAG repeat does not induce toxicity, suggesting that RNA is not contributing to the translated CAG phenotype. The major outcomes of this investigation into RNA pathogenesis have been published [226].

In addition to the results of the CAG/CAA experiment, further evidence against the RNA hypothesis comes from analysis of the effect of expressing the CAG repeat in the +2 reading frame (i.e. a GCA repeat) in the *Drosophila* eye. At the RNA level, this transcript is almost identical to that containing the CAG repeat and the propensity of the repeat to form a hairpin structure should still be the same. The only difference between the transcripts containing the $(CAG)_{52}$ repeat and $(GCA)_{\sim 50}$ repeat is the presence of an extra base before the repeat and the lack of one base after the repeat in the GCA transcript. However, an eye phenotype was observed upon expression of $(CAG)_{52}$ but not $(GCA)_{\sim 50}$. RT-PCR demonstrated that the GCA transcript is present, so lack of RNA expression is not the cause of the lack of phenotype. This observation again suggests that the CAG phenotype is not RNA-mediated, as if it were, the same phenotype would be expected upon expression of GCA.

Investigation of polyalanine toxicity

The results obtained here also provide indirect evidence against a contribution by polyalanine to the polyglutamine phenotype. However, further attempts to analyse whether polyalanine is toxic to *Drosophila* cells were hindered by an inability to determine whether the polyalanine protein was being expressed. Neither polyalanine nor polyleucine proteins were detectable by western analysis using the myc antibody. The reason for this is unclear. One possible explanation is that the proteins form aggregates which prevent them from entering the resolving gel. However, this seems unlikely given that no aggregates of polyalanine were visible in the eye sections, and the majority of the polyleucine protein was not present in aggregates. They were also not visible in the stacking gel, yet the entire sample was loaded into the gel, so it seems unlikely that lack of extraction or insolubility due to aggregation is the cause. Furthermore, the polyglutamine proteins formed tight aggregates that appeared to contain all of the visible protein, yet these proteins were detectable on the western blot.

An alternative explanation is that the proteins were not detected because they were not being expressed. However, this is not the case for polyleucine, as the protein was

visible in eye sections. Furthermore, the RT-PCR result demonstrates that the GCA repeat is being expressed, at least at the RNA level. A third explanation is that the polyleucine and polyalanine tracts form some sort of conformation that makes the myc epitope undetectable; evidence contradicting this comes from other studies where polyalanine and polyleucine proteins were detected using western analysis. In one of these, 30 repeats of alanine or leucine were detectable as bands on a western blot using an antibody against a C-terminal YFP tag; they were as easily visible as a polyglutamine protein containing 30 copies. Longer repeats were also detectable, with bands corresponding to proteins consisting of 70 alanines and 130 leucines visible in the stacking gel [225]. In another study of polyleucine proteins, a protein consisting of 291 leucines with a C-terminal myc tag was detectable by western (although in the stacking gel) using the 9E10 myc antibody [131]. These previous findings make it difficult to explain why polyleucine or polyalanine proteins could not be detected in this study.

Another puzzling aspect of the polyalanine experiments is the contraction of the repeat tract. The reason behind this is unclear. Contraction of a CGG repeat tract has been reported in *Drosophila* [188]; others have reported both instability [227] and stability of a CAG repeat [228] suggesting that repeat context may contribute to stability in the *Drosophila* system. Contraction of the repeat tract during generation of lines containing the $(CAG)_{99}$ construct was observed here, with 5 out of 16 lines containing a shorter repeat than the construct used for microinjection. However, this rate of contraction was much lower than that seen for the GCA repeat construct. At the DNA level, the GCA repeat is similar to the CAG repeat, with the only difference being the addition of one base-pair before the repeat and the deletion of one base-pair after. Therefore the same level of stability would be expected for the two types of repeat. This suggests that instability of the repeat at the DNA level is not the main cause of the contraction of the GCA repeat.

An alternative explanation for the lack of flies generated containing the full length GCA repeat is that expression from this construct is toxic to *Drosophila*. A low level of expression of UAS constructs in the absence of a GAL4 driver has not been reported in the literature, however leaky expression of the GCA repeat causing lethality could mean that only flies containing a contracted repeat survive to give rise to transgenic lines. If this were the case, in the surviving lines containing a contracted repeat, a severe phenotype would be expected in the presence of GAL4; however, no such

phenotype was observed. This lack of phenotype may be caused by lack of expression of the polyalanine protein due to its toxicity; it is possible that the lines generated were those in which the repeat construct was contracted and not expressed. This seems unlikely, as silencing of transgene expression at the level of translation has not been described previously, and there is no obvious mechanism by which such silencing could occur. However, toxicity and silencing remains a possible explanation for the repeat contractions and lack of detectable protein product observed.

Recent description of Drosophila expressing expanded polyalanine

Since the investigation of polyalanine toxicity in *Drosophila* described here was undertaken, another *Drosophila* model of expanded polyalanine disease has been described. In this case, a polyalanine tract consisting of 37 repeats tagged with enhanced green fluorescent protein (EGFP) was expressed in the fly under using the UAS-GAL4 system. Upon expression in the nervous system, the 37-alanine protein caused reduced survival to adulthood and early death, whilst expression in the eye using *eyeless-GAL4* led to rough eyes that were significantly reduced in size. The polyalanine protein formed inclusions in the eye disc and nervous system, and was detectable on a western blot using an antibody raised against GFP. Interestingly, an interaction between the 37-polyalanine protein and the GAL4 transcription factor was noted, with expression of polyalanine reducing the toxicity caused by GAL4 expression in the eye. GAL4 was found localised to polyalanine inclusions in cell culture, suggesting that the reduction of GAL4 toxicity in the fly may be caused by its sequestration into polyalanine aggregates [229].

It is unclear whether or not these findings shed any light on the difficulties encountered here in expressing expanded polyalanine proteins in *Drosophila*. The copy number of the alanine repeat used in this study was significantly longer (above 50), and the protein containing the repeat is predicted to be smaller than the 36 kDa protein consisting of EGFP fused to polyalanine described in the recent study. For these reasons, the toxicity of the protein described here is predicted to be greater, and thus if it were expressed, a more severe phenotype should result. Therefore, the results of the recent study, which demonstrate that expanded polyalanine proteins are toxic in *Drosophila*, support the conclusion that in this case the polyalanine protein was not expressed, and this is the reason for the lack of toxicity and lack of detection of the protein. It is not clear whether this lack of expression is due to an

interaction between the polyalanine protein and GAL4 – this explanation seems unlikely, as expression was achieved using GAL4 drivers in the previous study. However, it is possible that the longer length of the polyalanine protein described here increases its propensity to aggregate, and so it is more efficient at sequestering GAL4, resulting in decreased expression. Alternatively, the predicted increased toxicity of the longer alanine tract may have made generation of transgenic lines carrying the coding sequence of the protein impossible, as a small amount of leaky expression may be enough to induce lethality.

Recent CAG vs CAA results in cultured cells contradict those described here

Despite the difficulty encountered in generating transgenic flies expressing polyalanine, the CAA/CAG experiments still address the polyalanine hypothesis. In their main frame, both constructs encode polyglutamine, but if frameshifting occurs, only the CAG repeat can encode polyalanine. Therefore the observation that both repeats induced the same phenotype implies that polyalanine is not involved in pathogenesis. Following the publication of the results described here, a similar experiment comparing CAA and CAG repeats to determine if polyalanine is contributing to the CAG phenotype was reported [133]. In this case, $(CAA)_{73}$ or $(CAG)_{78}$ with a C-terminal epitope tag was transfected into COS-1 cells. Frameshifting to the +2 frame (and hence expression of polyalanine) occurred only within the CAG repeat, but not the CAA repeat or a shorter $(CAG)_{25}$ repeat. Cell death was assayed by light scattering, which represents a measurement of the morphological changes that occur in dying cells. This revealed a higher level of cell death in the cells expressing $(CAG)_{73}$ compared to cells expressing $(CAA)_{78}$. In addition, a higher level of cell death was seen in cells expressing $(CAG)_{73}$ undergoing a –1 frameshift than when no frameshift occurred, to the extent that the toxicity of $(CAG)_{73}$ without frameshifting was the same as the toxicity of $(CAG)_{25}$. The authors stated that this suggests the length of the polyglutamine repeat alone has no effect on cell toxicity in this assay. They concluded that the decreased toxicity of CAA repeats compared to CAG repeats is due to frameshifting that occurs within the CAG repeat [133].

The reason behind the conflicting results obtained in this study and those presented here is not clear. The simplest explanation is that frameshifting does not occur in *Drosophila*, perhaps for the same reason that expanded CAG tracts appear to be more stable in *Drosophila* than other organisms, even in the presence of mutations in

genes in the mismatch repair pathway, which is involved in repeat stability [228,230,231]. If frameshifting were not occurring in *Drosophila* as it does in COS-1 cells, then the toxicity of $(CAA)_{94}$ and $(CAG)_{99}$ would be the same in the fly, in line with the results described here. However, Toulouse *et al.* reported equal toxicity of long CAA and short CAG repeats, which was also the same as the toxicity of the long CAG repeat when frameshifting did not occur. Therefore, in *Drosophila* in the absence of frameshifting, the toxicity of $(CAG)_{99}$, $(CAA)_{94}$, $(CAG)_{52}$, $(CAA)_{20}$ and $(CAG)_{30}$ would be predicted to be the same; however, this was not the case. Thus a lack of frameshifting in *Drosophila* would not account for the different results obtained in these studies, and so the reason behind the differing results is unknown. It is possible that it represents some other disparity between the *Drosophila* and mammalian systems – if this is the case, increased toxicity of CAG over CAA in the mammalian system would indicate the possibility of a role for RNA that does not occur in the *Drosophila* system. Further investigation of RNA pathogenesis in *Drosophila* is described in Chapter 5.

Polyleucine toxicity in the Drosophila eye

In contrast to cell culture results and predictions based on hydrophobic nature of the amino acid, polyleucine seems to be less toxic in *Drosophila* than polyglutamine. This is based on the observation that only one line expressing $(CAG)_{99}$ in the eye did not show a phenotype, whereas more than half of the lines expressing $(CTG)_{86}$ did not show a phenotype. Also, the polyleucine phenotype appeared to be milder in general than the polyglutamine phenotype, with no pigment loss observed and no collapsing of the eye structure. Further experiments to investigate the nature of the polyleucine phenotype were not undertaken, largely due to the fact that no human diseases caused by polyleucine expansions have been described to date. However, this may change in the future, as HDL2 is caused by a CTG expansion that encodes polyleucine in one of its potential reading frames. In interpreting the results from the polyleucine experiment, it is important to note that the possibility has not been ruled out that the phenotype resulting from expression of the CTG repeat construct is RNA mediated, as CUG expansions in mRNA transcripts are the cause of DM and SCA8. Experiments undertaken to examine this will be described in Chapter 5.

Limitations of the Drosophila system

This study has utilised a *Drosophila* model of polyglutamine disease that has previously been well characterised. The neurodegeneration caused by expression of

expanded polyglutamine tracts in this model system is late onset and progressive [84], reflecting the nature of pathogenesis in the human disorders. However, in the *Drosophila* system, the expanded polyglutamine proteins are expressed at a high level driven by GMR-GAL4. This is necessary to ensure that the phenotype is visible during the short lifespan of the fly, but does not accurately reflect the human disease state. Thus, it is possible that this level of expression causes degeneration in *Drosophila* via a mechanism which is not involved in the human diseases or which plays a minor role. For example, the high level of polyglutamine protein in the eye may overwhelm the cellular response and induce toxicity in a manner that does not occur until late stages of the human diseases; RNA pathogenesis could be an initiating factor in human disease but does not have time to occur in the *Drosophila* system. This explanation seems unlikely but cannot be ruled out; therefore, care must be taken during interpretation of the results presented here, and the limitations of the model system taken into account.