



Trinucleotide repeats: a structural perspective

Bruno Almeida, Sara Fernandes[†], Isabel A. Abreu[†] and Sandra Macedo-Ribeiro*

Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal

Edited by:

Thomas M. Durcan, McGill University, Canada

Reviewed by:

Denis Soulet, Laval University, Canada
Thomas M. Durcan, McGill University, Canada

***Correspondence:**

Sandra Macedo-Ribeiro, Instituto de Biologia Molecular e Celular, Universidade do Porto, Rua do Campo Alegre 823, 4150-180 Porto, Portugal
e-mail: sribeiro@ibmc.up.pt

[†]Present address:

Sara Fernandes, Shannon ABC, Limerick Institute of Technology, Limerick, Ireland;
Isabel A. Abreu, GplantS, Instituto de Tecnologia Química e Biológica, Oeiras, Portugal.

Trinucleotide repeat (TNR) expansions are present in a wide range of genes involved in several neurological disorders, being directly involved in the molecular mechanisms underlying pathogenesis through modulation of gene expression and/or the function of the RNA or protein it encodes. Structural and functional information on the role of TNR sequences in RNA and protein is crucial to understand the effect of TNR expansions in neurodegeneration. Therefore, this review intends to provide to the reader a structural and functional view of TNR and encoded homopeptide expansions, with a particular emphasis on polyQ expansions and its role at inducing the self-assembly, aggregation and functional alterations of the carrier protein, which culminates in neuronal toxicity and cell death. Detail will be given to the Machado-Joseph Disease-causative and polyQ-containing protein, ataxin-3, providing clues for the impact of polyQ expansion and its flanking regions in the modulation of ataxin-3 molecular interactions, function, and aggregation.

Keywords: amino acid-repeats, microsatellites, protein complexes, protein aggregation, amyloid, protein structure

TRINUCLEOTIDE REPEATS AND HUMAN DISEASE

Trinucleotide repeat (TNR) expansions and their association with neurological disorders have been known for the past 20 years (La Spada et al., 1991). Expansion of CAG, GCG, CTG, CGG, and GAA repeats located in coding or non-coding sequences of different genes (summarized in **Table 1**; **Figures 1** and **2**) are associated with a diverse range of human monogenic diseases such as Spinobulbar Muscular Atrophy (SBMA, a.k.a. Kennedy disease), Huntington Disease (HD), Spinocerebellar Ataxias (SCAs), Oculopharyngeal Muscular Dystrophy (OPMD), Myotonic Type 1 (DM1), Fragile X-Associated Tremor Ataxia Syndrome (FXTAS), and Friedreich Ataxia (FRDA) (for a review see Orr and Zoghbi, 2007), with longer repeats being correlated with earlier age at onset and increased disease severity. These TNR are highly unstable and the repeat tract length can change between affected individuals within the same family and can be different in different tissues (La Spada, 1997; Brouwer et al., 2009). More interestingly, in the brain of patients affected by CAG expansions, differences in repeat instability have been found between specific cell types (Pearson et al., 2005; Gonitel et al., 2008; Lopez Castel et al., 2010). GCG repeats are usually shorter and reveal a higher stability in different tissues and across generations than CAG repeats. The dynamic nature of these DNA repeat expansions is a consequence of their capability to form different secondary structures, which interfere with the cellular mechanisms of replication, repair, recombination and transcription (for a recent review see Lopez Castel et al., 2010). The molecular mechanisms underlying pathogenesis in those disorders, either associated with mental retardation, neuronal, or muscular degeneration, might result from alterations in the levels of gene expression and/or the function of the RNA or protein it encodes, mechanisms that likely

act in concert to influence the pattern of selective cell toxicity. Some of those toxicity mechanisms will be briefly discussed below.

TRINUCLEOTIDE REPEATS AND RNA STRUCTURE

The formation of hairpin structures within the TNR RNA is related to the gain in RNA toxic function, the major pathogenic mechanism associated with CUG and CGG repeat expansions in non-coding regions of DM1 and FXTAS transcripts, which was also shown to contribute to pathogenesis in CAG repeat disorders such as HD and Machado-Joseph disease (MJD, a.k.a. SCA3) (reviewed in Krzyzosiak et al., 2012). These duplex structures, whose stability is positively correlated with the repeat size (Napierala and Krzyzosiak, 1997), sequester dsRNA binding proteins involved in mRNA splicing such as CUG-binding protein (CUGBP) and muscleblind protein 1 (MBNL1) (Miller et al., 2000), inducing aberrant splicing in affected cells, compromising multiple intracellular pathways, affecting cell-quality control regulation, and ultimately resulting in cell dysfunction (Li and Bonini, 2010). Structural studies on model trinucleotide CUG, CAG, and CGG repeats forming double-stranded chains revealed the features induced by periodic U-U, A-A, and G-G mismatches, and provided hints into the structural details of pathogenic RNAs that are recognized by RNA-binding proteins (Mooers et al., 2005; Kiliszek et al., 2010, 2011; Kumar et al., 2011; Parkesh et al., 2011). MBNL1 is composed of four zinc-containing RNA-binding domains arranged in two tandem segments, with the C-terminal zinc-finger pair displaying a GC-sequence recognition motif (Teplova and Patel, 2008) and interacting with the stem region of expanded CUG RNAs (Yuan et al., 2007). Electron microscopy analysis of MBNL1:CUG¹³⁶ complexes showed that the pathogenic dsRNA forms a scaffold

Table 1 | Human diseases associated with nucleotide repeat expansions (adapted from Messaied and Rouleau, 2009; Lopez Castel et al., 2010; Matos et al., 2011).

Disease name	Repeat type	Repeat location	Gene	Protein (UniProt identifier, number of residues)	Biological process ^a	Normal repeat length	Disease repeat length	Protein structure determined?
Spinal and bulbar muscular atrophy (SBMA)	CAG	Protein coding region (polyQ)	<i>AR</i>	Androgen receptor (P10275, 919 residues)	Transcription, transcription regulation	9–36	38–62	Residues 20–30 and 671–919 (PDB code 1xow)
Huntington's disease (HD)	CAG	Protein coding region (polyQ)	<i>HTT</i>	Huntingtin (P42858, 3142 residues)	Apoptosis	6–34	36–121	Residues 5–18 (3lrh), Residues 1–17 (2ld0, 2ld2), Residues 1–64 (3io4, 3io6, 3lor, 3lot, 3iou, 3iov, 3iow)
Dentatorubral-pallidousian atrophy (DRPLA)	CAG	Protein coding region (polyQ)	<i>ATN1</i>	atrophin 1 (P54259, 1190 residues)	Transcription, transcription regulation	7–34	49–88	No structural information
Spinocerebellar ataxia 1 (SCA1)	CAG	Protein coding region (polyQ)	<i>ATXN1</i>	ataxin 1 (P54253, 815 residues)	Transcription, transcription regulation	6–39	40–82	Residues 563–693 (1oa8)
Spinocerebellar ataxia 2 (SCA2)	CAG	Protein coding region (polyQ)	<i>ATXN2</i>	ataxin 2 (Q99700, 1313 residues)	No associated GO keywords for biological process	15–24	32–200	Residues 912–928 (3ktr)
Spinocerebellar ataxia 3 (SCA3)	CAG	Protein coding region (polyQ)	<i>ATXN3/MJD</i>	ataxin 3 (P54252, 364 residues)	Transcription, transcription regulation, Ub1 conjugation pathway	10–51	55–87	Residues 1–182 (1yzb), Residues 222–263 (2klz)
Spinocerebellar ataxia 6 (SCA6)	CAG	Protein coding region (polyQ)	<i>CACNA1A</i>	CACNA 1 _A , P/Q-type α 1A calcium channel subunit (O00555, 2505 residues)	Calcium transport, ion transport, transport	4–20	20–29	Residues 1955–1975 (3bxx)
Spinocerebellar ataxia 7 (SCA7)	CAG	Protein coding region (polyQ)	<i>ATXN7</i>	ataxin 7 (O15265, 892 residues)	Transcription, transcription regulation	4–35	37–306	Residues 330–401 (2kkr)
Spinocerebellar ataxia 17 (SCA17)	CAG	Protein coding region (polyQ)	<i>ATXN17</i>	TATA box binding protein (TBP) (P20226, 339 residues)	Transcription, transcription regulation, Host-virus interaction	25–42	47–63	Residues 159–337 (1cdw, 1c9b, 1jfi, 1nvp, 1tgh)
Multiple skeletal dysplasias (COMP)	GAC	Protein coding region (polyaspartate)	<i>COMP</i>	cartilage oligomeric matrix protein (a.k.a Thrombospondin-5) (P49747, 757 residues)	Apoptosis, cell adhesion	5	4, 6, 7	Residues 225–757 (3fby).
Synpolydactyly (HOXD13)	GCG	Protein coding region (polyA)	<i>HOXD13</i>	homeobox D13 (P35453, 343 residues)	Transcription, transcription regulation	15	22–29	No structural information

(Continued)

Table 1 | Continued

Disease name	Repeat type	Repeat location	Gene	Protein (UniProt identifier, number of residues)	Biological process ^a	Normal repeat length	Disease repeat length	Protein structure determined?
Oculopharyngeal Muscular Dystrophy (OPMD)	GCG	Protein coding region (polyA)	<i>PABPN1</i>	Polyadenylate-binding protein 2 (Q86U42, 306 residues)	mRNA processing	10	12–17	Residues 167–254 (3b4d, 3b4m, 3ucg)
Cleidocranial dysplasia (CBFA1)	GCG	Protein coding region (polyA)	<i>RUNX2</i>	Runt-related transcription factor 2 (Q13950, 521 residues)	Transcription; transcription regulation	17	27	No structural information
Holoprosencephaly (ZIC2)	GCG	Protein coding region (polyA)	<i>ZIC2</i>	Zinc-finger protein ZIC 2 (O95409, 532 residues)	Differentiation, neurogenesis, transcription, transcription regulation	15	25	No structural information
Hand-Foot-Genital Syndrome/HOXA13)	GCG	Protein coding region (polyA)	<i>HOXA13</i>	homeobox A13 (P31271, 388 residues)	Transcription, transcription regulation	18	24–26	No structural information
Blepharophimosis/ptosis/epicanthus inversus syndrome type II (FOXL2)	GCG	Protein coding region (polyA)	<i>FOXL2</i>	Forkhead box like 2 (P58012, 376 residues)	Differentiation, transcription, transcription regulation	14	22–24	Residues 322–328 (2I7z)
Infantile spasm syndrome (ARX)	GCG	Protein coding region (polyA)	<i>ARX</i>	Aristaless-related homeobox (Q96QS3, 562 residues)	Differentiation, neurogenesis, transcription, transcription regulation	10–16	17–23	No structural information
Myotonic dystrophy type 1 (DM1)	CTG	3' UTR	<i>DMPK</i>	Myotonic dystrophy protein kinase (DMPK) (Q09013, 639 residues)	No associated GO keywords for biological process	5–37	90–6500	Residues 11–420 (2vd5), Residues 460–537 (1wt6)
Friedreich ataxia (FRDA)	GAA	Intron	<i>FXN</i>	Frxataxin (Q16595, 210 residues)	Heme biosynthesis, Ion transport, Iron storage, Iron, transport	6–32	>200	Residues 88–210 (1ekg), Residues 91–210 (1ly7), Residues 82–210 (3s4m, 3s5d, 3s5e, 3s5f, 3t3j, 3t3k, 3t3l, 3t3t, 3t3x)
Spinocerebellar ataxia 8 (SCA8)	CTG	3' UTR	<i>ATXN8</i>	Ataxin-8 (a.k.a protein 1C2; (Present in SCA8-specific 1C2-positive intranuclear inclusions) (Q156A1, 80 residues)	Cell death	2–130	>110	No structural information

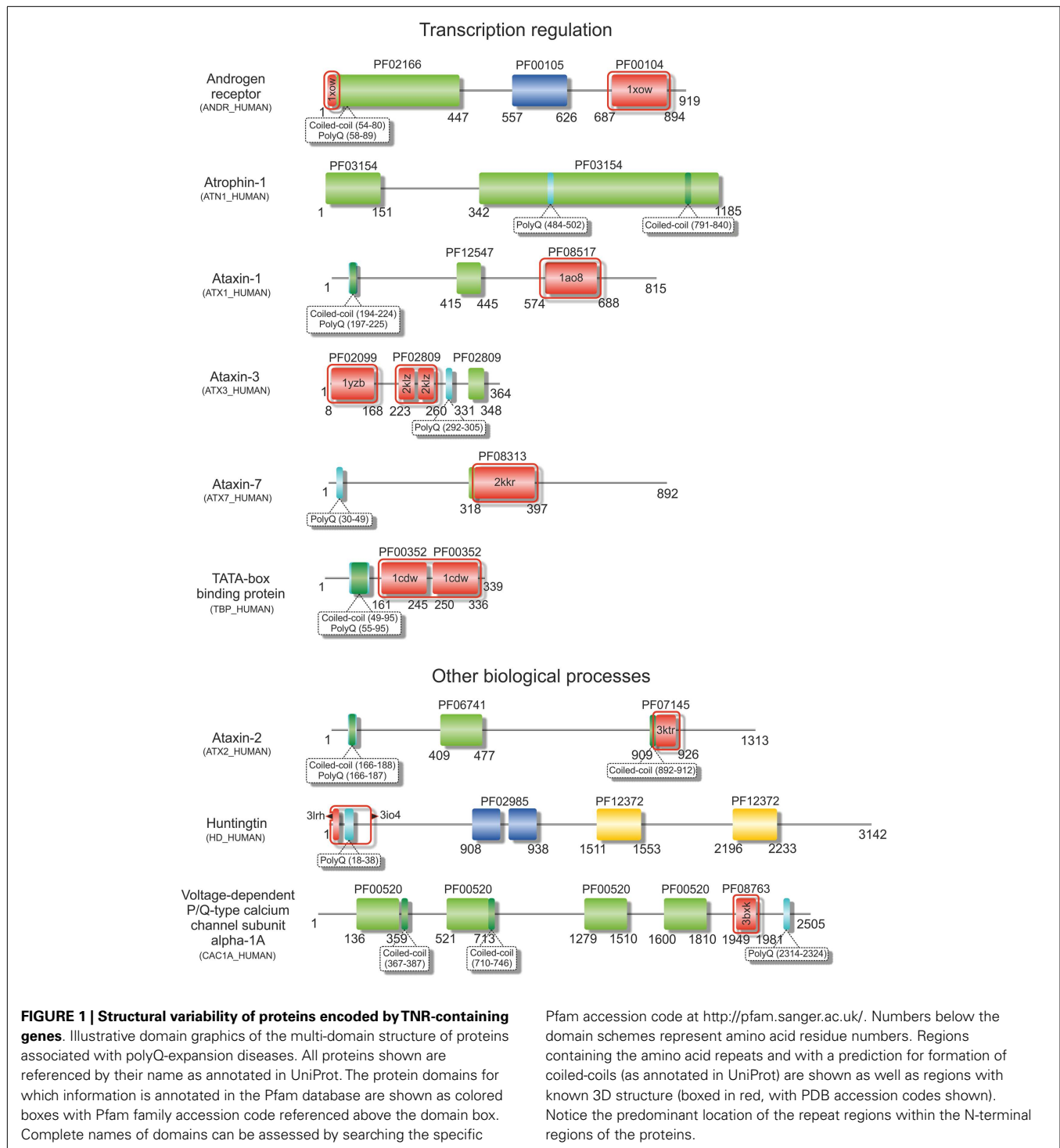
(Continued)

Table 1 | Continued

Disease name	Repeat type	Repeat location	Gene	Protein (UniProt identifier, number of residues)	Biological process ^a	Normal repeat length	Disease repeat length	Protein structure determined?
Spinocerebellar ataxia 12 (SCA12)	CAG	5' UTR	<i>PPP2R2B</i>	Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B β isoform (Q00005, 443 residues)	Apoptosis	7–45	55–78	No structural information
Huntington disease-like 2 (HDL2)	CAG	Alternative splice isoform 2 – polyA-expansion	<i>JPH3</i>	Junctophilin 3 (O8WXH2, 748 residues)	No associated GO keywords for biological process	6–27	51–57	No structural information
FRAXA: fragile X syndrome	CGG	5' UTR	<i>FMR1</i>	Fragile X mental retardation 1 protein (Q06787, 632 residues).	Transport; mRNA transport	6–52	230–2000	Residues 1–134 (2bkd), Residues 216–280 (2fmr), Residues 216–425 (2qnd), Residues 527–541 (2la5)
FXTAS: fragile X tremor/ataxia syndrome	CGG	5' UTR	<i>FMR1</i>	Fragile X mental retardation 1 protein (Q06787, 632 residues).	Transport; mRNA transport	6–52	59–230	Residues 1–134 (2bkd), Residues 216–280 (2fmr), Residues 216–425 (2qnd), Residues 527–541 (2la5)
FRAXE: fragile X syndrome	CGG	5' UTR	<i>FMR2</i>	Fragile X mental retardation 2 protein (P51816, 1311 residues)	mRNA processing, mRNA splicing	4–39	200–900	No structural information

UTR, untranslated region.

^a Biological Function based on Gene Ontology as annotated in UniProt.



with tandem spaced MBNL1 binding sites were MBNL1 oligomers with a ring-like structure can assemble, possibly leading to the formation of the ribonuclear foci identified in cell models of these TNR diseases (Yuan et al., 2007; de Mezer et al., 2011). The structure and stability of the TNR hairpin structures formed depends on the presence of interruptions as well as on the nature of the flanking regions. This might be related with the ability of individual

repeats to participate in the RNA toxicity mechanisms (Krzyzosiak et al., 2012).

In FRDA and FXTAS, pathogenesis results predominantly from decreased expression of the associated genes (*FXN* and *FMR1/FMR2*) caused by the expansion of GAA and CGG repeats, respectively, which results in loss of function of key proteins involved in iron-sulfur cluster biogenesis and mRNA translation

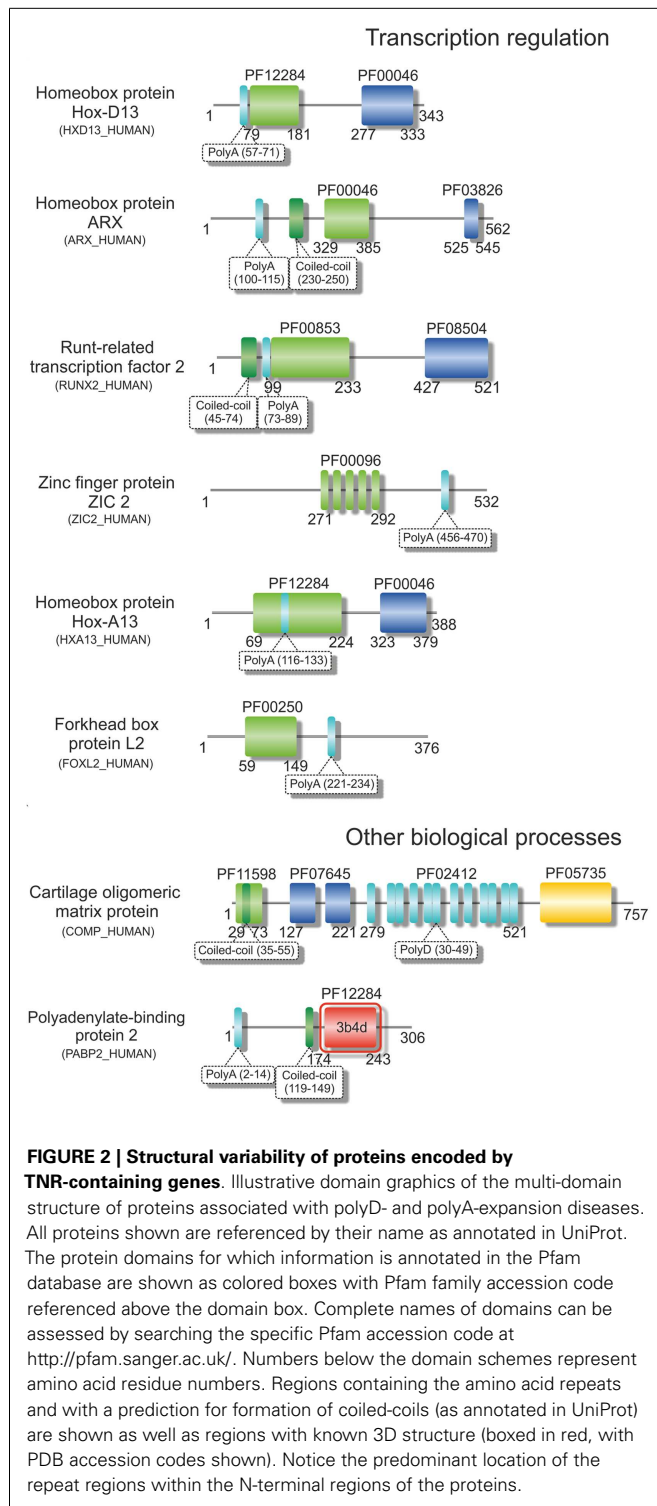


FIGURE 2 | Structural variability of proteins encoded by TNR-containing genes. Illustrative domain graphics of the multi-domain structure of proteins associated with polyD- and polyA-expansion diseases. All proteins shown are referenced by their name as annotated in UniProt. The protein domains for which information is annotated in the Pfam database are shown as colored boxes with Pfam family accession code referenced above the domain box. Complete names of domains can be assessed by searching the specific Pfam accession code at <http://pfam.sanger.ac.uk/>. Numbers below the domain schemes represent amino acid residue numbers. Regions containing the amino acid repeats and with a prediction for formation of coiled-coils (as annotated in UniProt) are shown as well as regions with known 3D structure (boxed in red, with PDB accession codes shown). Notice the predominant location of the repeat regions within the N-terminal regions of the proteins.

at synapses. Nevertheless, in FXTAS RNA toxicity is also proposed to play a role in pathogenesis (Li and Bonini, 2010). The recently discovered mechanisms of pathogenesis in spinocerebellar ataxia type 8 (SCA8) uncovered the extreme complexity of TNR disorders. In fact, SCA8 is caused by expansion of CTG/CAG repeats in the affected gene, which are transcribed bi-directionally leading

to the generation of expanded CUG and CAG-containing transcripts further translated into homopolymeric proteins, so that pathogenesis can be mediated by both RNA and protein toxicity (Merienne and Trottier, 2009). Curiously, recent data have highlighted the possibility of non-ATG translation across expanded TNR in all possible reading frames, which might further contribute to the generation of novel toxic proteins and RNAs adding to the multi-parametric character of the pathogenic mechanisms associated with TNR diseases (Li and Bonini, 2010; Pearson, 2011; Sicot et al., 2011).

TRINUCLEOTIDE REPEATS WITHIN PROTEIN CODING REGIONS

Over 20 years ago, the finding that the expansion of CAG repeats within the coding sequence of the androgen receptor gene was the genetic basis of SBMA (La Spada et al., 1991) represented a hallmark in the discovery of these novel dynamic mutations and their association with human disease. Some years later, the identification of intracellular inclusions containing the expanded proteins (Paulson et al., 1997) provided a clue to pathogenesis, directing research in the field into an extensive search for the mechanisms of polyQ-induced protein aggregation. The moderate expansion of GCG and CAG repeats, which are translated into polyA and polyQ tracts in the affected proteins (Figures 1 and 2), results in protein misfolding and aggregation, in accordance with a general, although not always unique, toxic gain of function mechanism of pathogenesis (Williams and Paulson, 2008). The appearance of insoluble cytoplasmic or nuclear inclusions enriched in the expanded polyA- or polyQ-containing protein constitutes a characteristic fingerprint of these diseases (Messaed and Rouleau, 2009; Orr, 2012a), regardless of their controversial role in pathogenesis. While the proteins containing polyA repeats are predominantly transcription factors with a role in development (see Table 1 and Amiel et al., 2004; Messaed and Rouleau, 2009), most of the proteins linked to polyQ-expansion diseases are involved in DNA-dependent regulation of transcription or neurogenesis and often contain multiple intermolecular partners (Butland et al., 2007). Despite the overall lack of sequence or structural homology, both polyQ- and polyA-repeat expansions are associated with formation of β -rich amyloid-like protein inclusions, and with the wider group of protein misfolding disorders. These inclusions are enriched in ubiquitin, proteasome subunits, and chaperones, and often recruit macromolecules that are part of the macromolecular interaction networks associated with the proteins' native functions (Williams and Paulson, 2008). As an example, the poly(A)-binding protein PABNP1 forms insoluble inclusions upon alanine expansion, co-aggregating together with poly(A)-mRNA, proteasome subunits, ubiquitin, heat-shock proteins, and SKIP, a transcription factor associated with muscle-specific gene expression (Brais, 2003; Tavanez et al., 2009; Winter et al., 2013).

The simplistic view of the predominant role of the inclusions in polyQ-induced pathogenesis was later challenged by the failure of this mechanism to explain the cell-specific vulnerability characteristic for each disease and by the identification of numerous examples of neuronal toxicity in the absence of visible intracellular inclusions (Arrasate et al., 2004). Indeed, the inclusions were shown to be fibrillar and display amyloid-like properties both *in vivo* and *in vitro* (Huang et al., 1998; Bevivino and Loll, 2001;

Sathasivam et al., 2010) and, in a mechanistic parallel with the pathogenic mechanisms proposed for “classical” amyloids, many studies suggested that the insoluble inclusions played a protective role, sequestering toxic, and misfolded protein conformers (Arrasate et al., 2004; Rub et al., 2006; Miller et al., 2010). Indeed, soluble intermediates in the aggregation pathway such as misfolded β -sheet rich polyQ protein monomers and oligomers have latter been identified and proposed to represent the major toxic species (Kayed et al., 2003; Gales et al., 2005; Nagai et al., 2007; Miller et al., 2011). Also, in OPMD, the primary toxic species are proposed to be the soluble variants of the expanded polyA-repeat protein PABPN1 (Messaed et al., 2007). It is currently accepted that in polyQ disorders the expanded region plays a role in inducing the self-assembly of the carrier protein, which engages in pathogenic interactions and leads to the formation of toxic monomers or oligomers (Takahashi et al., 2008; Weiss et al., 2008) latter converted to insoluble intracellular amyloid-like oligomers where both expanded and “normal” protein are sequestered along with other macromolecular partners (reviewed in Williams and Paulson, 2008; Matos et al., 2011; Costa and Paulson, 2012). As more biochemical data is gathered, more is understood about the role of amino acid expansions in modulating the interaction with macromolecular partners. As an example, expansion of the polyA tract in PABPN1 results in increased association with Hsp70 chaperones and type I arginine methyl transferases (Tavanez et al., 2009). This indicates that the distinct neuropathological features arising from this amino acid-repeat expansion might at least partially result from alterations on the native biological functions and macromolecular interactions of the carrier protein, which might vary in different intracellular environments.

Recent data have shown that expansion of polyA repeats is frequently associated with loss of normal function altering a multitude of cellular pathways with consequences in cell functionality (Amiel et al., 2004; Messaed and Rouleau, 2009), although protein aggregation might also play a dominant role in some of the polyA-associated disorders (Messaed and Rouleau, 2009; Winter et al., 2013). Studies with polyQ proteins have shown that pathogenesis might result from a subtle imbalance in the association of the mutant protein with multiple cellular partners and that toxicity and neuronal death could result from a combination of protein self-assembly and functional alterations (Friedman et al., 2007; Li et al., 2007b; Lim et al., 2008; Kratter and Finkbeiner, 2010; Orr, 2012b; Pastore and Temussi, 2012). In fact, neuronal death as a result of polyQ-expansion seems to resemble that of linker cell in *C. elegans* (Pilar and Landmesser, 1976; Chu-Wang and Oppenheim, 1978; Blum et al., 2012, 2013) which involves the polyQ protein pqn-4, pointing for a common mechanism for linker cell death, and neuronal death in polyQ diseases (Blum et al., 2013).

Polyglutamine diseases constitute a representative and largely studied group of neurodegenerative disorders where considerable amounts of data have been collected on the role of expanded polyQ for disease pathogenesis. However, given the proposed function of polyQ regions in mediating protein–protein interactions, which might be modulated by polyQ-expansion (Schaefer et al., 2012), the information on the role of these regions for native protein function, structure, and dynamics is still limited. Structural and functional information on the role of these repeat sequences

in protein function is crucial to better understand how expansion affects selected neuronal subpopulations. Below, we briefly discuss the current knowledge on the function and structure of polyQ repeats and their role on macromolecular interactions, and finally focus on the known structural and functional information on ataxin-3, the protein whose mutation causes MJD.

FUNCTION OF PolyQ ON PROTEIN–PROTEIN INTERACTIONS AND EVOLUTION

Until recently, the function of many amino acid-repeat-containing proteins and the role of homopeptide regions were somewhat obscure. However, several global analysis studies on single amino acid-repeat-containing proteins shed light onto their function and onto the biological significance of the repeated region, in particular of polyQ, the most prevalent amino acid repetition in humans (Alba and Guigo, 2004). It is now accepted that TNR, particularly those located within protein-coding regions, are considered important mutators providing the genetic variability required for driving evolution (King, 1994; Kashi et al., 1997; Kashi and King, 2006; Nithianantharajah and Hannan, 2007). In fact, simple or low-complexity amino acid-repeats are rare within prokaryotic but extremely abundant within eukaryotic proteins, particularly over-represented in *Plasmodium* (49–90% of the total proteome), *D. discoideum* (52%), *D. melanogaster* (20%), *C. elegans* (9%), and *H. sapiens* (14%) (Haerty and Golding, 2010). Among all homopolymeric repeats, the most common on eukaryotic proteins are glutamine, asparagine, alanine, and glutamate repeats (Faux et al., 2005). This seems to indicate that there has been a strong negative selection against the appearance of hydrophobic amino acid-repeats with high tendency to aggregate, such as polyisoleucine, poly-leucine, polyphenylalanine, and polyvaline (Oma et al., 2005, 2007).

The homopeptide regions seem to be particularly relevant for brain development and function, since these repeated regions can be found in various neurodevelopmental genes (Nithianantharajah and Hannan, 2007). Indeed, the sexual behavior of prairie voles (Hammock and Young, 2005), as well as human pair-bonding (Walum et al., 2008), seems to be dependent on the repeat length in the vasopressin 1A receptor gene. A wide study of the distribution and function of homopeptide-containing proteins could also demonstrate a clear trend in humans, *D. melanogaster*, and *C. elegans*, with the majority of homopeptide-containing proteins performing roles in transcription/translation and signaling processes and to a less extend in transport and adhesion processes (Faux et al., 2005). A similar profile was also found in a comparative analysis of proteins with amino acid-repeats in human and rodents (Alba and Guigo, 2004) and also on a comparative genomic study in domestic dogs, which unveiled an association between morphological variations and the length of the repeated region in the transcription factor-encoded genes *ALX4* and *RUNX2* (Fondon and Garner, 2004). Analysis of the human genome also revealed the existence of 64 CAG repeat-containing genes involved in biological processes such as regulation of transcription, binding of transcriptional co-activators and transcription factors, and in neurogenesis in general (Butland et al., 2007). Additionally, a detailed analysis of the human polyQ database (<http://pxgrid.med.monash.edu.au/polyq/>) (Robertson et al., 2011) also indicated that the

majority of polyQ-containing proteins display domains involved in development (Homeobox domain-containing proteins, Fibroblast growth factor receptor), chromatin remodeling (Bromodomain and PHD-containing proteins), and signal transduction (PDZ domain-containing proteins), all biological processes that are highly dependent on protein–protein interactions and associated with the formation of multicomponent protein complexes. As for humans, analysis of bovine polyQ proteins revealed an enrichment for large multi-domain transcriptional regulators (Whan et al., 2010).

It is currently accepted that the majority of repeat-containing proteins perform roles in processes that require the assembly of large multiprotein or protein/nucleic acid complexes (Faux et al., 2005; Hancock and Simon, 2005; Whan et al., 2010). Supporting this notion is the fact that homopolymeric amino acid-repeats are considered to be unstructured (Gojobori and Ueda, 2011) and that intrinsically unstructured regions are suggested to constitute macromolecular docking sites, which become structured only when bound to cognate ligand partners (Huntley and Golding, 2002; Simon and Hancock, 2009). In fact, “hub proteins” contain significantly longer and more frequent repeats or disordered regions, which facilitate binding to multiple partners (Dosztanyi et al., 2006). Recently, Fiumara et al. (2010) found an overrepresentation of coiled-coils domains in polyQ-containing proteins and in their interaction partners, which are able to form α -helical supersecondary structures, often inducing protein oligomerization (Parry et al., 2008). Thus, polyQ tracts due to their intrinsic structural flexibility, which is largely influenced by the flanking residues (see PolyQ: A Simple Sequence Repeat with a Polymorphic Structure below), may act as stabilizers of intra- and intermolecular protein interactions, possibly by extending a neighboring coiled-coil region to promote its interaction with a coiled-coil region in an interacting protein partner (Schaefer et al., 2012). A detailed analysis revealed heptad repeats typical of coiled-coils in regions flanking or overlapping polyQ stretches, whose disruption is sufficient to impair CHIP-huntingtin interaction, indicating that coiled-coils are crucial for polyQ-mediated protein contacts. Importantly, coiled-coils also seem to be important for the regulation of aggregation and insolubility of polyQ-containing proteins (see below and Fiumara et al., 2010) as recently proposed by Petrakis et al. (2012), which discovered a recurrent presence of coiled-coil domains in ataxin-1 misfolding enhancers, while such domains were not present in suppressors.

Based on the several observations on the function of polyQ-containing proteins it is suggested that a general function of polyQ, as for the majority of repeat sequences, is to aid in the assembly of macromolecular complexes, either through tethered distant domains or through interactions with the polyQ itself (Gerber et al., 1994; Korschen et al., 1999; Faux et al., 2005). By affecting protein interactions, and being present in particular functional classes such as transcription factors, polyQ is considered central to the evolution of this type of proteins and consequently crucial to the evolution of cellular signaling pathways (Hancock and Simon, 2005).

A structural analysis of polyQ repeats and its flanking domains as well as its role in protein aggregation will be discussed in greater detail in the next sections.

STRUCTURAL STUDIES ON PolyQ REPEATS

Since the discovery that polyQ repeats are associated with human neurodegenerative diseases that a huge effort has been made to determine the structure of polyQ and to understand how expansion of the repeat affects the structure of the carrier protein and/or the normal interaction with molecular partners. The first evidence from the aggregation-prone character of polyQ-rich proteins came from studies with glutamine-rich cereal storage proteins and synthetic glutamine polypeptides (Beckwith et al., 1965; Krull et al., 1965). After the discovery that a number of neurological disorders were triggered by expansion of a polyQ tract in different and unrelated proteins (La Spada et al., 1994), and before intracellular inclusions enriched in the polyQ-expanded protein were identified as a major fingerprint in these diseases (Davies et al., 1997; Paulson et al., 1997), Perutz (1994) anticipated that the expanded polyQ tract could mediate protein–protein interactions causing protein aggregation in neurons and recruiting other polyQ-rich proteins such as transcription factors leading to cellular dysfunction. Below, the structural features and self-assembly properties of polyQ sequences are briefly discussed (for a detailed review on the biophysical and structural features of polyQ, see Wetzel, 2012).

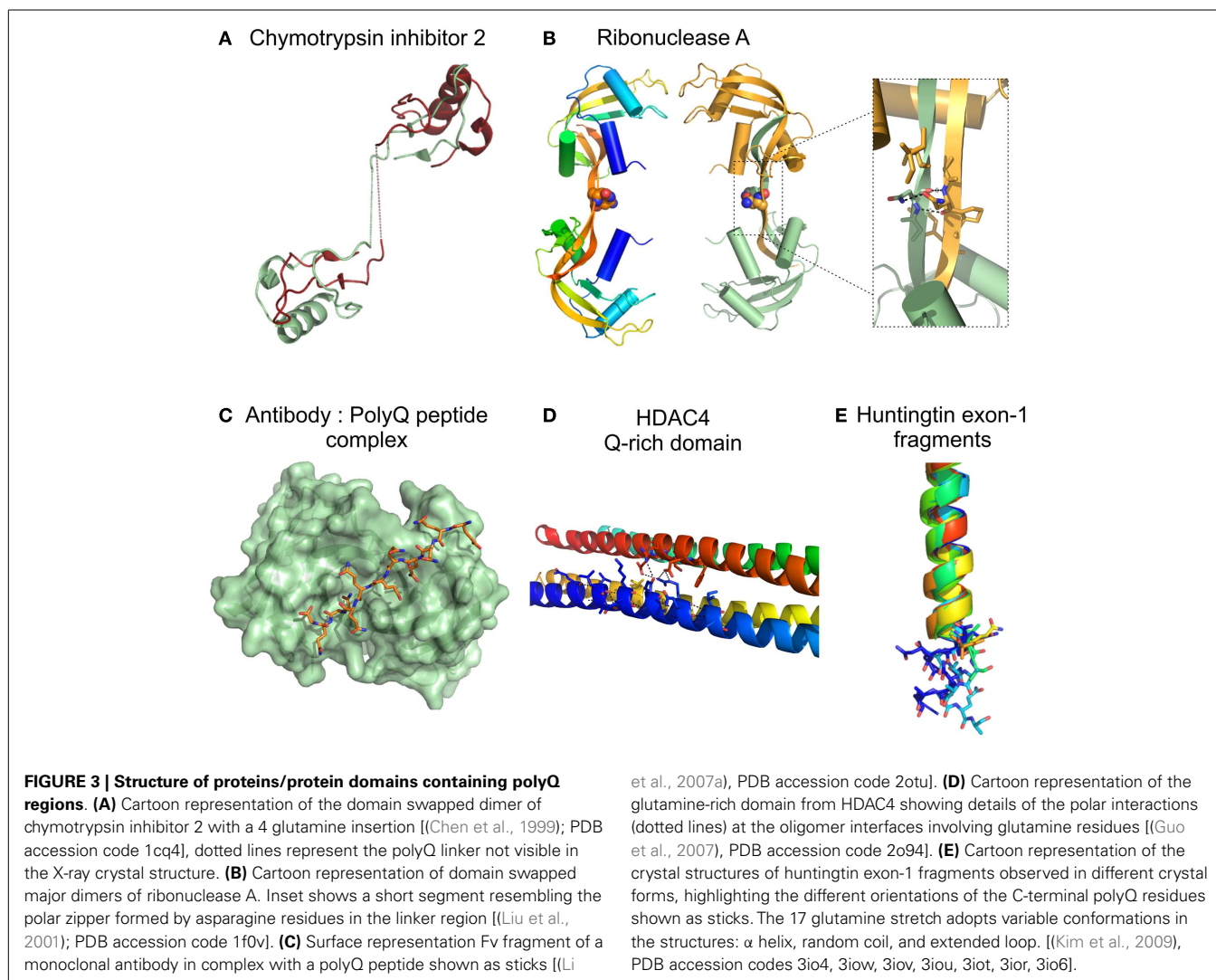
PolyQ: A SIMPLE SEQUENCE REPEAT WITH A POLYMORPHIC STRUCTURE

In order to elucidate the structure of the glutamine repeat and to uncover the structural changes induced by polyQ expansion, several strategies have been put forward including (a) the structural analysis of polyQ-containing peptides of different lengths, (b) the characterization of proteins of well-known structure after insertion of an exogenous polyQ repeat, and structural determination of (c) polyQ-antibody complexes, or (d) natural polyQ-rich proteins.

Using synthetic peptides containing 15 glutamine repeats, Perutz and coworkers proposed that polyQ stretches could self-associate forming hydrogen bonds between their side-chain amide groups and the main chain of a neighboring β -strand, to form cross- β structures (polar zippers) (Perutz, 1994). This study was followed by many reports where synthetic polyQ peptides were used as models of the biophysical properties of polyQ-rich proteins, which established that polyQ-containing peptides have a tendency toward self-assembly into amyloid-like structures (Chen et al., 2002a). Moreover, the results obtained *in vitro* reflected disease features observed *in vivo* such as the correlation between larger polyQ size, increased protein aggregation, and earlier disease onset (Chen et al., 2002b; Kar et al., 2011). Circular dichroism studies of polyQ peptides in solution have shown that their monomeric forms lack regular secondary structure (Altschuler et al., 1997; Klein et al., 2007) and additional biophysical experiments proposed that these peptides can adopt collapsed (Crick et al., 2006; Dougan et al., 2009; Peters-Libeu et al., 2012) or extended (Singh and Lapidus, 2008) coils in solution whose compactness was strongly correlated with the polyQ size (Walters and Murphy, 2009). The determination of the structure of monomeric polyQ peptides with atomic detail is however still lacking as a result of their intrinsic conformational flexibility and tendency to aggregate into heterogeneously sized β -rich oligomers. From the combination of experimental and theoretical methods a picture for polyQ

structure and aggregation is emerging, where the monomeric polyQ adopt an ensemble of conformations lacking regular secondary structures that assemble into β -structures in a polyQ-length dependent fashion (Vitalis et al., 2009; Walters and Murphy, 2009, 2011; Williamson et al., 2010; Kar et al., 2011). Divergent results proposing the existence of predominantly extended or collapsed conformations or the minimum size for polyQ aggregation are likely due to the differences in the introduction of variable flanking residues (Kar et al., 2011). They might result from the insertion of different polyQ tract interrupting residues (Walters and Murphy, 2011), or be a consequence of the protocols used for the preparation and disaggregation of the peptides used for the biophysical studies (Jayaraman et al., 2011). Most results obtained with these peptides do not generally take into account the possible effects of the protein context on the structural properties of the polyQ stretches, a particularly relevant feature considering that the role of non-polyQ domains in protein aggregation has been reported for ataxin-1 (de Chiara et al., 2005), ataxin-3 (Gales et al., 2005), and huntingtin (Tam et al., 2009; Thakur et al., 2009; Liebman and Meredith, 2010).

In a pioneer work, Stott et al. (1995) inserted a G-Q₁₀-G peptide into the inhibitory loop of chymotrypsin inhibitor 2 (CI2), a soluble small protein from barley seeds, showing that this CI2-polyQ chimera has an increased tendency for self-assembly. Even though a CI2 variant with four glutamines crystallized, the structure of the CI2-Q₄ dimer showed that the polyQ region was disordered and that oligomerization was mediated by domain swapping (Figure 3A) and not by direct polyQ association (Chen et al., 1999). A structure resembling the proposed polar zipper was later observed between two asparagines in the hinge loop of the major domain swapped dimer of bovine pancreatic ribonuclease A (Liu et al., 2001) (Figure 3B). Insertion of a 10 glutamine repeat within this hinge loop of ribonuclease A, resulted in domain swapping, oligomerization, and amyloid-like fiber formation, but strikingly the enzyme within the fibers was catalytically active, retaining its native fold (Sambashivan et al., 2005). However, although the structure of the domain swapped dimer was solved by X-ray crystallography, the repeat region was not visible in the electron density maps.



A first overview of a short polyQ stretch at atomic resolution resulted from the structure of a polyQ₁₀ peptide (GQ₁₀G) (**Figure 3C**) bound to MW1, an antibody against polyQ. This structure reveals that polyQ adopts an extended, coil-like structure in which contacts are made between side chains and/or main chain atoms of all 10 glutamines and the antibody-combining site (Li et al., 2007a). The peculiar structural features of these repeat-containing regions were also revealed by the crystallographic structure of a glutamine-rich domain of human histone deacetylase (HDAC4), that folds into a tetramer-forming straight α -helix (**Figure 3D**). The protein interfaces consist of multiple hydrophobic patches separated by polar interaction networks, in which clusters of glutamines engage in extensive intra- and interhelical interactions (Guo et al., 2007). Further details on the structure of polyQ were unveiled by the high-resolution crystal structures of huntingtin (HD) exon 1, containing 17 glutamines (Htt17Q) (Kim et al., 2009). Htt17Q in fusion with maltose-binding protein (MBP) folds into an amino-terminal α -helix followed by a polyQ₁₇ region that adopts multiple conformations in the different crystal forms, including α -helix, random coil, and extended loop, and a polyproline helix formed by the polyP₁₁ and mixed P/Q regions (**Figure 3E**). The authors suggested that the shallow equilibrium between α -helical, random coil, and extended conformations can be subtly altered by the size of polyQ sequence, the neighboring protein context, protein interactions, or by changes in cellular environment, and that this polymorphic behavior is a common characteristic of many amyloidogenic proteins (Kim et al., 2009).

SELF-ASSEMBLY AND AGGREGATION OF PolyQ REPEATS

The first approaches to characterize polyQ-induced protein aggregation and pathogenesis in the context of a full-length protein included the insertion of the polyQ peptides into well-known non-pathogenic protein carriers such as hypoxanthinephosphoribosyl transferase (HPRT), which resulted in a neurological phenotype mimicking that observed in mice expressing the mutant HD truncated protein (Ordway et al., 1997). *In vitro* studies aiming at better characterizing the structure and function of polyQ repeats in the context of full-length soluble proteins, included the insertion of ectopic polyQ stretches into well-characterized and soluble proteins such as CI2 (Stott et al., 1995; Chen et al., 1999), myoglobin (Mb) (Tanaka et al., 2001; Tobelmann and Murphy, 2011), glutathione S transferase (GST) (Masino et al., 2002; Bulone et al., 2006) and the B domain from *Staphylococcus aureus* Protein A (SpA) (Saunders et al., 2011). Fusion of the polyQ sequences with stable and soluble proteins moderates the intrinsic polyQ peptide aggregation propensity, but induces the self-assembly of carrier proteins into fibrillar amyloid-like structures, a nucleation-dependent process whose kinetics is directly proportional to the size of the inserted polyQ repeat. Likewise, polyQ peptides are able to seed the aggregation of intracellular soluble polyQ-containing proteins when added to cell cultures, conferring a heritable phenotype of self-sustaining seeding, resembling a prion-like mechanism (Ren et al., 2009), reviewed in Cushman et al. (2010).

The impact of the polyQ tract and its expansion on the perturbation of the structure of flanking sequences and domains is

critically dependent on the location of the amino acid-repeats, revealing impressive location-dependent changes in structural stability, and fibril morphology of the host proteins (Robertson et al., 2008; Saunders et al., 2011; Tobelmann and Murphy, 2011). Curiously, the studies with these model proteins showed that stability and structure of the carrier protein remained unaltered by polyQ expansion when the repeat was inserted at the N- or C-terminus of the structured domain (Robertson et al., 2008), mimicking the location of polyQ tracts in most disease-related proteins (**Figure 1**).

The role of the flanking regions in modulating protein fibril formation in polyQ disease proteins is well supported by experimental data (de Chiara et al., 2005; Gales et al., 2005; Bhattacharyya et al., 2006; Saunders and Bottomley, 2009; Tam et al., 2009; Thakur et al., 2009; Liebman and Meredith, 2010), in agreement with the knowledge that different polyQ-containing proteins have a diverse threshold for aggregation. For example, addition of a polyproline extension after the polyQ repeat slows down aggregation (Bhattacharyya et al., 2006), while protein domains outside the polyQ tract [e.g., Josephin domain (JD) of ataxin-3 and AHX domain of ataxin-1] have been shown to contribute to protein aggregation (Masino et al., 2004; de Chiara et al., 2005; Gales et al., 2005; Ellisdon et al., 2006, 2007). The multitude of data on the polyQ-induced aggregation of disease and non-disease-proteins highlights the complex interplay between the polyQ region and the adjacent protein domains. In light of the polymorphic nature of the polyQ and the modulation of its structural features by the protein context, two general mechanisms have been proposed for polyQ-mediated toxicity (Kim et al., 2009): (a) the expanded polyQ stretch adopts a novel conformation that mediates toxicity or is the precursor to toxic species; (b) intra- or intermolecular protein interactions mediated by expanded polyQ in the random coil conformation are sufficient to result in pathological effects. In both cases the affinity of the interactions involving the expanded polyQ region could be higher with selected target proteins, leading to a preference of the disease proteins for some of the protein partners, a fact that is in agreement with the hypothesis raised by Zuchner and Brundin (2008), which postulate that resistance to NMDA receptor-mediated excitotoxicity occurring in some mouse models for HD is a consequence of a differential binding of partner proteins, in a polyQ tract size dependent manner, to the proline-rich domain of huntingtin. In this context, differences in molecular interactions occurring in a cell- and tissue-specific manner would result in different toxicities according to particular cellular environments.

Given the above mentioned studies, it is nowadays clear that the polyQ region influences aggregation of proteins, but this process is highly dependent on the surrounding protein context. Therefore, even though the structural information on peptides and proteins with polyQ expansions is a useful guideline for the investigation of the pathogenic effects of polyQ expansion, each of the proteins involved in polyQ diseases shows distinctive characteristics, cellular roles, and structural properties causing difficulties in the formulation of structural hypothesis that could explain how different monomeric conformations of polyQ leads to various aggregated species and how they contribute to neurotoxicity.

PolyQ REPEATS IN ATAXIN-3 FUNCTION AND DYSFUNCTION

Machado-Joseph disease is an inherited neurodegenerative disorder of adult onset originally described in people of Portuguese Azorean descent but later shown to be the most common autosomal dominant spinocerebellar ataxia worldwide. Clinically, it is characterized by ataxia, ophthalmoplegia, and pyramidal signs, associated in variable degree with dystonia, spasticity, peripheral neuropathy, and amyotrophy (Coutinho and Andrade, 1978). Pathologically, the disorder is associated with degeneration of the deep nuclei of the cerebellum, pontine nuclei, subthalamic nuclei, substantia nigra, and spinocerebellar nuclei (Coutinho et al., 1982; Rosenberg, 1992; Margolis and Ross, 2001). It is caused by an expansion of a repetitive CAG tract within the *ATXN3* gene (Kawaguchi et al., 1994). While in the healthy population the number of CAG repeats ranges between 10 and 51, in MJD patients the length of ataxin-3 polyQ tract exceeds 55 consecutive residues. Ataxin-3 is a modular protein, located both in the nucleus and the cytoplasm (Perez et al., 1999; Antony et al., 2009; Macedo-Ribeiro et al., 2009), encompassing an N-terminal globular JD, with structural similarity to cysteine proteases (Scheel et al., 2003; Albrecht et al., 2004), followed by an extended tail composed of two ubiquitin interaction motifs (UIMs), the expandable polyQ tract, and a C-terminal region (Matos et al., 2011). The C-terminal region of ataxin-3 may contain a third UIM, depending on the splice variant (Goto et al., 1997), with the 3UIM isoform of ataxin-3 being predominantly found in the brain (Harris et al., 2010). Currently, the physiological function of ataxin-3, as well as the molecular mechanism by which expanded polyQ sequences causes selective neurodegeneration remain mostly unknown. However, since it is ubiquitously expressed and cell death is region specific, neurodegeneration is currently viewed as depending on sequence and structural features outside the ataxin-3 polyQ tract [reviewed in Matos et al. (2011) and references therein].

ATAXIN-3 BIOLOGICAL ROLES

ATXN3 orthologs have been identified in eukaryotic organisms including protozoans, plants, fungi, and animals (Albrecht et al., 2004; Costa et al., 2004; Rodrigues et al., 2007). Several functions have been ascertained to ataxin-3 based on studies with orthologs. Specifically, a role in cell structure and/or motility was proposed for mouse ataxin-3 as it is highly abundant in all types of muscle and in ciliated epithelial cells (Costa et al., 2004). In fact, ataxin-3 is able to interact with tubulin through its JD domain (Figure 4), with nM affinity (Mazzucchelli et al., 2009), which supports a role in cell structure. Interestingly, data on ataxin-3 *C. elegans* ortholog not only reinforces a function in structure/motility and signal transduction (Rodrigues et al., 2007), but also indicate a function in development as absence of *ATXN3* strongly modifies expression of several development-related genes. *ATXN3* knock-out animals showed no obvious deleterious phenotype, probably due to a putative redundant function between ataxin-3 and other JD-encoding proteins, such as ataxin-3-like protein, Josephin 1 and Josephin 2, all containing a typical cysteine protease catalytic triad. However the studies with *ATXN3* knock-out animals revealed an overall increase in the levels of ubiquitinated proteins (Schmitt et al., 2007) and signs of altered expression of core sets of genes associated with the ubiquitin-proteasome and signal transduction

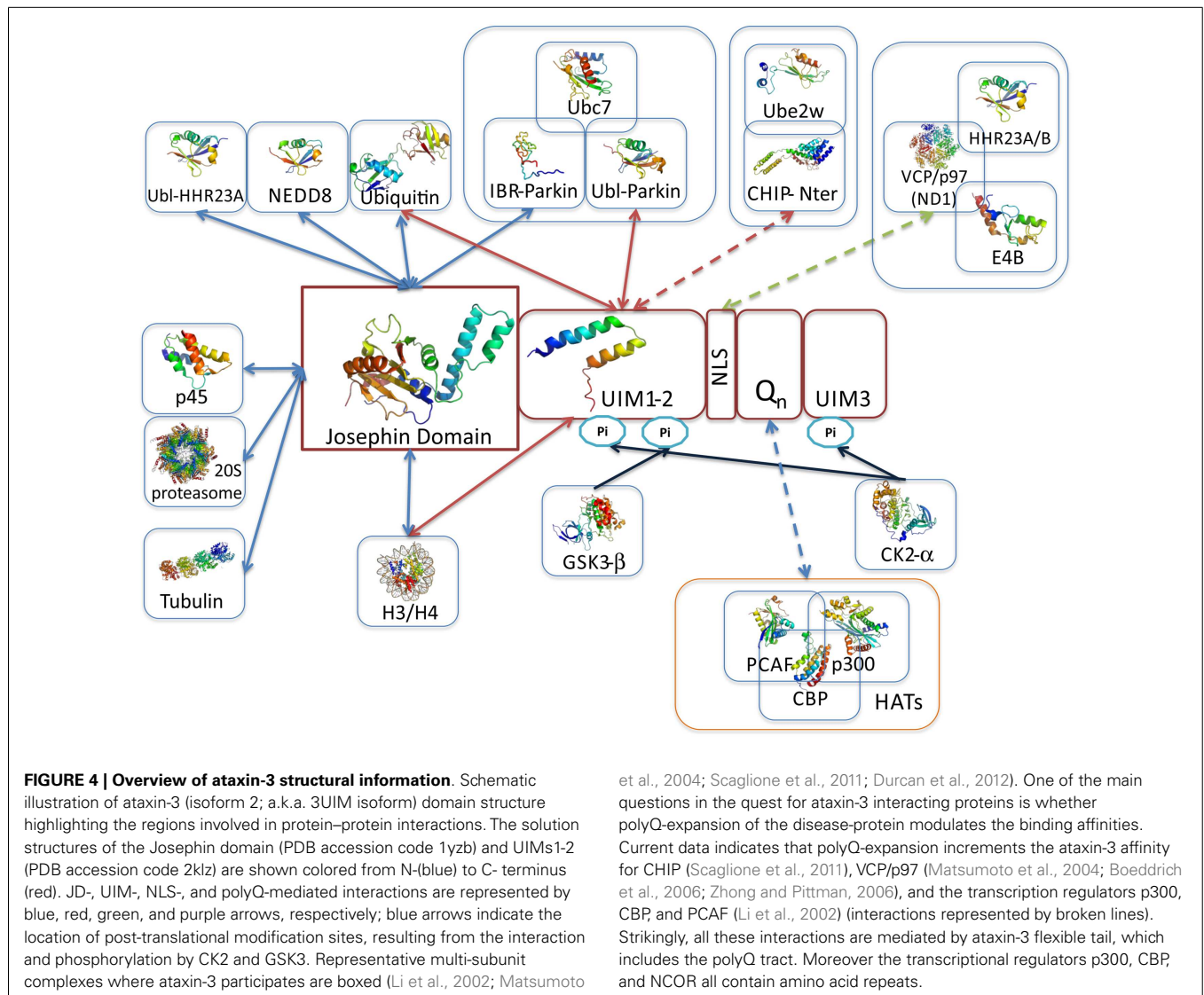
pathways (Rodrigues et al., 2007), pointing to a dual function of ataxin-3 in the ubiquitin-proteasome system and transcriptional regulation (Matos et al., 2011; Orr, 2012a).

Ataxin-3 function as transcriptional regulator

The putative role of ataxin-3 in transcriptional regulation is proposed to entail the modulation of histone acetylation and deacetylation at selected promoters. Ataxin-3 interacts with the major histone acetyltransferases cAMP-response-element binding protein (CREB)-binding protein (CBP), p300, and p300/CREB-binding protein-associated factor (KAT2B/PCAF, Figures 4 and 5), and is proposed to inhibit transcription in specific promoters (e.g., MMP-2 promoter) either by blocking access to histone acetylation sites or through recruitment of histone deacetylase 3 (HDAC3) and nuclear receptor co-repressor (NCOR1; Figures 4 and 5) (Li et al., 2002; Evert et al., 2006). Although, the interaction sites have not been mapped in detail for all these proteins, co-immunoprecipitation experiments showed that KAT2B/PCAF, p300, and CBP bind exclusively to the polyQ-containing C-terminal region of ataxin-3 (Figure 4), apparently in a polyQ-size dependent manner (Li et al., 2002). Experimental evidence also indicates that ataxin-3 forms part of a CREB-containing complex, although no direct interaction has been observed between the two proteins (Li et al., 2002). In contrast, the N-terminal region of ataxin-3 directly binds histones H3 and H4 (Table 2; Figure 4) (Li et al., 2002). Of note, p300 and CBP, as well as NCOR1, also encompass amino acid repetitions in its sequence. Interestingly, in huntingtin and in ataxin-1, polyQ interferes with CBP-activated gene transcription via interaction of their glutamine-rich domains (Shimohata et al., 2000; Nucifora et al., 2001) and mutant huntingtin targets specific components of the core transcriptional machinery, in a glutamine-tract length-sensitive manner (Zhai et al., 2005), pinpointing once again the role of the amino acid-repeat region in the establishment of protein-protein interactions.

Ataxin-3 molecular function: ubiquitin hydrolase

A role for ataxin-3 in ubiquitin-dependent pathways was proposed by bioinformatic analysis (Scheel et al., 2003; Albrecht et al., 2004), and its ability to bind and cleave poly-ubiquitin chains and polyubiquitinated proteins was later demonstrated experimentally (Burnett et al., 2003; Chai et al., 2004). Importantly, inhibition of ataxin-3 catalytic activity results in the increase of polyubiquitinated proteins, resembling the effects of proteasome inhibition (Berke et al., 2005), indicating that ataxin-3 is involved with proteins targeted for proteasomal degradation. The function of ataxin-3 in the ubiquitin-proteasome system was further supported by the identification of its association with the ubiquitin-like domain of the human homologs of the yeast DNA repair protein Rad23, HHR23A, and HHR23B (Wang et al., 2000; Doss-Pepe et al., 2003; Nicastro et al., 2005, 2009), with valosin-containing protein (VCP)/p97 (Hirabayashi et al., 2001; Doss-Pepe et al., 2003; Boeddrich et al., 2006; Zhong and Pittman, 2006), and with the ubiquitin ligase E4B (Matsumoto et al., 2004) (Figures 4 and 5). Strikingly, the weak direct association between ataxin-3 and E4B is strongly reinforced by the addition of VCP/p97, indicating that these proteins form part of a higher order macromolecular



complex to regulate the degradation of misfolded ER proteins (Matsumoto et al., 2004; Zhong and Pittman, 2006) (Figure 5).

Biochemical studies showed that ataxin-3 displays a strong preference for chains containing four or more ubiquitins (Chai et al., 2004) and that full-length ataxin-3 and its JD both display proteolytic activity toward either linear substrates containing a single ubiquitin molecule (Burnett et al., 2003; Chow et al., 2004b; Weeks et al., 2011) or K48/K63-linked poly-ubiquitin chains (Winborn et al., 2008; Todi et al., 2009), displaying also the capacity to bind the ubiquitin-like protein NEDD8 in a substrate-like fashion (Ferro et al., 2007). Moreover, ataxin-3-like protein, Josephin 1 and Josephin 2, also display ubiquitin protease activity (Tzvetkov and Breuer, 2007; Weeks et al., 2011), although the relative activities are highly variable in spite of their high sequence similarity. Characterization of ataxin-3 ubiquitin hydrolase activity has also revealed that the full-length protein preferentially cleaves Lys-63-linked and mixed-linkage chains with more than four ubiquitins (Burnett et al., 2003; Winborn et al., 2008). This specificity is dictated

by the UIMs, as the isolated JD shows a preference toward the disassembly of Lys-48-linked chains (Nicastro et al., 2009, 2010). Altogether, this indicates that ataxin-3 ubiquitin hydrolase activity is likely to be associated with delivery of target substrates to the proteasome rather than with their rescue from degradation, as it happens with most of the other deubiquitinases (Ventii and Wilkinson, 2008; Matos et al., 2011; Scaglione et al., 2011). Interestingly, ubiquitin hydrolase activity of ataxin-3 is not affected by polyQ expansion and both normal and expanded ataxin-3 are able to increase the cellular levels of a short-lived GFP normally degraded by the ubiquitin-proteasome pathway (Burnett et al., 2003).

The 3D structures for JD alone or in the presence of ubiquitin as well as that of the tandem UIM1-UIM2 have already been determined (Mao et al., 2005; Nicastro et al., 2005, 2009; Song et al., 2010), giving a structural perspective on the ubiquitin hydrolase function of ataxin-3. The JD contains two ubiquitin binding sites, both of hydrophobic nature, with site 1 being negatively charged to

Table 2 | Human ataxin-3 associated proteins.

Ataxin-3 interacting protein (UniProt accession code)	Protein name	Direct interaction?	Interaction domains		Reference
			Ataxin-3	Partner protein	
CELL-QUALITY CONTROL (PROTEIN HOMEOSTASIS)					
HHR23A/B (P54725/P54727)	UV excision repair protein RAD23 homolog A/B	Yes, kD (JD:Ub) = 12 μ M	JD	Ubiquitin-like (Ub) N-terminal domain	Wang et al. (2000), Doss-Pepe et al. (2003), Nicastro et al. (2005, 2009)
Poly-ubiquitin (P0CG48/P0CG47)	Polyubiquitin- C/Polyubiquitin-B	Yes, kD (atxn3:K48- tetraUb) = 0.2 μ M, kD (atxn3:Ub) = 50 μ M	UIMs, JD	K48- and K63-linked Ub (\geq 4 Ub), K48-linked diUb	Burnett et al. (2003), Doss-Pepe et al. (2003), Chai et al. (2004), Nicastro et al. (2009, 2010)
Ubiquilin-1 (Q9UMX0)	Protein linking IAP with cytoskeleton 1	n.d.	n.d.	n.d.	Heir et al. (2006)
NEDD8 (O15843)	Ubiquitin-like protein Nedd8	Yes	JD	NEDD8	Ferro et al. (2007)
Parkin (Q60260)	E3 ubiquitin-protein ligase parkin	Yes	JD, UIMs	IBR domain, Ubiquitin-like (Ub) domain	Durcan et al. (2011, 2012)
Ubc7 (P62253)	Ubiquitin-conjugating enzyme E2 G1	Yes (transient interaction detected using cross-linking reagents)	n.d.	n.d.	Durcan et al. (2012)
p45 (P62195)	26S proteasome regulatory subunit 8	Yes	N-terminal atxn3 region (residues 1–133)	n.d.	Wang et al. (2007)
20S Proteasome (P25786, P25787, P25788, P25789, P28066, P60900, O14818, P20618, P49721, P49720, P28070, P28074, P28072, Q99436)	Proteasome subunits α types 1-7 and β types 1-7	n.d.	N-terminal atxn3 region (residues 1–150)	n.d.	Doss-Pepe et al. (2003)
CHIP (Q9UNE7)	E3 ubiquitin-protein ligase CHIP	Yes, kD (atxn3:CHIP) = 2.2 μ M, kD (atxn3:Ub-CHIP) = 0.1 μ M	Atxn3 C-terminus (residues 133–357)	CHIP N-terminus	Jana et al. (2005), Scaglione et al. (2011)
VCP/p97 (P55072)	Transitional endoplasmic reticulum ATPase	Yes	Residues 277–281 (includes arginine/lysine-rich NLS)	N domain, residues 1–199	Hirabayashi et al. (2001), Doss-Pepe et al. (2003), Matsumoto et al. (2004, 7) Boeddrich et al. (2006), and Zhong and Pittman (2006)
E4B (O95155)	Ubiquitin conjugation factor E4 B	Yes (with 79Q-ataxin-3)	n.d.	n.d.	Matsumoto et al. (2004)

(Continued)

Table 2 | Continued

Ataxin-3 interacting protein (UniProt accession code)	Protein name	Direct interaction?	Interaction domains		Reference
			Ataxin-3	Partner protein	
OTUB2 (Q96DC9)	Ubiquitin thioesterase OTUB2	n.d.	n.d.	n.d.	Sowa et al. (2009)
USP13 (Q92995)	Ubiquitin carboxyl-terminal hydrolase 13	n.d.	n.d.	n.d.	Sowa et al. (2009)
KCTD10 (Q9H3F6)	BTB/POZ domain-containing adapter for CUL3-mediated RhoA degradation protein 3	n.d.	n.d.	n.d.	Sowa et al. (2009)
Tubulin dimer (Q71U36/P68363)	Tubulin α -1A, Tubulin β -2B	Yes, kD (atxn3:tubulin) = 50–70 nM	JD	n.d.	Mazzucchelli et al. (2009)
Dynein (Q9Y6G9)	Cytoplasmic dynein 1 light intermediate chain 1	n.d.	n.d.	n.d.	Burnett and Pittman (2005)
HDAC6 (Q9UBN7)	Histone deacetylase 6	n.d.	n.d.	n.d.	Burnett and Pittman (2005)
TRANSCRIPTIONAL REGULATION					
p300 (Q09472)	Histone acetyltransferase p300	Yes	PolyQ-containing C terminus of atxn3 (residues 288–354)	n.d.	Li et al. (2002)
CBP (Q92793)	cAMP-response-element binding protein (CREB)-binding protein	Yes	PolyQ-containing C terminus of atxn3 (residues 288–354)	n.d.	Li et al. (2002)
PCAF (Q92831)	p300/CREB-binding protein-associated factor: histone acetyltransferase KAT2B	Yes	PolyQ-containing C terminus of atxn3 (residues 288–354)	n.d.	Li et al. (2002)
Histone H3/H4 (P68431/P62805)	Histone	Yes	JD + UJM1 and 2 (residues 1–288)	n.d.	Li et al. (2002)
HDAC3 (O15379)	histone deacetylase 3	Yes	n.d.	n.d.	Evert et al. (2006)
NCOR1 (O75376)	Nuclear receptor corepressor 1	n.d.	n.d.	n.d.	Evert et al. (2006)
MAML3 (Q96JK9)	Mastermind-like protein 3	n.d.	n.d.	n.d.	Ravasi et al. (2010)
EWSR1 (Q01844)	RNA-binding protein EWS	n.d.	n.d.	n.d.	Vinayagam et al. (2011)

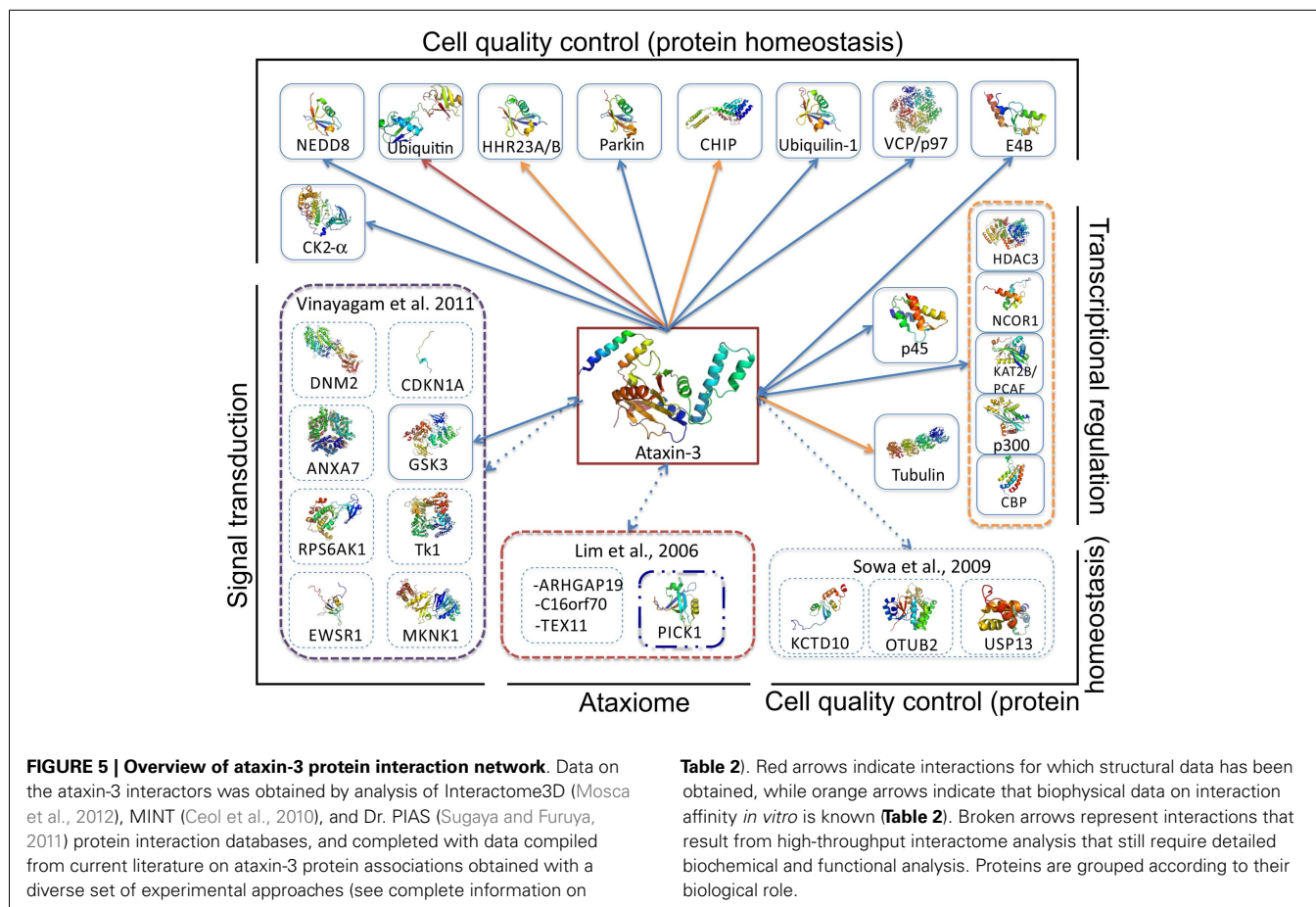
(Continued)

Table 2 | Continued

Ataxin-3 interacting protein (UniProt accession code)	Protein name	Direct interaction?	Interaction domains		Reference
			Ataxin-3	Partner protein	
SIGNAL TRANSDUCTION					
CK2 (P19784)	Casein kinase II subunit α	Yes	n.d.	n.d.	Tao et al. (2008), Mueller et al. (2009)
GSK3B (P49841)	Glycogen synthase kinase-3 β	Yes	n.d.	n.d.	Fei et al. (2007), Vinayagam et al. (2011)
DNM2 (P50570)	Dynammin-2	n.d.	n.d.	n.d.	Vinayagam et al. (2011)
CDKN1A (P38936)	Cyclin-dependent kinase inhibitor 1	n.d.	n.d.	n.d.	Vinayagam et al. (2011)
ANXA7 (P20073)	Annexin A7	n.d.	n.d.	n.d.	Vinayagam et al. (2011)
RPS6AK1 (Q15418)	Ribosomal protein S6 kinase α -1	n.d.	n.d.	n.d.	Vinayagam et al. (2011)
TK1 (P04183)	Thymidine kinase, cytosolic	n.d.	n.d.	n.d.	Vinayagam et al. (2011)
MKNK1 (Q9BUB5)	MAP kinase-interacting serine/threonine-protein kinase 1	n.d.	n.d.	n.d.	Vinayagam et al. (2011)
ATAXIOME					
TEX11 (Q81YF3)	Testis-expressed sequence 11 protein	n.d.	n.d.	n.d.	Lim et al. (2006)
C16orf70 (Q9BSU1)	UPF0183 protein C16orf70	n.d.	n.d.	n.d.	Lim et al. (2006)
ARHGAP19 (Q14CB8)	Rho GTPase-activating protein 19	n.d.	n.d.	n.d.	Lim et al. (2006)
PICK1 (Q9NRD5)	PRKCA-binding protein	n.d.	n.d.	n.d.	Lim et al. (2006)

Boxes shaded in gray represent associations identified in high-throughput interactome screenings.

Atxn3, ataxin-3; IBR, In Between Ring fingers; JD, Josephin domain; n. d., not determined; NLS, nuclear localization sequence; Ub, ubiquitin; UBA, ubiquitin associated domain; Ubl, ubiquitin-like domain; UJM, ubiquitin-interacting motifs.



facilitate docking of the positively charged ubiquitin C-terminus close to the catalytic site. Binding of ubiquitin to site 1 is of crucial importance for both JD and full-length ataxin-3 activity as ubiquitin hydrolase (Nicastro et al., 2010). Site 2 confers ubiquitin-chain linkage preference to ataxin-3 and it overlaps with the surface for interaction of the ubiquitin-like domain in HHR23B (Nicastro et al., 2005, 2010). Solution structure for the two UIMs (UIM1 and UIM2), which are separated by a short 2 amino acid spacer, revealed that they fold into two α -helices separated by a flexible linker (Song et al., 2010). Upon ubiquitin binding, this structure adopts a typical helix-loop-helix folding pattern, where hydrophobic interactions dominate the complex formation (Song et al., 2010). When in tandem, UIM1 and UIM2 show higher binding affinity for mono- or poly-ubiquitin than individual UIMs (Song et al., 2010), suggesting a cooperative binding mechanism (Song et al., 2010). The effect of the presence of UIM3 in ataxin-3 binding affinity for ubiquitin has not been shown, but its role in ubiquitin chain binding and recognition is unlikely to be of relevance to ataxin-3 activity, since no differences in proteolytic activity were identified when the 2UIM and 3UIM isoforms were compared. In the model proposed for ataxin-3 ubiquitin chain proteolysis, the UIMs (UIM1-UIM2) select and recruit poly-ubiquitin substrates, presenting them to the catalytic JD for cleavage (Mao et al., 2005).

Even though ataxin-3 functions as ubiquitin hydrolase, its proteolytic activity is rather low, indicating that either ataxin-3/JD

requires additional factors (post-translational modifications, cofactors, intracellular interactions) to exhibit significant proteolytic activity or the substrates used *in vitro* so far are not optimal. Interestingly, only three amino acid mutations are sufficient to significantly increase the proteolytic activity of ataxin-3, to a value close to that of ataxin-3-like protein (Weeks et al., 2011). Under physiological conditions, one candidate for an activating signal is mono-ubiquitination at K117, which has been shown to increase the enzyme's rate of cleavage of Lys-63 linked substrates (Todi et al., 2009). However, the molecular mechanism by which ubiquitination increases enzyme activity is not still clear, nor is it known whether other cellular signals (e.g., phosphorylation by CK2 or GSK3b; Fei et al., 2007; Tao et al., 2008) may also modulate the activity of ataxin-3. Interestingly the JD-containing protein, Josephin 1 was also demonstrated to cleave ubiquitin chains only after it is mono-ubiquitinated (Seki et al., 2013). The regulation of ataxin-3 activity through ubiquitination might depend on the interaction of ataxin-3 with several E3 ubiquitin ligases (Durcan and Fon, 2013), such as the C-terminus of 70 kDa heat-shock protein (Hsp70)-interacting protein (CHIP), parkin, and E4B (Figure 5), since all were shown to promote ataxin-3 ubiquitination and regulate its degradation by the proteasome (Matsumoto et al., 2004; Jana et al., 2005; Miller et al., 2005). Association of ataxin-3 with CHIP is a multistep process regulated by mono-ubiquitination of the N-terminal region of

CHIP by the E2-conjugating enzyme Ube2w, and occurs through the region encompassing polyQ and UIM1 and 2 (Jana et al., 2005) (Figure 4). As observed for other interactions involving the C-terminal region of ataxin-3, the ataxin-3-CHIP complex is affected by polyQ expansion and the polyQ-expanded protein displays a sixfold increase in binding affinity (Scaglione et al., 2011). The presence of ataxin-3 in multicomponent E3-ligase complexes is also supported by the identification of a direct interaction with parkin, an association that stabilizes the interaction between parkin and the E2-conjugating enzyme Ubc7 (Durcan et al., 2011). In contrast with what is observed in the ataxin-3:CHIP complex, ataxin-3 association with parkin remains unaltered by polyQ expansion (Durcan et al., 2012) (Figure 4). However, we still do not understand the mechanisms that regulate shuttling of ataxin-3 between these functional complexes or how its distribution is modulated by polyQ expansion. Further biochemical studies are required to establish the correlation between these macromolecular interactions and their relevance for ataxin-3 aggregation and neurodegeneration in MJD patients

ATAXIN-3 AGGREGATION: A MULTISTEP PATHWAY MODULATED BY THE PROTEIN CONTEXT

A characteristic hallmark of MJD and other polyQ-expansion diseases is the appearance of intracellular inclusions enriched in the disease protein and containing components from the cell-quality control machinery (e.g., ubiquitin, proteasome subunits, and chaperones), indicating that these diseases form part of the larger family of protein misfolding disorders (Williams and Paulson, 2008). Early *in vitro* studies showed that expansion of the polyQ tract within the pathological range induced formation of insoluble β -rich fibrils with the capacity to bind amyloid-specific dyes (Bevino and Loll, 2001). Later it was demonstrated that non-pathological ataxin-3 could also form insoluble fibrillar aggregates upon destabilization of its structure by temperature, pressure or denaturing agents (Marchal et al., 2003; Chow et al., 2004c). Unexpectedly, under partially destabilizing conditions the JD was also able to form insoluble fibrils, indicating that even though polyQ repeats are themselves toxic, the structure of ataxin-3, besides polyQ, has an important role in aggregation and fibril formation (Masino et al., 2004). Structural studies have shown that in ataxin-3, the addition of a polyQ tract destabilized the protein but polyQ expansion within the pathogenic range did not further induce significant structural changes (Chow et al., 2004a). In fact, alterations in ataxin-3 stability were later shown not to be a requirement for amyloid fibril formation since both non-expanded ataxin-3 and the isolated JD were shown to self-assemble and form fibrils under near-physiological conditions (Gales et al., 2005). Since kinetics of aggregation is slower for non-expanded ataxin-3 (Ellisdon et al., 2007), early events in the oligomerization pathway could be identified, with the formation of JD-mediated dimers proposed to be the first step in the ataxin-3 self-assembly pathway (Gales et al., 2005; Masino et al., 2011; Scarff et al., 2012). In a striking parallel with the classical amyloids the kinetics of fibril formation follows a nucleation-dependent polymerization mechanism, where these intermediate species might represent the most toxic species (Kayed et al., 2003; Laganowsky et al., 2012).

The current model for polyQ-expanded ataxin-3 aggregation consists of two steps. A JD-dependent step that leads to the formation of SDS-soluble protofibrils, followed by the formation of detergent-resistant fibrils similar to those found in MJD, where the polyQ-expanded tracts play a key role (Ellisdon et al., 2006, 2007). On the contrary, non-expanded polyQ ataxin-3 undergoes a single step aggregation event resulting in SDS-soluble aggregates, which does not involve the polyQ tract (Ellisdon et al., 2006) but directly depends on conformational changes outside the polyQ repeat. In the initial stages of aggregation, JD retains a native-like secondary structure, but is deployed of catalytic activity pointing to a subtle conformational change before fibril assembly (Masino et al., 2011). Although the JD plays a central role in this aggregation process, recent studies show that the ataxin-3 flexible region (Masino et al., 2003; Scarff et al., 2012), encompassing UIM1, and UIM2 also contributes to aggregation of the full-length protein (Santambrogio et al., 2012). This multistep aggregation modulated by the polyQ protein context seems to be common in other polyQ disorders such as HD (Thakur et al., 2009) and SCA1 (de Chiara et al., 2005). The presence of an expanded polyQ tract leads subsequently to rapid inter-fibril association and formation of large, highly stable amyloid-like fibrils. This indicates that the non-polyQ domains predispose ataxin-3 for aggregation and that the presence of a pathological polyQ tract introduces an additional step resulting in the formation of a highly stable amyloid-like aggregate.

This complex aggregation mechanism, involving domains outside the amino acid-repeat region, is also found in PABPN1 (Tavanez et al., 2005; Winter et al., 2013), a multi-domain polyA-containing protein consisting of an N-terminal segment containing the alanine repeat, a coiled-coil domain, a RNA recognition motif (RRM), and a C-terminal domain. As observed for ataxin-3, the propensity of PABPN1 to aggregate and form amyloid fibrils (Scheuermann et al., 2003) is not exclusive of the expanded protein and can also be observed in the non-disease protein (Tavanez et al., 2005; Rohrberg et al., 2008). This indicates that polyA-flanking regions/domains influence the aggregation process, and that this is likely a broader mechanism common in homeopeptide repeat-containing proteins. In fact, analysis of PABPN1 sequence indicates higher scores for aggregation propensity within the RRM domain (Tavanez et al., 2009), with mutations in the RRM being sufficient to prevent aggregate formation (Tavanez et al., 2005).

Aggregation of homeopeptide repeat-containing proteins is therefore a multiparametric process that culminates in cell-specific degeneration, whose toxicity might be explained by the context-dependent molecular interactions and post-translational modifications. The relation between protein interactions, function, and aggregation will be discussed below, with a particular focus on the polyQ-containing protein, ataxin-3.

ATAXIN-3 FUNCTION VS. AGGREGATION

Several studies have been focused on the search for specific ataxin-3 interactors, some of which have identified direct physical association between the protein partners and provided clues into ataxin-3 biological role (Table 2 and references herein; Figure 5). Recently, different high throughput interactome screens focused on the search for protein complexes associated with ubiquitin hydrolases

(Sowa et al., 2009), ataxia-related proteins (Lim et al., 2006), signal transduction pathways (Vinayagam et al., 2011), and transcriptional regulation (Ravasi et al., 2010), contributing with data on novel putative ataxin-3 binding proteins (Table 2; Figures 4 and 5).

An overview of the current data on the ataxin-3 interactome shows that a large number of interactions map to the catalytic JD. The intrinsic tendency of JD to self-associate involves the hydrophobic patches on its surface, which overlap with the functionally relevant ubiquitin binding sites 1 and 2 (Matos et al., 2011; Pastore and Temussi, 2012), providing a direct link between protein function and aggregation and exposing a role for intracellular interactors, such as ubiquitin, in protecting against ataxin-3 self-assembly (Gales et al., 2005; Masino et al., 2011; Matos et al., 2011; Pastore and Temussi, 2012). Likewise, in the polyA-containing protein PABPN1 the RRM domain responsible for the interaction with the mRNA (Banerjee et al., 2013) and with polyadenylate-specific RNA polymerase is also involved in the aggregation process (Winter et al., 2013). In addition, the fact that heat-shock proteins including Hsp70, and type I arginine methyl transferases (PRMT1 and PRMT3) associate preferentially with expanded PABPN1 raises the question whereas proteotoxicity of expanded PABPN1 might also be caused by altered protein networking. Destabilization of the ataxin-3 JD by specific mutations revealed that any conformational change in this region is directly linked with aggregation of the full-length protein (Saunders et al., 2011), emphasizing the putative therapeutic potential associated with the identification of macro-molecules with the ability to stabilize this N-terminal region. In fact, interaction of JD with protein partners (Masino et al., 2011) or chaperones (Robertson et al., 2010) is sufficient to induce stabilization of JD leading to a reduction in ataxin-3 self-assembling properties (Matos et al., 2011; Pastore and Temussi, 2012).

Concerning the C-terminal flexible tail of ataxin-3, the interaction with VCP/p97 is probably the one for which more experimental data is available (Hirabayashi et al., 2001; Doss-Pepe et al., 2003; Matsumoto et al., 2004; Boeddrich et al., 2006; Zhong and Pittman, 2006). This association is dependent on the arginine/lysine-rich motifs close to the polyQ tract of ataxin-3 (Boeddrich et al., 2006), and several reports point to a stronger interaction with the disease-protein containing longer polyQ stretches (Wang et al., 2000; Matsumoto et al., 2004; Boeddrich et al., 2006; Zhong and Pittman, 2006). Functional interaction with VCP/p97 is able to modulate the fibrillogenesis of a C-terminal fragment of expanded ataxin-3 (71Q) in a concentration-dependent manner, with equimolar concentrations of VCP/p97 stimulating fibrillogenesis, while a four-fold excess of VCP prevented aggregation (Boeddrich et al., 2006). This study provides clues toward the role of interactors targeting the C-terminal region of ataxin-3 as modulators of its oligomerization properties, indicating that this mostly unstructured region (Masino et al., 2003; Scarff et al., 2012) may also represent a bridge between physiological interactions, function, and aggregation. In the field of polyQ disorders, the search for protein interactors is an active area of research, uncovering novel macromolecular partners often acting as disease modifiers (Goehler et al., 2004; Kaltenbach et al., 2007; McGurk and Bonini, 2012). A recent screen for modifiers of ataxin-1 aggregation and toxicity in mammalian cells showed that toxicity enhancers often contained coiled-coil

domains. Importantly, coiled-coil formation by ataxin-3 polyQ stretch and its flanking domains were also predicted (Fiumara et al., 2010), however neither the consequences of its expansion nor its functional properties were experimentally assessed. Since coiled-coil structures are known to be involved in protein-protein interactions (Parry et al., 2008; Fiumara et al., 2010), it would be interesting to access the role of protein interactions mediated by the polyQ region of ataxin-3 in regulating its aggregation behavior (Figure 4). In fact, some of the interactors associated with ataxin-3 nuclear functions, rich in polyQ- (p300, NCOR, CBP), and polyA-repeats (NCOR), target this region and are predicted to bind better to the expanded protein (Figures 4 and 5). Therefore it is tempting to speculate that these anomalously stronger interactions with the expanded protein in the nucleus might be associated with increased aggregation and toxicity observed when a strong nuclear localization sequence (NLS) is fused to ataxin-3 (Perez et al., 1998; Bichelmeier et al., 2007; Macedo-Ribeiro et al., 2009) or when the protein shuttles to the nucleus upon increased cellular stress (Reina et al., 2010). Heat-shock induces phosphorylation of a serine residue located on ataxin-3 JD and drives the protein to the nucleus leading to its dissociation from the cytosolic VCP/HHR23A complex (Reina et al., 2010), highlighting the dynamic nature of ataxin-3 partition between macromolecular assemblies and making it tempting to speculate that polyQ expansion might affect this distribution.

CONCLUDING REMARKS

Trinucleotide repeats are keen for driving evolution by providing genetic variability, with homeopeptide-encoded regions being crucial for the establishment of protein interactions. However, as unstable regions, expansion of the homeopeptide regions might occur, being responsible for several neurodegenerative and muscular diseases. Homeopeptide repeats such as polyQ and polyA seem to influence and to drive the repeat-containing protein toward self-assembly and aggregation. On the other hand, structural studies were able to reveal that aggregation of homeopeptide-containing proteins also depends on the homeopeptide protein context with additional protein domains playing a role in a multi-domain self-assembly mechanism. Differential expression of protein isoforms generated by alternative splicing, post-translational modifications and, additionally, differences in macromolecular interactions are currently advanced as hypotheses that, by their ability to modulate protein function and aggregation, could explain the cell-specific toxicity of the homeopeptide-expanded proteins.

Ataxin-3 is an excellent example of a repeat-containing protein that, upon polyQ-expansion, does not undergo drastic structural and functional changes, but achieves an increased tendency toward self-assembly and aggregation. The protein has deubiquitinase activity and plays a role in the cell-quality control system, and in transcriptional regulation. As a result of its modular multi-domain structure, ataxin-3 engages in multiple macromolecular interactions and several evidences show that it associates with several functional multiprotein complexes, in some cases in a polyQ-dependent manner. The structural and mechanistic details regulating ataxin-3 redistribution between different cellular machineries are still unclear, although post-translational modifications of the protein subunits assembled in

these complexes are likely to play a role. Different cellular events such as proteotoxic stress or aging might unbalance the association of expanded ataxin-3 with its molecular partners and contribute to the alteration of ataxin-3 normal cellular functions. Since ataxin-3 self-assembly is a complex process that involves several protein domains, including JD, the relocalization of the protein to different complexes might induce the exposure of aggregation-prone regions and lead to the appearance of the characteristic intracellular protein inclusions.

Since macromolecular interactions seem to be either protective or exacerbate aggregation of the homopeptide-containing proteins, they might be targeted therapeutically. However, an in-depth understanding about the effect of homopeptide-expansion in the function of the containing-protein and on the interactions with molecular partners is required, in order to understand how they contribute to neurodegeneration. The combination of biochemical and computational approaches to the identification of disease-protein interaction networks is critical for defining their

normal function, identifying new markers for disease prognosis and also for the development of tools to selectively target those interactions with potentially reduced side effects.

ACKNOWLEDGMENTS

We thank Pedro J. B. Pereira and Ana Luísa Carvalho, for critical reading of the manuscript. Research in SM-R lab is funded by FEDER funds through the Operational Competitiveness Programme – Compete and by National Funds through FCT – Fundação para a Ciência e a Tecnologia under the projects FCOMP-01-0124-FEDER-022718 (PEst-C/SAU/LA0002/2011), FCOMP-01-0124-FEDER-007075 (PTDC/BIA-PRO/70455/2006), FCOMP-01-0124-FEDER-009031 (PTDC/BIA-PRO/100059/2008), and FCOMP-01-0124-FEDER-015860 (PTDC/SAU-NMC/110602/2009). Bruno Almeida, Sara Fernandes, and Isabel A. Abreu acknowledge the financial support from FCT through fellowships SFRH/BPD/70783/2010 SFRH/BPD/77009/2011, and Ciência 2008 Programme, respectively.

REFERENCES

- Alba, M. M., and Guigo, R. (2004). Comparative analysis of amino acid repeats in rodents and humans. *Genome Res.* 14, 549–554. doi:10.1101/gr.1925704
- Albrecht, M., Golatta, M., Wullner, U., and Lengauer, T. (2004). Structural and functional analysis of ataxin-2 and ataxin-3. *Eur. J. Biochem.* 271, 3155–3170. doi:10.1111/j.1432-1033.2004.04245.x
- Altschuler, E. L., Hud, N. V., Mazrimas, J. A., and Rupp, B. (1997). Random coil conformation for extended polyglutamine stretches in aqueous soluble monomeric peptides. *J. Pept. Res.* 50, 73–75. doi:10.1111/j.1399-3011.1997.tb00622.x
- Amiel, J., Trochet, D., Clement-Ziza, M., Munnich, A., and Lyonnet, S. (2004). Polyalanine expansions in human. *Hum. Mol. Genet.* 2, R235–243. doi:10.1093/hmg/ddh251
- Antony, P. M., Mantele, S., Mollenkopf, P., Boy, J., Kehlenbach, R. H., Riess, O., et al. (2009). Identification and functional dissection of localization signals within ataxin-3. *Neurobiol. Dis.* 36, 280–292. doi:10.1016/j.nbd.2009.07.020
- Arrasate, M., Mitra, S., Schweitzer, E. S., Segal, M. R., and Finkbeiner, S. (2004). Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* 431, 805–810. doi:10.1038/nature02998
- Banerjee, A., Apponi, L. H., Pavlath, G. K., and Corbett, A. H. (2013). PABPN1: molecular function and muscle disease. *FEBS J.* doi:10.1111/febs.12294
- Beckwith, A. C., Wall, J. S., and Jordan, R. W. (1965). Reversible reduction and reoxidation of the disulfide bonds in wheat gliadin. *Arch. Biochem. Biophys.* 112, 16–24. doi:10.1016/0003-9861(65)90004-4
- Berke, S. J., Chai, Y., Marrs, G. L., Wen, H., and Paulson, H. L. (2005). Defining the role of ubiquitin-interacting motifs in the polyglutamine disease protein, ataxin-3. *J. Biol. Chem.* 280, 32026–32034. doi:10.1074/jbc.M506084200
- Bevivino, A. E., and Loll, P. J. (2001). An expanded glutamine repeat destabilizes native ataxin-3 structure and mediates formation of parallel beta-fibrils. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11955–11960. doi:10.1073/pnas.211305198
- Bhattacharyya, A., Thakur, A. K., Chellgren, V. M., Thiagarajan, G., Williams, A. D., Chellgren, B. W., et al. (2006). Oligoproline effects on polyglutamine conformation and aggregation. *J. Mol. Biol.* 355, 524–535. doi:10.1016/j.jmb.2005.10.053
- Bichelmeier, U., Schmidt, T., Hubener, J., Boy, J., Ruttiger, L., Habig, K., et al. (2007). Nuclear localization of ataxin-3 is required for the manifestation of symptoms in SCA3: in vivo evidence. *J. Neurosci.* 27, 7418–7428. doi:10.1523/JNEUROSCI.4540-06.2007
- Blum, E. S., Abraham, M. C., Yoshimura, S., Lu, Y., and Shaham, S. (2012). Control of nonapoptotic developmental cell death in *Caenorhabditis elegans* by a polyglutamine-repeat protein. *Science* 335, 970–973. doi:10.1126/science.1215156
- Blum, E. S., Schwendeman, A. R., and Shaham, S. (2013). PolyQ disease: misfiring of a developmental cell death program? *Trends Cell Biol.* 23, 168–174. doi:10.1016/j.tcb.2012.11.003
- Boeddrich, A., Gaumer, S., Haacke, A., Tzvetkov, N., Albrecht, M., Evert, B. O., et al. (2006). An arginine/lysine-rich motif is crucial for VCP/p97-mediated modulation of ataxin-3 fibrillogenesis. *EMBO J.* 25, 1547–1558. doi:10.1038/sj.emboj.7601043
- Brais, B. (2003). Oculopharyngeal muscular dystrophy: a late-onset polyalanine disease. *Cytogenet. Genome Res.* 100, 252–260. doi:10.1159/000072861
- Brouwer, J. R., Willemsen, R., and Oostra, B. A. (2009). Microsatellite repeat instability and neurological disease. *Bioessays* 31, 71–83. doi:10.1002/bies.080122
- Bulone, D., Masino, L., Thomas, D. J., San Biagio, P. L., and Pastore, A. (2006). The interplay between PolyQ and protein context delays aggregation by forming a reservoir of protofibrils. *PLoS ONE* 1:e111. doi:10.1371/journal.pone.0000111
- Burnett, B., Li, F., and Pittman, R. N. (2003). The polyglutamine neurodegenerative protein ataxin-3 binds polyubiquitylated proteins and has ubiquitin protease activity. *Hum. Mol. Genet.* 12, 3195–3205. doi:10.1093/hmg/ddg344
- Burnett, B. G., and Pittman, R. N. (2005). The polyglutamine neurodegenerative protein ataxin 3 regulates aggresome formation. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4330–4335. doi:10.1073/pnas.0407252102
- Butland, S. L., Devon, R. S., Huang, Y., Mead, C. L., Meynert, A. M., Neal, S. J., et al. (2007). CAG-encoded polyglutamine length polymorphism in the human genome. *BMC Genomics* 8:126. doi:10.1186/1471-2164-8-126
- Ceol, A., Chatr Aryamontri, A., Licata, L., Peluso, D., Briganti, L., Perfetto, L., et al. (2010). MINT, the molecular interaction database: 2009 update. *Nucleic Acids Res.* 38, D532–539. doi:10.1093/nar/gkp983
- Chai, Y., Berke, S. S., Cohen, R. E., and Paulson, H. L. (2004). Poly-ubiquitin binding by the polyglutamine disease protein ataxin-3 links its normal function to protein surveillance pathways. *J. Biol. Chem.* 279, 3605–3611. doi:10.1074/jbc.M310939200
- Chen, S., Berthelie, V., Hamilton, J. B., O’nuallain, B., and Wetzel, R. (2002a). Amyloid-like features of polyglutamine aggregates and their assembly kinetics. *Biochemistry* 41, 7391–7399. doi:10.1021/bi011772q
- Chen, S., Ferrone, F. A., and Wetzel, R. (2002b). Huntington’s disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc. Natl. Acad. Sci. U.S.A.* 99, 11884–11889. doi:10.1073/pnas.182276099
- Chen, Y. W., Stott, K., and Perutz, M. F. (1999). Crystal structure of a dimeric chymotrypsin inhibitor 2 mutant containing an inserted glutamine repeat. *Proc. Natl. Acad. Sci. U.S.A.* 96, 1257–1261. doi:10.1073/pnas.96.4.1257
- Chow, M. K., Ellisdon, A. M., Cabrita, L. D., and Bottomley, S. P. (2004a). Polyglutamine expansion in ataxin-3 does not affect protein stability: implications for misfolding and disease. *J. Biol. Chem.* 279, 47643–47651. doi:10.1074/jbc.M405799200
- Chow, M. K., Mackay, J. P., Whistock, J. C., Scanlon, M. J., and Bottomley, S. P. (2004b). Structural and functional analysis of the

- Josephin domain of the polyglutamine protein ataxin-3. *Biochem. Biophys. Res. Commun.* 322, 387–394. doi:10.1016/j.bbrc.2004.07.131
- Chow, M. K., Paulson, H. L., and Bottomley, S. P. (2004c). Destabilization of a non-pathological variant of ataxin-3 results in fibrillogenesis via a partially folded intermediate: a model for misfolding in polyglutamine disease. *J. Mol. Biol.* 335, 333–341. doi:10.1016/j.jmb.2003.08.064
- Chu-Wang, I. W., and Oppenheim, R. W. (1978). Cell death of motoneurons in the chick embryo spinal cord. I. A light and electron microscopic study of naturally occurring and induced cell loss during development. *J. Comp. Neurol.* 177, 33–57. doi:10.1002/cne.901770105
- Costa, M. C., Gomes-Da-Silva, J., Miranda, C. J., Sequeiros, J., Santos, M. M., and Maciel, P. (2004). Genomic structure, promoter activity, and developmental expression of the mouse homologue of the Machado-Joseph disease (MJD) gene. *Genomics* 84, 361–373. doi:10.1016/j.ygeno.2004.02.012
- Costa Mdo, C., and Paulson, H. L. (2012). Toward understanding Machado-Joseph disease. *Prog. Neurobiol.* 97, 239–257. doi:10.1016/j.pneurobio.2011.11.006
- Coutinho, P., and Andrade, C. (1978). Autosomal dominant system degeneration in Portuguese families of the Azores Islands. A new genetic disorder involving cerebellar, pyramidal, extrapyramidal and spinal cord motor functions. *Neurology* 28, 703–709. doi:10.1212/WNL.28.7.703
- Coutinho, P., Guimaraes, A., and Scaravilli, F. (1982). The pathology of Machado-Joseph disease. Report of a possible homozygous case. *Acta Neuropathol.* 58, 48–54. doi:10.1007/BF00692697
- Crick, S. L., Jayaraman, M., Frieden, C., Wetzel, R., and Pappu, R. V. (2006). Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc. Natl. Acad. Sci. U.S.A.* 103, 16764–16769. doi:10.1073/pnas.0608175103
- Cushman, M., Johnson, B. S., King, O. D., Gitler, A. D., and Shorter, J. (2010). Prion-like disorders: blurring the divide between transmissibility and infectivity. *J. Cell. Sci.* 123, 1191–1201. doi:10.1242/jcs.051672
- Davies, S. W., Turmaine, M., Cozens, B. A., Difiglia, M., Sharp, A. H., Ross, C. A., et al. (1997). Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell* 90, 537–548. doi:10.1016/S0092-8674(00)80513-9
- de Chiara, C., Menon, R. P., Dal Pia, F., Calder, L., and Pastore, A. (2005). Polyglutamine is not all: the functional role of the AXH domain in the ataxin-1 protein. *J. Mol. Biol.* 354, 883–893. doi:10.1016/j.jmb.2005.09.083
- de Mezer, M., Wojciechowska, M., Napierala, M., Sobczak, K., and Krzyzosiak, W. J. (2011). Mutant CAG repeats of Huntingtin transcript fold into hairpins, form nuclear foci and are targets for RNA interference. *Nucleic Acids Res.* 39, 3852–3863. doi:10.1093/nar/gkq1323
- Doss-Pepe, E. W., Stenroos, E. S., Johnson, W. G., and Madura, K. (2003). Ataxin-3 interactions with rad23 and valosin-containing protein and its associations with ubiquitin chains and the proteasome are consistent with a role in ubiquitin-mediated proteolysis. *Mol. Cell Biol.* 23, 6469–6483. doi:10.1128/MCB.23.18.6469-6483.2003
- Dosztanyi, Z., Chen, J., Dunker, A. K., Simon, I., and Tompa, P. (2006). Disorder and sequence repeats in hub proteins and their implications for network evolution. *J. Proteome Res.* 5, 2985–2995. doi:10.1021/pr060171o
- Dougan, L., Li, J., Badilla, C. L., Berne, B. J., and Fernandez, J. M. (2009). Single homopolymer peptide chains collapse into mechanically rigid conformations. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12605–12610. doi:10.1073/pnas.0900678106
- Durcan, T. M., and Fon, E. A. (2013). Ataxin-3 and its e3 partners: implications for Machado-Joseph disease. *Front. Neurol.* 4:46. doi:10.3389/fneur.2013.00046
- Durcan, T. M., Kontogianna, M., Bedard, N., Wing, S. S., and Fon, E. A. (2012). Ataxin-3 deubiquitination is coupled to Parkin ubiquitination via E2 ubiquitin-conjugating enzyme. *J. Biol. Chem.* 287, 531–541. doi:10.1074/jbc.M111.288449
- Durcan, T. M., Kontogianna, M., Thorarindottir, T., Fallon, L., Williams, A. J., Djarmati, A., et al. (2011). The Machado-Joseph disease-associated mutant form of ataxin-3 regulates parkin ubiquitination and stability. *Hum. Mol. Genet.* 20, 141–154. doi:10.1093/hmg/ddq452
- Ellisdon, A. M., Pearce, M. C., and Bottomley, S. P. (2007). Mechanisms of ataxin-3 misfolding and fibril formation: kinetic analysis of a disease-associated polyglutamine protein. *J. Mol. Biol.* 368, 595–605. doi:10.1016/j.jmb.2007.02.058
- Ellisdon, A. M., Thomas, B., and Bottomley, S. P. (2006). The two-stage pathway of ataxin-3 fibrillogenesis involves a polyglutamine-independent step. *J. Biol. Chem.* 281, 16888–16896. doi:10.1074/jbc.M601470200
- Evert, B. O., Araujo, J., Vieira-Saecker, A. M., De Vos, R. A., Harendza, S., Klockgether, T., et al. (2006). Ataxin-3 represses transcription via chromatin binding, interaction with histone deacetylase 3, and histone deacetylation. *J. Neurosci.* 26, 11474–11486. doi:10.1523/JNEUROSCI.2053-06.2006
- Faux, N. G., Bottomley, S. P., Lesk, A. M., Irving, J. A., Morrison, J. R., De La Banda, M. G., et al. (2005). Functional insights from the distribution and role of homopeptide repeat-containing proteins. *Genome Res.* 15, 537–551. doi:10.1101/gr.3096505
- Fei, E., Jia, N., Zhang, T., Ma, X., Wang, H., Liu, C., et al. (2007). Phosphorylation of ataxin-3 by glycogen synthase kinase 3beta at serine 256 regulates the aggregation of ataxin-3. *Biochem. Biophys. Res. Commun.* 357, 487–492. doi:10.1016/j.bbrc.2007.03.160
- Ferro, A., Carvalho, A. L., Teixeira-Castro, A., Almeida, C., Tome, R. J., Cortes, L., et al. (2007). NEDD8: a new ataxin-3 interactor. *Biochim. Biophys. Acta* 1773, 1619–1627. doi:10.1016/j.bbamcr.2007.07.012
- Fiumara, F., Fioriti, L., Kandel, E. R., and Hendrickson, W. A. (2010). Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell* 143, 1121–1135. doi:10.1016/j.cell.2010.11.042
- Fondon, J. W. III, and Garner, H. R. (2004). Molecular origins of rapid and continuous morphological evolution. *Proc. Natl. Acad. Sci. U.S.A.* 101, 18058–18063. doi:10.1073/pnas.0408118101
- Friedman, M. J., Shah, A. G., Fang, Z. H., Ward, E. G., Warren, S. T., Li, S., et al. (2007). Polyglutamine domain modulates the TBP-TFIIB interaction: implications for its normal function and neurodegeneration. *Nat. Neurosci.* 10, 1519–1528. doi:10.1038/nn2011
- Gales, L., Cortes, L., Almeida, C., Melo, C. V., Costa, M. C., Maciel, P., et al. (2005). Towards a structural understanding of the fibrillization pathway in Machado-Joseph's disease: trapping early oligomers of non-expanded ataxin-3. *J. Mol. Biol.* 353, 642–654. doi:10.1016/j.jmb.2005.08.061
- Gerber, H. P., Seipel, K., Georgiev, O., Hofferer, M., Hugg, M., Rusconi, S., et al. (1994). Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* 263, 808–811. doi:10.1126/science.8303297
- Goehler, H., Lalowski, M., Stelzl, U., Waelter, S., Stroedicke, M., Worm, U., et al. (2004). A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease. *Mol. Cell* 15, 853–865. doi:10.1016/j.molcel.2004.09.016
- Gojobori, J., and Ueda, S. (2011). Elevated evolutionary rate in genes with homopolymeric amino acid repeats constituting nondisordered structure. *Mol. Biol. Evol.* 28, 543–550. doi:10.1093/molbev/msq225
- Gonitell, R., Moffitt, H., Sathasivam, K., Woodman, B., Detloff, P. J., Faull, R. L., et al. (2008). DNA instability in postmitotic neurons. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3467–3472. doi:10.1073/pnas.0800048105
- Goto, J., Watanabe, M., Ichikawa, Y., Yee, S. B., Ihara, N., Endo, K., et al. (1997). Machado-Joseph disease gene products carrying different carboxyl termini. *Neurosci. Res.* 28, 373–377. doi:10.1016/S0168-0102(97)00056-4
- Guo, L., Han, A., Bates, D. L., Cao, J., and Chen, L. (2007). Crystal structure of a conserved N-terminal domain of histone deacetylase 4 reveals functional insights into glutamine-rich domains. *Proc. Natl. Acad. Sci. U.S.A.* 104, 4297–4302. doi:10.1073/pnas.0608041104
- Haerty, W., and Golding, G. B. (2010). Low-complexity sequences and single amino acid repeats: not just “junk” peptide sequences. *Genome* 53, 753–762. doi:10.1139/g10-063
- Hammock, E. A., and Young, L. J. (2005). Microsatellite instability generates diversity in brain and sociobehavioral traits. *Science* 308, 1630–1634. doi:10.1126/science.1111427
- Hancock, J. M., and Simon, M. (2005). Simple sequence repeats in proteins and their significance for network evolution. *Gene* 345, 113–118. doi:10.1016/j.gene.2004.11.023
- Harris, G. M., Dodelzon, K., Gong, L., Gonzalez-Alegre, P., and Paulson, H. L. (2010). Splice isoforms of the polyglutamine disease protein ataxin-3 exhibit similar enzymatic yet different aggregation

- properties. *PLoS ONE* 5:e13695. doi:10.1371/journal.pone.0013695
- Heir, R., Ablasou, C., Dumontier, E., Elliott, M., Fagotto-Kaufmann, C., and Bedford, F. K. (2006). The UBL domain of PLIC-1 regulates aggresome formation. *EMBO Rep.* 7, 1252–1258. doi:10.1038/sj.embor.7400823
- Hirabayashi, M., Inoue, K., Tanaka, K., Nakadate, K., Ohsawa, Y., Kamei, Y., et al. (2001). VCP/p97 in abnormal protein aggregates, cytoplasmic vacuoles, and cell death, phenotypes relevant to neurodegeneration. *Cell Death Differ.* 8, 977–984. doi:10.1038/sj.cdd.4400907
- Huang, C. C., Faber, P. W., Persichetti, F., Mittal, V., Vonsattel, J. P., Macdonald, M. E., et al. (1998). Amyloid formation by mutant huntingtin: threshold, progressivity and recruitment of normal polyglutamine proteins. *Somat. Cell Mol. Genet.* 24, 217–233. doi:10.1023/B:SCAM.0000007124.19463.e5
- Huntley, M. A., and Golding, G. B. (2002). Simple sequences are rare in the Protein Data Bank. *Proteins* 48, 134–140. doi:10.1002/prot.10150
- Jana, N. R., Dikshit, P., Goswami, A., Kotliarova, S., Murata, S., Tanaka, K., et al. (2005). Co-chaperone CHIP associates with expanded polyglutamine protein and promotes their degradation by proteasomes. *J. Biol. Chem.* 280, 11635–11640. doi:10.1074/jbc.M412042200
- Jayaraman, M., Thakur, A. K., Kar, K., Kodali, R., and Wetzel, R. (2011). Assays for studying nucleated aggregation of polyglutamine proteins. *Methods* 53, 246–254. doi:10.1016/j.ymeth.2011.01.001
- Kaltenbach, L. S., Romero, E., Becklin, R. R., Chettier, R., Bell, R., Phansalkar, A., et al. (2007). Huntingtin interacting proteins are genetic modifiers of neurodegeneration. *PLoS Genet.* 3:e82. doi:10.1371/journal.pgen.0030082
- Kar, K., Jayaraman, M., Sahoo, B., Kodali, R., and Wetzel, R. (2011). Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent. *Nat. Struct. Mol. Biol.* 18, 328–336. doi:10.1038/nsmb.1992
- Kashi, Y., King, D., and Soller, M. (1997). Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet.* 13, 74–78. doi:10.1016/S0168-9525(97)01008-1
- Kashi, Y., and King, D. G. (2006). Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.* 22, 253–259. doi:10.1016/j.tig.2006.03.005
- Kawaguchi, Y., Okamoto, T., Tanikawa, M., Aizawa, M., Inoue, M., Katayama, S., et al. (1994). CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat. Genet.* 8, 221–228. doi:10.1038/ng1194-221
- Kayed, R., Head, E., Thompson, J. L., McIntire, T. M., Milton, S. C., Cotman, C. W., et al. (2003). Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* 300, 486–489. doi:10.1126/science.1079469
- Kiliszek, A., Kierzek, R., Krzyzosiak, W. J., and Rypniewski, W. (2010). Atomic resolution structure of CAG RNA repeats: structural insights and implications for the trinucleotide repeat expansion diseases. *Nucleic Acids Res.* 38, 8370–8376. doi:10.1093/nar/gkq700
- Kiliszek, A., Kierzek, R., Krzyzosiak, W. J., and Rypniewski, W. (2011). Crystal structures of CGG RNA repeats with implications for fragile X-associated tremor ataxia syndrome. *Nucleic Acids Res.* 39, 7308–7315. doi:10.1093/nar/gkr368
- Kim, M. W., Chelliah, Y., Kim, S. W., Otwinowski, Z., and Bezprozvanny, I. (2009). Secondary structure of Huntingtin amino-terminal region. *Structure* 17, 1205–1212. doi:10.1016/j.str.2009.08.002
- King, D. G. (1994). Triple repeat DNA as a highly mutable regulatory mechanism. *Science* 263, 595–596. doi:10.1126/science.263.5147.595-b
- Klein, F. A., Pastore, A., Masino, L., Zeder-Lutz, G., Nierengarten, H., Oulad-Abdelghani, M., et al. (2007). Pathogenic and non-pathogenic polyglutamine tracts have similar structural properties: towards a length-dependent toxicity gradient. *J. Mol. Biol.* 371, 235–244. doi:10.1016/j.jmb.2007.05.028
- Korschen, H. G., Beyermann, M., Muller, F., Heck, M., Vantler, M., Koch, K. W., et al. (1999). Interaction of glutamic-acid-rich proteins with the cGMP signalling pathway in rod photoreceptors. *Nature* 400, 761–766. doi:10.1038/23468
- Kratter, I. H., and Finkbeiner, S. (2010). PolyQ disease: too many Qs, too much function? *Neuron* 67, 897–899. doi:10.1016/j.neuron.2010.09.012
- Krull, L. H., Wall, J. S., Zobel, H., and Dimler, R. J. (1965). Synthetic polypeptides containing side-chain amide groups: water-insoluble polymers. *Biochemistry* 4, 626–633. doi:10.1021/bi00880a003
- Krzyzosiak, W. J., Sobczak, K., Wojciechowska, M., Fiszer, A., Mykowska, A., and Kozlowski, P. (2012). Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic Acids Res.* 40, 11–26. doi:10.1093/nar/gkr729
- Kumar, A., Park, H., Fang, P., Parkesh, R., Guo, M., Nettles, K. W., et al. (2011). Myotonic dystrophy type 1 RNA crystal structures reveal heterogeneous 1×1 nucleotide UU internal loop conformations. *Biochemistry* 50, 9928–9935. doi:10.1021/bi2013068
- La Spada, A. R. (1997). Trinucleotide repeat instability: genetic features and molecular mechanisms. *Brain Pathol.* 7, 943–963. doi:10.1111/j.1750-3639.1997.tb00895.x
- La Spada, A. R., Paulson, H. L., and Fischbeck, K. H. (1994). Trinucleotide repeat expansion in neurological disease. *Ann. Neurol.* 36, 814–822. doi:10.1002/ana.410360604
- La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E., and Fischbeck, K. H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 352, 77–79. doi:10.1038/352077a0
- Laganowsky, A., Liu, C., Sawaya, M. R., Whitelegge, J. P., Park, J., Zhao, M., et al. (2012). Atomic view of a toxic amyloid small oligomer. *Science* 335, 1228–1231. doi:10.1126/science.1213151
- Li, F., Macfarlan, T., Pittman, R. N., and Chakravarti, D. (2002). Ataxin-3 is a histone-binding protein with two independent transcriptional corepressor activities. *J. Biol. Chem.* 277, 45004–45012. doi:10.1074/jbc.M205259200
- Li, L. B., and Bonini, N. M. (2010). Roles of trinucleotide-repeat RNA in neurological disease and degeneration. *Trends Neurosci.* 33, 292–298. doi:10.1016/j.tins.2010.03.004
- Li, P., Huey-Tubman, K. E., Gao, T., Li, X., West, A. P. Jr., Bennett, M. J., et al. (2007a). The structure of a polyQ-anti-polyQ complex reveals binding according to a linear lattice model. *Nat. Struct. Mol. Biol.* 14, 381–387. doi:10.1038/nsmb.1234
- Li, X. J., Friedman, M., and Li, S. (2007b). Interacting proteins as genetic modifiers of Huntington disease. *Trends Genet.* 23, 531–533. doi:10.1016/j.tig.2007.07.007
- Liebman, S. W., and Meredith, S. C. (2010). Protein folding: sticky N17 speeds huntingtin pile-up. *Nat. Chem. Biol.* 6, 7–8. doi:10.1038/nchembio.279
- Lim, J., Crespo-Barreto, J., Jafar-Nejad, P., Bowman, A. B., Richman, R., Hill, D. E., et al. (2008). Opposing effects of polyglutamine expansion on native protein complexes contribute to SCA1. *Nature* 452, 713–718. doi:10.1038/nature06731
- Lim, J., Hao, T., Shaw, C., Patel, A. J., Szabo, G., Rual, J. F., et al. (2006). A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell* 125, 801–814. doi:10.1016/j.cell.2006.03.032
- Liu, Y., Gotte, G., Libonati, M., and Eisenberg, D. (2001). A domain-swapped RNase A dimer with implications for amyloid formation. *Nat. Struct. Biol.* 8, 211–214. doi:10.1038/84941
- Lopez Castel, A., Cleary, J. D., and Pearson, C. E. (2010). Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol.* 11, 165–170. doi:10.1038/nrm2854
- Macedo-Ribeiro, S., Cortes, L., Maciel, P., and Carvalho, A. L. (2009). Nucleocytoplasmic shuttling activity of ataxin-3. *PLoS ONE* 4:e5834. doi:10.1371/journal.pone.0005834
- Mao, Y., Senic-Matuglia, F., Di Fiore, P. P., Polo, S., Hodsdon, M. E., and De Camilli, P. (2005). Deubiquitinating function of ataxin-3: insights from the Josephin domain. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12700–12705. doi:10.1073/pnas.0506344102
- Marchal, S., Shehi, E., Harricane, M. C., Fusi, P., Heitz, F., Tortora, P., et al. (2003). Structural instability and fibrillar aggregation of non-expanded human ataxin-3 revealed under high pressure and temperature. *J. Biol. Chem.* 278, 31554–31563. doi:10.1074/jbc.M304205200
- Margolis, R. L., and Ross, C. A. (2001). Expansion explosion: new clues to the pathogenesis of repeat expansion neurodegenerative diseases. *Trends Mol. Med.* 7, 479–482. doi:10.1016/S1471-4914(01)02179-7
- Masino, L., Kelly, G., Leonard, K., Trotter, Y., and Pastore, A. (2002). Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins. *FEBS Lett.* 513, 267–272. doi:10.1016/S0014-5793(02)02335-9
- Masino, L., Musi, V., Menon, R. P., Fusi, P., Kelly, G., Frenkiel, T. A., et al. (2003). Domain architecture of the polyglutamine protein ataxin-3: a globular domain followed by a flexible tail. *FEBS*

- Lett.* 549, 21–25. doi:10.1016/S0014-5793(03)00748-8
- Masino, L., Nicastrò, G., Calder, L., Vendruscolo, M., and Pastore, A. (2011). Functional interactions as a survival strategy against abnormal aggregation. *FASEB J.* 25, 45–54. doi:10.1096/fj.10-161208
- Masino, L., Nicastrò, G., Menon, R. P., Dal Piaz, F., Calder, L., and Pastore, A. (2004). Characterization of the structure and the amyloidogenic properties of the Josephin domain of the polyglutamine-containing protein ataxin-3. *J. Mol. Biol.* 344, 1021–1035. doi:10.1016/j.jmb.2004.09.065
- Matos, C. A., De Macedo-Ribeiro, S., and Carvalho, A. L. (2011). Polyglutamine diseases: the special case of ataxin-3 and Machado-Joseph disease. *Prog. Neurobiol.* 95, 26–48. doi:10.1016/j.pneurobio.2011.06.007
- Matsumoto, M., Yada, M., Hatakeyama, S., Ishimoto, H., Tanimura, T., Tsuji, S., et al. (2004). Molecular clearance of ataxin-3 is regulated by a mammalian E4. *EMBO J.* 23, 659–669. doi:10.1038/sj.emboj.7600081
- Mazzucchi, S., De Palma, A., Riva, M., D'urzo, A., Pozzi, C., Pastori, V., et al. (2009). Proteomic and biochemical analyses unveil tight interaction of ataxin-3 with tubulin. *Int. J. Biochem. Cell Biol.* 41, 2485–2492. doi:10.1016/j.biocel.2009.08.003
- McGurk, L., and Bonini, N. M. (2012). Protein interacting with C kinase (PICK1) is a suppressor of spinocerebellar ataxia 3-associated neurodegeneration in *Drosophila*. *Hum. Mol. Genet.* 21, 76–84. doi:10.1093/hmg/ddr439
- Merienne, K., and Trotter, Y. (2009). SCA8 CAG/CTG expansions, a tale of two TOXICities: a unique or common case? *PLoS Genet.* 5:e1000593. doi:10.1371/journal.pgen.1000593
- Messaed, C., Dion, P. A., Abu-Baker, A., Rochefort, D., Laganiere, J., Brais, B., et al. (2007). Soluble expanded PABPN1 promotes cell death in oculopharyngeal muscular dystrophy. *Neurobiol. Dis.* 26, 546–557. doi:10.1016/j.nbd.2007.02.004
- Messaed, C., and Rouleau, G. A. (2009). Molecular mechanisms underlying polyalanine diseases. *Neurobiol. Dis.* 34, 397–405. doi:10.1016/j.nbd.2009.02.013
- Miller, J., Arrasate, M., Brooks, E., Libeu, C. P., Legleiter, J., Hatters, D., et al. (2011). Identifying polyglutamine protein species in situ that best predict neurodegeneration. *Nat. Chem. Biol.* 7, 925–934. doi:10.1038/nchembio.694
- Miller, J., Arrasate, M., Shaby, B. A., Mitra, S., Masliyah, E., and Finkbeiner, S. (2010). Quantitative relationships between huntingtin levels, polyglutamine length, inclusion body formation, and neuronal death provide novel insight into huntingtin's disease molecular pathogenesis. *J. Neurosci.* 30, 10541–10550. doi:10.1523/JNEUROSCI.0146-10.2010
- Miller, J. W., Urbinati, C. R., Teng-Ummuay, P., Stenberg, M. G., Byrne, B. J., Thornton, C. A., et al. (2000). Recruitment of human muscle-blind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J.* 19, 4439–4448. doi:10.1093/emboj/19.17.4439
- Miller, V. M., Nelson, R. F., Gouvion, C. M., Williams, A., Rodriguez-Lebron, E., Harper, S. Q., et al. (2005). CHIP suppresses polyglutamine aggregation and toxicity in vitro and in vivo. *J. Neurosci.* 25, 9152–9161. doi:10.1523/JNEUROSCI.3001-05.2005
- Mooers, B. H., Logue, J. S., and Berglund, J. A. (2005). The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. *Proc. Natl. Acad. Sci. U.S.A.* 102, 16626–16631. doi:10.1073/pnas.0505873102
- Mosca, R., Ceol, A., and Aloy, P. (2012). Interactome3D: adding structural details to protein networks. *Nat. Methods* 10, 47–53. doi:10.1038/nmeth.2289
- Mueller, T., Breuer, P., Schmitt, I., Walter, J., Evert, B. O., and Wullner, U. (2009). CK2-dependent phosphorylation determines cellular localization and stability of ataxin-3. *Hum. Mol. Genet.* 18, 3334–3343. doi:10.1093/hmg/ddp274
- Nagai, Y., Inui, T., Popiel, H. A., Fujikake, N., Hasegawa, K., Urade, Y., et al. (2007). A toxic monomeric conformer of the polyglutamine protein. *Nat. Struct. Mol. Biol.* 14, 332–340. doi:10.1038/nsmb1215
- Napierala, M., and Krzyzosiak, W. J. (1997). CUG repeats present in myotonin kinase RNA form metastable “slippery” hairpins. *J. Biol. Chem.* 272, 31079–31085. doi:10.1074/jbc.272.49.31079
- Nicastrò, G., Masino, L., Esposito, V., Menon, R. P., De Simone, A., Fraternali, F., et al. (2009). Josephin domain of ataxin-3 contains two distinct ubiquitin-binding sites. *Biopolymers* 91, 1203–1214. doi:10.1002/bip.21210
- Nicastrò, G., Menon, R. P., Masino, L., Knowles, P. P., McDonald, N. Q., and Pastore, A. (2005). The solution structure of the Josephin domain of ataxin-3: structural determinants for molecular recognition. *Proc. Natl. Acad. Sci. U.S.A.* 102, 10493–10498. doi:10.1073/pnas.0501732102
- Nicastrò, G., Todi, S. V., Karaca, E., Bonvin, A. M., Paulson, H. L., and Pastore, A. (2010). Understanding the role of the Josephin domain in the PolyUb binding and cleavage properties of ataxin-3. *PLoS ONE* 5:e12430. doi:10.1371/journal.pone.0012430
- Nithianantharajah, J., and Hannan, A. J. (2007). Dynamic mutations as digital genetic modulators of brain development, function and dysfunction. *Bioessays* 29, 525–535. doi:10.1002/bies.20589
- Nucifora, F. C. Jr., Sasaki, M., Peters, M. F., Huang, H., Cooper, J. K., Yamada, M., et al. (2001). Interference by huntingtin and atrophin-1 with cbp-mediated transcription leading to cellular toxicity. *Science* 291, 2423–2428. doi:10.1126/science.1056784
- Oma, Y., Kino, Y., Sasagawa, N., and Ishiura, S. (2005). Comparative analysis of the cytotoxicity of homopolymeric amino acids. *Biochim. Biophys. Acta* 1748, 174–179. doi:10.1016/j.bbapap.2004.12.017
- Oma, Y., Kino, Y., Toriumi, K., Sasagawa, N., and Ishiura, S. (2007). Interactions between homopolymeric amino acids (HPAAs). *Protein Sci.* 16, 2195–2204. doi:10.1110/ps.072955307
- Orday, J. M., Tallaksengreen, S., Gutekunst, C. A., Bernstein, E. M., Cearley, J. A., Wiener, H. W., et al. (1997). Ectopically expressed CAG repeats cause intranuclear inclusions and a progressive late onset neurological phenotype in the mouse. *Cell* 91, 753–763. doi:10.1016/S0092-8674(00)80464-X
- Orr, H. T. (2012a). Cell biology of spinocerebellar ataxia. *J. Cell Biol.* 197, 167–177. doi:10.1083/jcb.201105092
- Orr, H. T. (2012b). Polyglutamine neurodegeneration: expanded glutamines enhance native functions. *Curr. Opin. Genet. Dev.* 22, 251–255. doi:10.1016/j.gde.2012.01.001
- Orr, H. T., and Zoghbi, H. Y. (2007). Trinucleotide repeat disorders. *Annu. Rev. Neurosci.* 30, 575–621. doi:10.1146/annurev.neuro.29.051605.113042
- Parkesh, R., Fountain, M., and Disney, M. D. (2011). NMR spectroscopy and molecular dynamics simulation of r(CCGCUGCGG) reveal a dynamic UU internal loop found in myotonic dystrophy type 1. *Biochemistry* 50, 599–601. doi:10.1021/bi101896j
- Parry, D. A., Fraser, R. D., and Squire, J. M. (2008). Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure. *J. Struct. Biol.* 163, 258–269. doi:10.1016/j.jsb.2008.01.016
- Pastore, A., and Temussi, P. A. (2012). The two faces of Janus: functional interactions and protein aggregation. *Curr. Opin. Struct. Biol.* 22, 30–37. doi:10.1016/j.sbi.2011.11.007
- Paulson, H. L., Perez, M. K., Trotter, Y., Trojanowski, J. Q., Subramony, S. H., Das, S. S., et al. (1997). Intranuclear inclusions of expanded polyglutamine protein in spinocerebellar ataxia type 3. *Neuron* 19, 333–344. doi:10.1016/S0896-6273(00)80943-5
- Pearson, C. E. (2011). Repeat associated non-ATG translation initiation: one DNA, two transcripts, seven reading frames, potentially nine toxic entities! *PLoS Genet.* 7:e1002018. doi:10.1371/journal.pgen.1002018
- Pearson, C. E., Nichol Edamura, K., and Cleary, J. D. (2005). Repeat instability: mechanisms of dynamic mutations. *Nat. Rev. Genet.* 6, 729–742. doi:10.1038/nrg1689
- Perez, M. K., Paulson, H. L., Pendse, S. J., Saionz, S. J., Bonini, N. M., and Pittman, R. N. (1998). Recruitment and the role of nuclear localization in polyglutamine-mediated aggregation. *J. Cell Biol.* 143, 1457–1470. doi:10.1083/jcb.143.6.1457
- Perez, M. K., Paulson, H. L., and Pittman, R. N. (1999). Ataxin-3 with an altered conformation that exposes the polyglutamine domain is associated with the nuclear matrix. *Hum. Mol. Genet.* 8, 2377–2385. doi:10.1093/hmg/8.13.2377
- Perutz, M. (1994). Polar zippers: their role in human disease. *Protein Sci.* 3, 1629–1637. doi:10.1002/pro.5560031002
- Peters-Libeu, C., Miller, J., Rutenber, E., Newhouse, Y., Krishnan, P., Cheung, K., et al. (2012). Disease-associated polyglutamine stretches in monomeric huntingtin adopt a compact structure. *J. Mol. Biol.* 421, 587–600. doi:10.1016/j.jmb.2012.01.034
- Petrakis, S., Rasko, T., Russ, J., Friedrich, R. P., Stroedicke, M., Riechers, S. P., et al. (2012). Identification of human proteins that modify misfolding and proteotoxicity of pathogenic

- ataxin-1. *PLoS Genet.* 8:e1002897. doi:10.1371/journal.pgen.1002897
- Pilar, G., and Landmesser, L. (1976). Ultrastructural differences during embryonic cell death in normal and peripherally deprived ciliary ganglia. *J. Cell Biol.* 68, 339–356. doi:10.1083/jcb.68.2.339
- Ravasi, T., Suzuki, H., Cannistraci, C. V., Katayama, S., Bajic, V. B., Tan, K., et al. (2010). An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* 140, 744–752. doi:10.1016/j.cell.2010.01.044
- Reina, C. P., Zhong, X., and Pittman, R. N. (2010). Proteotoxic stress increases nuclear localization of ataxin-3. *Hum. Mol. Genet.* 19, 235–249. doi:10.1093/hmg/ddp482
- Ren, P. H., Lauckner, J. E., Kachirskaja, I., Heuser, J. E., Melki, R., and Kopito, R. R. (2009). Cytoplasmic penetration and persistent infection of mammalian cells by polyglutamine aggregates. *Nat. Cell Biol.* 11, 219–225. doi:10.1038/ncb1830
- Robertson, A. L., Bate, M. A., Androulakis, S. G., Bottomley, S. P., and Buckle, A. M. (2011). PolyQ: a database describing the sequence and domain context of polyglutamine repeats in proteins. *Nucleic Acids Res.* 39, D272–D276. doi:10.1093/nar/gkq1100
- Robertson, A. L., Headey, S. J., Saunders, H. M., Ecroyd, H., Scanlon, M. J., Carver, J. A., et al. (2010). Small heat-shock proteins interact with a flanking domain to suppress polyglutamine aggregation. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10424–10429. doi:10.1073/pnas.0914773107
- Robertson, A. L., Horne, J., Ellisdon, A. M., Thomas, B., Scanlon, M. J., and Bottomley, S. P. (2008). The structural impact of a polyglutamine tract is location-dependent. *Biophys. J.* 95, 5922–5930. doi:10.1529/biophysj.108.138487
- Rodrigues, A. J., Coppola, G., Santos, C., Costa Mdo, C., Ailion, M., Sequeiros, J., et al. (2007). Functional genomics and biochemical characterization of the Machado-Joseph disease protein ataxin-3. *FASEB J.* 21, 1126–1136. doi:10.1096/fj.06-7002com
- Rohrberg, J., Sachs, R., Lodderstedt, G., Sackewitz, M., Balbach, J., and Schwarz, E. (2008). Monitoring fibril formation of the N-terminal domain of PABPN1 carrying an alanine repeat by tryptophan fluorescence and real-time NMR. *FEBS Lett.* 582, 1587–1592. doi:10.1016/j.febslet.2008.04.002
- Rosenberg, R. N. (1992). Machado-Joseph disease: an autosomal dominant motor system degeneration. *Mov. Disord.* 7, 193–203. doi:10.1002/mds.870070302
- Rub, U., De Vos, R. A., Brunt, E. R., Sebesteny, T., Schols, L., Auburger, G., et al. (2006). Spinocerebellar ataxia type 3 (SCA3): thalamic neurodegeneration occurs independently from thalamic ataxin-3 immunopositive neuronal intranuclear inclusions. *Brain Pathol.* 16, 218–227. doi:10.1111/j.1750-3639.2006.00022.x
- Sambashivan, S., Liu, Y., Sawaya, M. R., Gingery, M., and Eisenberg, D. (2005). Amyloid-like fibrils of ribonuclease A with three-dimensional domain-swapped and native-like structure. *Nature* 437, 266–269. doi:10.1038/nature03916
- Santambrogio, C., Frana, A. M., Natalello, A., Papaleo, E., Regonesi, M. E., Doglia, S. M., et al. (2012). The role of the central flexible region on the aggregation and conformational properties of human ataxin-3. *FEBS J.* 279, 451–463. doi:10.1111/j.1742-4658.2011.08438.x
- Sathasivam, K., Lane, A., Legleiter, J., Warley, A., Woodman, B., Finkbeiner, S., et al. (2010). Identical oligomeric and fibrillar structures captured from the brains of R6/2 and knock-in mouse models of Huntington's disease. *Hum. Mol. Genet.* 19, 65–78. doi:10.1093/hmg/ddp467
- Saunders, H. M., and Bottomley, S. P. (2009). Multi-domain misfolding: understanding the aggregation pathway of polyglutamine proteins. *Protein Eng. Des. Sel.* 22, 447–451. doi:10.1093/protein/gzp033
- Saunders, H. M., Gilis, D., Rومان, M., Dehouck, Y., Robertson, A. L., and Bottomley, S. P. (2011). Flanking domain stability modulates the aggregation kinetics of a polyglutamine disease protein. *Protein Sci.* 20, 1675–1681. doi:10.1002/pro.698
- Scaglione, K. M., Zavadzky, E., Todi, S. V., Patury, S., Xu, P., Rodriguez-Lebron, E., et al. (2011). Ube2w and ataxin-3 coordinately regulate the ubiquitin ligase CHIP. *Mol. Cell* 43, 599–612. doi:10.1016/j.molcel.2011.05.036
- Scarff, C. A., Sicorello, A., Tomé, R. J. L., Ashcroft, A. E., Radford, S. E., and Macedo-Ribeiro, S. (2012). A tale of a tail: structural insights into the conformational properties of the polyglutamine protein ataxin-3. *Int. J. Mass Spectrom.* doi:10.1016/j.ijms.2012.08.032
- Schaefer, M. H., Wanker, E. E., and Andrade-Navarro, M. A. (2012). Evolution and function of CAG/polyglutamine repeats in protein-protein interaction networks. *Nucleic Acids Res.* 40, 4273–4287. doi:10.1093/nar/gks011
- Scheel, H., Tomiuk, S., and Hofmann, K. (2003). Elucidation of ataxin-3 and ataxin-7 function by integrative bioinformatics. *Hum. Mol. Genet.* 12, 2845–2852. doi:10.1093/hmg/ddg297
- Scheuermann, T., Schulz, B., Blume, A., Wähle, E., Rudolph, R., and Schwarz, E. (2003). Trinucleotide expansions leading to an extended poly-L-alanine segment in the poly (A) binding protein PABPN1 cause fibril formation. *Protein Sci.* 12, 2685–2692. doi:10.1110/ps.03214703
- Schmitt, I., Linden, M., Khazneh, H., Evert, B. O., Breuer, P., Klockgether, T., et al. (2007). Inactivation of the mouse Atxn3 (ataxin-3) gene increases protein ubiquitination. *Biochem. Biophys. Res. Commun.* 362, 734–739. doi:10.1016/j.bbrc.2007.08.062
- Seki, T., Gong, L., Williams, A. J., Sakai, N., Todi, S. V., and Paulson, H. L. (2013). JosD1, a membrane-targeted deubiquitinating enzyme, is activated by ubiquitination and regulates membrane dynamics, cell motility and endocytosis. *J. Biol. Chem.* doi:10.1074/jbc.M113.463406
- Shimohata, T., Nakajima, T., Yamada, M., Uchida, C., Onodera, O., Naruse, S., et al. (2000). Expanded polyglutamine stretches interact with TAFIII130, interfering with CREB-dependent transcription. *Nat. Genet.* 26, 29–36. doi:10.1038/79139
- Sicot, G., Gourdon, G., and Gomes-Pereira, M. (2011). Myotonic dystrophy, when simple repeats reveal complex pathogenic entities: new findings and future challenges. *Hum. Mol. Genet.* 20, R116–123. doi:10.1093/hmg/ddr343
- Simon, M., and Hancock, J. M. (2009). Tandem and cryptic amino acid repeats accumulate in disordered regions of proteins. *Genome Biol.* 10, R59. doi:10.1186/gb-2009-10-6-r59
- Singh, V. R., and Lapidus, L. J. (2008). The intrinsic stiffness of polyglutamine peptides. *J. Phys. Chem. B* 112, 13172–13176. doi:10.1021/jp805636p
- Song, A. X., Zhou, C. J., Peng, Y., Gao, X. C., Zhou, Z. R., Fu, Q. S., et al. (2010). Structural transformation of the tandem ubiquitin-interacting motifs in ataxin-3 and their cooperative interactions with ubiquitin chains. *PLoS ONE* 5:e13202. doi:10.1371/journal.pone.0013202
- Sowa, M. E., Bennett, E. J., Gygi, S. P., and Harper, J. W. (2009). Defining the human deubiquitinating enzyme interaction landscape. *Cell* 138, 389–403. doi:10.1016/j.cell.2009.04.042
- Stott, K., Blackburn, J. M., Butler, P. J., and Perutz, M. (1995). Incorporation of glutamine repeats makes protein oligomerize: implications for neurodegenerative diseases. *Proc. Natl. Acad. Sci. U.S.A.* 92, 6509–6513. doi:10.1073/pnas.92.14.6509
- Sugaya, N., and Furuya, T. (2011). Dr. PIAS: an integrative system for assessing the druggability of protein-protein interactions. *BMC Bioinformatics* 12:50. doi:10.1186/1471-2105-12-50
- Takahashi, T., Kikuchi, S., Katada, S., Nagai, Y., Nishizawa, M., and Onodera, O. (2008). Soluble polyglutamine oligomers formed prior to inclusion body formation are cytotoxic. *Hum. Mol. Genet.* 17, 345–356. doi:10.1093/hmg/ddm311
- Tam, S., Spiess, C., Auyeung, W., Joachimiak, L., Chen, B., Poirier, M. A., et al. (2009). The chaperonin TRiC blocks a huntingtin sequence element that promotes the conformational switch to aggregation. *Nat. Struct. Mol. Biol.* 16, 1279–1285. doi:10.1038/nsmb.1700
- Tanaka, M., Morishima, I., Akagi, T., Hashikawa, T., and Nukina, N. (2001). Intra- and intermolecular beta-pleated sheet formation in glutamine-repeat inserted myoglobin as a model for polyglutamine diseases. *J. Biol. Chem.* 276, 45470–45475. doi:10.1074/jbc.M107502200
- Tao, R. S., Fei, E. K., Ying, Z., Wang, H. F., and Wang, G. H. (2008). Casein kinase 2 interacts with and phosphorylates ataxin-3. *Neurosci. Bull.* 24, 271–277. doi:10.1007/s12264-008-0605-5
- Tavanez, J. P., Bengoechea, R., Berciano, M. T., Lafarga, M., Carmo-Fonseca, M., and Enguita, F. J. (2009). Hsp70 chaperones and type I PRMTs are sequestered at intranuclear inclusions caused by polyalanine expansions in PABPN1. *PLoS ONE* 4:e6418. doi:10.1371/journal.pone.0006418
- Tavanez, J. P., Calado, P., Braga, J., Lafarga, M., and Carmo-Fonseca, M. (2005). In vivo aggregation properties of the nuclear poly(A)-binding protein PABPN1. *RNA* 11, 752–762. doi:10.1261/rna.7217105
- Teplova, M., and Patel, D. J. (2008). Structural insights into RNA recognition by the alternative-splicing

- regulator muscleblind-like MBNL1. *Nat. Struct. Mol. Biol.* 15, 1343–1351. doi:10.1038/nsmb.1519
- Thakur, A. K., Jayaraman, M., Mishra, R., Thakur, M., Chellgren, V. M., Byeon, I. J., et al. (2009). Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism. *Nat. Struct. Mol. Biol.* 16, 380–389. doi:10.1038/nsmb.1570
- Tobelman, M. D., and Murphy, R. M. (2011). Location trumps length: polyglutamine-mediated changes in folding and aggregation of a host protein. *Biophys. J.* 100, 2773–2782. doi:10.1016/j.bpj.2011.04.028
- Todi, S. V., Winborn, B. J., Scaglione, K. M., Blount, J. R., Travis, S. M., and Paulson, H. L. (2009). Ubiquitination directly enhances activity of the deubiquitinating enzyme ataxin-3. *EMBO J.* 28, 372–382. doi:10.1038/emboj.2008.289
- Tzvetkov, N., and Breuer, P. (2007). Josephin domain-containing proteins from a variety of species are active de-ubiquitination enzymes. *Biol. Chem.* 388, 973–978. doi:10.1515/BC.2007.107
- Ventii, K. H., and Wilkinson, K. D. (2008). Protein partners of deubiquitinating enzymes. *Biochem. J.* 414, 161–175. doi:10.1042/BJ20080798
- Vinayagam, A., Stelzl, U., Foulle, R., Plassmann, S., Zenkner, M., Timm, J., et al. (2011). A directed protein interaction network for investigating intracellular signal transduction. *Sci. Signal.* 4, rs8. doi:10.1126/scisignal.2001699
- Vitalis, A., Lyle, N., and Pappu, R. V. (2009). Thermodynamics of beta-sheet formation in polyglutamine. *Biophys. J.* 97, 303–311. doi:10.1016/j.bpj.2009.05.003
- Walters, R. H., and Murphy, R. M. (2009). Examining polyglutamine peptide length: a connection between collapsed conformations and increased aggregation. *J. Mol. Biol.* 393, 978–992. doi:10.1016/j.jmb.2009.08.034
- Walters, R. H., and Murphy, R. M. (2011). Aggregation kinetics of interrupted polyglutamine peptides. *J. Mol. Biol.* 412, 505–519. doi:10.1016/j.jmb.2011.07.003
- Walum, H., Westberg, L., Henningsson, S., Neiderhiser, J. M., Reiss, D., Igl, W., et al. (2008). Genetic variation in the vasopressin receptor 1a gene (AVPR1A) associates with pair-bonding behavior in humans. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14153–14156. doi:10.1073/pnas.0803081105
- Wang, G., Sawai, N., Kotliarova, S., Kanazawa, I., and Nukina, N. (2000). Ataxin-3, the MJD1 gene product, interacts with the two human homologs of yeast DNA repair protein RAD23, HHR23A and HHR23B. *Hum. Mol. Genet.* 9, 1795–1803. doi:10.1093/hmg/9.12.1795
- Wang, H., Jia, N., Fei, E., Wang, Z., Liu, C., Zhang, T., et al. (2007). p45, an ATPase subunit of the 19S proteasome, targets the polyglutamine disease protein ataxin-3 to the proteasome. *J. Neurochem.* 101, 1651–1661. doi:10.1111/j.1471-4159.2007.04460.x
- Weeks, S. D., Grasty, K. C., Hernandez-Cuebas, L., and Loll, P. J. (2011). Crystal structure of a Josephin-ubiquitin complex: evolutionary restraints on ataxin-3 deubiquitinating activity. *J. Biol. Chem.* 286, 4555–4565. doi:10.1074/jbc.M110.177360
- Weiss, A., Klein, C., Woodman, B., Sathasivam, K., Bibbel, M., Regulier, E., et al. (2008). Sensitive biochemical aggregate detection reveals aggregation onset before symptom development in cellular and murine models of Huntington's disease. *J. Neurochem.* 104, 846–858.
- Wetzel, R. (2012). Physical chemistry of polyglutamine: intriguing tales of a monotonous sequence. *J. Mol. Biol.* 421, 466–490. doi:10.1016/j.jmb.2012.01.030
- Whan, V., Hobbs, M., Mcwilliam, S., Lynn, D. J., Lutzow, Y. S., Khatkar, M., et al. (2010). Bovine proteins containing poly-glutamine repeats are often polymorphic and enriched for components of transcriptional regulatory complexes. *BMC Genomics* 11:654. doi:10.1186/1471-2164-11-654
- Williams, A. J., and Paulson, H. L. (2008). Polyglutamine neurodegeneration: protein misfolding revisited. *Trends Neurosci.* 31, 521–528. doi:10.1016/j.tins.2008.07.004
- Williamson, T. E., Vitalis, A., Crick, S. L., and Pappu, R. V. (2010). Modulation of polyglutamine conformations and dimer formation by the N-terminus of huntingtin. *J. Mol. Biol.* 396, 1295–1309. doi:10.1016/j.jmb.2009.12.017
- Winborn, B. J., Travis, S. M., Todi, S. V., Scaglione, K. M., Xu, P., Williams, A. J., et al. (2008). The deubiquitinating enzyme ataxin-3, a polyglutamine disease protein, edits Lys63 linkages in mixed linkage ubiquitin chains. *J. Biol. Chem.* 283, 26436–26443. doi:10.1074/jbc.M803692200
- Winter, R., Liebold, J., and Schwarz, E. (2013). The unresolved puzzle why alanine extensions cause disease. *Biol. Chem.* doi:10.1515/hsz-2013-0112
- Yuan, Y., Compton, S. A., Sobczak, K., Stenberg, M. G., Thornton, C. A., Griffith, J. D., et al. (2007). Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res.* 35, 5474–5486. doi:10.1093/nar/gkm601
- Zhai, W., Jeong, H., Cui, L., Krainc, D., and Tjian, R. (2005). In vitro analysis of huntingtin-mediated transcriptional repression reveals multiple transcription factor targets. *Cell* 123, 1241–1253. doi:10.1016/j.cell.2005.10.030
- Zhong, X., and Pittman, R. N. (2006). Ataxin-3 binds VCP/p97 and regulates retrotranslocation of ERAD substrates. *Hum. Mol. Genet.* 15, 2409–2420. doi:10.1093/hmg/ddl164
- Zuchner, T., and Brundin, P. (2008). Mutant huntingtin can paradoxically protect neurons from death. *Cell Death Differ.* 15, 435–442. doi:10.1038/sj.cdd.4402261

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 08 March 2013; accepted: 04 June 2013; published online: 20 June 2013.

Citation: Almeida B, Fernandes S, Abreu IA and Macedo-Ribeiro S (2013) Trinucleotide repeats: a structural perspective. *Front. Neurol.* 4:76. doi: 10.3389/fneur.2013.00076

This article was submitted to *Frontiers in Neurodegeneration*, a specialty of *Frontiers in Neurology*.

Copyright © 2013 Almeida, Fernandes, Abreu and Macedo-Ribeiro. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.