Polyglutamine disease proteins: Commonalities and differences in interaction profiles and pathological effects

Megan Bonsor | Orchid Ammar | Sigrid Schnoegl | Erich E. Wanker | Eduardo Silva Ramos

Department of Neuroproteomics, Max Delbrück Center for Molecular Medicine, Berlin, Germany

Correspondence
Erich Wanker and Eduardo Silva Ramos, Department of Neuroproteomics, Max Delbrück Center for Molecular Medicine, Robert-Rösse-Straße 10, 13125 Berlin, Germany. Email: erich.w@mdc-berlin.de and eduardo.silvaramos@mdc-berlin.de

Funding information
Huntington’s Disease Society of America; Berman-Topper Family HD Career Development Fellowship; CHDI Foundation

Abstract
Currently, nine polyglutamine (polyQ) expansion diseases are known. They include spinocerebellar ataxias (SCA1, 2, 3, 6, 7, 17), spinal and bulbar muscular atrophy (SBMA), dentatorubral-pallidoluysian atrophy (DRPLA), and Huntington’s disease (HD). At the root of these neurodegenerative diseases are trinucleotide repeat mutations in coding regions of different genes, which lead to the production of proteins with elongated polyQ tracts. While the causative proteins differ in structure and molecular mass, the expanded polyQ domains drive pathogenesis in all these diseases. PolyQ tracts mediate the association of proteins leading to the formation of protein complexes involved in gene expression regulation, RNA processing, membrane trafficking, and signal transduction. In this review, we discuss commonalities and differences among the nine polyQ proteins focusing on their structure and function as well as the pathological features of the respective diseases. We present insights from AlphaFold-predicted structural models and discuss the biological roles of polyQ-containing proteins. Lastly, we explore reported protein–protein interaction networks to highlight shared protein interactions and their potential relevance in disease development.

Abbreviations: AR, androgen receptor; ASO, anti-sense oligonucleotide; ATAC, Ada-two-A-containing; ATN1, atrophin 1; ATXN1, ataxin 1; ATXN2, ataxin 2; ATXN3, ataxin 3; ATXN7, ataxin 7; BAIAP2, brain-specific angiogenesis inhibitor 1-associated protein 2; BECN1, beclin 1; CACNA1A, calcium voltage-gated channel subunit alpha 1 A (Cav2.1); CACNB1, calcium voltage-gated channel subunit alpha 1 B; CACNA1C, calcium voltage-gated channel subunit alpha 1 C; CAG, cytosine–adenine–guanine; CALM1, calmodulin 1; CALM2, calmodulin 2; CALM3, calmodulin 3; CBP, CREB-binding protein or CREBBP; CIC, capicua; cryo-EM, cryo-electron microscopy; C-terminal, carboxyl terminal; EP400, E3A binding protein 400; F8A1, 40-kD a huntingtin-associated protein; GAPDH, glyceraldehyde-3-phosphatereductase; G3BP1, GTPase-activating protein-binding protein 1; GGN5, general control non-depressible 5; GTF, general transcription factor; HSF1, heat shock protein 1; HSP90, heat shock protein 90; HSP90AA1, heat shock protein 90 family A member 1; HSP90AB1, heat shock protein 90 family A member 1; HTT, huntingtin; IP3, inositol triphosphate receptor; ITPR1, inositol 1,4,5-trisphosphate receptor type 1; KAT2B, lysine acetyltransferase 2B; KDM1A, lysine-specific histone demethylase 1A; LBD, ligand binding domain; LIG3, DNA ligase 3; LSD1, lysine-specific histone demethylase 1; MDM2, mouse double minute 2 homolog; MED15, mediator complex subunit 15; mHTT, huntingtin-N-terminal fragments; MN1, MN proto-oncogene, transcriptional regulator; MSH3, MutS homolog 3; NCOA1, nuclear receptor corepressor 1; NCOA2, nuclear receptor corepressor 2; NCOA4, nuclear receptor coactivator 1; NCOA4, nuclear receptor coactivator 4; NS15, nuclear localization sequences; NURF1, nuclear receptor subfamily 1 group D member 1; p53, tumor protein p53; PABPC1, poly(A)-binding protein 1; pLDOT, per-residue model confidence score; POLR2A, RNA polymerase II subunit A; polyQ, polyglutamine; POLR2F, POLR2G, POLR2H, POU class 3 homeobox 2; PPI, protein–protein interaction; Rab5, Rab5A, member RAS oncogene family; RAD23A, RAD23 homolog A, nucleotide excision repair protein; RAD23B, RAD23 homolog B, nucleotide excision repair protein; RBFFOX1, RNA binding fox-1 homolog 1; RBP, RNA-binding protein; RNP, ribonucleoprotein; SBMA, spinal and bulbar muscular atrophy; SCA, spinocerebellar ataxia; SETD2, SET domain-containing 2; SFGF29, SAGA complex associated factor 29; SH3, SRC homology 3 domain; SH3GL2, SH3 domain containing GRB2 like 2, endophilin A3; SNAI, single nucleotide polymorphism; SP1, Sp1 transcription factor; SRC, SRC proto-oncogene, non-receptor tyrosine kinase; STAGA, SPT3-TAF1I21-GCN5L. acetylase; TADA3, transcriptional adaptor 3; TADA2A, transcriptional adaptor 2A; TAF, TAF-related factor; TATA-box binding protein associated factor 9; TAF12, TATA-box binding protein associated factor 12; TAF5L, TATA-box binding protein associated factor 5L; TBP, TATA-binding protein; TCEG1, transcription elongation regulator 1; TFIIB, transcription factor IB; TIA1, TIA1 cytosolic granule associated RNA-binding protein; TRRAP, transformation/transcription domain-associated protein; VMA1, vesicular monoamine transporter.

Megan Bonsor and Orchid Ammar contributed equally.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Authors. Proteomics published by Wiley-VCH GmbH.
INTRODUCTION

Polyglutamine (polyQ) diseases are a group of neurodegenerative diseases predominately caused by cytosine–adenine–guanine (CAG) trinucleotide repeat expansion mutations in distinct gene coding regions [1]. They include six spinocerebellar ataxia (SCA) types (SCA1, SCA2, SCA3, SCA6, SCA7, and SCA17) [2, 3], spinal and bulbar muscular atrophy (SBMA) [4], dentatorubral-pallidoluysian atrophy (DRPLA) [5], and Huntington’s disease (HD) [6]. In the first half of this review, we will discuss genetic, clinical, and pathogenic aspects of all nine polyQ diseases. In the second half, we will focus on protein–protein interaction (PPI) networks associated with polyQ-containing disease proteins to elucidate commonalities and differences among disease processes, and to predict disease-driving mechanisms.

1.1 CAG repeat expansion mutations in protein-coding regions of genes cause neurodegeneration and disease

A common feature of polyQ diseases is that the CAG expansions in the respective disease genes are located in coding regions. This leads to the production of the polyQ-expanded proteins ataxin 1 (ATXN1), ataxin 2 (ATXN2), ataxin 3 (ATXN3), calcium voltage-gated channel subunit alpha1 A (Cav2.1; CACNA1A), ataxin 7 (ATXN7), TATA-binding protein (TBP), androgen receptor (AR), atrophin 1 (ATN1), and huntingtin (HTT), in both neuronal and non-neuronal cells (Table 1). However, the nine genes differ in their normal CAG repeat lengths, as well as their pathogenic lengths (Table 1). In most of the affected genes, a CAG repeat length of up to 30 CAGs is tolerated without disease manifestation. Interestingly, however, the genes CACNA1A and ATXN7 can only tolerate between 4 and 18 CAGs without the appearance of disease symptoms (Table 1), suggesting that even relatively short polyQ tracts in disease proteins can induce neurotoxicity. An important common feature of all polyQ diseases is the strong correlation between the length of CAG repeats and disease onset. The longer the CAG repeat tract, the earlier the disease onset and the more rapid the manifestation of symptoms [7, 8]. In the case of HD, it was recently reported that the number of uninterrupted CAGs indeed determines the timing of disease onset [9].

1.2 PolyQ diseases are rare, inherited diseases

PolyQ diseases are epidemiologically rare, affecting less than 50 individuals per 100,000. They all have autosomal dominant inheritance, apart from SBMA, which is X-linked [8]. SCAs can result from CAG repeat expansions, as well as from other genetic mutations and non-genetic causes. To this day, 40 different SCA subtypes have been described [10]. Pathologically expanded polyQ tracts account for 36%–82% of all SCA cases with genetic mutations. Globally, SCA3 is the most common SCA subtype, accounting for 21% of all cases [11]. DRPLA,

TABLE 1 Details of polyQ repeat expansion diseases, the affected proteins, protein symbols, and the range of non-pathogenic and pathogenic repeat lengths.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Affected protein</th>
<th>Protein symbol</th>
<th>Normal repeat length</th>
<th>Pathogenic repeat length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spinocerebellar ataxia 1 (SCA1)</td>
<td>Ataxin 1</td>
<td>ATXN1</td>
<td>6–34</td>
<td>39–88</td>
</tr>
<tr>
<td>Spinocerebellar ataxia 2 (SCA2)</td>
<td>Ataxin 2</td>
<td>ATXN2</td>
<td>14–31</td>
<td>32–77</td>
</tr>
<tr>
<td>Spinocerebellar ataxia 3 (SCA3)</td>
<td>Ataxin 3</td>
<td>ATXN3</td>
<td>12–40</td>
<td>55–86</td>
</tr>
<tr>
<td>Spinocerebellar ataxia 6 (SCA6)</td>
<td>Calcium voltage-gated channel subunit alpha1 A (Cav2.1)</td>
<td>CACNA1A</td>
<td>4–18</td>
<td>21–33</td>
</tr>
<tr>
<td>Spinocerebellar ataxia 7 (SCA7)</td>
<td>Ataxin 7</td>
<td>ATXN7</td>
<td>7–18</td>
<td>37–200</td>
</tr>
<tr>
<td>Spinocerebellar ataxia 17 (SCA17)</td>
<td>TATA-box binding protein</td>
<td>TBP</td>
<td>25–43</td>
<td>45–63</td>
</tr>
<tr>
<td>Spinal and bulbar muscular atrophy (SBMA)</td>
<td>Androgen receptor</td>
<td>AR</td>
<td>6–36</td>
<td>38–70</td>
</tr>
<tr>
<td>Dentatorubral-pallidoluysian atrophy (DRPLA)</td>
<td>Atrophin-1</td>
<td>ATN1</td>
<td>3–38</td>
<td>49–88</td>
</tr>
<tr>
<td>Huntington’s disease (HD)</td>
<td>Huntingtin</td>
<td>HTT</td>
<td>6–35</td>
<td>36–121</td>
</tr>
</tbody>
</table>

Adapted from [7] and with results from [300]. ATXN1, ataxin 1; ATXN2, ataxin 2; ATXN3, ataxin 3; CACNA1A, calcium voltage-gated channel subunit alpha1 A (Cav2.1); ATXN7, ataxin 7; TBP, TATA-binding protein; ATN1, atrophin 1; HTT, huntingtin.
PolyQ diseases vary in age of onset

The average age of onset for most polyQ diseases is around 30 years of age [19–21]. However, cases of early or late onset have also been reported. For example, DRPLA patients with disease onset at 15–19 years [22] and also at ~60 years have been described. The average age of onset of SBMA is in the mid-40s, but the disease may manifest as early as 18 and as late as 64 [23]. In HD patients, onset typically occurs between 30 and 50 years of age, but cases of juvenile HD with an onset as early as 18 months of age have been reported, usually when the CAG repeat is drastically expanded [24]. In some rare cases, individuals with long CAG repeats may still exhibit a late onset, suggesting that other genetic and/or environmental factors can significantly influence disease onset. Recent genome-wide association studies (GWAS) have identified numerous genetic modifiers that either shorten or delay the onset of HD [25, 26]. One such modifier is MutS homolog 3 (MSH3); a key gene involved in DNA mismatch repair. A single nucleotide polymorphism (SNP) in MSH3 is associated with slower disease progression in HD patients [25]. Knockout of MSH3 in HD mouse models reduced somatic expansion of CAG repeats and mutant HTT (mHTT) aggregation [27, 28], supporting its relevance as a genetic modifier.

PolyQ diseases exhibit both shared and unique clinical phenotypes

In addition to similar genetic inheritance patterns and mostly adult onset, polyQ diseases also share impaired muscle function, as a result of the degeneration of different regions of the central nervous system [7]. Some diseases also show cognitive impairment, such as SCAs (1, 2, 3, 6, 7, and 17) [29], HD [30], and DRPLA [31], while speech difficulties are often observed in SCAs [32], SBMA [33], and HD [34]. Various features, however, are unique to the individual diseases. For example, DRPLA patients display choreoathetosis—involuntary twitching, corneal endothelial degeneration, and cervical dystonia [35–38]. These symptoms have not been reported in other polyQ diseases. In SBMA, clinical phenotypes which are the direct result of androgen insensitivity include gynecomastia, testicular atrophy, reduced fertility, and erectile dysfunction [39]. HD patients may exhibit an early decline in cognitive functions such as memory, attention, episodic learning, and emotional processing [40–42]. Moreover, psychiatric symptoms often precede diagnosis by 10 years in HD; these include depression, anxiety, and obsessive compulsiveness [43].

PolyQ diseases display tissue-specific neurodegeneration

The most prominent neuro-pathological feature of SCAs caused by polyQ-expanded proteins (Table 1) is atrophy of Purkinje cells in the cerebellum [44, 45], accompanied by atrophy of the brain stem, cerebral cortex, and spinal cord [2, 46, 47]. MRI studies showed high-intensity lesions in the cerebral white matter, brainstem, and thalamus of DRPLA patients, indicating atrophy of the spinal cord and cerebellum [22, 48]. Patients affected by SBMA suffer from limb and bulbar muscle atrophy, as well as a loss of lower motor neurons in the spinal cord and brainstem motor nuclei [49]. Moreover, a recent neuroimaging study of SBMA patients has reported structural changes in the cerebellum [50]. The characteristic neuropathological feature of HD is a progressive loss of the GABA-ergic medium spiny neurons within the striatum [51]. MRI and post-mortem studies of HD patient brains also found neurodegeneration in the cerebral cortex, thalamus, cerebellum and brain stem [52, 53]. These findings cumulatively suggest that the cerebellum is particularly vulnerable to polyQ-driven neurodegeneration.

PolyQ diseases display protein aggregation as a common molecular signature

Protein aggregates are a pathological hallmark of polyQ diseases, suggesting that the process of protein misfolding and aggregation plays a role in disease [54]. Protein aggregation collectively refers to the formation of various aberrant protein structures including oligomers, protofibrils, fibrils, and inclusion bodies, found in patient cells [55, 56]. Such protein aggregates have been detected in intranuclear as well as cytoplasmic regions of the cell [57]. Various studies have reported strong links between aggregates and disease pathology [58–60], while others propose that aggregated polyQ-expanded proteins represent a protective mechanism [61]. Furthermore, it has been proposed that the soluble polyQ-expanded protein, rather than its insoluble aggregated species, acts as a driver of cytotoxicity [57, 62].

Several studies have observed polyQ-expanded protein aggregates in SCAs 1, 2, 3, 6, 7, and 17 [63–68]. These aggregates can localize to the cytoplasm but are more commonly observed within the nuclei of neurons and are often referred to as nuclear inclusion bodies [69–72]. In DRPLA patients, intranuclear inclusions have been detected predominantly in cerebellar dentate nuclei, but are also observed throughout different brain regions [73, 74]. Likewise, nuclear inclusion bodies have been reported in brains of SMBA patients [75]. In HD, highly aggregation-prone mutant huntingtin N-terminal fragments (mHTTNT) were shown to promote the formation of misfolded toxic oligomers and/or small fibrils in the brains of HD patients [76, 77]. These mHTTNT...
structures are detectable before symptom onset and function as seeds for the formation of larger mHTT NT aggregates [77]. Recently, residue-specific NMR investigations have revealed that the polyQ tract in soluble N-terminal HTT fragments adopts an α-helical structure, which becomes longer and more stable with increasing polyQ length [78]. These α-helical structures are suggested to drive the formation of stable oligomeric structures, ultimately leading to the assembly of amyloidogenic fibrils [79]. Different HTT aggregate conformations induced cytotoxicity when introduced into cell models [80]. Closer examination of polyQ aggregates highlights the sequestering of other proteins, such as ubiquitin, chaperone proteins, proteasome proteins, and transcription factors [81, 82], within the aggregates. For example, the CREB-binding protein CBP (CBP), tumor protein p53 (p53), and mSin3a transcription factors were found in polyQ-expanded HTT aggregates [83–86]. The presence of CBP was also observed in nuclear inclusions of SBMA patient tissues [87]. Sequestering of the general transcription factor IIB (TFIIB) into inclusions has also been reported in models of SCA17 [88]. Interestingly, TBP, the polyQ protein behind SCA17, has been found in nuclear inclusions of SCA3 and DRPLA disease brains [89, 90]. Additionally, inclusions from SCA3 and DRPLA disease brains are enriched with the transcription factors Sp1 and TAFII130 [90].

### 1.7 PolyQ disease proteins are structurally distinct

SCA1, SCA2, SCA3, SCA7, SCA6, SCA17, SBMA, DRPLA, and HD are caused by mutations in the genes ATXN1, ATXN2, ATXN3, ATXN7, CACNA1A, TBP, AR, ATN1, and HTT, respectively [3–6, 91–95]. They do not share sequence motifs, except for the polyQ-expanded region. Structural information would be conducive to a better understanding of how the expanded polyQ tracts produced from CAG/CAA repeats affect PPIs. To date, there is no high-resolution structural data available for the polyQ-containing disease proteins, except for HTT. About 90% of the open reading frame of full-length HTT was structurally resolved by cryo-electron microscopy (cryo-EM) [96]. However, advances in machine learning are enabling researchers to make structural predictions for proteins which are difficult to crystallize. One such advancement is the structure prediction tool AlphaFold [97]. Since its release in 2021, various improvements have been made to make predictions faster and more accessible [98]. We searched the AlphaFold Protein Structure Database for the polyQ proteins and specifically examined the structural configuration of the polyQ region for each protein. Presently, all polyQ proteins, except HTT, can be found in the AlphaFold database (Figure 1A–I). The absence of HTT might be attributed to its large size, which generally requires greater-than-average computing power. Nonetheless, there are two Cryo-EM structures of HTT published; however, both lack the polyQ region [96, 99]. In this review, we used the PDB 6X9O structure with a 2.6 Å resolution for comparisons [99]. Overall, the predicted models are structurally distinct. ATXN3, CACNA1A, TBP, and HTT were predicted to have a high degree of three-dimensional organization as exhibited by distinct structural domains (Figure 1C, E, F, and I). ATXN1, ATXN2, ATXN7, ATN1, and to some extent AR, were predicted to have only a few folded domains, while a large fraction of the amino acid sequence was disordered (Figure 1A, B, D, and H).

### 1.8 The polyQ region in disease proteins is predicted to fold into an α-helix

Previous experimental work has suggested that the polyQ region is flexible, adopting multiple conformations including a random-coil [100–102], beta-sheet [103, 104], or alpha (α)-helical structure [102, 105–107]. Recently, studies using nuclear magnetic resonance (NMR), circular dichroism (CD), and x-ray crystallography have cemented the notion that polyQ regions predominantly adopt α-helical structures in both polyQ disease proteins [108–112] and non-disease proteins [113].

Regarding the features of this α-helical structure, studies of the N-terminal fragment of AR have highlighted that there is a notable increase in helical propensity upon polyQ expansion [110, 111, 114]. Similar findings have been reported for long polyQ tracts of HTT exon1 [78]. The increase in helicity is dictated by unconventional side-chain-to-main-chain hydrogen bonds, donated by glutamine side chains to the main-chain carboxyl of neighboring residues in the polyQ region, which stabilize the α-helix [111, 115]. Interestingly, the strength of these intramolecular AR polyQ interactions is dependent on the ability of the acceptor residue to shield the glutamine side chains from competing interactions with water molecules, suggesting that the α-helical structure of the polyQ region is not only residue-specific but also influenced by the local cellular environment [111]. The α-helical structure, in both the AR and HTT exon1, is not uniformly distributed throughout the length of the polyQ region; rather it has been found to decrease in helicity from N- to C-terminus [110, 111, 116, 117]. This may be due to the regions flanking the polyQ stretch and their effects on the helical structure. Recent NMR and CD studies of the N-terminal fragment of AR have demonstrated a notable increase in helical propensity upon polyQ expansion which is initiated by an N-terminal leucine motif [110, 114, 115].

The poly-Q flanking regions may also influence the helical propensity of the polyQ domain. Leucine residues can be found enriched at the N-terminus of proteins with longer polyQ tracts [115]. Leucine residues are hydrophobic and they form strong hydrogen bonds with the hydrophilic glutamines in the N-terminus of the polyQ tract and thus strengthen N-terminal helicity [111, 118]. Conversely, regions rich in prolines were reported to occur typically at the C-terminus of the polyQ tract [115, 119]. Interestingly, a flanking proline-rich region has been demonstrated to reduce the stability of the polyQ region and delay aggregate formation [107, 120–123]. The conservation of both N-terminal leucine and C-terminal proline residues is hypothesized to be an evolutionarily conserved “capping” mechanism which sterically protects the α-helical polyQ domain against a structural transition into a beta-sheet secondary conformation, preventing subsequent fibril assembly and aggregation [107, 115, 124, 125].
FIGURE 1  Predicted AlphaFold structures of polyQ disease proteins. (A) Predicted AlphaFold protein model of full-length ATXN1 (Human; AF-P54253), (B) ATXN2 (Human; AF-Q99700), (C) ATXN3 (Human; AF-P54252), (D) ATXN7 (Human; AF-O15265), (E) CACNA1A (Human; AF-O00555), (F) TBP (Human; AF-P20226), (G) AR (Human; AF-P10275), and (H) ATN1 (Human; AF-P54259). (I) Predicted AlphaFold protein model HTT (HTTQ21(1–414)) from amino acids residues 1 to 413, containing 21 polyglutamines. Predicted HTTQ21(1–414) AlphaFold model is shown superimposed with cryo-EM determined HTT-HAP40 protein structure shown in gray (protein data bank ID 6X90, 2.60 Å resolution [99], where the polyQ region was not determined in the cryo-EM structure. HTTQ21(1-414) model aligns to a high degree to cryoEM structure. Residues framed by black rectangle represent the wildtype polyQ region. Scale bar represents pLDDT values derived from the AlphaFold prediction and represents a per-residue confidence measure [97]: pLDDT > 90, high accuracy; 90 > pLDDT > 70 modelled well; 70 > pLDDT > 50 low confidence; pLDDT < 50 poor accuracy. AR, androgen receptor; ATN1, atrophin 1; ATXN1, ataxin 1; ATXN2, ataxin 2; ATXN3, ataxin 3; ATXN7, ataxin 7; CACNA1A, calcium voltage-gated channel subunit alpha 1A (Cav2.1); cryo-EM, cryo-electron microscopy; HTT, huntingtin; pLDDT, per-residue model confidence score; polyQ, polyglutamine; TBP, TATA-binding protein.
To compare the predicted folding of the polyQ region for all proteins, we performed an AlphaFold prediction with the first 413 residues of HTT that contain a polyQ21 region and superimposed it onto the PDB 6X9O cryo-EM structure [99], which lacks the polyQ region. This prediction generated a structure with high per-residue confidence, while the polyQ region, as expected, showed lower confidence (Figure 1). Comparing all polyQ regions, we found that the polyQ tract was disordered only in the predicted CACNA1A protein (Figure 1E). In strong contrast, α-helical structures were obtained for the different polyQ tracts in all other predicted models, supporting the hypothesis that polyQ-containing regions in proteins may have a high propensity to form α-helical structures [78, 126–128].

These predicted models also revealed that polyQ regions are typically flanked by disordered regions, suggesting their ability to associate with polyQ tracts or other domains present in interacting protein molecules. This flexibility may facilitate the formation of stable coiled-coil structures and promote oligomer assembly. Indeed, previous studies have proposed that polyQ domains induce protein association through stable α-helical coiled-coil structures [113, 129]. Furthermore, the presence of coiled-coil structures within polyQ proteins themselves suggests a close relationship between polyQ domains and coiled coils [112, 113, 130]. The coiled-coil domains (CCDs) of the proximal sequences strongly affect the structures and properties of the polyQ tracts, influencing and often increasing their aggregation potential [113]. In fact, the 17-residue N-terminus directly flanking the polyQ sequence in HTT has been shown to greatly enhance polyQ aggregation via an oligomerization process driven by helical interactions [131].

This helical coiled-coil conformation is proposed to be an intermediate in the aggregation process [113]. On the other hand, an alanine-rich region preceding the polyQ tract in ATXN7, which normally forms a stable α-helix itself and can initiate the formation of α-helical polyQ tracts, is suggested to suppress the aggregation of ATXN7 polyQ tracts [132]. Taken together, the AlphaFold predictions, combined with multiple studies into the structures adopted by polyQ tracts of different proteins, show not only a high likelihood for polyQ regions to fold into α-helices but also that this folding is often dictated by the amino acids flanking the polyQ tracts. The flanking sequences can have either suppressive or enhancing effects on the aggregation propensities of long polyQ regions in disease proteins.

Interestingly, when structurally mapped to full proteins, the polyQ domain consistently resides at the protein surface, supporting its potential role in mediating PPIs [112]. Thus, the α-helical structure of the polyQ domain may be seen as an extension of the helical structures flanking it, thereby extending the polyQ proteins’ coiled-coil secondary structure which may be key in triggering the formation of PPIs.

## 1.10 PolyQ disease proteins interact with multiple other cellular proteins

Proteins in cells form complex macromolecular assemblies to carry out their specific tasks in biological systems, such as regulation of gene expression, protein degradation, or energy production. Thus, knowledge about PPIs of polyQ disease proteins is important to understanding genotype-to-phenotype relationships and the development of human diseases [146, 147]. To gain information about interaction partners of polyQ disease proteins, four widely used PPI databases were queried (BioGRID [148], IntAct [149], STRING [150], and HIPPIE [151]). Based on the database queried, different numbers of human PPIs were obtained for each protein. For STRING, which allows the display of a maximum of 500 interactors and requires specifying a confidence score, we used the default of 0.4, that is, medium confidence.

Within all PPI databases, HTT and AR had the highest number of protein interactors, while ATN1 and ATXN7 had the lowest (Figure 2A). For HTT, both BioGRID and STRING showed a lower number of interactors (452 and 470, respectively), while HIPPIE and the IntAct database contained the most PPI annotations (1099 and 1196, respectively). However, when a specific HTT PPI database (OMNI) [152] was queried, a total of 3392 PPIs were found to be annotated across all organisms, with 260 recorded as human HTT interactors (Figure 2B). This suggests that the vast majority of HTT interactors are derived from studies using mouse, yeast, or other models. Based on the BioGRID database, the AR protein has the second highest number of protein interactors annotated, at 952 (Figure 2A). This is followed by ATXN1 with 630 interactors and ATXN3 with 515. For TBP and AR, the STRING database shows 500 interactors, the maximum number of results it can display. For ATXN2 (301), CACNA1A (219), ATXN7 (157), and ATN1 (179), STRING showed fewer interactors compared to the other polyQ

## 1.9 PolyQ expansions influence cellular localization of polyQ disease proteins

An intriguing aspect of the polyQ disease proteins is that nearly all localize to the nucleus. While ATXN1, TBP, and ATXN2 are primarily localized to the nucleus [133–135], ATXN3 and ATXN7 are predominantly found in the cytoplasm, but generally shuttle between the cytoplasm and the nucleus [69]. However, polyQ-expansion of ATXN3 and ATXN7 leads to their retention in the nucleus [136, 137]. In the case of ATXN3, nuclear localization marks a crucial shift in the manifestation of SCA3 symptoms [138]. Similarly, AR moves between the cytoplasm and nucleus [139]. Upon polyQ expansion, it too becomes sequestered in the nucleus, potentially due to interactions between the polyQ domain and the nuclear pore complex [140]. Among the polyQ disease proteins, HTT and ATN1 both exhibit a dual presence in the cytoplasm and the nucleus [141, 142]. mHTT tends to accumulate in both the nucleus and cytoplasm; with early stage HD showing predominantly cytoplasmic aggregates, while juvenile HD patients exhibit nuclear aggregates [59, 143]. ATN1 is mainly retained in the nucleus following polyQ expansion [142]. In contrast to all other polyQ disease proteins, CACNA1A, a transmembrane subunit of the P/Q type calcium channel Cav2.1, is exclusively associated with cellular membranes [144]. PolyQ expansions in CACNA1A result in rare cytoplasmic and nuclear aggregate formation [145].
Proteomic scales of polyQ disease proteins based on PPI databases. (A) PPI databases (BioGRID, HIPPIE, IntAct, STRING) were queried for the total number of human protein interactors. For BioGRID, HIPPIE, and IntAct no confidence filtering was performed. For the online site of STRING, a confidence score of 0.4 and a maximum of 500 interactors were selection criteria. STRING is limited to displaying 500 total interactors. The bold dashed line denotes 500 interactors, the red dashed line 100 interactors. (B) The HTT PPI database OMNI was queried to obtain the total number of protein interactors for human and all organisms (e.g. mouse, yeast, human, etc.). Total HTT protein interactors from (A) were included for comparison. HTT, huntingtin; PPI, protein–protein interaction; polyQ, polyglutamine.

1.11 PolyQ disease proteins have diverse biological functions

The number of reported protein interactors does not permit assertions about the confidence of interactions, involvement in specific pathways, or functional enrichment. Since STRING provides interaction confidence scoring, we queried the database to generate protein network maps for polyQ disease proteins based on the following criteria: human data, physical subnetwork, interaction confidence score of 0.7 or higher, and maximum size of 25 nodes. The resulting maps were subjected to an enrichment analysis for GO biological process using ShinyGO [153] to obtain an overview of the main functions for each polyQ disease protein (Figure 3A–H).

An enrichment analysis for ATN1 was not possible due to the limited number of PPIs within the network. Analysis of the ATXN1, ATXN2, ATXN7, and TBP protein networks highlighted an enrichment for PPIs involved in gene expression (Figure 3A, B, D, and F). Specifically, ATXN1 PPIs were associated with RNA processing (GO:0006396) and negative regulation of transcription by RNA polymerase II (GO:0000122), implying a role in suppressing transcription (Figure 3A). Similarly, ATXN2 PPIs were linked to the assembly of stress granules (GO:0034063) and P bodies (GO:0033962), which sequester mRNA [154], suggesting involvement in the negative regulation of transcription (Figure 3B). Conversely, TBP PPIs predominantly relate to RNA polymerase II preinitiation complex assembly (GO:0051123), indicating a role in the initiation of transcription (Figure 3F). Similarly, ATXN7 PPIs were strongly associated with histone acetylation (GO:0043966 and GO:0016573), indicating a role in encouraging transcription by altering chromatin accessibility (Figure 3D).

In contrast, enrichment analysis of the ATXN3, CACNA1A, and AR networks emphasized their involvement in specific biological processes (Figure 3C, E, and G). For instance, ATXN3 PPIs showed a notable enrichment for aggrephagy proteins (GO:0035973) and protein deubiquitination (GO:0016579), suggesting a significant role in protein clearance mechanisms (Figure 3C). CACNA1A PPIs were primarily associated with calcium ion transmembrane transport (GO:0070588) (Figure 3E), while AR PPIs were linked to steroid hormone-related processes (GO:0030518, GO:0071383, and GO:0048545) (Figure 3G). Lastly, HTT PPIs were chiefly enriched in vesicular localization and trafficking (GO:0051650, GO:0051648, and GO:0099518), indicating the involvement of HTT in regulating vesicular transport (Figure 3H). Additionally, HTT PPIs showed enrichment of protein stability regulation (GO:0031647) and metabolic pathways/processes that break down molecules (GO:0009895, GO:0031329, and GO:0051248), such as autophagy (GO:00010506).

Most of the polyQ disease protein networks, namely those of ATXN1, ATXN2, ATXN7, and TBP, exhibited a significant involvement in gene expression regulation, notably through histone modification (GO:0016570), regulation of transcription from RNA polymerase II promoters (GO:0000122, GO:0006367), or via regulation of translation (GO:0006412). Although not represented as an enriched process, HTT has been found to modulate transcription [128, 155], and the AR protein is a well-characterized ligand-dependent nuclear transcription.
PolyQ disease proteins exhibit diverse biological functions. GO biological processes enrichment analysis for (A) ATXN1, (B) ATXN2, (C) ATXN3, (D) ATXN7, (E) CACNA1A, (F) TBP, (G) AR, and (H) HTT using ShinyGO 0.77 [153]. Protein networks used for enrichment analysis are based on the STRING PPI database using the following criteria to query for high-confidence proteins: human physical subnetwork with a maximum 25 interactors and a confidence score of 0.7. The analyzed PPI networks are shown in Figure 4. Lollipop graphs show fold enrichment in the x-axis; size of circular node indicates number of total genes that constitute the GO biological process. Color represents false discovery rate (FDR) as log10 (Blue = low, Green = high). AR, androgen receptor; ATXN1, ataxin 1; ATXN2, ataxin 2; ATXN3, ataxin 3; ATXN7, ataxin 7; CACNA1A, calcium voltage-gated channel subunit alpha 1A (Cav2.1); HTT, huntingtin; polyQ, polyglutamine; PPI, protein–protein interaction; TBP, TATA-binding protein.
factor [156]. Moreover, ATXN3 was found to control transcription by regulating chromatin structure [157] and the CACNA1A bicistronic gene was found to encode for α1ACT, a transcription factor containing the polyQ region that promotes cerebellar development [158]. This finding highlights shared biological processes among the polyQ disease proteins and aligns with existing literature which has described polyQ proteins as pivotal regulators of transcription [159–161].

2 | PROTEIN NETWORKS

2.1 | ATXN1 binding to the transcriptional repressor CIC drives pathogenesis in SCA1 models

ATXN1 is a protein implicated in two primary pathways related to gene expression [162] and RNA metabolism [163]. Examining physical human PPIs with a high confidence using STRING reveals a network of ten ATXN1 interactors (Figure 4A). Among the interactors is the transcriptional repressor ccapia (CIC). CIC was shown to interact directly with native ATXN1 to form a complex in vivo [162]. This ATXN1/CIC complex is highly expressed in the nuclei of Purkinje cells, which are the most vulnerable neurons affected in SCA1 [162]. PolyQ expansion of ATXN1 alters CIC’s repressor activity through a toxic gain of function, resulting in increased repression of CIC target genes [162, 164]. Intriguingly, a 50% reduction in CIC levels in ATXN1 154Q mice alleviates behavioral symptoms and Purkinje cell loss [165]. Moreover, preventing ATXN1/CIC binding with ATXN1 mutations showed a partial rescue of cerebellar phenotypes, motor coordination, respiration, and lifespan in SCA1 mouse models [164, 166]. These findings strongly implicate the ATXN1/CIC complex in contributing to polyQ-mediated proteotoxicity in SCA1.

2.2 | ATXN2 is a stress granule protein implicated in RNA processing

Assessing human PPIs with a high confidence using STRING reveals a network of 14 ATXN2 interactors (Figure 4B). Among these interactors is an endocytic adaptor called SH3 domain containing GRB2 like 2, endophilin A2 (SH3GL2). ATXN2 has been shown to be involved in endocytic receptor cycling [167], where it associates with the endocytic adaptors endophilin A1 and A3 (SH3GL2 and SH3 domain containing GRB2 like 2, endophilin A3 [SH3GL3]). This leads to the formation of an endocytosis complex that negatively regulates the endocytic internalization of the epidermal growth factor receptor at the plasma membrane [168]. Another interactor is the GTPase-activating protein-binding protein 1 (G3BP1) (Figure 4B). G3BP1 plays a significant role in the formation of stress granules [169], as well as in mRNA stabilization [170], and degradation [171].

A noteworthy ATXN2 interactor is the poly(A)-binding protein 1 (PABPC1) (Figure 4B), the C-terminal region of which consists of five α-helices [172]. Studies show that PABPC1 regulates mRNA degradation and maturation by interacting with the poly(A) tail of mRNAs [173]. The interacting proteins ATXN2 and PABPC1 were found to be localized to stress granules [174]. The interaction between these two proteins was also reported to be mediated by a C-terminal intrinsically disordered region in ATXN2. Furthermore, it was shown that the ATXN2-PABPC1 interaction promotes lengthening of poly(A) tails in target mRNAs, supporting a functional role of ATXN2 in transcriptional regulation [175]. Interestingly, polyQ-expanded ATXN2 has been found to sequester PABPC1 into inclusion bodies with protein aggregates [176], suggesting that a loss function of the ATXN2-PABPC1 protein complex plays a role in disease development. Lastly, it is also interesting to note that the ATXN2 protein network (Figure 4B) includes interactions between ATXN2 and proteins that regulate alternative splicing such as cytosolic granule associated RNA binding protein (TIA1) [177], RNA binding fox-1 homolog 1 (RBFOX1) [178], and the RNA helicase DDX6 [179] further affirning the involvement of ATXN2 in maintaining the overall homeostasis of RNA in the cell.

2.3 | ATXN3’s polyQ domain and deubiquitinating activity are essential for BECN1 binding and promotion of autophagy

ATXN3 contains a conserved Josephin domain that facilitates the deubiquitination of proteins with polyubiquitin chains. Interestingly, polyQ-expansions in ATXN3 reduce its deubiquitinase activity [180]. Among the STRING protein network of high-confidence ATXN3 interactors are several proteins involved in protein-ubiquitination such as parkin [181], ubiquitin [182], and ubiquilin-1 [183]. Moreover, ATXN3 interacts with RAD23 homolog A, nucleotide excision repair protein (RAD23A) and RAD23 homolog B, nucleotide excision repair protein (RAD23B), which are subunits of the ubiquitin-proteasome system [184–186] (Figure 4C).

Another high-confidence interactor of ATXN3 is beclin 1 (BECN1), a key inducer of autophagy [187] (Figure 4C). It was found that BECN1 interacts with the polyQ tract of ATXN3 [180]. This polyQ-mediated interaction facilitates the deubiquitination of BECN1, protecting it from proteasomal degradation and allowing autophagosome biogenesis to proceed [180]. The expression of polyQ-expanded ATXN3 results in increased binding to BECN1 but reduces BECN1 deubiquitination, which leads to impaired starvation-induced autophagy [180]. Moreover, it was shown that wildtype and mutant ATXN3 compete for binding with BECN1, with the mutant ATXN3 negatively interfering with the wildtype ATXN3-BECN1 interaction [188]. Interestingly, BECN1 contains a CCD and forms an anti-parallel coiled-coil homodimer in the absence of interactions with other CCD-containing or α-helical proteins [189]. Given that polyQ domains are known to adopt α-helical structures with a propensity for coiled-coil interactions [113], it is plausible that the expanded polyQ domain of mutant ATXN3 might exhibit stronger coiled-coil-mediated interactions with BECN1 compared to wild-type ATXN3. Conversely, overexpression of BECN1 was shown to lead to the clearance of mutant ATXN3 [190], the prevention of mutant...
FIGURE 4 STRING high-confidence protein networks of polyQ disease proteins. Protein networks for (A) ATXN1, (B) ATXN2, (C) ATXN3, (D) ATXN7, (E) TBP, (F) CACNA1A, (G) ATN1, (H) AR, and (I) HTT. Nodes denote proteins and the edges represent interactions within the nodes. Protein networks based on the STRING PPI database were queried for high-confidence proteins using the following criteria: human physical subnetwork with a maximum of 25 interactors and a confidence score of 0.7. AR, androgen receptor; ATN1, atrophin 1; ATXN1, ataxin 1; ATXN2, ataxin 2; ATXN3, ataxin 3; ATXN7, ataxin 7; CACNA1A, calcium voltage-gated channel subunit alpha1 A (Cav2.1); HTT, huntingtin; polyQ, polyglutamine; PPI, protein–protein interaction; TBP, TATA-binding protein.
ATXN7 is a key component of the STAGA transcriptional coactivator complex

The ATXN7 protein network is composed of highly connected proteins involved in transcription regulation (Figure 4D). Core structural components of the SPT3-TAFII31-GCNSL acetylase (STAGA) complex are represented in the ATXN7 protein network (TATA-box binding protein associated factor 9 [TAF9], transformation/transcription domain associated protein [TRRAP], TATA-box binding protein associated factor 12 [TAF12], and TATA-box binding protein associated factor 5L [TAF5L]) (Figure 4D). The STAGA complex is responsible for both acetylation and deubiquitinating histones, thereby controlling transcription, with ATXN7 as one of its vital components [192–195]. ATXN7 plays an important role in the assembly and stability of the STAGA complex and functions as a transcriptional coactivator [196]. Studies have found that ATXN7 interacts with TAF9, TRRAP, TAF12, and TAF5L through its C-terminus [196]. Expansion of the polyQ domain in ATXN7 disrupts the assembly of a fully functional complex, reducing its ability to acetylate histone H3. This suggests that polyQ-expanded ATXN7 perturbs the general control non- depressive 5 (GCNS5)-mediated histone acetyltransferase activity of the complex, leading to transcriptional dysfunction in neurons of SCA7 patients [196].

Expansion of TBP’s polyQ domain disrupts transcriptional activity

Examining the TBP PPI network, one can see that TBP interacts directly with multiple general transcription factors (GTFs) and TBP-associated factors (TAFs) (Figure 4E). Moreover, this protein network is highly interconnected, suggesting that all TBP-associated proteins are functionally linked. In fact, TBP together with 13 TAFs (TAF1-13) constitute the TFIIID, a large multiprotein assembly that plays a crucial role in recruiting other subunits of the transcription factor complex to assemble the core pre-initiation complex required for transcription initiation [197, 198]. One noteworthy interaction partner is the transcription initiation factor IIB (GTF2B) protein [199], for which TBP serves as the DNA-binding subunit. Studies have shown that TBP binds GTF2B via its C-terminal domain in order to form pre-initiation complexes [200, 201]. Intriguingly, nuclear inclusions with polyQ-expanded TBP were found to contain GTF2B; and co-immunoprecipitation experiments revealed enhanced binding of polyQ-expanded TBP with GTF2B compared to non-pathogenic controls [88]. This suggests that abnormal TBP/GTF2B interactions affect promoter occupancy and exert downstream effects on the transcription of damage control proteins, such as heat shock proteins [88].

CACNA1A plays a role in calcium regulation and its bicistronic gene product functions as a transcription factor

The CACNA1A gene encodes an alpha1a subunit of the P/Q type voltage-sensitive calcium channel (VSCC) known as Ca(V)2.1 [202]. VSCCs are transmembrane channels that regulate the entry of calcium ions into cells, playing essential roles in calcium-dependent processes such as muscle contraction, hormone or neurotransmitter release, gene expression, and cell death [203]. It was determined that CACNA1A is responsible for pore formation, facilitating the transport of calcium in and out of Purkinje cells [204].

Unsurprisingly, the majority of the high-confidence PPIs are involved in calcium signaling pathways (Figure 4F). These include calcium voltage-gated channel subunit alpha 1B (CACNA1B), C (CACNA1C), and CACNA1E, which regulate calcium binding and the pore-forming function of CACNA1A [203]. Additionally, CACNA1A interacts with calmodulin, a protein encoded by three separate genes: calmodulin 1 (CALM1), calmodulin 2 (CALM2), and calmodulin 3 (CALM3) [205], which play a crucial role in the calcium signal transduction pathway and can modulate the activity of the Ca(V)2.1 channel [206].

To date, studies have mostly focused on the general perturbation of the P/Q-type calcium channel caused by the polyQ expansion in the CACNA1A protein [207–209]. Further investigations exploring the polyQ dependency of CACNA1A interactions are warranted. A particular focus can be placed on the interactors of the bicistronic CACNA1A gene product that expresses a transcription factor called α1ACT. This CACNA1A-associated protein was shown to coordinate a gene expression program involved in Purkinje cell development [158]. The α1ACT transcription factor also contains a polyQ domain and when expanded, can cause SCA6 pathology [158, 210]. Thus, besides investigating the polyQ-expanded CACNA1A protein and its interactions, unveiling the α1ACT interactome may yield mechanistic insights into the disease process and facilitate the development of therapeutic approaches.

ATN1 has links to transcription factors and insulin signaling pathways

The specific subcellular activities of ATN1 currently remain largely unknown. However, there is some evidence pointing toward a role of this polyQ-containing protein as a transcriptional co-repressor [211–213] and as a facilitator of early striatum development [214]. When assessing the physical human PPIs network of ATN1, only two high-confidence validated PPIs were identified (Figure 4G): nuclear receptor subfamily 2 group E member 1 (NR2E1) and brain-specific angiogenesis inhibitor 1-associated protein 2 (BAIAP2). Notably, these two proteins are not associated with one another but only share ATN1 as a common interactor (Figure 4G).

NR2E1 is a DNA-binding protein [215] that belongs to the orphan nuclear receptor family, a group of ligand-dependent transcription factors.
AR is a ligand-dependent transcription factor [216, 217]. NR2E1 mediates transcriptional repression through histone modification via interactions with lysine-specific histone demethylase 1 (LSD1) [218]. Research has shown that the ligand binding domain (LBD) of NR2E1 is required for interaction with the C terminus of ATN1 [213]. Deletion of the last 45 amino acids of the LBD domain abolished the transcriptional repression activities of NR2E1. This shows that ATN1 binding to NR2E1 is important for NR2E1-mediated transcriptional repression. However, the impact of pathogenic ATN1 polyQ expansions on its interaction with NR2E1 remains to be studied.

The second interactor, BAIAP2 (also known as insulin receptor substrate protein 53 – IRSp53), is an adaptor protein that connects membrane-bound G-proteins to effector proteins in the cytoplasm [219] (Figure 4G). It is ubiquitously expressed in excitatory synapses and has been theorized to be important for synapse development and plasticity [219]. Moreover, insulin receptor substrates generally form pre-assembled complexes and play a role in signal transduction pathways and actin organization [220]. Via its SH3 domain, BAIAP2 binds to a proline-rich region in ATN1 that is close to the ATN1 polyQ region [221]. Galactose-responsive transcription factor GAL4 (GAL4)-yeast two-hybrid experiments demonstrated that when ATN1 harbored an expanded polyQ tract it can still interact with BAIAP2 but with significantly reduced beta-galactosidase activities [221]. These results suggest that the polyQ expansion might reduce the stability of the interaction between BAIAP2 and ATN1, ultimately resulting in impairment of the insulin growth factor 1 signaling pathway.

2.8 | AR is a ligand-dependent transcription factor requiring heat-shock protein binding for its function

The AR protein network generated using STRING (Figure 4H) resulted in 25 high-confidence interactors that can be categorized into three distinct groups of differing functions. The first group includes post-translational modifiers, such as mouse double minute 2 homolog (MDM2); the second includes nuclear receptor coregulators, nuclear receptor corepressor 1 (NCOR1), 2 (NCOR2), NCOA1, NCOA2, and NCOA4. The third group comprises molecular chaperones, including multiple heat shock proteins (Figure 4H). Two notable interactors, SRC and the lysine-specific histone demethylase 1A (KDM1A), have been found to exhibit enhanced activity upon polyQ expansion in AR [222, 223]. Some of the heat shock proteins represented in the network are the heat-inducible molecular chaperone heat shock protein 90 (HSP90), along with its isoforms heat shock protein 90 family class A member 1 (HSP90AA1) and heat shock protein 90 family class B member 1 (HSP90AB1) (Figure 4H). Members of the HSP90 protein family facilitate the refolding of misfolded proteins, prevent aggregation, and significantly contribute to the overall maintenance of cellular proteostasis [224]. Upon binding to a hormone ligand, AR undergoes a conformational change driven by HSP90, adopting a DNA-binding-competent conformation [225]. The AR is then translocated to the nucleus in an HSP90-dependent manner [225]. Therefore, HSP90 plays an indispensable role in ensuring the proper functioning of AR. Studies with SBMA mouse models identified an increased occurrence of polyQ-expanded AR in complex with HSP90 and its co-chaperone p23, compared to wild-type AR [226]. Structural studies have revealed a model of HSP90-p23 which binds to the glucocorticoid steroid hormone receptor (GR) via a 13-residue α-helical tail [227]. This tail was previously believed to be an unstructured region at the C-terminus of p23. Notably, the residues involved in this p23:GR interface are conserved across different steroid hormone receptors, including AR in vertebrates, and are essential for HSP90 binding [227]. Therefore, it is plausible that the increased occurrence of HSP90 in complex with AR may be influenced by an enhanced interaction between the α-helical tail of p23 and the α-helical structure of the polyQ region of AR, possibly through coiled-coil interactions. Interestingly, treatment with the HSP90 inhibitor 17-AAG resulted in a preferential degradation of both monomeric and aggregated mutant AR by the proteasome, ameliorating SBMA motor phenotypes [229]. Modulation of polyQ-induced AR toxicity through HSP90 inhibitors may therefore be a potential therapeutic avenue for treating SBMA [228].

2.9 | HTT plays a role in vesicular trafficking and transcriptional regulation

HTT is a highly conserved and ubiquitously expressed protein with diverse subcellular functions [229, 230]. Evidence implicates HTT in axonal transport [231], vesicular trafficking [232], cell division [233], endocytosis [234], autophagy [235], and transcriptional regulation [236]. This multifaceted function of HTT is also reflected in its STRING-generated high-confidence PPI network (Figure 4I). Firstly, multiple interactors of HTT are proteins involved in vesicular trafficking and cytoskeletal organization, such as 40-kDa huntingtin-associated protein HAP40 (F8A1), SH3GL2, and SET domain-containing 2 (SETD2) [168, 237, 238]. Interestingly, although these proteins have similar functions, they do not actually associate with one another (Figure 4I). F8A1, commonly referred to as HAP40, is a protein whose function is not yet fully understood. Studies in cell models have shown that HTT and HAP40 form a complex and that their expression levels within the cell are coupled. Also, structural studies on the complex revealed that they form a stable heterodimer [99]. This complex can function as a RAB5A, member RAS oncogene family (Rab) effector, regulating cytoskeleton-dependent endosome formation and dynamics [237]. These results suggest that the HTT/HAP40 complex is implicated in regulating cytoskeleton-dependent endosome dynamics.

Another interaction partner of HTT is the methyltransferase SETD2 protein. It is a lysine methyltransferase responsible for posttranslational modifications on histone H3 [239] but was recently found to also regulate microtubule dynamics via trimethylation of alpha-tubulin [240]. SETD2 was also shown to methylate actin through its interaction with HTT and the huntingtin interacting protein 1 related protein (HIP1R) [238], which is also a high-confidence HTT interactor (Figure 4I). Disruption of the interactions between SETD2-HTT-HIP1R resulted in inhibition of actin filament methylation, culminating in defective actin polymerization and impaired cell


migrating [238]. Interestingly, the HTT PPI network shows no direct interaction between SETD2 and HIP1R (Figure 4I). This most likely means that the formation of this complex is highly dependent on HTT being the link between SETD2 and HIP1R.

The HTT protein network also contains multiple interaction partners whose main functions are transcriptional regulation, such as p53, CREBBP (also known as CBP), lysine acetyltransferase 2B (KAT2B), transcription elongation regulator 1 (TCERG1), and SP1 (Figure 4I). Some of these interactors have also been shown to interact with one another (Figure 4I). HTT and p53 directly interact via the N-terminus of HTT and the C-terminus of p53, regardless of the polyQ length in HTT [86]. However, upon polyQ expansion, the transcription factor activity of p53 is greatly diminished [86]. It is important to note that p53 transcription is coactivated by CBP, a polyQ-containing transcription factor. CBP interacts with residues 1–73 in p53 [241].

Interestingly, this region in p53 is structurally similar to the expanded HTT exon1, suggesting that CBP and HTT potentially compete for p53 binding in cells. In cell-based assays, recruitment of CBP into inclusion bodies with pathogenic HTTx1 aggregates has been reported [85]. Furthermore, there is a noticeable decrease in transcription of p53-regulated promoters when pathogenic HTTx1 aggregates are formed in cells; and transcription factors, such as CBP or p53, are recruited into inclusion bodies [86]. Moreover, studies have shown that mHTT can also directly interact with multiple histone acetyltransferases, including KAT2B, and associates with p53 repressing its transcriptional activity [242], as evidenced by their tight association within the HTT PPI network (Figure 4I). All in all, mHTT, p53, and CBP have been shown to co-aggregate and polyQ expansions in N-terminal HTT fragments have been shown to cause a significant repression of p53 transcriptional activity. These molecular changes may eventually lead to cellular toxicity and neuronal dysfunction in HD patients [83].

By functionally analyzing the HTT gene promoter, researchers have been able to identify the transcriptional regulator Sp1, which also interacts with the HTT protein [243, 244]. Strikingly, Sp1 is also a validated high-confidence interactor of CBP and p53 [Figure 4I], supporting the link between HTT and proteins that regulate gene transcription. Sp1 is responsible for regulating the transcription of multiple housekeeping genes and is required for normal embryonic development [245]. Thus, HTT with expanded polyQ tracts, through its direct interactions with Sp1 (Figure 4I), may disrupt the transcriptional activity of Sp1 [246]. However, later studies with HD transgenic mice did not confirm this hypothesis [247]. These conflicting results indicate that there is still a gap in the molecular understanding of the HTT-Sp1 interaction and its relevance for HD development.

Lastly, transcriptional elongation regulator 1 (TCERG1) was identified as a direct interaction partner of HTT (Figure 4I). This protein is highly conserved and regulates the expression of multiple genes by coupling transcriptional elongation with splicing [248]. Other investigations showed that TCERG1 overexpression can indeed rescue mutant HTT neurotoxicity and delay striatal cell death [249]. Interestingly the TCERG1 gene contains a short tandem repeat tract of 38 hexanucleotides (CAGGCC). Recent studies revealed that there is a significant correlation between the age of onset of HD and the sum of the repeat lengths from both alleles in the TCERG1 gene [250]. It is important to note that TCERG1 was identified in a GWAS to be a genetic modifier of HD [9, 250]. More specifically, the longer the repeat hexamer of TCERG1 is, the earlier the age of disease onset. However, more research is required to better understand how TCERG1 influences the onset of HD and how its interaction with mHTT affects pathogenesis.

2.10 | HTT is highly connected to other polyQ disease protein networks

Our GO term enrichment analysis suggested that protein interaction networks with a central polyQ disease protein share common protein interactors (Figure 3). Therefore, we examined direct and indirect links between individual polyQ disease proteins, such as HTT, ATXN1, ATXN3, or TBP, by merging and analyzing the high-confidence STRING-based protein networks. The CACNA1A and ATN1 protein networks stood out because they lacked any common interactors with the other seven networks (Figure 5A). This may be explained by the fact that both polyQ proteins have very specific subcellular functions in comparison to the others. Furthermore, their partner proteins do not connect (Figure 4). It may also well be that the interaction networks of CACNA1A and ATN1 are still too small to find overlapping PPIs. For ATN1, only two high-confidence interaction partners have been described to date (Figure 5A), indicating that it is the most understudied protein among the polyQ disease proteins in terms of its interactions. A literature review for additional ATN1 interactors yielded only glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as a potential common interactor with HTT, albeit this interaction remains to be validated [251].

A common interactor, inositol 1,4,5-trisphosphate receptor type 1 (ITPR1), was detected for HTT, ATXN2, and ATXN3. ITPR1 is a member of the inositol trisphosphate receptor (IP3)-receptor family that forms a calcium channel in the endoplasmic reticulum [252]. This protein is highly expressed in Purkinje cells and experimental evidence has confirmed that deletions in the ITPR1 gene can cause SCA (SCA15) [253]. Perturbations in IP3 have also been reported in other neurodegenerative diseases, such as familial Alzheimer’s disease, autosomal dominant sensoaty ataxia, and SCA29 [254]. The IP3 receptor was shown to function as a nodal point for cellular signaling pathways, such as secretion, gene expression, and cell growth [255]. Regarding its interactions with ATXN2 and ATXN3, ITPR1 appears to show a preference for the expanded polyQ proteins over the wild type [256, 257]. Moreover, polyQ-expanded HTT specifically binds to and activates the ITPR1 calcium channel [258, 259]. The interaction patterns of ITPR1 highlight a common interactor among the different polyQ disease proteins, suggesting that perturbed calcium signaling in neurons may play a role in the development of neurodegenerative diseases, such as SCA2, SCA3, and HD.

Additionally, the merged networks revealed HTT as central protein interaction hub that holds direct and indirect connections to six other...
polyQ disease proteins (ATXN1, ATXN2, ATXN3, ATXN7, TBP, and AR). The majority of proteins linking HTT to other polyQ proteins are proteins with a functional role in gene transcription activation/repression, such as CREBBP (Figure 5A). CREBBP functions as a histone acetyltransferase with transcriptional co-activator activity and promotes transcription of various target genes [260]. It was found to associate with the promoter regions of over 16,000 genes [261]. Additionally, CREBBP can bind to TBP and GTF2B (a TBP interactor) [260], as well as HTT [86]. Moreover, CREBBP binding to CREB1, a common HTT and AR interactor, has been described [262]. Another histone acetyltransferase protein that connects HTT to ATXN7 and TBP is KAT2B. In fact, KAT2B is a subunit of a histone acetyltransferase...
module of the human Ada-two-A-containing (ATAC) complex, a coactivator complex composed of SAGA complex associated factor 29 (SGF29), transcriptional adaptor 3 (TADA3), and transcriptional adaptor 2A (TADA2A) [263]. This histone acetyltransferase module is also present in the STAGA complex, another coactivator complex, of which TAF9, TAF10, and TAF12 are key components [263]. The interactions of KAT2B not only link ATXN7 directly to TBP but also indirectly to HTT. Lastly, another interactor linking HTT to the TBP and ATXN7 is TAF4, a subunit of the basal transcription factor IID. This protein was shown to bind to HTT [246, 264], exhibits high connectivity to TBP interactors and shares partner proteins with ATXN7. Overall, our network analysis highlights the involvement of polyQ-containing disease proteins in transcription regulation and underscores the importance of HTT as a potential hub protein among the other polyQ disease proteins. A case in point is the report of a transcription-coupled DNA repair complex, which contains the proteins HTT, ATXN3, CBR DNA ligase 3 (LIG3), RNA polymerase II subunit A (POLR2A), and other basic transcription factors, for the repair of DNA lesions during transcription elongation [264].

3  CONCLUDING REMARKS

3.1  PolyQ domains in disease proteins likely are flexible, α-helical structures

The public release of AlphaFold is enabling researchers to make structural predictions for proteins and their interaction partners [265, 266]. Most importantly, it can provide high-accuracy predictions of specific folded domains in largely unfolded proteins. The AlphaFold predictions presented in this review indicate significant structural differences among the different polyQ disease proteins (Figure 1A–I); while, notably, the polyQ tract in eight of nine proteins was predicted to adopt a rigid α-helical fold. This suggests that an α-helical fold is the predominant secondary structure in nearly all polyQ-containing disease proteins. Analytical chemistry studies support the notion that polyQ domains in soluble proteins adopt a stable α-helical conformation [111, 117, 267]. Experimental evidence that polyQ domains in proteins are largely unstructured regions has also been presented [268–271]. Intriguingly, AlphaFold predicted that the α-helical polyQ domains in most of the polyQ proteins (except for CACNA1A) are stand-alone structures flanked by disordered regions. This facilitates high flexibility in the three-dimensional space and suggests that α-helical polyQ domains in disease proteins are unlikely to be part of other folded domains. It seems reasonable to speculate that polyQ tracts have a high propensity to associate with other proteins due to their unstructured flanking regions. This may explain why many more PPIs have been detected with N-terminal HTT fragments that contain a polyQ domain than with folded C-terminal HTT fragments lacking such a flexible region [272, 273]. Thus, polyQ expansions in disease proteins may either stabilize or perturb interactions with other proteins, leading to neuronal dysfunction and disease.

3.2  The molecular function of α-helical polyQ domains is still unclear

The AlphaFold predictions suggest that polyQ domains are flexible protein domains with stable α-helical polyQ tracts in soluble proteins. This indicates that they might have specific molecular functions that are perturbed when polyQ tracts get abnormally expanded in disease. To this day, specific molecular functions have not been assigned to polyQ domains in proteins. However, early inquiries into the function of glutamine-rich domains suggested that they serve as structural motifs for transcription activation [274]. Another study found that glutamine-rich stretches are present predominantly in transcriptional regulatory proteins [275] (e.g., POU class 3 homeobox 2 [POU3F2], CBP, E1A binding protein P400 [EP400], mediator complex subunit 15 [MED15], MN1 proto-oncogene, transcriptional regulator [MN1], NOCA2, and NOCA3), and that the length of glutamine-rich stretches correlates with the ability to stimulate transcription [275]. Moreover, it was determined that the highest levels of transcriptional activity can be reached with constructs having between 6 and 34 glutamine residues; while longer stretches exhibited reduced transcriptional activation [275]. In brief, polyQ domains may serve as transcription activation domains and abnormal polyQ expansions in these domains might impair their specific molecular function.

3.3  PolyQ domains may facilitate the association of proteins through formation of coiled-coil structures

In line with the proposed role of polyQ domain-containing proteins in transcription, we found an enrichment of polyQ disease protein interactors predominantly in nuclear processes involved in gene expression regulation, histone modification, and mRNA processing (Figure 3). This raises the question of how the α-helical structure of a polyQ domain relates to the shared nuclear localization and transcriptional activity of polyQ disease proteins.

It was shown previously that α-helical domains have a high propensity to self-assemble, either within the same protein or between different proteins, to form super-secondary coiled-coil structures [276]. These stable coiled-coil interactions play a crucial role in many subcellular processes as well as in transcriptional regulation [277]. Coiled-coil structures have been found to control subcellular protein localization by interacting with and masking subcellular compartment targeting sequences, such as nuclear localization sequences (NLS) [278]. Furthermore, due to their long flexible nature, many coiled-coil proteins have been found to span multiple subcellular compartments and traverse large distances within the cell [277, 279]. Therefore, mutations in α-helical regions, such as polyQ expansions, may disrupt or enable coiled-coil interactions resulting in altered subcellular compartment targeting or function.

Coiled-coil forming motifs are frequently found within DNA and RNA binding proteins [280, 281]. Commonly located adjacent to DNA
and RNA binding domains, coiled-coil forming motifs regulate protein oligomerization and subsequent gene expression [281, 282]. Interestingly, recent findings have highlighted coiled-coil forming motifs as being present in 36% of all RNA-binding proteins (RBPs); which are essential regulators of RNA splicing, localization, and translation [281]. PolyQ proteins have been found to associate with RBPs and RBP-containing non-membrane bound compartments known as ribonuclear protein (RNP) granules [283–287]. Therefore, coiled-coil forming interactions between polyQ proteins and DNA and RNA binding proteins may facilitate gene expression.

Overall, the striking similarities in polyQ disease protein localization and function may be due, in part, to their shared α-helical polyQ domain. It is plausible that polyQ proteins might engage in PPIs by forming coiled-coil interactions via the helical polyQ domains [129]. Notably, proteins with an affinity for expanded polyQ tracts, such as BECN1 and HSP90, contain α-helical structures which are essential for mediating PPIs [189, 227]. This suggests that coiled-coil interactions are a potential binding modality to join polyQ tract-containing proteins in cells. Although additional research is necessary to fully elucidate the role of coiled-coil interactions in polyQ diseases, emerging evidence suggests that these interactions may initiate the aggregation of polyQ-containing proteins in disease [104, 105].

3.4 | On therapeutic perspectives

As of now, there are no cures or approved disease-modifying drugs for polyQ diseases. However, different therapeutic avenues are currently being explored for most, if not all, of the polyQ diseases. There have been promising results from clinical trials with Riluzole, a drug approved for the treatment of ALS—a neurodegenerative disease without polyQ protein at its origin—that showed improvement of cerebellar symptoms in SCA patients [288, 289]. Treatment of SBMA patients with leuprorelin acetate, a manufactured analogue of gonadotropin-releasing hormone, significantly improved swallowing function in a phase 2 clinical trial [290]. Tetrabenazine, an FDA-approved vesicular monoamine transporter (VMAT) inhibitor, and the related Valbenazine are used to reduce chorea in HD patients [291, 292].

Other treatment strategies have focused on anti-sense oligonucleotides (ASOs). Pre-clinical studies have yielded promising results using ASOs to downregulate transcripts encoding polyQ proteins in SCA1, SCA2, SCA3, and SCA7 mouse models [293–296]. For HD, there are several ASOs in clinical trials; the ASO Tominersen targets both normal and mutant HTT RNA, while the ASO WVE-003 is specific for mutant HTT [297]. The ASO VO659 (ClinicalTrials.gov Identifier: NCT05822908) can target long CAG repeats and has the potential for targeting different polyQ disease proteins (e.g., ATXN1 and ATXN3). The clinical trial failures for Tominersen (GENERATION HD1) and the other ASOs (WVE-120101 and WVE-120102) highlight the challenges of ASOs as a treatment option for patients [298].

This review highlights polyQ disease proteins and their networks, putative toxic gain-of-function mechanisms of polyQ-expanded proteins, formation of protein aggregates, and perturbations in gene regulation. One area of research that has not been explored enough for potential treatments is targeting PPIs of polyQ disease proteins. This gap might stem from the lack of high-confidence interactors that are extensively characterized in terms of binding sites, binding affinity strengths, polyQ dependency, and relevance for modifying pathology. From our assessment of PPI data, some promising candidates would meet the above criteria. One notable example is the interaction between ATXN1 and CIC. The ATXN1/CIC complex is highly expressed in the most vulnerable neurons affected in disease [162]; and polyQ expansion of ATXN1 alters CIC’s repressor activity, resulting in a higher level of repression of CIC target genes [162, 164]. Moreover, mutations at the interface of the ATXN1-CIC complex were shown to prevent SCA1 pathogenesis [164]. Further studies, aiming to target the ATXN1-CIC interface through small molecules, peptide mimetics, or protein degradation technologies might pave the way for a new therapeutic approach for SCA1. Given the importance of the normal protein, the ideal approach would be to specifically target the mutant protein. In the case of the AR, the oral protein degrader ARV-110 is in clinical trials for prostate cancer [299]. This degrader, however, lacks mutant AR specificity and is unlikely to be suitable for the treatment of SBMA. Instead, the development of HSP90 or SRC kinase inhibitors (both AR interactors) might help to restore motor function in SBMA patients, as it has been demonstrated previously for SBMA mouse models [222, 226]. Targeting extended polyQ tracts in disease proteins and modulating their abnormal association with partner proteins might be a universal therapeutic strategy for all polyQ diseases described in this review.

ACKNOWLEDGMENTS

The authors would like to thank CHDI for their funding to study disease relevant HTT protein interactions to E.E.W. The authors also thank the Huntington’s Disease Society of America for funding support through the Berman-Topper Family HD Career Development Fellowship (to E.S.R.).

Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors have declared no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Erich E. Wanker https://orcid.org/0000-0001-8072-1630
Eduardo Silva Ramos https://orcid.org/0000-0002-8520-9107

REFERENCES


DRPLA proteins selectively interact with the enzyme GAPDH. Nature Medicine, 2, 347–350. https://doi.org/10.1038/nm0396-347


How to cite this article: Bonsor, M., Ammar, O., Schnoegel, S., Wanker, E. E., & Silva Ramos, E. (2024). Polyglutamine disease proteins: Commonalities and differences in interaction profiles and pathological effects. Proteomics, 24, e2300114. https://doi.org/10.1002/pron.2300114