Flanking regions determine the structure of the poly-glutamine homo- repeat in huntingtin through mechanisms common among glutamine-rich human proteins

Annika N Urbanek, Matija Popovic, Anna Morató, Alejandro N Estaña, Carlos A Elena-Real, Pablo Mier, Aurélie Fournet, Frédéric Allemand, Stéphane Delbecq, Miguel A Andrade-Navarro, et al.

Flanking regions determine the structure of the poly-glutamine homo-repeat in huntingtin through mechanisms common among glutamine-rich human proteins

Annika Urbanek[1,6], Matija Popovic[1,6], Anna Morató[1], Alejandro Estaña[1,2], Carlos A. Elena-Real[1], Pablo Mier[3], Aurélie Fournet[1], Frédéric Allemand[1], Stephane Delbecq[4], Miguel A. Andrade-Navarro[3], Juan Cortés[2], Nathalie Sibille[1], Pau Bernadó[1,5,*]

[1] Centre de Biochimie Structurale (CBS), INSERM, CNRS, Université de Montpellier. 34090 Montpellier, France.

[2] LAAS-CNRS, Université de Toulouse, CNRS. 31400 Toulouse, France.

[3] Institute of Organismic and Molecular Evolution, Faculty of Biology, Johannes Gutenberg University of Mainz. 55128 Mainz, Germany.

[4] Laboratoire de Biologie Cellulaire et Moléculaire (LBCM-EA4558 Vaccination Antiparasitaire), UFR Pharmacie, Université de Montpellier. 34090 Montpellier, France.

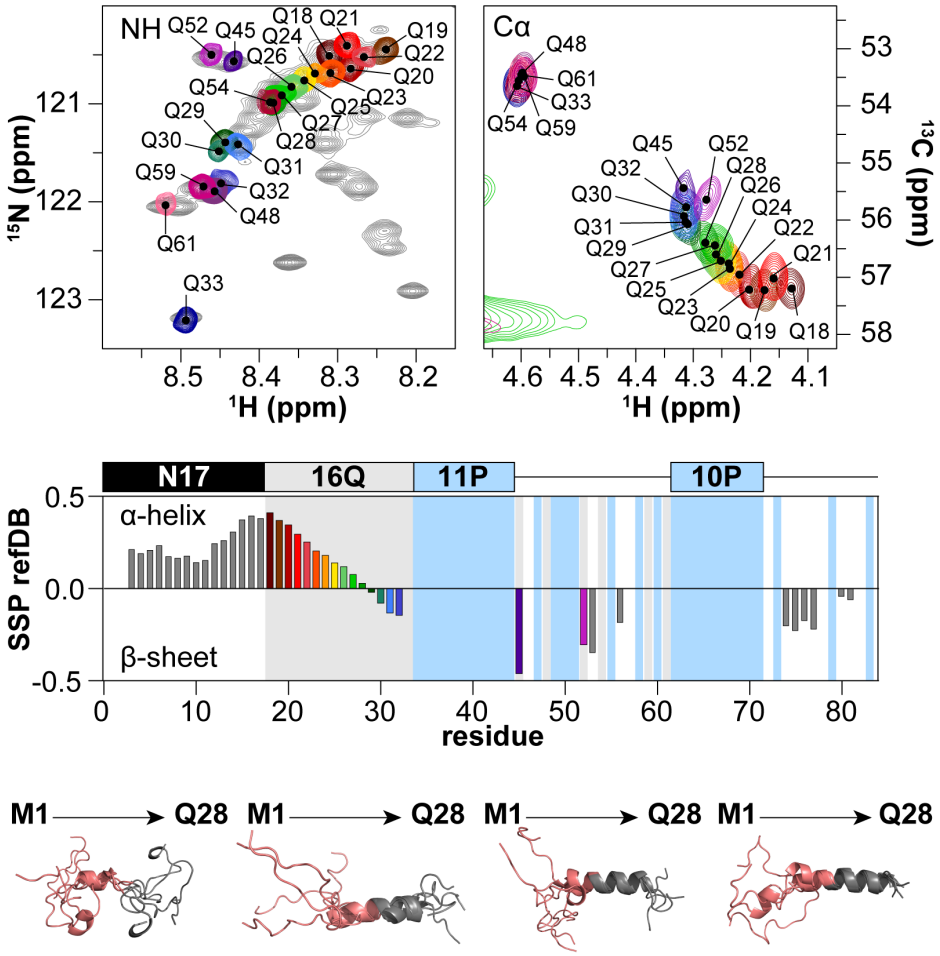
[5] Lead Contact

[6] These authors contributed equally

[*]Correspondence: Pau Bernadó (pau.bernado@cbs.cnrs.fr)

Summary

The causative agent of Huntington's disease, the poly-Q homo-repeat in the N-terminal region of huntingtin (httex1), is flanked by a 17-residue-long fragment (N17) and a proline-rich region (PRR), which by poorly understood mechanisms promote and inhibit the aggregation propensity of the protein, respectively. Based on experimental data obtained from site-specifically labeled NMR samples, we derived an ensemble model of httex1 that identified both flanking regions as opposing poly-Q secondary structure promoters. While N17 triggers helicity through a promiscuous hydrogen bond network involving the side chains of the first glutamines in the poly-Q tract, the PRR promotes extended conformations in neighboring glutamines. Furthermore, a bioinformatics analysis of the human proteome showed that these structural traits are present in many human glutamine-rich proteins and that they are more prevalent in proteins with longer poly-Q tracts. Taken together, these observations provide the structural bases to understand previous biophysical and functional data on httex1.



Introduction

Huntington's disease (HD) is a hereditary neurodegenerative disorder caused by an expansion of CAG triplet repeats beyond a pathological threshold. For HD, this expansion is located in the first exon of the huntingtin gene and results in an abnormally long poly-glutamine (poly-Q) tract within the N-terminus of the huntingtin protein (httex1)(Walker, 2007). When the number of consecutive glutamines exceeds 35, the resulting mutant protein forms large cytoplasmic and nuclear aggregates, a hallmark of HD, and causes neuronal degeneration, especially affecting the neurons of the striatum(DiFiglia et al., 1997; Hosp et al., 2017; Orr, 2001; Wanker, 2000). Aggregation, disease risk and age of onset correlate with the length of the poly-Q tract(Walker, 2007; Wanker, 2000). Interestingly, the aggregates predominantly contain mutant httex1 fragments, instead of the full-length protein, which comprises 3,142 amino acids in the non-pathogenic form. Indeed, it has been shown that the httex1 fragment alone is enough to reproduce the HD symptoms in mice(Mangiarini et al., 1996).

While the httex1 aggregation mechanism and the resulting β-sheet amyloid fibrils have been thoroughly characterized(Fiumara et al., 2010; Hoop et al., 2016; Isas et al., 2015; Jayaraman et al., 2012; Scherzinger et al., 1997; Shen et al., 2016), the structural bases of the pathological threshold and the mechanisms by which the native form of mutant httex1 give rise to toxicity and cell death are still poorly understood. Some clues regarding aggregation and pathogenicity of mutant httex1 have been found in the flanking regions of the poly-Q tract. The N-terminal domain, composed of 17 residues (N17) (Figure 1a), enhances aggregation of longer poly-Q tracts *in vitro* and *in vivo* and has been shown to form an amphipathic helix that interacts with membranes and chaperones(Ceccon et al., 2018; Kotler et al., 2019; Michalek et al., 2013; Scherzinger et al., 1997; Shen et al., 2016; Tam et al., 2009; Thakur et al., 2009). Moreover, post-translational modifications of N17 modulate huntingtin function, translocation, aggregation, and toxicity(Ansaloni et al., 2014; Atwal et al., 2011; Chiki et al., 2017; Ehrnhoefer et al., 2011; Mishra et al., 2012; Steffan et al., 2004). The poly-Q region is followed by a poly-proline (poly-P) tract of 11 consecutive prolines, which is part of the proline-rich region (PRR) containing 31 prolines in total (Figure 1a). In contrast to N17, the poly-P tract has a protective effect against aggregation *in vitro* and *in vivo,* but is necessary for the formation of visible aggregates in cells(Bhattacharyya et al., 2006; Dehay and Bertolotti, 2006; Shen et al., 2016; Steffan et al., 2004). This effect is directional, as N-terminal poly-P tracts do not attenuate the aggregation of poly-Q peptides(Bhattacharyya et al., 2006). It has also been shown that the flanking regions differently shape the aggregation pathways of pathological httex1, define the structure and stability of fibrils, and modulate its neuronal toxicity(Shen et al., 2016).

Two models linking poly-Q abnormal expansion and cytotoxicity have been proposed(Feng et al., 2018). The 'toxic structure' model proposes the appearance of a distinct toxic conformation when the tract expands beyond the pathological threshold(Miller et al., 2011; Nucifora et al., 2012; Peters-Libeu et al., 2012). The second model, the so-called 'linear lattice' model, suggests that even short poly-Qs

are inherently toxic and httex1 toxicity systematically increases with the tract length(Klein et al., 2013; Li et al., 2007; Owens et al., 2015). Evidence for both models has been obtained using monoclonal antibodies in cells expressing httex1 of different lengths(Bennett et al., 2002; Klein et al., 2013; Li et al., 2007; Owens et al., 2015; Peters-Libeu et al., 2012). However, this approach provides a very indirect perspective on httex1 conformations, and higher resolution information is required to discriminate between both hypotheses(Feng et al., 2018).

In a recent study, combining single-molecule FRET (smFRET) data with atomistic simulations, no sharp conformational change of monomeric httex1 around the pathological threshold could be observed, but rather a continuous global compaction with increasing poly-Q length induced by the interaction between N17 and the poly-Q tract was suggested(Newcombe et al., 2018; Warner et al., 2017). Recent circular dichroism (CD) and electronic paramagnetic resonance (EPR) experiments report on a systematic increase of the helical propensity and rigidity in httex1 when the poly-Q tract length increases(Bravo-Arredondo et al., 2018; Fodale et al., 2014). Observations from these *in vitro* studies are in coherence with the 'linear lattice' model. However, they only focused on the overall properties of the protein and could not probe httex1 at atomic resolution. Nuclear magnetic resonance (NMR) is the most suitable technique to provide a high-resolution picture of the conformational preferences of flexible proteins and structural characteristics of subpopulations of toxic conformers(Milles et al., 2018). However, NMR studies of httex1 are inherently challenging due to its strong compositional bias, which impedes residue-specific assignment and the measurements of structural constraints. Due to this challenge, only incomplete observations regarding the conformational preferences of the poly-Q and the flanking regions have been reported(Bravo-Arredondo et al., 2018; Newcombe et al., 2018; Thakur et al., 2009). All these NMR studies, independently of the poly-Q length, indicate a transient helical propensity encompassing N17 and the homo-repeat. Current structural models of httex1 suggest a compact overall arrangement in which N17 and the poly-Q tract interact through fuzzy contacts while the PRR sticks out. These tadpole-like structures display a systematic increase of the surface area with the length of the tract, also in line with the 'linear lattice' toxicity model(Newcombe et al., 2018; Warner et al., 2017). However, these models are based on sparse data or single conformation structural modeling.

In order to overcome the previously mentioned challenges, we have recently developed a methodology to site-specifically incorporate a single [^{15}N, ^{13}C]-labeled glutamine into proteins, and thereby obtain simplified NMR spectra(Urbanek et al., 2018). By systematically applying this site-specific isotopic labeling (SSIL) strategy, which combines cell-free protein expression(Kigawa et al., 1999) and nonsense suppression(Wang et al., 2006), we have obtained the NMR assignment at nearly physiological conditions of all non-proline residues in a httex1 construct containing 16 consecutive glutamines (H16). The ensemble modeling of the resulting chemical shifts demonstrated the presence of multiple, partially formed α-helical regions initiated in N17 and involving fragments of the poly-Q tract of different lengths. The application of SSIL to N17 and PRR mutants demonstrated that the

distinct conformational features of both flanking regions are propagated into the poly-Q tract, which acts as a conformationally versatile polypeptide. These observations provide the structural determinants underlying the key role of flanking regions in modulating the aggregation properties of httex1(Bhattacharyya et al., 2006; Jayaraman et al., 2012).

Results

Glutamine NMR scanning of H16

The monomeric httex1 that we characterized contained 16 glutamines in the poly-Q tract and another six in the PRR (Figure 1a). We produced H16 samples with glutamine-specific isotopic labeling using the SSIL strategy previously developed in our group(Urbanek et al., 2018). To streamline the preparation of the 22 H16 NMR samples, we first made sure that all samples could be prepared with similar efficiency by scanning all the TAG-mutated H16-sfGFP plasmids in a 96-well plate after addition of 10 µM glutamine loaded $tRNA_{CUA}$ (Figure 1b). All positions showed fluorescence intensities of ~30% of the positive control (H16 without amber stop codon), indicating that the efficiency of the incorporation of the labeled glutamine is independent of the specific sequence and the yield is similar to those achieved in other studies(Ellman et al., 1992; Peuker et al., 2016). Once the suppression efficiency was verified at a small scale, the CF reaction volume was increased to 5 mL to produce the NMR samples.

The ^{15}N-HSQC of H16 displayed the typical features of poly-Q-containing proteins(Baias et al., 2017; Bravo-Arredondo et al., 2018; Eftekharzadeh et al., 2016; Newcombe et al., 2018). While peaks from N17 and the PRR are well dispersed, a large density of unresolved peaks corresponding to glutamine residues was observed (Figure 2a). In order to disentangle this massive overlap we measured ^{15}N- and ^{13}C-HSQCs of the SSIL H16 samples containing a single [^{15}N, ^{13}C]-labeled glutamine. As observed in Figure 2b, the glutamines adjacent to N17 (Q18-Q21) appear in the upfield region of the poly-Q density without any specific trend. The following glutamines (Q22-Q28) display a consistent ^{1}H and ^{15}N downfield shift, indicating a systematic structural change along the homo-repeat. A large deshielding effect is subsequently observed for Q29, Q30 and Q31, which are strongly overlapped. Finally, the last two glutamines of the tract, Q32 and Q33, display isolated peaks induced by the proximity of the downstream poly-P. The chemical shifts of glutamines in the PRR are more dispersed due to their different neighboring residues. Cα-Hα correlations measured in the same SSIL samples follow similar trends along the poly-Q tract (Figure 2c).

α-helical propensity in N17 and the poly-Q tract

The Cα and Cβ chemical shifts measured for all glutamines in this study and the previously reported assignment of H16(Urbanek et al., 2018) allowed the determination of the structural propensities of H16. The secondary chemical shift (SCS) analysis using a neighbor-corrected random coil

database(Nielsen and Mulder, 2018) indicates that both N17 and the poly-Q tract are enriched in α-helical conformations, although this propensity is not homogeneous (Figure 2d). Helicity increases along N17, reaching its maximum at the first glutamine, Q18, and subsequently decreases smoothly. A transition is observed at Q29, which adopts a small and negative SCS value. This extends to the following three glutamines, indicating the presence of random coil or slightly extended conformations. This conformational transition is pinpointed in the secondary structure propensity (SSP) analysis(Marsh et al., 2006; Zhang et al., 2003) (Figure 2e). Note that the helical propensity of the N-terminal part of H16 remains below 40%, in agreement with similar analyses using an httex1 fragment with 17 glutamines and the partially assigned httex1 with 25 glutamines(Baias et al., 2017; Newcombe et al., 2018). The C-terminal region of H16 presents negative SCS values, probably reflecting the enrichment in polyproline-II conformations induced by the large number of prolines(Isas et al., 2015).

The ensemble model of H16 reveals a conformational equilibrium involving multiple α-helices

The ensemble structure of H16 was investigated by combining the backbone NMR chemical shifts and a recently developed approach to build realistic ensemble models of intrinsically disordered proteins(Estaña et al., 2019). Briefly, our method appends residues, which are considered to be either fully disordered or partially structured, to build the complete chain without steric clashes. For fully disordered residues, amino acid specific ϕ/ψ angles defining the residue conformation are randomly selected from the database, disregarding their flanking residues. For partially structured residues, the nature and the conformation of the flanking residues are taken into account when selecting the conformation of the incorporated residue (see detailed explanation of the algorithm in the original publication(Estaña et al., 2019)). Two families of ensembles were built to investigate the conformational influence of both flanking regions of H16. For the first family (N→C ensembles), starting with the ^{10}AFESLKSF^{17} region of N17 as partially structured, multiple ensembles of 5,000 conformations were built by successively including an increasing number of glutamines in the poly-Q tract (from Q18 to Q33) as partially structured, while the rest of the chain was considered to be fully disordered. Note that in the partially structured building strategy secondary structural elements are propagated due to the neighboring effects. An equivalent strategy was followed for the second family of ensembles (N←C ensembles) for which glutamines were considered successively as partially structured from the poly-P tract (from Q33 to Q18). For the resulting 17 ensembles of each family, and after building the side chains with the program SCWRL4(Krivov et al., 2009), averaged Cα and Cβ chemical shifts were computed with SPARTA+(Shen and Bax, 2010) and compared with the experimental ones (Figure S1).

Theoretical Cα chemical shifts for the poly-Q tract present different values for regions built as partially structured (influenced by the flanking regions) or disordered. Three Cα CS plateaus are observed corresponding to α-helical, extended and random coil conformations, and transitions are observed between regions built as disordered and influenced by the flanking regions (Figure S1). Not

surprisingly, the Cβ chemical shifts turned out to be less sensitive to the presence of structured regions in the homo-repeat region (Figure S1). These simulations indicate that flanking regions induce a distinct conformational bias to the neighboring glutamines. While N17 induces helical conformations with Cα chemical shift values larger than those usually observed for a random coil (N→C ensembles), the poly-P tract enriches the ensemble with extended conformations with smaller Cα chemical shift values compared to a random coil (N←C ensembles). However, the simulated conformational ensembles fail to reproduce the chemical shifts measured in H16, indicating that our simple sampling strategy cannot simultaneously describe the structural influence exerted by both flanking regions.

A third ensemble model of H16 was built by reweighting the populations of the pre-computed ensembles, using the experimental Cα and Cβ chemical shifts as constraints. In order to capture the influence of the flanking regions, glutamines within the tract were divided into two groups: those influenced by N17 and those influenced by the poly-P tract, whose chemical shifts were fitted with the N→C and N←C ensembles, respectively. The limit between both families was systematically explored, reaching an optimal description of the experimental chemical shifts when Q28 was chosen as the last residue structurally connected with N17 (Figure S1). Importantly, the optimization, which was performed through a Monte-Carlo procedure, was repeated multiple times always yielded equivalent populations. The resulting ensemble nicely described the complete Cα and Cβ CS profiles for H16 (Figure S2). Importantly, the systematic decrease of the Cα chemical shifts along the poly-Q tract and the flat profile observed for the Cβ chemical shifts were well reproduced, indicating that the refined ensemble captures the structural features of the homo-repeat and the distinct conformational perturbations exerted by both flanking regions.

The conformational properties of the optimized ensemble were subsequently investigated in detail. First, we explored the conformational preferences of individual glutamines using Ramachandran plots (Figures 3 and S3a). While the first four glutamines of the tract (Q18-Q21) displayed a strong enrichment in helical conformations (Figure 3a), the last four (Q30-Q33) preferred extended ones (Figure 3b). The conformational preferences along the tract, calculated from the derived ensemble, indicate a systematic decrease in the helical population from ~65% (Q18) to ~50% (Q28) (Figure 3c). In line with the NMR measurements (Figure 2), a sharp conformational transition is observed for Q29, which is the first residue displaying a preference for extended conformations.

The cooperativity between the residue-specific conformations to form stable α-helices was analyzed using the secondary structure map (SS-map) tool(Iglesias et al., 2013). The fragment encompassing N17 and the poly-Q tract can be described as a complex equilibrium of multiple co-existing α-helices of variable length (Figure 3d). The core of this family of helical structures includes the last four residues of N17 and the first two glutamines of the homo-repeat. The last residues of N17 act as nucleation points for the helices that afterward extend to include a variable number of glutamines of the tract, giving a triangular shape to the SS-map. According to our analysis, no α-helices are nucleated within the poly-Q tract and, as a consequence, helices involving inner glutamines belong

only to lowly populated long helical elements. This is shown in Figures 3e and S3b, which display representative conformations and the α-helical fragments of the four sub-ensembles selected to describe the NMR CSs. Three of these ensembles present α-helices that encompass the last residues of N17 and the first residues of the poly-Q. No persistent turns in the residues connecting both domains are observed, which would otherwise yield a strong signature in the chemical shift profile. As a consequence, H16 should be considered as an elongated flexible particle, in contrast to the previously proposed compact tadpole-like model(Newcombe et al., 2018; Warner et al., 2017).

Glutamine side chains indicate a structural coupling of N17 and the poly-Q tract

According to our ensemble model, the last four residues of N17 are strongly linked to the first two glutamines of the poly-Q tract. However, the model, which is based on backbone CSs, does not unveil the structural bases of this structural connection. Benefitting from the lack of signal overlap in the ^{13}C-HSQC of the SSIL samples, glutamine-specific Cβ-H_2 and Cγ-H_2 correlations could be analyzed (Figures 4 and S4). As expected for a flexible protein, the majority of glutamines in H16 display two correlation peaks for Cβ-H_2 and a single one for Cγ-H_2, indicating increased mobility along the side chain (Figure S4). Interestingly, the first four glutamines, Q18-Q21, present different spectroscopic features. While Q18 and Q19 display a single peak for Cβ-H_2 and Cγ-H_2, these correlations are split in two for Q20 and Q21 (Figure 4a). Most probably, the splitting of Cγ-H_2 is caused by the rigidification of the glutamine side chains, which results in a different chemical environment for the two diastereotopic Hγ atoms. This rigidification likely originates from the formation of a hydrogen bond between the side chain amide group and the backbone of a neighboring residue. Notice that similar spectroscopic features were observed in a recent characterization of the androgen receptor (AR) N-terminal domain fragments hosting poly-Q tracts of different lengths(Escobedo et al., 2019).

In order to substantiate this hypothesis and profiting that Q20 and Q21 Nε-H_{21} displayed isolated peaks (see below), we determined the temperature coefficients (σH^N/T) for these two atoms in a ^{15}N-labeled H16 sample (Figure S4). We derived σH^N/T values of -4.1 and -3.5 ppb/K for Q20 and Q21 $H^N\varepsilon_{21}$, respectively. These values are less negative than the threshold value, −4.5 ppb/K, suggesting their participation in a hydrogen bond(Baxter and Williamson, 1997). Conversely, we obtained σH^N/T values of -5.8 and -6.2 for Q32 and Q54 $H^N\varepsilon_{21}$, respectively, confirming the singularity of the first glutamines of the tract.

Multiple α-helical N-capping hydrogen bonding networks involving glutamine side chains have been described(Dasgupta and Bell, 1993; Newell, 2015; Richardson and Richardson, 1988; Seale et al., 1994). In the AR study, the authors propose a bifurcated hydrogen bond where the amide backbone of residue *i*-4 simultaneously forms hydrogen bonds with the backbone and the side chain of glutamine in position *i*(Escobedo et al., 2019). Indeed, in this novel mechanism, the side chain hydrogen bond further stabilizes the canonical (*i-4* → *i*) backbone helical hydrogen network. This interaction would be protected by the hydrophobic side chain of residue *i-4*, a leucine in AR(Gao et al., 2009).

According to this model and in the context of huntingtin, the side chain amide groups of Q20 and Q21 would form hydrogen bonds with S16 and F17, respectively, the latter one being the most stable interaction according to the extent of the Cγ-H_2 splitting. The Nε-H_{21} peaks for Q20 and Q21, which appear clearly shifted from the other side-chain peaks, further substantiate this feature (Figure 4b). The frequency shift for these two peaks cannot be only attributed to the involvement of these two atoms in an α-helical hydrogen bond, whose signature is a ^{15}N upfield shift(Escobedo et al., 2019). An alternative explanation is the ring current effects exerted by F17 that, upon formation of the canonical hydrogen bond with Q21, places its side chain in the proximity of Q21 Nε-H_{21} and to lesser extent to Q20 Nε-H_{21}. Note that the magnitude of the ring current shift is difficult to anticipate as it depends on persistence and the orientation of the aromatic side chain with respect to the shifted atom. Conversely, Q18 Nε-H_{21}, which is adjacent to F17 in the sequence, is not affected by the presence of the aromatic side chain. This last observation, which is in line with the protective role of the phenylalanine hydrophobic side chain, underpins the structural coupling between the N17 and the poly-Q tract through a hydrogen bonds network.

Mutants reveal the effects of N17 side chains on structural coupling

In order to further investigate the structural bases of the connection between the N17 and the poly-Q domains, we designed three H16 mutants in which the last residues of N17 (^{14}LKSF^{17}) were mutated to ^{14}LKGG^{17}, ^{14}LLLF^{17} and ^{14}LKAA^{17} (Figures 1a and 5). The LKGG and LKAA mutants were designed to weaken to different extents the hydrogen bond network found in wild-type (wt) httex1, while the LLLF would strengthen the network. The ^{15}N-HSQC spectrum of the LKGG mutant presented very clear differences with respect to the wt one, especially in the glutamine region (Figures 5a and S5). The relatively disperse glutamine peaks of wt H16 coalesced in a broad, high-intensity downfield-shifted peak. Furthermore, the dispersion of the Nε2-H_2 side chain signals in LKGG-H16 was dramatically reduced (Figure S5). These observations demonstrated that the helical nature of the poly-Q tract is lost when mutating the last two residues of N17 to glycine. The origin of the dramatic structural changes was investigated using the SSIL strategy by isotopically labeling residues Q18, Q20 and Q21 of LKGG-H16 (Figure 5b). Compared to H16, the three residues present very different peak positions in both spectra. While the N-H correlation of Q18 was strongly influenced by the neighboring glycines, Q20 and Q21 appeared shifted downfield, in the same position as the broad glutamine peak (Figure 5a). Cα-Hα correlation peaks for these three residues were strongly shifted towards a less helical region of the spectrum (Figure 5b). The SCS analysis of these three residues indicated that the helicity was severely reduced compared to wt H16 but not completely abolished, indicating that the poly-Q tract is slightly helical for this mutant (Figure 5l). For the three glutamines, Cβ-H_2 and Cγ-H_2 correlations presented a doublet and a singlet, respectively (Figure 5c), indicating the loss of the hydrogen bonds connecting the N17 to the poly-Q. However, it was unclear whether the absence of this structural coupling affected the inherent helical tendency of N17. To resolve this point

we assigned the N17 region of LKGG-H16, using traditional 3D-NMR experiments, and computed the SCSs (Figure S5d). Comparison of the wt and LKGG-H16 SCS analyses showed that the double point mutation is resulting in a bidirectional loss of helicity, impacting the last six residues of N17 as well as the following glutamines.

The mutant LLLF-H16 was designed to provide new sites to the first glutamines of the tract to form side chain hydrogen bonds and thus strengthen the helical tendency of the homo-repeat. Glutamine peaks of the LLLF-H16 ^{15}N-HSQC spectrum presented an additional upfield density that was attributed to an increased helical content in this mutant (Figure 5d). SSIL samples for Q18, Q20 and Q21 displayed important chemical shift changes in both the N-H and the Cα-Hα correlations (Figure 5e). In fact, Q20 and Q21 N-H and Cα-Hα peaks appear shifted towards more helical conformations with respect to the wt. Unfortunately, the Cα-Hα peak for Q18 could not be observed, most probably due to a folding/unfolding process in the μs to ms dynamic regime that broadens the peak beyond detection. The SCS analysis showed a strong α-helical increase for Q20 and Q21, substantiating the above-mentioned qualitative observations regarding the helical increase for this mutant (Figure 5l). Despite their overall low intensity, the Cβ-H_2 and Cγ-H_2 peaks demonstrate a stronger structural coupling between N17 and the poly-Q tract. The Cγ-H_2 splitting of Q20 and Q21 is larger than that observed in the wt. Cβ-H_2 presents a single peak for Q18 and Q20, something occurring only for Q18 and Q19 in the wt (Figure 5f, 4a), indicating a stronger hydrogen bond network involving additional residues. Therefore, the LLLF-H16 mutant unambiguously links the strength of the hydrogen bond network between N17 and the first glutamines of the homo-repeat with the persistence and stability of the resulting α-helices.

The third mutant, LKAA-H16, was designed to display an intermediate behavior with respect to the other two. Alanine is a helical promoter amino acid but its side chain is smaller than those of leucine and phenylalanine. The LKAA-H16 ^{15}N-HSQC spectrum was similar to the wt one, although less density was observed in the upfield part of the glutamine spectral region (Figure 5g and S5). The Cα-Hα peaks for Q18, Q20 and Q21 were shifted downfield in the ^{1}H dimension with respect to those of the wt (Figure 5h). This feature was quantified in the SCS analysis, which indicates a decrease in the helical tendency for Q18 and Q20, while Q21 remained almost unchanged. Exploration of the side chains of these three residues suggested some clues to this observation. Interestingly, only Q18 presented two Cγ-H_2 peaks, indicating a hydrogen bond between the side chain of this residue and the backbone of L14. Therefore, the structural connectivity is modified in LKAA-H16 by exchanging the two side chain hydrogen bonds present in the wt by a new one involving the first glutamine of the tract and concomitantly a decrease of the helical tendency for this mutant.

The inspection of the Nε-H_2 peaks of the suppressed samples further substantiates the structural model of the hydrogen bond connection (Figure S5c). Q21 Nε-H_{21} peak of LLLF-H16 displays a stronger upfield shift in the ^{1}H dimension than in the wt, suggesting more persistent ring current effect by F17 aromatic ring caused by the formation of a more stable hydrogen bond. This enhanced stabilization of

the α-helix is also manifested in the Q18 Nε-H_{21} peak that now appears strongly upfield shifted in the ^{15}N dimension. In LKAA-H16, where F17 is mutated by an alanine, the Nε-H_{21} peaks of Q18, Q20 and Q21 are not displaced in the ^{1}H dimension despite the fact that they are involved in an α-helix, demonstrating that the ring current effects are at the origin of the unusual frequencies of Nε-H_{21} atoms in httex1.

The poly-P C-terminal flanking region breaks the helical tendency of the glutamine homo-repeat

In order to explore the structural connection between the poly-Q and the poly-P homo-repeats, we designed a mutant with five glycines between these tracts (H16-5G), aiming to structurally uncouple them (Figure 1a)(Bhattacharyya et al., 2006). This mutant yielded a very similar ^{15}N-HSQC spectrum to that of H16, with glutamine peaks displaying an equivalent level of dispersion (Figure 5j and S5). No relevant differences were observed in the backbone or side chain correlations between both spectra, suggesting that the presence of the five glycines does not perturb the overall structure of H16. Nevertheless, we prepared an SSIL H16-5G sample with [^{15}N, ^{13}C]-glutamine in position Q30, which lies in the non-helical part of the poly-Q tract of H16, to investigate structural changes resulting from uncoupling both homo-repeats at residue level. In comparison with the wt, the N-H correlations were shifted upfield, whereas the Cα-Hα correlations were shifted downfield in the ^{1}H and upfield in the ^{13}C dimension (Figure 5k). This observation suggested an increase in the helical tendency of this residue in the new context, which was quantitatively proven by SCS analysis. Q30 adopts a positive SCS value in H16-5G while in the wt this residue has a slightly negative value (Figure 5l). This observation demonstrates that the poly-P tract in httex1 exerts a strong conformational perturbation on the neighboring glutamines by enriching the ensemble with extended conformations, which break the inherent helical propensity of the poly-Q.

Sequence analyses of poly-Q flanking regions in human proteins

In a previous bioinformatics analysis it was shown that leucines, prolines and histidines were especially enriched in the flanking regions of human poly-Q tracts(Ramazzotti et al., 2012). While leucine and histidine were similarly enriched on both sides, proline displayed a preference for the C-flanking region. We complemented this study by exploring whether the compositional bias in the flanking regions was poly-Q length dependent. For that, four hundred fragments with ten or more glutamine residues and containing a maximum of two non-glutamine residues were collected from 309 different human proteins, and the ten preceding (-10 to -1) and succeeding (+1 to +10) residues were compositionally analyzed. Figure S6 shows that using our poly-Q definition (maximum of 2 non-glutamine residues in fragments of 10 or more glutamine residues), we obtain similar results as those derived by Ramazzotti *et al.*(Ramazzotti et al., 2012), with leucine, proline and to a lesser extent histidine and alanine being enriched in poly-Q flanking regions, as well as the positional asymmetry of proline. Interestingly, using our poly-Q definition we identify an enhanced enrichment of leucines in

the N-flanking region compared with the C-flanking one. Note that a less restrictive definition of the homo-repeat to include larger glutamine-rich regions was used in the previous study and this could lead to changes in the enrichment levels.

We then analyzed the effect of the length of the glutamine homo-repeats on the above-described compositional biases by selecting pure glutamine stretches. The leucine population in position -1 increases with the length of the poly-Q tract, reaching a maximum of 30.0% when the number of consecutive glutamines in the tract is seven or more, and it is slightly reduced for longer homo-repeats (Figure 6a). Interestingly, positions from -2 to -4 also display a similar length dependency, although the enrichment is less prominent than in position -1. The population of prolines in the C-flanking region systematically increases with the length of the poly-Q tract. The maximum of the enrichment occurs at position +1 that extends over the complete region, while it remains close to the background in the N-flanking region (Figure 6b).

Next, we explored the secondary structure propensity in the N-flanking region of long human poly-Q tracts with a recently developed approach (manuscript in preparation) based on the previously mentioned large database of three-residue fragments(Estaña et al., 2019). Briefly, the residue-specific conformational bias was evaluated accounting for the effects exerted by the preceding and succeeding amino acids. Then, the percentage of α-helical, extended or other conformations was derived. The position-specific percentages obtained for each family were averaged in increasing sections of the N-flanking regions and reported as notched box plots in Figure 6c. For each fragment, the α-helical conformation was preferred with median values ranging from 50.2% to 70.6%, while the preference for extended or other conformations was always lower than 25%. Interestingly, the α-helical tendency presents its largest percentage when close to the poly-Q homo-repeat (residues -1 and -2), and systematically decreases when more residues of the N-flanking region are incorporated in the analysis.

In summary, these sequence analyses indicate that the structural and compositional characteristics observed in httex1 flanking regions are shared by a large number of other human poly-Q-containing proteins. This observation suggests that the structure-mediated functional mechanisms found for httex1 in the present study are common to many other human glutamine-rich proteins.

Discussion

In this study, we demonstrate that the previously developed SSIL strategy(Urbanek et al., 2018) can be systematically applied to investigate poly-Q tracts, one of the most abundant homo-repeats in eukaryotes(Jorda and Kajava, 2010; Lobanov and Galzitskaya, 2012; Mier et al., 2017), and to connect their structural features with their specialized biological functions. The NMR analysis of the SSIL samples demonstrates that H16 is disordered, but hosts an important level of helicity that is initiated in N17, reaching the maximum at the beginning of the poly-Q tract and smoothly vanishing along the homo-repeat. A conformational ensemble model refined from experimental data recapitulates this non-

uniform helical propensity as an equilibrium of multiple canonical helices of different lengths. All these helices start in N17 and extend towards the poly-Q tract, comprising an increasing number of glutamines. Q28 is the last glutamine influenced by the α-helical tendency, and subsequent glutamines present random coil or slightly extended conformations. The enrichment in α-helical conformations in httex1 is in agreement with crystallographic structures(De Genst et al., 2015; Kim et al., 2009) and NMR data(Baias et al., 2017; Newcombe et al., 2018). However, the non-homogeneous helicity can only be captured when an ensemble representation is used, as done in the present study.

Our NMR measurements demonstrate that N17 has an inherent α-helical tendency that is transferred to the glutamine homo-repeat through a hydrogen bond network involving glutamine side chains. Although the structure of this network cannot be unambiguously resolved with our NMR data, a recent study on the poly-Q tract of the AR demonstrates that glutamine side chains form hydrogen bonds with hydrophobic residues in the *i-4* position, reinforcing the canonical $CO_{i-4} \rightarrow H_{N,i}$ backbone hydrogen bond(Escobedo et al., 2019). In this study it was suggested that the large and hydrophobic residues in the *i-4* position were key for the formation of the bifurcated hydrogen bond by protecting it from water molecules. In the context of H16, the last two residues of N17, ^{16}SF^{17}, would play the main role in stabilizing and propagating the helix within the poly-Q tract. We have validated this model by monitoring the side chain CSs of three mutants in which we modified the last residues in N17 and profiting the chemical shift changes induced by the ring current effects of F17 to spatially close atoms. While the LLLF-H16 mutant strengthens the structural coupling between N17 and the poly-Q tract, the LKGG-H16 mutant is unable to form the hydrogen bond network. Interestingly, LKAA-H16 provides evidence of the malleability of this helical propagation. For this mutant, hydrogen bonds involving ^{20}QQ^{21} are hampered by the absence of large hydrophobic amino acids in positions *i-4* and, instead, this mutant utilizes L14 and Q18 to trigger the structural coupling between both regions. In addition, these results highlight that the conformational nature of the residues involved in the hydrogen bond network is important. In that sense, despite not forming bifurcate hydrogen bonds, the inherent helical propensity of alanines is required to connect N17 with the poly-Q tract, a phenomenon that is not observed in the LKGG-H16 mutant. These observations suggest that the residue preceding the poly-Q tract (position -1 according to our nomenclature) is the preferred one to trigger helicity in the homo-repeat. Consequently, the large population of leucines in this position found here and in a previous bioinformatics analysis of eukaryotic proteomes strongly suggests the generality of helical induction in poly-Q tracts through side chain hydrogen bonds(Ramazzotti et al., 2012). Interestingly, this enrichment increases for poly-Q tracts with seven or more consecutive glutamines (Figure 6a). Altogether, these observations point towards a general structure/function relationship for poly-Q fragments involving long α-helices of variable length and stability, depending on the residues preceding the tract. This observation is in line with the recurrent presence of coiled-coils in protein fragments containing poly-Q tracts as well as in their corresponding partners(Fiumara et al., 2010).

Multiple post-translational modifications have been described for N17, including phosphorylation, acetylation, ubiquitination and SUMOylation, and it has been shown that their presence perturbs the function, aggregation properties and toxicity of huntingtin(Chiki et al., 2017; Ehrnhoefer et al., 2011). According to our observations, modifications that decrease the helical propensity of N17 or break the hydrogen-bond network will induce an increase in disorder in the poly-Q tract. In a recent study, it was demonstrated that mono-phosphorylation on S13 or S16 and di-phosphorylation strongly disrupt N17 helicity. Interestingly, these post-translationally modified forms of httex1 are less prone to aggregation than the unmodified form(DeGuire et al., 2018). These observations can now be rationalized in the light of our results, indicating a strong link between the level of structure, aggregation and modulation through post-translational modifications.

It is well known that due to the limited conformational variability and the inability to form hydrogen bonds, proline is considered to be a structure-breaking residue with the capacity to extend its structural influence towards neighboring residues(Theillet et al., 2013). Previous CD experiments on httex1-mimicking peptides demonstrated the enrichment of polyproline-II conformations in poly-Q tracts preceding poly-P(Darnell et al., 2007). Here, we could demonstrate this effect at residue level through the NMR-driven molecular modeling of httex1 and by monitoring the CS changes in the H16-5G mutant. Moreover, our NMR analysis enables the assessment of the extent of structural perturbation exerted by the poly-P over the poly-Q tract. The last five glutamines of the tract preferentially adopt random coil or slightly extended conformations due to the influence of the proline tract(MacArthur and Thornton, 1991). However, this influence extends much further and causes the smooth decay of the helicity along most of the poly-Q tract in H16. Indeed, recent CD experiments as well as partial NMR assignments of httex1 variants with longer homo-repeats show that the helical content of httex1 systematically increases with the length of the poly-Q(Bravo-Arredondo et al., 2018; Fodale et al., 2014; Newcombe et al., 2018). The ensemble of these observations suggests that the perturbation exerted by the poly-P tract has a defined range of influence and, therefore, the poly-Q homo-repeat remains helical in the region preceding the perturbed segment. According to the ensemble of these studies, we can estimate that the conformational influence of the poly-P tract extends to the last 13 glutamines of httex1. Glutamines lying in this perturbed region sense a distinct structural influence from both sides, the helical propagation from the N-terminus and the helix-breaking tendency from the C-terminus. These opposing influences are captured in a different balance between α-helix and extended conformations in the individual Ramachandran plots displayed in Figures 3a,b and S3.

Sequence analyses also demonstrate that the presence of prolines at the C-terminal flanking region of glutamine-rich segments is common in eukaryotic proteins and especially significant in the positions immediately adjacent to poly-Q tracts(Ramazzotti et al., 2012). Here we show that in human proteins the extent of this proline compositional bias is poly-Q length dependent, meaning that proteins having longer poly-Q tracts have a higher probability to be followed by prolines. Interestingly, an examination of huntingtin orthologs shows that the poly-P occurs only in species with four or more

consecutive glutamines, suggesting that these two homo-repeats have coevolved(Schaefer et al., 2012; Tartari et al., 2008). The consecutive presence of glutamine and proline repeats is also observed in ataxin-2 and ataxin-7, two proteins whose abnormal poly-Q expansion causes spinocerebellar ataxias SCA2 and SCA7, respectively(Darling and Uversky, 2017). This concatenation of glutamine- and proline-rich regions in unrelated proteins from different organisms suggests a strong selective pressure at the molecular level and a common structure/function mechanism(Ramazzotti et al., 2012). For many of these proteins this mechanism might be the protection from aggregation of the expanded poly-Q tracts that arises from the conformational influence exerted by proline-rich regions. Prolines at the C-terminus shorten the length of the helical fragments of the poly-Q tract, reducing the stability of the intermolecular interactions and the subsequent aggregation.

Our results point to an overall extended structure of httex1 that is in contrast to the tadpole-like model where N17 and the poly-Q tract form a compact structure stabilized by fuzzy contacts from which the semi-rigid PRR sticks out(Newcombe et al., 2018; Warner et al., 2017). The compact httex1 structure has been derived from computational studies and sparse distance restrains derived from smFRET(Warner et al., 2017; Williamson et al., 2010). Although our experimental data do not report on long-range contacts, the hydrogen network involving N17 and the poly-Q tract, as well as the absence of the spectroscopic features of a turn in the interphase between both domains, strongly privileges the extended model over the compact one. Despite the overall extendedness, our data show that httex1 remains highly disordered, especially the last glutamines of the poly-Q tract and the PRR region. This flexibility would allow transient contacts between remote parts of the protein that could be at the origin of the long-range contacts observed in smFRET experiments(Warner et al., 2017). This extended structure supports the 'linear lattice' model of toxicity in which the number of exposed glutamines increases with the length of the tract. However, the emergence of a toxic conformation, appearing after the formation of soluble oligomers as previously suggested(Shen et al., 2016), is also compatible with our model, which focuses in the monomeric form of httex1.

The relatively low stability of helical conformations observed in H16, where multiple low-populated helices of different length co-exist, most probably regulates the capacity of httex1 to recognize its partners through coiled-coil interactions(Fiumara et al., 2010). When the number of glutamines exceeds the pathological threshold, the protective effect of prolines does not impede the presence of long α-helices. We can speculate that these long poly-Q helices could form coiled-coil interactions with other non-biological partners, sequester them and perturb natural signaling or metabolic pathways. This phenomenon could explain the long list of symptoms observed in HD patients(Walker, 2007). In terms of the oligomerization capacity, longer helices can form more stable assemblies, which could eventually nucleate the formation of the β-stranded amyloidogenic fibrils found in patients' brains. The correlation between poly-Q length, disease severity and age of onset could be explained by the enhanced stability of these long poly-Q oligomers.

From a practical point of view, our observations warn about the use of isolated poly-Q peptides

disregarding the sequence context to predict the biophysical/structural behavior and the aggregation propensity of glutamine-rich proteins(Crick et al., 2006; Walters and Murphy, 2009). We demonstrate that the chemical and structural features of poly-Q flanking regions govern the conformational behavior of the homo-repeat. Therefore, biophysical studies on poly-Q containing proteins must be performed with fragments including the relevant neighboring elements. With the SSIL approach these protein-specific properties can be now addressed at high resolution in order to unveil among other features the origin of the different pathological thresholds observed in poly-Q related diseases(Zoghbi and Orr, 2000).

Altogether, our data demonstrates that the poly-Q tract in httex1 is exposed to opposing structural effects from both flanking regions. Notably, the enrichment in hydrophobic residues and the α-helical conformations in the N-flanking region, as well as the downstream enrichment in prolines, are shared by many eukaryotic glutamine-rich proteins. This suggests that many proteins exploit these structural properties, which are centered on the structural flexibility and versatility of poly-Q tracts, in order to perform specific biological functions while avoiding aggregation and toxicity.

Acknowledgements

The authors thank Gottfried Otting for providing the BL21 (DE3) Star::RF1-CBD3 strain and Grayson Gerlich for reading the manuscript. This work was supported by the European Research Council under the European Union's H2020 Framework Programme (2014-2020) / ERC Grant agreement n° [648030], Labex EpiGenMed, an « Investissements d'avenir » program (ANR-10-LABX-12-01) awarded to PB, and GPCteR (ANR-17-CE11-0022-01) to NS. The CBS is a member of France-BioImaging (FBI) and the French Infrastructure for Integrated Structural Biology (FRISBI), 2 national infrastructures supported by the French National Research Agency (ANR-10-INBS-04-01 and ANR-10-INBS-05, respectively). AU is supported by a grant from the Fondation pour la Recherche Médicale (SPF20150934061). The authors thank Lionel Imbert, IBS cell-free facility, for his technical help and valuable advice. This work used the Cell-Free facility at the Grenoble Instruct Centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from Instruct (PID: 1552) within the Grenoble Partnership for Structural Biology (PSB). This work benefited from the HPC resources of the CALMIP supercomputing center under the allocation 2016-P16032.

Author Contributions

P.B. and S.D. conceived the project. F.A. and A.F. designed and cloned all the constructs and produced GLN4. P.B., A.U. and A.M. designed and analyzed the suppression experiments. A.U., M.P., A.M. and C.A.E-R. prepared the protein NMR samples. N.S. designed, and M.P., A.U., C.A.E-

R. and N.S. performed and analyzed the NMR experiments. A.E. performed the computational experiments and A.E., J.C., and P.B. analyzed these data. P.M. and M.A.A-N. performed the bioinformatics analyses. A.U. and P.B. wrote the paper with input from all coauthors.

Declaration of Interests

The authors declare no conflict of interest.

STAR Methods

Resource availability

Plasmids generated in this study are available from the Lead Contact without. This study did not generate new unique reagents.

Data and Code availability

This study did not generate any complete datasets or code.

Experimental Model and Subject Details

BL21 (DE3)

Starter cultures in LB medium supplemented with kanamycin (50 µg/mL) were incubated at 37ºC overnight and used for inoculation of the expression culture. For protein expression, cells were cultured overnight in ZYM 5052 auto-inducing medium supplemented with 50 µg/mL kanamycin in shaker flasks at 25ºC with constitutive shaking. Cells were harvested the next day.

BL21 Star (DE3)::RF1-CBD_3

In BL21 Star (DE3)::RF1-CBD_3 cells the genomic release factor 1 (RF1) is tagged with three chitin binding domains (CBD_3). Starter cultures in Z-medium supplemented with kanamycin (50 µg/mL) were incubated at 37ºC until an OD_{600} of ~1 was reached and used to inoculate the fermenter. Cultures to obtain lysate were grown at 37°C in a fermenter with 3 L of Z-medium with added 110 mM glucose, 10 mg/L thiamine, 1 mM $MgSO_4$ and 50 µg/mL kanamycin. When the OD_{600} reached ~1, 1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce T7 RNA polymerase synthesis. The cells were harvested in the mid-log phase to proceed with lysate preparation.

Key Resources Table

Contact for Reagent and Resource Sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Pau Bernadó (pau.bernado@cbs.cnrs.fr).

Method Details

Huntingtin exon1 constructs

All plasmids were prepared as previously described(Urbanek et al., 2018). Briefly, synthetic genes of wild-type huntingtin exon1 with 16 consecutive glutamines (H16) or H16 carrying the amber codon (TAG) instead of the glutamine codon, e.g. Q18 (H16Q18), were ordered from Integrated DNA Technologies (IDT). Following this strategy, 22 amber mutants were ordered: 16 within the poly-Q tract and six outside. Synthetic genes of the structural mutants (LKGG-H16, LKAA-H16, LLLF-H16 and H16-5G) and their corresponding amber codon mutants (Q18, Q20, Q21 and Q30) were ordered from GeneArt®. The synthetic genes were cleaved by NcoI and KpnI endonucleases and cloned into pIVEX 2.3d by an In-Fusion® (Clontech) reaction, giving rise to pIVEX-H16-3C-sfGFP-His_6 and mutants. The sequence of all plasmids was confirmed by sequencing by GENEWIZ®.

Preparation of glutamine ligase GLN4

Glutamine ligase GLN4 from *Saccharomyces cerevisiae* was expressed in *E. coli* BL21 (DE3) cells. To this end, a synthetic gene coding for GLN4, based on UniProt ID P13188, was ordered from IDT and subcloned into pET22 between the NdeI and XhoI restriction sites to yield the pET22-Gln4 vector. The final GLN4 construct carried a HRV 3C protease recognition site followed by GST-His_6 at its C-terminus. Cells were grown overnight at 25°C in ZYM 5052 auto-inducing medium(Studier, 2005) supplemented with 50 µg/mL kanamycin and harvested by centrifugation (6,000 xg, 20 min, 4°C). The pellet was resuspended in 20 mM Tris-HCl pH 7.5, 300 mM NaCl and 2 mM β-mercaptoethanol (GLN4 buffer A) supplemented with a cOmplete™ EDTA free protease inhibitor tablet (Roche) and lysed by sonication. The lysate was cleared by centrifugation (40,000 xg, 30 min, 4°C) and imidazole was added to a final concentration of 10 mM before loading it onto a gravity affinity column (Ni sepharose 6 FF 5 mL, GE Life Sciences) equilibrated with GLN4 buffer B (GLN4 buffer A + 10 mM imidazole). The column was washed with 50 mL GLN4 buffer B and the target protein was eluted with GLN4 buffer C (GLN4 buffer A + 250 mM imidazole). Fractions were analyzed by SDS-PAGE and fractions containing GLN4 were pooled and dialyzed against GLN4 buffer A overnight at 4°C. To further purify the protein, the dialysate was loaded on a 5 mL gravity GST column (glutathione sepharose 4B, GE Life Sciences) equilibrated in GLN4 buffer A. The resin was washed with GLN4 buffer A and GLN4 was eluted with GLN4 buffer D (GLN4 buffer A + 10 mM glutathione). Protein fractions were analyzed by SDS-PAGE and fractions containing GLN4

were pooled, dialyzed against GLN4 buffer E (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM DTT) and concentrated to 6 mg/mL with Vivaspin centrifugal concentrators (Sartorius Stedim Biotech). Aliquots were stored at -20ºC.

Lysate preparation

Lysate was based on the *Escherichia coli* strain BL21 Star (DE3)::RF1-CBD_3, a gift from Gottfried Otting (Australian National University, Canberra, Australia)(Loscha et al., 2012). *E. coli* lysates were prepared as described by Apponyi et al. and Loscha et al., but with slight modifications(Apponyi et al., 2008; Loscha et al., 2012). The cells were grown at 37°C in a fermenter with 3 L of Z-medium (41.2 mM potassium phosphate monobasic, 166 mM potassium phosphate dibasic, 10 g/L yeast extract) with added 110 mM glucose, 10 mg/L thiamine, 1 mM $MgSO_4$ and 50 µg/mL kanamycin. When the OD_{600} reached ~1, 1 mM IPTG was added to induce the expression of T7 RNA polymerase. Cells were harvested in the mid-log phase (OD_{600} ~3-4) and washed with S30 α buffer (10 mM Tris-acetate, pH 8.2, 16 mM potassium acetate, 14 mM magnesium acetate, 0.5 mM PMSF, 1 mM DTT and 7.2 mM β-mercaptoethanol) before storing the pellets at -80°C. The thawed cells were suspended in S30 α buffer (1.3 mL of buffer per gram of cells) and disrupted in a French press cell (Emulsiflex C-3, Avestin) at a constant pressure of 20,000 psi. The lysate was cleared by centrifugation twice (30 min; 30,000 xg; 4°C) before genomic chitin-tagged release factor 1 (RF1-CBD_3) was removed from the lysate by passing it over chitin resin (New England Biolabs). The lysate was then dialyzed against buffer S30 β (S30 α buffer without PMSF and β-mercaptoethanol) using SpectraPor 4 dialysis tubing (12-14 kDa MWCO, Spectrum Laboratories Inc.) for 3x 1 hour. In a next step, the lysate was dialyzed against 50% PEG 6000 in S30 buffer until the volume of the extract was reduced to half. Residual traces of PEG were removed by a short dialysis against S30 β (~15 min) before changing to buffer S30 γ (S30 β with 400 mM NaCl) overnight. All dialysis steps were performed at 4°C. Subsequently, the dialysis tubes were placed into a 250 mL Pyrex glass bottle filled with pre-warmed buffer S30 γ (42°C), and incubated for 45 minutes at 42°C in a water-bath with gentle shaking. The lysate was then dialyzed against buffer S30 β for 4 h. The extract was cleared by a final centrifugation (10 min; 30,000 xg; 4°C) and the supernatant was aliquoted, flash frozen and stored at -80°C.

Preparation and aminoacylation of suppressor $tRNA_{CUA}$

The artificial suppressor $tRNA_{CUA}$ (5' GGUCCUAUAG UGUAGUGGUU AUCACUUUCG GUUCUAAUCC GAACAACCCC AGUUCGAAUC CGGGUGGGAC CUCCA 3') was transcribed *in vitro* and purified by phenol-chloroform extraction. Prior to use, the suppressor $tRNA_{CUA}$ was refolded in 100 mM HEPES-KOH pH 7.5, 10 mM KCl at 70°C for 5 min and a final concentration of 5 mM $MgCl_2$ was added just before the reaction was placed on ice. The refolded $tRNA_{CUA}$ was then aminoacylated with [^{15}N, ^{13}C]-glutamine (CortecNet) in a standard aminoacylation reaction: 20 µM $tRNA_{CUA}$, 0.5 µM GLN4, 0.1 mM [^{15}N, ^{13}C]-Gln in 100 mM HEPES-KOH pH 7.5, 10 mM KCl,

20 mM $MgCl_2$, 1 mM DTT and 10 mM ATP(Walker and Fredrick, 2008). After incubation at 37°C for 1 hour GLN4 was removed by addition of glutathione beads and loaded suppressor $tRNA_{CUA}$ was precipitated with 300 mM sodium acetate pH 5.2 and 2.5 volumes of 96% EtOH at -80°C and stored as dry pellets at -20°C. Successful loading was confirmed by urea-PAGE (6.5% acrylamide 19:1, 8 M urea, 100 mM sodium acetate pH 5.2)(Walker and Fredrick, 2008).

Standard batch mode cell-free expression conditions

Cell-free protein expression was performed in batch mode as described by Apponyi *et al*.(Apponyi et al., 2008). Briefly, the standard batch mode reaction mixture consisted of the following components: 55 mM HEPES-KOH (pH 7.5), 1.2 mM ATP, 0.8 mM each of CTP, GTP and UTP, 1.7 mM DTT, 0.175 mg/mL *E. coli* total tRNA mixture (from strain MRE600), 0.64 mM cAMP, 27.5 mM ammonium acetate, 68 µM 1-5-formyl-5,6,7,8-tetrahydrofolic acid (folinic acid), 1 mM of each of the 20 amino acids, 80 mM creatine phosphate (CP), 250 µg/mL creatine kinase (CK), plasmid (16 µg/mL) and 22.5% (v/v) S30 extract. The concentrations of magnesium acetate (5 - 20 mM) and potassium glutamate (60 - 200 mM) were adjusted for each new batch of S30 extract. A titration of both compounds was performed to obtain the maximum yield. The reactions were carried out in a reaction volume of 50 µL dispensed in 96-well plates and were incubated at 23°C for 5 hours.

Cell-free H16Qx position screen

Plasmids of all 22 amber mutants of wild-type H16 were tested for possible position specific effects of the amber codon placement on the suppression efficiency at a final concentration of 10 µM $tRNA_{CUA}$. The time-course of H16 protein synthesis was monitored using a fluorescence read-out (sfGFP) and a plate reader/incubator (Gen5, BioTek Instruments, 485 nm (excitation), 528 nm (emission)). Assays were carried out as triplicates in a reaction volume of 50 µL dispensed in 96-well plates. The reactions were incubated at 23°C for 5 hours.

Preparation of NMR samples

Samples for NMR studies were produced at 5-15 mL scale and incubated at 23°C and 750 rpm in a thermomixer for 5 hours. Uniformly labeled NMR samples were obtained by substituting the standard amino acid mix with 3 mg/mL [^{15}N, ^{13}C]-labeled ISOGRO®(Kigawa et al., 1999) (an algal extract lacking four amino acids: Asn, Cys, Gln and Trp) and additionally supplying [^{15}N, ^{13}C]-labeled Asn, Cys, Gln and Trp (1 mM each). Furthermore, potassium glutamate was substituted by 80 mM potassium acetate to enable the labeling of glutamates. To produce site-specifically labeled samples, 10 µM of [^{15}N, ^{13}C]-Gln suppressor $tRNA_{CUA}$ were added to the standard batch mode reaction mixture (see above).

Protein purification

The cell-free reaction was thawed on ice and diluted 2-3 fold with buffer A (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 5 mM imidazole) before loading onto a Ni gravity-flow column of 1 mL bed volume (cOmplete™ His-Tag Purification Resin, Sigma Aldrich). The column was washed with buffer B (50 mM Tris-HCl pH 7.5, 1000 mM NaCl, 5 mM imidazole) and the target protein was eluted with buffer C (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 250 mM imidazole). Elution fractions were checked under UV light and fluorescent fractions were pooled and dialyzed against NMR buffer (20 mM BisTris-HCl pH 6.5, 150 mM NaCl) at 4°C using SpectraPor 1 MWCO 6-8 kDa dialysis tubing (Spectrum Labs). Dialyzed protein was then concentrated with 10 kDa MWCO Vivaspin centrifugal concentrators (3500 x g, 4°C) (Sartorius). Protein concentrations were determined by means of fluorescence using an sfGFP calibration curve. Final NMR sample concentrations ranged from 4 to 11 µM. Protein integrity was analyzed by SDS-PAGE.

NMR experiments and data analysis

All NMR samples contained final concentrations of 10% D_2O and 0.5 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Experiments were performed at 293 K on a Bruker Avance III spectrometer equipped with a cryogenic triple resonance probe and Z gradient coil, operating at a ^{1}H frequency of 700 MHz or 800 MHz. ^{15}N-HSQC and ^{13}C-HSQC were acquired for each sample in order to determine amide ($^{1}H_N$ and ^{15}N) and aliphatic ($^{1}H_{aliphatic}$ and $^{13}C_{aliphatic}$) chemical shifts, respectively. Spectra acquisition parameters were set up depending on the sample concentration and the magnet strength. ^{15}N-HSQC spectra were acquired for 8 to 20 hours using 256-512 scans, 88-128 increments and a spectral width of 21 ppm in the indirect dimension. ^{13}C-HSQC spectra were acquired for 10 to 24 hours using 256-512 scans, 96-128 increments and a spectral width of 60 ppm in the indirect dimension. All spectra were processed with TopSpin v3.5 (Bruker Biospin) and analyzed using CCPN-Analysis software(Vranken et al., 2005). Chemical shifts were referenced with respect to the H_2O signal relative to DSS using the $^{1}H/X$ frequency ratio of the zero point according to Markley *et al*.(Markley et al., 1998).

Random coil chemical shifts were predicted using POTENCI, a pH, temperature and neighbor corrected IDP library (https://st-protein02.chem.au.dk/potenci/)(Nielsen and Mulder, 2018). Secondary chemical shifts (SCS) were obtained by subtracting the predicted value from the experimental one ($SCS=\delta_{exp}-\delta_{pred}$). For better reliability of the results regarding possible referencing errors, we used the combined C_α and C_β secondary chemical shifts ($SCS(C_\alpha)-SCS(C_\beta)$). In addition, secondary structure propensities (SSPs) were calculated using the script developed by Marsh *et al*.(Marsh et al., 2006) and the refDB database(Zhang et al., 2003).

Model building and experimental ensemble optimization

Ensemble models for the two families capturing the conformational influences of the flanking regions, N→C and N←C, were constructed with the algorithm described in reference(Estaña et al., 2019),

which uses a curated database of three-residue fragments extracted from high-resolution protein structures. The averaged Cα and Cβ CSs for the 34 ensembles, 17 for each family, were computed with SPARTA+(Shen and Bax, 2010) and used to refine a final ensemble in agreement with the experimental data. Concretely, the optimized ensemble model of H16 was built by reweighting the populations of the pre-computed ensembles, minimizing the error with respect to the experimental Cα and Cβ CSs. In order to capture the influence of the flanking regions, glutamines within the tract were divided into two groups: those influenced by N17 and those influenced by the poly-P tract, whose chemical shifts were fitted with the N→C and N←C ensembles, respectively. The limit between both families was systematically explored by computing the agreement between the experimental and optimized CSs through a χ_i^2 value. An optimal description of the complete CS profile was obtained when Q28 was chosen as the last residue structurally connected with N17. Finally, an ensemble of 50,000 structures was built using the optimized weights and it was used to analyze the residue-specific Ramachandran propensities and the secondary structure population using SS-map(Iglesias et al., 2013).

Quantification and statistical analysis

In the cell-free H16Qx position screen data is represented as mean of triplicates ± SD.

Notch plots were generated using the library "ggplot2" in R.

References

Ansaloni, A., Wang, Z.-M., Jeong, J.S., Ruggeri, F.S., Dietler, G., and Lashuel, H.A. (2014). One-pot semisynthesis of exon 1 of the Huntingtin protein: new tools for elucidating the role of posttranslational modifications in the pathogenesis of Huntington's disease. Angew. Chem. Int. Ed. Engl. *53*, 1928–1933.

Apponyi, M.A., Ozawa, K., Dixon, N.E., and Otting, G. (2008). Cell-free protein synthesis for analysis by NMR spectroscopy. In Structural Proteomics. Methods in Molecular Biology™, B. Kobe, M. Guss, and T. Huber, eds. (Humana Press), pp. 257–268.

Atwal, R.S., Desmond, C.R., Caron, N., Maiuri, T., Xia, J., Sipione, S., and Truant, R. (2011). Kinase inhibitors modulate huntingtin cell localization and toxicity. Nat. Chem. Biol. *7*, 453–460.

Baias, M., Smith, P.E.S., Shen, K., Joachimiak, L.A., Żerko, S., Koźmiński, W., Frydman, J., and Frydman, L. (2017). Structure and dynamics of the huntingtin exon-1 N-terminus: A solution NMR perspective. J. Am. Chem. Soc. *139*, 1168–1176.

Baxter, N.J., and Williamson, M.P. (1997). Temperature dependence of 1H chemical shifts in proteins. J. Biomol. NMR *9*, 359–369.

Bennett, M.J., Huey-Tubman, K.E., Herr, A.B., West, A.P., Ross, S.A., and Bjorkman, P.J. (2002). A linear lattice model for polyglutamine in CAG-expansion diseases. Proc. Natl. Acad. Sci. U. S. A. *99*, 11634–11639.

Bhattacharyya, A., Thakur, A.K., Chellgren, V.M., Thiagarajan, G., Williams, A.D., Chellgren, B.W., Creamer, T.P., and Wetzel, R. (2006). Oligoproline effects on polyglutamine conformation and aggregation. J. Mol. Biol. *355*, 524–535.

Bravo-Arredondo, J.M., Kegulian, N.C., Schmidt, T., Pandey, N.K., Situ, A.J., Ulmer, T.S., and Langen, R. (2018). The folding equilibrium of huntingtin exon 1 monomer depends on its polyglutamine tract. J. Biol. Chem. *293*, 19613–19623.

Ceccon, A., Schmidt, T., Tugarinov, V., Kotler, S.A., Schwieters, C.D., and Clore, G.M. (2018). Interaction of huntingtin exon-1 peptides with lipid-based micellar nanoparticles probed by solution NMR and Q-band pulsed EPR. J. Am. Chem. Soc. *140*, 6199–6202.

Chiki, A., DeGuire, S.M., Ruggeri, F.S., Sanfelice, D., Ansaloni, A., Wang, Z.-M., Cendrowska, U., Burai, R., Vieweg, S., Pastore, A., et al. (2017). Mutant exon1 huntingtin aggregation is regulated by T3 phosphorylation-induced structural changes and crosstalk between T3 phosphorylation and acetylation at K6. Angew. Chem. Int. Ed. Engl. *56*, 5202–5207.

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.

Darling, A.L., and Uversky, V.N. (2017). Intrinsic disorder in proteins with pathogenic repeat expansions. Molecules *22*, 2027.

Darnell, G., Orgel, J.P.R.O., Pahl, R., and Meredith, S.C. (2007). Flanking polyproline sequences

inhibit beta-sheet structure in polyglutamine segments by inducing PPII-like helix structure. J. Mol. Biol. *374*, 688–704.

Dasgupta, S., and Bell, J.A. (1993). Design of helix ends. Amino acid preferences, hydrogen bonding and electrostatic interactions. Int. J. Pept. Protein Res. *41*, 499–511.

DeGuire, S.M., Ruggeri, F.S., Fares, M.-B., Chiki, A., Cendrowska, U., Dietler, G., and Lashuel, H.A. (2018). N-terminal Huntingtin (Htt) phosphorylation is a molecular switch regulating Htt aggregation, helical conformation, internalization, and nuclear targeting. J. Biol. Chem. *293*, 18540–18558.

Dehay, B., and Bertolotti, A. (2006). Critical role of the proline-rich region in huntingtin for aggregation and cytotoxicity in yeast. J. Biol. Chem. *281*, 35608–35615.

DiFiglia, M., Sapp, E., Chase, K.O., Davies, S.W., Bates, G.P., Vonsattel, J.P., and Aronin, N. (1997). Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. Science *277*, 1990–1993.

Eftekharzadeh, B., Piai, A., Chiesa, G., Mungianu, D., García, J., Pierattelli, R., Felli, I.C., and Salvatella, X. (2016). Sequence context influences the structure and aggregation behavior of a polyQ tract. Biophys. J. *110*, 2361–2366.

Ehrnhoefer, D.E., Sutton, L., and Hayden, M.R. (2011). Small changes, big impact: posttranslational modifications and function of huntingtin in Huntington disease. Neuroscientist *17*, 475–492.

Ellman, J.A., Volkman, B.F., Mendel, D., Schultz, P.G., and Wemmer, D.E. (1992). Site-specific isotopic labeling of proteins for NMR studies. J. Am. Chem. Soc. *114*, 7959–7961.

Escobedo, A., Topal, B., Kunze, M.B.A., Aranda, J., Chiesa, G., Mungianu, D., Bernardo-Seisdedos, G., Eftekharzadeh, B., Gairí, M., Pierattelli, R., et al. (2019). Side chain to main chain hydrogen bonds stabilize a polyglutamine helix in a transcription factor. Nat. Commun. *10*, 2034.

Estaña, A., Sibille, N., Delaforge, E., Vaisset, M., Cortés, J., and Bernadó, P. (2019). Realistic ensemble models of intrinsically disordered proteins using a structure-encoding coil database. Structure *27*, 381–391.

Feng, X., Luo, S., and Lu, B. (2018). Conformation polymorphism of polyglutamine proteins. Trends Biochem. Sci. *43*, 424–435.

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.

Fodale, V., Kegulian, N.C., Verani, M., Cariulo, C., Azzollini, L., Petricca, L., Daldin, M., Boggio, R., Padova, A., Kuhn, R., et al. (2014). Polyglutamine- and temperature-dependent conformational rigidity in mutant huntingtin revealed by immunoassays and circular dichroism spectroscopy. PLoS One *9*, e112262.

Gao, J., Bosco, D.A., Powers, E.T., and Kelly, J.W. (2009). Localized thermodynamic coupling between hydrogen bonding and microenvironment polarity substantially stabilizes proteins. Nat. Struct. Mol. Biol. *16*, 684–690.

De Genst, E., Chirgadze, D.Y., Klein, F.A.C., Butler, D.C., Matak-Vinković, D., Trottier, Y., Huston,

J.S., Messer, A., and Dobson, C.M. (2015). Structure of a single-chain Fv bound to the 17 N-terminal residues of huntingtin provides insights into pathogenic amyloid formation and suppression. J. Mol. Biol. *427*, 2166–2178.

Hoop, C.L., Lin, H.-K., Kar, K., Magyarfalvi, G., Lamley, J.M., Boatz, J.C., Mandal, A., Lewandowski, J.R., Wetzel, R., and van der Wel, P.C.A. (2016). Huntingtin exon 1 fibrils feature an interdigitated β-hairpin-based polyglutamine core. Proc. Natl. Acad. Sci. U. S. A. *113*, 1546–1551.

Hosp, F., Gutiérrez-Ángel, S., Schaefer, M.H., Cox, J., Meissner, F., Hipp, M.S., Hartl, F.-U., Klein, R., Dudanova, I., and Mann, M. (2017). Spatiotemporal proteomic profiling of huntington's disease inclusions reveals widespread loss of protein function. Cell Rep. *21*, 2291–2303.

Iglesias, J., Sanchez-Martínez, M., and Crehuet, R. (2013). SS-map: Visualizing cooperative secondary structure elements in protein ensembles. Intrinsically Disord. Proteins *1*, e25323.

Isas, J.M., Langen, R., and Siemer, A.B. (2015). Solid-state nuclear magnetic resonance on the static and dynamic domains of huntingtin exon-1 fibrils. Biochemistry *54*, 3942–3949.

Jayaraman, M., Kodali, R., Sahoo, B., Thakur, A.K., Mayasundari, A., Mishra, R., Peterson, C.B., and Wetzel, R. (2012). Slow amyloid nucleation via α-helix-rich oligomeric intermediates in short polyglutamine-containing huntingtin fragments. J. Mol. Biol. *415*, 881–899.

Jorda, J., and Kajava, A. V (2010). Protein homorepeats. Adv. Protein Chem. Struct. Biol. *79*, 59–88.

Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T., and Yokoyama, S. (1999). Cell-free production and stable-isotope labeling of milligram quantities of proteins. FEBS Lett. *442*, 15–19.

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.

Klein, F.A.C., Zeder-Lutz, G., Cousido-Siah, A., Mitschler, A., Katz, A., Eberling, P., Mandel, J.-L., Podjarny, A., and Trottier, Y. (2013). Linear and extended: a common polyglutamine conformation recognized by the three antibodies MW1, 1C2 and 3B5H10. Hum. Mol. Genet. *22*, 4215–4223.

Kotler, S.A., Tugarinov, V., Schmidt, T., Ceccon, A., Libich, D.S., Ghirlando, R., Schwieters, C.D., and Clore, G.M. (2019). Probing initial transient oligomerization events facilitating Huntingtin fibril nucleation at atomic resolution by relaxation-based NMR. Proc. Natl. Acad. Sci. U. S. A. *116*, 3562–3571.

Krivov, G.G., Shapovalov, M. V, and Dunbrack, R.L. (2009). Improved prediction of protein side-chain conformations with SCWRL4. Proteins *77*, 778–795.

Li, P., Huey-Tubman, K.E., Gao, T., Li, X., West, A.P., Bennett, M.J., and Bjorkman, P.J. (2007). The structure of a polyQ-anti-polyQ complex reveals binding according to a linear lattice model. Nat. Struct. Mol. Biol. *14*, 381–387.

Lobanov, M.Y., and Galzitskaya, O. V (2012). Occurrence of disordered patterns and homorepeats in eukaryotic and bacterial proteomes. Mol. Biosyst. *8*, 327–337.

Loscha, K. V, Herlt, A.J., Qi, R., Huber, T., Ozawa, K., and Otting, G. (2012). Multiple-site labeling of proteins with unnatural amino acids. Angew. Chem. Int. Ed. Engl. *51*, 2243–2246.

MacArthur, M.W., and Thornton, J.M. (1991). Influence of proline residues on protein conformation. J. Mol. Biol. *218*, 397–412.

Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetherington, C., Lawton, M., Trottier, Y., Lehrach, H., Davies, S.W., et al. (1996). Exon I of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. Cell *87*, 493–506.

Markley, J.L., Bax, A., Arata, Y., Hilbers, C.W., Kaptein, R., Sykes, B.D., Wright, P.E., and Wüthrich, K. (1998). Recommendations for the presentation of NMR structures of proteins and nucleic acids. J. Mol. Biol. *280*, 933–952.

Marsh, J.A., Singh, V.K., Jia, Z., and Forman-Kay, J.D. (2006). Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. Protein Sci. *15*, 2795–2804.

Michalek, M., Salnikov, E.S., and Bechinger, B. (2013). Structure and topology of the huntingtin 1-17 membrane anchor by a combined solution and solid-state NMR approach. Biophys. J. *105*, 699–710.

Mier, P., Alanis-Lobato, G., and Andrade-Navarro, M.A. (2017). Context characterization of amino acid homorepeats using evolution, position, and order. Proteins *85*, 709–719.

Miller, J., Arrasate, M., Brooks, E., Libeu, C.P., Legleiter, J., Hatters, D., Curtis, J., Cheung, K., Krishnan, P., Mitra, S., et al. (2011). Identifying polyglutamine protein species in situ that best predict neurodegeneration. Nat. Chem. Biol. *7*, 925–934.

Milles, S., Salvi, N., Blackledge, M., and Jensen, M.R. (2018). Characterization of intrinsically disordered proteins and their dynamic complexes: From in vitro to cell-like environments. Prog. Nucl. Magn. Reson. Spectrosc. *109*, 79–100.

Mishra, R., Hoop, C.L., Kodali, R., Sahoo, B., van der Wel, P.C.A., and Wetzel, R. (2012). Serine phosphorylation suppresses huntingtin amyloid accumulation by altering protein aggregation properties. J. Mol. Biol. *424*, 1–14.

Newcombe, E.A., Ruff, K.M., Sethi, A., Ormsby, A.R., Ramdzan, Y.M., Fox, A., Purcell, A.W., Gooley, P.R., Pappu, R. V, and Hatters, D.M. (2018). Tadpole-like conformations of huntingtin exon 1 are characterized by conformational heterogeneity that persists regardless of polyglutamine length. J. Mol. Biol. *430*, 1442–1458.

Newell, N.E. (2015). Mapping side chain interactions at protein helix termini. BMC Bioinformatics *16*, 231.

Nielsen, J.T., and Mulder, F.A.A. (2018). POTENCI: prediction of temperature, neighbor and pH-corrected chemical shifts for intrinsically disordered proteins. J. Biomol. NMR *70*, 141–165.

Nucifora, L.G., Burke, K.A., Feng, X., Arbez, N., Zhu, S., Miller, J., Yang, G., Ratovitski, T., Delannoy, M., Muchowski, P.J., et al. (2012). Identification of novel potentially toxic oligomers formed in vitro from mammalian-derived expanded huntingtin exon-1 protein. J. Biol. Chem. *287*, 16017–16028.

Orr, H.T. (2001). Beyond the Qs in the polyglutamine diseases. Genes Dev. *15*, 925–932.

Owens, G.E., New, D.M., West, A.P., and Bjorkman, P.J. (2015). Anti-polyQ antibodies recognize a short polyQ stretch in both normal and mutant huntingtin exon 1. J. Mol. Biol. *427*, 2507–2519.

Peters-Libeu, C., Miller, J., Rutenber, E., Newhouse, Y., Krishnan, P., Cheung, K., Hatters, D., Brooks, E., Widjaja, K., Tran, T., et al. (2012). Disease-associated polyglutamine stretches in monomeric huntingtin adopt a compact structure. J. Mol. Biol. *421*, 587–600.

Peuker, S., Andersson, H., Gustavsson, E., Maiti, K.S., Kania, R., Karim, A., Niebling, S., Pedersen, A., Erdelyi, M., and Westenhoff, S. (2016). Efficient isotope editing of proteins for site-directed vibrational spectroscopy. J. Am. Chem. Soc. *138*, 2312–2318.

Ramazzotti, M., Monsellier, E., Kamoun, C., Degl'Innocenti, D., and Melki, R. (2012). Polyglutamine repeats are associated to specific sequence biases that are conserved among eukaryotes. PLoS One *7*, e30824.

Richardson, J.S., and Richardson, D.C. (1988). Amino acid preferences for specific locations at the ends of alpha helices. Science *240*, 1648–1652.

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.

Scherzinger, E., Lurz, R., Turmaine, M., Mangiarini, L., Hollenbach, B., Hasenbank, R., Bates, G.P., Davies, S.W., Lehrach, H., and Wanker, E.E. (1997). Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. Cell *90*, 549–558.

Seale, J.W., Srinivasan, R., and Rose, G.D. (1994). Sequence determinants of the capping box, a stabilizing motif at the N-termini of alpha-helices. Protein Sci. *3*, 1741–1745.

Shen, Y., and Bax, A. (2010). SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. J. Biomol. NMR *48*, 13–22.

Shen, K., Calamini, B., Fauerbach, J.A., Ma, B., Shahmoradian, S.H., Serrano Lachapel, I.L., Chiu, W., Lo, D.C., and Frydman, J. (2016). Control of the structural landscape and neuronal proteotoxicity of mutant huntingtin by domains flanking the polyQ tract. Elife *5*, 1–29.

Steffan, J.S., Agrawal, N., Pallos, J., Rockabrand, E., Trotman, L.C., Slepko, N., Illes, K., Lukacsovich, T., Zhu, Y.-Z., Cattaneo, E., et al. (2004). SUMO modification of huntingtin and huntington's disease pathology. Science *304*, 100–104.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. *41*, 207–234.

Tam, S., Spiess, C., Auyeung, W., Joachimiak, L., Chen, B., Poirier, M.A., and Frydman, J. (2009). The chaperonin TRiC blocks a huntingtin sequence element that promotes the conformational switch to aggregation. Nat. Struct. Mol. Biol. *16*, 1279–1285.

Tartari, M., Gissi, C., Lo Sardo, V., Zuccato, C., Picardi, E., Pesole, G., and Cattaneo, E. (2008). Phylogenetic comparison of huntingtin homologues reveals the appearance of a primitive polyQ in sea

urchin. Mol. Biol. Evol. *25*, 330–338.

Thakur, A.K., Jayaraman, M., Mishra, R., Thakur, M., Chellgren, V.M., Byeon, I.-J.L., Anjum, D.H., Kodali, R., Creamer, T.P., Conway, J.F., et al. (2009). Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism. Nat. Struct. Mol. Biol. *16*, 380–389.

Theillet, F.-X., Kalmar, L., Tompa, P., Han, K.-H., Selenko, P., Dunker, A.K., Daughdrill, G.W., and Uversky, V.N. (2013). The alphabet of intrinsic disorder: I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins. Intrinsically Disord. Proteins *1*, e24360.

Urbanek, A., Morató, A., Allemand, F., Delaforge, E., Fournet, A., Popovic, M., Delbecq, S., Sibille, N., and Bernadó, P. (2018). A general strategy to access structural information at atomic resolution in polyglutamine homorepeats. Angew. Chem. Int. Ed. Engl. *57*, 3598–3601.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins *59*, 687–696.

Walker, F.O. (2007). Huntington's disease. Lancet (London, England) *369*, 218–228.

Walker, S.E., and Fredrick, K. (2008). Preparation and evaluation of acylated tRNAs. Methods *44*, 81–86.

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.

Wang, L., Xie, J., and Schultz, P.G. (2006). Expanding the genetic code. Annu. Rev. Biophys. Biomol. Struct. *35*, 225–249.

Wanker, E.E. (2000). Protein aggregation and pathogenesis of Huntington's disease: mechanisms and correlations. Biol. Chem. *381*, 937–942.

Warner, J.B., Ruff, K.M., Tan, P.S., Lemke, E.A., Pappu, R. V, and Lashuel, H.A. (2017). Monomeric huntingtin exon 1 has similar overall structural features for wild-type and pathological polyglutamine lengths. J. Am. Chem. Soc. *139*, 14456–14469.

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.

Zhang, H., Neal, S., and Wishart, D.S. (2003). RefDB: a database of uniformly referenced protein chemical shifts. J. Biomol. NMR *25*, 173–195.

Zoghbi, H.Y., and Orr, H.T. (2000). Glutamine repeats and neurodegeneration. Annu. Rev. Neurosci. *23*, 217–247.

Figure 1.

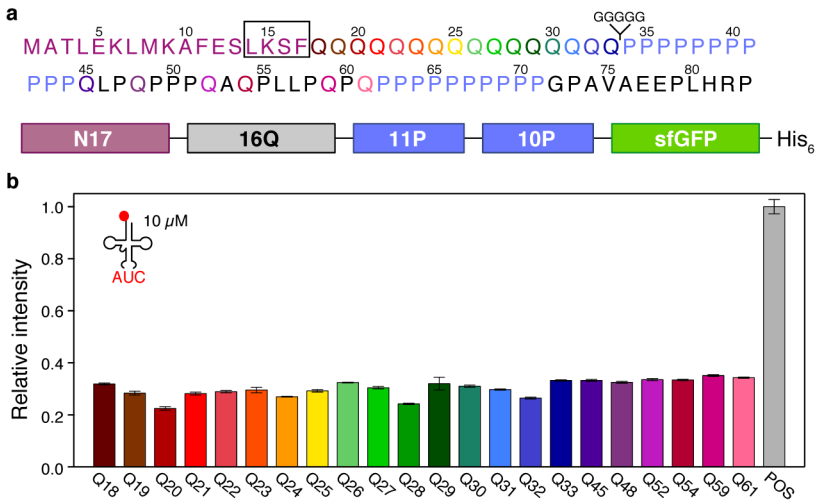


Figure 1. Glutamine SSIL scanning of H16. (**a**) Sequence of H16 and scheme of the sfGFP-fused construct used in this study. The color code identifies the individual glutamines throughout the study. The box encompassing residues ^{14}LKSF^{17} identifies the residues mutated to probe the structural connection between N17 and the poly-Q tract. The position of the insertion of glycines between the poly-Q and the PPR to structurally disconnect both regions is also displayed. (**b**) A scan probing the suppression efficiency using 10 µM loaded $tRNA_{CUA}$ showed no strong position-specific effects. The experiments were repeated three times.

Figure 2

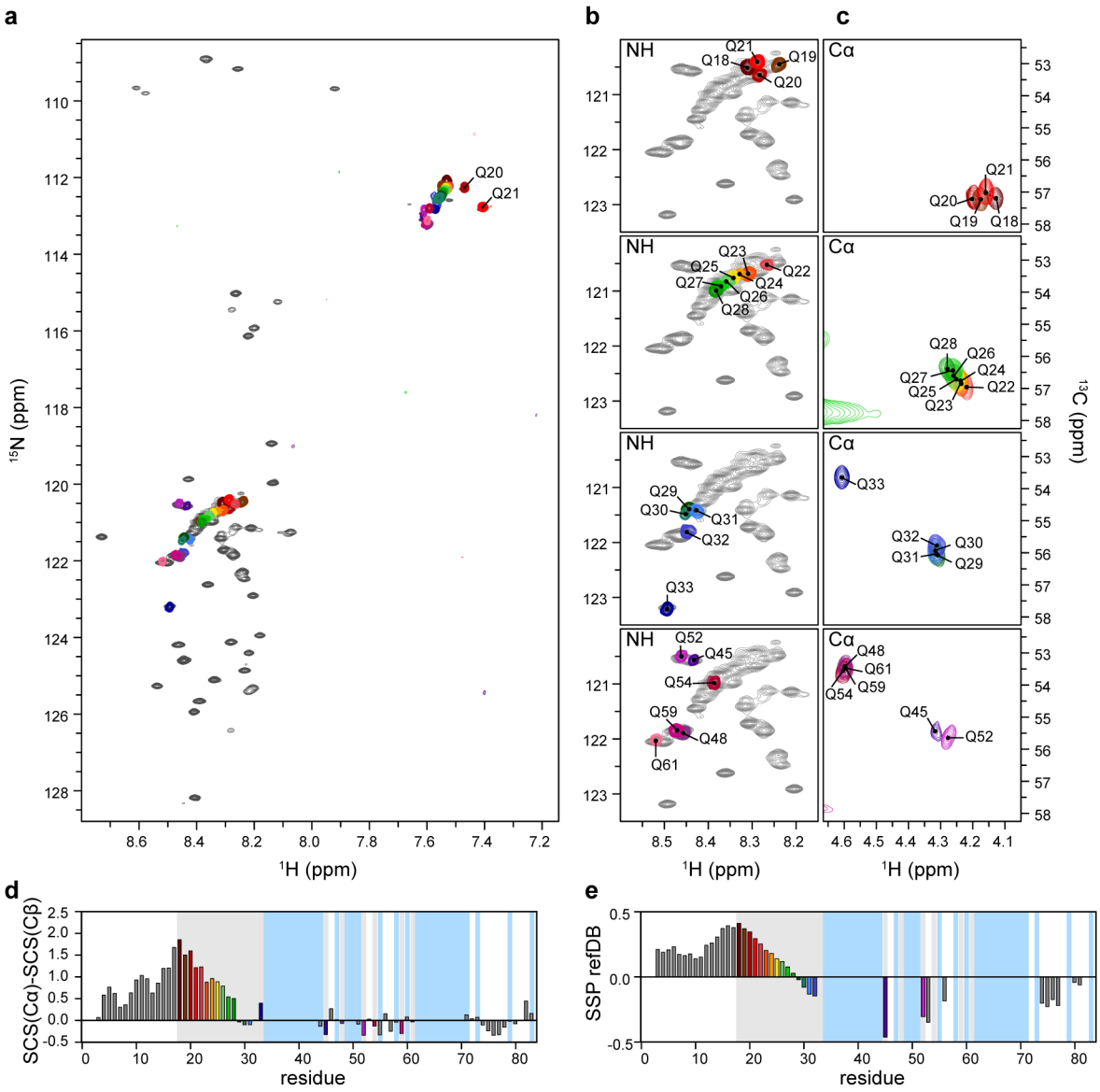


Figure 2. NMR analysis of H16. (**a**) Overlay of fully labeled H16 (grey) with individually colored SSIL ^{15}N-HSQC spectra. (**b**) Zoomed ^{15}N-HSQC overlay showing the poly-Q region with different glutamine clusters (Q18-Q21; Q22-Q28; Q29-Q33; and PRR glutamines). (**c**) Zoomed ^{13}C-HSQC overlay showing the poly-Q region with the same glutamine clusters as in (**b**). (**d**) Secondary chemical shift analysis of H16 using experimental Cα and Cβ chemical shifts and a neighbor-corrected random-coil library(Nielsen and Mulder, 2018) and (**e**) secondary structure propensity plot(Marsh et al., 2006; Zhang et al., 2003). The positions of glutamine and proline residues in the primary sequence are highlighted in grey and blue, respectively. Prolines and residues followed by prolines were not considered in the SSP refDB analysis.

Figure 3

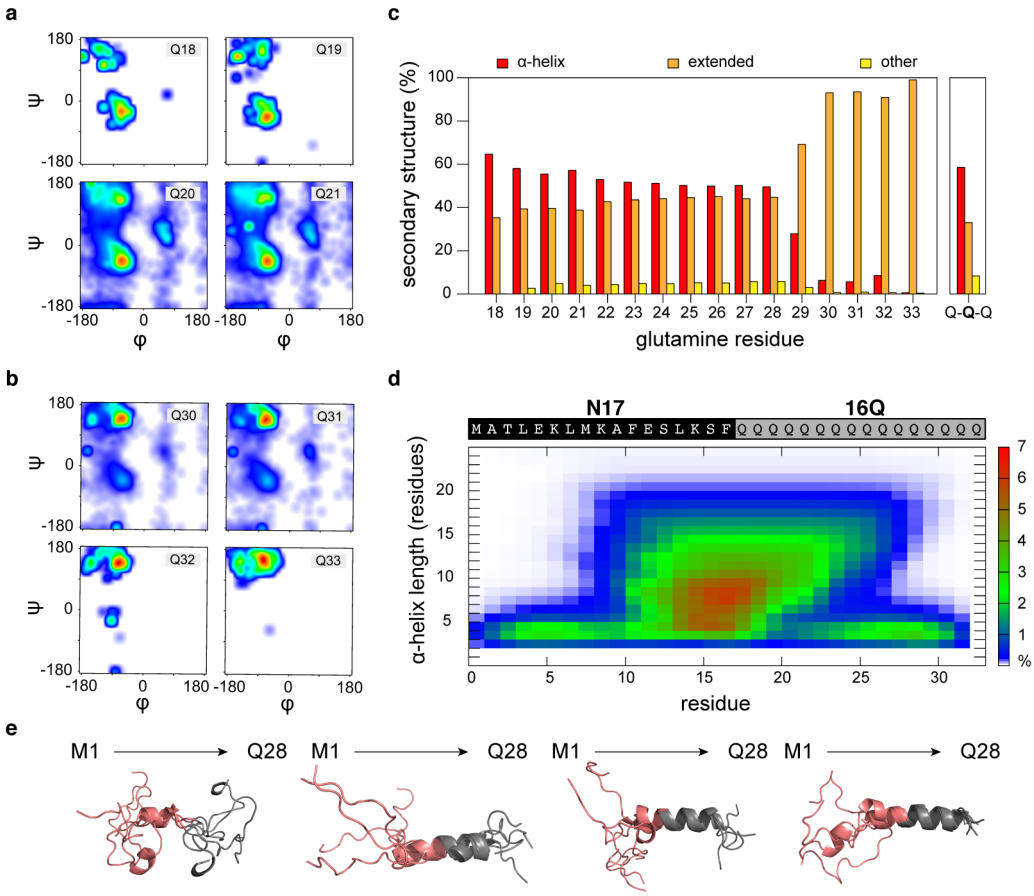


Figure 3. NMR-derived ensemble model of H16. Residue-specific Ramachandran plots for (**a**) Q18-Q21 and (**b**) Q30-Q33 obtained from the optimized ensemble. See also Figures S1 and S2. (**c**) Population of α-helix, extended and *other* conformations calculated from the Ramachandran plot for all glutamines in H16 (see Figure S3a). The side panel Q-**Q**-Q shows these populations for glutamine tri-peptides present in a coil database(Estaña et al., 2019). (**d**) Secondary structure map (SS-map) displaying the length and the residues encompassing the α-helices found in the N-terminal region of the optimized ensemble model of H16. The color code (right) indicates the population of the α-helices. (**e**) Representative conformations of the four ensembles used to describe the NMR CSs measured for H16. Only the region from M1 to Q28, optimized with the N→C ensembles, is displayed. The SS-maps for these ensembles are displayed in Figure S3b.

Figure 4

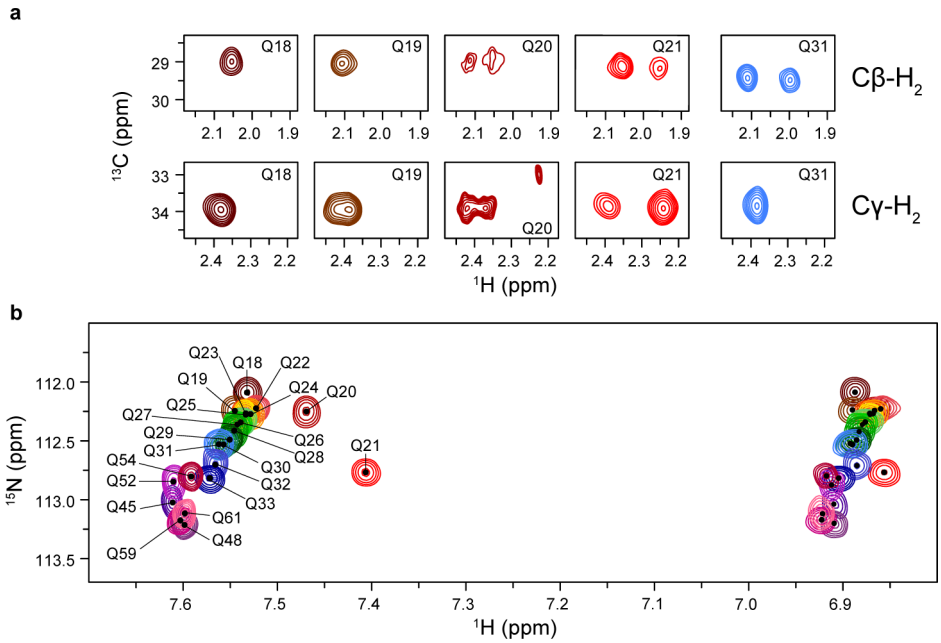


Figure 4. Side chain NMR scanning. (**a**) Cβ-H_2 and Cγ-H_2 regions of the ^{13}C-NMR spectra of glutamines Q18, Q19, Q20, Q21, and Q31. The spectra of Q31 display the standard behavior of disordered glutamines with a doublet and singlet for Cβ-H_2 and Cγ-H_2, respectively. (**b**) Zoom on the Nε2-Hε2 side chain region of the ^{15}N-HSQC spectra measured for all glutamines in H16. Q20 and Q21 do not follow the trend displayed by the other glutamines due to their implication in the formation of hydrogen bonds. See also Figure S4.

Figure 5

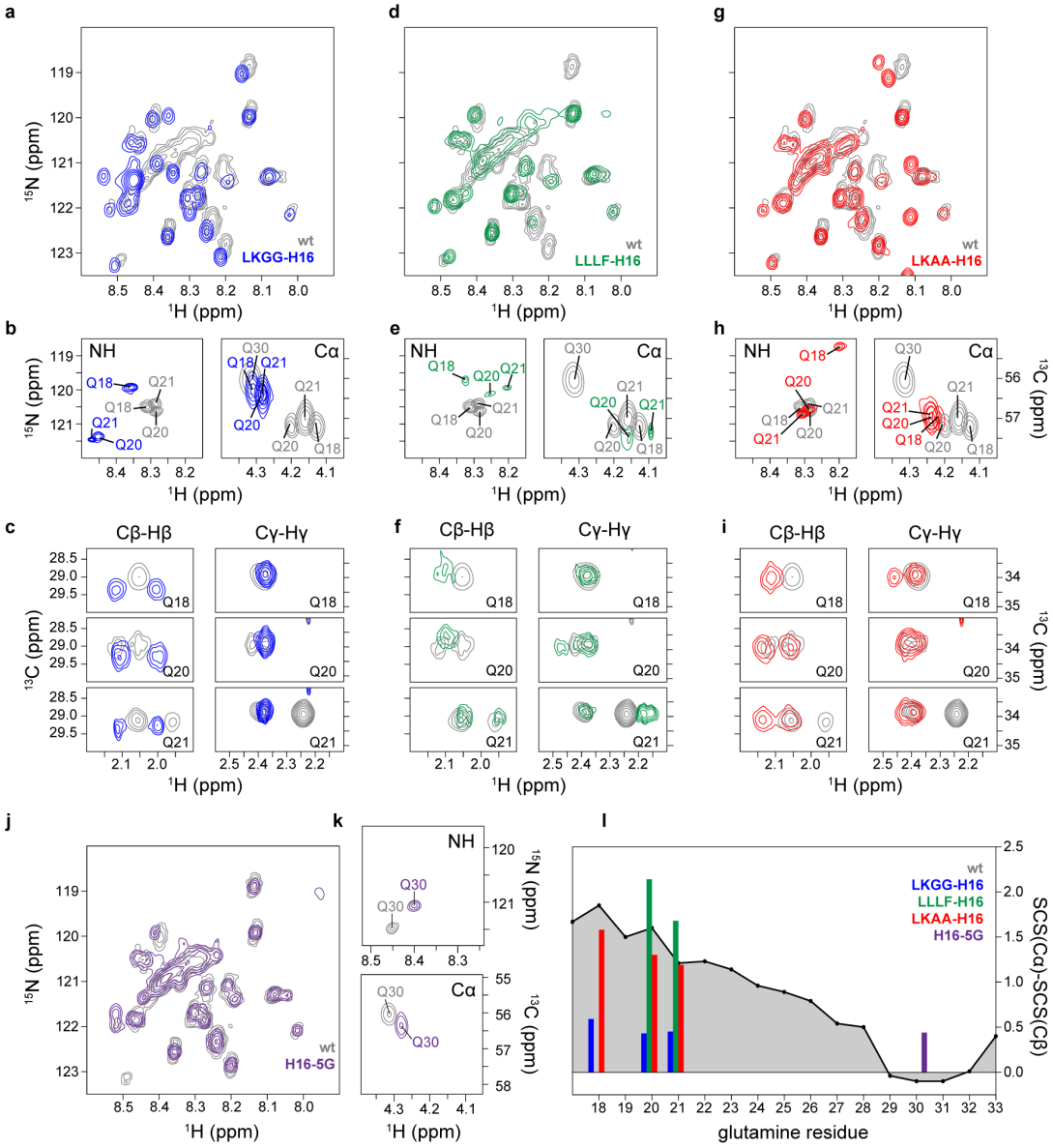


Figure 5. SSIL analyses of the structural effects of the flanking regions on the poly-Q tract of H16. Overlay of the glutamine region of the ^{15}N-HSQC spectra of fully labeled wt H16 (grey) with the N17 mutants LKGG-H16 (**a, blue**), LLLF-H16 (**d, green**), and LKAA-H16 (**g, red**). The same color-code was used throughout the figure. Zoomed overlays of the ^{15}N- and ^{13}C-HSQCs for site-specifically labeled Q18, Q20 and Q21 of wt H16 (grey) with LKGG-H16 (**b**), LLLF-H16 (**e**) and LKAA-H16 (**h**). Cβ-H_2 and Cγ-H_2 NMR peaks of the Q18, Q20 and Q21 glutamine side chains of the three N17 mutants LKGG-H16 (**c**), LLLF-H16 (**f**) and LKAA-H16 (**i**) compared with those obtained for the wt (grey). Zoomed ^{15}N- and ^{13}C-HSQC spectra for the H16-5G mutant, which probes the structural perturbation exerted by the poly-P tract, displaying the N-H glutamine region (**j, purple**) overlaid with the wt (grey), and the SSIL spectra measured for Q30 (**k**). (**l**) Histogram of the SCS analyses for the different SSIL samples of the structural mutants measured: Q18, Q20 and Q21 for the LKGG-H16, LLLF-H16 and LKAA-H16 mutants, and Q30 for the H16-5G mutant. The SSIL-derived SCS values are compared to those obtained for the wt (grey area). Note that no SCS value was derived for Q18 in the LLLF-H16 mutant due to the absence of the Cα-Hα peak in the ^{13}C-HSQC. See also Figure S5.

Figure 6

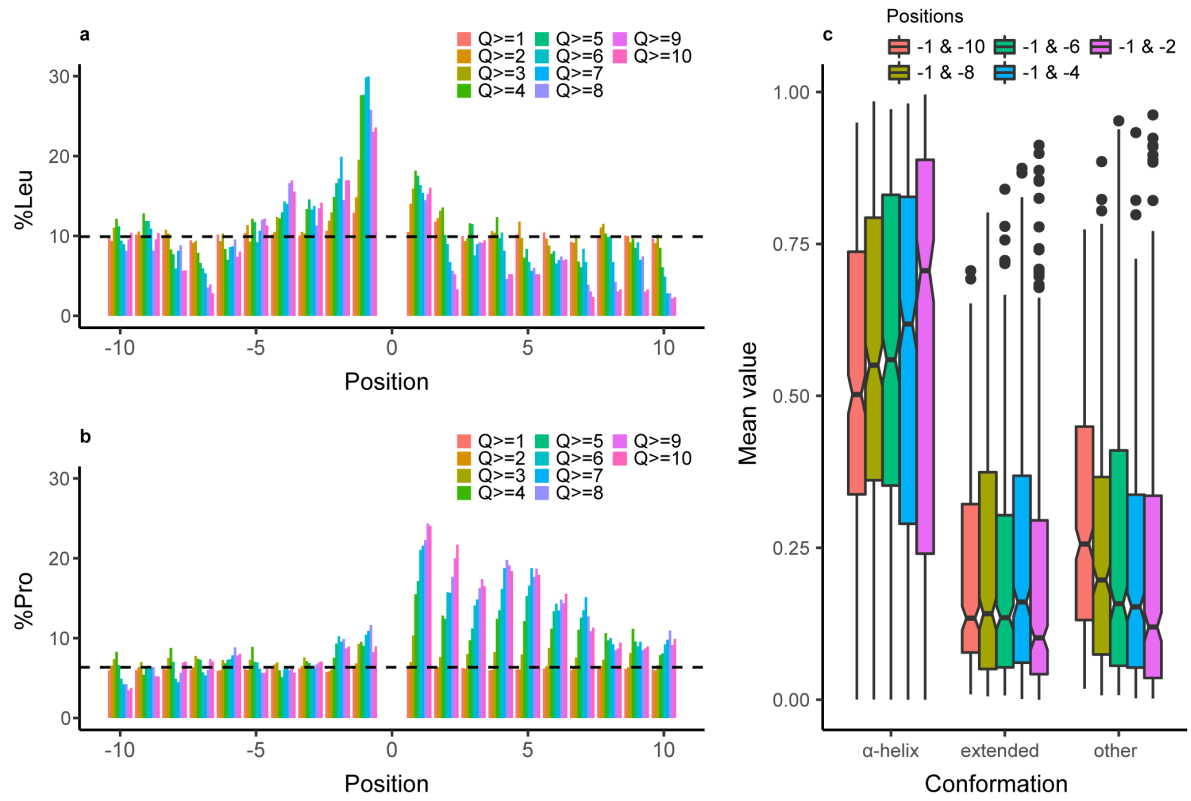


Figure 6. Primary and secondary structure context of human glutamine-rich proteins. (**a**) Leucine and (**b**) proline abundance per position in region -10 to +10 of poly-Q regions in the context of pure glutamine stretches of variable length. Horizontal dashed lines correspond to the percentage of leucines (9.9%) and prolines (6.3%) found in the human proteome. An analysis of all 20 natural amino acids is displayed in Figure S6. (**c**) Secondary structural prediction (α-helix, extended and others) per two-residue block in the N-terminal flanking regions of glutamine-rich fragments.