Transient interdomain interactions modulate the monomeric structural ensemble and oligomerization landscape of Huntingtin Exon 1

Priyesh Mohanty[1*], Tien Minh Phan[1] and Jeetain Mittal[1,2,3*]

[1]Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843, USA

[2]Department of Chemistry, Texas A&M University, College Station, TX 77843, USA

[3]Interdisciplinary Graduate Program in Genetics and Genomics, Texas A&M University, College Station, TX 77843, USA

*Correspondence: Priyesh Mohanty - priyeshm@tamu.edu, Jeetain Mittal - jeetain@tamu.edu

Abstract

Polyglutamine expansion (≥ 36 residues) within the N-terminal exon-1 of Huntingtin (Httex1) leads to Huntington's disease, a neurogenerative condition marked by the presence of intranuclear Htt inclusions. Notably, the polyglutamine tract in Httex1 is flanked by an N-terminal coiled-coil domain - N17 (17 amino acids), which undergoes self-association to promote the formation of soluble Httex1 oligomers and brings the aggregation-prone polyQ tracts in close spatial proximity. However, the mechanisms underlying the subsequent conversion of soluble oligomers into insoluble β-rich aggregates with increasing polyQ length, remain unclear. Current knowledge suggests that expansion of the polyQ tract increases its helicity, and this favors its oligomerization and aggregation. In addition, studies utilizing conformation-specific antibodies and a stable coiled-coil heterotetrametric system fused to polyQ indicate that domain "cross-talk" (i.e., interdomain interactions) may be necessary to efficiently promote the emergence of toxic conformations (in monomers and oligomers) and fibrillar aggregation. Here, we performed extensive atomistic molecular dynamics (MD) simulations (aggregate time ~ 0.7 ms) of N17-polyQ fragments to uncover the interplay between structural transformation and domain "cross-talk" on the monomeric structural ensemble and oligomerization landscape of Httex1. Our simulation ensembles of N17-polyQ monomers validated against ^{13}C NMR chemical shifts indicated that in addition to elevated α-helicity, polyQ expansion also favors transient, interdomain (N17-polyQ) interactions which result in the emergence of β-conformations. Further, interdomain interactions decreased the overall stability of N17-mediated dimers by counteracting the stabilizing effect of increased α-helicity and promoted a heterogenous oligomerization landscape on the sub-microsecond timescale. Overall, our study uncovers the significance of domain "cross-talk" in modulating the monomeric conformational ensemble and oligomerization landscape of Httex1 to favor the formation of amyloid aggregates.

Introduction

There are nine polyglutamine (polyQ)-associated neurodegenerative disorders which are known to arise as a result of CAG trinucleotide repeat expansion within the polyglutamine-encoding tract of a specific gene (Ross 2002; Shao and Diamond 2007; Adegbuyiro et al. 2017). Huntington's disease (HD) is a polyglutamine-related disorder marked by the presence intranuclear inclusions (aggregates) of the protein - Huntingtin (Htt) in striatal neurons (DiFiglia et al. 1997; Sieradzan et al. 1999). At present, there is no cure available for HD. (Ross et al. 2014; Tabrizi et al. 2020). Huntingtin (Htt) is a 348 kDa multifunctional protein which shuttles between nucleus and cytoplasm, is broadly involved in the development of the nervous system, brain-derived neutrophic factor (BDNF) production and transport, and cell adhesion (Bates et al. 2015). The loss of native Htt function and its toxic gain-of-function via length expansion and aggregation of the polyglutamine tract (>36 residues) within the N-terminal Htt exon-1 (Httex1) is associated with HD (Gunawardena et al. 2003; Li et al. 2016). The length of polyQ expansion inversely correlates with the age of disease onset (Duyao et al. 1993). Notably, Httex1 fragments generated through proteolytic cleavage are observed in intranuclear inclusions (Kim et al. 1999) and *in vivo* expression of Httex1 was found to reproduce the keys features of HD in mice models (Yamamoto et al. 2000; Yang et al. 2020). Not surprisingly, the structural and conformational changes associated with polyglutamine expansion (Crick et al. 2006; Nagai et al. 2007; Warner et al. 2017) and the precise role of flanking regions in Httex1 oligomerization (Sánchez et al. 2003), fibrillar aggregation (Crick et al. 2013) and cellular toxicity (Cisbani and Cicchetti 2012) has been the subject of intense debate and investigations (Hatters 2008; Matlahov and van der Wel 2019).

Httex1 is an intrinsically disordered polypeptide (91 amino acids) comprising of (i) an N-terminal α-helical domain (N17: 17 amino acids), a central polyglutamine (polyQ$_n$) tract which is aggregation-prone and a C-terminal proline rich domain (PRD: 51 amino acids) **(Fig. 1A)**. The polyglutamine flanking domains (N17/PRD) exert opposing effects on fibrillar aggregation; N17 promotes fibrillation and suppresses the formation of non-fibrillar species, while PRD generally disfavors aggregation and promotes the formation of soluble oligomers (Crick et al. 2013; Shen et al. 2016; Chen and Wolynes 2017; Phan and Schmit 2020). In terms of the classical nucleation model for polyglutamine amyloid formation (Bhattacharyya et al. 2005), Wetzel and colleagues determined that the size of the critical nucleus required to template the formation of β-rich aggregates of dilysine-flanked polyglutamine peptides *in vitro* depends upon the length of polyQ: the nuclei for >Q$_{25}$ are monomeric while the smaller fragments are observed to form larger nuclei (n=2-4) (Kar et al. 2011). These findings were further corroborated in a computational study by Wolynes and co-workers based on simulated free energy landscapes of polyglutamine

aggregation computed using a coarse-grained protein model (Chen et al. 2016). Computational modeling also suggests that the nucleation of smaller polyQ fragments promotes elongation and fibrillation by reducing the conformational entropy of interacting monomers in a concentration-dependent manner (Phan and Schmit 2022). Recently, Halfmann and colleagues confirmed the formation of a monomeric nucleus *in vivo* for the pathogenic polyQ$_{60}$ fragment and determined its structural features using an integrative modeling approach (Kandola et al. 2023a).

N17 significantly enhances the rate of polyQ fibrillation *in vitro and in vivo* through a complex, multistep mechanism involving small (low molecular weight) prenucleation oligomers which undergo further rearrangement to promote amyloid nucleation (Thakur et al. 2009; Tam et al. 2009). Small oligomers of pathogenic Httex1-Q46 were shown to be a heterogeneous species *in vitro,* comprising of 7-11 subunits (Pandey et al. 2018). The formation of small oligomers proceeds through a tetrameric α-helical intermediate (Jayaraman et al. 2012a)**.** A kinetically competing, N17-independent aggregation pathway also exists for Huntingtin exon 1 which resembles that of simple polyglutamine and dominates upon disruption of N17 association or its partial excision (Jayaraman et al. 2012b). Notably, solid-state NMR experiments concluded that the core of Httex1 fibrils was structurally similar to simple polyglutamine fragments**,** comprising of an inter-digitating hydrogen bond network formed between adjacent layers of polyQ β-hairpins through glutamine sidechains while the flanking N17 remains helical (Sivanandam et al. 2011; Hoop et al. 2016; van der Wel 2024).

While higher-order oligomerization through N17 significantly increases the local concentration of polyQ tracts (‘proximity’ model), such a mechanism alone fails to explain the enhanced rate of polyQ aggregation as forcing a high local concentration of polyQ by fusing it to homo-oligomeric partners suppresses its fibrillar aggregation *in vitro* (Kokona et al. 2014) and *in vivo* (Kandola et al. 2023a). Two additional models were proposed to explain the conversion of Httex1 oligomers into aggregates via polyQ expansion (Kokona et al. 2014): (i) ‘transformation’ model, i.e., the propagation of N17 helical secondary structure into the polyQ tract and, (ii) ‘domain cross-talk’ model, i.e., tertiary interdomain interactions between N17 and polyQ domains. A model of Httex1 amyloid aggregation based on photocross-linking experiments proposes that long-range N17-polyQ interactions result in a conformational switch which promotes rearrangement of the initial N17-mediated oligomers to form aggregation-competent nuclei (Tam et al. 2009). The relevance of interdomain interactions for Httex1 aggregation is further supported by *in vivo* experiments where the selective binding of a polyQ β-hairpin antibody to mutant Httex1 oligomers (Nucifora et al. 2012) was favorable when atleast one of the flanking regions (N17 or PRD) were present (Shen et al. 2016).

Clore and co-workers determined the structural model of a lowly-populated tetrameric intermediate formed by the slowly-aggregating N17-polyQ$_7$ construct based on transient interactions detected by intermolecular PRE experiments and a simulated annealing protocol (Kotler et al. 2019). In comparison, the substantially faster aggregation of longer N17-polyQ constructs (>Q$_7$) renders them less amenable to high resolution NMR experiments for studying conformational exchange and detecting interdomain interactions occurring within their oligomeric assemblies (Kotler et al. 2019; Ceccon et al. 2022). Alternatively, molecular dynamics (MD) simulations have the potential to complement solution experiments and provide a high-resolution picture of the conformational landscape and underlying mechanisms which facilitate the conversion of α-helical oligomers into amyloid aggregates. Towards this goal, numerous studies have attempted to investigate the structural and conformational changes in Httex1 monomers upon polyQ expansion using atomistic MD simulations (Moldovean and Chiş 2020). Nevertheless, these studies report conflicting results regarding the nature of pathogenic conformations due to: (i) inherent imbalances in secondary structure propensities (i.e., relative stabilities of α-helical versus β-conformations) for conventional protein force fields (Best et al. 2008; Mittal and Best 2010; Lindorff-Larsen et al. 2012) and, (ii) the tendency to generate excessively collapsed ensembles for unfolded and disordered proteins when simulated in implicit or primitive (three-site) explicit water models (Rauscher et al. 2015; Henriques et al. 2015). However, recent improvements in MD force fields for proteins which correctly balance secondary structure propensities coupled with the adoption of accurate four-site water models, allow for the generation of accurate disordered protein ensembles which exhibit considerably improved agreement with biophysical experiments such as NMR, FRET and SAXS (Best and Mittal 2010; Best et al. 2014; Huang and MacKerell 2018; Robustelli et al. 2018).

In this study, we performed extensive MD simulations (aggregate time ~ 0.7 ms) of N17-polyQ fragments in monomeric and dimeric states to assess the mechanisms which underlie the conversion of Httex1 oligomers into β-rich aggregates. Using a state-of-the-art protein force field (Best et al. 2014), our simulations yielded conformational ensembles in good agreement with residue-specific helical populations inferred from high-resolution NMR experiments with site-specific isotope labeling of polyQ residues (Urbanek et al. 2020a; Elena-Real et al. 2023a). With increasing polyQ length, we observed longer α-helices extending further into the polyglutamine tract along with the emergence of transient β-sheet conformations (<2% total population) across multiple trajectories. Importantly, stable β-sheet conformations were not observed for a Q46-only fragment on a comparable simulation timescale, despite its low propensity for α-helix formation.

These findings imply that interactions between N17 and polyQ domains are directly implicated in the enhancing the formation of β-sheet conformations which could act as precursors in the folding pathway enroute to the critical nucleus. Finally, we investigated the interplay between structural transformation (i.e., elevated polyQ helicity) and interdomain interactions in dimers stabilized via N17-mediated interactions with increasing polyQ length. We observed that an increase in polyQ length (7 to 16) allowed transient, intra and intermolecular N17-polyQ interactions to effectively outcompete the stabilizing effect of α-helical transformation and promote a heterogenous dimer ensemble. Since N17-driven helical propagation into polyQ is structurally incompatible with its fibrillation, the observations from oligomer simulations collectively suggest that polyQ expansion favors a dynamic ensemble of N17 dimers and higher-order oligomers which may reorganize faster to promote interactions between proximal polyQ regions, facilitate nucleation and/or growth of β-rich species. In summary, our simulations coupled with prior experimental findings point towards an essential role for interdomain (N17-polyQ) interactions in modulating the structural and oligomerization landscape of Httex1 to favor pathological aggregation.

Results

Validation of N17-Q_{16} simulation ensembles against NMR chemical shifts

The α-helical structure of N17 is associated with enhanced oligomerization and aggregation of Httex1 compared to their simple polyQ counterparts. Recent structural studies employing NMR spectroscopy with site-specific isotope labeling resolved the residue-level α-helical propensities of amino acids within the N17 and polyQ regions in Httex1 constructs with increasing polyQ length below (Q_{16}) and above the pathogenic threshold (Q_{46}) (Urbanek et al. 2020b; Elena-Real et al. 2023b). Such high resolution, residue-specific information serves an invaluable reference data set for the structural validation of Httex1 ensembles generated by MD force fields. We chose a balanced protein force field - AMBER03ws (Best et al. 2014)**,** to generate conformational ensembles of N17-$(Q)_n$ monomers and dimers using multi-microsecond and enhanced sampling simulations (**Methods, Table S1**). Previous studies by us and others have firmly established the ability of AMBER03ws to accurately describe the global dimensions and local structural features of a wide variety of intrinsically disordered polypeptides (Robustelli et al. 2018; Zerze et al. 2019).

We first performed multi-microsecond simulations of the N17-Q_{16} fragment with a minimal C-terminal stretch of five proline residues (P_5). The mean probability distribution for the radius of gyration across three independent trajectories exhibited low standard errors, implying convergence of the N17-Q_{16} ensembles in terms of their global dimensions **(Fig. S1A)**. To validate the structural accuracy of the N17-Q_{16} ensemble, we first compared it with per-residue α-helical

propensities computed from ^{13}C chemical shifts using the SSP algorithm (Marsh et al. 2006). Overall, mean per-residue helical fraction of N17-polyQ$_{16}$ were found to be in excellent agreement with NMR across both N17 and Q$_{16}$ regions **(Fig. 1B)**. Further, the direct comparison of $^{13}C_{\alpha/\beta}$ chemical shifts predicted using the SPARTA+ program (See Materials and Methods) indicated that the root mean square deviation (RMSD) over all residues with respect to experiment are within the prediction error **(Fig. S1B)**. Finally, the comparison of per-residue ^{13}C secondary chemical shifts showed positive values in the N17 and polyQ region which were comparable to experiments thereby establishing the structural accuracy of our ensemble **(Fig. S1C)**. As negative controls, we also simulated Q$_{16}$-P$_5$ and Q$_{16}$ fragments; both exhibited substantially lower α-helicity compared to N17-polyQ$_{16}$ which is consistent with experiment **(Fig. 1B)**.

NMR experiments indicated that a four residue motif in N17 (^{14}LSKF^{17}) structurally connects N17 and the downstream polyQ residues (Urbanek et al. 2020b). Notably, these experiments revealed that ^{14}LKSF^{17} uniquely favored the formation of bifurcated hydrogen bonds involving glutamine sidechains (i→i-4) in the adjacent polyQ region (aa:18-24) which could contribute to the elevated α-helical propensity (Urbanek et al. 2020a). Similarly, an L$_4$ motif flanking the Androgen receptor polyQ tract was found to promote bifurcated hydrogen bonds which stabilized its α-helical structure (Escobedo et al. 2019). Based on site-specific labeling of a few glutamine residues, it was observed that substitution of the wild-type motif - ^{14}LKSF^{17} to either ^{14}LKAA^{17} or ^{14}LLLF^{17} enhanced helical stability while ^{14}LKGG^{17} substitution significantly reduced the helicity of the polyQ region. We simulated all three ^{14}LKSF^{17} variants to generate the per-residue helicity profiles of all mutants and assessed whether our force field could correctly capture the expected trends in their α-helicity profiles. The helix-promoting mutants showed an increase in α-helicity **(Fig. 1C)** starting from the C-terminal region of N17 (aa:11-17) leading into the polyQ region (aa:18-23). In contrast, ^{14}LKGG^{17} disrupted the structural connectivity between N17 and polyQ regions, resulting in a complete loss of α-helical structure in the Q$_{16}$ tract **(Fig. 1C)** and a corresponding increase in the population of coil conformations **(Fig. S1D)**. We also analyzed our ensembles for the prevalence of bifurcated hydrogen bonds in the polyQ tract among α-helical conformers (helix fraction >30% of total residues). While the overall occupancy of these hydrogen bonds was low, the observed trend correlated with their residue-level α-helical fractions: ^{14}LLLF^{17} variant showed the higher hydrogen bond occupancies (>5%) in the central region (Q18/19) compared to ^{14}LKAA^{17} and the wild-type sequence - ^{14}LKSF^{17} **(Figure S1E)**.

Overall, the above comparison of N17-Q$_{16}$ ensembles with high-resolution solution NMR data firmly establishes the suitability of our chosen model to generate structurally accurate

ensembles of N17-polyQ fragments, assess the effect of point mutations on α-helix stability and analyze the mechanisms of helix initiation and propagation, as discussed in following section.

Mechanism of α-helix nucleation and propagation into the polyQ tract

The high structural accuracy of our N17-Q_{16} ensemble motivated us to analyze the mechanisms of helix nucleation and propagation into the Q_{16} region which accounts for the observed residue-level α-helical populations. Accordingly, we first visualized the variation in secondary structure across both domains over the time course of the three independent trajectories (**Fig. 1D, S2A**). Upon careful inspection, we observed that α-helices generally appeared to nucleate in the region spanning the C-terminal N17 region (aa:11-17) and the adjacent Q_{16} tract (aa:18-23) and could further propagate in a bidirectional manner to form longer helices which persisted for several microseconds. However, these helices were less stable and did not persist for more than a microsecond. The N-terminal region of N17 (aa:3-10) formed transient helical structures (<1 μs) which were weakly associated with the formation of longer helices, suggestive of its structural decoupling from the downstream N17-polyQ sequence. Occasionally, N-terminal N17 also appeared to act as a nucleation site while also forming short α-helices which partially extended into the C-terminal N17 region **(Fig. S2A, bottom panel)**. In both N17-Q_{16} and Q_{16} trajectories, short α-helices (<8 residues) appeared within different stretches of polyQ sequence and usually persisted for less than 0.2 μs. In the case of Q_{16} trajectories, longer α-helices were also observed, although these typically did not persist beyond 1 μs, highlighting the critical requirement of N17 domain for stable α-helix nucleation and propagation into the polyQ tract (**Fig. 1D, S2B**).

The observations from time course analysis of N17-Q_{16} are summarized in the form of a secondary structure (SS) map (Iglesias et al. 2013) which describes the sequence-level cooperativity underlying α-helix formation by reporting on the helical populations for a range of possible lengths. The SS map reveals that α-helical populations were the highest for single turns (<5 residues) in either N- or C-terminal region of N17, followed by a short α-helix in the central region (aa:11-23) which can propagate into further into the polyQ tract to form longer helices (**Fig. 1E**). In comparison, the ^{14}LLLF^{17} variant showed a clear increase in the stability of longer α-helices (upto 15 residues) initiated from the central region as indicated their higher populations in SS map **(Fig. 1F).** The structural decoupling of the N-terminal N17 region was evident in N17-Q_{16} wild-type based on its reduced prevalence in longer helices (>8 residues) compared to the downstream N17-polyQ sequence and even more pronounced in case of the ^{14}LLLF^{17} variant.

Overall, the time course analysis of secondary structure formation and SS maps clearly establish a mechanism of helix initiation which is favored from the central region (aa:11-23) and its propagation into the polyQ tract leads to longer helices while the N-terminal N17 region (aa:3-10) has a low α-helix propensity and can be structurally uncoupled from the downstream sequence.

Emergence of β-sheet conformations upon polyQ length expansion

In vitro experiments indicate that the critical nucleus size for polyglutamine aggregation reduces from a multimer (2-4) to monomer in the case of longer fragments (>Q_{25}) (Kar et al. 2011). Based on these observations, an outstanding question with is how polyQ expansion influences the structural ensemble of the Httex1 monomer and induces conformational changes in the polyQ tract to favor aggregation. Current evidence suggest that the formation of β-hairpin structure in expanded Httex1 strongly favors amyloid aggregation. Firstly, a monoclonal antibody (3B5H10) recognizes a two-stranded β-hairpin conformation in the polyQ tract of mutant Httex1 monomer and its formation *in vivo* effectively predicts neurodegeneration (Miller et al. 2011; Peters-Libeu et al. 2012a). Secondly, the aggregation rate of a non-pathogenic Httex1 fragment can be enhanced to levels comparable to that of a pathogenic fragment through the introduction of short β-hairpin-forming motifs into polyglutamine sequences (Kar et al. 2013). Despite the importance of β-hairpin conformations in Httex1 aggregation and toxicity, NMR and FRET experiments indicate an increase in α-helicity and rigidification of the polyQ tract with length expansion (Caron et al. 2013; Bravo-Arredondo et al. 2018; Elena-Real et al. 2023a, b). Further, the elevated helicity in the pathogenic Httex1-Q46 also correlated with an enhanced rate of aggregation *in vitro* and *in vivo* (Elena-Real et al. 2023a).

To address the underlying interplay between α-helix and β-structure formation in the context of polyQ expansion, we performed multi-microsecond simulations (**Table S1**) of N17-polyQ fragments with increasing polyQ length: $Q_{24/32/46}$ and compared their structural ensembles. Consistent with NMR experiments, the per-residue α-helix fractions increased in a polyQ length-dependent manner (Q_{16} to Q_{46}) (**Fig. S2A**). The elevated α-helix fraction was predominantly associated within residues beyond the central polyQ tract, implying long-range structural cooperativity across the N17 and expanded polyQ. In the case of N17-Q_{46}, however, multi-microsecond simulations (aggregate duration~105 μs) yielded an ensemble with lower (15-20%) α-helicity per-residue in the polyQ tract compared to NMR estimates (**Fig. S3A**) despite the choice of diverse initial conformations and the low error estimates for R_g distributions across the

trajectories (**Fig. S3B**). With parallel tempering (PT-WTE) simulations (750 ns per replica), we observed improved agreement with NMR for α-helicity within the polyQ tract (**Fig. 2A, S6A**). These observations highlight the difficulty associated with adequate sampling of structural transitions for IDPs with α-helical propensity using multi-microsecond simulations, due to long helix nucleation times (De Sancho and Best 2011). The computed SS map of the N17-Q_{46} ensemble from the PT-WTE temperature replica (293.15 K) clearly indicates the elevated populations of longer helices (25-30 residues) which formed via propagation from the central region as observed for N17-Q_{16} (**Fig. 1E**).

A closer inspection of the independent trajectories revealed the formation of transient two-stranded β-sheets (See Methods) for all three expanded N17-polyQ constructs (**Fig. 2C S4, S5A-B**) which was not observed in N17-Q_{16}, visualized in the two-dimensional intramolecular contact maps computed for each of the ensembles (**Fig. 2D, S6**). For N17-$Q_{24,}$ a two-stranded β-sheet structure formed in two trajectories (total population~0.5%) (**Fig. S3A**) while only a β-bridge structure involving N17 and polyQ was also observed in another trajectory. For N17-Q_{32}, two out of three trajectories showed the formation of two-stranded β-sheet structures (total population ~ 1.8%). Among the six N17-Q_{46} trajectories, β-sheet conformations involving the polyQ tract were observed in five trajectories (six conformations) with an aggregate population of 1.8% (**Fig. 2C, S5A-B**). Interestingly, half of the β-conformations of N17-Q_{46} were directly stabilized by hydrogen-bonding with N17, the other half involved hydrogen bonding within the polyQ region (**Fig. 2E-F, S5C**). Similar to the unbiased N17-Q_{46} trajectories, a low population of β-sheet conformations (1.1%) were also observed in the PT-WTE 293 K replica trajectory (**Fig. S6B**), further confirming their unstable nature compared to α-helical and coil conformations.

The direct participation of N17 in the formation of β-bridges/sheets (**Fig. 2F, left**) raises a possibility that emergence of β-sheet conformations in N17-Q_{46} may can be enhanced through transient interdomain interactions. In support of this idea, the contact map of a β-sheet conformation stabilized within the polyQ region also showed weak N17-polyQ interdomain interactions (**Fig. 2F, right)**. As stated earlier, the conformation-specific Httex1 antibody (3B5H10) selectively binds to a two-stranded β-hairpin conformation formed within polyQ tract of pathogenic Httex1 (Peters-Libeu et al. 2012a). Notably, the antibody bound to the polyQ tract *in vivo* only when either of the Httex1 flanking domains (N17/PRD) were present (Shen et al. 2016). These observations suggest that polyQ flanking regions such as N17 can modulate the polyQ conformational ensemble and promote the emergence of pathogenic conformations such as intramolecular β-hairpins which favor amyloid aggregation (Kar et al. 2013). To directly test the

idea that N17 promotes β-conformations in the polyQ tract through interdomain interactions, we performed multi-microsecond (triplicate) simulations of a simple Q_{46} polypeptide for comparable duration to N17-Q_{46}. The analysis of secondary structure variation over the time course of the Q_{46} trajectories (**Fig. S8**) indicated an absence of β-sheet/bridge conformations which were stable on timescales comparable to those seen for N17-Q_{46}, thereby confirming the ability of interdomain N17-polyQ interactions to promote transient chain compaction and form potentially toxic β-conformations.

Inter-domain interactions modulate the stability of N17-mediated Httex1 oligomers

NMR experiments studies indicate that the stereospecific tetramerization of Httex1 via the helical N17 region is essential for efficient nucleation and aggregation (Kotler et al. 2019). Interestingly, the transformation of prenucleation tetramers into amyloid fibrils remains unfavorable for polyQ length < 8 residues (Jayaraman et al. 2012a). Therefore, a mere increase in the local concentration of polyQ tracts through oligomerization is insufficient to explain its aggregation behavior. Two additional models were invoked based on *in vitro* experiments to explain the underlying mechanism (Kokona et al. 2014): (i) transformation model, i.e., an increase in α-helical structure and (ii) domain "cross-talk" model, i.e., inter-domain interactions. The precise atomic details of the steps leading to the conversion of oligomers into aggregates upon polyQ expansion currently remain difficult to through solution NMR experiments due to rapid aggregation and insufficient data collection.

Based on PRE measurements and simulated annealing (Kotler et al. 2019), a structural model of α-helical tetrameric (<1% population) was determined for a N17-Q_7 construct (high solubility) and found to comprise of two anti-parallel N17 dimer units stabilized by hydrophobic interactions (**Fig. 3A**). The formation of the tetrameric intermediate of Httex1 initially involves weak dimerization (K_d ~ 0.1-10 mM) through the N17 domain as detected by NMR and analytical ultracentrifugation (Kotler et al. 2019). The dimeric ensemble was determined to comprise of "productive" (i.e., anti-parallel) dimers which can effectively assemble into tetramers and "non-productive" dimers which cannot undergo tetramerization. We simulated four "productive" dimer variants on the microsecond timescale: N17, N17-polyQ$_7$, N17-polyQ$_{16}$ and its ^{14}LKGG^{17} variant, to tease apart the interplay between monomeric α-helix formation and inter-domain contact dynamics on the stability of the N17 dimer and evaluate the plausibility of the proposed models to explain oligomer-mediated aggregation.

The simulations of N17-polyQ$_{7/16}$ dimers showed an oligomerization-dependent stabilization of α-helical structure in N17 **(Fig. 3B, S9A-C)** which is consistent with the CD and

NMR experiments (Jayaraman et al. 2012a; Kotler et al. 2019)**.** To characterize the association free energy landscape of the dimer variants, two-dimensional PMF plots were computed using two order parameters chosen: (i) The total number of contacts between N17 monomers which distinguishes between native-like dimers, alternate dimers, and weakly-associated states, and (ii) the inter-helical (N17) angle to characterize the diversity of monomer orientations in addition to the native (NMR) dimer basin (**Fig**. **3C**). The N17-Q_7 dimer maintained its stability over microsecond timescale; none of the trajectories showed dissociation of the dimer (**Fig. 3C, S10A** - top row, left) on the simulated timescale (~2.3 μs per trajectory). Interestingly, however, the bound dimer showed considerable structural heterogeneity, comprising an array of monomer orientations ranging from antiparallel (native-like), perpendicular, and parallel. Further, additional energy basins (favorable regions) were observed for parallel orientations. As the parallel dimer orientations are not compatible with the structural model of the tetrameric species, we speculate that these may represent the "unproductive" dimers which do not promote higher-order oligomerization and subsequent aggregation. Despite the wide range of sampled dimer orientations, the N17-Q_7 monomer units formed stable α-helices (**Fig. 3B, S9A**).

The expansion of polyQ from 7 to 16 resulted in a substantial reduction of the dimer stability as evident by the loss of numerous native-like and parallel dimer energy basins (**Fig. 3C** - bottom row, left). Three out of six trajectories showed complete dissociation of the dimer followed by multiple weak reassociation events (**Fig S10C**) which show a reduced number of contacts (<25) between N17 domains and α-helix stability (**Fig. S9B, C**) compared to the bound ensemble. To assess whether the domain "cross-talk" model (Kokona et al. 2014) could explain the reduced stability of N17-polyQ$_{7/16}$, we computed contact maps of both N17-polyQ$_7$/Q$_{16}$ dimer ensembles. In support of the domain "cross-talk" model, contact map analysis indicated an extensive network of inter-domain N17-polyQ interactions for N17-polyQ$_{16}$ (**Fig. S9E, F**). Notably, the expansion of the polyQ tract to 16 residues also caused helix propagation into the first few glutamine residues, as indicated by their elevated α-helical populations **(Fig. 3B)**.

To determine the effect of polyQ helical transformation on the stability of the bound dimer ensemble, we simulated two dimer variants: N17 and N17-Q_{16} dimer variant with ^{14}LKGG^{17} substitution. ^{14}LKGG^{17} disrupts α-helical propagation **(Fig. 1C)** and reduces the rate of aggregation for pathogenic Httex1-Q46 (Elena-Real et al. 2023a). The N17 dimer was found to be highly unstable compared to N17-Q_7; five out of the six trajectories resulted in complete dissociation (**Fig. 3C, 310B** - top row/right) due to the lower α-helicity of N17 in the absence of a downstream polyQ tract. Similarly, among the ^{14}LKGG^{17} trajectories, five out of six trajectories showed complete dissociation followed by weak reassociation (**Fig S10D)**. Accordingly, the

resulting dimerization landscape shows a lower population of stable N17-mediated dimers (**Fig. 3C** bottom row, right) and a concomitant increase in the population of weakly-associated dimers. These results support the interpretation that structural transformation (i.e., increase in α-helicity) of the monomer promotes N17 dimer stability and provides a plausible explanation for the reduced aggregation propensity of Httex1-Q46 ^{14}LKGG^{17} variant.

The above findings collectively suggest that the main effect of polyQ expansion is to destabilize N17-mediated oligomers through enhanced interdomain N17-polyQ interactions. These results also suggest that there is likely an optimal affinity associated with N17-polyQ helical transformation which favors polyQ fibrillation, consistent with the inhibitory effect of stable coiled-coil oligomerization on polyQ aggregation (Kokona et al. 2014; Kandola et al. 2023b). Importantly, the reduced stability of N17-dimers may promote homotypic interactions between the proximal polyQ tracts (Williamson et al. 2010) required for the emergence of aggregation-competent nuclei and formation of amyloid protofibrils. Further, the higher population of weakly-associated dimers and monomers seen in the case of N17-polyQ$_{16}$ may provide an effective reservoir of molecules for the elongation of fibrils which form initially via productive oligomerization (Jayaraman et al. 2012a).

Discussion

Emerging evidence from biophysical investigations suggest that there is a complex interplay between the N-terminal and C-terminal flanking sequence composition and structure, and polyQ length in modulating the overall aggregation propensity of polyQ proteins. Among the polyQ proteins implicated in neurodegenerative disorders, three of the proteins: Huntingtin exon 1 (Httex1), Androgen Receptor (AR), and Ataxin-7 (Atx-7) contain hydrophobic N-terminal flanking sequences which induce and progressively increase the α-helical propensity in the polyQ tract with length expansion (Eftekharzadeh et al. 2016; Hong et al. 2019; Urbanek et al. 2020b). However, the increased α-helical propensity in the polyQ region appears to have dissimilar effects on the aggregation of these polyQ-containing sequences. While both AR and Atx-7 showed suppressed polyQ aggregation for polyQ-25/33 (Eftekharzadeh et al. 2016; Hong et al. 2019), aggregation of longer polyQ sequences (Q46-72) is favored in the case of Httex1 (Elena-Real et al. 2023b). It is therefore important to decipher the role of polyQ flanking sequences in order to obtain mechanistic insights into the amyloid aggregation of polyQ proteins and develop novel therapeutics to treat their associated-neurodegenerative disorders.

The rapid aggregation kinetics of pathogenic Httex1 fragments limit the ability to achieve a complete atomic-level characterization of the role of flanking domains on structural and

oligomerization landscape through experimental approaches alone. Further, conflicting results from previous MD simulation studies (Długosz and Trylska 2011; Kang et al. 2017; Khaled et al. 2023) due to force field and water model inaccuracies has led to a lack of clarity regarding the precise molecular mechanisms of Httex1 oligomerization and pathogenic aggregation. Our study provides a comprehensive characterization of the structural ensembles of normal and pathogenic N17-polyQ fragments using the AMBER03ws force field which was carefully tuned to achieve balanced secondary structure propensities and protein-water interactions (Best and Mittal 2010; Best et al. 2014). Importantly, our simulation ensembles correctly captured the length-dependent increase in helical from Q_{16} to Q_{46}, achieving good quantitative agreement with NMR in the case of N17-$Q_{16/46}$ fragments. Therefore, a distinguishing feature of this computational study is the extensive structural validation of N17-polyQ simulation ensembles against recent high-resolution NMR experiments (Urbanek et al. 2020b; Elena-Real et al. 2023b), which considerably enhances the reliability of our predictions. In addition, our simulations also captured the effect of point mutations in the hydrophobic four-residue motif (^{14}LKSF^{17}) which structurally couples N17 with the polyQ tract (Urbanek et al. 2020b). The N17 domain is also subject to numerous post-translational modifications (PTMs) which alter its α-helical structure and impacts the cellular localization, oligomerization, and aggregation of Htt. (Mishra et al. 2012a; Gu et al. 2015; DiGiovanni et al. 2016; DeGuire et al. 2018). Our chosen model and simulation methodology demonstrated here opens the way for accurate predictions of the effect of point mutations and PTMs on the structure and conformation of Httex1 and other polyQ proteins which contain N-terminal hydrophobic motifs adjacent to the polyQ tract.

Two opposing models are proposed to explain the coupling between structure and binding properties of polyglutamine with increasing length, namely the (i) “linear lattice” model and, (ii) “conformational emergence” model. The linear lattice model proposes that the enhanced multivalency (number of ligand-binding sites) of polyglutamine upon length expansion coupled with the prevalence of an extended, random coil structural ensemble (abundant in N17-Q_{46}/Q_{46} simulations) explains its aberrant cellular interactions and increased binding affinity towards anti-polyQ antibodies (Bennett et al. 2002; Klein et al. 2013; Owens et al. 2015). Nevertheless, several studies also point towards the existence of β-hairpin conformations formed in pathogenic polyglutamine (Nagai et al. 2007) and Httex1 (Miller et al. 2011; Peters-Libeu et al. 2012b; Kim 2013; Shen et al. 2016) in monomeric and oligomeric species. Interestingly, our results appear to reconcile the two opposing models in the case of pathogenic Httex1, by suggesting that the increased multivalency can also promote intramolecular interdomain (N17-polyQ) interactions **(Fig. 4)** which may give rise to “toxic” β-conformations. Alternatively, such interactions may also

promote polyglutamine “folding” to form the critical β-hairpin nucleus, which can template the formation of amyloids through a kinetically competing N17-independent pathway (Jayaraman et al. 2012b). It is important to note, however, that β-conformers observed in our simulations constitute a low population species (<2%) in the conformational landscape of expanded Httex1, implying that such conformations remain highly unfavorable relative to helical and coil conformations. These observations are consistent with a large free energy barrier associated with polyglutamine nucleation (ΔG_{Q47} = 12.2 kcal mol^{-1}) and the delayed onset (>40 years) of HD symptoms for nanomolar protein concentrations found within cells (Chen et al. 2002).

We provide a high-resolution view into the competing effects of domain “cross-talk” and structural “transformation’ due to polyQ length expansion, on N17-mediated dimerization of the Httex1. Based on the reduction in dimer stability of N17-Q_{16} in simulations and its expectedly faster aggregation kinetics compared to N17-Q_7 (Jayaraman et al. 2012a), we build upon the previous models of oligomerization-dependent Httex1 aggregation and propose that polyQ expansion (>7) favors interdomain N17-polyQ interactions which outcompete the stabilizing effect of α-helical transformation on N17-mediated oligomerization **(Fig. 4)**. This results in a heterogenous landscape of dimerization and by extension, higher-order oligomerization. The capacity for heterogenous oligomerization upon polyQ expansion may promote a greater diversity of intermolecular interactions once the local concentration of polyQ tracts are increased through N17-mediated oligomerization. These include homotypic interactions between proximal polyQ tracts which are likely critical for the emergence of nuclei from the “excited” pre-nucleation tetramers and culminates in the formation of “ground-state” protofibrils as proposed in a recent kinetic NMR study of Httex1-Q_{35} aggregation (Torricella et al. 2024). Further, the increased population of weakly-stabilized dimers arising due to increased interdomain interactions may contribute to rapid fibril growth by contributing free monomers which was proposed earlier (Jayaraman et al. 2012a). Our findings directly support the idea that modulating a combination of both homo- (N17) (Mishra et al. 2012b) and heterodomain (N17-polyQ) interactions (Tam et al. 2009) in Httrex1 oligomers may serve as a promising therapeutic strategy for the inhibition of pathological self-assembly.

Acknowledgments

This work was supported by NIH/NIGMS R35GM153388. Atomistic MD simulations were conducted with the advanced computing resources provided by Texas A&M High Performance Research Computing. We are thankful to Pau Bernado and Amin Sagar for providing us with the ^{13}C NMR chemical shift data to validate the atomistic simulations presented in this work and their critical feedback on the manuscript during its preparation. We would also like to gratefully acknowledge Randal Halfmann and Shriram Venkatesan for fruitful discussions, and their critical feedback during the initial writing process and preparation of the final manuscript.

Materials and Methods

Initial structure, force field choice and system setup. For N17-polyQ$_{16/24/32}$ monomer, simulations were initiated from three different coil configurations generated based on the Flexible Meccano algorithm (Ozenne et al. 2012) using the ProtSA webserver (Estrada et al. 2009). For PT-WTE simulations N17-polyQ$_{46}$, a single coil conformation was chosen among the three used for unbiased simulations. The three partially-helical, initial conformations of N17-Q$_{46}$ were obtained from the first microsecond of a coil trajectory. For dimer simulations, the initial complex was extracted from the lowest energy NMR model of the N17 tetramer (PDB 6N8C). PolyQ sequences (Q$_{7/16}$) were added to the monomer units of the N17 dimer using MODELLER (Webb and Sali 2016). The initial configurations were solvated in octahedral boxes with 150 mM NaCl, and additional counter ions added to achieve electroneutrality. The polypeptide and water topologies were described using parameters from the AMBER03ws force field (Best et al. 2014) (https://bitbucket.org/jeetain/all-atom_ff_refinements/src/master/) and TIP4P/2005 water model (Abascal and Vega 2005) respectively. For Na^+ and Cl^- ions, modified LJ parameters (Luo and Roux 2010) were used to improve ion solubility. Details of box dimensions and duration of simulation trajectories for each N17-polyQ monomer and dimer trajectories are provided in **Table S1**.

Conventional MD Simulation Protocol. To remove any steric clashes, the solvated polypeptide systems were first relaxed by performing energy minimization using the steepest descent algorithm in GROMACS-2020.4 (Páll et al. 2020). Following minimization, the systems were simulated for 100 ps using Nose-Hoover thermostat (Evans and Holian 1985) with a coupling constant (τ_c) of 1 ps to achieve temperature equilibration at 293.15 K. Next, a 100 ps simulation was conducted using the Berendsen barostat (Berendsen et al. 1984) with isotropic coupling and

a coupling constant (τ_p) of 5 ps for pressure control, to achieve pressure equilibration at 1 bar. Production simulations were performed in the NVT ensemble was performed using the Langevin Middle Integrator (friction coefficient = 1 ps^{-1}) (Zhang et al. 2019) in OpenMM-7.5 (Eastman et al. 2017). Short-range nonbonded interactions were computed using a cutoff radius of 0.9 nm, while long-range electrostatics were described with the PME method (Darden et al. 1993; Essmann et al. 1995). The hydrogen masses were increased by 1.5 times to enable a simulation timestep of 4 fs (Hopkins et al. 2015) for the production runs, and hydrogen-containing bonds were constrained using the SHAKE algorithm (Ryckaert et al. 1977).

Parallel Tempering Well-Tempered Ensemble (PT-WTE) MD Simulation Protocol and Assessment of Convergence. Following energy minimization, temperature, and pressure equilibration of the system at 293.15 K and 1 bar, temperature replicas were prepared and equilibrated in the NVT ensemble without exchange for 10 ns at higher temperatures using the Langevin integrator (friction coefficient = 1 ps^{-1}). A total of 16 temperature replicas were generated based on a geometric progression that spanned a range from 293 to 500 K. Long-range electrostatic interactions were modeled using the Particle Mesh Ewald method (Darden et al. 1993; Essmann et al. 1995) with a real space distance cutoff of 0.9 nm. The hydrogen mass repartitioning scheme (Hopkins et al., 2015) was used to reduce the computational cost. The LINCS algorithm (Hess et al., 1997) was used to constrain all bonds, allowing a time step of 5 fs. Following temperature equilibration of the replicas, production runs were performed in the NVT ensemble with bias potentials (bias factor = 72) deposited every 4 ps to increase fluctuations in the system potential energy (collective variable) resulting in a well-temperature ensemble (WTE) with improved frequency of conformational exchange between adjacent replicas (Bonomi and Parrinello 2010; Deighan et al. 2012), which was attempted every 1 ps. The initial height and width of the Gaussian potential were set to 1.5 and 460.0 kJ mol^{-1}, respectively. Exchange probabilities between adjacent replicas varied from 17-30%.

The first 150 ns from the 293 K replica was discarded as equilibration time, during which the initial height of the gaussian potential dropped below 0.25 kJ mol^{-1}, and sufficient overlap in the potential energy distributions between adjacent temperature replicas was observed. The initial configurations for each replica underwent several round trips through the entire temperature space, which indicates efficient conformational exchange between replicas. The convergence of global chain dimensions was determined based on the radius of gyration (R_g) for the 293 K ensemble. The low errors associated with their probability distributions averaged over 150 ns intervals indicate that the N17-Q_{46} ensemble is well-converged in terms of its global dimensions.

***Trajectory Analysis*.** Secondary structure calculations were performed based on the DSSP library (Kabsch and Sander 1983) using the gmx do_dssp program. The calculation of NMR chemical shifts and comparison with experiments was performed using SPARTA+ as described in our previous study (Mohanty et al. 2023). The calculation of secondary structure (SS) maps for assessing the α-helicity cooperativity of N17-polyQ regions was performed using a modified version of the python script provided as part of the original study to include the DSSP definition for identification of α-helices. For the detection of bifurcated hydrogen bonds (B-HBs), gmx hbond in combination with in-house python scripts were used to compute the fraction of trajectory frames in which glutamine sidechain to backbone (i $\rightarrow$ i-4) hydrogen bonds coexisted with backbone-backbone (i-4 $\rightarrow$ i) hydrogen bonds. B-HB analysis was carried out only on frames in which >30% of the residues adopted an α-helical conformation. For the analysis and visualization of β-sheet conformations, only frames where >4 residues adopted a β-sheet conformations were extracted from N17-Q_{46} unbiased and PT-WTE trajectories. For contact analysis, two residues i and j with sequence separation greater 3 (|i-j|>3) were in contact if any two heavy atoms of the residues were within 0.6 nm of each other. The chosen distance cutoff has been extensively tested in our previous studies (Zheng et al. 2020; Mohanty et al. 2023, 2024) and is large enough to include the diverse contact modes such as hydrogen bonds (<0.35 nm), van der Waal interactions and salt-bridges (<0.6 nm). For the analysis of N17-polyQ dimerization, the minimum distance, and orders parameters (number of N17-N17 contacts and inter-helical angle) for 2D-PMFs **(Fig. 3C)** were computed using gmx mindist and gmx gangle.

References

Abascal JLF, Vega C (2005) A general purpose model for the condensed phases of water: TIP4P/2005. J Chem Phys 123:234505. https://doi.org/10.1063/1.2121687

Adegbuyiro A, Sedighi F, Pilkington AW, et al (2017) Proteins Containing Expanded Polyglutamine Tracts and Neurodegenerative Disease. Biochemistry 56:1199–1217. https://doi.org/10.1021/acs.biochem.6b00936

Bates GP, Dorsey R, Gusella JF, et al (2015) Huntington disease. Nat Rev Dis Primers 1:15005. https://doi.org/10.1038/nrdp.2015.5

Bennett MJ, Huey-Tubman KE, Herr AB, et al (2002) A linear lattice model for polyglutamine in CAG-expansion diseases. Proceedings of the National Academy of Sciences 99:11634–11639. https://doi.org/10.1073/pnas.182393899

Berendsen HJC, Postma JPM, van Gunsteren WF, et al (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690. https://doi.org/10.1063/1.448118

Best RB, Buchete N-V, Hummer G (2008) Are current molecular dynamics force fields too helical? Biophys J 95:L07-9. https://doi.org/10.1529/biophysj.108.132696

Best RB, Mittal J (2010) Protein Simulations with an Optimized Water Model: Cooperative Helix Formation and Temperature-Induced Unfolded State Collapse. J Phys Chem B 114:14916–14923. https://doi.org/10.1021/jp108618d

Best RB, Zheng W, Mittal J (2014) Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. J Chem Theory Comput 10:5113–5124. https://doi.org/10.1021/ct500569b

Bhattacharyya AM, Thakur AK, Wetzel R (2005) Polyglutamine aggregation nucleation: Thermodynamics of a highly unfavorable protein folding reaction. Proceedings of the National Academy of Sciences 102:15400–15405. https://doi.org/10.1073/pnas.0501651102

Bonomi M, Parrinello M (2010) Enhanced Sampling in the Well-Tempered Ensemble. Phys Rev Lett 104:190601. https://doi.org/10.1103/PhysRevLett.104.190601

Bravo-Arredondo JM, Kegulian NC, Schmidt T, et al (2018) The folding equilibrium of huntingtin exon 1 monomer depends on its polyglutamine tract. Journal of Biological Chemistry 293:19613–19623. https://doi.org/10.1074/jbc.RA118.004808

Caron NS, Desmond CR, Xia J, Truant R (2013) Polyglutamine domain flexibility mediates the proximity between flanking sequences in huntingtin. Proceedings of the National Academy of Sciences 110:14610–14615. https://doi.org/10.1073/pnas.1301342110

Ceccon A, Tugarinov V, Torricella F, Clore GM (2022) Quantitative NMR analysis of the kinetics of prenucleation oligomerization and aggregation of pathogenic huntingtin exon-1 protein. Proceedings of the National Academy of Sciences 119:. https://doi.org/10.1073/pnas.2207690119

Chen M, Tsai M, Zheng W, Wolynes PG (2016) The Aggregation Free Energy Landscapes of Polyglutamine Repeats. J Am Chem Soc 138:15197–15203. https://doi.org/10.1021/jacs.6b08665

Chen M, Wolynes PG (2017) Aggregation landscapes of Huntingtin exon 1 protein fragments and the critical repeat length for the onset of Huntington's disease. Proceedings of the National Academy of Sciences 114:4406–4411. https://doi.org/10.1073/pnas.1702237114

Chen S, Ferrone FA, Wetzel R (2002) Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. Proceedings of the National Academy of Sciences 99:11884–11889. https://doi.org/10.1073/pnas.182276099

Cisbani G, Cicchetti F (2012) An in vitro perspective on the molecular mechanisms underlying mutant huntingtin protein toxicity. Cell Death Dis 3:e382–e382. https://doi.org/10.1038/cddis.2012.121

Crick SL, Jayaraman M, Frieden C, et al (2006) Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. Proceedings of the National Academy of Sciences 103:16764–16769. https://doi.org/10.1073/pnas.0608175103

Crick SL, Ruff KM, Garai K, et al (2013) Unmasking the roles of N- and C-terminal flanking sequences from exon 1 of huntingtin as modulators of polyglutamine aggregation. Proceedings of the National Academy of Sciences 110:20075–20080. https://doi.org/10.1073/pnas.1320626110

Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N ·log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092. https://doi.org/10.1063/1.464397

De Sancho D, Best RB (2011) What Is the Time Scale for α-Helix Nucleation? J Am Chem Soc 133:6809–6816. https://doi.org/10.1021/ja200834s

DeGuire SM, Ruggeri FS, Fares M-B, et al (2018) N-terminal Huntingtin (Htt) phosphorylation is a molecular switch regulating Htt aggregation, helical conformation, internalization, and nuclear targeting. Journal of Biological Chemistry 293:18540–18558. https://doi.org/10.1074/jbc.RA118.004621

Deighan M, Bonomi M, Pfaendtner J (2012) Efficient simulation of explicitly solvated proteins in the well-tempered ensemble. J Chem Theory Comput 8:2189–2192. https://doi.org/10.1021/ct300297t

DiFiglia M, Sapp E, Chase KO, et al (1997) Aggregation of Huntingtin in Neuronal Intranuclear Inclusions and Dystrophic Neurites in Brain. Science (1979) 277:1990–1993. https://doi.org/10.1126/science.277.5334.1990

DiGiovanni LF, Mocle AJ, Xia J, Truant R (2016) Huntingtin N17 domain is a reactive oxygen species sensor regulating huntingtin phosphorylation and localization. Hum Mol Genet 25:3937–3945. https://doi.org/10.1093/hmg/ddw234

Długosz M, Trylska J (2011) Secondary Structures of Native and Pathogenic Huntingtin N-Terminal Fragments. J Phys Chem B 115:11597–11608. https://doi.org/10.1021/jp206373g

Duyao M, Ambrose C, Myers R, et al (1993) Trinucleotide repeat length instability and age of onset in Huntington's disease. Nat Genet 4:387–392. https://doi.org/10.1038/ng0893-387

Eastman P, Swails J, Chodera JD, et al (2017) OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. PLoS Comput Biol 13:e1005659. https://doi.org/10.1371/journal.pcbi.1005659

Eftekharzadeh B, Piai A, Chiesa G, et al (2016) Sequence Context Influences the Structure and Aggregation Behavior of a PolyQ Tract. Biophys J 110:2361–2366. https://doi.org/10.1016/j.bpj.2016.04.022

Elena-Real CA, Sagar A, Urbanek A, et al (2023a) The structure of pathogenic huntingtin exon 1 defines the bases of its aggregation propensity. Nat Struct Mol Biol 30:309–320. https://doi.org/10.1038/s41594-023-00920-0

Elena-Real CA, Urbanek A, Lund XL, et al (2023b) Multi-site-specific isotopic labeling accelerates high-resolution structural investigations of pathogenic huntingtin exon-1. Structure 31:644-650.e5. https://doi.org/10.1016/j.str.2023.04.003

Escobedo A, Topal B, Kunze MBA, et al (2019) Side chain to main chain hydrogen bonds stabilize a polyglutamine helix in a transcription factor. Nat Commun 10:2034. https://doi.org/10.1038/s41467-019-09923-2

Essmann U, Perera L, Berkowitz ML, et al (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577–8593. https://doi.org/10.1063/1.470117

Estrada J, Bernadó P, Blackledge M, Sancho J (2009) ProtSA: a web application for calculating sequence specific protein solvent accessibilities in the unfolded ensemble. BMC Bioinformatics 10:104. https://doi.org/10.1186/1471-2105-10-104

Evans DJ, Holian BL (1985) The Nose–Hoover thermostat. J Chem Phys 83:4069–4074. https://doi.org/10.1063/1.449071

Gu X, Cantle JP, Greiner ER, et al (2015) N17 Modifies Mutant Huntingtin Nuclear Pathogenesis and Severity of Disease in HD BAC Transgenic Mice. Neuron 85:726–741. https://doi.org/10.1016/j.neuron.2015.01.008

Gunawardena S, Her L-S, Brusch RG, et al (2003) Disruption of Axonal Transport by Loss of Huntingtin or Expression of Pathogenic PolyQ Proteins in Drosophila. Neuron 40:25–40. https://doi.org/10.1016/S0896-6273(03)00594-4

Hatters DM (2008) Protein misfolding inside cells: The case of huntingtin and Huntington's disease. IUBMB Life 60:724–728. https://doi.org/10.1002/iub.111
Henriques J, Cragnell C, Skepö M (2015) Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. J Chem Theory Comput 11:3420–3431. https://doi.org/10.1021/ct501178z
Hong J-Y, Wang D-D, Xue W, et al (2019) Structural and dynamic studies reveal that the Ala-rich region of ataxin-7 initiates α-helix formation of the polyQ tract but suppresses its aggregation. Sci Rep 9:7481. https://doi.org/10.1038/s41598-019-43926-9
Hoop CL, Lin H-K, Kar K, et al (2016) Huntingtin exon 1 fibrils feature an interdigitated β-hairpin–based polyglutamine core. Proceedings of the National Academy of Sciences 113:1546–1551. https://doi.org/10.1073/pnas.1521933113
Hopkins CW, Le Grand S, Walker RC, Roitberg AE (2015) Long-time-step molecular dynamics through hydrogen mass repartitioning. J Chem Theory Comput 11:1864–1874. https://doi.org/10.1021/ct5010406
Huang J, MacKerell AD (2018) Force field development and simulations of intrinsically disordered proteins. Curr Opin Struct Biol 48:40–48. https://doi.org/10.1016/j.sbi.2017.10.008
Iglesias J, Sanchez-Martínez M, Crehuet R (2013) SS-map. Intrinsically Disord Proteins 1:e25323. https://doi.org/10.4161/idp.25323
Jayaraman M, Kodali R, Sahoo B, et al (2012a) Slow Amyloid Nucleation via α-Helix-Rich Oligomeric Intermediates in Short Polyglutamine-Containing Huntingtin Fragments. J Mol Biol 415:881–899. https://doi.org/10.1016/j.jmb.2011.12.010
Jayaraman M, Mishra R, Kodali R, et al (2012b) Kinetically Competing Huntingtin Aggregation Pathways Control Amyloid Polymorphism and Properties. Biochemistry 51:2706–2716. https://doi.org/10.1021/bi3000929
Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637. https://doi.org/10.1002/bip.360221211
Kandola T, Venkatesan S, Zhang J, et al (2023a) Pathologic polyglutamine aggregation begins with a self-poisoning polymer crystal. Elife 12:. https://doi.org/10.7554/eLife.86939
Kandola T, Venkatesan S, Zhang J, et al (2023b) Pathologic polyglutamine aggregation begins with a self-poisoning polymer crystal. bioRxiv. https://doi.org/10.1101/2023.03.20.533418
Kang H, Vázquez FX, Zhang L, et al (2017) Emerging β-Sheet Rich Conformations in Supercompact Huntingtin Exon-1 Mutant Structures. J Am Chem Soc 139:8820–8827. https://doi.org/10.1021/jacs.7b00838
Kar K, Hoop CL, Drombosky KW, et al (2013) β-Hairpin-Mediated Nucleation of Polyglutamine Amyloid Formation. J Mol Biol 425:1183–1197. https://doi.org/10.1016/j.jmb.2013.01.016
Kar K, Jayaraman M, Sahoo B, et al (2011) Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent. Nat Struct Mol Biol 18:328–336. https://doi.org/10.1038/nsmb.1992
Khaled M, Strodel B, Sayyed-Ahmad A (2023) Comparative molecular dynamics simulations of pathogenic and non-pathogenic huntingtin protein monomers and dimers. Front Mol Biosci 10:. https://doi.org/10.3389/fmolb.2023.1143353
Kim M (2013) Beta conformation of polyglutamine track revealed by a crystal structure of Huntingtin N-terminal region with insertion of three histidine residues. Prion 7:221–228. https://doi.org/10.4161/pri.23807
Kim M, Lee H-S, LaForet G, et al (1999) Mutant Huntingtin Expression in Clonal Striatal Cells: Dissociation of Inclusion Formation and Neuronal Survival by Caspase Inhibition. The Journal of Neuroscience 19:964–973. https://doi.org/10.1523/JNEUROSCI.19-03-00964.1999
Klein FAC, Zeder-Lutz G, Cousido-Siah A, et al (2013) Linear and extended: a common polyglutamine conformation recognized by the three antibodies MW1, 1C2 and 3B5H10. Hum Mol Genet 22:4215–4223. https://doi.org/10.1093/hmg/ddt273
Kokona B, Rosenthal ZP, Fairman R (2014) Role of the Coiled-Coil Structural Motif in Polyglutamine Aggregation. Biochemistry 53:6738–6746. https://doi.org/10.1021/bi500449a

Kotler SA, Tugarinov V, Schmidt T, et al (2019) Probing initial transient oligomerization events facilitating Huntingtin fibril nucleation at atomic resolution by relaxation-based NMR. Proceedings of the National Academy of Sciences 116:3562–3571. https://doi.org/10.1073/pnas.1821216116

Li L, Liu H, Dong P, et al (2016) Real-time imaging of Huntingtin aggregates diverting target search and gene transcription. Elife 5:. https://doi.org/10.7554/eLife.17056

Lindorff-Larsen K, Maragakis P, Piana S, et al (2012) Systematic Validation of Protein Force Fields against Experimental Data. PLoS One 7:e32131. https://doi.org/10.1371/journal.pone.0032131

Luo Y, Roux B (2010) Simulation of Osmotic Pressure in Concentrated Aqueous Salt Solutions. J Phys Chem Lett 1:183–189. https://doi.org/10.1021/jz900079w

Marsh JA, Singh VK, Jia Z, Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. Protein Science 15:2795–2804. https://doi.org/10.1110/ps.062465306

Matlahov I, van der Wel PC (2019) Conformational studies of pathogenic expanded polyglutamine protein deposits from Huntington's disease. Exp Biol Med 244:1584–1595. https://doi.org/10.1177/1535370219856620

Miller J, Arrasate M, Brooks E, et al (2011) Identifying polyglutamine protein species in situ that best predict neurodegeneration. Nat Chem Biol 7:925–934. https://doi.org/10.1038/nchembio.694

Mishra R, Hoop CL, Kodali R, et al (2012a) Serine Phosphorylation Suppresses Huntingtin Amyloid Accumulation by Altering Protein Aggregation Properties. J Mol Biol 424:1–14. https://doi.org/10.1016/j.jmb.2012.09.011

Mishra R, Jayaraman M, Roland BP, et al (2012b) Inhibiting the Nucleation of Amyloid Structure in a Huntingtin Fragment by Targeting α-Helix-Rich Oligomeric Intermediates. J Mol Biol 415:900–917. https://doi.org/10.1016/j.jmb.2011.12.011

Mittal J, Best RB (2010) Tackling Force-Field Bias in Protein Folding Simulations: Folding of Villin HP35 and Pin WW Domains in Explicit Water. Biophys J 99:L26–L28. https://doi.org/10.1016/j.bpj.2010.05.005

Mohanty P, Rizuan A, Kim YC, et al (2024) A complex network of interdomain interactions underlies the conformational ensemble of monomeric <scp>TDP</scp> -43 and modulates its phase behavior. Protein Science 33:. https://doi.org/10.1002/pro.4891

Mohanty P, Shenoy J, Rizuan A, et al (2023) A synergy between site-specific and transient interactions drives the phase separation of a disordered, low-complexity domain. Proceedings of the National Academy of Sciences 120:. https://doi.org/10.1073/pnas.2305625120

Moldovean SN, Chiş V (2020) Molecular Dynamics Simulations Applied to Structural and Dynamical Transitions of the Huntingtin Protein: A Review. ACS Chem Neurosci 11:105–120. https://doi.org/10.1021/acschemneuro.9b00561

Nagai Y, Inui T, Popiel HA, et al (2007) A toxic monomeric conformer of the polyglutamine protein. Nat Struct Mol Biol 14:332–340. https://doi.org/10.1038/nsmb1215

Nucifora LG, Burke KA, Feng X, et al (2012) Identification of Novel Potentially Toxic Oligomers Formed in Vitro from Mammalian-derived Expanded huntingtin Exon-1 Protein. Journal of Biological Chemistry 287:16017–16028. https://doi.org/10.1074/jbc.M111.252577

Owens GE, New DM, West AP, Bjorkman PJ (2015) Anti-PolyQ Antibodies Recognize a Short PolyQ Stretch in Both Normal and Mutant Huntingtin Exon 1. J Mol Biol 427:2507–2519. https://doi.org/10.1016/j.jmb.2015.05.023

Ozenne V, Bauer F, Salmon L, et al (2012) Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. Bioinformatics 28:1463–1470. https://doi.org/10.1093/bioinformatics/bts172

Páll S, Zhmurov A, Bauer P, et al (2020) Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. J Chem Phys 153:134110. https://doi.org/10.1063/5.0018516

Pandey NK, Isas JM, Rawat A, et al (2018) The 17-residue-long N terminus in huntingtin controls stepwise aggregation in solution and on membranes via different mechanisms. Journal of Biological Chemistry 293:2597–2605. https://doi.org/10.1074/jbc.M117.813667

Peters-Libeu C, Miller J, Rutenber E, et al (2012a) Disease-Associated Polyglutamine Stretches in Monomeric Huntingtin Adopt a Compact Structure. J Mol Biol 421:587–600. https://doi.org/10.1016/j.jmb.2012.01.034
Peters-Libeu C, Miller J, Rutenber E, et al (2012b) Disease-Associated Polyglutamine Stretches in Monomeric Huntingtin Adopt a Compact Structure. J Mol Biol 421:587–600. https://doi.org/10.1016/j.jmb.2012.01.034
Phan TM, Schmit JD (2022) Conformational entropy limits the transition from nucleation to elongation in amyloid aggregation. Biophys J 121:2931–2939. https://doi.org/10.1016/j.bpj.2022.06.031
Phan TTM, Schmit JD (2020) Thermodynamics of Huntingtin Aggregation. Biophys J 118:2989–2996. https://doi.org/10.1016/j.bpj.2020.05.013
Rauscher S, Gapsys V, Gajda MJ, et al (2015) Structural ensembles of intrinsically disordered proteins depend strongly on force field: A comparison to experiment. J Chem Theory Comput 11:5513–5524. https://doi.org/10.1021/acs.jctc.5b00736
Robustelli P, Piana S, Shaw DE (2018) Developing a molecular dynamics force field for both folded and disordered protein states. Proceedings of the National Academy of Sciences 115:. https://doi.org/10.1073/pnas.1800690115
Ross CA (2002) Polyglutamine Pathogenesis. Neuron 35:819–822. https://doi.org/10.1016/S0896-6273(02)00872-3
Ross CA, Aylward EH, Wild EJ, et al (2014) Huntington disease: natural history, biomarkers and prospects for therapeutics. Nat Rev Neurol 10:204–216. https://doi.org/10.1038/nrneurol.2014.24
Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341. https://doi.org/10.1016/0021-9991(77)90098-5
Sánchez I, Mahlke C, Yuan J (2003) Pivotal role of oligomerization in expanded polyglutamine neurodegenerative disorders. Nature 421:373–379. https://doi.org/10.1038/nature01301
Shao J, Diamond MI (2007) Polyglutamine diseases: emerging concepts in pathogenesis and therapy. Hum Mol Genet 16:R115–R123. https://doi.org/10.1093/hmg/ddm213
Shen K, Calamini B, Fauerbach JA, et al (2016) Control of the structural landscape and neuronal proteotoxicity of mutant Huntingtin by domains flanking the polyQ tract. Elife 5:. https://doi.org/10.7554/eLife.18065
Sieradzan KA, Mechan AO, Jones L, et al (1999) Huntington's Disease Intranuclear Inclusions Contain Truncated, Ubiquitinated Huntingtin Protein. Exp Neurol 156:92–99. https://doi.org/10.1006/exnr.1998.7005
Sivanandam VN, Jayaraman M, Hoop CL, et al (2011) The Aggregation-Enhancing Huntingtin N-Terminus Is Helical in Amyloid Fibrils. J Am Chem Soc 133:4558–4566. https://doi.org/10.1021/ja110715f
Tabrizi SJ, Flower MD, Ross CA, Wild EJ (2020) Huntington disease: new insights into molecular pathogenesis and therapeutic opportunities. Nat Rev Neurol 16:529–546. https://doi.org/10.1038/s41582-020-0389-4
Tam S, Spiess C, Auyeung W, et al (2009) The chaperonin TRiC blocks a huntingtin sequence element that promotes the conformational switch to aggregation. Nat Struct Mol Biol 16:1279–1285. https://doi.org/10.1038/nsmb.1700
Tang WS, Fawzi NL, Mittal J (2020) Refining All-Atom Protein Force Fields for Polar-Rich, Prion-like, Low-Complexity Intrinsically Disordered Proteins. J Phys Chem B 124:9505–9512. https://doi.org/10.1021/acs.jpcb.0c07545
Thakur AK, Jayaraman M, Mishra R, et al (2009) Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism. Nat Struct Mol Biol 16:380–389. https://doi.org/10.1038/nsmb.1570
Torricella F, Tugarinov V, Clore GM (2024) Nucleation of Huntingtin Aggregation Proceeds via Conformational Conversion of Pre-Formed, Sparsely-Populated Tetramers. Advanced Science. https://doi.org/10.1002/advs.202309217

Urbanek A, Popovic M, Morató A, et al (2020a) Flanking Regions Determine the Structure of the Poly-Glutamine in Huntingtin through Mechanisms Common among Glutamine-Rich Human Proteins. Structure 28:733-746.e5. https://doi.org/10.1016/j.str.2020.04.008
Urbanek A, Popovic M, Morató A, et al (2020b) Flanking Regions Determine the Structure of the Poly-Glutamine in Huntingtin through Mechanisms Common among Glutamine-Rich Human Proteins. Structure 28:733-746.e5. https://doi.org/10.1016/j.str.2020.04.008
van der Wel PCA (2024) Solid-state nuclear magnetic resonance in the structural study of polyglutamine aggregation. Biochem Soc Trans. https://doi.org/10.1042/BST20230731
Warner JB, Ruff KM, Tan PS, et al (2017) Monomeric Huntingtin Exon 1 Has Similar Overall Structural Features for Wild-Type and Pathological Polyglutamine Lengths. J Am Chem Soc 139:14456–14469. https://doi.org/10.1021/jacs.7b06659
Webb B, Sali A (2016) Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinformatics 54:. https://doi.org/10.1002/cpbi.3
Williamson TE, Vitalis A, Crick SL, Pappu R V. (2010) Modulation of Polyglutamine Conformations and Dimer Formation by the N-Terminus of Huntingtin. J Mol Biol 396:1295–1309. https://doi.org/10.1016/j.jmb.2009.12.017
Yamamoto A, Lucas JJ, Hen R (2000) Reversal of Neuropathology and Motor Dysfunction in a Conditional Model of Huntington's Disease. Cell 101:57–66. https://doi.org/10.1016/S0092-8674(00)80623-6
Yang H, Yang S, Jing L, et al (2020) Truncation of mutant huntingtin in knock-in mice demonstrates exon1 huntingtin is a key pathogenic form. Nat Commun 11:2582. https://doi.org/10.1038/s41467-020-16318-1
Zerze GH, Zheng W, Best RB, Mittal J (2019) Evolution of All-Atom Protein Force Fields to Improve Local and Global Properties. J Phys Chem Lett 10:2227–2234. https://doi.org/10.1021/acs.jpclett.9b00850
Zhang Z, Liu X, Yan K, et al (2019) Unified Efficient Thermostat Scheme for the Canonical Ensemble with Holonomic or Isokinetic Constraints via Molecular Dynamics. J Phys Chem A 123:6056–6079. https://doi.org/10.1021/acs.jpca.9b02771
Zheng W, Dignon GL, Jovic N, et al (2020) Molecular Details of Protein Condensates Probed by Microsecond Long Atomistic Simulations. J Phys Chem B 124:11671–11679. https://doi.org/10.1021/acs.jpcb.0c10489

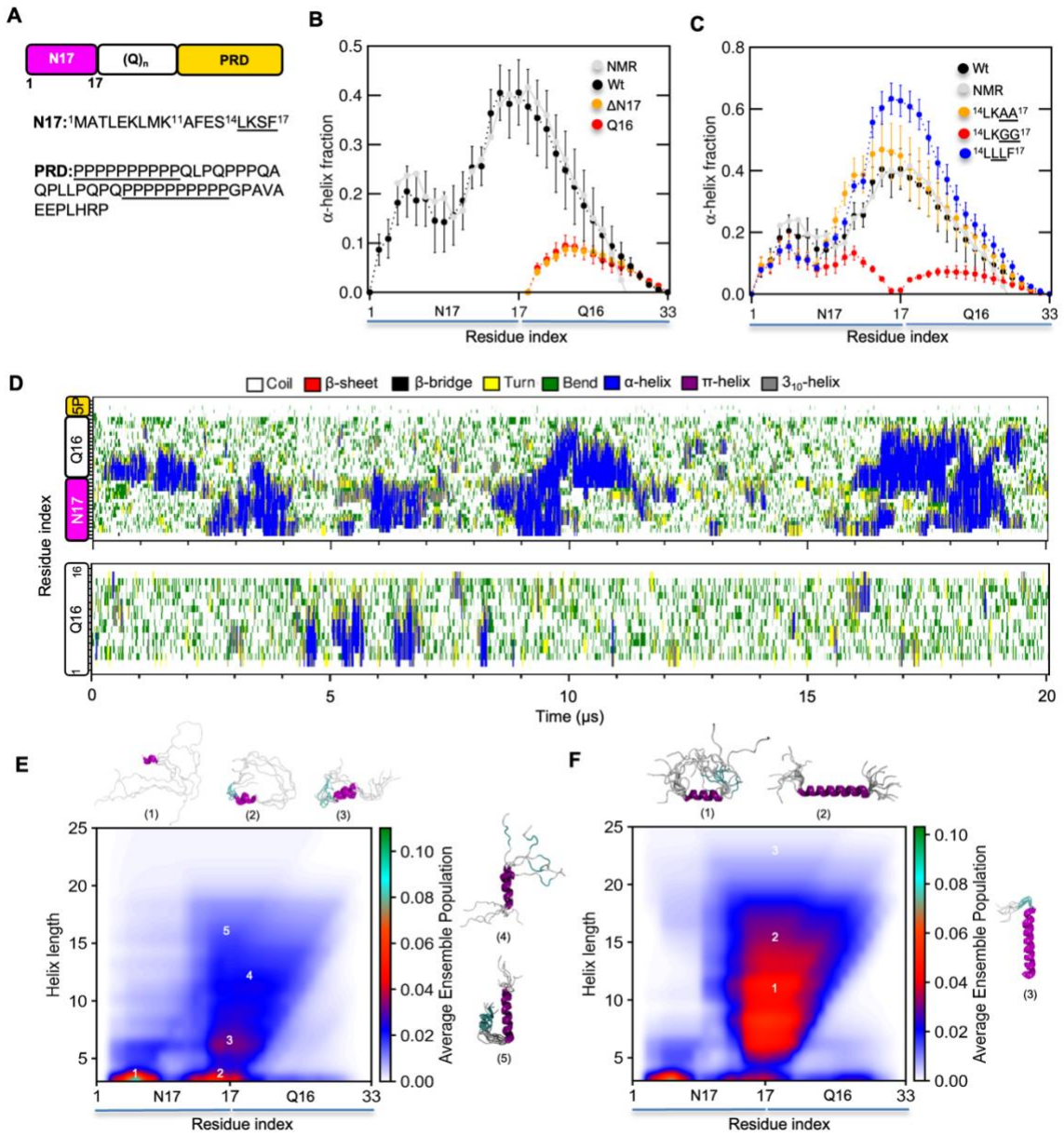

Figure 1. Conformational ensemble of N17-Q_{16}. **A.** Schematic showing the organization of Httex1 into three regions - N17, polyQ and PRD (top), and the corresponding sequences of N17/PRD domains (bottom). The length of polyQ region in the wild-type Huntingtin protein is ~23 residues. **B.** Comparison of per-residue α-helix fractions for N17-Q_{16} wild-type and its deletion constructs with respect to NMR. Helical fractions were computed using the DSSP algorithm and compared to SSP scores for the helical propensity (grey) computed from experimental ^{13}C chemical shifts. **C.** Comparison of per-residue α-helix fractions for N17-Q_{16} wild-type and ^{14}LKSF^{17} mutants with respect to NMR. **D.** Secondary structure variation as a function of time in a representative N17-Q_{16} (top) and Q_{16} (bottom) trajectory. **E.** SS-map of N17-Q_{16} wild-type computed from an aggregate trajectory (~63 μs) indicating the probability of various helical length across N17 and polyQ regions. Representative α-helical conformations corresponding to numbered regions of the SS-map are shown in cartoon representation. **F.** Same as in E for ^{14}LLLF^{17} mutant.

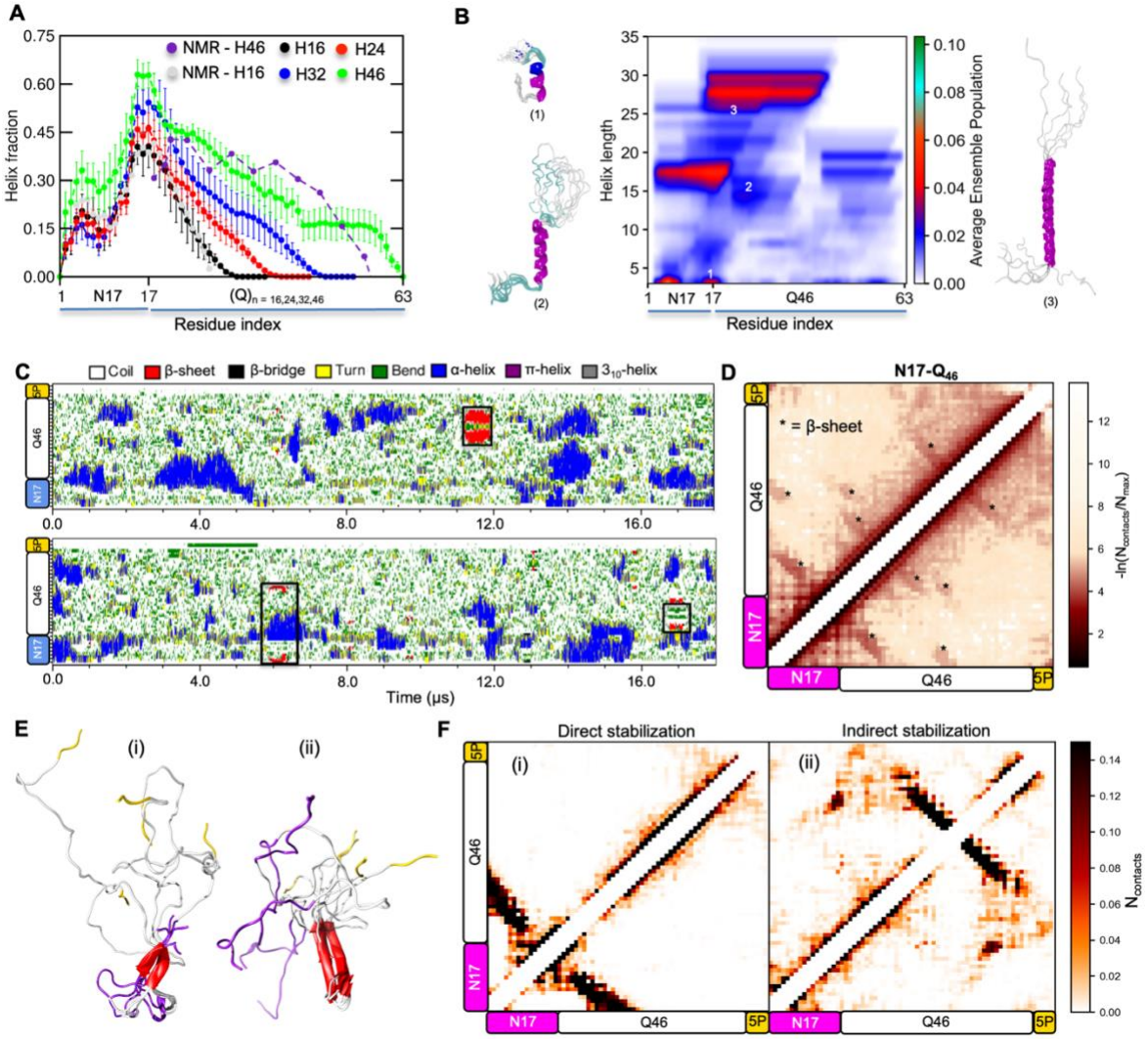

Figure 2. Conformational ensemble of N17-Q_{46}. A. Comparison of per-residue α-helix fractions for N17-polyQ with increasing polyQ length with against NMR-derived helical propensities. **B.** SS-map of N17-Q_{46} wild-type (left) from the PT-WTE aggregate trajectory (750 ns) indicating the probability of helices formed of varying lengths across N17 and polyQ regions. Representative helical conformations corresponding to numbered regions of the SS-map are shown in cartoon representation. **C.** Secondary structure variation as a function of time for two representative N17-Q_{46} trajectories showing the formation of two-stranded β-sheet structures (black boxes). **D.** Two-dimensional intramolecular contact maps calculated over six N17-Q_{46} independent trajectories indicating the low population of β-conformations (marked as *) relative to α-helices. **E**. Representative β-sheet structures (5 in total) from the trajectory periods highlighted in C (black boxes) involving either direct (i) and indirect (ii) contacts between the N17 and Q_{46} domains. The coloring scheme for the structures is as follows: N17 - purple, Q46 - white and 5P - Gold, β-strand - red. **F.** Contact maps calculated for the two β-sheet structural ensembles shown in **E**.

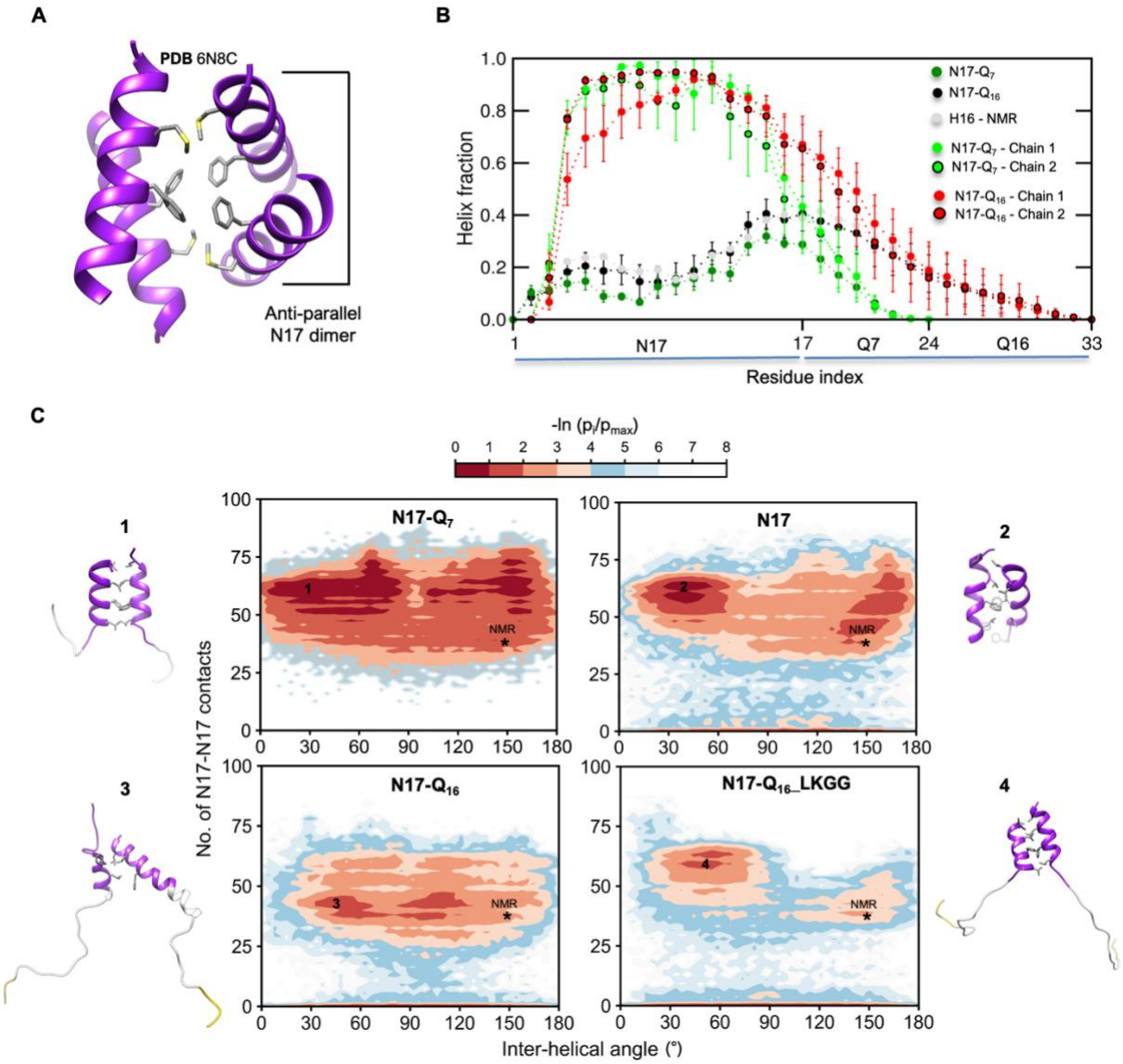

Figure 3. Competition between structural transformation and interdomain interactions shape the dimerization landscape of N17-polyQ. A. NMR structure of Httex1-Q7 (H7) tetramer showing interactions (sidechains shown as sticks) in the hydrophobic core. The anti-parallel dimer was extracted from the structure for MD simulations. **B**. Comparison of per-residue α-helix fractions for H7/16 chains from dimer simulations against monomer simulations and NMR. The comparison indicates significantly elevated helicity of N17 region in H7/16 dimers compared to their respective monomers. **C.** Two-dimensional PMF (potential of mean force) plots as a function of two order parameters - (i) total number of intermolecular N17 contacts and (ii) the inter-helical angle (°) between N17 domain, which characterize the association free energy landscape for four N17 dimer variants. The plots shows the combined data derived from 6 independent dimer trajectories (~2.3 μs) for each variants. The position of NMR - N17 dimer (initial structure) is marked on the PMF plots as (*). Representative dimer MD structures from the basins marked (1, 2) in PMF plots are shown on the left with interfacial hydrophobic residues shown as grey sticks.

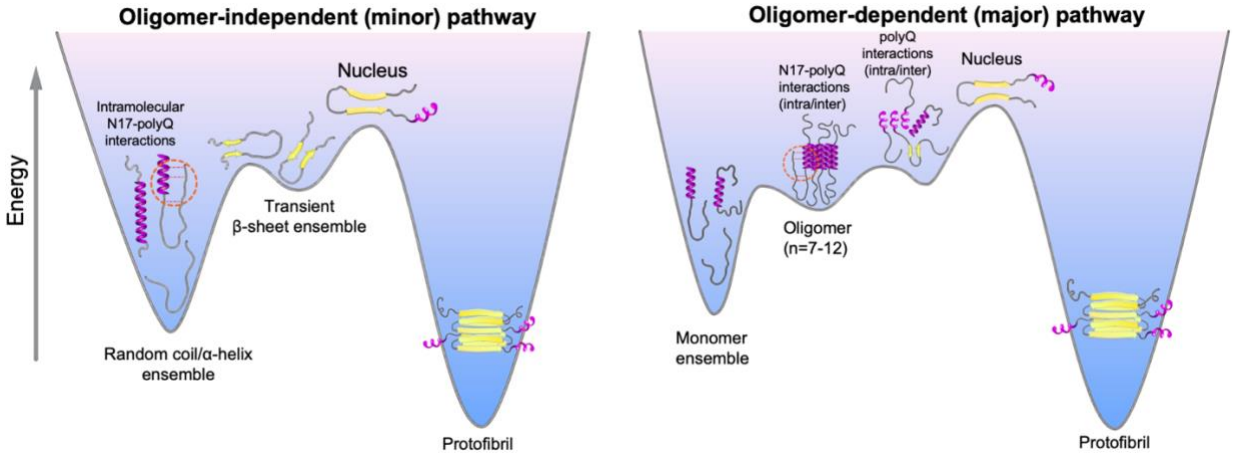

Figure 4. polyQ expansion promotes "domain crosstalk" which modulates the structural ensemble and oligomerization landscape of Httex1 to favor fibrillar aggregation. Schematic illustrating the effect of polyQ length expansion on the conformation and oligomerization of Httex1. polyQ expansion promotes inter-domain N17-polyQ interactions. These interactions can (i) induce the emergence of transient β-sheet conformers in monomers which may fold to form the critical nucleus and initiate protofibril formation (oligomer-independent pathway), and (ii) counteract stabilization of N17-mediated oligomers to favor polyQ-polyQ interactions which lead to nucleation and fibrillation. The nucleus size is shown as a monomer (n=1) and is representative of mutant Httex1 (≥36).