

Unraveling the Molecular Complexity of N-Terminus Huntingtin Oligomers: Insights into Polymorphic Structures

Neha Nanajkar,[§] Abhilash Sahoo,[§] and Silvina Matysiak*



Cite This: <https://doi.org/10.1021/acs.jpcb.4c03274>



Read Online

ACCESS |



Metrics & More

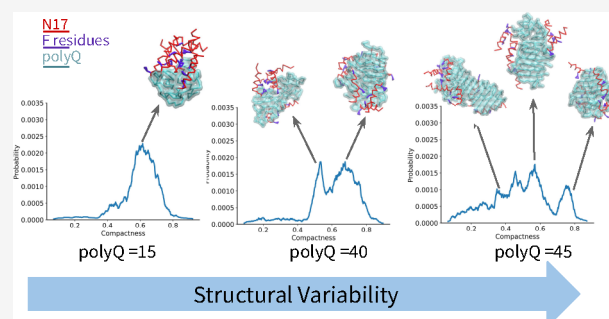


Article Recommendations



Supporting Information

ABSTRACT: Huntington's disease (HD) is a fatal neurodegenerative disorder resulting from an abnormal expansion of polyglutamine (polyQ) repeats in the N-terminus of the huntingtin protein. When the polyQ tract surpasses 35 repeats, the mutated protein undergoes misfolding, culminating in the formation of intracellular aggregates. Research in mouse models suggests that HD pathogenesis involves the aggregation of N-terminal fragments of the huntingtin protein (htt). These early oligomeric assemblies of htt, exhibiting diverse characteristics during aggregation, are implicated as potential toxic entities in HD. However, a consensus on their specific structures remains elusive. Understanding the heterogeneous nature of htt oligomers provides crucial insights into disease mechanisms, emphasizing the need to identify various oligomeric conformations as potential therapeutic targets. Employing coarse-grained molecular dynamics, our study aims to elucidate the mechanisms governing the aggregation process and resultant aggregate architectures of htt. The polyQ tract within htt is flanked by two regions: an N-terminal domain (N17) and a short C-terminal proline-rich segment. We conducted self-assembly simulations involving five distinct N17 + polyQ systems with polyQ lengths ranging from 7 to 45, utilizing the ProMPT force field. Prolongation of the polyQ domain correlates with an increase in β -sheet-rich structures. Longer polyQ lengths favor intramolecular β -sheets over intermolecular interactions due to the folding of the elongated polyQ domain into hairpin-rich conformations. Importantly, variations in polyQ length significantly influence resulting oligomeric structures. Shorter polyQ domains lead to N17 domain aggregation, forming a hydrophobic core, while longer polyQ lengths introduce a competition between N17 hydrophobic interactions and polyQ polar interactions, resulting in densely packed polyQ cores with outwardly distributed N17 domains. Additionally, at extended polyQ lengths, we observe distinct oligomeric conformations with varying degrees of N17 bundling. These findings can help explain the toxic gain-of-function that htt with expanded polyQ acquires.



INTRODUCTION

The huntingtin protein is a complex molecule with multiple domains, playing vital roles in various cellular processes such as axonal transport, signaling pathways, and protein interactions.^{1–3} Notably, abnormalities in the glutamine repeat region (polyQ) located at the N-terminus of the huntingtin protein (htt) are associated with Huntington's Disease (HD), a progressive neurodegenerative condition characterized by deteriorating physical and cognitive functions. In individuals without the disease, polyQ length ranges between 6 and 34 (htt^{wt}), but clinical pathogenicity occurs when the length exceeds a threshold of 35Q residues (htt^{mutant}).⁴ These elongated polyQ segments undergo misfolding, leading to the formation of intracellular aggregates, which are hallmark features of HD pathology. The length of the expanded polyQ domain correlates with earlier disease onset and increased disease severity, highlighting its significance in the progression of HD.

While Huntington's Disease is among 9 recognized polyQ-based disorders, our understanding of the aggregation process

and its underlying mechanisms remains incomplete.⁵ PolyQ presents unique challenges for experimental study due to its rapid and length-dependent aggregation kinetics, conformational flexibility, and insolubility.^{6,7} Despite these difficulties, researchers have made progress in uncovering several structural characteristics of polyQ. PolyQ can self-assemble into polar zippers, where glutamine side chains stabilize β -sheets.⁸ Additionally, evidence suggests the presence of coiled-coil structures, α -helices, and β -helical conformations within the polyQ domain,^{9–11} illustrating its flexible nature. In solution, polyQ tends to be insoluble and exhibits a preference for self-interaction, resulting in collapsed and more globular structures

Received: May 17, 2024

Revised: July 24, 2024

Accepted: July 25, 2024

as polyQ length increases.^{12,13} The structural and dynamic properties of polyQ are significantly influenced by its flanking sequences, particularly the N-terminal N17 domain and the C-terminal proline-rich domain (PRD).^{14–16} Studies indicate that N17 may adopt a coiled-coil state or an amphipathic helix, promoting htt aggregation.^{17,18} The helical nature of N17 can extend into the polyQ domain, initiating a region of helicity before polyQ adopts an extended structure.¹⁴

Numerous aggregation models have been proposed, yet a consensus regarding the structure of aggregates remains elusive. Tryptophan fluorescence experiments on htt with varying polyQ lengths suggested that aggregates comprise an amyloid core involving both the polyQ and N17 domains.¹⁹ Similarly, other studies propose ternary and quaternary interactions between N17 and polyQ at the fibril core, with the PRD extending from it.^{20,21} Conversely, ssNMR analysis of the N17Q₃₀P₁₀K₂ fragment indicated that the amyloid core solely comprises the polyQ domain, while the N17 domain resides peripherally, more exposed to solvent.²² Chen and Wolynes' implicit solvent htt model for N17Q₂₀ suggests N17 domain bundling, while the polyQ region forms trailing interpeptide β -sheets.²³ Recent evidence suggests that smaller, polymorphic prefibrillar oligomers, rather than mature fibrils, are the neurotoxic species in neurodegenerative disorders. However, characterizing them experimentally and structurally remains challenging due to their transient nature and the coexistence of monomeric, oligomeric, and fibrillar species.

A thorough comprehension of the aggregation process and oligomeric structures is imperative for the development of therapeutics targeting proteins. While techniques like cryo-EM and NMR have been instrumental in elucidating the morphology and polymorphisms of htt fibrils, molecular details of the oligomers remain elusive.^{20,22,24} Molecular dynamics simulations offer a promising avenue to bridge these knowledge gaps, providing insights into the underlying molecular mechanisms.

Biological processes such as protein folding and aggregation occur over time scales inaccessible to unbiased all-atom molecular dynamics simulations. Hence, coarse-grained molecular dynamics, a reduced-resolution approach, becomes indispensable. By grouping local atoms into coarse-grained beads, this method yields smoother free energy landscapes, enabling simulations of larger systems over biologically relevant time scales.

In this study, we utilized our in-house developed ProMPT coarse-grained force field,²⁵ featuring free-energy-based interaction parametrization and explicit dipoles to represent polar regions of the protein. Incorporating dipoles enables the force field to simulate protein structural transitions in response to the changes in local environment.

Our investigation focused on examining the monomeric and oligomeric states of both wild-type and mutant lengths of htt using coarse-grained molecular dynamics. We constructed five N17 + polyQ systems, comprising three wild-type (7Q, 15Q, and 35Q) and two mutant (40Q and 45Q) systems. Our results revealed a propensity for intramolecular β -sheets as the polyQ length increased. We observed bundling of the N-terminal N17 domain in aggregate systems, which diminished with longer polyQ lengths. Furthermore, we characterized the influence of polyQ length on oligomeric morphology and the multiplicity of distinct oligomeric conformations. These variations in oligomeric morphology likely contribute to the observed polymorphism in fibrillar htt structures.

Coarse-grained molecular dynamics were used to examine the aggregation process and oligomer morphologies of both htt^{wt} and htt^{mutant}. We observed bundling of the N17 domain in the aggregate systems, which reduces upon increase in polyQ length. At mutant polyQ lengths, there is a heightened variability in N17 bundling with N17 locating at the periphery of the aggregate. In contrast, N17 is located in the center of the oligomer for htt^{wt}. In addition, distinct oligomeric conformations are observed for htt^{mutant} in contrast to htt^{wt}. We suspect these variations in oligomeric morphology may be a structural basis for the polymorphism observed in fibrillar htt structures.

Our observations revealed the bundling of the N17 domain in aggregate systems, a phenomenon that diminishes with increasing polyQ length. Notably, at mutant polyQ lengths, we observed heightened variability in N17 bundling, with N17 predominantly located at the periphery of the aggregate. In contrast, in htt^{wt} systems, N17 tends to occupy the oligomer core. Additionally, we identified distinct oligomeric conformations in htt^{mutant} in contrast to htt^{wt}. These variations in oligomeric morphology likely serve as a structural basis for the polymorphism observed in fibrillar htt structures.

METHODS

All simulations were carried out on GROMACS 2019.4 simulator,²⁶ visualized on VMD 1.9.4,²⁷ and analyzed using in-house scripts in MDAnalysis version 1.1.1.²⁸ Reweighting of simulation data was accomplished using the Multiscale Bennett Acceptance Ratio method, implemented in the pyMBAR package.²⁹

Force Field. We employed Protein Model with Polarizability and Transferability (ProMPT) for all simulations, as detailed in our previous work.²⁵ ProMPT is a coarse grained mini-protein force field based on the popular Martini model³⁰ that features free-energy based interaction parametrization and follows a 4:1 mapping. To account for orientation changes of atoms beneath polar CG beads, we incorporate flexible dipoles formed by oppositely charged dummy particles within each CG bead. These dipoles interact through Coulombic potentials, introducing structural polarization in the model. The addition of dipoles allows the force field to generate protein structural transitions in response to the changing local environment, making it well-equipped to capture changes in secondary structure and aggregation processes. ProMPT has undergone validation on various small proteins exhibiting diverse structural characteristics.³¹ In conjunction, we use the polarizable Martini water model, which maps 4 water molecules into one water bead.³² The coarse-grained representation of the N17 domain is illustrated in SI 1, Figure S1.

Simulation Parameters. In our simulations, we perform simulations on the N-terminal N17 domain with varying lengths of polyQ. The sequence used for the N17 domain is MATLEKLMKAFESLKSF, and this is succeeded by a polyQ stretch. The polyQ region was modeled to have 7, 15, 35, 40, and 45 Q residues. To characterize the aggregation process and oligomeric structure, six peptides were placed in a 12 nm box, initially in an unfolded state. In this model, a helical dihedral potential with competing minima was applied to the N17 backbone beads, utilizing a force constant of 10 kcal/mol (see SI 1, Figure S2.). Experimental evidence suggests that N17 adopts an unstructured or coil-like structure at low concentrations,¹⁹ while at high concentrations, it forms an

amphipathic helix.³³ No dihedral potential was applied to the polyQ domain.

Restraints were applied to hold the peptides in place while the system underwent steepest descent energy minimization for 5000 steps, followed by a 5000 step NPT equilibration. Sodium and chloride ions, sourced from the MARTINI model, were introduced into the simulation box to achieve an effective salt concentration of 125 mM, mirroring typical experimental conditions for htt studies. After the addition of salt ions, a second round of energy minimization and equilibration was performed, consisting of 5000 steps each. Upon release of the constraints, a canonical production run extended for 350 ns, employing a time step of 10 fs.

The peptide systems were simulated at a range of temperatures spanning from 380 to 500 K, with increments of 20 K, resulting in 7 temperatures for each peptide system. Running simulations at higher temperatures offers the advantage of enabling the system to sample infrequent, higher energy configurations that may be inaccessible at lower temperatures. Varying the temperature enables a more thorough exploration of the free energy landscape, a common technique in molecular dynamics simulations.^{34–36}

It is important to note that the coarse-grained temperatures employed in this study are not directly comparable to real or experimental temperatures. Molecular simulation methods, whether all-atom or coarse-grained, are approximated versions of the systems they seek to model. While these are necessary simplifications, they do affect the accuracy of the system. Temperature can be expressed as a measure of the kinetic energy of a system, and the kinetic energy of a system depends on the number of particles in it. However, since the number of particles in a coarse-grained system is a simplification, it cannot be compared with real life temperature, rather should be viewed as a descriptor of the system. Since these simulations were conducted under the NVT ensemble, the solvent density remained consistent across all temperatures. The Nosé–Hoover thermostat, with a time constant of 1 ps, was utilized to maintain fixed temperatures. A relative dielectric constant of 2.5 was applied, and electrostatic calculations were performed using Particle Mesh Ewald. Constraints were evaluated using the LINCS algorithm, and the neighbor list was updated every 10 steps.

Analysis Methods. To determine the probability of initial interactions between peptides, we focused on the first 90 ns of simulation time, corresponding to the period during which oligomer formation occurred. Interactions between peptides were categorized as N17–N17, Q–Q, or N17–Q driven. The averages and error bars were calculated by varying the distance cutoffs (6–8 Å).

For subsequent analyses, we analyzed the last 260 ns of the simulations. This ensured that structural characterization was performed on segments of the trajectory where the complete oligomer had formed. When studying β -sheet structures using all atom simulations, H-bonds between the chains of the β -sheet are the main focus. However, due to the coarse-graining of our protein model, we cannot measure H-bonds directly. Instead, we look for dipole–dipole interactions between the positive and negative dummy charges of the protein backbone bead. Positive and negative dummy bead interactions across β -sheet chains are a dipole interaction, similar to the hydrogen bonds that are responsible for β -sheet structure. Thus, interpeptide sheet contacts were quantified based on the number of contiguous contacts between the polyQ domains of

different peptides, disregarding interactions within a single peptide. Intrapeptide sheet contacts referred to nonhelical contiguous contacts within the polyQ domain of a single peptide. In both inter- and intrapeptide sheet contacts, a minimum of 4 contiguous contacts was required to be classified as a sheet. The number of interpeptide and intrapeptide contacts was then normalized by the maximum possible contacts in each system.

The hydrophobic solvent accessible surface area (hSASA) was computed using the built-in *gmx sasa* tool from GROMACS. Specifically, hydrophobic residues F, L, M, and V were selected for analysis.

To determine the number of interacting N17 domains (N_{int}), a distance matrix was constructed using the backbone and side chain beads of the N17 domain. Residues within 7 Å of each other were defined as being in contact. In order to filter out transient interactions of N17 domains, interactions were considered significant if there were at least 4 or more interactions between the domains.

To examine polymorphism in htt, trajectories were filtered based on the value of N_{int} . Compactness was calculated as the ratio of the smallest and largest eigenvalues of the moment of inertia tensors, inspired by the implementation in ATRANET.³⁷

RESULTS AND DISCUSSION

Aggregation Pathway and Structural Characteristics.

To elucidate the mechanisms driving aggregation, we assessed the likelihood of the initial interaction between the N17 and polyQ domains of the peptides (Figure 1). Previous studies

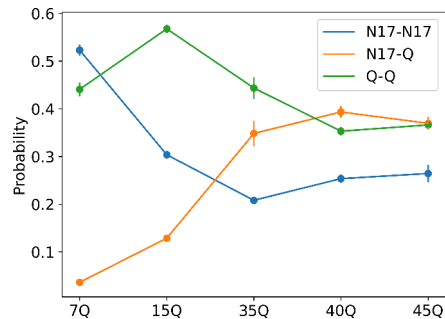


Figure 1. Fraction of interpeptide interactions occurring through N17–N17, N17–Q and Q–Q interactions.

have suggested that the aggregation of polyQ is facilitated by the self-associating properties of N17. Consistently, in the 7Q system, initial interpeptide interactions predominantly involve N17–N17 or Q–Q driven first interactions, with both domains preferring self-interaction. As the polyQ length increases, we note decreases in N17–N17 and Q–Q interactions, and a concurrent increase in interdomain interactions.

Next, we delved into the structural characteristics of the peptide systems. Detailed information on monomeric structures is available in SI 2. Assessment of secondary structure characteristics in the oligomers was conducted by examining the helical fraction and β -sheet fraction. We observed that the N17 domain predominantly adopts helical conformations, with a helical fraction of 0.9, while the polyQ domain exhibits a helical fraction of approximately 0.2 across all polyQ lengths, as depicted in SI 1, Figure S3. In terms of β -

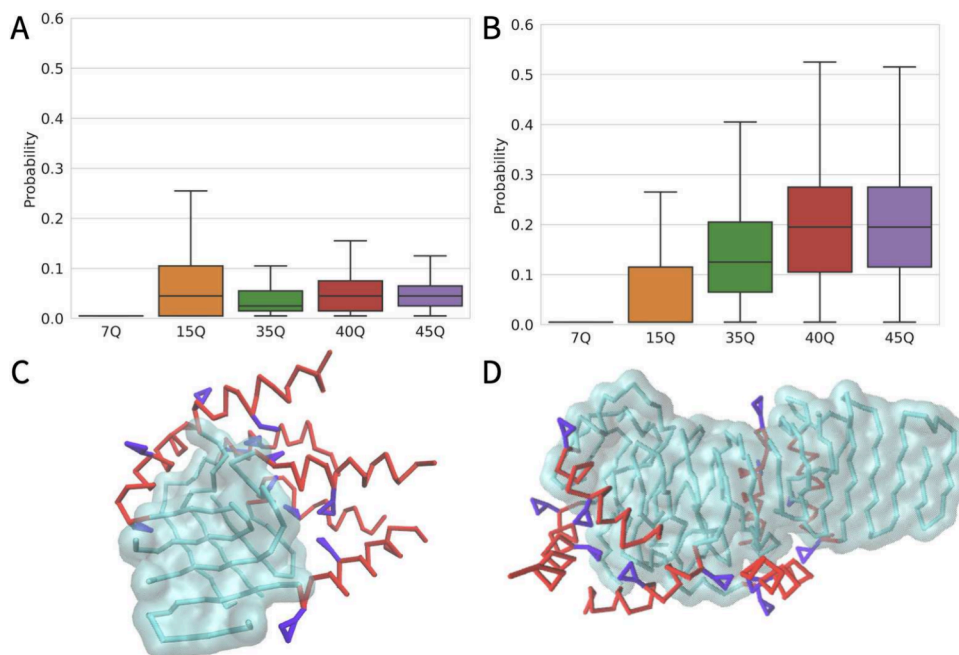


Figure 2. β -Sheet preferences of the polyQ domain, reweighted to 380 K. (A) Box plot of the distribution of interpeptide β -sheet fractions. (B) Box plot distribution of intrapeptide β -sheet fraction. Boxes denote the quartiles and the whiskers denote the ends of the distribution. (C) Snapshot of interpeptide β -sheets observed in the 15Q system. (D) Snapshot of intrapeptide β -sheets observed in the 45Q system. N17, polyQ domain, and F side chain residues are colored in red, cyan, and violet, respectively.

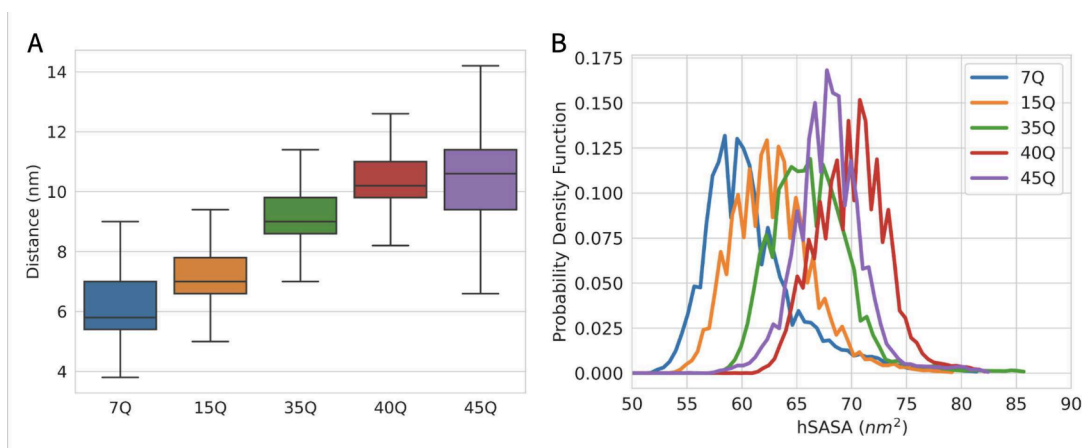


Figure 3. Gradual loss of the hydrophobic core with increasing polyQ length occurs due to the repositioning of N17 to the periphery of the oligomer. (A) Box plot of the distance between the center of mass of F side chain beads from the center of mass of the whole aggregate. Boxes denote the quartiles and the whiskers denote the ends of the distribution. (B) Hydrophobic solvent accessible surface area of the oligomers of all peptide systems.

sheet formation (Figure 2), minimal to no β -sheet formation was observed in the 7Q system due to the short length of the polyQ domain. In the 15Q system, interpeptide β -sheets were favored over intrapeptide β -sheets. A snapshot of the 15Q system (Figure 2C) illustrates the alignment of polyQ domains between peptides, forming interpeptide β -sheets, albeit with less frequent intrapeptide β -sheet formation. However, as the polyQ length increases (polyQ \geq 35), the preference shifts toward intrapeptide β -sheets, as demonstrated in a snapshot of the 45Q system (Figure 2D).

These observations are consistent with the findings of Yushchenko et al.,³⁸ who utilized FTIR spectroscopy to analyze htt fragments with different polyQ lengths, and with the research by Bugg et al.,²⁰ which emphasized the prevalence of antiparallel β -sheets. Moreover, various biophysical

investigations have highlighted the polyQ domain's conformational flexibility, demonstrating its ability to adopt helices, coiled coils, and β -sheet-rich structures.^{9–11,39}

It is noteworthy that our simulations did not indicate the presence of coiled coils, contrary to the findings of Chen and Wolynes,²³ who observed extensive interpeptide β -sheets in simulations involving N17Q₂₀ and identified coiled-coil structures in the N17 domain. The snapshots provided in Figure 2 illustrate that larger polyQ domains tend to collapse upon themselves, resulting in an increase in intrapeptide β -sheets and the formation of hairpin-like structures. This preference for more collapsed configurations is evident not only in the oligomeric systems but also in the monomeric systems (see SI 2). Several studies have indicated that longer polyQ domains tend to adopt more collapsed, globular

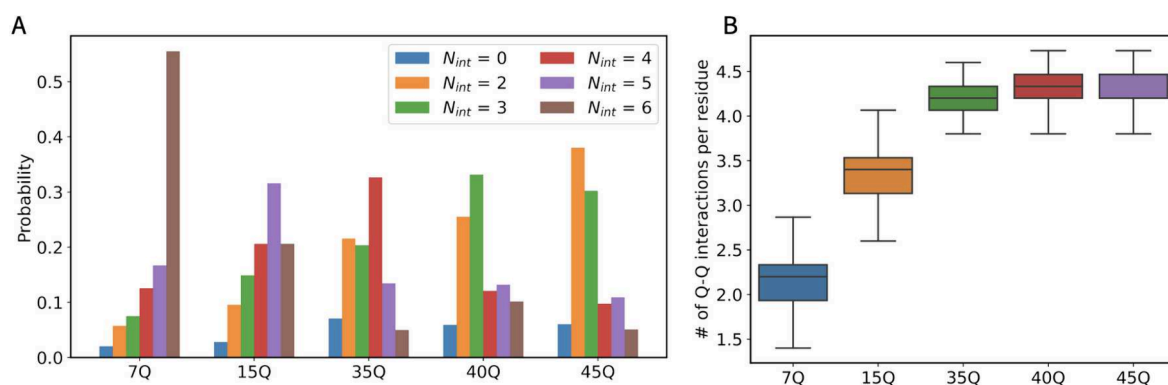


Figure 4. Oligomer morphology depends on polyQ length. (A) Probabilities of the number of interacting N17 domains (N_{int}) in each protein system. (B) Box plot of the average number of Q–Q interactions per residue for each peptide system. Boxes denote the quartiles and the whiskers denote the ends of the distribution.

structures.^{13,40} Recently, Kang et al. demonstrated this phenomenon with varying lengths of polyQ in htt.¹³

Competition between Hydrophobic and Polar Interactions. Upon visual examination of the oligomers, we observed the presence of a hydrophobic core in systems with shorter polyQ lengths. However, as the polyQ length increased, hydrophobic interactions diminished, resulting in the loss of the hydrophobic core. To quantify this trend, we plotted the average distance of phenylalanine (F) beads from the oligomer's center of mass (Figure 3A). As anticipated, the distance between F residues and the oligomer's center of mass increased with longer polyQ lengths, indicating that F and other hydrophobic residues in the N17 domain were pushed toward the periphery of the oligomer, while the core comprised predominantly the polyQ domain.

Furthermore, to assess the environment surrounding the N17 domain, we examined the hydrophobic solvent accessible surface area (hSASA) (Figure 3B). A gradual increase in hSASA was observed with longer polyQ lengths, suggesting that as the polyQ length increased, the hydrophobic residues in the N17 domain became progressively more exposed to solvent.

The repositioning of N17 to the periphery of the aggregate can be attributed to a competition between hydrophobic interactions within the N17 domain and polar interactions involving the Q residues. In systems with shorter polyQ lengths, the dominance of hydrophobic interactions among N17 domains leads to their bundling. However, as the polyQ length increases, the influence of polar Q–Q interactions becomes more prominent, diminishing the bundling of N17 domains due to a reduction in the dominance of hydrophobic interactions.

The same trend is evident in the radius of gyration of the phenylalanine (F) side chains, as depicted in SI 1, Figure S4, which increases with polyQ length. At shorter polyQ lengths, the hydrophobic residues tend to face inward due to N17 bundling, resulting in lower values of the radius of gyration. Conversely, in systems with longer polyQ domains and consequently reduced N17 bundling, the F side chains become more exposed to solvent, leading to higher values of the radius of gyration. Short videos depicting the final oligomeric structures for polyQ lengths of 7 and 45 are provided in Video 1 and Video 2, respectively. Previous studies on the N17Q₃₀P₁₀K₂ fragment of htt using solid-state nuclear

magnetic resonance (ssNMR) techniques have suggested an external distribution of N17.²²

Variations in Oligomer Morphology and Observed Polymorphisms. To gain insight into the possible conformations of oligomeric structures, we evaluated the number of interacting N17 domains within the oligomer, denoted as N_{int} . With 6 peptides in our simulation, N_{int} can range from 0 to 6, except for 1. Figure 4A illustrates the probability distribution of various N_{int} values for each peptide system. At shorter polyQ lengths, there is a high probability of observing $N_{int} = 6$, indicating extensive bundling between N17 domains within the oligomer. For instance, in the 7Q system, the probability of $N_{int} = 6$ is approximately 0.55. However, as the polyQ length increases, this probability decreases, with the 45Q system showing the highest probability for $N_{int} = 2$.

For htt^{mutant}, we observe a heightened variability in the number of interacting N17 domains compared to htt^{wt}, indicating multiple oligomeric conformations with distinct N17 distributions on the oligomeric surface. Previous experimental studies have demonstrated that N17 engages in tertiary and quaternary interactions involving both hydrophobic and hydrophilic interfaces at longer polyQ lengths.²⁰ This aligns with the nonspecific bundling of N17 observed in our simulations, reflected in the high variability of N_{int} and solvation of F side chains. Despite its amphipathic nature driving N17 to bundle with neighboring domains through hydrophobic interactions, this tendency is hindered by increasing polyQ length.

Recent cryo-electron microscopy (cryo-EM) studies by Nazarov et al. have suggested that the core of htt fibrils consists of heterogeneous structures within the polyQ core, while the N17 and proline-rich domain (PRD) remain highly mobile around the polyQ core. Although concrete conclusions await a 3D cryo-EM structure, the reduction in N_{int} observed in our simulations with long polyQ aligns strongly with their findings.²⁴

In tandem with the fluctuations in N_{int} , there is an augmentation in the average number of Q–Q interactions per glutamine residue, as seen in Figure 4B. In polyQ systems with $Q \geq 35$, each glutamine residue interacts, on average, with 4 or more other glutamine residues. This is indicative of a vast polar, dipole bonding network between the glutamine residues, consistent with the polyQ domain's reported capacity to form hydrogen bonds (which are a kind of dipole interaction) via its side chains and its propensity for self-association.⁸ These

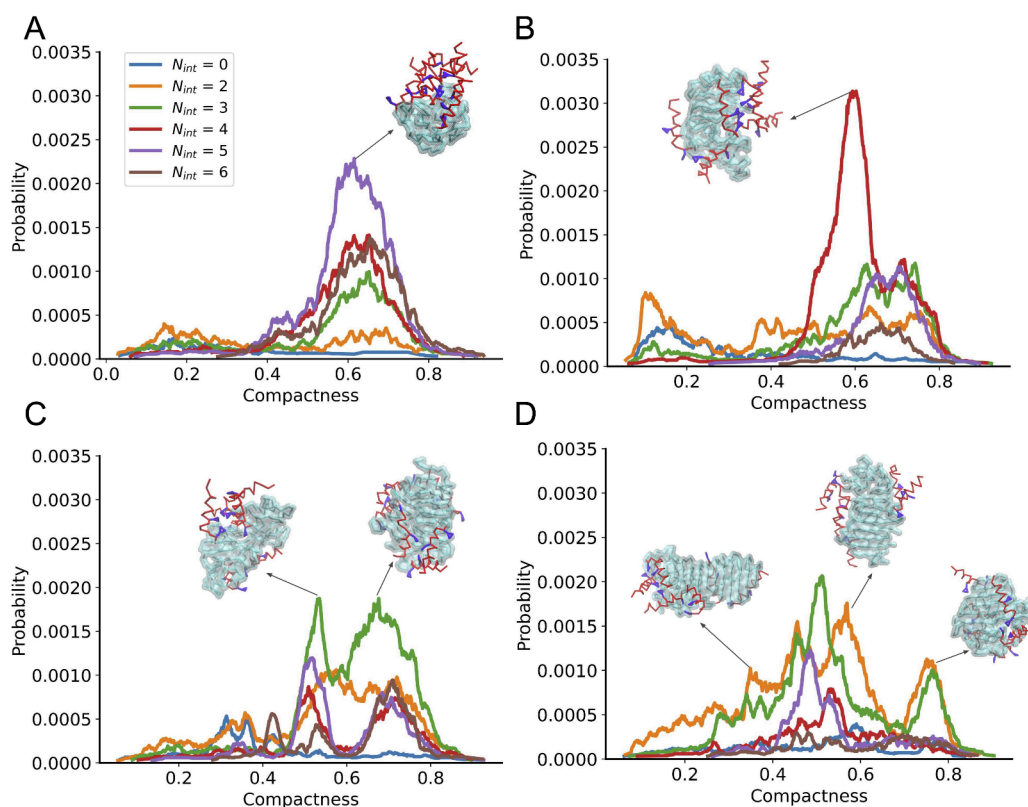


Figure 5. Flexibility in oligomer shape is indicative of polymorphism observed in systems with long polyQ lengths. Oligomer compactness distributions for (A) 15Q, (B) 35Q, (C) 40Q, and (D) 45Q systems at different values of the number of interacting N17 domains (N_{int}). N17, polyQ, and F side chains are represented in red, cyan, and violet, respectively.

trends align with the results of the monomeric simulations, where the average number of Q–Q interactions per glutamine residue increases, while the average number of water beads interacting with each glutamine residue reduces (see SI 2, Figure S4). These findings are in agreement with simulations of polyQ performed by Wen et al. and a growing body of work that posits water is a poor solvent for polyQ.^{41–43}

To further elucidate the diverse oligomeric conformations characterized by distinct N17 locations, we analyze the shapes of the oligomers, as illustrated in Figure 5. A compactness value of 1 indicates a spherical aggregate, while a value closer to 0 represents a prolate/extended conformation.

The compactness of oligomers reveals a notable correlation with the length of the polyQ segment. In the 15Q system (Figure 5A), a singular peak around a compactness value of approximately 0.58 is consistently observed across all N_{int} values, indicating a predominant polyQ oligomeric conformation unaffected by the number of interacting N17 domains. The inset snapshot illustrates a representative image for the 15Q system with $N_{\text{int}} = 5$, showing bundling of the N17 domain resulting in a hydrophobic core, evident from the close proximity of F side chains (violet). The 7Q system demonstrates a similar trend (see SI 1, Figure S5).

As the polyQ length increases, the oligomer can adopt more extended conformations, as evidenced by the broadening of the compactness distributions. The 40Q system presents a bimodal distribution for $N_{\text{int}} > 2$, suggesting two major conformations adopted by the polyQ domain in the oligomeric structure. Representative snapshots corresponding to the peaks of $N_{\text{int}} = 3$ are presented. In the 45Q system, there is an overall increase in the variety of shapes the oligomer can take on, as evident

from the multiple peaks in the compactness distribution. The snapshots presented in Figure 5D correspond to regions of $N_{\text{int}} = 2$. The oligomer can adopt a range of conformations, from spherical to elongated shapes. The ensembles with low values of N_{int} ($N_{\text{int}} < 4$) have broader distributions, adopting conformations that can be either spherical or elongated. An enhanced variability in the shape of the oligomer is observed because low N_{int} values are observed at longer lengths of polyQ (Figure 4A).

The polymorphic nature of htt emerges predominantly at longer polyQ lengths, enabling the oligomer to adopt multiple conformational states corresponding to varying levels of compactness and distinct positioning of N17 domains. Conversely, in oligomers with shorter polyQ lengths, characterized by primarily high N_{int} values, the range of conformational polymorphisms is considerably restricted, resulting in overlapping compactness distributions.

These findings underscore the conformational flexibility and polymorphic nature of htt protein fragments. At shorter polyQ lengths, the oligomers adopt more uniform structures, as evidenced by the probability of having high N_{int} (i.e., finding all N17 domains interacting) as seen in Figure 4, and the low variability in compactness (Figure 5). Here, the hydrophobic interactions between N17 domains lead to a more invariable oligomer. In contrast, systems with longer polyQ lengths experience two contrasting forces—hydrophobic interactions between N17 and the polar interactions between the Q residues. These opposing interactions result in oligomers that can adopt a wider variety of conformations, as seen by the numerous peaks in the compactness distributions. So far, the current knowledge on polymorphism has come from studying

fibrillar structures. We hypothesize that these results may provide insight into the molecular mechanisms behind the varying amyloid states.

CONCLUSIONS

We observe a shift in the driving forces of association as polyQ length varies: shorter polyQ lengths predominantly exhibit N17–N17 interactions, whereas longer polyQ lengths show an increased prevalence of Q–Q/N17–Q interactions between peptides. Our findings align well with existing experimental studies, where the polyQ domain demonstrates a propensity to form helices and sheets. Moreover, the length of the polyQ sequence dictates the preference for intermolecular sheet contacts over intramolecular ones, particularly favored at longer polyQ lengths. These results corroborate spectroscopic findings by Yushchenko et al.³⁸

We have revealed how oligomeric morphology varies with the length of the polyQ segment. Short polyQ lengths lead to oligomers with a distinct hydrophobic core, driven by the bundling of N17 domains where their hydrophobic sides face inward. As the polyQ segment elongates, N17 bundling diminishes, and the oligomer core transitions to interlinked glutamine regions, while N17 relocates to the aggregate surface. This finding aligns with previous studies indicating tertiary and quaternary interactions among N17 domains²⁰ and N17 positioning on the periphery of the amyloid core.²² The repositioning of N17 results from a competition between its hydrophobic effects and the polar interactions of the polyQ tract, with hydrophobic interactions dominating at shorter lengths. We speculate that these insights could shed light on the heightened toxicity observed in htt^{mutant} aggregates.

A number of studies have indicated that in Huntington's disease, the protein undergoes a toxic gain-of-function by interfering with normal function. One such mechanism involves the sequestration of transcription factors like CREB binding protein by mutant htt.⁴⁴ Co-aggregation of htt with CREB, increased cell toxicity, and depletion of CREB binding protein has been reported.^{45,46} Notably, our findings reveal that in htt^{mutant} oligomers, N17 is significantly more solvent exposed (see Figure 3). This heightened exposure, particularly of the hydrophobic F residues in the N17 domain, could create favorable binding sites. We hypothesize that these oligomers may bind to essential enzymes and proteins, leading to the depletion of cellular resources crucial for maintaining homeostasis.

Furthermore, our study demonstrates that htt^{mutant} oligomeric structures exhibit polymorphic characteristics. A measure of oligomer shape was employed to compare ensembles with similar N_{int} values (representing the number of interacting N17 domains) within peptide systems. For systems with longer polyQ (≥ 35), multiple shapes are observed from spherical to elongated. In contrast, polyQ < 35 predominantly displayed overlapping compactness distributions for all N_{int} values, indicating that htt^{wt} oligomers mainly adopt a single conformation that remains consistent across changes in N_{int} .

Amyloid fibrils are known to exhibit various conformations from the same protein sequence, allowing them to propagate through the recruitment of additional molecules.^{47,48} This process typically involves the exposure of hydrophobic residues to the solvent. In on-pathway folding, these hydrophobic interactions stabilize the protein core, whereas in off-pathway folding, they can recruit other proteins.^{49,50} We hypothesize

that the reduction in N17 bundling may contribute to the polymorphisms observed in htt^{mutant} fibrils.

Moreover, N17 has been identified as one of the membrane-binding domains of htt, facilitating its subcellular localization to various organelles.⁵¹ It contains a nuclear export sequence that directs its translocation into the cytoplasm.⁵² Chaibva et al. discovered N17's ability to sense membrane curvature.⁵³ Therefore, the external distribution of N17 in htt^{mutant} oligomers could explain a potential gain-of-function mechanism in diseased states, particularly its interaction with membranes. Our future research will involve characterizing htt^{wt} and htt^{mutant} oligomers in association with lipid membranes of varying curvatures and compositions.

Although our results are in alignment with existing studies on htt, it is necessary to highlight some limitations of our approach. In this study, we examine the structural characteristics of the oligomeric state of htt with 6 peptides. However, studying larger oligomers is beyond the scope of our current approach. We can speculate that with increasing oligomer size, a greater degree of polymorphism will be observed. In addition, our current simulations do not account for the effects of the proline rich domain (PRD) on the structural polymorphisms observed in htt. The PRD is known to be a very flexible region that can inhibit the rate of htt aggregation. Our future works will involve the examination of the effect the PRD has on the polymorphic nature of htt.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcb.4c03274>.

Information on the force field, protein representation, and oligomeric structure (PDF)

Relevant analysis for the monomeric systems (PDF)

Oligomeric structure of the 7Q system (MP4)

Oligomeric structure of the 45Q system (MP4)

AUTHOR INFORMATION

Corresponding Author

Silvina Matysiak – Fischell Department of Bioengineering, University of Maryland, College Park, Maryland 20740, United States; orcid.org/0000-0003-3824-9787; Phone: +1(301) 405-0313; Email: matysiak@umd.edu

Authors

Neha Nanajkar – Department of Biology, University of Maryland, College Park, Maryland 20740, United States

Abhilash Sahoo – Center for Computational Biology and Center for Computational Mathematics, Flatiron Institute, New York, New York 10010, United States

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jpcb.4c03274>

Author Contributions

[§]N.N. and A.S. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation under grants CHE-1454948 and CHE-2202281. The authors acknowledge the University of Maryland super-

computing resources (<http://hpcc.umd.edu>) made available for conducting the research reported in this paper.

REFERENCES

- (1) DiFiglia, M.; Sapp, E.; Chase, K.; Schwarz, C.; Meloni, A.; Young, C.; Martin, E.; Vonsattel, J.-P.; Carraway, R.; Reeves, S. A.; et al. Huntingtin is a cytoplasmic protein associated with vesicles in human and rat brain neurons. *Neuron* **1995**, *14*, 1075–1081.
- (2) Zuccato, C.; Tartari, M.; Crotti, A.; Goffredo, D.; Valenza, M.; Conti, L.; Cataudella, T.; Leavitt, B. R.; Hayden, M. R.; Timmusk, T.; et al. Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. *Nature genetics* **2003**, *35*, 76–83.
- (3) Schulte, J.; Littleton, J. T. The biological function of the Huntingtin protein and its relevance to Huntington's Disease pathology. *Current trends in neurology* **2011**, *5*, 65.
- (4) Donaldson, J.; Powell, S.; Rickards, N.; Holmans, P.; Jones, L. What is the pathogenic CAG expansion length in Huntington's disease? *Journal of Huntington's Disease* **2021**, *10*, 175–202.
- (5) Cohen-Carmon, D.; Meshorer, E. Polyglutamine (polyQ) disorders: the chromatin connection. *Nucleus* **2012**, *3*, 433–441.
- (6) Temussi, P. A.; Masino, L.; Pastore, A. From Alzheimer to Huntington: why is a structural understanding so difficult? *EMBO journal* **2003**, *22*, 355–361.
- (7) Kim, M. W.; Chelliah, Y.; Kim, S. W.; Otwinowski, Z.; Bezprozvanny, I. Secondary structure of Huntingtin amino-terminal region. *Structure* **2009**, *17*, 1205–1212.
- (8) Perutz, M. F.; Johnson, T.; Suzuki, M.; Finch, J. T. Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 5355–5358.
- (9) Chellgren, B. W.; Miller, A.-F.; Creamer, T. P. Evidence for polyproline II helical structure in short polyglutamine tracts. *Journal of molecular biology* **2006**, *361*, 362–371.
- (10) Masino, L.; Kelly, G.; Leonard, K.; Trottier, Y.; Pastore, A. Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins. *Febs Letters* **2002**, *513*, 267–272.
- (11) Khare, S. D.; Ding, F.; Gwanmesia, K. N.; Dokholyan, N. V. Molecular origin of polyglutamine aggregation in neurodegenerative diseases. *PLoS computational biology* **2005**, *1*, No. e30.
- (12) Warner, J. B., IV; Ruff, K. M.; Tan, P. S.; Lemke, E. A.; Pappu, R. V.; Lashuel, H. A. Monomeric huntingtin exon 1 has similar overall structural features for wild-type and pathological polyglutamine lengths. *J. Am. Chem. Soc.* **2017**, *139*, 14456–14469.
- (13) Kang, H.; Vázquez, F. X.; Zhang, L.; Das, P.; Toledo-Sherman, L.; Luan, B.; Levitt, M.; Zhou, R. Emerging β -sheet rich conformations in supercompact huntingtin exon-1 mutant structures. *J. Am. Chem. Soc.* **2017**, *139*, 8820–8827.
- (14) Urbanek, A.; Popovic, M.; Morató, A.; Estaña, A.; Elena-Real, C. A.; Mier, P.; Fournet, A.; Allemand, F.; Delbecq, S.; Andrade-Navarro, M. A.; et al. Flanking regions determine the structure of the poly-glutamine in huntingtin through mechanisms common among glutamine-rich human proteins. *Structure* **2020**, *28*, 733–746.
- (15) Lakhani, V. V.; Ding, F.; Dokholyan, N. V. Polyglutamine induced misfolding of huntingtin exon1 is modulated by the flanking sequences. *PLoS computational biology* **2010**, *6*, No. e1000772.
- (16) Crick, S. L.; Ruff, K. M.; Garai, K.; Frieden, C.; Pappu, R. V. Unmasking the roles of N- and C-terminal flanking sequences from exon 1 of huntingtin as modulators of polyglutamine aggregation. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 20075–20080.
- (17) Kelley, N. W.; Huang, X.; Tam, S.; Spiess, C.; Frydman, J.; Pande, V. S. The predicted structure of the headpiece of the Huntingtin protein and its implications on Huntingtin aggregation. *Journal of molecular biology* **2009**, *388*, 919–927.
- (18) Michalek, M.; Salnikov, E. S.; Werten, S.; Bechinger, B. Membrane interactions of the amphipathic amino terminus of huntingtin. *Biochemistry* **2013**, *52*, 847–858.
- (19) Thakur, A. K.; Jayaraman, M.; Mishra, R.; Thakur, M.; Chellgren, V. M.; L. Byeon, I.-J.; Anjum, D. H.; Kodali, R.; Creamer, T. P.; Conway, J. F.; et al. others Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism. *Nature structural molecular biology* **2009**, *16*, 380–389.
- (20) Bugg, C. W.; Isas, J. M.; Fischer, T.; Patterson, P. H.; Langen, R. Structural features and domain organization of huntingtin fibrils. *J. Biol. Chem.* **2012**, *287*, 31739–31746.
- (21) Hoop, C. L.; Lin, H.-K.; Kar, K.; Magyarfalvi, G.; Lamley, J. M.; Boatz, J. C.; Mandal, A.; Lewandowski, J. R.; Wetzel, R.; Van Der Wel, P. C. Huntingtin exon 1 fibrils feature an interdigitated β -hairpin-based polyglutamine core. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 1546–1551.
- (22) Sivanandam, V.; Jayaraman, M.; Hoop, C. L.; Kodali, R.; Wetzel, R.; van der Wel, P. C. The aggregation-enhancing huntingtin N-terminus is helical in amyloid fibrils. *J. Am. Chem. Soc.* **2011**, *133*, 4558–4566.
- (23) Chen, M.; Wolynes, P. G. Aggregation landscapes of Huntingtin exon 1 protein fragments and the critical repeat length for the onset of Huntington's disease. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 4406–4411.
- (24) Nazarov, S.; Chiki, A.; Boudeffa, D.; Lashuel, H. A. Structural basis of huntingtin fibril polymorphism revealed by cryogenic electron microscopy of exon 1 HTT fibrils. *J. Am. Chem. Soc.* **2022**, *144*, 10723–10735.
- (25) Sahoo, A.; Lee, P.-Y.; Matysiak, S. Transferable and Polarizable Coarse Grained Model for Proteins- ProMPT. *J. Chem. Theory Comput.* **2022**, *18*, 5046–5055.
- (26) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.
- (27) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (28) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J.; Melo, M. N.; Seyler, S. L.; Domanski, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; et al. MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations. *Proceedings of the 15th python in science conference*. 2016; p 105.
- (29) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
- (30) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.
- (31) Lee, P.-Y.; Sahoo, A.; Matysiak, S. Folding and modulation of the helical conformation of Glycophorin A by point mutations. *Phys. Chem. Chem. Phys.* **2023**, *25*, 10885–10893.
- (32) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. Polarizable water model for the coarse-grained MARTINI force field. *PLoS computational biology* **2010**, *6*, No. e1000810.
- (33) Jayaraman, M.; Kodali, R.; Sahoo, B.; Thakur, A. K.; Mayasundari, A.; Mishra, R.; Peterson, C. B.; Wetzel, R. Slow amyloid nucleation via α -helix-rich oligomeric intermediates in short polyglutamine-containing huntingtin fragments. *Journal of molecular biology* **2012**, *415*, 881–899.
- (34) Toofanny, R. D.; Daggett, V. Understanding protein unfolding from molecular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 405–423.
- (35) Khurshid, B.; Rehman, A. U.; Luo, R.; Khan, A.; Wadood, A.; Anwar, J. Heparin-Assisted Amyloidogenesis Uncovered through Molecular Dynamics Simulations. *ACS omega* **2022**, *7*, 15132–15144.
- (36) Cino, E. A.; Choy, W.-Y.; Karttunen, M. Comparison of secondary structure formation using 10 different force fields in microsecond molecular dynamics simulations. *J. Chem. Theory Comput.* **2012**, *8*, 2725–2740.
- (37) Schäffler, M.; Khaled, M.; Strodel, B. ATRANET—Automated generation of transition networks for the structural characterization of intrinsically disordered proteins. *Methods* **2022**, *206*, 18–26.

- (38) Yushchenko, T.; Deuerling, E.; Hauser, K. Insights into the aggregation mechanism of PolyQ proteins with different glutamine repeat lengths. *Biophysical journal* **2018**, *114*, 1847–1857.
- (39) Fiumara, F.; Fioriti, L.; Kandel, E. R.; Hendrickson, W. A. Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell* **2010**, *143*, 1121–1135.
- (40) Dougan, L.; Li, J.; Badilla, C. L.; Berne, B.; Fernandez, J. M. Single homopolypeptide chains collapse into mechanically rigid conformations. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12605–12610.
- (41) Wen, J.; Scoles, D. R.; Facelli, J. C. Molecular dynamics analysis of the aggregation propensity of polyglutamine segments. *PloS one* **2017**, *12*, No. e0178333.
- (42) Crick, S. L.; Jayaraman, M.; Frieden, C.; Wetzel, R.; Pappu, R. V. Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 16764–16769.
- (43) Vitalis, A.; Wang, X.; Pappu, R. V. Atomistic simulations of the effects of polyglutamine chain length and solvent quality on conformational equilibria and spontaneous homodimerization. *Journal of molecular biology* **2008**, *384*, 279–297.
- (44) Landles, C.; Bates, G. P. Huntingtin and the molecular pathogenesis of Huntington's disease: Fourth in Molecular Medicine Review Series. *EMBO reports* **2004**, *5*, 958–963.
- (45) Steffan, J. S.; Kazantsev, A.; Spasic-Boskovic, O.; Greenwald, M.; Zhu, Y.-Z.; Gohler, H.; Wanker, E. E.; Bates, G. P.; Housman, D. E.; Thompson, L. M. The Huntington's disease protein interacts with p53 and CREB-binding protein and represses transcription. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 6763–6768.
- (46) Jiang, H.; Poirier, M. A.; Liang, Y.; Pei, Z.; Weiskittel, C. E.; Smith, W. W.; DeFranco, D. B.; Ross, C. A. Depletion of CBP is directly linked with cellular toxicity caused by mutant huntingtin. *Neurobiology of disease* **2006**, *23*, 543–551.
- (47) Tycko, R. Amyloid polymorphism: structural basis and neurobiological relevance. *Neuron* **2015**, *86*, 632–645.
- (48) Petkova, A. T.; Leapman, R. D.; Guo, Z.; Yau, W.-M.; Mattson, M. P.; Tycko, R. Self-propagating, molecular-level polymorphism in Alzheimer's β -amyloid fibrils. *Science* **2005**, *307*, 262–265.
- (49) Adamcik, J.; Mezzenga, R. Amyloid polymorphism in the protein folding and aggregation energy landscape. *Angew. Chem., Int. Ed.* **2018**, *57*, 8370–8382.
- (50) Paravastu, A. K.; Leapman, R. D.; Yau, W.-M.; Tycko, R. Molecular structural basis for polymorphism in Alzheimer's β -amyloid fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 18349–18354.
- (51) Rockabrand, E.; Slepko, N.; Pantalone, A.; Nukala, V. N.; Kazantsev, A.; Marsh, J. L.; Sullivan, P. G.; Steffan, J. S.; Sensi, S. L.; Thompson, L. M. The first 17 amino acids of Huntingtin modulate its sub-cellular localization, aggregation and effects on calcium homeostasis. *Human molecular genetics* **2007**, *16*, 61–77.
- (52) Zheng, Z.; Li, A.; Holmes, B. B.; Marasa, J. C.; Diamond, M. I. An N-terminal nuclear export signal regulates trafficking and aggregation of Huntingtin (Htt) protein exon 1. *J. Biol. Chem.* **2013**, *288*, 6063–6071.
- (53) Chaibva, M.; Burke, K. A.; Legleiter, J. Curvature enhances binding and aggregation of huntingtin at lipid membranes. *Biochemistry* **2014**, *53*, 2355–2365.