frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 10 April 2023
DOI 10.3389/fmolb.2023.1143353

OPEN ACCESS

EDITED BY
Alfonso De Simone,
University of Naples Federico II, Italy

REVIEWED BY
Phuong Nguyen,
UPR9080 Laboratoire de Biochimie Théorique (LBT), France
Davide Mercadante,
The University of Auckland, New Zealand

*CORRESPONDENCE
Birgit Strodel,
b.strodel@fz-juelich.de
Abdallah Sayyed-Ahmad,
asayyeda@birzeit.edu

SPECIALTY SECTION
This article was submitted to Protein Folding, Misfolding and Degradation, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 12 January 2023
ACCEPTED 29 March 2023
PUBLISHED 10 April 2023

CITATION
Khaled M, Strodel B and Sayyed-Ahmad A (2023), Comparative molecular dynamics simulations of pathogenic and non-pathogenic huntingtin protein monomers and dimers.
Front. Mol. Biosci. 10:1143353.
doi: 10.3389/fmolb.2023.1143353

Comparative molecular dynamics simulations of pathogenic and non-pathogenic huntingtin protein monomers and dimers

Mohammed Khaled[1], Birgit Strodel[1,2]* and Abdallah Sayyed-Ahmad[3]*

[1]Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Forschungszentrum Jülich, Jülich, Germany, [2]Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, Düsseldorf, Germany, [3]Department of Physics, Birzeit University, Birzeit, Palestine

Polyglutamine expansion at the N-terminus of the huntingtin protein exon 1 (Htt-ex1) is closely associated with a number of neurodegenerative diseases, which result from the aggregation of the increased polyQ repeat. However, the underlying structures and aggregation mechanism are still poorly understood. We performed microsecond-long all-atom molecular dynamics simulations to study the folding and dimerization of Htt-ex1 (about 100 residues) with non-pathogenic and pathogenic polyQ lengths, and uncovered substantial differences. The non-pathogenic monomer adopts a long α-helix that includes most of the polyQ residues, which forms the interaction interface for dimerization, and a PPII-turn-PPII motif in the proline-rich region. In the pathogenic monomer, the polyQ region is disordered, leading to compact structures with many intra-protein interactions and the formation of short β-sheets. Dimerization can proceed *via* different modes, where those involving the N-terminal headpiece bury more hydrophobic residues and are thus more stable. Moreover, in the pathogenic Htt-ex1 dimers the proline-rich region interacts with the polyQ region, which slows the formation of β-sheets.

KEYWORDS

polyglutamine, huntingtin, molecular dynamics, oligomer, aggregation

1 Introduction

Huntington's disease (HD) is an inherited neurodegenerative disease caused by an abnormal expansion in the polyglutamine (polyQ) tract of the N-terminal Huntingtin (Htt-ex1) protein. The elongated polyQ tract mutation is caused by the expansions of nucleotide CAG repeats in exon-1 of the HD gene that encodes the elongated polyQ tract within the Htt-ex1 protein. Furthermore, the expansion of polyQ tracts and their toxicity are correlated to the age of the onset of HD (MacDonald et al., 1993). In HD, Htt-ex1 becomes pathogenic beyond a threshold of 36 glutamine repeats (Takeuchi and Nagai, 2017). The full-length Htt-ex1 is over 3,100 amino acids expressed in all mammalian cells including nerve cells in the brain where it has higher concentrations. Htt-ex1 interacts with a wide range of proteins involved in many cellular processes (Gusella and MacDonald, 2000; Li and Li, 2004; Saudou and Humbert, 2016), but the exact structure and function of Htt-ex1 are still poorly understood (Mangiarini et al., 1996). Nevertheless, HD onset and progression are associated with the misfolding of Htt-ex1 which eventually forms amyloid aggregates

Htt-Q_{23} sequence

Nt_{17}	$polyQ_{23}$	$polyP_{11}$	PRD	$polyP_{10}$	
MATLEKLMKAFESLKSF	-Q23-	-P11-	QLPQPPPQAQPLLPQPQ	-P10-	GPAVAEEPLHRP
17	40	51	68	78	90

Htt-Q_{48} sequence

Nt_{17}	$polyQ_{48}$	$polyP_{11}$	PRD	$polyP_{10}$	
MATLEKLMKAFESLKSF	-Q48-	-P11-	QLPQPPPQAQPLLPQPQ	-P10-	GPAVAEEPLHRP
17	65	76	93	103	115

FIGURE 1
Sequence of Htt-Q_{23} and Htt-Q_{48} studied in this work. The Nt_{17} region is highlighted in red, polyQ in blue, and the PRD is shown in yellow for $polyP_{11}$ and $polyP_{10}$ and rose for the other PRD parts.

that are connected to neuronal cell death (Perutz et al., 1994). Toxic aggregation of Htt-ex1 into amyloid cannot only be observed *in vivo*, but can also be reproduced *in vitro* (Nagai et al., 2007). Several hypotheses have been suggested to explain the aggregation behavior and thereby the toxicity of Htt-ex1. One of these hypotheses proposes that pathogenic Htt-ex1 accumulates into insoluble aggregates in neurons as amyloid fibrillar structures (Miller et al., 2011). Other hypotheses suggest that Htt-ex1 monomers or oligomers with extended polyQ tracts interact with other cellular proteins and alter their functions, which leads to neuronal cell death (Williams and Paulson, 2008; Gkekas et al., 2021).

Importantly, the expansion of the polyQ tract is also associated with many other inherited neurodegenerative diseases (Williams and Paulson, 2008). Thus, several studies have been conducted to investigate the structures and aggregation mechanisms of isolated polyQ peptides. Experiments under different solutions conditions found that isolated polyQ tracts could sample various conformations as collapsed structures (Crick et al., 2006) with random coils (Chen et al., 2001; Klein et al., 2007), α-helix (Bhattacharyya et al., 2006), β-sheets (Nagai et al., 2007; Darnell et al., 2009) and PPII helix (Chellgren et al., 2006; Darnell et al., 2007; Darnell et al., 2009). The conformational flexibility of polyQ tracts has also been confirmed by computational modeling using molecular dynamics (MD) simulations, as was comprehensively reviewed by Moldovean and Chis (Moldovean and Chiş, 2019). Of particular note are the studies (Długosz and Trylska, 2011; Vöpel et al., 2017; Priya and Gromiha, 2019) as they did not only simulate isolated polyQ stretches as done in the majority of the other simulation studies, but included the first 17 N-terminal residues directly preceding the polyQ sequence and some of the following C-terminal residues. Despite the various experimental and simulation efforts, there is no consensus yet on the preferred polyQ structure in solution. Almost all secondary structures have been suggested, ranging from α-helical structures (Elena-Real et al., 2022) to coil and β-sheets (Moldovean and Chiş, 2019). Nonetheless, there is ample evidence that indicates that longer polyQ increases the β-sheet propensity and aggregation rates (Thakur and Wetzel, 2002; Klein et al., 2007; Sivanandam et al., 2011).

The Htt-ex1 consists of three regions (Figure 1): the N-terminal 17 amino acid region (Nt_{17}), the polyQ tract region, followed by a proline-rich region (PRD). The N-terminal and proline-rich regions surrounding the polyQ region have critical roles in modulating the aggregation mechanism as has been illustrated in computational and experimental studies (Chow et al., 2012; Qin et al., 2004; Duennwald et al., 2006a; Duennwald et al., 2006b). The Nt_{17} region is known to accelerate the aggregation by forming prefibrillar spherical oligomers with Nt_{17} at their core (Tam et al., 2009; Thakur et al., 2009). Experimentally determined structures of Nt_{17} suggested that it folds into an amphipathic α-helix (Kim et al., 2009; Kim, 2013; Michalek et al., 2013). In addition, it has been suggested that the α-helix Nt_{17} may initialize the Htt-ex1 aggregation by forming α-helix rich oligomers (Hoop et al., 2014; Sahoo et al., 2014; Pandey et al., 2018). In general, it appears that Nt_{17} is mostly helical in Htt-ex1 aggregates (Sivanandam et al., 2011; Jayaraman et al., 2012; Hoop et al., 2014). However, nuclear magnetic resonance (NMR) spectroscopy showed that structures of Nt_{17} are intrinsically disordered and adopt different conformations (Thakur et al., 2009). The PRD, on the other hand, is found to decrease the stability and rate of amyloid-like aggregation without changing the fundamental mechanism of aggregation (Bhattacharyya et al., 2006). NMR and electron paramagnetic resonance (EPR) spectroscopy showed that the PRD tends to adopt similar structures with rich PPII helical structures in both monomers and fibril structures (Bugg et al., 2012; Isas et al., 2015).

Proteins with glutamine repeats are structurally unstable. This makes it difficult to resolve the structure of regions surrounding the polyQ tracts (Takeuchi and Nagai, 2017). As of today, only structures of the Htt-ex1 containing 17 glutamine repeats were resolved by X-ray crystallography (Kim et al., 2009). Thus, various computational studies of model Htt-ex1 have been conducted to illustrate the effect of polyQ tract length on protein structure (Moldovean and Chiş, 2019). Other studies provided insights into the effects of Htt-ex1 flanking domains on its structure (Długosz and Trylska, 2011; Vöpel et al., 2017; Priya and Gromiha, 2019). By adding or removing these domains, the PRD was found to destabilize the protein and inhibited β-sheet formation (Lakhani et al., 2010; Williamson et al., 2010). Kang and coworkers carried out MD simulations of full-length Htt-ex1 monomers with different polyQ lengths. Their results suggest a positive correlation between polyQ tract length and the β-sheet

content (Kang et al., 2017). Finally, the polyQ region was found to be the main driver for Htt-ex1 aggregation (Williamson et al., 2010).

In this work, we performed all-atom MD simulations to investigate the structural properties of Htt-Q_n monomers and dimers with n = 23 for non-pathogenic Htt-ex1 and n = 48 for disease-causing Htt-ex1. In addition to including the Nt_{17} region, we also added 50 amino acids to model the impact of the PRD on the polyQ region (Figure 1). With this, our simulated systems are longer than those usually simulated. In addition, we used the Charmm36m force field, which has been adapted to proteins with low-complexity sequences (Huang et al., 2017) and has been shown to be useful in predicting the conformational ensemble of intrinsically disordered proteins (Robustelli et al., 2018; Paul et al., 2021) and amyloid aggregates (Samantray et al., 2020), but has not yet been applied to Htt-ex1 proteins. Furthermore, we performed simulations on the microsecond time scale, which is longer than in most preceding simulations. Our study therefore provides updated insights into the combined effects of Nt_{17}, PRD, and polyQ tract length on the Htt-ex1 structure and aggregation.

2 Materials and methods

2.1 Initial structures and system preparations

The two simulated sequences are MATLEKLMKAFESLKSF-Q_n-P_{11}-QLPQPPPQAQPLLPQPQ-P_{10}-GPAVAEEPLHRP with n = 23 and n = 48 glutamine residues (Figure 1). The X-ray crystal structure of Htt-ex1 (PDB ID: 3IOT) (Kim et al., 2009) consists of an α-helix for Nt_{17} and the α-helix extends up to 15 glutamines into the $polyQ_{17}$ tract. For the $polyP_{11}$ region, extended loops, random coil conformations and also PPII-helix formation were observed. However, since the structure of the full-length Htt-ex1 protein is not completely resolved, the initial structures were built using the Avogadro software (Hanwell et al., 2012) and PyMOL (Schrödinger, 2022) by adding the peptide fragment Q_{11}-PRD and Q_{36}-PRD to the N-terminal part Nt_{17}-Q_{12} taken from the PDB structure 3IOT for modeling the Htt-Q_{23} and Htt-Q_{48}, respectively. This resulted in extended structures for both proteins, which were solvated with water and then collapsed through initial 20 ns MD simulations to reduce the system size in the production MD simulations.

2.2 MD simulations

All MD simulations were carried out using GROMACS 2020 (Van Der Spoel et al., 2005; Abraham et al., 2015) as MD program, employing the atomistic force field Charmm36m along with the Charmm-modified TIP3P water model (Huang et al., 2017). The initial (already collapsed) monomer structures were solvated in a ~2,000 nm^3 cubic box, including 150 mM Na^+ and Cl^- ions and overall system neutralization, leading to about 200,000 atoms. The energy of each system was initially minimized using the steepest descent algorithm (Müller and Brown, 1979; Zhang, 2015), followed by 1.0 ns equilibration and then the production runs under NPT ensemble conditions. The pressure and temperature were maintained at 1.0 bar and 298 K using the Parrinello-Rahman pressure coupling method (Parrinello and Rahman, 1981; Parrinello and Rahman, 1982) and a velocity-rescale thermostat method (Bussi et al., 2007), respectively. All simulations implied periodic boundary conditions applied in all directions and using a cutoff distance of 1.2 nm for the calculation of the non-bonded interactions in real space. The electrostatic interactions were computed using the particle mesh Ewald (PME) method (Essmann et al., 1995). The LINCS algorithm was applied (Hess et al., 1997) to constrain bond lengths, and a leapfrog integrator (Van Gunsteren and Berendsen, 1988) was used to integrate the equations of motions with a 2 fs time step. For all simulations it was checked that there were no self-interactions of the proteins with their periodic images during the course of the simulations as a result of conformational reorientations.

For each monomer system, 5 × 1.0 μs simulations were carried out. Simulation 2 was initiated from the final snapshot of the first simulation, while simulations 3 through 5 were started from the central structure of the most populated conformational cluster of the respective preceding simulation. For the analysis, the first simulation per sequence was discarded due to the bias of the extended initial structures, despite the initial 20 ns MD simulation to collapse these structures. This bias is visible in the time traces of various observables (Supplementary Figures S1–S3). To assesses the convergence of the monomer simulations, the autocorrelation functions of various structural properties were calculated for the individual 1.0 μs runs (Supplementary Figure S4). They all tend to zero within 500 ns, indicating loss of memory from initial structures. Nonetheless, while general convergence of the individual trajectories is found, certain fluctuations of the various structural properties persist, which is an expected behavior for intrinsically disordered proteins (Paul et al., 2021). We further conducted the Augmented Dickey Fuller (ADF) test as a unit-root test to assess the stationarity of the monomer data. The ADF test was applied to the concatenated time series of various observables (Supplementary Table S1). The obtained p-values for all the tested observables are below 0.05, and the test statistics are also smaller than the critical values, thereby rejecting the null hypothesis and indicating that the data are stationary.

For both proteins, 3 × 2.0 μs dimer simulations were carried out in a cubic simulation box with a volume of about 1,925 nm^3 and a total number of 190,000 atoms. The six most dominant conformational clusters determined from the previous monomer simulations were used as starting structures for the monomers in the dimer simulations. The simulations were initiated by randomly orienting two monomers in the simulation box, ensuring a minimum distance of 5 nm between them.

2.3 Analysis

The MD simulations were analyzed using different tools, which were invoked from GROMACS or from the MDAnalysis Python package (Michaud-Agrawal et al., 2011; Gowers et al., 2016). For the analysis of the monomers, the 4 × 1.0 μs simulations per protein (ignoring the initial simulation) were concatenated. In the case of the dimers, the 3 × 2.0 μs simulations per protein were joined for analysis. The determination of the secondary structure was done with the DSSP (Define Secondary Structure of Proteins)

program (Kabsch and Sander, 1983) as available *via* GROMACS. The PPII-helix structures were determined based on the dihedral angles of the protein backbone, ϕ and ψ, which fall into the range of $-104 \leq \phi \leq -46$ and $116 \leq \psi \leq 174$ for a PPII helix (Mansiaux et al., 2011; Yu et al., 2021). Previous studies have demonstrated that the assignment of the PPII helix using either DSSP-PPII or the backbone dihedral-angles approach provide highly similar results (Jephthah et al., 2021; McIvor et al., 2022). Conformational clustering of the trajectories was accomplished using the algorithm by Daura et al. (Daura et al., 1999), which is a nearest neighbor algorithm, using the root mean square deviation (RMSD) between all trajectory snapshots together with an RMSD cutoff of 0.5 nm to assign the neighbors. For the RMSD calculations, only the Cα atoms were used for both the alignment and actual calculation. To determine the contacts between residues, we calculated minimum distances for all residue pairs, within the proteins and also between the proteins in the case of the dimers. The time-averaged distances between the residues are presented as distance matrices. The contacts were further analyzed based on their interaction type, i.e., hydrophobic, H-bond, or salt-bridge interactions, using the CONAN software (Mercadante et al., 2018). This tool was also employed to identify correlations in the motions of the proteins in their monomeric form. Further analysis involved the calculation of the radius of gyration (R_g) of the proteins, the distance between the Cα atoms of the protein termini (denoted as end-to-end distance or d_{ee}), and the root mean square fluctuations (RMSFs) of the Cα atoms after alignment to the time-averaged structure. The free energy as a function of R_g and d_{ee} was calculated as $\Delta G(R_g, d_{ee}) = -k_B T[\ln P(R_g, d_{ee}) - \ln P_{\max}(R_g, d_{ee})]$, where k_B is the Boltzmann constant, T is room temperature, $P(R_g, d_{ee})$ is the probability of the protein to have given values (R_g, d_{ee}), and $P_{\max}(R_g, d_{ee})$ is the maximum of that probability distribution. The free energy was also determined as a function of the first two principal components, PC1 and PC2, that were determined from a principal component analysis (PCA) using the Cartesian coordinates of the Cα atoms.

The solvent-accessible surface area (SASA) was calculated for both the whole proteins and selected residues. The SASA of the proteins was further distinguished into the hydrophobic and polar solvent-accessible surface areas using the residue sets (Met, Ala, Phe, Leu, Pro, Gly, Val) and (Asp, Glu, His, Lys, Arg, Ser, Thr, Gln), respectively. The standard errors of the SASA values were determined *via* block averaging with 4 blocks, each containing 1,000 data points.

3 Results and discussion

3.1 The pathogenic Htt-ex1 monomer is more compact and has a higher β-sheet propensity than its non-pathogenic counterpart

In this section, we illustrate the structural differences between Htt-Q_{23} and Htt-Q_{48} monomers by the analysis of various structural properties, including the RMSD, R_g, d_{ee}, secondary structure, and the SASA. The evolution of these structural properties are presented in Supplementary Figures S1–S3.

3.1.1 Secondary structure and structural flexibility

Figures 2A, B indicate significant differences in the secondary structure profiles of the two proteins. Htt-Q_{48} has a greater tendency to adopt β-sheet, bend and turn conformations, while Htt-Q_{23} seems to favor α-helix and random coil conformations. The first ten residues of the Nt_{17} region of Htt-Q_{48} form a stable α-helix, which is contrary to Htt-Q_{23} where these N-terminal residues prefer disordered structures with some bends and turns. However, the following residues of Htt-Q_{23} form an α-helix that extends up to 16 glutamines into the polyQ region. The last five residues of $polyQ_{23}$ are mostly disordered with a negligible β-sheet content between residues 36–39. The polyP regions of Htt-Q_{23} adopt straight PPII helical structures, which is revealed by an analysis of the Ramachandran angles (Supplementary Figure S5). To further evaluate the propensities of the individual residues to adopt a PPII helix, we calculated the backbone dihedral angles ϕ and ψ and used them to calulate the PPII-helix probabilities (Supplementary Figure S6). These probabilities confirm a high propensity of PPII-helix formation in $polyP_{11}$ and $polyP_{10}$ of both proteins. In particular, the average ϕ and ψ angles are very similar for the residues of these regions. A rather high probability of PPII-helix formation is also found for most of the other residues of the PRD of both Htt-Q_{23} and Htt-Q_{48}, yet with a higher number of kinks and turns present in the PRD of Htt-Q_{48}. The α-helical regions encompassing the second half of Nt_{17} and most of the polyQ region of Htt-Q_{23} and most of the Nt_{17} region of Htt-Q_{48} have zero propensity to adopt PPII-helix structures. This suggests a more disordered polyQ region in Htt-Q_{48} than in Htt-Q_{23}, which in the former can also adopt ϕ and ψ angles that fall within the PPII-helix region with an average probability of 44% (Supplementary Figure S6).

These observations are consistent with previous NMR solution results which showed that the wild-type Huntingtin protein Nt_{17} has a propensity to adopt helical structures that extended to the polyQ domains (Baias et al., 2017; Newcombe et al., 2018). The N-terminal helix in Htt-Q_{48} is mainly limited to the Nt_{17} region, while the majority of its polyQ region forms a random coil or a bend/turn conformation. However, the terminal residues of the extended polyQ region shows a high propensity to adopt β-sheet structures. This higher β-sheet content compared to Htt-Q_{23} is consistent with previous experimental and MD simulation results (Nagai et al., 2007; Heck et al., 2014; Kang et al., 2017). These studies suggested that the β-sheet content within the polyQ region is associated with longer polyQ lengths and may play a role in amyloid fibril formation (Nagai et al., 2007; Kang et al., 2017). Experiments with synthesized β-sheet in the polyQ domains showed an increased rate of aggregation (Kar et al., 2013). This suggests that the β-sheet conformations within the polyQ region represent the aggregation-prone structures giving rise to amyloid fibrils (Kar et al., 2013).

We continued with a structural cluster analysis to see how the different secondary structures are arranged in the proteins. The six most populated clusters per protein are shown in Figures 2C, D. Five of the shown cluster structures of Htt-Q_{48} have β-sheet conformations that involve polyQ regions (residues 22–25 and 62–64) and parts of the PRD (residues 86–87), with two or three β-strands per sheet. However, in the fourth cluster structure coil conformations outside the helical Nt_{17} region dominate, indicating that the β-sheet formation in Htt-Q_{48} is transient. Htt-Q_{23} samples a

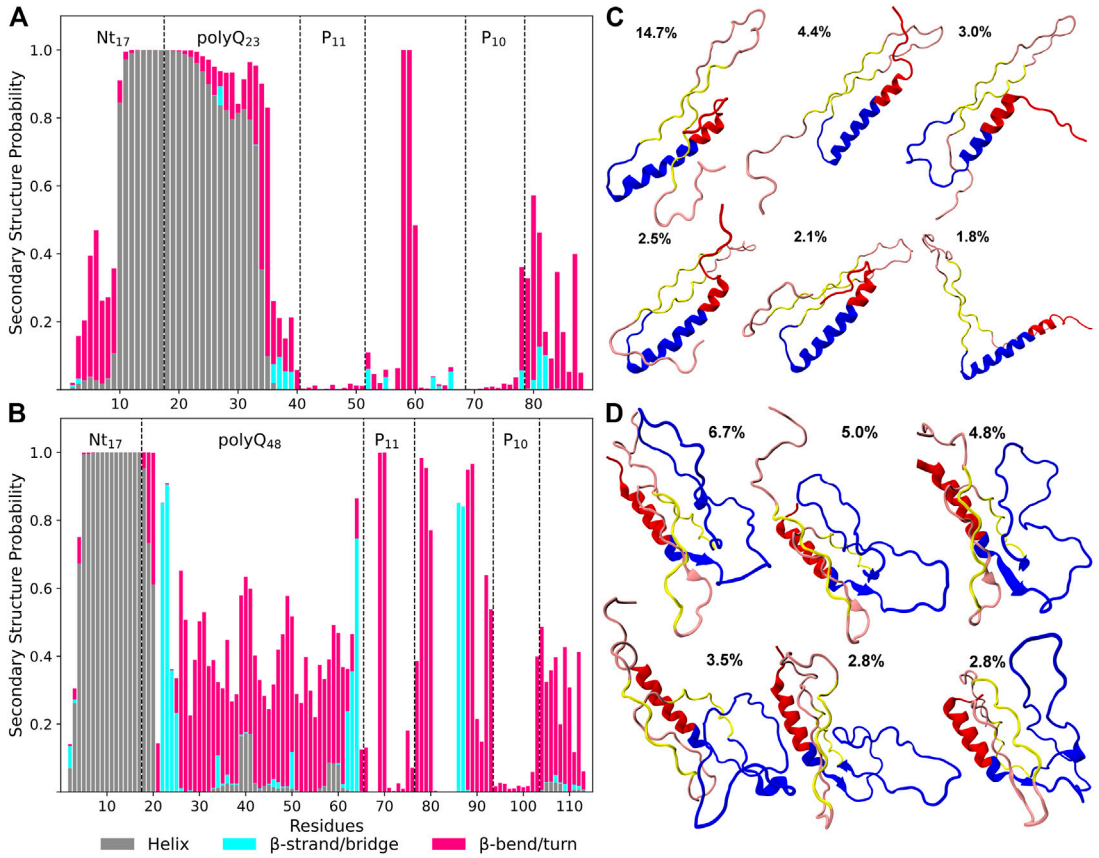

FIGURE 2
Secondary structure preferences per residue of Htt-Q_{23} **(A)** and Htt-Q_{48} **(B)** monomers. The bars represent the additive secondary structure probabilities consisting of α-helix (gray), β-strand/bridge (cyan), and β-turn/bend (magenta). The difference to 1.0 presents the other structures, in particular PPII and random coil. Representative structures were determined from conformational clustering using an RMSD cutoff of 0.5 nm for Htt-Q_{23} **(C)** and Htt-Q_{48} **(D)**. The population of each cluster is given. The Nt_{17} region is shown in red, polyQ in blue, $polyP_{11}$ and $polyP_{10}$ in yellow, and the rest of the PRD in rose.

long helix in the Nt_{17} and most of the polyQ region. The whole protein is less collapsed than Htt-Q_{48}, but rather extended, which results from the long N-terminal α-helix plus the two polyP stretches in PPII conformation that prefer to align in an antiparallel arrangement in Htt-Q_{23}. This PPII-turn-PPII arrangement is also made possible because residues Pro58/Glu59 between $polyP_{11}$ and $polyP_{10}$ allow for a turn to be formed with 100% probability. Apart from that bend, only some further bends and turns are formed in the last ten residues of the PRD of Htt-Q_{23}, while in Htt-Q_{48} several residues along the PRD sequence have a high bend/turn propensity. Together with the disordered polyQ region, this gives the longer protein a more collapsed and also more disordered shape. The Htt-Q_{23} results suggest that the region reaching from P_{11} to P_{10} strongly affects the folding of this protein by forming PPII helices that also have an ordering effect on the polyQ region, leading to helix formation in most of the 23 glutamine residues.

The different structural flexibilities of Htt-Q_{23} and Htt-Q_{48} are also visible in the RMSFs of their Cα atoms (Figure 3A). The RMSFs reveal for Htt-Q_{23} a higher flexibility in the Nt_{17}, where the first ten residues did not form a helix as in Htt-Q_{48}, while the C-terminal of the PRD region is flexible in both proteins. However, the rest of the Htt-Q_{23} is more stable than Htt-Q_{48}. In the latter, the polyQ region is particularly flexible, especially in its middle. These results suggest that longer polyQ tracts stabilize the N-terminal half of Nt_{17}, while inducing disorder and flexibility in the rest of the protein.

3.1.2 Intra-protein contacts and solvent accessibility

To further define the structural differences, we calculated the average distances between the protein residues to characterize the coupling between different domains (Figure 3B). The distance maps confirm the tendency of Htt-Q_{48} to adopt more compact structures as indicated by the significant contacts among Htt-Q_{48} residues. These contacts enable β-sheet formation in Htt-Q_{48} to take place. The contact map of Htt-Q_{23} is defined by the long α-helix in the N-terminal protein part and the alignment between $polyP_{11}$ and $polyP_{10}$ in the C-terminal half. The alignment of the polyP regions in Htt-Q_{23} is stabilized by hydrophobic interactions (Supplementary Figure S7). In addition, the polyQ region sporadically interacts with $polyP_{10}$ and the C-terminal residues (79–90). Conversely, the Nt_{17} of Htt-Q_{48} tends to interact with $polyP_{11}$, the middle of PRD (77–87)—both mediated by hydrophobic interactions –, and the C-terminal (98–110). Moreover, the polyQ region of Htt-Q_{48} can also interact with PRD domains excluding the PRD C-terminal, which involves

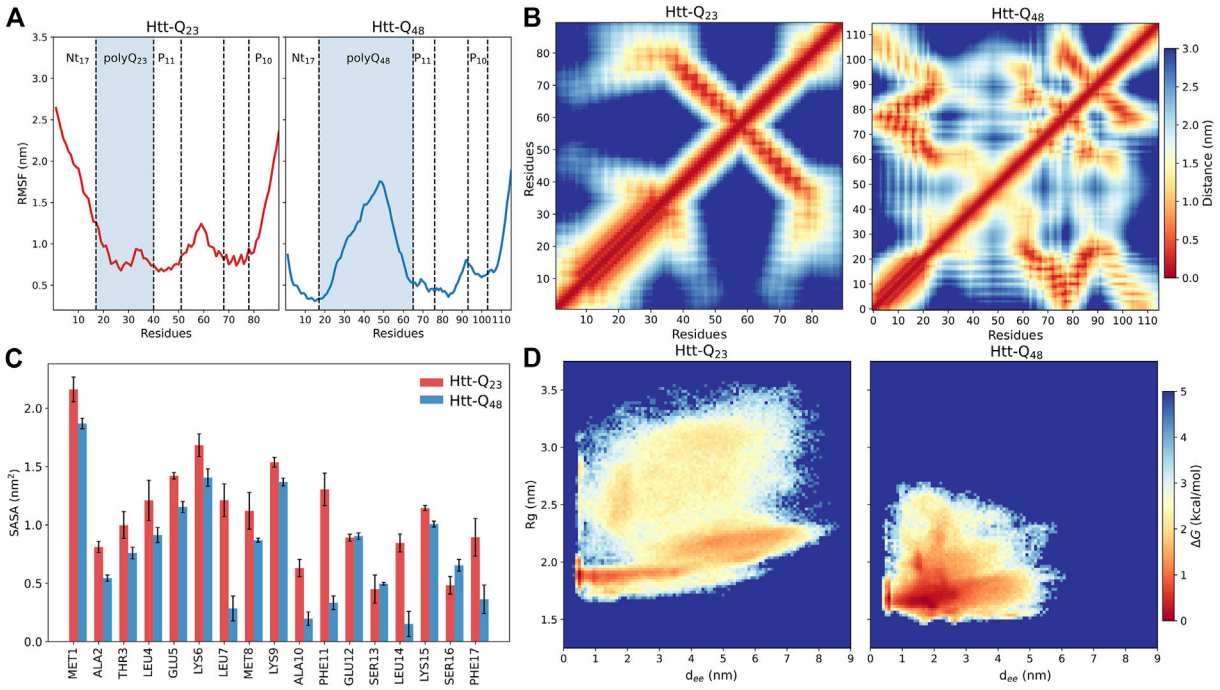

FIGURE 3
Structural properties of Htt-Q_{23} and Htt-Q_{48} monomers. **(A)** The RMSFs of the Cα atoms of Htt-Q_{23} (left) and Htt-Q_{48} (right). The polyQ regions are highlighted by a blue shade. **(B)** Intra-protein distance matrices of Htt-Q_{23} (left) and Htt-Q_{48} (right). The color bar on the right shows the average residue-residue distance (in nm). **(C)** The average SASA of the Nt_{17} residues (with standard errors) for Htt-Q_{23} (red) and Htt-Q_{48} (blue). **(D)** The free energy surface of Htt-Q_{23} (left) and Htt-Q_{48} (right) was calculated using the radius of gyration and the end-to-end distance. The color bar shows the value of the free energy (ΔG) in kcal/mol.

H-bond formation. Strikingly, the polyP regions interact with different domains of Htt-Q_{48}. A remarkable difference between the two proteins is that the Nt_{17} region of Htt-Q_{48} shows strong contacts with PRD regions, which is hardly the case in Htt-Q_{23}. The stability of the intra-protein interaction was further assessed by calculating their probability expressed as percent lifetime (Supplementary Figure S8). This analysis reveals that the hydrophobic interactions between the polyP regions in Htt-Q_{23} and those involving Nt_{17} in Htt-Q_{48} are the most stable interactions. The other very stable interactions are the H-bonds in α-helices that are present in either protein. The terminal interactions in Htt-Q_{48} seem to stabilize the helix in the first ten Nt_{17} residues, and also induce bends/turns with some helical propensity in the last ten C-terminal residues. It has been shown that the helical structure of Nt_{17} is characterized by amphipathic properties (Kelley et al., 2009; Długosz and Trylska, 2011). Thus, the hydrophobic residues of the Nt_{17} minimize their contact with water by interacting with the hydrophobic $polyP_{11}$ and the following PRD mixed region. The different intra-protein interactions in the two proteins also manifest themselves in different coupling between their domains, as revealed by dynamic cross-correlation maps that uncover correlations in the motions (Supplementary Figure S9). In Htt-Q_{23}, all residues spanning from the start of $polyP_{11}$ to the end of $polyP_{10}$ move in a correlated fashion, as they are coupled *via* their strong intra-protein interactions, while this region moves anticorrelated to residues 1–40 containing Nt_{17} and $polyQ_{23}$. In Htt-Q_{48} the correlated and anticorrelated motions are more fragmented along the sequence as a result of the more distributed intra-protein contacts. Most notably, the motions of the $polyQ_{48}$ are mainly anticorrelated to the motions of all other residues on either side of that region.

An analysis of the SASA per residue of Nt_{17} confirms that in Htt-Q_{48} these residues are better shielded from water; this difference is particularly noticeable for the hydrophobic residues Leu4, Leu7, Ala10, Phe11, Leu14, and Phe17 (Figure 3C). This is consistent with previous observations that the Nt_{17}, $polyP_{11}$ and mixed region of Htt-ex1 form a hydrophobic core (Williamson et al., 2010; Długosz and Trylska, 2011) and is confirmed by the hydrophobic and polar SASA of the whole proteins. Htt-Q_{48} exposes more polar residues and prefers to bury the hydrophobic ones compared to Htt-Q_{23} (Supplementary Figure S10A). However, the polar SASA results are not directly comparable, as Htt-Q_{48} has 25 more polar residues than Htt-Q_{23}. Supplementary Figure S10B shows that the polyQ stretch of Htt-Q_{48} is solvent-exposed, shielding the hydrophobic core from the solvent from one side, while on the other side the shielding is done by the hydrophilic face of the N-terminal helix. In Htt-Q_{23}, on the other hand, the hydrophobic residues cluster around the polyP PPII helices and interact with the hydrophobic residues of the Nt_{17} helix (Supplementary Figure S10B), leaving them more solvent-exposed as the 23 glutamine residues are not enough to cover them in that geometry. Nonetheless, this protein structure is stable as the pronounced intra-protein contacts attest. Previous studies suggested that the increase of the interaction surface of the polyQ tract could enhance the protein-protein interactions in the

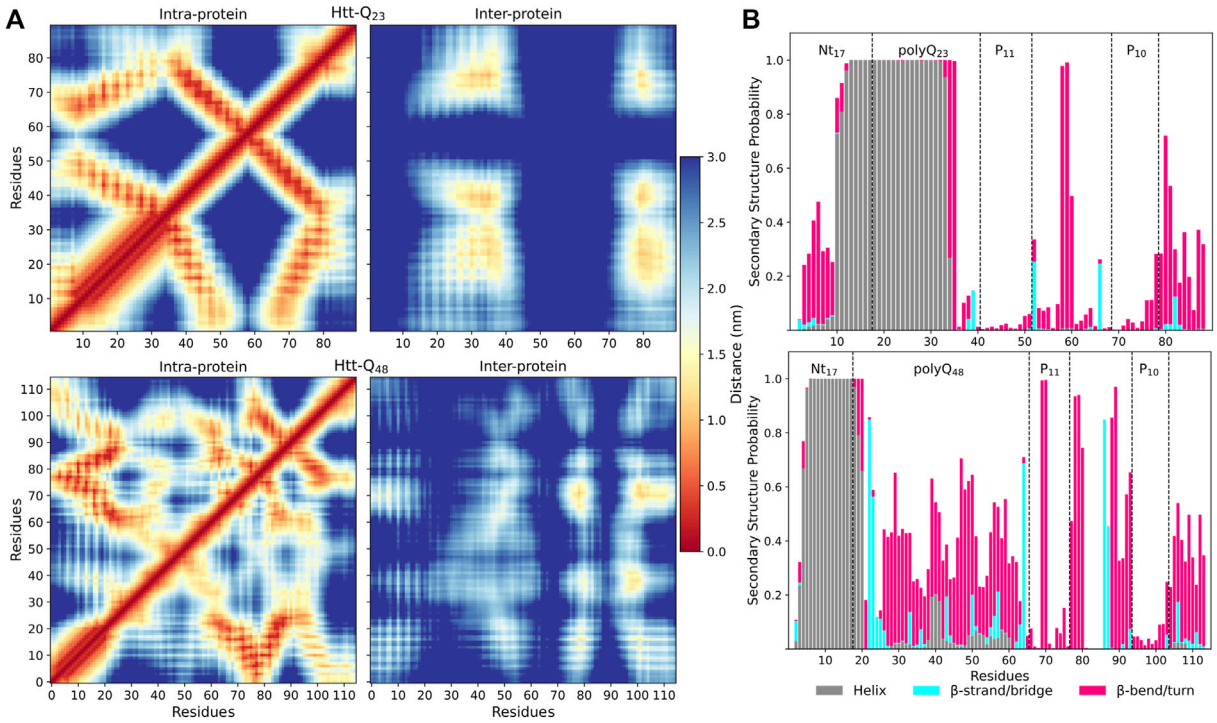

FIGURE 4
Residue-residue contacts and secondary structures in Htt-Q_{23} and Htt-Q_{48} dimers. **(A)** The intra- and inter-protein contacts between residues for Htt-Q_{23} (top) and Htt-Q_{48} (bottom) dimers. The intra-protein contacts within protein 1 are shown below the main diagonal and within protein 2 above the main diagonal. The color bar on the right shows the average distances (in nm). **(B)** The time- and protein-averaged probability of secondary structure per residue for Htt-Q_{23} (top) and Htt-Q_{48} (bottom) dimers. The bars represent the additive secondary structure probabilities consisting of α-helix (gray), β-strand/bridge (cyan), and β-turn/bend (magenta). The difference to 1.0 presents the other structures, in particular PPII and random coil.

cell by enhancing the binding (Warner et al., 2017; Newcombe et al., 2018), which according to the solvent exposure of the polyQ tract seen here for Htt-Q_{48} seems plausible. Overall, the intra-protein interactions of Htt-ex1 are highly associated with the increase of polyQ length, and are mainly governed by the increased protein flexibility in that region, allowing the formation of a hydrophobic core in Htt-Q_{48}.

3.1.3 Conformational variability

To highlight the conformational differences between the two proteins and get better insights into how the elongated polyQ tract affects the conformational energy landscape, we calculated the free energy surface (FES) using the radius of gyration and the end-to-end distance as order parameters. Figure 3D illustrates the structural differences between the two proteins. The FES of Htt-Q_{23} shows that it covers a wider range of the conformational (R_g, d_{ee}) space than Htt-Q_{48}, which populates regions with more compact structures, confirming our observations made thus far. Htt-Q_{23} prefers conformations with larger R_g values and end-to-end distances due to its elongated shape. Its greater flexibility results mainly from the larger conformational freedom of the first ten residues, that together with the C-terminal motions lead to a large distribution of the R_g and d_{ee} values in Htt-Q_{23}. As a result, there are structures where the two termini are very close to each other (e.g., clusters 1 and 5 in Figure 2C), structures where they are directed away from each other (cluster 2), and the intermediate cases (clusters 3, 4 and 6). Nonetheless, it should be reiterated that the overall fold of Htt-Q_{23} defined by the N-terminal helix and the PPII-turn-PPII motif reaching from polyP$_{11}$ to polyP$_{10}$ is rather stable despite the FES suggesting otherwise. Since d_{ee} and R_g are partially correlated, although in Htt-Q_{23} small d_{ee} values are possible for large R_g and the converse is also true, i.e., small R_g with large d_{ee} values, we also created FESs as a function of the two main pincipal (orthogonal) components, PC1 and PC2 (Supplementary Figure S11A). These suggest that the two proteins cover an equally large conformational space. However, the two FESs cannot be directly compared because the principal components of Htt-Q_{23} and Htt-Q_{48} are not identical, as the two proteins have different sizes. Projection of the individual trajectories onto PC1-PC2 space (Supplementary Figure S11B) shows that Htt-Q_{48} tends toward more compact structures with increasing simulation time, whereas the shorter protein continues to explore the entire conformational space. In summary, the main conclusion from the FES analysis is that Htt-Q_{48} prefers to form more collapsed structures, which is in agreement with its numerous intra-protein contacts. Several experiments and simulations studies suggested that the polyQ tracts prefer to adopt collapsed conformations due to their poor water solubility (Kang et al., 2017). However, this only partly agrees to our findings as we find the long polyQ tract of Htt-Q_{48} to shield the more hydrophobic residues from the water, giving rise to more collapsed conformations than on average seen for Htt-Q_{23} where the polyQ tract is not long enough to provide sufficient shielding and forms a helix instead.

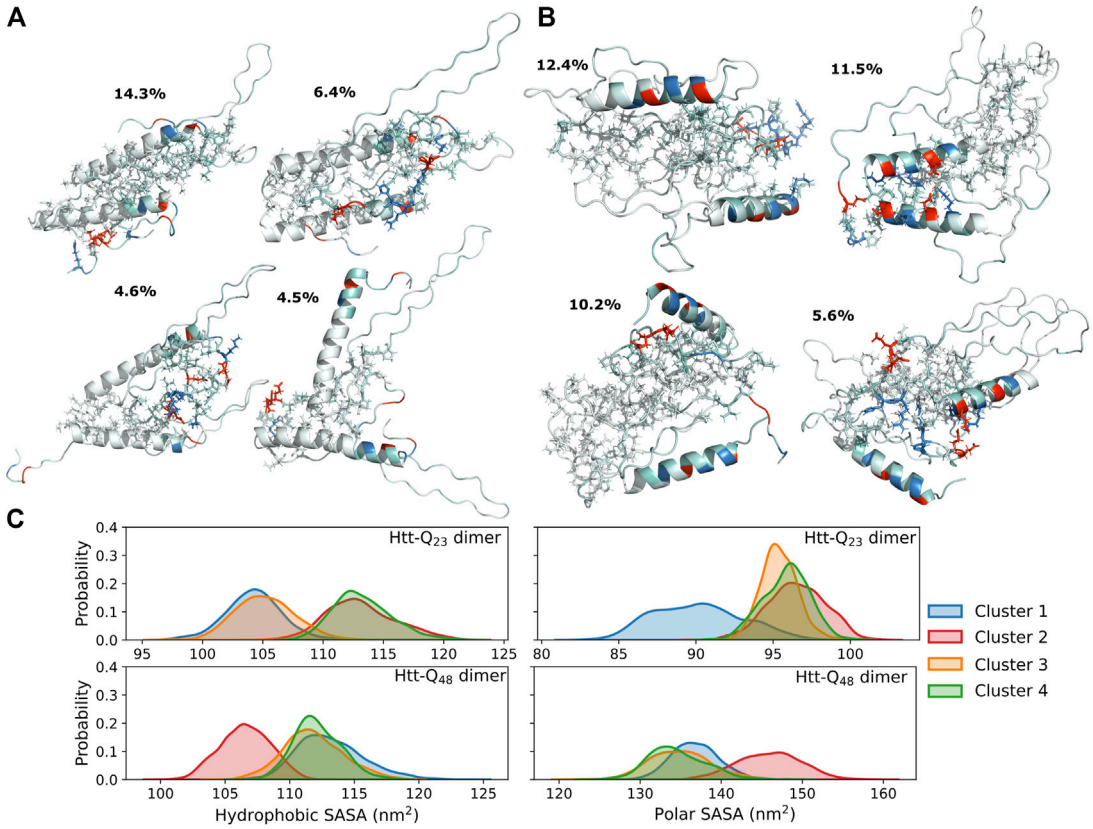

FIGURE 5
Structures and SASAs of dimers. The four most populated cluster structures for the Htt-Q_{23} **(A)** and Htt-Q_{48} **(B)** dimer. The proteins are shown as cartoons and colored red for the negatively charged residues, blue for the positively charged ones, and white otherwise. The side chains at the protein interfaces are shown as sticks. **(C)** The distribution of the hydrophobic (right) and polar (left) SASAs of the four most populated clusters for Htt-Q_{23} (top) and Htt-Q_{48} (bottom) dimers are shown. The results for the four clusters are shown in different colors according to the color mapping on the right.

3.2 Non-pathogenic and pathogenic Htt-ex1 dimerize in distinct patterns

We investigated the dimerization between the Htt-ex1 monomers to obtain information on the first oligomerization step. For either protein, we observed an association between the monomers within ~100 ns of the simulation time (Supplementary Figure S12). The dimerization modes are analyzed in more detail in the following.

3.2.1 Interaction interfaces and dimer structures

To understand how the proteins interact with each other, we calculated intra- and inter-protein residue-residue distance matrices (Figure 4A). They reveal different interaction patterns within the Htt-Q_{23} and Htt-Q_{48} dimers. In Htt-Q_{23}, there are four pronounced interaction regions between the two proteins: i) polyQ/polyP$_{11}$ with polyQ/polyP$_{11}$, ii) polyQ/polyP$_{11}$ with polyP$_{10}$/PRD (79–90), iii) polyP$_{10}$/PRD(79–90) with polyQ/plolyP$_{11}$, iv)polyP$_{10}$/PRD(79–90) with ployP$_{10}$/PRD(79–90). Nt$_{17}$ and PRD(52–68) are not involved at all in the dimerization of Htt-Q_{23}. Htt-Q_{48}, on the other hand, features intermolecular interactions across the whole sequence, which are, however, less clear-cut. This indicates that different dimer configurations are possible for Htt-Q_{48}. The most prominent interactions involve the polyP regions, polyP with mixed regions in PRD, and polyP with polyQ. Furthermore, the polyQ is seen to interact with the whole protein, while the Nt$_{17}$ shows interactions with polyP$_{11}$, polyQ, and the C-terminal of the PRD.

The intra-protein interactions in the two proteins of the Htt-Q_{48} dimers are very similar to the ones found for the corresponding monomer (Figure 3B). This implies that the aggregation did not cause major structural transitions, which is confirmed by the secondary structure analysis (Figure 4B vs. Figure 2A). A small difference in Htt-Q_{23} is that dimerization stabilized the Nt$_{17}$/polyQ helix, while in Htt-Q_{48} a small increase in the tendency for β-sheet formation is observed across the polyQ region. For more β-sheet formation as a result of Htt-Q_{48} dimerization to occur, we expect that more simulation time than the 6×2 μs sampled here is required. In Htt-Q_{23}, in addition to the intra-protein contacts that were already present in the monomer, two more contact areas within the proteins emerged: the Nt$_{17}$ and first half of the polyQ region interacts with both polyP$_{11}$ and polyP$_{10}$. This speaks for small reorientations of the secondary-structure elements with respect to each other. In order to further understand the interaction interfaces between the proteins, we performed structural clustering and analyzed the hydrophobic and the polar SASAs of the dominant

dimer clusters. Figures 5A, B show the representative structures of the four most populated clusters. In general, the interaction interfaces between the Htt-Q_{23} proteins are mainly governed by the interactions of the partly amphipathic Nt_{17}/polyQ helices. The two helices tend to be in a parallel orientation to each other, and their interactions add helical stability. The polyP regions, on the other hand, are hardly involved in the interaction interface and instead point away from it. On the other hand, the Htt-Q_{48} dimer clusters reveal a more diverse aggregation behavior with different interaction interfaces where the polyP and the PRD take part in them.

3.2.2 Comparison with experimental observations

Our observations are in agreement with atomic force microscopy (AFM) results that identified oligomeric aggregates of Htt-Q_{20} to be structurally different from those of Htt-ex1 with disease-causing polyQ lengths (Legleiter et al., 2010). The Htt-Q_{20} oligomers did also not transit to form fibril structures. In another study it was shown that the addition of isolated polyP to Htt-ex1 could block the initial aggregation by interacting with Nt_{17} and reduce the ability of headpiece dimerization (Arndt et al., 2019). This speaks for the general involvement of Nt_{17} in the aggregation of Htt-ex1, which is also seen here for both Htt-Q_{23} and Htt-Q_{48}. Another possible aggregation-modifying effect of added polyP is that it can also interact with the polyQ region and stabilize it by promoting it to form PPII helix (Darnell et al., 2007; Darnell et al., 2009), which could prevent the nucleation in the oligomers or promote aggregation pathways that are not mediated by the Nt_{17} headpieces (Arndt et al., 2019). Pandey et al. proposed that the Htt-ex1 aggregation is initiated *via* the Nt_{17} headpiece, suggesting that the aggregation pathway of synthetic Htt-ex1 protein into oligomers starts with the self-assembly into an α-helix rich oligomer with helical Nt_{17} in the core, followed by β-sheet formation within the polyQ tracts. The final and slowest step is the structural maturation of the PRD (Pandey et al., 2018). Through the simulations performed here, we were able to capture the event of Nt_{17} headpiece interactions for Htt-Q_{48} (Figure 5B, cluster 2). Dimerization *via* the Nt_{17} helices results in a reduction of the hydrophobic surface area, while it increases the polar surface area, compared to the other dimerization modes (Figure 5C). This difference is expected to make the helix-mediated dimer more stable. Mutational studies have shown that altering the hydrophobic residues of Nt_{17} to negatively charged residues or by serine phosphorylation, strongly affects the aggregation kinetics by slowing down the aggregation (Gu et al., 2009; Thakur et al., 2009; Mishra et al., 2012). However, Htt-ex1 aggregation was not prevented by these amino-acid changes. In addition to that, according to the aggregation mechanism investigated by Williamson et al., metastable aggregates start to form *via* Nt_{17} packed cores and the polyQ regions being excluded from these cores. These initial aggregates develop then to amyloid nuclei composed of Nt_{17} headpieces and polyQ tracts (Williamson et al., 2010). Based on our dimer simulation results, this aggregation mechanism appears fully plausible. We further find that longer polyQ tracts are needed for amyloid formation as for shorter polyQ sequences, the Nt_{17} helix extends far into the polyQ region and is resistant against β-sheet formation. This observation is in agreement with recent NMR spectroscopy results (Elena-Real et al., 2022).

4 Conclusion

In this work, we carried out multiple MD simulations to investigate the monomer and dimer structures of a pathogenic Htt-ex1 protein (Htt-Q_{48}) and a non-pathogenic (Htt-Q_{23}) counterpart. We found rather different monomer structures for the two proteins. One of the most relevant distinctions is that in Htt-Q_{23} the N-terminal helix involves not only Nt_{17} but also the majority of the $polyQ_{23}$ residues. This can be explained by the rather high helix propensity of glutamine, which comes seventh on the helix propensity scale of the twenty proteinogenic amino acids (Pace and Scholtz, 1998). However, for longer polyQ sequences as in Htt-Q_{48}, this propensity is not sufficient and the polyQ tract becomes disordered. Here, it should also be considered that the length of an α-helix always results from a competition between α-helix folding, unfolding into a random coil and the formation of higher-order tertiary structures, and that this competition leads to naturally favored α-helix lengths of 9–17 amino acids (Qin et al., 2013). It is thus fully understandable that with increasing polyQ length, this region becomes disordered. As a result, more intra-protein interactions develop, which in turn can lead to β-sheet formation, even though glutamine has only a medium propensity to be in a β-sheet (Shingate and Sowdhamini, 2012). Here, we observed an increase in bend/turn conformations across the polyQ and PRD regions of Htt-Q_{48}, which yielded compact protein structures involving small intra-protein β-sheets. In both proteins, the polyP regions mainly adopted PPII structures. However, in Htt-Q_{48} they were disrupted by kinks, which is in line with the compact Htt-Q_{48} conformations, whereas in Htt-Q_{23} the $polyP_{11}$ and $polyP_{10}$ are straight segments that are aligned antiparallel to each other. As a result, Htt-Q_{23} has an elongated shape resulting from the 25 residue-long N-terminal helix and the PPII-turn-PPII motif. This overall fold appeared rather stable and structural fluctuations mainly derived from the first and last ten residues of Htt-Q_{23}.

With regard to dimerization, we found that for both Htt-Q_{23} and Htt-Q_{48} this process involves the N-terminal helix, yet to a larger degree in Htt-Q_{23}. Here, the helix-helix interaction further stabilized the helix, especially in its second part that involves residues from the $polyQ_{23}$ region, which explains why this region is resistant against β-sheet formation. Moreover, interaction through the amphipathic α-helix Nt_{17} headpieces minimizes the hydrophobic surface, making this dimer conformation particularly stable. In Htt-Q_{48}, on the other hand, different dimerization modes were observed, of which the one involving the two N-terminal helices is one possibility. However, we expect the Nt_{17} helix mediated dimer to be more stable than the others due to the larger burial of hydrophobic residues. An important aspect of Htt-Q_{48} dimerization is that the disordered polyQ regions are considerably involved in the process, allowing β-sheets to form on the long time scale (not simulated here). However, this structural transition is slowed down by visible interactions with the PRD. For both Htt-Q_{23} and Htt-Q_{48} the polyP regions are not involved in the dimerization, though their avoidance of the dimer interaction interface is more pronounced in Htt-Q_{23} due to its particular elongated protein shape, allowing the polyP regions to keep away from it.

Overall, we observed that polyQ extension affects the conformational ensemble of Htt-ex1 already at the monomer level and has a direct impact on dimerization. Our results provide insights into the roles of the flanking domains of polyQ in modulating the conformations and the aggregation pathways, thereby providing explanations for several experimental findings at the atomistic level.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

MK, AS-A, and BS contributed to design of the study. MK performed and analyzed the simulations. MK wrote the first draft of the manuscript. All authors contributed to the manuscript revision, read, and approved.

Funding

The project was funded *via* the Palestinian-German Science Bridge financed by the German Federal Ministry of Education and Research (BMBF).

Acknowledgments

The authors gratefully acknowledge the computing time granted through JARA on the supercomputer Jureca at Forschungszentrum Jülich (project AMYLOID-MSM).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2023.1143353/full#supplementary-material

References

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1, 19–25. doi:10.1016/j.softx.2015.06.001

Arndt, J. R., Chaibva, M., Beasley, M., Kiani Karanji, A., Ghassabi Kondalaji, S., Khakinejad, M., et al. (2019). Nucleation inhibition of huntingtin protein (htt) by polyproline ppii helices: A potential interaction with the n-terminal α-helical region of htt. *Biochemistry* 59, 436–449. doi:10.1021/acs.biochem.9b00689

Baias, M., Smith, P. E., Shen, K., Joachimiak, L. A., Zerko, S., Kozminski, W., et al. (2017). Structure and dynamics of the huntingtin exon-1 n-terminus: A solution nmr perspective. *J. Am. Chem. Soc.* 139, 1168–1176. doi:10.1021/jacs.6b10893

Bhattacharyya, A., Thakur, A. K., Chellgren, V. M., Thiagarajan, G., Williams, A. D., Chellgren, B. W., et al. (2006). Oligoproline effects on polyglutamine conformation and aggregation. *J. Mol. Biol.* 355, 524–535. doi:10.1016/j.jmb.2005.10.053

Bugg, C. W., Isas, J. M., Fischer, T., Patterson, P. H., and Langen, R. (2012). Structural features and domain organization of huntingtin fibrils. *J. Biol. Chem.* 287, 31739–31746. doi:10.1074/jbc.M112.353839

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *Chem. Phys.* 126, 014101. doi:10.1063/1.2408420

Chellgren, B. W., Miller, A.-F., and Creamer, T. P. (2006). Evidence for polyproline ii helical structure in short polyglutamine tracts. *J. Mol. Biol.* 361, 362–371. doi:10.1016/j.jmb.2006.06.044

Chen, S., Berthelier, V., Yang, W., and Wetzel, R. (2001). Polyglutamine aggregation behavior *in vitro* supports a recruitment mechanism of cytotoxicity. *J. Mol. Biol.* 311, 173–182. doi:10.1006/jmbi.2001.4850

Chow, W. N. V., Luk, H. W., Chan, H. Y. E., and Lau, K.-F. (2012). Degradation of mutant huntingtin via the ubiquitin/proteasome system is modulated by fe65. *Biochem. J.* 443, 681–689. doi:10.1042/BJ20112175

Crick, S. L., Jayaraman, M., Frieden, C., Wetzel, R., and Pappu, R. V. (2006). Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc. Natl. Acad. Sci. U.S.A.* 103, 16764–16769. doi:10.1073/pnas.0608175103

Darnell, G., Orgel, J. P., Pahl, R., and Meredith, S. C. (2007). Flanking polyproline sequences inhibit β-sheet structure in polyglutamine segments by inducing ppii-like helix structure. *J. Mol. Biol.* 374, 688–704. doi:10.1016/j.jmb.2007.09.023

Darnell, G. D., Derryberry, J., Kurutz, J. W., and Meredith, S. C. (2009). Mechanism of cis-inhibition of polyq fibrillation by polyp: Ppii oligomers and the hydrophobic effect. *Biophys. J.* 97, 2295–2305. doi:10.1016/j.bpj.2009.07.062

Daura, X., Gademann, K., Jaun, B., Seebach, D., Gunsteren, W., and Mark, A. (1999). Peptide folding: When simulation meets experiment. *Angew. Chem. Int. Ed.* 38, 236–240. doi:10.1002/(SICI)1521-3773(19990115)38:1/2⟨236::AID-ANIE236⟩3.0.CO;2-M

Długosz, M., and Trylska, J. (2011). Secondary structures of native and pathogenic huntingtin n-terminal fragments. *J. Phys. Chem. B* 115, 11597–11608. doi:10.1021/jp206373g

Duennwald, M. L., Jagadish, S., Giorgini, F., Muchowski, P. J., and Lindquist, S. (2006a). A network of protein interactions determines polyglutamine toxicity. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11051–11056. doi:10.1073/pnas.0604548103

Duennwald, M. L., Jagadish, S., Muchowski, P. J., and Lindquist, S. (2006b). Flanking sequences profoundly alter polyglutamine toxicity in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11045–11050. doi:10.1073/pnas.0604547103

Elena-Real, C. A., Sagar, A., Urbanek, A., Popovic, M., Morató, A., Estaña, A., et al. (2022). The structure of pathogenic huntingtin exon-1 defines the bases of its aggregation propensity. *bioRxiv*. doi:10.1101/2022.10.25.513661

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). A smooth particle mesh ewald method. *Chem. Phys.* 103, 8577–8593. doi:10.1063/1.470117

Gkekas, I., Gioran, A., Boziki, M. K., Grigoriadis, N., Chondrogianni, N., and Petrakis, S. (2021). Oxidative stress and neurodegeneration: Interconnected processes in polyq diseases. *Antioxidants* 10, 1450. doi:10.3390/antiox10091450

Gowers, R. J., Linke, M., Barnoud, J., Reddy, T. J. E., Melo, M. N., Seyler, S. L., et al. (2016). "MDAnalysis: A Python package for the rapid analysis of molecular dynamics simulations," in Proceedings of the 15th Python in Science Conference. Editors S. Benthall and S. Rostrup, 98–105. doi:10.25080/Majora-629e541a-00e

Gu, X., Greiner, E. R., Mishra, R., Kodali, R., Osmand, A., Finkbeiner, S., et al. (2009). Serines 13 and 16 are critical determinants of full-length human mutant huntingtin induced disease pathogenesis in hd mice. *Neuron* 64, 828–840. doi:10.1016/j.neuron.2009.11.020

Gusella, J. F., and MacDonald, M. E. (2000). Molecular genetics: Unmasking polyglutamine triggers in neurodegenerative disease. *Nat. Rev. Neurosci.* 1, 109–115. doi:10.1038/35039051

Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., and Hutchison, G. R. (2012). Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics.* 4, 17. doi:10.1186/1758-2946-4-17

Heck, B. S., Doll, F., and Hauser, K. (2014). Length-dependent conformational transitions of polyglutamine repeats as molecular origin of fibril initiation. *Biophys. Chem.* 185, 47–57. doi:10.1016/j.bpc.2013.11.008

Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997). Lincs: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 18, 1463–1472. doi:10.1002/(sici)1096-987x(199709)18:12<1463::aid-jcc4>3.0.co;2-h

Hoop, C. L., Lin, H.-K., Kar, K., Hou, Z., Poirier, M. A., Wetzel, R., et al. (2014). Polyglutamine amyloid core boundaries and flanking domain dynamics in huntingtin fragment fibrils determined by solid-state nuclear magnetic resonance. *Biochemistry* 53, 6653–6666. doi:10.1021/bi501010q

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B. L., et al. (2017). Charmm36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods.* 14, 71–73. doi:10.1038/nmeth.4067

Isas, J. M., Langen, R., and Siemer, A. B. (2015). Solid-state nuclear magnetic resonance on the static and dynamic domains of huntingtin exon-1 fibrils. *Biochemistry* 54, 3942–3949. doi:10.1021/acs.biochem.5b00281

Jayaraman, M., Kodali, R., Sahoo, B., Thakur, A. K., Mayasundari, A., Mishra, R., et al. (2012). Slow amyloid nucleation via α-helix-rich oligomeric intermediates in short polyglutamine-containing huntingtin fragments. *J. Mol. Biol.* 415, 881–899. doi:10.1016/j.jmb.2011.12.010

Jephthah, S., Pesce, F., Lindorff-Larsen, K., and Skepö, M. (2021). Force field effects in simulations of flexible peptides with varying polyproline ii propensity. *J. Chem. Theory Comput.* 17, 6634–6646. doi:10.1021/acs.jctc.1c00408

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637. doi:10.1002/bip.360221211

Kang, H., Vázquez, F. X., Zhang, L., Das, P., Toledo-Sherman, L., Luan, B., et al. (2017). Emerging β-sheet rich conformations in supercompact huntingtin exon-1 mutant structures. *J. Am. Chem. Soc.* 139, 8820–8827. doi:10.1021/jacs.7b00838

Kar, K., Hoop, C. L., Drombosky, K. W., Baker, M. A., Kodali, R., Arduini, I., et al. (2013). β-hairpin-mediated nucleation of polyglutamine amyloid formation. *J. Mol. Biol.* 425, 1183–1197. doi:10.1016/j.jmb.2013.01.016

Kelley, N. W., Huang, X., Tam, S., Spiess, C., Frydman, J., and Pande, V. S. (2009). The predicted structure of the headpiece of the huntingtin protein and its implications on huntingtin aggregation. *J. Mol. Biol.* 388, 919–927. doi:10.1016/j.jmb.2009.01.032

Kim, M. W., Chelliah, Y., Kim, S. W., Otwinowski, Z., and Bezprozvanny, I. (2009). Secondary structure of huntingtin amino-terminal region. *Struc* 17, 1205–1212. doi:10.1016/j.str.2009.08.002

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. doi:10.4161/pri.23807

Klein, F. A., Pastore, A., Masino, L., Zeder-Lutz, G., Nierengarten, H., Oulad-Abdelghani, M., et al. (2007). Pathogenic and non-pathogenic polyglutamine tracts have similar structural properties: Towards a length-dependent toxicity gradient. *J. Mol. Biol.* 371, 235–244. doi:10.1016/j.jmb.2007.05.028

Lakhani, V. V., Ding, F., and Dokholyan, N. V. (2010). Polyglutamine induced misfolding of huntingtin exon1 is modulated by the flanking sequences. *PLoS Comput. Biol.* 6, e1000772. doi:10.1371/journal.pcbi.1000772

Legleiter, J., Mitchell, E., Lotz, G. P., Sapp, E., Ng, C., DiFiglia, M., et al. (2010). Mutant huntingtin fragments form oligomers in a polyglutamine length-dependent manner *in vitro* and *in vivo*. *J. Biol. Chem.* 285, 14777–14790. doi:10.1074/jbc.M109.093708

Li, S.-H., and Li, X.-J. (2004). Huntingtin–protein interactions and the pathogenesis of huntington's disease. *Trends Genet.* 20, 146–154. doi:10.1016/j.tig.2004.01.008

MacDonald, M. E., Ambrose, C. M., Duyao, M. P., Myers, R. H., Lin, C., Srinidhi, L., et al. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on huntington's disease chromosomes. *Cell* 72, 971–983. doi:10.1016/0092-8674(93)90585-E

Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetherington, C., et al. (1996). Exon 1 of the hd gene with an expanded cag repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* 87, 493–506. doi:10.1016/S0092-8674(00)81369-0

Mansiaux, Y., Joseph, A. P., Gelly, J.-C., and de Brevern, A. G. (2011). Assignment of polyproline ii conformation and analysis of sequence – structure relationship. *PLOS ONE* 6, e18401–e18415. doi:10.1371/journal.pone.0018401

McIvor, J. A. P., Larsen, D. S., and Mercadante, D. (2022). Simulating polyproline ii-helix-rich peptides with the latest kirkwood–buff force field: A direct comparison with amber and charmm. *J. Am. Phys. Chem. B* 126, 7833–7846. doi:10.1021/acs.jpcb.2c03837

Mercadante, D., Gräter, F., and Daday, C. (2018). Conan: A tool to decode dynamical information from molecular interaction maps. *Biophys. J.* 114, 1267–1273. doi:10.1016/j.bpj.2018.01.033

Michalek, M., Salnikov, E. S., and Bechinger, B. (2013). Structure and topology of the huntingtin 1–17 membrane anchor by a combined solution and solid-state nmr approach. *Biophys. J.* 105, 699–710. doi:10.1016/j.bpj.2013.06.030

Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., and Beckstein, O. (2011). Mdanalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* 32, 2319–2327. doi:10.1002/jcc.21787

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

Mishra, R., Hoop, C. L., Kodali, R., Sahoo, B., van der Wel, P. C., and Wetzel, R. (2012). Serine phosphorylation suppresses huntingtin amyloid accumulation by altering protein aggregation properties. *J. Mol. Biol.* 424, 1–14. doi:10.1016/j.jmb.2012.09.011

Moldovean, S. N., and Chiş, V. (2019). Molecular dynamics simulations applied to structural and dynamical transitions of the huntingtin protein: A review. *ACS Chem. Neurosci.* 11, 105–120. doi:10.1021/acschemneuro.9b00561

Müller, K., and Brown, L. D. (1979). Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theor. Chim. Acta* 53, 75–93. doi:10.1007/BF00547608

Nagai, Y., Inui, T., Popiel, H. A., Fujikake, N., Hasegawa, K., Urade, Y., et al. (2007). A toxic monomeric conformer of the polyglutamine protein. *Nat. Struct. Mol. Biol.* 14, 332–340. doi:10.1038/nsmb1215

Newcombe, E. A., Ruff, K. M., Sethi, A., Ormsby, A. R., Ramdzan, Y. M., Fox, A., et al. (2018). Tadpole-like conformations of huntingtin exon 1 are characterized by conformational heterogeneity that persists regardless of polyglutamine length. *J. Mol. Biol.* 430, 1442–1458. doi:10.1016/j.jmb.2018.03.031

Pace, C. N., and Scholtz, J. M. (1998). A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.* 75, 422–427. doi:10.1016/S0006-3495(98)77529-0

Pandey, N. K., Isas, J. M., Rawat, A., Lee, R. V., Langen, J., Pandey, P., et al. (2018). The 17-residue-long n terminus in huntingtin controls stepwise aggregation in solution and on membranes via different mechanisms. *J. Biol. Chem.* 293, 2597–2605. doi:10.1074/jbc.M117.813667

Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 52, 7182–7190. doi:10.1063/1.328693

Parrinello, M., and Rahman, A. (1982). Strain fluctuations and elastic constants. *Chem. Phys.* 76, 2662–2666. doi:10.1063/1.443248

Paul, A., Samantray, S., Anteghini, M., Khaled, M., and Strodel, B. (2021). Thermodynamics and kinetics of the amyloid-β peptide revealed by markov state models based on md data in agreement with experiment. *Chem. Sci.* 12, 6652–6669. doi:10.1039/D0SC04657D

Perutz, M. F., Johnson, T., Suzuki, M., and Finch, J. T. (1994). Glutamine repeats as polar zippers: Their possible role in inherited neurodegenerative diseases. *Proc. Natl. Acad. Sci. U.S.A.* 91, 5355–5358. doi:10.1073/pnas.91.12.5355

Priya, S. B., and Gromiha, M. M. (2019). Structural insights into the aggregation mechanism of huntingtin exon 1 protein fragment with different polyq-lengths. *J. Cell. Biochem.* 120, 10519–10529. doi:10.1002/jcb.28338

Qin, Z.-H., Wang, Y., Sapp, E., Cuiffo, B., Wanker, E., Hayden, M. R., et al. (2004). Huntingtin bodies sequester vesicle-associated proteins by a polyproline-dependent interaction. *J. Neurosci.* 24, 269–281. doi:10.1523/JNEUROSCI.1409-03.2004

Qin, Z., Fabre, A., and Buehler, M. (2013). Structure and mechanism of maximum stability of isolated alpha-helical protein domains at a critical length scale. *Eur. Phys. J. E* 36, 53. doi:10.1140/epje/i2013-13053-8

Robustelli, P., Piana, S., and Shaw, D. (2018). Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U.S.A.* 115, E4758–E4766. doi:10.1073/pnas.1800690115

Sahoo, B., Singer, D., Kodali, R., Zuchner, T., and Wetzel, R. (2014). Aggregation behavior of chemically synthesized, full-length huntingtin exon1. *Biochem. J.* 53, 3897–3907. doi:10.1021/bi500300c

Samantray, S., Yin, F., Kav, B., and Strodel, B. (2020). Different force fields give rise to different amyloid aggregation pathways in molecular dynamics simulations. *J. Chem. Inf. Model.* 60, 6462–6475. PMID: 33174726. doi:10.1021/acs.jcim.0c01063

Saudou, F., and Humbert, S. (2016). The biology of huntingtin. *Neuron* 89, 910–926. doi:10.1016/j.neuron.2016.02.003

Schrödinger, L. L. C. (2022). *The pymol molecular graphics system*. Schrödinger. version 2.5.0.

Shingate, P., and Sowdhamini, R. (2012). Analysis of domain-swapped oligomers reveals local sequence preferences and structural imprints at the linker regions and swapped interfaces. *PloS one* 7, e39305. doi:10.1371/journal.pone.0039305

Sivanandam, V., Jayaraman, M., Hoop, C. L., Kodali, R., Wetzel, R., and van der Wel, P. C. (2011). The aggregation-enhancing huntingtin n-terminus is helical in amyloid fibrils. *J. Am. Chem. Soc.* 133, 4558–4566. doi:10.1021/ja110715f

Takeuchi, T., and Nagai, Y. (2017). Protein misfolding and aggregation as a therapeutic target for polyglutamine diseases. *Brain Sci.* 7, 128. doi:10.3390/brainsci7100128

Tam, S., Spiess, C., Auyeung, W., Joachimiak, L., Chen, B., Poirier, M. A., et al. (2009). The chaperonin tric blocks a huntingtin sequence element that promotes the conformational switch to aggregation. *Nat. Struct. Mol. Biol.* 16, 1279–1285. doi:10.1038/nsmb.1700

Thakur, A. K., and Wetzel, R. (2002). Mutational analysis of the structural organization of polyglutamine aggregates. *Proc. Natl. Acad. Sci. U.S.A.* 99, 17014–17019. doi:10.1073/pnas.252523899

Thakur, A. K., Jayaraman, M., Mishra, R., Thakur, M., Chellgren, V. M., Byeon, I.-J. L., et al. (2009). Polyglutamine disruption of the huntingtin exon 1 n terminus triggers a complex aggregation mechanism. *Nat. Struct. Mol. Biol.* 16, 380–389. doi:10.1038/nsmb.1570

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005). Gromacs: Fast, flexible, and free. *J. Comput. Chem.* 26, 1701–1718. doi:10.1002/jcc.20291

Van Gunsteren, W. F., and Berendsen, H. J. (1988). A leap-frog algorithm for stochastic dynamics. *Mol. Simul.* 1, 173–185. doi:10.1080/08927028808080941

Vöpel, T., Bravo-Rodriguez, K., Mittal, S., Vachharajani, S., Gnutt, D., Sharma, A., et al. (2017). Inhibition of huntingtin exon-1 aggregation by the molecular tweezer clr01. *J. Am. Chem. Soc.* 139, 5640–5643. doi:10.1021/jacs.6b11039

Warner, J. B., IV, Ruff, K. M., Tan, P. S., Lemke, E. A., Pappu, R. V., and Lashuel, H. A. (2017). Monomeric huntingtin exon 1 has similar overall structural features for wild-type and pathological polyglutamine lengths. *J. Am. Chem. Soc.* 139, 14456–14469. doi:10.1021/jacs.7b06659

Williams, A. J., and Paulson, H. L. (2008). Polyglutamine neurodegeneration: Protein misfolding revisited. *Trends Neurosci.* 31, 521–528. doi:10.1016/j.tins.2008.07.004

Williamson, T. E., Vitalis, A., Crick, S. L., and Pappu, R. V. (2010). Modulation of polyglutamine conformations and dimer formation by the n-terminus of huntingtin. *J. Mol. Biol.* 396, 1295–1309. doi:10.1016/j.jmb.2009.12.017

Yu, L., Li, D.-W., and Brüschweiler, R. (2021). Systematic differences between current molecular dynamics force fields to represent local properties of intrinsically disordered proteins. *J. Am. Phys. Chem. B* 125, 798–804. PMID: 33444020. doi:10.1021/acs.jpcb.0c10078

Zhang, J. (2015). The hybrid idea of (energy minimization) optimization methods applied to study prion protein structures focusing on the beta2-alpha2 loop. *Biochem. Pharmacol.* 4, 2167–0501. doi:10.4172/2167-0501.1000175