

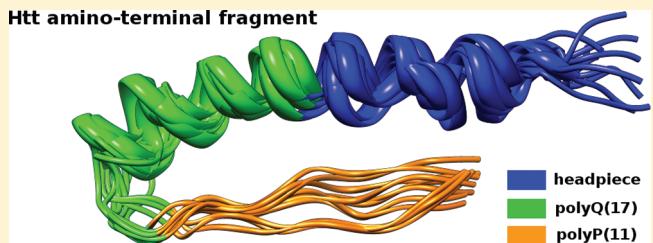
Secondary Structures of Native and Pathogenic Huntingtin N-Terminal Fragments

Maciej Długosz* and Joanna Trylska

Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Żwirki i Wigury 93, Warsaw 02-089, Poland

 Supporting Information

ABSTRACT: Huntington's disease is a neurodegenerative disorder caused by a polyglutamine (polyQ) expansion in the N-terminal fragment of the Huntingtin (Htt) protein. Structural properties of Htt N-terminal regions and the molecular mechanism leading to protein aggregation have not been fully explained yet. We performed all-atom replica exchange molecular dynamics to investigate the structures of Htt N-terminal parts with polyQ tracts of nonpathogenic and pathogenic lengths. The monomers were composed of the headpiece (17 N-terminal residues), a polyQ tract (polyQ₍₁₇₎ for native and polyQ₍₅₅₎ for pathogenic sequence), and a polyP₍₁₁₎ region, followed by 17 amino acids of mixed sequence. We found that corresponding regions in both fragments fold to similar secondary structures; the headpiece and polyQ stretch adopt mainly α -helical conformations, and polyP₍₁₁₎ forms the PP II-type helix. The native N-terminal fragment is more compact and stabilized by hydrophobic interactions between the surface of polyP₍₁₁₎ and the amphipathic helix of the headpiece. In the pathogenic fragment the headpiece is solvent exposed and does not interact with polyP₍₁₁₎. The predicted structure of the native N-terminal fragment agrees with the X-ray structure of the Htt first exon containing polyQ₍₁₇₎. The structure of the pathogenic fragment adheres to an aggregation model that is mediated by the Htt headpiece.



INTRODUCTION

Several neurodegenerative diseases, such as Huntington disease (HD), are associated with expanded CAG tracts¹ leading to production of the polyglutamine (polyQ) elongated proteins. HD is an inherited and lethal disorder caused by an extended polyQ domain in the amino-terminal fragment of the Huntingtin protein (Htt). The polyQ expansion in Htt leads to formation of intracellular aggregates. Protein aggregation is believed to play a key role in the disease. There is a connection between the aggregation process, accumulation of aggregates, and neuronal death; however, the mechanism is not clear. Also, the exact brain pathology that the disease induces is yet unknown. A correlation exists between the mutant polyQ length, severity of the disease, and the age of onset. A critical threshold of about 35–40 glutamines is required to exhibit disease symptoms² and neuronal intranuclear inclusions. In general, the age of patients who show first symptoms of HD is inversely proportional to the length of the polyQ expansion.³

The Htt protein is expressed in all mammalian cells. Its structure and exact function are unknown, but Htt plays a key role in nerve cells and is highly expressed in the brain. Htt also interacts with other proteins involved in transport, transcription, and signal processing. Even though the Htt protein is composed of over 3100 residues with a mass of about 350 kDa, expressing only the first exon of the Htt gene is enough to observe HD pathogenesis in mice and toxicity in cells.⁴ Exon I alone also produces toxicity in yeast.⁵ Moreover, the neuronal inclusions

extracted postmortem from HD patients consist mainly of N-terminal Htt fragments (of about 40 kDa) and not of the full-length Htt proteins.⁶ This fact suggests that the truncated protein itself may be toxic to cells.

Even though there is a correlation between the aggregation process and toxicity to neuronal cells, the detailed molecular mechanism of HD pathogenesis has not been clarified yet. Different hypotheses have been suggested to explain the aggregation scenario and cytotoxicity. One belief is that monomeric Htt proteins with extended polyQ tracts oligomerize and accumulate in neurons in a form of intracellular aggregates.⁷ Another suggestion states that both monomers and oligomers of polyQ-extended Htt interact with other cellular proteins, resulting in their altered function, which ultimately leads to neuronal dysfunction and cell death.⁸ Also, the above-described presence of mainly N-terminal Htt fragments in neuronal inclusions⁶ indicates a mechanism based on preferential proteolytic cleavage of mutant Htt. The Htt protein is cleaved by a cysteine protease (caspase) which becomes activated as the cell initiates programmed cell death and the cleavage rate grows with the length of the polyQ tract.⁹ A number of other CAG disease proteins require cleavage of the polyQ-containing fragment prior to aggregation.¹⁰ Therefore, the generation of truncated toxic

Received: July 6, 2011

Revised: August 30, 2011

Published: September 12, 2011

protein fragments containing the expanded polyQ tract may also be a probable mechanism of the disease onset.

On the protein level, typically, the misfolding or conformational changes in the polyQ extensions of mutant Htt proteins are believed to be central for the development of HD.^{11–15} However, models of HD pathogenesis that involve amino acids neighboring polyQ in the Htt sequence were also proposed.^{16–20} It was also postulated that nonpathogenic and pathogenic polyQ tracts have similar structural properties in solution.^{21–23} Therefore, the questions of what triggers the aggregation process and how complex is this mechanism still remain unanswered. Overall, the process of Htt aggregation is complex and multistep, passing from monomers through soluble and insoluble aggregates to fibrils and neuronal inclusions.

Much of the knowledge on polyQ-dependent aggregation has been derived from model peptides. Numerous studies on the isolated polyQ monomers and oligomers have been conducted. Still, the structural properties of monomeric polyQ are unclear and inconsistent. Experiments on single polyQ tracts of different lengths, modified for solubility, suggest their propensity for various configurations: a random coil,^{21,22,24} a collapsed structure,²⁵ an α -helix,²⁶ a β -sheet,^{13,27} or even a PP II helix.^{27–29} Also, a variety of polyQ tract structures—random coils, collapsed globules, disordered compact conformations, and extended or helical conformations—were predicted using computational modeling.^{18,30–37} From the above studies, one can conclude that the isolated polyQ tract exhibits conformational flexibility and its structure is sensitive to the local environment. It seems that, prior to aggregation, polyQ samples various conformations under different solution conditions.

Even if isolated polyQ tracts do not have an identified ordered structure, there is a generally confirmed belief that polyQ aggregates adopt structures rich in β -content^{20,38–42} with longer polyQ peptides aggregating faster.^{22,43} Peptides as short as 6–8 polyQ^{29,41} can form aggregates; this polyQ number is much lower than the threshold for the disease manifestation.²¹ Nevertheless, there are conflicting concepts about the intermediates that lead to the oligomerization of polyQ monomers. For example, in experiments of Lee et al.⁴⁴ synthetic K₂–Q₂₃–K₂ monomers were shown to initially assemble into soluble linear aggregates with no regular secondary structure and eventually maturing into insoluble ones. The soluble oligomers were believed to be precursors of the mature fibrils. Armen et al.³⁰ proposed that fibril formation occurs through an α -sheet structure as an aggregation nucleus. Molecular dynamics simulations of the aggregation of polyQ peptides between 16 and 48 residues long^{34,45} suggested that polyQ self-assemble to annular structures formed of β -sheets.

The aggregation process in the cell is much more complex than for isolated polyQ tracts and does not depend solely on the polyQ tract but also its environment. To mimic the environment surrounding the polyQ regions in full-length proteins, some authors fused other soluble proteins to polyQ. Some studies have shown that such fusing modulates the rate of aggregation and the types of intermediates formed, as well as subcellular localization.^{46–48} For example, Bulone et al.⁴⁷ studied a model system composed of polyQ₂₂ or polyQ₄₁ fused to a soluble protein glutathione S-transferase and found that this protein carrier changes the aggregation pathway of polyQ by leading to formation of soluble intermediates before the formation of insoluble large aggregates. Masino et al.⁴⁹ examined the solution structure of polyQ in a similar soluble system and found that

polyQ is highly exposed to solvent and in a random coil conformation. Ignatova and Giersch⁵⁰ fused Htt exon 1 residues to a globular protein (CRABP I) and found that expanding the polyQ region to above 35 glutamines results in structural perturbation of the CRABP I protein and enhances the aggregation of such a chimera.

Obviously, it is crucial to investigate the context of the Htt protein itself, which needs to be included if one wants to understand which processes lead to aggregation and eventually to the disease. The question arises if and how the residues neighboring the polyQ stretch and the entire Htt context modulate the aggregation (see, e.g., the review in ref 51). In the first exon of the native Htt the polyQ tract follows a 17 amino acid long N-terminal mixed flanking sequence—the Htt headpiece. From the other end, the polyQ tract neighbors with a polyP stretch followed by a mixed sequence also rich in Q and P.

The role of flanking residues in the aggregation of polyQ was confirmed by several studies on model peptides. Experiments of Thakur et al.¹⁶ used synthetic peptides containing the Htt headpiece, polyQ of various lengths, and polyP. The authors showed that the addition of the Htt headpiece to polyQ tract facilitates aggregation and the addition of the polyP tract suppresses aggregation. They confirmed that polyQ aggregation is modulated by its flanking sequences and most importantly facilitated by the N-terminal headpiece. Tam et al.¹⁷ have also shown that isolated polyQ stretches aggregate very slowly without its flanking N-terminal 17 residues. They proved the role of the Htt headpiece in aggregation by showing that the chaperonin TRiC sequesters aggregation by binding to the Htt headpiece and not to the polyQ tract itself. Rockabrand et al.⁴⁸ performed experiments on a polypeptide composed of the Htt headpiece, polyQ (of different lengths), and polyP fused to the green fluorescent protein. The authors found that the deletion of the Htt headpiece reduced aggregation, suggesting that the first 17 residues play an important role in this process. The role of the Htt headpiece in stimulating the aggregation and altering the aggregation mechanism is now widely recognized.^{16,17,19,20} Apparently, the headpiece enhances rates of aggregation^{16,17,48} by mediating the rapid formation of spherical oligomers with Htt headpieces at their core.¹⁶

The decrease of polyQ aggregation rate by attaching C-terminally the polyP tract was also shown by Bhattacharyya et al.²⁶ in experiments on synthetic K₂Q₄₀P₁₀K₂ peptides. Introduction of polyP after polyQ in synthetic peptides decreases both the rate of aggregation and the stability of amyloid-like aggregates. Sequences shorter than six prolines and placing polyP at the N-terminus do not show this suppressing effect. Darnell et al.²⁹ in experiments on peptides containing polyQ–polyP tracts also found a protective role of the polyP sequence; polyP placed C-terminally to polyQ prevented aggregation by increasing the threshold from 9 to about 15 polyQ. Duennwald et al.⁵² studied the role of polyQ flanking residues in a yeast cell model. They used various polyQ constructs fused to the green fluorescent protein and found that polyP positioned from the carboxyl terminus of polyQ reduces toxicity in yeast cells,⁵² suggesting again a protective effect of the proline-rich region. On the contrary, positioning polyP at the amino terminus of polyQ unmasks toxicity.

In summary, the above studies show that the headpiece enhances and the polyP tract slows the aggregation of polyQ. Perhaps these polyP regions are universally conserved to

prevent aggregation by stabilizing an aggregation incompetent conformation of the monomer. Also, since the disease proteins require proteolytic cleavage from the host protein prior to aggregation, the large Htt protein context may even inhibit aggregation.

Very few simulation studies accounted for the residues neighboring polyQ tracts. Williamson et al.¹⁸ carried out simulations to determine the secondary structure of monomeric polyQ tracts of different lengths in the presence of the Htt headpiece. They showed that the polyQ helical content decreases when the number of Q in the tract increases. The headpiece segment was found to suppress the intrinsic propensity of polyQ aggregation, and the association was found to be governed mainly by polyQ domains. In a recent work of Lakhani et al.⁵³ polyQ tracts of 23, 36, 40, and 47 repeat lengths were studied with full flanking regions, without polyP sequences, and in isolated polyQ peptides. The authors found a positive correlation between the length of the polyQ expansion and its probability to form a β -rich misfolded state. On the basis of the simulations, they also concluded that the flanking sequences affect the formation of β -sheet structures in the polyQ region.

The structure of the Htt protein has not been resolved. However, the structure of the Htt first exon containing a repeat of 17 glutamines (Htt17Q-EX1) was determined using X-ray crystallography.⁵⁴ The structure of Htt17Q-EX1 consists of an amino-terminal α -helix and polyP helix II formed in the polyP₍₁₁₎ region. An α -helix (of maximum 15 glutamines depending on the crystal), extended loop, and random coil conformations were evidenced in the polyQ₍₁₇₎ tract. The experimental X-ray structure of any pathogenic Htt fragment is still unknown.

Due to contradictory hypotheses of HD pathogenesis and a lack of 3D structure of the Htt protein, it seems crucial to determine the structural features of polyQ tracts in solution and in the environment including the polyQ neighboring residues. In the present work we investigated the secondary and tertiary structures of the native and pathogenic variant of the Htt N-terminal region (i.e., including the polyQ flanking residues). We studied the structural properties of the following isolated Htt amino-terminal fragments, with the lengths of polyQ repeats below (17) and above (55) the threshold needed for HD manifestation: MATLEKLMKAFESLKSF-polyQ₍₁₇₎-polyP₍₁₁₎-LPQPPQAAQPLLPQPQ and MATLEKLMKAFESLKSF-polyQ₍₅₅₎-polyP₍₁₁₎-LPQPPQAAQPLLPQPQ corresponding, respectively, to 62 and 101 N-terminal amino acids of the native and mutated Htt. The goal was also to determine the interactions between polyQ and its neighboring regions. To determine the accessible conformations of native and pathogenic Htt fragments, we performed all-atom molecular dynamics (MD) using the canonical replica exchange molecular dynamics (REMD) technique.^{55–60}

We compare the calculated structural properties of pathogenic and nonpathogenic N-terminal fragments of Htt with experiments and discuss them in the context of proposed HD pathogenesis mechanisms. Our studies discuss the differences in the interactions of the headpiece and polyP stretch with the polyQ tract and with each other in the native and mutant Htt fragments. We discuss how the flanking sequences may modulate the aggregation mechanism and show the dominant effect of hydrophobic interactions in this modulation. The predicted structures of both variants corroborate with an aggregation mechanism that is initiated and mediated by the Htt headpieces as proposed in ref 16.

METHODS

Molecular Dynamics Simulations. The studied sequences are labeled as follows: N17Q^{Htt} (nonpathogenic fragment, MATLEKLMKhyphen-qj;AFESLKSF-polyQ₍₁₇₎-polyP₍₁₁₎-LPQ-PPPQAQPLLPQPQ) and N55Q^{Htt} (pathogenic fragment, MATLEKLMKAFESLKSF-polyQ₍₅₅₎-polyP₍₁₁₎-LPQPPP-QAQPLLPQPQ). The C-terminal carboxylate groups were neutralized by amidation.

REMD simulations were conducted using the AMBER 9 package⁶¹ with the FF03 force field.⁶² The FF03 force field was previously used with both explicit and implicit description of solvent to study folding of small peptides and proteins^{59,60,63} and in the simulations of β -sheet aggregation.⁶⁴ Solvation effects were modeled implicitly with the generalized Born model (GBn) with nonpolar contribution to the solvation free energy of Mongan et al.⁶⁵ Dielectric constants of 1 and 78.54 were assigned to solutes and solvent, respectively. The surface tension coefficient was set to 0.005 (kcal/mol) Å⁻² and the salt concentration to 0.15 M. A cutoff of 18 Å was applied to evaluate the nonbonded interactions. Temperature was maintained with Langevin dynamics protocol with a collision frequency of 1 ps⁻¹. The time step was 2 fs, and all bonds involving hydrogens were constrained with the SHAKE algorithm with a tolerance of 10⁻⁴ Å. For the sake of comparison we also performed standard, constant temperature MD (CTMD) simulations at 300 K, using the same setup.

Initial configurations of N17Q^{Htt} and N55Q^{Htt} were built as fully extended peptide chains and were subjected to gradual heating to 500 K during 200 ps. Such uncorrelated structures were then energy minimized and used to initiate the CTMD and REMD runs. Snapshots from CTMD trajectories were collected in 1 ps intervals. During REMD, eight replicas were simulated over a range of roughly exponentially spaced temperatures from 290 to 377.1 K: 290.0, 301.1, 312.6, 324.6, 337.0, 349.8, 363.2, and 377.1 K. Exchange attempts were made every 2 ps. The acceptance exchange ratios between neighbor temperatures ranged between ~20 and ~30% (Figure S1 shows the energy overlap between different replicas). Initial attempts to use a wider temperature range showed a considerable drop in the replica–swap acceptance ratio indicating that for the studied systems the heat capacity is temperature-dependent (see Figure S2 of the Supporting Information). Each REMD trajectory length was 125 ns summing to a total of 1 μ s simulation time. Snapshots from REMD trajectories were collected every 1 ps, and the first 25 ns was discarded for analysis. We validated the sampling efficiency and convergence of REMD and checked whether individual copies of the simulated systems traversed the whole studied temperature range (see Figure S3, Supporting Information). Additionally, we split the temperature–space trajectory of each simulated copy into two parts and for each part computed the copy's residence times at different temperatures. The obtained distributions were largely converged (see Figure S3).

Additionally, we performed MD simulations in explicit solvent with NAMD⁶⁶ and the Amber FF03 force field⁶² in an effort to judge the reliability of the described above implicit solvent simulations. Uncorrelated (without any secondary structure elements) conformation of the N17Q^{Htt} was immersed in a truncated octahedron box of TIP3P waters (the thickness of the resulting water layer around the protein was at least 15 Å in each direction). Explicit ions of the NaCl salt were placed inside the box to both neutralize the system and create physiological conditions of 0.15 M ionic strength. Periodic boundary conditions

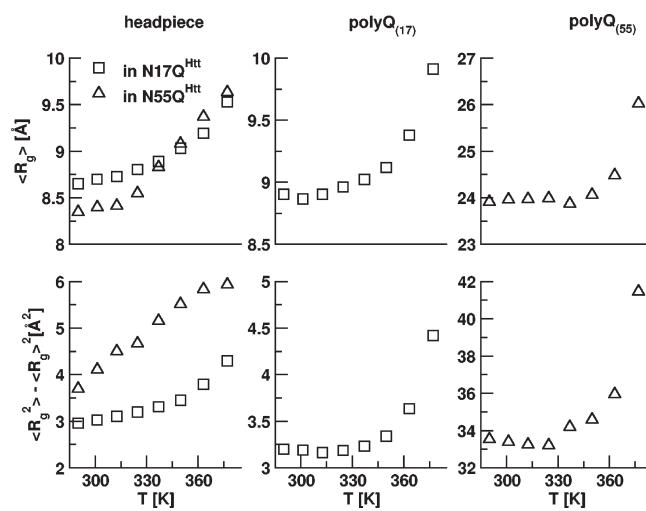


Figure 1. Average gyration radii ($\langle R_g \rangle$, top) and their variance (bottom) computed for the headpiece and polyQ tracts of $N17Q^{Htt}$ and $N55Q^{Htt}$ for configurations sampled by REMD at different temperatures.

were applied. A simulation production phase was performed in the *NPT* ensemble with a constant pressure of 1 bar regulated by the Langevin piston method⁶⁷ and a constant temperature of 300 K ensured via the Langevin thermostat. To allow for a longer, 2 fs, integration time step, the SHAKE⁶⁸ algorithm was used. For long-range interactions, the particle mesh Ewald method⁶⁹ was applied with a grid spacing of about 1.0 Å; a 10 Å short-range cutoff was used for nonbonded interactions.

Data Analysis. The trajectories were analyzed using the *ptra* module of AMBER⁶¹ and various GROMACS⁷⁰ tools. Molecular graphics were done with the UCSF Chimera software.⁷¹

To generate reduced and representative for a given temperature structural ensembles, we performed root-mean-square deviation (rmsd) conformational clustering, based on the algorithm described previously by Daura et al.⁷² The GROMACS⁷⁰ *g_cluster* utility was employed. The structures from REMD trajectories were extracted every 10 ps, resulting in 10 000 configurations for each temperature. The rmsd clustering was performed on a subset of 44 N-terminal residues of $N17Q^{Htt}$ (ATLEKLMKA-FESLKSF—polyQ₍₁₇₎—polyP₍₁₁₎) and 82 N-terminal residues of $N55Q^{Htt}$ (ATLEKLMKA-FESLKSF—polyQ₍₅₅₎—polyP₍₁₁₎), excluding the Met 1 residue. The selected residues were clustered using as a similarity criterion the rmsd of all main-chain heavy atoms. Cutoffs of 2 and 2.5 Å were chosen for $N17Q^{Htt}$ and $N55Q^{Htt}$, respectively. We describe in detail the structural properties of the Htt headpiece, polyQ_(17/55), and polyP₍₁₁₎ tracts because we do not observe well-defined secondary structures in the carboxy-terminal regions of the studied Htt fragments.

RESULTS AND DISCUSSION

Thermal Stability. Gyration Radii. Figure 1 shows the average gyration radii ($\langle R_g \rangle$) of the Htt headpiece and polyQ_(17/55) tracts as a function of temperature, calculated for their conformations sampled in REMD.

In all cases, $\langle R_g \rangle$ remains leveled in the temperature range of 290–330 K and eventually increases with temperature because at high temperatures the structures expand. The R_g variance also increases with temperature. Larger R_g fluctuations are observed for the pathogenic variant $N55Q^{Htt}$ for both the headpiece and

polyQ tract. The R_g of the headpiece in the native fragment undergoes smaller fluctuations than in the pathogenic one, suggesting that the headpiece is more stable in the $N17Q^{Htt}$ variant. This result agrees with the previously reported data showing the dependence of the headpiece $\langle R_g \rangle$ on the polyQ length.¹⁸ Overall, we can conclude that the Htt fragment with a longer polyQ tract, $N55Q^{Htt}$ is less thermally stable.

The time course of R_g for both Htt fragments from CTMD is shown in Figure S4 in the Supporting Information. In general, for both Htt variants, the R_g calculated at low temperatures (290 and 301.1 K) in REMD are in accord with R_g computed from CTMD at 300 K. Also, larger R_g fluctuations are observed for the pathogenic monomer in CTMD.

Secondary Structure Content. The results of REMD suggest that the corresponding sequences in $N17Q^{Htt}$ and $N55Q^{Htt}$ adopt similar secondary structures. Secondary structure contents for the Htt headpiece, polyQ_(17/55), and polyP₍₁₁₎ tracts are presented in Figure 2 and discussed in the following subsections.

Htt Headpiece. In both $N17Q^{Htt}$ and $N55Q^{Htt}$ the Htt headpiece (residues 1–17) tends to form an α -helix with a large α -helical content characteristic especially for residues 4–17. The secondary structure content also suggests a possible turn for residues Met 8, Lys 9, and Ala 10. We find that Htt headpieces either form a straight, continuous α -helix or a two-helix bundle. While the propensity for forming an α -helix decreases upon increasing temperature, the turn content for residues 8–10 remains almost constant up to 377.1 K. The turn content is larger for $N55Q^{Htt}$ (~30%) than for $N17Q^{Htt}$ (~20%), signifying that the headpiece is more prone to form a two-helix bundle in the pathogenic variant. In $N17Q^{Htt}$ and $N55Q^{Htt}$ the 3–10 helix and turn are observed in the boundary regions of the Htt headpiece, suggesting its conformational flexibility. Indeed, the average number of helical residues in the Met 1–Phe 17 region changes upon raising the temperature, from 8 to 6.3 in the case of $N17Q^{Htt}$ and from 6.8 to 5.6 in case of $N55Q^{Htt}$. This decrease in the helical content of the headpiece upon raising the temperature is also shown in Figure 3. Overall, the helical content observed in the headpiece is larger for the native fragment.

According to experimental studies, the Htt headpiece either is a random coil^{16,18} or has some α -helical content.^{16–18,54,73} In a recent study of Sivanandam et al.²⁰ the magic angle spinning solid-state NMR technique was used to investigate the properties of the Htt N-terminal fragment aggregates. The authors described that the headpiece in mature aggregates is composed in part of a well-defined helix. Additionally, on the basis of Fourier transform infrared spectroscopy (FTIR) measurements, Sivanandam et al.²⁰ suggested that the helical conformation is also the dominant feature of the headpiece at the initial stage of oligomerization of the Htt N-terminal fragments.

The computational studies also predict different structures of the headpiece depending on the neighboring context.^{18,53,63} Kelley et al.⁶³ observed in explicit solvent REMD a dominant α -helical conformation of the isolated Htt headpiece, as well as its ability to form a two-helix bundle. On the other hand, Williamson et al.¹⁸ who investigated a series of constructs containing the Htt headpiece and polyQ of various lengths (but without the polyP₍₁₁₎ tracts) did not observe a substantial helical content in the headpiece. The authors applied Monte Carlo simulations with the ABSINTH implicit solvation model⁷⁴ and found that a fractional α -helical content in the Htt headpiece for constructs with polyQ tracts longer than 15 residues was below 15% and almost vanished at 385 K. According to the results of Lakhani et al.⁵³

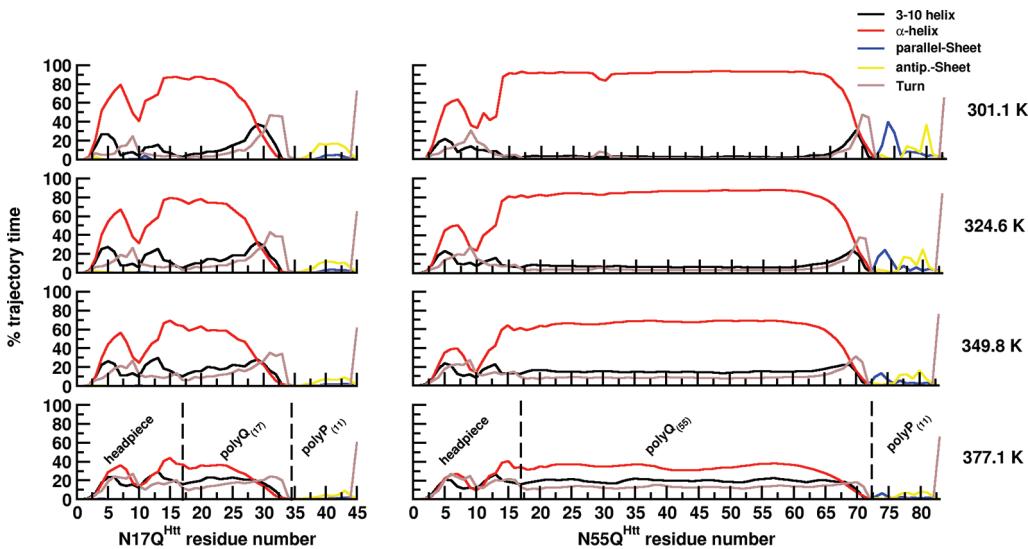


Figure 2. Secondary structure content by residue for $\text{N}17\text{Q}^{\text{Htt}}$ and $\text{N}55\text{Q}^{\text{Htt}}$ derived from REMD at different temperatures.

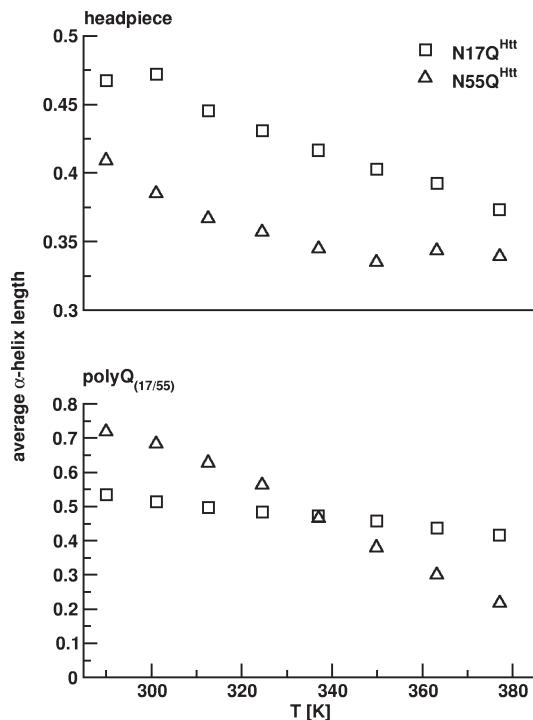


Figure 3. Average α -helical contents in the headpiece (top) and $\text{polyQ}_{(17/55)}$ (bottom) in the $\text{N}17\text{Q}^{\text{Htt}}$ and $\text{N}55\text{Q}^{\text{Htt}}$ fragments derived from REMD at different temperatures. The Y axes are normalized to 1, which corresponds to a helix composed of all residues in the shown fragment.

who applied discrete molecular dynamics simulations⁷⁵ employing the EEF1 implicit solvent model⁷⁶ (for various constructs including the headpiece, polyQ, and polyP tracts), the Htt headpiece most likely adopts a random coil (random coil conformations were observed for over 50% of simulation time regardless of the polyQ length). However, the authors also found that in the headpiece an α -helix or a β -strand content are possible. In this study the probability of forming a particular secondary structure of the headpiece was correlated with the

length of the neighboring polyQ tract. The likelihood of forming a β -sheet grew with the polyQ length. Headpiece residues sampled β -sheet dihedral angles more frequently than α -helical dihedral angles when the length of the polyQ tract was above 23.⁵³ The tendency to adopt a β -sheet conformation was even more pronounced when polyP regions were removed from the studied constructs.⁵³

On the basis of the above literature, it is clear that the computational secondary structure predictions of the Htt headpiece depend on the applied simulation methodology. Both in Kelley et al.⁶³ and our work, the all-atom AMBER FF03 force field was used which may slightly overstabilize α -helices⁷⁷ when used with explicit solvent. However, the FF03 force field used together with the GBn⁶⁵ implicit solvation model in REMD, on the contrary, did not stabilize helical structures.⁷⁸ Moreover, the authors claimed that the FF03 parameters for dihedral torsions are rather biased toward the β -structure when used with implicit solvent.⁷⁸ The balance in the backbone conformational preferences of a particular force field and the solvation model may certainly be the source of differences. However, we note that the crystal structure of the fragment of the first Htt17Q exon⁵⁴ displays in the Htt headpiece region a straight, continuous 17 amino acid long α -helix, which is in accord with our results and those of Kelley et al.,⁶³ as well as with some experiments.^{16–18,73}

Thakur et al.¹⁶ performed fluorescence resonance energy-transfer (FRET) experiments to determine the end-to-end distance of the Htt headpiece dependent on the polyQ length. These FRET distances were about 24, 25.5, 31.5, and 34 Å for the headpiece-Q₃, headpiece-Q₂₀P₁₀, headpiece-Q₃₇P₁₀, and the extended form of the headpiece-Q₃, respectively. The authors concluded that the headpiece is compact and acquires a collapsed structure. However, these FRET distances do not rule out a folded helical state of the headpiece. Assuming that each residue in an α -helix contributes about 1.5 Å to the helix longitudinal length, the 17-residue fragment would have an end-to-end distance of about 25.5 Å. This agrees with the value obtained for the headpiece-Q₂₀P₁₀ fragment. A similar in content native fragment studied by us contains a helix in the headpiece which is additionally stabilized by the PPII helix of the following polyP strand. FRET also shows that the headpiece end-to-end distance

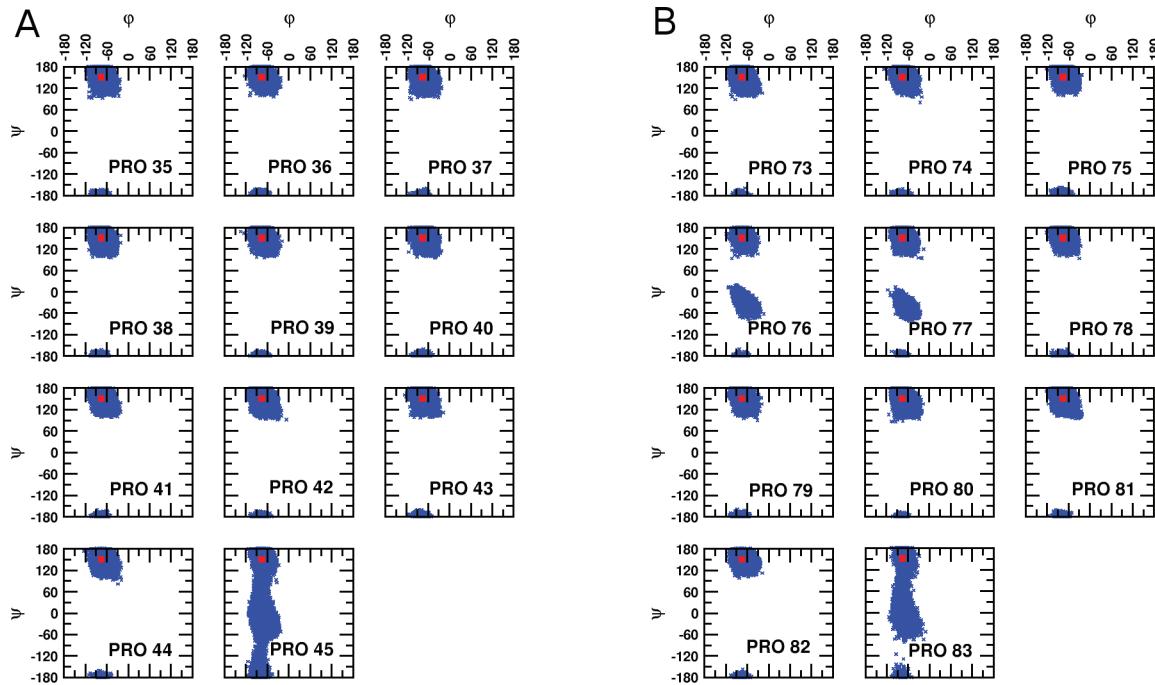


Figure 4. Ramachandran plots for the polyP₍₁₁₎ segment of N17Q^{Htt} (A) and N55Q^{Htt} (B) for structures sampled in REMD. A red dot denotes an ideal conformation of a PP II helix.

increases by about 8–9 Å for the length of the polyQ tract of 37. This suggests that the headpiece is less structured and partially unfolded for the longer polyQ. This fact is also in accord with our data that show that the headpiece is more structurally variable in the pathogenic fragment and contains much less helical content. Also, the expanded polyQ domain was found to destabilize a folded protein domain,⁵⁰ which agrees with the fact that both the headpiece and polyP in the pathogenic fragment are more structurally variable than in the native fragment.

PolyP₍₁₁₎ Flanking Region. The polyP₍₁₁₎ region acquires an extended conformation both in N17Q^{Htt} (residues 35–45) and in N55Q^{Htt} (residues 73–83). According to Ramachandran plots presented in Figure 4, the dominant conformation of the polyP₍₁₁₎ fragment is the PP II helix (with $\phi = -75^\circ$ and $\psi = 150^\circ$). In N17Q^{Htt} the PP II helix is predominantly straight and stabilized by hydrophobic interactions with the surface of the α -helix formed in the Htt headpiece. In N55Q^{Htt} this PP II helix adopts a distorted, kinked configuration evidenced by additional allowed regions of Pro 76 and Pro 77 in the Ramachandran map. The polyP₍₁₁₎ fragment was also resolved as a PP II helix by crystallography,⁵⁴ adopting either a kinked or straight configuration. Indeed, the proline stretches are known to favor a PP II helix.^{26,27,29}

PolyQ₍₁₇₎ and PolyQ₍₅₅₎ Tracts. In our REMD simulations the polyQ tracts, in both nonpathogenic and pathogenic variants, contain α -helical structures (Figure 2 and Figure 3). However, the α -helical content decreases upon raising the temperature from 290 to 377 K; in N17Q^{Htt} the average number of helical Q residues changes from 9 (53% of all glutamines in the polyQ tract) to 7 (41%) and in N55Q^{Htt} from 39 (71%) to 11.5 (20%). The longer α -helix adopted by the polyQ₍₅₅₎ stretch is thus more conformationally variable and most of it diminishes at higher temperatures. In both monomers, the α -helical content decreases toward the C-end of the polyQ stretches. Both in

polyQ₍₁₇₎ and polyQ₍₅₅₎ at low temperatures the C-terminal glutamines adopt a 3–10 type helix. In polyQ₍₁₇₎, upon raising the temperature, the 3–10 helix propensity increases for glutamines no. 18–27 and decreases for 28–34. A similar decrease of the 3–10 helix propensity is observed for C-terminal glutamines in polyQ₍₅₅₎.

At low temperatures there is also a significant probability of a turn involving the last seven to nine glutamines in both polyQ₍₁₇₎ (residues 25–34) and polyQ₍₅₅₎ (residues 63–72). Such extended conformation is at least partially imposed by the PP II helix that forms in the polyP₍₁₁₎ region. The proline stretches apart from favoring a PP II helix themselves are known to influence the conformation of the sequences preceding the polyP stretch.^{79,80} The fact that a polyP region can stabilize an extended conformation of a polyQ stretch was previously described in circular dichroism and NMR studies of synthetic polyQ–polyP peptides.^{27,29} It was also shown that adding a C-terminal polyP fragment to a polyQ stretch prevents the formation of an α -helix in the polyQ region.²⁶ In the crystal structure of the first Htt17Q exon, the last four glutamines of the polyQ stretch also acquire an extended configuration.⁵⁴ According to our simulations, introducing the polyP sequences following the polyQ stretches affects the configuration of a similar number of glutamines in N17Q^{Htt} and N55Q^{Htt}. In both cases the turn content diminishes upon raising the temperature (Figure 2).

Conformational flexibility observed here for the C-terminally located glutamines of the polyQ tracts corroborates with the crystal structures of the Htt exon 1⁵⁴ because the terminal fragment of the polyQ stretch was not completely resolved in this region which suggests its significant conformational freedom. The secondary structures in the polyQ₍₁₇₎ region observed in the Htt17Q-EX1 crystals include an α -helix (up to 15 residues long), a random coil, and an extended loop⁵⁴ (see Figure S5, Supporting Information). While no evidence of a β -structure in the polyQ

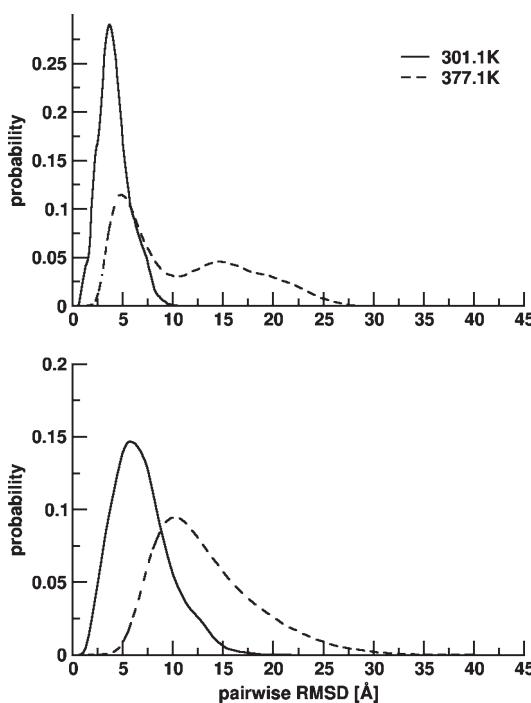


Figure 5. Histograms of the pairwise rmsd computed for configurations of $\text{N}17\text{Q}^{\text{Htt}}$ (top) and $\text{N}55\text{Q}^{\text{Htt}}$ (bottom) derived from REMD.

tract exists in the crystals of the first Htt17Q exon, according to the work of Lakhani et al.⁵³ even in the constructs with polyQ tracts well below the pathogenic threshold the probability of β -sheet dihedral angles in polyQ residues was quite substantial ($\sim 30\%$). Interestingly, this probability was almost constant for varying number of Q repeats. Recently, Singer et al.²³ synthesized the Htt exon 1 polypeptides of native and pathogenic lengths (90 amino acids with polyQ₍₂₃₎ and 109 amino acids with polyQ₍₄₂₎). In circular dichroism spectroscopy experiments, in both fragments, they saw weak helical content (in the phosphate buffer with the presence of 50% TFE to stabilize secondary structure) and excluded any β -sheets. The helical content was higher for the longer fragment; however, the spectra suggest that both α -helices were not fully developed. In the phosphate buffer, the spectra for both fragments were similar and indicated a stabilized turn. The data of Singer et al.²³ are in agreement with the crystallographic ones⁵⁴ and our REMD simulations. Again, without making an attempt to judge the performance of different approaches, we note that in the case of MD simulations the details of a particular force field and the solvation model might be an issue when one tries to predict structural properties of proteins. Therefore, a careful comparison of methods and experimental data is needed.

Structural Clusters. To obtain the most representative REMD configurations of $\text{N}17\text{Q}^{\text{Htt}}$ and $\text{N}55\text{Q}^{\text{Htt}}$, we performed the clustering analysis as described in Methods. Figure 5 presents the distributions of the pairwise rmsd constructed for trajectory conformations that were selected for clustering.

These distributions show that at 301.1 K the most common rmsd values are $\sim 3.5 \text{ \AA}$ for $\text{N}17\text{Q}^{\text{Htt}}$ and $\sim 5 \text{ \AA}$ for $\text{N}55\text{Q}^{\text{Htt}}$. Since meaningful clusters should be confined in small configurational volumes, we clustered configurations with the rmsd criteria of 2 \AA for $\text{N}17\text{Q}^{\text{Htt}}$ and 2.5 \AA for $\text{N}55\text{Q}^{\text{Htt}}$. Additionally, we assumed that clusters representative for certain energetic minima should be significantly populated, so we describe only the clusters

Table 1. Populations ($P_{\text{N}17\text{Q}^{\text{Htt}}}$, $P_{\text{N}55\text{Q}^{\text{Htt}}}$) of the Most Occupied Clusters of $\text{N}17\text{Q}^{\text{Htt}}$ and $\text{N}55\text{Q}^{\text{Htt}}$ at Different Temperatures

T (K)	$P_{\text{N}17\text{Q}^{\text{Htt}}}$ (%)	$P_{\text{N}55\text{Q}^{\text{Htt}}}$ (%)	T (K)	$P_{\text{N}17\text{Q}^{\text{Htt}}}$ (%)	$P_{\text{N}55\text{Q}^{\text{Htt}}}$ (%)
290.0	23.35	28.50	349.8	4.81	0.65
301.1	25.10	14.70	363.2	2.58	0.24
312.6	19.00	6.40	377.1	1.10	0.10
324.6	13.35	3.85			
337.0	9.05	1.84			

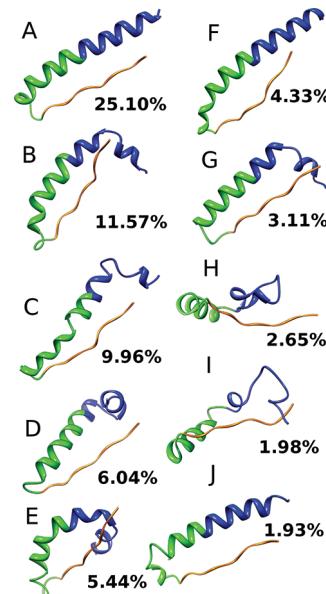


Figure 6. Middle structures of the 10 most populated clusters observed at 301.1 K for $\text{N}17\text{Q}^{\text{Htt}}$. The ordering (A–J) reflects the decreasing number of cluster members with populations denoted in percent. Colors: blue, the Htt headpiece (17 N-terminal amino acids); green, polyQ₍₁₇₎; orange, polyP₍₁₁₎.

with relative populations above 2%. Figure 5 also shows that at 301.1 K the pairwise rmsd distribution obtained for $\text{N}55\text{Q}^{\text{Htt}}$ is substantially wider than for $\text{N}17\text{Q}^{\text{Htt}}$ because the longer fragment has more conformational freedom than the shorter one. At 377.1 K the distribution for $\text{N}17\text{Q}^{\text{Htt}}$ is clearly bimodal (with a narrow peak that overlaps with the distribution obtained at 301.1 K) and for $\text{N}55\text{Q}^{\text{Htt}}$ the distribution is single-peaked and wider with a maximum around 11 \AA . This suggests that for the native Htt fragment at high temperatures the low-temperature conformations still exist in the ensemble of allowed conformations.

Table 1 presents the temperature-dependent occupancy of the most populated cluster for the nonpathogenic and pathogenic monomer forms. The number of members in the most populated cluster decreases while raising the temperature, and this decrease is larger for $\text{N}55\text{Q}^{\text{Htt}}$. Along with the data in Figure 5 these observations confirm that the structure of the pathogenic Htt amino-terminal fragment is less conformationally stable than its nonpathogenic form.

Figure 6 and Figure 7 show middle structures of the 10 largest clusters derived from the $\text{N}17\text{Q}^{\text{Htt}}$ and $\text{N}55\text{Q}^{\text{Htt}}$ REMD ensembles.

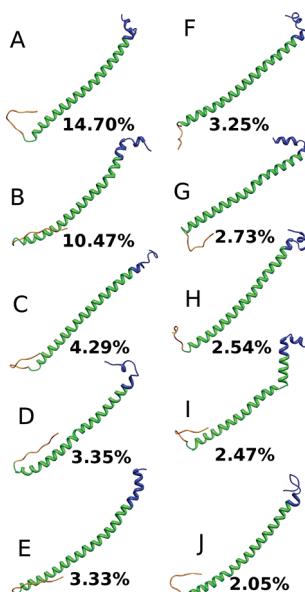


Figure 7. Middle structures of the 10 most populated clusters observed at 301.1 K for $\text{N}55\text{Q}^{\text{Htt}}$. The ordering (A–J) reflects the decreasing number of cluster members with cluster occupancy given in percent. Colors: blue, the Htt headpiece (17 N-terminal amino acids); green, poly $\text{Q}_{(55)}$; orange, poly $\text{P}_{(11)}$.

At 301.1 K these 10 clusters contain about 72% of all sampled configurations of $\text{N}17\text{Q}^{\text{Htt}}$ and about 50% of $\text{N}55\text{Q}^{\text{Htt}}$ (even though a larger clustering cutoff was used for the latter). In the most occupied cluster of $\text{N}17\text{Q}^{\text{Htt}}$, shown in Figure 6A, the Htt headpiece and poly $\text{Q}_{(17)}$ (up to Gln 29) form a straight α -helix. Only the C-end of the poly $\text{Q}_{(17)}$ sequence is distorted and forms an extended turn. A straight α -helix is also characteristic for less populated clusters F and J (Figure 6). In the remaining clusters, the Htt headpiece acquires either a two-helix bundle or a distorted, partially extended conformation. In all clusters, the poly $\text{Q}_{(17)}$ region forms primarily an α -helix of different lengths that starts at its amino end. However, in some clusters we also observe a more tight 3–10 helix; see, for example, Figure 6C. In general, the helical fragments of the poly $\text{Q}_{(17)}$ sequence are followed by turns. The poly $\text{P}_{(11)}$ region forms a straight PP II helix in all the clusters.

The conformations of the most populated cluster in $\text{N}17\text{Q}^{\text{Htt}}$ closely resemble those seen in the Htt17Q-EX1 X-ray structure. In all seven crystals used to determine the Htt17Q-EX1 structure (PDB IDs:⁵⁴ 3IOR, 3IO6, 3IOT, 3IOU, 3IOR, 3IOV, and 3IOW; see Figure S5) there is a straight α -helix in the Htt headpiece and poly $\text{Q}_{(17)}$ regions. The last resolved residue of the poly $\text{Q}_{(17)}$ α -helix ranges in different crystals from Gln 18 to Gln 32. The PP II helices of poly $\text{P}_{(11)}$ are straight in three crystals and kinked in the other three (in one of the crystals the poly $\text{P}_{(11)}$ region was not resolved). We note, however, that the two-helix bundle configuration of the Htt headpiece observed in some of the REMD clusters was not seen in the crystal structures. Also, the native first exon of Htt with 17 glutamines was crystallized with a maltose binding fusion protein⁵⁴ (MBP). The N-terminal MBP was connected with the Htt sequence using a linker of three alanines. Moreover, the MBP-Htt17Q-EX1 formed a trimer in all crystals (which is an artifact of crystallization⁵⁴), with N-terminal helices of Htt17Q-EX1 closely packed, even though in solution MBP-Htt17Q-EX1 is a monomer.⁵⁴ Therefore, it is possible that

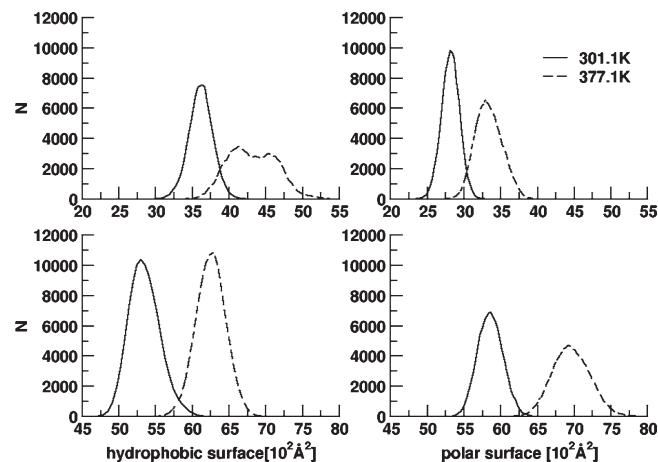


Figure 8. Distributions of hydrophobic and polar solvent accessible surface areas of $\text{N}17\text{Q}^{\text{Htt}}$ (top) and $\text{N}55\text{Q}^{\text{Htt}}$ (bottom) constructed for configurations derived from REMD at 301.1 and 377.1 K.

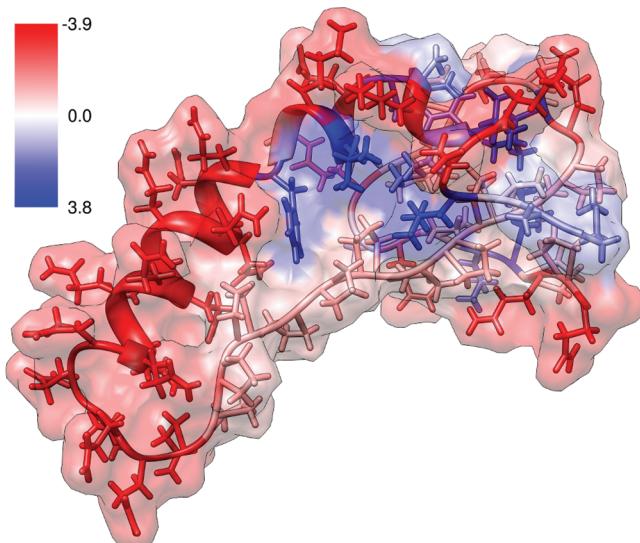


Figure 9. Exemplary low-temperature structure of $\text{N}17\text{Q}^{\text{Htt}}$. Amino acids are represented as the cartoon and licorice and surrounded by their solvent accessible surface. Coloring is based on the Kyte–Doolittle hydrophobicity scale⁸¹ with hydrophobic residues in blue and polar in red.

the exclusively straight configuration of the Htt headpiece α -helix is a consequence of the crystal packing.

In the most populated cluster of the pathogenic $\text{N}55\text{Q}^{\text{Htt}}$ (Figure 7A), the Htt headpiece forms a two-helix bundle, followed by an α -helix in the poly $\text{Q}_{(55)}$ tract, and a kinked PP II helix in poly $\text{P}_{(11)}$. Overall, the poly $\text{Q}_{(55)}$ tract helix is followed by a turn in the bordering poly Q /poly P region. The kink in the poly $\text{Q}_{(55)}$ α -helix (Figure 7I) is consistent with a small spike in the turn propensity observed at low temperatures for Gln 29 in Figure 2. The two-helix bundle observed in the headpiece is characteristic of the $\text{N}55\text{Q}^{\text{Htt}}$ fragment (Figure 7A,B,D,F,G–I). The poly $\text{P}_{(11)}$ sequence forms a PP II helix which is either straight (Figure 7D–F) or kinked (Figure 7A–C,G–J).

Hydrophobic Surface. Since hydrophobicity is a crucial property determining molecular stability and aggregation properties, we analyzed the contributions of hydrophobic interactions

Table 2. Average Solvent Accessible Surfaces and Their Standard Deviations for Residues of the Htt Headpiece in N17Q^{Htt} and N55Q^{Htt} Derived from the 301.1 K Replica

residue	SASA _{N17Q^{Htt}} (Å ²)	SASA _{N55Q^{Htt}} (Å ²)	residue	SASA _{N17Q^{Htt}} (Å ²)	SASA _{N55Q^{Htt}} (Å ²)
Met 1	195 ± 24	203 ± 21	Ala 10	43 ± 23	69 ± 17
Ala 2	71 ± 15	78 ± 14	Phe 11	118 ± 46	135 ± 30
Thr 3	89 ± 28	102 ± 22	Glu 12	117 ± 21	122 ± 17
Leu 4	94 ± 34	122 ± 20	Ser 13	29 ± 20	53 ± 20
Glu 5	118 ± 23	123 ± 32	Leu 14	83 ± 29	115 ± 20
Lys 6	146 ± 24	163 ± 19	Lys 15	143 ± 21	150 ± 19
Leu 7	79 ± 30	106 ± 27	Ser 16	47 ± 19	55 ± 18
Met 8	90 ± 37	122 ± 30	Phe 17	75 ± 30	120 ± 29
Lys 9	149 ± 24	165 ± 18			

to the stability of the obtained secondary and tertiary structures. Figure 8 shows the distributions of polar and hydrophobic solvent accessible surface areas (SASA) for the short N17Q^{Htt} and expanded N55Q^{Htt} fragments derived from the REMD trajectories.

Overall, for the N55Q^{Htt} variant both hydrophobic and polar SASA undergo larger changes upon raising the temperature, which is in accord with the postulated-above lower thermal stability of the pathogenic fragment. Moreover, we observe a bimodal character of the hydrophobic SASA distribution for N17Q^{Htt} at 377.1 K which confirms the presence of a hydrophobic core that remains at least partially stable at high temperatures. The formation of such a hydrophobic core is presented in Figure 9. This core is also evidenced by clustering analysis that shows the interactions between the headpiece and polyP helix.

The distribution and positions of hydrophobic residues in the headpiece are important. It was shown that the helix formed by the Htt headpiece is amphipathic.⁶³ Also, mutations of the headpiece residues which reduce its hydrophobic character, but do not change the charged residues, abrogate the rapid aggregation of the Htt exon 1 peptides.¹⁶ The Htt headpiece hydrophobic residues tend to minimize their solvent-exposed surface; their interactions with the polyP₍₁₁₎ tract, as well as with the mixed sequence following polyP₍₁₁₎, lead to the formation of a hydrophobic core evidenced by Figure 8.

The formation of the hydrophobic core was also observed in the explicit solvent MD simulation (for details see Figure S6 in the Supporting Information). Even though the secondary structures are not fully developed in the explicit-solvent simulations due to their short time scale, some helical content is visible in the headpiece and polyQ parts and the polyP region shows the PP II conformation. Moreover, one may already notice that the headpiece minimizes its SASA by interacting with the polyP helix.

In the pathogenic N55Q^{Htt} fragment, we do not observe such a hydrophobic core since the distribution of the hydrophobic SASA at 377.1 K is uniform (Figure 8). Additionally, the average SASA of the headpiece residues are smaller in the N17Q^{Htt} fragment than in N55Q^{Htt} (see Table 2). Differences are especially pronounced for the following hydrophobic residues: Leu 4, Leu 7, Ala 10, Phe 11, Leu 14, and Phe 17.

Headpiece Dimerization. In the experiments performed on synthetic peptide models of Htt exon 1, Thakur et al.¹⁶ proposed the aggregation mechanism that is initiated by self-aggregation of the Htt headpieces. According to this mechanism, investigated further by Williamson et al.,¹⁸ metastable aggregates composed of cores of loosely packed Htt headpieces are first formed. The

polyQ tracts are excluded from these cores. Next, amyloid nuclei consisting both of headpieces and polyQ tracts develop from these initial aggregates. A similar in nature mechanism proposing a role of the 17-residue flanking headpiece was also recently described by Liebman et al.¹⁹

To check the possibility of oligomerization of the headpieces of the exon 1 fragment as proposed by Thakur et al.¹⁶ and Liebman and Meredith,¹⁹ we performed MD simulations of the dimer consisting of two Htt headpieces. A CTMD simulation at 300 K was initiated using extended starting structures of the Htt headpieces in a parallel orientation. During the simulation both headpieces adopted helical conformations and formed a stable dimer. Moreover, as shown in Figure 10 there is a change in hydrophobic surface area upon formation of the dimer. The distribution of the hydrophobic surface is shifted toward smaller surface values. No such effect is seen for the polar surface showing that the headpieces interact through the hydrophobic part of the helix. The interactions of the headpieces through the hydrophobic surface are in accord with a mutational study performed by Thakur et al.¹⁶ and described above. This suggests that the headpiece provides some sequence-specific hydrophobic contribution to aggregation.

The structural studies performed by us corroborate with the mechanism mediated by the Htt headpieces because in the pathogenic fragment the headpieces are exposed to solvent and may initiate self-aggregation. In the native fragment the headpiece interacts with the polyP region and is not available for intermolecular interactions with another headpiece, which is also in accord with the observation that polyP positioned C-terminally to polyQ suppresses aggregation.^{16,26} When the polyP tract is positioned N-terminally to polyQ, such a pattern does not protect from aggregation.²⁶ Also, even though single headpieces and these connected to short polyQ tracts do aggregate, this process is slow.¹⁶ Aggregation is much faster when long polyQ₃₅ is connected either C- or N-terminally to the Htt headpiece. Also, aggregation is faster than for single polyQ₃₅ tracts, suggesting that headpieces enhance aggregation even more for the longer polyQ tracts. It seems that only long polyQ tracts expose the headpieces to the solvent and additionally may contribute to stabilizing its secondary structure. The headpiece–polyQ strands with polyQ tracts up to 15 glutamines may form a two-helix bundle or another structure where polyQ interacts with the headpiece and slows the overall aggregation.

In summary, the headpieces interacting with the PP II helix in the native fragment are not exposed to interactions with each other, contrary to the headpiece exposure observed in the pathogenic variant. The hydrophobic core presented in Figure 9

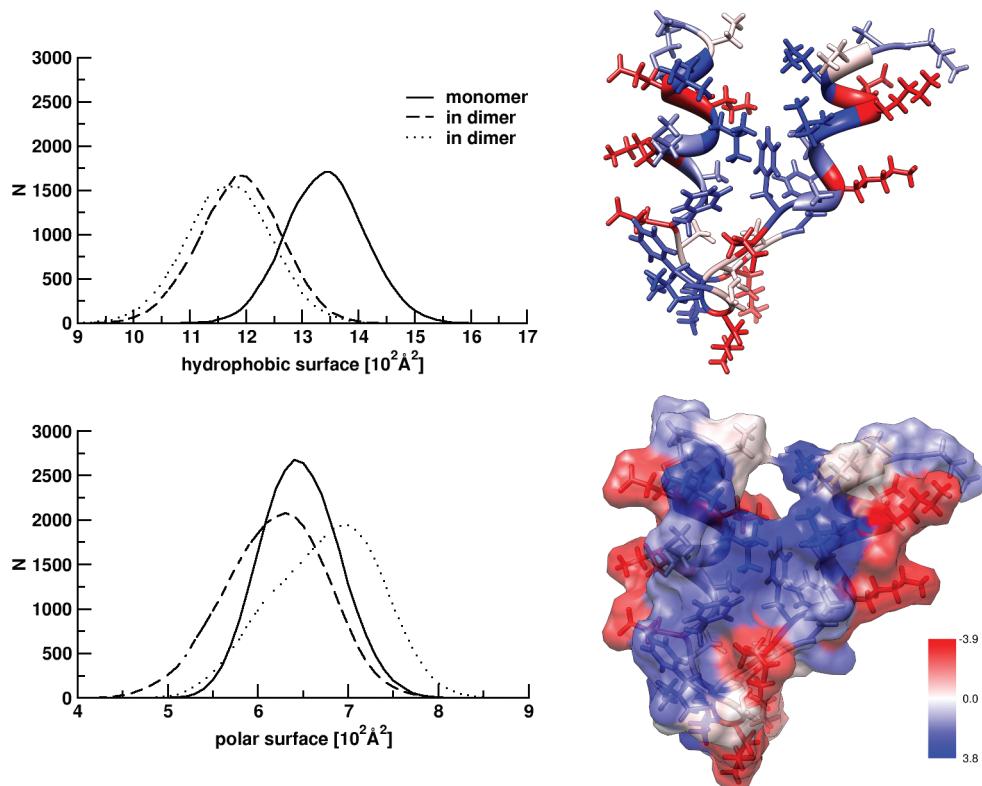


Figure 10. Histograms of the hydrophobic and polar surface of the Htt headpiece in the monomer and dimer. The full-atom and surface model of the headpiece dimer from CTMD simulations is also shown. The surface is colored according to the hydrophobicity scale of Kyte–Doolittle⁸¹ with hydrophobic surfaces in blue and polar in red.

and Figure S6 confirms that it is more difficult for the headpieces in the native fragment to interact with each other.

CONCLUSION

We performed replica-exchange molecular dynamics simulations to predict the structures of native and pathogenic Htt amino-terminal fragments. The mostly populated structural clusters of the native Htt N-terminal region derived from REMD at low temperatures closely resemble X-ray structures of the Htt17Q-EX1.⁵⁴ We find that corresponding sequences of the headpiece, polyQ, and polyP tracts of both Htt fragments fold to similar secondary structures. In the most populated state of N17Q^{Htt} the Htt headpiece forms a straight α -helix, but in N55Q^{Htt} a two-helix bundle is more probable. For the polyQ regions, α -helical configurations are dominant in both fragments but the helical content decreases upon increasing the temperature. Additionally, the polyP₍₁₁₎ sequence in N17Q^{Htt} forms mainly a straight PP II helix, and in the most populated cluster of N55Q^{Htt} this PP II helix is kinked. We find that in N17Q^{Htt}, contrary to N55Q^{Htt}, the PP II helix and the Htt amino-terminal α -helix are stabilized by hydrophobic interactions. In the elongated N55Q^{Htt} the lack of such hydrophobic interactions between the N-terminal α -helix and polyP₍₁₁₎ changes their structural preferences. Recent MD simulations⁶³ predicted that the most populated state of the isolated Htt headpiece is a two-helix bundle which corroborates with our observations.

The central idea of common hypotheses put forward to explain the HD pathogenesis is that the expanded polyQ tracts cause misfolding in monomeric proteins.^{8,13} These misfolded,

polyQ-extended monomers are likely to form oligomers that are toxic to neurons¹⁴ or involved in aberrant interactions with other cellular proteins causing neuronal dysfunction and ultimately cell death.⁸ The structural properties of N17Q^{Htt} and N55Q^{Htt} N-terminal fragments predicted in our REMD do not conform to the misfolding hypothesis and do not confirm a polyQ tract length dependent critical conformational transition.⁸² However, more recently, in the experiments conducted for synthetic peptide models of Htt exon 1, Thakur et al.¹⁶ proposed a different, complex aggregation mechanism that involves the Htt headpiece. According to this mechanism, investigated further by Williamson et al.,¹⁸ metastable aggregates are initially formed with cores consisting of Htt headpieces. The polyQ tracts are excluded from these cores. Next, amyloid nuclei consisting both of Htt headpieces and polyQ tracts develop from these initial aggregates. A similar in nature mechanism proposing a role of the 17-residue flanking headpiece was also recently described in ref 19.

The structures of native and pathogenic variants of Htt N-terminal fragments predicted here along with the simulations of dimerization of headpieces support the above headpiece-mediated aggregation mechanism. We observe that the native Htt fragment is more compact and conformationally stable due to the hydrophobic core composed of the Htt headpiece amphiphatic α -helix and the polyP₍₁₁₎ PP II helix. The pathogenic fragment lacks these hydrophobic interactions, and, therefore, the Htt headpiece in N55Q^{Htt} is more exposed to solvent. Such a solvent exposed Htt headpiece (attached at the beginning of a long α -helical polyQ₍₅₅₎ fragment) is likely to interact with the N-termini of other polyQ-expanded Htt monomers via hydrophobic interactions which stabilizes initial aggregates. According

to our simulations native and pathogenic variants possess similar secondary structures in the polyQ region. One may thus speculate that the postulated conformational transition in the expanded polyQ stretch takes place after the initial nucleus has formed and that the final aggregates are indeed rich in β -sheets as described by Thakur et al.¹⁶ and Williamson et al.¹⁸ However, while these previous works suggest that the Htt headpiece exists predominantly as a compact coil that can be unfolded by longer polyQ fragments and thus form initial aggregates more easily, we do not observe unfolding of the Htt headpiece by the elongated polyQ tract. Our main conclusion regarding the role of expanded polyQ tracts in Htt aggregation is that they simply solvent expose the α -helical Htt headpiece by disrupting its hydrophobic interactions with the polyP tract and destabilizing the hydrophobic core formed in the native Htt N-terminal region. Such exposure to solvent makes the headpiece minimize its hydrophobic surface by self-aggregation.

■ ASSOCIATED CONTENT

S Supporting Information. Potential energy distributions for different replicas (Figure S1), heat capacities as a function of temperature (Figure S2), single replica residence times at different temperatures (Figure S3), rmsd and gyration radii plotted as a function of MD simulation time at 300 K (Figure S4), cartoon representation of the X-ray structure of the Htt exon 1 fragment 2IOW (Figure S5), and snapshots from explicit solvent MD simulations for N17Q^{Htt} (Figure S6). This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: mdlugosz@icm.edu.pl. Tel.: +48 22 5540 818. Fax: +48 22 5540 801.

■ ACKNOWLEDGMENT

We thank Beata Wielgus-Kutrowska, Marta Narczyk, and Agnieszka Bzowska for helpful discussions. This work was supported by the Foundation for Polish Science (Focus program) and University of Warsaw (Grants BST and G31-4). J.T. acknowledges support from Polish Ministry of Science and Higher Education (Grant N N301 245236) and the Foundation for Polish Science Team project (TEAM/2009-3/8) cofinanced by the European Regional Development Fund operated within the Innovative Economy Operational Program.

■ REFERENCES

- (1) Orr, H. T.; Zoghbi, H. Y. *Annu. Rev. Neurosci.* **2007**, *30*, 575–621.
- (2) Tobin, A. J.; Singer, E. R. *Trends Cell Biol.* **2000**, *10*, 531–536.
- (3) Bates, G. P. *Nat. Rev. Genet.* **2005**, *6*, 766–773.
- (4) Mangiarini, L.; Sathasivam, K.; Seller, M.; Cozens, B.; Harper, A.; Hetherington, C.; Lawton, M.; Trottier, Y.; Lehrach, H.; Davies, S. W.; Bates, G. P. *Cell* **1996**, *87*, 493–506.
- (5) Meriin, A. B.; Zhang, X.; He, X.; Newnam, G. P.; Chernoff, Y. O.; Sherman, M. Y. *J. Cell Biol.* **2002**, *157*, 997–1004.
- (6) DiFiglia, M.; Sapp, E.; Chase, K. O.; Davies, S. W.; Bates, G. P.; Vonsattel, J. P.; Aronin, N. *Science* **1997**, *277*, 1990–1993.
- (7) Michalik, A.; Broeckhoven, C. V. *Hum. Mol. Genet.* **2003**, *12*, R173–R186.
- (8) Williams, A. J.; Paulson, H. L. *Trends Neurosci.* **2008**, *31*, 521–528.
- (9) Goldberg, Y.; Nicholson, D.; Rasper, D.; Kalchman, M.; Koide, H.; Graham, R.; Bromm, M.; Kazemi-Esfarjani, P.; Thornberry, N.; Vaillancourt, J.; Hayden, M. *Nat. Genet.* **1996**, *13*, 442–449.
- (10) Wellington, C. L.; Ellerby, L. M.; Hackam, A. S.; Margolis, R. L.; Trifiro, M. A.; Singaraja, R.; McCutcheon, K.; Salvesen, G. S.; Propp, S. S.; Bromm, R.; et al. *J. Biol. Chem.* **1998**, *273*, 9158–9167.
- (11) Gusella, J. F.; MacDonald, M. E. *Nat. Rev. Neurosci.* **2000**, *1*, 109–115.
- (12) Ross, C. A. *Neuron* **2002**, *35*, 819–822.
- (13) Nagai, Y.; Inui, T.; Popiel, H. A.; Fujikake, N.; Hasegawa, K.; Urade, Y.; Goto, Y.; Naiki, H.; Toda, T. *Nat. Struct. Mol. Biol.* **2007**, *14*, 332–340.
- (14) Truant, R.; Atwal, R. S.; Desmond, C.; Munsie, L.; Tran, T. *FEBS J.* **2008**, *275*, 4252–4262.
- (15) Sugaya, K.; Matsubara, S. *Mol. Neurodegener.* **2009**, *4*, 29–40.
- (16) Thakur, A. K.; Jayaraman, M.; Mishra, R.; Thakur, M.; Chellgren, V. M.; Byeon, I. L.; Anjum, D. H.; Kodali, R.; Creamer, T. P.; Conway, J. F.; Gronenborn, A. M.; Wetzel, R. *Nat. Struct. Biol.* **2009**, *16*, 380–389.
- (17) Tam, S.; Spiess, C.; Auyeung, W.; Joachimiak, L.; Chen, B.; Poirier, M. A.; Frydman, J. *Nat. Struct. Mol. Biol.* **2009**, *16*, 1279–1286.
- (18) Williamson, T. E.; Vitalis, A.; Crick, S. L.; Pappu, R. *J. Mol. Biol.* **2010**, *396*, 1295–1309.
- (19) Lieberman, S. W.; Meredith, S. C. *Nat. Chem. Biol.* **2010**, *6*, 7–8.
- (20) Sivanandam, V. N.; Jayaraman, M.; Hoop, C. L.; Kodali, R.; Wetzel, R.; van der Wel, P. C. A. *J. Am. Chem. Soc.* **2011**, *133*, 4558–4566.
- (21) Chen, S.; Berthelier, V.; Yang, W.; Wetzel, R. *J. Mol. Biol.* **2001**, *311*, 173–182.
- (22) Klein, F. A. C.; Pastore, A.; Massino, L.; Zeder-Lutz, G.; Nierengarten, H.; Oulad-Abdelghani, M.; Altschuh, D.; Mandel, J.-L.; Trottier, Y. *J. Mol. Biol.* **2007**, *371*, 235–244.
- (23) Singer, D.; Zauner, T.; Genz, M.; Hoffmann, R.; Zuchner, T. *J. Pept. Sci.* **2010**, *16*, 358–363.
- (24) Altschuler, E. L.; Hud, N. V.; Mazrimas, J. A.; Rupp, B. *J. Pept. Res.* **1997**, *50*, 73–75.
- (25) Crick, S. L.; Jayaraman, M.; Frieden, C.; Wetzel, R.; Pappu, R. V. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 16764–16769.
- (26) Bhattacharyya, A.; Thakur, A. K.; Chellgren, V. M.; Thiagarajan, G.; Williams, A. D.; Chellgren, B. W.; Creamer, T. P.; Wetzel, R. *J. Mol. Biol.* **2006**, *355*, 524–535.
- (27) Darnell, G. D.; Derryberry, J. M.; Kurutz, J. W.; Meredith, S. C. *Biophys. J.* **2009**, *97*, 2295–2305.
- (28) Chellgren, B. W.; Miller, A.-F.; Creamer, T. P. *J. Mol. Biol.* **2006**, *361*, 362–371.
- (29) Darnell, G.; Orgel, J. P.; Pahl, R.; Meredith, S. C. *J. Mol. Biol.* **2007**, *374*, 688–704.
- (30) Armen, R. S.; Bernard, B. M.; Day, R.; Alonso, D. O.; Daggett, V. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13433–13438.
- (31) Khare, S. D.; Ding, F.; Gwanmesia, K. N.; Dokholyan, N. V. *PLoS Comput. Biol.* **2005**, *1*, 230–235.
- (32) Tsukamoto, K.; Shimizu, H.; Ishida, T.; Akiyama, Y.; Nukina, N. *J. Mol. Struct.* **2006**, *778*, 85–95.
- (33) Wang, X. L.; Vitalis, A.; Wyczalkowski, M. A.; Pappu, R. V. *Proteins* **2006**, *63*, 297–311.
- (34) Marchut, A. J.; Hall, C. K. *Proteins: Struct., Funct., Bioinf.* **2007**, *66*, 96–109.
- (35) Vitalis, A.; Wang, X.; Pappu, R. V. *J. Mol. Biol.* **2009**, *103*, 16764–16769.
- (36) Laghaei, R.; Mousseau, N. *J. Chem. Phys.* **2010**, *132*, No. 165102.
- (37) Zhang, Q. C.; Yeh, T.; Leyva, A.; Frank, L. G.; Miller, J.; Kim, Y. E.; Langen, R.; Finkbeiner, S.; Amzel, M. L.; Ross, C. A.; Poirier, M. A. *J. Biol. Chem.* **2011**, *286*, 8188–8196.
- (38) Perutz, M. F.; Finch, J. T.; Berriman, J.; Lesk, A. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 5591–5595.
- (39) Perutz, M. F. *Curr. Opin. Struct. Biol.* **1996**, *6*, 848–858.
- (40) Thakur, A. K.; Wetzel, R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 17014–17019.
- (41) Sharma, D.; Shinchuk, L. M.; Inouye, H.; Wetzel, R.; Kirschner, D. A. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 398–411.

- (42) Takahashi, T.; Kikuchi, S.; Katada, S.; Nagai, Y.; Nishizawa, M.; Onodera, O. *Hum. Mol. Genet.* **2008**, *17*, 345–356.
- (43) Chen, S.; Ferrone, F. A.; Wetzel, R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11884–11889.
- (44) Lee, C. C.; Walters, R. H.; Murphy, R. M. *Biochemistry* **2007**, *46*, 12810–12820.
- (45) Marchut, A. J.; Hall, C. K. *Biophys. J.* **2006**, *90*, 4574–4584.
- (46) Nozaki, K.; Onodera, O. *Neuroreport* **2001**, *12*, 3357–3364.
- (47) Bulone, D.; Masino, L.; Thomas, D. J.; San Biagio, P. L.; Pastore, A. *PLoS ONE* **2006**, *1*, No. e111.
- (48) Rockabrand, E.; Slepko, N.; Pantalone, A.; Nukala, V. N.; Kazantsevand, A.; Marshand, J. L.; Sullivan, P. G.; Steffan, J. S.; Sensi, S. L.; Thompson, L. M. *Hum. Mol. Genet.* **2007**, *16*, 61–77.
- (49) Masino, L.; Kelley, G.; Leonard, K.; Trottier, Y.; Pastore, A. *FEBS Lett.* **2002**, *513*, 267–272.
- (50) Ignatova, Z.; Giersch, L. M. *J. Biol. Chem.* **2006**, *281*, 12959–12967.
- (51) Robertson, A. L.; Bottomley, S. P. *Curr. Med. Chem.* **2010**, *17*, 3058–3068.
- (52) Duennwald, M. L.; Jagadish, S.; Muchowski, P. J.; Lindquist, S. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 11045–11050.
- (53) Lakhani, V. V.; Ding, F.; Dokholyan, N. V. *PLOS Comput. Biol.* **2010**, *6*, No. e1000772.
- (54) Kim, M. W.; Chelliah, Y.; Kim, S. W.; Otwinowski, Z.; Bezprozvanny, I. *Structure* **2009**, *17*, 1205–1212.
- (55) Hansmann, U. H. E.; Okamoto, Y.; Eisenmenger, F. *Chem. Phys. Lett.* **1996**, *259*, 321–330.
- (56) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Biopolymers* **2001**, *60*, 96–123.
- (57) Nymeyer, H.; Gnanakaran, S.; Garcia, A. E. *Methods Enzymol.* **2004**, *383*, 119–149.
- (58) Cheng, X.; Cui, G.; Hornak, V.; Simmerling, C. *J. Phys. Chem. B* **2005**, *109*, 8220–8230.
- (59) Lei, H.; Wu, C.; Liu, H.; Duan, Y. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 4925–4930.
- (60) Chowdhury, S.; Lei, H.; Duan, Y. *J. Phys. Chem. B* **2005**, *109*, 9073–9081.
- (61) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (62) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (63) Kelley, N. W.; Huang, X.; Tam, S.; Spiess, C.; Frydman, J.; Pande, V. S. *J. Mol. Biol.* **2009**, *388*, 919–927.
- (64) Wu, C.; Lei, H.; Duan, Y. *J. Am. Chem. Soc.* **2005**, *127*, 13530–13537.
- (65) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. *J. Chem. Theory Comput.* **2007**, *3*, 156–169.
- (66) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (67) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613–4621.
- (68) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (69) Darden, T.; Perera, L.; Li, L.; Pedersen, L. *Structure* **1999**, *7*, R55–R60.
- (70) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (71) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- (72) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angew. Chem., Int. Ed.* **1999**, *38*, 236–240.
- (73) Atwal, R. S.; Xia, J.; Pinchev, D.; Taylor, J.; Epanad, R. M.; Truant, R. *Hum. Mol. Genet.* **2007**, *16*, 2600–2615.
- (74) Vitalis, A.; Pappu, R. V. *J. Comput. Chem.* **2008**, *30*, 673–699.
- (75) Ding, F.; Nie, H.; Dokholyan, N. V. *Structure* **2008**, *16*, 1010–1018.
- (76) Lazaridis, T.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 133–152.
- (77) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.
- (78) Shell, M. S.; Ritterson, R.; Dill, K. A. *J. Phys. Chem. B* **2008**, *112*, 6878–6886.
- (79) Reiersen, H.; Rees, A. R. *Trends Biochem. Sci.* **2001**, *26*, 679–684.
- (80) Hinderaker, M. P.; Raines, R. T. *Protein Sci.* **2003**, *12*, 1188–1194.
- (81) Kyte, J.; Doolittle, R. F. *J. Mol. Biol.* **1982**, *157*, 105–132.
- (82) Chen, S.; Ferrone, F. A.; Wetzel, R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11884–11889.