Probing initial transient oligomerization events facilitating Huntingtin fibril nucleation at atomic resolution by relaxation-based NMR

Samuel A. Kotler[a], Vitali Tugarinov[a], Thomas Schmidt[a], Alberto Ceccon[a], David S. Libich[a,1], Rodolfo Ghirlando[b], Charles D. Schwieters[c], and G. Marius Clore[a,2]

[a]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520; [b]Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0530; and [c]Imaging Sciences Laboratory Center for Information Technology, National Institutes of Health, Bethesda, MD 20892-5624

Contributed by G. Marius Clore, December 28, 2018 (sent for review December 13, 2018; reviewed by Lewis E. Kay and David J. Weber)

The N-terminal region of the huntingtin protein, encoded by exon-1, comprises an amphiphilic domain (htt^{NT}), a polyglutamine (Q$_n$) tract, and a proline-rich sequence. Polyglutamine expansion results in an aggregation-prone protein responsible for Huntington's disease. Here, we study the earliest events involved in oligomerization of a minimalistic construct, htt^{NT}Q$_7$, which remains largely monomeric over a sufficiently long period of time to permit detailed quantitative NMR analysis of the kinetics and structure of sparsely populated (≲2%) oligomeric states, yet still eventually forms fibrils. Global fitting of concentration-dependent relaxation dispersion, transverse relaxation in the rotating frame, and exchange-induced chemical shift data reveals a bifurcated assembly mechanism in which the NMR observable monomeric species either self-associates to form a productive dimer (τ_{ex} ∼ 30 μs, K_{diss} ∼ 0.1 M) that goes on to form a tetramer (τ_{ex} ≲ 25 μs; K_{diss} ∼ 22 μM), or exchanges with a "nonproductive" dimer that does not oligomerize further (τ_{ex} ∼ 400 μs; K_{diss} ∼ 0.3 M). The excited state backbone chemical shifts are indicative of a contiguous helix (residues 3–17) in the productive dimer/tetramer, with only partial helical character in the nonproductive dimer. A structural model of the productive dimer/tetramer was obtained by simulated annealing driven by intermolecular paramagnetic relaxation enhancement data. The tetramer comprises a D_2 symmetric dimer of dimers with largely hydrophobic packing between the helical subunits. The structural model, validated by EPR distance measurements, illuminates the role of the htt^{NT} domain in the earliest stages of prenucleation and oligomerization, before fibril formation.

prenucleation | kinetics | relaxation-based NMR | EPR | structure

Polyglutamine [poly(Q)] expansion, arising from mutations that extend the length of glutamine-encoding CAG repeats, is associated with a number of neurodegenerative diseases (1), the best known of which is Huntington's disease, an autosomal dominant condition characterized pathologically by widespread neuronal degeneration and clinically by involuntary jerky movements (hence the name Huntington's chorea), dementia, and ultimately death (2, 3). The gene responsible is Huntingtin (*HTT*), and the poly(Q) tract is encoded within exon 1 (3). Although the Huntingtin protein (htt) is very large (∼350 kDa), proteolysis (4) and/or incomplete mRNA splicing of *HTT* (5) generates mutated N-terminal fragments that aggregate to form neuronal inclusion bodies in pathological states (6). The N-terminal fragment (htt^{Ex1}), corresponding to the first exon of *HTT*, comprises three domains (3): a 16-residue N-terminal amphiphilic sequence (htt^{NT}), a poly(Q) tract of variable length, and a proline-rich domain (Fig. 1*A*). The mean length of the poly(Q) repeat ranges from 17 to 20; lengths of 36 or greater result in Huntington's disease (3), with the age of clinical onset being inversely correlated to the length of the poly(Q) segment (2, 7).

Htt^{Ex1} self-assembles to form prefibrillar oligomers and fibrils, both of which are neurotoxic in cell culture (2, 8). The rate of poly(Q) aggregation is greatly accelerated by the presence of the htt^{NT} sequence and down-regulated by the proline-rich domain (9, 10). Solid-state NMR has shown that htt^{Ex1} fibrils consist of a static polyglutamine β-hairpin/β-sheet core, connected by interdigitating glutamine side chains (11–13) and surrounded by htt^{NT} helices of intermediate dynamics (14–16). It has been postulated that the htt^{NT} sequence self-associates to form a helical coiled coil, and mutations engineered to enhance coiled-coil propensity promote aggregation, while those that disrupt coiled-coil formation have the reverse effect (17). These results, as well as other biophysical studies (18), have led to the hypothesis that the htt^{NT} domain of htt^{Ex1}, as well as of other poly(Q) proteins, is a functional switch for the initial nucleation event that eventually leads to protofibril and fibril formation (17).

The earliest transient association events in htt^{Ex1} protofibril and fibril assembly, especially those involving sparsely populated, spectroscopically invisible states, are difficult to study as htt^{Ex1} aggregates relatively rapidly in vitro even for nonpathogenic poly(Q) lengths (18). However, amyloid nucleation of htt^{NT}-poly(Q) peptides (htt^{NT}Q$_n$) can be slowed down by reducing the length of

Significance

Huntington's disease is a fatal neurodegenerative condition arising from polyglutamine expansion within the Huntingtin protein leading to fibril accumulation in neurons. The initial multimerization events occur on the submillisecond timescale and involve sparsely populated species that can be probed at atomic resolution by NMR. Using a minimalistic construct comprising the N-terminal amphiphilic domain and seven glutamines, we uncover a branched oligomerization pathway, one leading to a tetramer comprising a dimer of coiled-coil helical dimers, and the other resulting in a nonproductive, partially helical, dimer. The results illuminate the contribution of the N-terminal amphiphilic domain in prenucleation events that precede fibril formation.

Author contributions: S.A.K., V.T., T.S., A.C., R.G., C.D.S., and G.M.C. designed research; S.A.K., V.T., T.S., A.C., D.S.L., R.G., C.D.S., and G.M.C. performed research; D.S.L. contributed new reagents/analytic tools; S.A.K., V.T., T.S., R.G., C.D.S., and G.M.C. analyzed data; and S.A.K., V.T., and G.M.C. wrote the paper.

Reviewers: L.E.K., University of Toronto; and D.J.W., University of Maryland School of Medicine.

The authors declare no conflict of interest.

Data deposition: Atomic coordinates of the tetramer, as well as experimental restraints, have been deposited in the Protein Data Bank, www.wwpdb.org (PDB ID code 6N8C), and backbone assignments for the tetramer have been deposited in the Biological Magnetic Resonance Data Bank, www.bmrb.wisc.edu/ (BMRB ID code 30545).

[1]Present address: Department of Biochemistry and Structural Biology, University of Texas Health Science Center at San Antonio, San Antonio, TX 78229.

[2]To whom correspondence should be addressed. Email: mariusc@mail.nih.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1821216116/-/DCSupplemental.

Published online February 11, 2019.

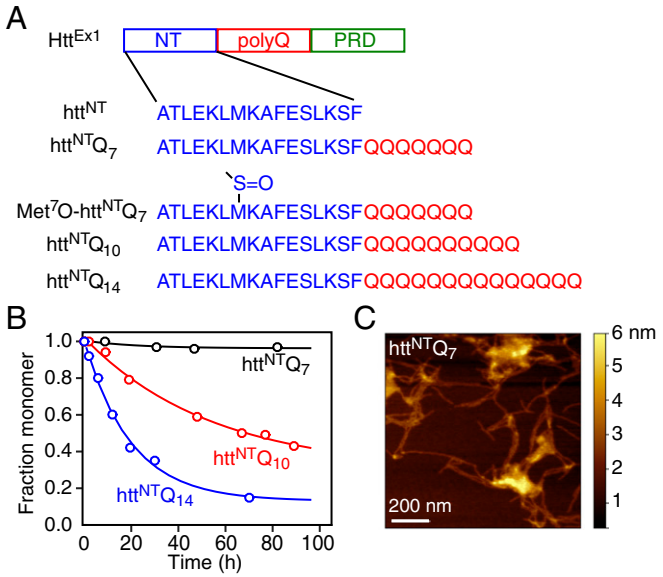

Fig. 1. Characterization of htt$^{NT}Q_n$. (*A*) Organization of htt^{Ex1} and htt$^{NT}Q_n$ sequences used in the current work. NT, N-terminal amphiphilic sequence; poly(Q), polyglutamine domain; PRD, proline-rich domain. (*B*) Fraction of monomer as a function of incubation time at 10 °C, measured from the intensity of the amide proton envelope in the first increment of a ^{1}H–^{15}N HSQC spectrum, for htt$^{NT}Q_n$ with $n = 7$ (1 mM), 10 (100 μM), and 14 (100 μM). (*C*) Atomic force microscopy image of fibrils of htt$^{NT}Q_7$ obtained after incubation for 3 weeks at 10 °C of 0.6 mM htt$^{NT}Q_7$ in 20 mM phosphate buffer, pH 6.5, and 50 mM NaCl.

the poly(Q) repeat (19). Here, we take advantage of a construct comprising only seven glutamines, htt$^{NT}Q_7$, to probe the kinetics and structural features of highly transient multimeric species (with lifetimes less than 1 ms) formed during the earliest association events by quantitative analysis of concentration-dependent Carr–Purcell–Meiboom–Gill (CPMG) relaxation dispersion (20, 21), rotating frame $R_{1\rho}$ dispersion (22), transverse relaxation in the rotating frame and exchange-induced chemical shift (23) data, as well as NMR intermolecular paramagnetic relaxation enhancement (PRE) measurements (24, 25) and pulse double electron–electron resonance (DEER) (26) and continuous-wave (CW) (27) EPR. We demonstrate the existence of two branching pathways: the first involves self-association to a tetrameric helical bundle via a helical coiled-coil dimer; the second results in the formation of a nonproductive dimer (with partial helical character). Oxidation of the single methionine at position 7 of the htt^{NT} region abolishes all association, as does removal of the Q_7 sequence. These results provide a detailed mechanistic picture of oligomer assembly occurring at the very earliest stages along the aggregation pathway of htt^{Ex1} that eventually leads to fibril nucleation and formation.

Results and Discussion

Seven Glutamines Are Sufficient to Induce Slow Fibril Formation. Given the necessity of stable samples over prolonged periods of time needed for detailed analysis of exchange dynamics by NMR, we focused on four N-terminal htt^{Ex1} peptide fragments (Fig. 1*A*): the 16-residue N-terminal peptide, htt^{NT}, with no glutamine repeats [as the N-terminal methionine is likely cleaved in vivo (28)]; a 23-residue peptide with seven glutamines C-terminal to the htt^{NT} sequence, htt$^{NT}Q_7$ (*SI Appendix*, Fig. S1); (Met^{7}O)-htt$^{NT}Q_7$ in which the side chain of Met7 was oxidized to a sulfoxide by mild treatment with hydrogen peroxide (*SI Appendix*, Fig. S2); and 26- and 30-residue peptides with 10 and 14 glutamines C-terminal to the htt^{NT} sequence, htt$^{NT}Q_{10}$ and htt$^{NT}Q_{14}$. The ^{1}H–^{15}N correlation spectra of all constructs are typical of intrinsically disordered peptides with very limited backbone amide proton spectral dispersion (*SI Appendix*, Figs. S1*A* and S2), and the backbone secondary chemical shifts of htt$^{NT}Q_7$ give no indication for any significant secondary structure propensity (*SI Appendix*, Fig. S1*B*) under the experimental conditions employed throughout the current work (5 or 10 °C, 20 mM sodium phosphate, pH 6.5, 50 mM NaCl). We followed aggregation by monitoring the decrease in intensity of the ^{15}N-labeled monomer ^{1}H–^{15}N cross-peaks arising from the formation of very high–molecular-weight species whose resonances are broadened beyond the limits of detection. Both htt^{NT} and (Met^{7}O)-htt$^{NT}Q_7$ are stable monomers under these conditions. Also, htt$^{NT}Q_7$ is largely monomeric over a period of 3 weeks from initial dissolution at 10 °C [see analytical ultracentrifugation (AUC) data in *SI Appendix*, Fig. S3] with a 1 mM solution displaying only a minimal decrease in ^{1}H–^{15}N cross-peak intensities (>95% monomer, Fig. 1*B*). Only after about 3 weeks can the presence of amyloid-like fibrils be detected by atomic force microscopy (Fig. 1*C*). Under the same conditions, however, 100 μM solutions of htt$^{NT}Q_{10}$ and htt$^{NT}Q_{14}$ aggregate too rapidly at 10 °C for detailed NMR studies with monomer half-lives of ∼65 and ∼18 h, respectively (Fig. 1*B*). Thus, htt$^{NT}Q_7$ provides an ideal system for probing monomer–oligomer exchange dynamics on the submillisecond timescale by relaxation dispersion NMR under conditions where the populations of the oligomeric states are low.

Relaxation Dispersion, Rotating Frame Transverse Relaxation, and Exchange-Induced Shifts. The htt$^{NT}Q_7$ construct displays concentration-dependent ^{13}Cα exchange-line broadening (^{13}Cα-R_{ex}), ^{13}Cα exchange-induced shifts (^{13}Cα-δ_{ex}), and ^{15}N transverse relaxation rates (^{15}N-R_2^{1kHz}) measured from ^{15}N-$R_{1\rho}$ experiments recorded with a 1-kHz radiofrequency field (RF) spin lock (Fig. 2). In contrast, neither htt^{NT} nor (Met^{7}O)-htt$^{NT}Q_7$ show any change in these parameters with concentration. Thus, htt$^{NT}Q_7$ must undergo rapid interconversion between the major monomeric species and minor, sparsely populated, higher-order oligomeric species. In addition, ^{15}N-ΔR_2^{1kHz} is correlated to ^{15}N-δ_{ex}^{2} (Fig. 2*D* and *SI Appendix*, Figs. S4 and S5). Thus, the increase in ^{15}N-R_2^{1kHz} as a function of concentration is a hallmark of a fast exchange process that is not suppressed by the 1-kHz RF spin-lock field, and can be attributed to exchange between species with different chemical shifts (chemical exchange line broadening) on the submillisecond timescale, as opposed to lifetime line broadening arising from exchange between species with very different molecular weights and tumbling times (29). Furthermore, ^{15}N dark state exchange saturation transfer experiments (30) show no evidence of any detectable exchange process between the NMR visible monomer and large molecular weight species.

To quantitatively explore the exchange phenomena for htt$^{NT}Q_7$, we collected ^{13}Cα (Fig. 3*A*) and ^{15}N (Fig. 4*A*) CPMG relaxation dispersion experiments at three (0.4, 0.75, and 1 mM) concentrations, ^{15}N $R_{1\rho}$ dispersion experiments at one concentration (1 mM, Fig. 4*B*), ^{13}Cα (Fig. 3*B* and *SI Appendix*, Fig. S6) and ^{15}N exchange-induced shifts (Fig. 4*C*, *Top* and *SI Appendix*, Fig. S4), and ^{15}N-R_2^{1kHz} data (Fig. 4*C*, *Bottom*, and *SI Appendix*, Fig. S5) at 13 concentrations from 50 μM to 1.2 mM. (Note that the number and range of concentrations used for the CPMG relaxation dispersion experiments was limited by signal-to-noise considerations; further CPMG relaxation dispersion experiments could not be carried out above 1 mM as some aggregation, resulting in a 10–15% loss of NMR signal intensity, occurs over a period of 48 h at 1.2 mM, close to the solubility limit of htt$^{NT}Q_7$; exchange-induced shift and ^{15}N-R_2^{1kHz} data, however, can be recorded in a short period of time, allowing access to concentrations up to 1.2 mM.) These data, summarized in Table 1, were fit simultaneously by solving the appropriate McConnell equations (31) corresponding to a given kinetic scheme, optimizing the relevant rate constants, the differences in chemical

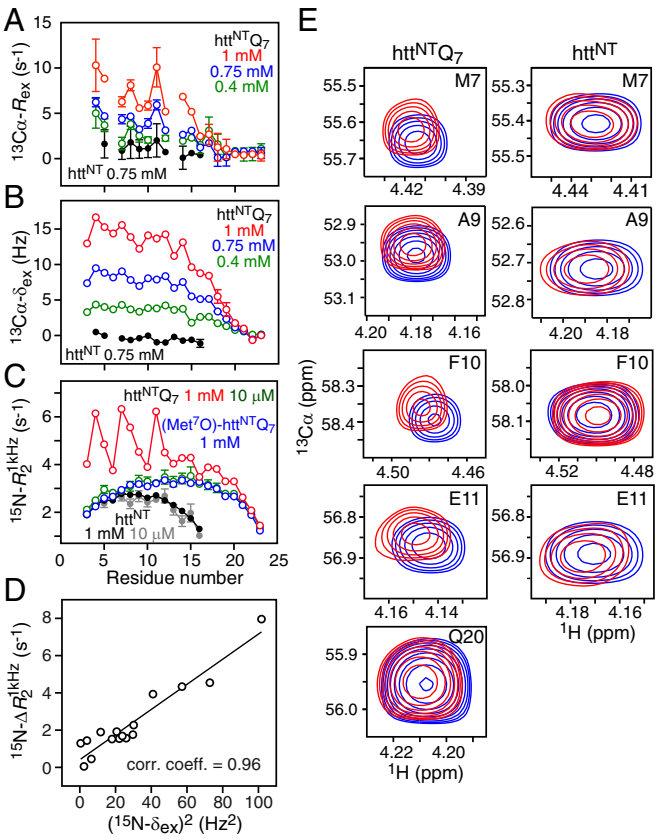

Fig. 2. Concentration-dependent exchange dynamics of $htt^{NT}Q_7$ observed by NMR. (*A*) $^{13}C\alpha$ exchange line-broadening measured by CPMG relaxation dispersion ($^{13}C\alpha\text{-}R_{ex} = R_{2,eff}^{25Hz} - R_{2,eff}^{2,000Hz}$) at 600 MHz and (*B*) $^{13}C\alpha$ exchange induced shifts ($^{13}C\alpha\text{-}\delta_{ex} = \delta_{obs} - \delta_{ref}$) for $htt^{NT}Q_7$ (800 MHz) at three concentrations (0.4, 0.75, and 1 mM) and htt^{NT} (900 MHz) at one concentration (0.75 mM) and 5 °C. For $htt^{NT}Q_7$, the reference shifts, δ_{ref}, in the absence of exchange were determined by fitting the concentration dependence of the observed shifts, δ_{obs}, for each residue to a second-order polynomial (*SI Appendix, SI Materials and Methods*); for htt^{NT}, the reference shifts are obtained from the spectrum recorded on a 20 μM sample. (*C*) $^{15}N\text{-}R_2^{1kHz}$ transverse relaxation rate profiles obtained from $^{15}N\text{-}R_{1\rho}$ measurements at 600 MHz and 10 °C recorded with a spin-lock radiofrequency field strength of 1 kHz for $htt^{NT}Q_7$ (10 μM and 1 mM), (Met^7O)-$htt^{NT}Q_7$ (1 mM), and htt^{NT} (10 μM and 1 mM). (*D*) Correlation of $^{15}N\text{-}R_2^{1kHz}$ and $(^{15}N\text{-}\delta_{ex})^2$ for $htt^{NT}Q_7$ at 800 MHz and 5 °C. (*E*) Expansion of selected regions of the 900-MHz ^1H–^{13}C constant time HSQC spectra of 0.8 mM (blue) and 20 μM (red) samples of $htt^{NT}Q_7$ (*Left*) and htt^{NT} (*Right*) at 5 °C. Error bars represent 1 SD (when not shown, they are within the circles representing the experimental data).

shifts between major and minor species for each residue, and the transverse relaxation rates for each residue at each concentration (see *SI Appendix* for full details of the relevant kinetic equations and of the global fitting procedure; *SI Appendix*, Eqs. **S1–S6**).

The distinct curvature in the concentration dependence of the exchange-induced shift (Figs. 3*B* and 4*C* and *SI Appendix*, Figs. S4 and S6–S9) and $^{15}N\text{-}R_2^{1kHz}$ (Fig. 4*C* and *SI Appendix*, Fig. S5) data are indicative of the formation of higher-order (>2) oligomers. The simplest model that can fit all of the NMR data simultaneously is depicted in Fig. 5*A* and involves a branching pathway. The monomer, **P**, is the NMR observable species. In the first branch, **P** self-associates to form a productive dimer, $\mathbf{P_2}$, that further self-associates to form a tetramer, $\mathbf{P_4}$. In the second branch, **P** self-associates to form a "nonproductive" dimer, $\mathbf{P_2^*}$, which does not undergo further oligomerization.

Alternative simpler kinetic models were also considered but eliminated on the basis of goodness-of-fit in the simultaneous analysis of all available concentration-dependent NMR data, and/or the ability of a given kinetic scheme to provide physically plausible parameters of exchange, transverse relaxation rates (R_2^P) of the major (monomeric) species and chemical shifts of the minor (oligomeric) species. Two-site exchange models such as $\mathbf{P} \leftrightarrow \mathbf{P_2}$, $\mathbf{P} \leftrightarrow \mathbf{P_3}$, and $\mathbf{P} \leftrightarrow \mathbf{P_4}$, as well as a three-site branched exchange model ($\mathbf{P_2} \leftrightarrow \mathbf{P} \leftrightarrow \mathbf{P_4}$), fail to satisfy the relaxation dispersion and exchange-induced shift data simultaneously, and do not reproduce the concentration dependence of the exchange-induced shifts. Although the $\mathbf{P} \leftrightarrow \mathbf{P_2}$ two-site exchange model can approximately fit the available CPMG relaxation dispersion data alone, unphysically high fitted values of R_2^P are obtained at higher peptide concentrations, implying the existence of another (very fast) process that is not accounted for by two-site exchange. Three-site exchange models comprising an on-pathway dimeric intermediate leading either to a trimer or tetramer (32) were also investigated. The latter fit the experimental data slightly worse than the bifurcated four-site exchange model shown in Fig. 5 but, more importantly, can be disqualified as they result in (*i*) unrealistically high populations of the intermediate dimeric species ($\mathbf{P_2}$), typically exceeding ~10% at the highest concentration (1.2 mM) used in the NMR experiments, that are inconsistent with other biophysical data obtained in the present work (e.g., analytical ultracentrifugation; *SI Appendix*, Fig. S3), and (*ii*) unphysically large (by absolute magnitude) values of the resulting changes in $^{13}C\alpha$ and ^{15}N chemical shifts ($\Delta\omega$) of at least one of the involved oligomeric species.

Because of the complexity of the scheme in Fig. 5, we made the simplifying assumption that the $^{13}C\alpha$ and ^{15}N chemical shifts of the productive dimer and tetramer are the same. This is reasonable as $^{13}C\alpha$ and ^{15}N backbone shifts are largely affected by secondary structure (33) which would be predicted to be the same in the productive dimer and tetramer, if the tetramer is a dimer of dimers. Exchange between the monomer ($\mathbf{P}$) and nonproductive dimer ($\mathbf{P_2^*}$) occurs on a timescale of ~400 μs and is largely probed by CPMG relaxation dispersion; the exchange processes between the monomer ($\mathbf{P}$) and the productive dimer ($\mathbf{P_2}$) and between the productive dimer ($\mathbf{P_2}$) and the tetramer ($\mathbf{P_4}$) are approximately an order of magnitude faster and occur on timescales of ~30 and $\lesssim$ 25 μs, respectively, that are largely characterized by $R_{1\rho}$ dispersion, transverse relaxation in the rotating frame, and exchange-induced chemical shifts (23). The effects of the interplay of the fast and slow processes on both the CPMG relaxation dispersion profiles and the concentration dependence of the exchange-induced chemical shifts is discussed in detail in *SI Appendix* and *SI Appendix*, Fig. S10. Hence, a combination of these NMR techniques in conjunction with simultaneous analysis of all concentration-dependent data are necessary for a detailed characterization of the kinetics of oligomerization depicted in Fig. 5.

Initial fits indicated that the association (k_2) and dissociation (k_{-2}) rate constants for the $\mathbf{P_2} \leftrightarrow \mathbf{P_4}$ exchange process could not be determined independently with any degree of certainty. We therefore carried out a grid search in which k_{-2} was held fixed at values spanning from 10^4 to 10^5 s^{-1}, while optimizing all of the remaining parameters. The results are summarized in Fig. 6. Briefly, χ^2 increases as the value of k_{-2} falls below about 4×10^4 s^{-1}; for values of $k_{-2} \gtrsim 4 \times 10^4 s^{-1}$, χ^2 remains constant, and the values of k_1 and k_{-1} for the $\mathbf{P_2} \leftrightarrow \mathbf{P_4}$ exchange remain essentially unchanged. k_2 and k_{-2} are highly correlated (correlation coefficient of 0.99), and the values of k_3 and k_{-3} for the $\mathbf{P} \leftrightarrow \mathbf{P_2^*}$ exchange remain stable for all values of k_{-2}. Finally, the equilibrium dissociation constants, K_1^{diss}, K_2^{diss}, and K_3^{diss} for the $\mathbf{P} \leftrightarrow \mathbf{P_2}$, $\mathbf{P_2} \leftrightarrow \mathbf{P_4}$, and $\mathbf{P} \leftrightarrow \mathbf{P_2^*}$ processes, respectively, are stable for values of $k_{-2} \gtrsim 4 \times 10^4 s^{-1}$.

The overall exchange rate between monomer and tetramer, $k_{ex}^{P \leftrightarrow P_4}$, is approximately equal to k_{-2} at low $htt^{NT}Q_7$ concentrations

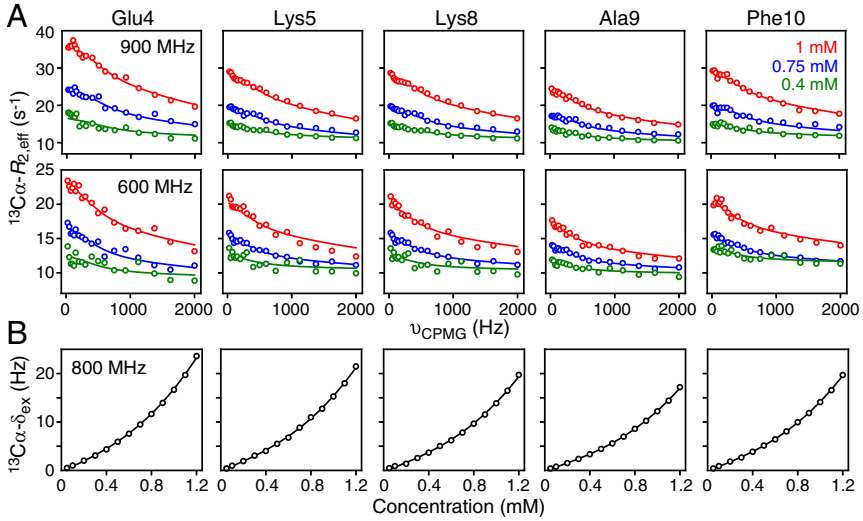

Fig. 3. Quantitative analysis of the concentration dependence of ^{13}Cα CPMG relaxation dispersion profiles and ^{13}Cα exchange-induced shifts for htt^{NT}Q$_7$. Representative (*A*) ^{13}Cα CPMG relaxation dispersion profiles at 900 MHz (*Top*) and 600 MHz (*Bottom*), and (*B*) ^{13}Cα exchange-induced chemical shifts (δ_{ex}) at 800 MHz. The experimental data, recorded at 5 °C, are shown as circles. The best-fit curves from a global fit to the kinetic model shown in Fig. 5 are represented by the continuous lines. The CPMG relaxation dispersion data were recorded on a ^{13}Cα/^{15}N-labeled sample, while the exchange-induced shift data were recorded on a uniformly ^{13}C/^{15}N-labeled sample.

(<0.3 mM). At higher peptide concentrations, $k_{ex}^{P\leftrightarrow P_4}$ can be shown to be well approximated by k_{-2}/n, where n varies from 1 to 4 for peptide concentrations ranging from 0.1 to 2.0 mM (and almost exactly 2 at the highest concentration of 1.2 mM used in the NMR experiments) for k_{-2} within the range of ∼40,000 to ∼60,000 s^{-1}. The counterintuitive decrease in $k_{ex}^{P\leftrightarrow P_4}$ at higher concentrations is a consequence of the fact that the overall dissociation process from tetramer to monomer, as well as the association process from monomer to tetramer, are concentration dependent.

The concentration dependence of the populations of $\mathbf{P_2}$, $\mathbf{P_4}$, and $\mathbf{P_2^*}$ simulated from the optimized values of the equilibrium dissociation constants is shown in Fig. 5*B*. The populations of $\mathbf{P_2}$ and $\mathbf{P_2^*}$ plateaus at ∼3.8% and 1.6%, respectively, at a total peptide concentration of 5 mM, and then decay very slowly as the concentration increases further. The concentration dependence of the population of $\mathbf{P_4}$, on the other hand, is characterized by an initial lag phase followed by a rapid increase; at 5 mM total peptide concentration, the population of $\mathbf{P_4}$ reaches ∼29%. At the highest concentration used in the NMR experiments, the populations of $\mathbf{P_4}$, $\mathbf{P_2}$, and $\mathbf{P_2^*}$ are ∼2.0%, 1.9%, and 0.9%, respectively (Fig. 5*B*, *Inset*).

In addition to kinetic information, the relaxation dispersion and exchange-induced shift data provide structural information in the form of ^{13}Cα (Fig. 7*B*) and ^{15}N (*SI Appendix*, Fig. S11) chemical shift differences (Δω) relative to the intrinsically disordered monomer. The ^{13}Cα-Δω values for residues 3–17 of $\mathbf{P_2}$/$\mathbf{P_4}$ range from +1 to +3 ppm (downfield shifts) with residues 3–12 and 14–15 having ^{13}Cα-Δω values in excess of +2 ppm, indicative of a contiguous helix. The latter is consistent with the negative values of Δω (upfield shifts) for ^{15}N, ^{13}Cβ, and ^{1}Hα nuclei of residues 3–17 (*SI Appendix*, Fig. S11), as well as the secondary structure assignment based on the backbone shifts using the program Talos (34). The ^{13}Cα-Δω values for $\mathbf{P_2^*}$ (off-path; Fig. 7*B*) are 50–100% smaller than the corresponding values for $\mathbf{P_2}$/$\mathbf{P_4}$ (on-path; Fig. 7*B*) but still represent downfield shifts, indicative of partial helix formation, for example due to an ensemble of partial helical coiled-coil conformations with different registers and hence different degrees of overlap.

Modulation of Dimerization and Tetramerization Equilibria. Since htt^{NT}Q$_7$ was expressed as a fusion protein with the Ig-binding domain of streptococcal protein G (GB1) separated by a 10-residue linker, we also investigated the dimerization and tetramerization properties of the GB1-htt^{NT}Q$_7$ fusion protein, as well as of two variants thereof with nitroxide (R1) spin labels covalently linked to engineered cysteines (S12C or S15C) in the htt^{NT}Q$_7$ portion of the fusion protein. The latter were generated by conjugation of the cysteines to *S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1*H*-pyrrol-3-yl) methyl methanesulfonothioate. GB1-htt^{NT}Q$_7$ and GB1-htt^{NT}Q$_7$ (S12C-R1) were characterized by sedimentation equilibrium AUC (*SI Appendix*, Figs. S12–S14), while GB1-htt^{NT}Q$_7$ (S15C-R1) was investigated by sedimentation velocity AUC (*SI Appendix*, Fig. S15). The results are summarized in Table 2.

The GB1-htt^{NT}Q$_7$ (*SI Appendix*, Figs. S12 and S13) and GB1-htt^{NT}Q$_7$ (S12C-R1) (*SI Appendix*, Fig. S14) sedimentation equilibrium AUC data can only be fit to a reversible monomer–dimer–tetramer model with significant occupancy of the tetramer. At the highest concentrations used in the analysis (which are higher than the loading concentrations), ∼0.8 mM for GB1-htt^{NT}Q$_7$ and ∼0.4 mM for GB1-htt^{NT}Q$_7$ (S12C-R1), the tetramer populations (in monomer units) reach values of ∼40% (*SI Appendix*, Fig. S13*B*) and ∼60% (*SI Appendix*, Fig. S14*B*), respectively. Interestingly, the dimer–tetramer equilibrium for these two constructs ($K_2^{diss} \sim 8-9$ μM) is similar to that for htt^{NT}Q$_7$ ($K_2^{diss} \sim 22$ μM). However, the values of K_1^{diss} are ∼15- and 200-fold lower for GB1-htt^{NT}Q$_7$ (∼9 mM) and GB1-htt^{NT}Q$_7$ (S12C-R1) (0.58 mM), respectively, relative to that for htt^{NT}Q$_7$ (∼0.12 M). Thus, the extent of tetramerization for these three constructs is predominantly determined by the stability of the dimer relative to the monomer. The increased stabilization of the dimer relative to the monomer as a consequence of the introduction of GB1 fused to the N terminus of htt^{NT}Q$_7$ is presumably due to transient, weak interactions between the GB1 globular domain of one subunit and the polyglutamine tail of the other.

In contrast, sedimentation velocity AUC data indicate that GB1-htt^{NT}Q$_7$ (S15C-R1) only undergoes dimerization with $K_1^{diss} \sim 0.22$mM (*SI Appendix*, Fig. S15), a factor of ∼3 lower than that of htt^{NT}Q$_7$ (S12C-R1). No evidence of any tetramer is apparent in the sedimentation velocity data.

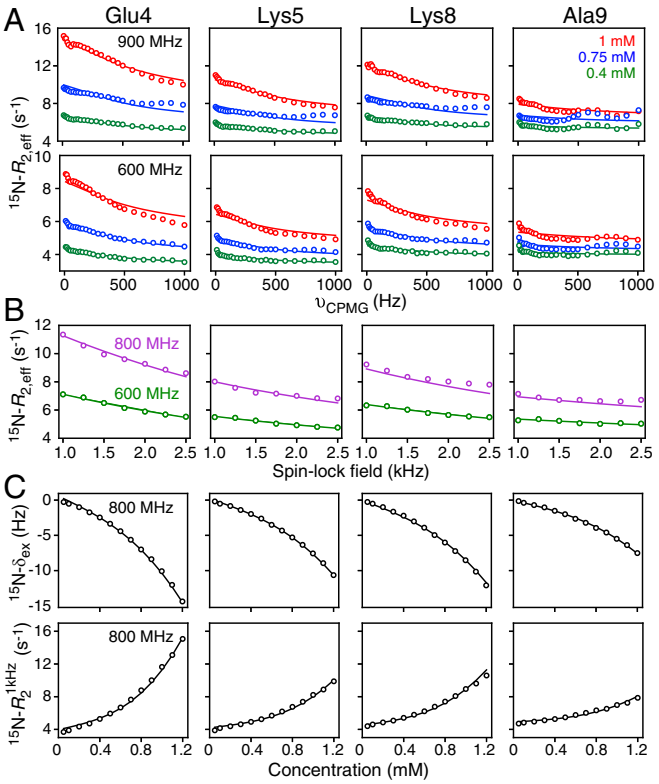

Fig. 4. Quantitative analysis of the concentration dependence of ^{15}N-relaxation dispersion profiles and exchange-induced shifts for htt^{NT}Q$_7$. Representative (*A*) ^{15}N-CPMG relaxation dispersion profiles at 900 MHz (*Top*) and 600 MHz (*Bottom*) at three concentrations (1, 0.75, and 0.4 mM), (*B*) ^{15}N-$R_{1\rho}$ relaxation dispersion profiles at 800 and 600 MHz at a concentration of 1 mM, and (*C*) ^{15}N-δ_{ex} exchange-induced chemical shifts (*Top*) and ^{15}N-R_2^{1kHz} (*Bottom*) at 800 MHz. The experimental data, recorded at 5 °C, are shown as circles. The best-fit curves from a global fit to the kinetic model shown in Fig. 5 are represented by the continuous lines. The relaxation dispersion data were recorded on a ^{13}Cα/^{15}N-labeled sample, while the exchange-induced shift and R_2^{1kHz} data were recorded on a uniformly ^{13}C/^{15}N-labeled sample.

Thus, apparently innocuous modifications, whether involving a fusion construct with a small globular protein at the N terminus, nitroxide spin-labeling within the htt^{NT} portion of the constructs, or even oxidation of Met7 (Fig. 2) can have a profound effect on both dimerization and tetramerization of htt^{NT}Q$_7$.

PRE-Based Structural Modeling of the Productive Dimer and Tetramer of htt^{NT}Q$_7$. To derive a structural model of the htt^{NT}Q$_7$ dimer and tetramer, we made use of intermolecular PRE measurements. When exchange is fast on the PRE relaxation time scale, intermolecular PREs observed on NMR visible ^{15}N-labeled htt^{NT}Q$_7$ in the presence of a small amount of a paramagnetically labeled htt^{NT}Q$_7$ derivative at natural isotopic abundance, yield structural information on the sparsely populated, NMR invisible, oligomeric states (25). Note that the PRE experiments make use of a ^{1}H–^{15}N correlation-based pulse scheme, and hence PREs arising from proximity to the paramagnetic label are only detected on the ^{1}H$_N$ protons of ^{15}N-labeled htt^{NT}Q$_7$ and not on the paramagnetically labeled derivative, which is at natural isotopic abundance.

Because of solubility and purification issues related to nitroxide-labeled htt^{NT}Q$_7$ constructs, we made use of the GB1-htt^{NT}Q$_7$ fusion protein labeled with a nitroxide (R1) at either S15C or S12C, both of which lie on the hydrophilic face of an amphiphilic helix (Fig. 7*A*). Experiments were carried out at a total peptide concentration of 0.6 mM with a molar ratio of 1:40 spin-labeled GB1-htt^{NT}Q$_7$ to ^{15}N-labeled htt^{NT}Q$_7$, and the resulting intermolecular PRE profiles are shown in Fig. 7*C*. The large excess of ^{15}N-labeled htt^{NT}Q$_7$ ensures that the transient dimers and tetramers either contain only a single nitroxide-labeled subunit, or at best two nitroxide-labeled subunits owing to dimer enrichment of nitroxide-labeled GB1-htt^{NT}Q$_7$ as a result of their low dimerization dissociation constants relative to htt^{NT}Q$_7$ (0.2–0.6 mM versus ∼0.1 M, respectively; Table 2). The latter will only yield PREs between dimers (and not within dimers), and hence the interdimer PRE profile will be the same whether one or two subunits within one of the dimers are nitroxide labeled.

Three control experiments were carried out. PREs were measured on a 1:1 (as well as 1:9) mixture of htt^{NT}Q$_7$ (S15C-R1) to ^{15}N-labeled htt^{NT}Q$_7$ (*SI Appendix*, Fig. S16*A*) and found to be highly correlated ($R = 0.98$) to those measured on a 1:40 mixture of GB1-htt^{NT}Q$_7$ (S15C-R1) to ^{15}N-labeled htt^{NT}Q$_7$ (*SI Appendix*, Fig. S16*B*), indicating that the oligomeric structures are unaffected by the presence of the GB1 solubility tag on the nitroxide-labeled subunit. PRE experiments were also recorded on samples of ^{15}N-labeled htt^{NT}Q$_7$ in the presence of nitroxide-labeled GB1 or (Met^{7}O)-htt^{NT}Q$_7$ (which does not oligomerize). No intermolecular PREs were detected in the latter two samples (*SI Appendix*, Fig. S16*A*), indicating that the PREs observed on ^{15}N-labeled htt^{NT}Q$_7$ in the presence of nitroxide-labeled constructs of GB1-htt^{NT}Q$_7$ are the result of specific association into well-defined multimers and do not arise from either nonspecific binding of the nitroxide label itself or to solvent PREs (25). In addition, the PRE data were collected at 10 °C, instead of 5 °C used for the other NMR experiments, to remove spectral overlap and permit PREs to be obtained for all residues within the helix. The PRE profiles, however, are the same at the two temperatures (*SI Appendix*, Fig. S17).

Table 1. Experimental data used in global fits

Experiment	Magnetic field, MHz	Residue number*	Concentrations, mM
^{13}Cα-CPMG relaxation dispersion	600, 900	4, 5, 8, 9, 10, 11, 12, 14, 16, 17	0.4, 0.75, 1.0
^{13}Cα-δ_{ex}	800	4, 5, 8, 9, 10, 11, 12, 14, 16, 17	0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2
^{15}N-CPMG relaxation dispersion	600, 900	4, 5, 8, 9, 12, 15, 17	0.4, 0.75, 1
^{15}N-$R_{1\rho}$ relaxation dispersion	600, 800	4, 5, 8, 9, 12, 15, 17	1.0
^{15}N-δ_{ex}	800	4, 5, 8, 9, 12, 15, 17	0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2
^{15}N-R_2^{1kHz}	800	4, 5, 8, 9, 12, 15, 17	0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2

*Residues not included in the global fit are those where there is either chemical shift overlap precluding accurate measurement of relaxation dispersion data or where the dispersions are small.

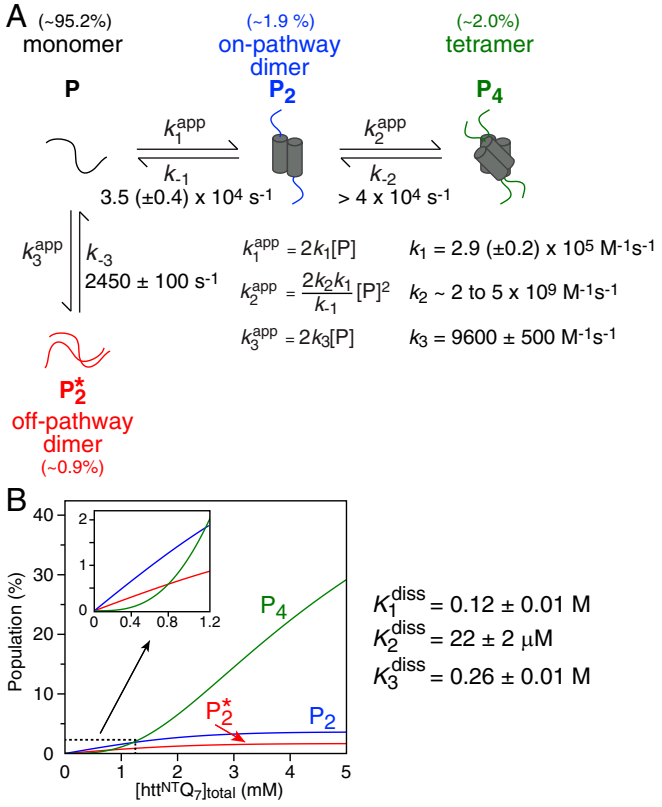

Fig. 5. Kinetic model used to fit the relaxation dispersion and exchange-induced shift data. (*A*) The kinetic model involves two pathways, one leading to the formation of a tetramer ($\mathbf{P_4}$) via a productive dimer ($\mathbf{P_2}$), and the other to a nonproductive dimer ($\mathbf{P_2^*}$). See *SI Appendix* for full details of the relevant kinetic equations (*SI Appendix*, Eqs. **S1–S5**). The expressions for the concentration dependence of the apparent pseudo–first-order association rate constants, as well as the optimized values of the association and dissociation rate constants, are shown. The population of the various species at $[htt^{NT}Q_7] = 1.2$ mM, the highest concentration used in the NMR experiments, is provided above each species. (*B*) Simulation of the species' populations (in monomer units) as a function of total peptide concentration using the rate constants determined from the global best-fits to the NMR data. The *Inset* shows an expanded view of the dependence over the concentration range of the experiments (50 μM to 1.2 mM).

A qualitative inspection of the intermolecular PREs arising from nitroxide labels at S12C-R1 and S15C-R1 suggests a mixture of antiparallel and parallel helical arrangements as both PRE profiles follow an oscillatory pattern with similar magnitude at both the N- and C-terminal ends of the helix extending from residue 3 to residue 17 (Fig. 7*C*). In addition, the magnitude of the PREs decays beyond residue 17 and is small for the C-terminal three glutamines, confirming the conclusions from $^{13}C\alpha$ chemical shifts (Fig. 7*B*) that the Q_7 sequence becomes progressively disordered.

To model the structure of the $htt^{NT}Q_7$ productive dimer ($\mathbf{P_2}$) and tetramer ($\mathbf{P_4}$), we made use of PRE-driven simulated annealing calculations using the program Xplor-NIH (35, 36). Full details of these calculations are provided in *SI Appendix*. Briefly, conjoined rigid body/torsion angle simulated annealing was employed with the backbone of residues 3–17 treated as a rigid idealized α-helix and the side chains given torsion degrees of freedom. In addition, strict symmetry was employed such that explicit degrees of freedom are only allowed for one subunit, the protomer, while the dimer-forming partner is computed by a 180° rotation about the z axis, and the full tetramer by rotating the dimer 180° along the axis of a line described by $z = 0$ and $x = y$. The dimer is C_2 symmetric, and the tetramer is a dimer of dimers related to one another by D_2 symmetry. The packing of the subunits is encoded in the absolute position in space of the protomer. Because the PRE restraints are limited, we assume that the structure of the productive dimer is exactly the same as that of the dimer within the tetramer. Of note, we also assume that the contribution of the unproductive dimer ($\mathbf{P_2^*}$; Fig. 5) to the observable PREs is negligible as its population is lower than those of the productive dimer and tetramer, and furthermore it is highly likely that $\mathbf{P_2^*}$ comprises an ensemble of states and therefore any potential PREs arising from this ensemble are absorbed into the PRE background.

There are two additional complications that were addressed in computing the dimer/tetramer structure from the PRE data (see *SI Appendix* for details). First, we do not know the populations of the transient heterodimer and heterotetramers giving rise to the intermolecular PREs, nor the correlation times of the species. For this reason, we used a target function that minimizes the correlation coefficient between observed and calculated PREs (*SI Appendix*, Eqs. **S7** and **S8**), rather than the sum squares of differences between observed and calculated PREs. Second, we do not know the relative population of dimer and tetramer. We therefore calculated a set of 11 structural ensembles ranging from 0% dimer/100% tetramer to 100% dimer/0% tetramer by appropriate weighting of the PREs, which takes into account the ratio of total correlation times expected for the dimer and tetramer. The exact ratio is not critical, and in the final calculations we assumed that the total correlation time (dominated by the rotational correlation time as the electron correlation time for a nitroxide is in excess of 100 ns) for the tetramer is 50% larger than that of the dimer. (Note that the heterotetramer is not double the molecular weight of the heterodimer.)

The results of the PRE-driven simulated annealing calculations are shown in Fig. 8 and *SI Appendix*, Fig. S18. A superposition of the lowest energy structures calculated with dimer to tetramer partitioning ranging from 90%:10% to 0%:100% is shown in Fig. 8*A*. The Cα atomic rms difference to the mean coordinate positions is 0.6 Å, and the overall correlation coefficient between observed and calculated PREs is 0.95 (Fig. 8*B*). Within each ensemble over this dimer-to-tetramer partitioning range, the Cα atomic rms precision for the 10 lowest energy structures ranges from ~0.5 to ~1 Å, and the average Cα rms

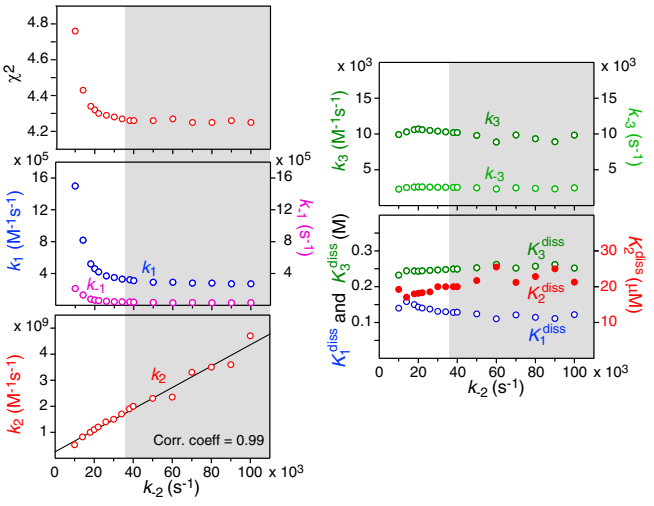

Fig. 6. Global fitting results to all of the NMR data on $htt^{NT}Q_7$ using a grid search on the value of the rate constant k_{-2} for the dissociation of tetramers ($\mathbf{P_4}$) into dimers ($\mathbf{P_2}$). The panels show the dependence of χ^2 and the optimized values of the rate constants and equilibrium dissociation constants as a function of fixed values of k_{-2}. The gray shaded region represents the range of k_{-2} where χ^2 is at a plateau minimum.

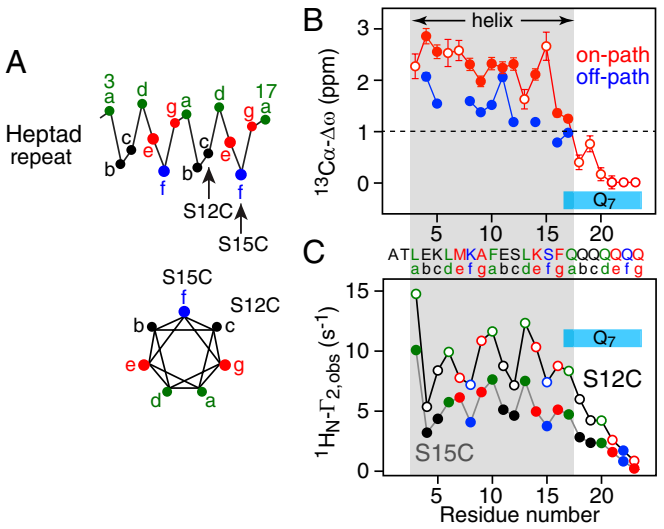

Fig. 7. ^{13}Cα-Δω profiles and intermolecular $^{1}H_N$-PRE data for htt$^{NT}Q_7$. (*A*) Heptad repeat parallel (*Left*) and orthogonal (*Right*) to the helix axis with positions *a* and *d* in green, *e* and *g* in red, *b* and *c* in black, and *f* in blue. (*B*) Calculated ^{13}Cα-Δω for the productive dimer/tetramer (red) and the nonproductive dimer (blue) obtained from the global fits to the relaxation dispersion and exchange-induced shift data. The filled-in circles indicate residues used in the global fits; the ^{13}Cα-Δω values calculated a posteriori from the exchange-induced shifts for residues not included in the global fits (either because chemical shift overlap precluded accurate measurement of dispersions or because the dispersions were very small) are represented by open circles (*SI Appendix*). (*C*) Intermolecular PRE $^{1}H_N$-Γ_2 profiles observed at 10 °C on ^{15}N-labeled htt$^{NT}Q_7$ in the presence of nitroxide spin-labeled S15C-R1 (filled-in circles) and S12C-R1 (open circles) GB1-htt$^{NT}Q_7$ at a molar ratio of 1:40 nitroxide spin-labeled to ^{15}N-labeled peptide. The color coding of the residues follows that of the heptad repeat shown in *A*. The total peptide concentration was 0.6 mM.

difference of the 100 lowest energy structures is ∼0.9 Å. Thus, the structural model is robust over a wide range of dimer to tetramer partitioning. The structures calculated with 100% dimer converge poorly and fail to satisfy the intermolecular PRE data (Fig. 8*C* and *SI Appendix*, Fig. S18). The overall PRE correlation coefficient and PRE Q-factor between observed and calculated PREs remains stable over dimer:tetramer ratios of 70%:30% to 0%:100%, but at 80% dimer and above, a clear decrease in correlation coefficient and Q-factor becomes apparent (Fig. 8*C* and *SI Appendix*, Fig. S18).

The structure of the tetramer is a dimer of dimers (red/light red and blue/light blue in Fig. 8*A*) related by D_2 symmetry, reminiscent of the helical components of the p53 tetramerization domain (37). The interhelical angle between the two helices of the dimer ranges from 145 to 155°, and the two interhelical angles between the dimers are ∼90 and 100°. (For the structures calculated with 100% dimer, which do not fit the PRE data, the angle between the helices of the dimer is increased to ∼170°.)

Side-chain packing between the helices of the dimer and between the dimer of dimers is shown in Fig. 9, respectively. Packing between helices of the dimer is hydrophobic and involves the side chains of Leu6, Ala9, Phe10, Leu13, and Phe16. The involvement of Phe16 at the dimer interface provides a rational basis for understanding why htt^{NT} alone undergoes no detectable self-association (Fig. 2): for the latter to occur, Phe16 must be part of the helix, which presumably requires a few additional C-terminal residues for stabilization. In this regard, we note that not only does the helix extend out to residue 17, but it is possible from both the ^{13}Cα (Fig. 7*B*) and ^{15}N (*SI Appendix*, Fig. S11) chemical shifts of the dimer/tetramer, as well as the PRE data (Fig. 7*C*), that residues 18, 19, and possibly even 20 may adopt partial helical character. The surface exposed side chains are hydrophilic (Glu4, Lys8, Glu11, Ser12, and Ser15), characteristic of an amphiphilic helix. The interface between the two dimers is predominantly hydrophobic (Leu3, Leu6, Met7, and Phe10), supplemented by a potential electrostatic interaction between Glu11 and Lys14. The presence of Met7 at the interface of the two dimers is also of interest since oxidation of the side chain Met7 to a sulfoxide abrogates self-association (Fig. 2*C*), presumably by inhibiting tetramer formation through the introduction of a bulky hydrophilic group within the tetramer core, potentially leading to steric clash between interacting methionine residues (Fig. 9*B*).

Using EPR to Independently Validate the Structure of the Dimer Unit Within the htt$^{NT}Q_7$ Tetramer. As noted above, sedimentation velocity AUC indicates that GB1-htt$^{NT}Q_7$ (S15C-R1) only forms a dimer (Table 2 and *SI Appendix*, Fig. S15). To validate the intersubunit/intradimer distance between S15C-R1 labels in the PRE-derived structures, we used Q-band pulsed DEER EPR (26) at 50 K (Fig. 10). The raw and background-corrected DEER echo curves are shown in Fig. 10*A*. Inversion modulation DEER (IM-DEER) experiments (*SI Appendix*, Fig. S19) in which the normalized modulation depth, Δ/Δ_{max}, is measured as a function of the ELDOR pulse flip angle (38) confirms that GB1-htt$^{NT}Q_7$ (S15C-R1) is dimeric with no evidence for the presence of any trimer or tetramer (Fig. 10*B*). The equilibrium dissociation constant for the dimer, K_{dimer}, obtained by best fitting the modulation depth as a function of concentration (39) is ∼0.5 mM (Fig. 10*C*), only a factor of ∼2 higher than that obtained by AUC (Table 2), which is unsurprising since the sample is flash frozen and requires the presence of glycerol to prevent clustering of molecules during the freezing process. As an additional control, oxidation of Met7 to a sulfoxide (by mild treatment with H_2O_2) hugely reduces the DEER modulation depth (*SI Appendix*, Fig. S20), indicating that GB1-(Met^{7}O)-htt$^{NT}Q_7$ (S15C-R1) does not undergo self-association, fully consistent with the solution NMR data on (Met^{7}O)-htt$^{NT}Q_7$ (Fig. 2*B*).

Table 2. Monomer/dimer and dimer/tetramer equilibrium dissociation constants for various constructs of htt$^{NT}Q_7$

		Equilibrium dissociation constants	
		Monomer/dimer	Dimer/tetramer
Sample	Method	K_1^{diss}	K_2^{diss}
htt$^{NT}Q_7$	NMR*	0.12 ± 0.01 M	22 ± 2 μM
GB1-htt$^{NT}Q_7$	AUC†	9 ± 0.3 mM	8^{+14}_{-2} μM
GB1-htt$^{NT}Q_7$ (S15C-R1)	AUC‡	0.22 ± 0.01 mM	ND§
	EPR¶	0.49 ± 0.05 mM	ND§
GB1-htt$^{NT}Q_7$ (S12C-R1)	AUC†	0.58 ± 0.13 mM	9 ± 3 μM
	EPR¶	0.41 ± 0.04 mM	14^{+6}_{-5} μM

*From global fit to ^{15}N and ^{13}Cα relaxation dispersion experiments, ^{15}N and ^{13}Cα exchange-induced shifts and R_2^{1kHz} data over a range of concentrations (Figs. 3–5 and *SI Appendix*, Figs. S4–S6). NMR experiments were conducted at 5 °C in 20 mM phosphate (pH 6.5), 50 mM NaCl, and 10% D_2O (vol/vol). The equilibrium dissociation constant for the off-pathway monomer/dimer equilibrium (K_3^{diss}) is 0.26 ± 0.01 M (Fig. 5).

†Sedimentation equilibrium experiments were conducted at 20 °C in 50 mM sodium phosphate (pH 6.5) with 150 mM NaCl for GB1-htt$^{NT}Q_7$ and 100 mM NaCl for GB1-htt$^{NT}Q_7$ (*SI Appendix*, Fig. S12*C*).

‡Sedimentation velocity experiments were conducted at 20 °C in 50 mM sodium phosphate (pH 6.5) and 100 mM NaCl.

§ND, not detectable. No tetrameric species was detectable for GB1-htt$^{NT}Q_7$ (S15C-R1) either by sedimentation velocity AUC or EPR (Fig. 10*B*).

¶The DEER modulation depth as a function of concentration was measured by Q-band EPR at 50 K (Fig. 10*C*). Samples at room temperature were flash frozen in liquid nitrogen. The buffer was the same as in the corresponding AUC experiments with the addition of 30% (vol/vol) glycerol.

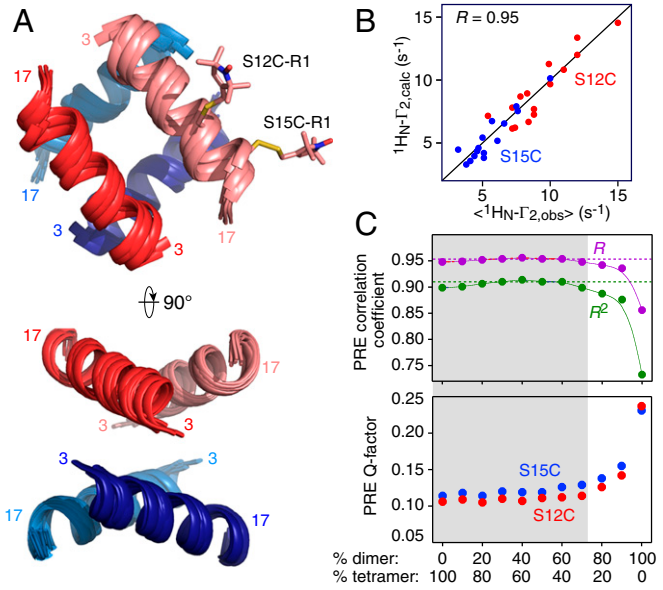

Fig. 8. PRE-based structure of the htt^{NT}Q$_7$ tetramer. (*A*) Two orthogonal views showing ribbon diagrams of a superposition of the lowest energy, PRE-based structures from each ensemble calculated with dimer to tetramer partitioning ranging from 0%:100% to 90%:10%. The tetramer is a dimer of dimers, and the structure of the dimer (one shown in red and pink, and the other in blue and light blue) is assumed to be the same in the dimer and tetramer. The helix that extends from residues 3–17 is treated as a rigid body. (*B*) Correlation between observed and calculated PREs. The latter are the average of the calculated values of the lowest energy structures from each ensemble computed with dimer:tetramer populations ranging from 0%:100% to 90%:10%. The Pearson correlation coefficient *R* has a value of 0.95. Equivalent plots for the individual ensembles calculated at the different dimer-to-tetramer ratios are shown in *SI Appendix*, Fig. S18. (*C*) Plot of PRE correlation coefficient (*Top*, with *R* and R^2 shown in purple and green, respectively) and PRE Q factor (*Bottom*, with the values for S12C-R1 and S15C-R1 shown in red and blue, respectively) as a function of dimer-to-tetramer partitioning (1 SD error bars lie within the circles). The PRE Q factor is given by $\left\{\sum_i [\Gamma_2^{obs}(i)-\Gamma_2^{calc}(i)]^2/\sum_i \Gamma_2^{obs}(i)^2\right\}^{1/2}$ (46). The gray shaded area indicates the likely range of partitioning of dimer-to-tetramer populations. The structures calculated with 100% dimer do not satisfy the experimental PRE data; the structures calculated with 80% and 90% dimer exhibit some degradation in agreement with the PRE data (as measured both by the correlation coefficient and the Q factor), although they are structurally very similar to the structures calculated with lower dimer populations.

Deconvolution of the DEER echo curves for GB1-htt^{NT}Q$_7$ (S15C-R1) obtained over a range of concentrations from 60 μM to 2.2 mM (*SI Appendix*, Fig. S21) using validated Tikhonov regularization (40) yields a *P*(*r*) distance distribution with a mean value of 26.5 Å, which compares very well with the average *P*(*r*) distribution calculated from the 10 lowest energy structures with dimer to tetramer partitioning ranging from 90%:10% to 0%:100% (Fig. 10*D*). When the mean intersubunit/intradimer S15C-R1/S15C-R1 distance is calculated for the 10 lowest energy structures within each ensemble obtained at the different dimer-to-tetramer ratios, it can be seen that the calculated mean distance ranges from 26.8 Å for the structures calculated with 100% and 80% tetramer, to 27.8 Å for those calculated with 70–90% dimer, in excellent agreement with the DEER results (Fig. 10*E* and *SI Appendix*, Fig. S22). For the ensemble calculated with 100% dimer, however, agreement is poor and the mean distance is considerably longer (30 Å), as expected since the helices in the 100% dimer structure ensemble are close to fully antiparallel.

GB1-htt^{NT}Q$_7$ (S12C-R1) is largely tetrameric as evidenced both by AUC (*SI Appendix*, Fig. S14) and from best fitting the DEER modulation depth as a function of concentration (Table 2). Unfortunately, DEER cannot be used to determine distances between S12C-R1 labels within the tetramer as the predicted intersubunit/intradimer distance is less than 15 Å, and hence the spins are in the strong dipolar coupling limit, beyond the range of applicability of DEER (26). Furthermore, the presence of a strongly coupled spin pair effectively precludes measurement of the longer interdimer distances in a multispin system. However, the short intersubunit/intradimer distance between S12C-R1 labels can be probed by CW EPR at room temperature by measuring line broadening arising from strong dipolar interactions between spins separated within a range of 8–20 Å (27). Comparison of the X-band EPR spectrum of monomeric GB1-htt^{NT}Q$_7$ (S12C-R1) at low concentration (10 μM) with largely tetrameric GB1-htt^{NT}Q$_7$ (S12C-R1) at high concentration (3 mM) reveals considerable line broadening corresponding to a mean distance of 13 Å (Fig. 11*A*) with a *P*(*r*) distribution [derived by quantitative analysis using the program ShortDistances (27)] that is in excellent agreement with that calculated from the PRE-based structures (Fig. 11*B*).

Concluding Remarks

The N-terminal htt^{NT} sequence plays a critical role in facilitating aggregation and fibril formation of poly(Q) tracts within the N-terminal domain of huntingtin, encoded by exon 1 of the huntingtin gene (41). While previous investigations established that the htt^{NT} sequence has helical propensity (18, 42–44), the initial events involved in htt^{NT} multimerization that eventually lead to polyglutamine fibril nucleation and the structures of the multimers were unknown. Using a minimalistic construct, htt^{NT}Q$_7$, that remains predominantly monomeric over a sufficiently long period of time to enable detailed quantitative NMR measurements, but still eventually forms fibril, we were able to resolve at atomic resolution the early transient events in multimerization involving sparsely populated (<2%) oligomeric states, based on global analysis of the concentration dependence of an array of relaxation-based NMR measurements, including CPMG and $R_{1\rho}$ relaxation dispersion, transverse relaxation in the rotating frame, and exchange-induced chemical shifts.

The NMR data on htt^{NT}Q$_7$ reveal the existence of a branching pathway, one path leading to a productive dimer followed by a tetramer on a timescale of 20–30 μs, and the other to a non-productive dimer on an order of magnitude slower timescale (Fig. 5). Both dimers are very unstable with monomer–dimer equilibrium dissociation constants of 0.1–0.2 M, in contrast to

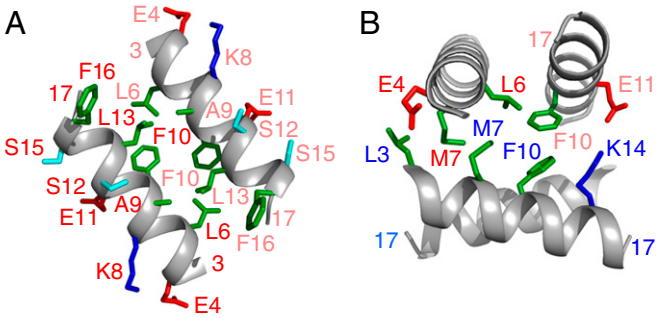

Fig. 9. Side-chain interactions within and between dimer units of the htt^{NT}Q$_7$ tetramer. (*A*) Contacts between subunits of the dimer (labeling of side chain for one of the subunits is in dark red, and for the other in light red). (*B*) Contacts within the tetramer between the two helices of one dimer and one helix of the other dimer. Labeling within the two dimer units of the tetramer is in dark red/light red and dark blue/light blue. Residue color coding is as follows: hydrophobic, green; positively charged, blue; negatively charged, red; hydrophilic, cyan.

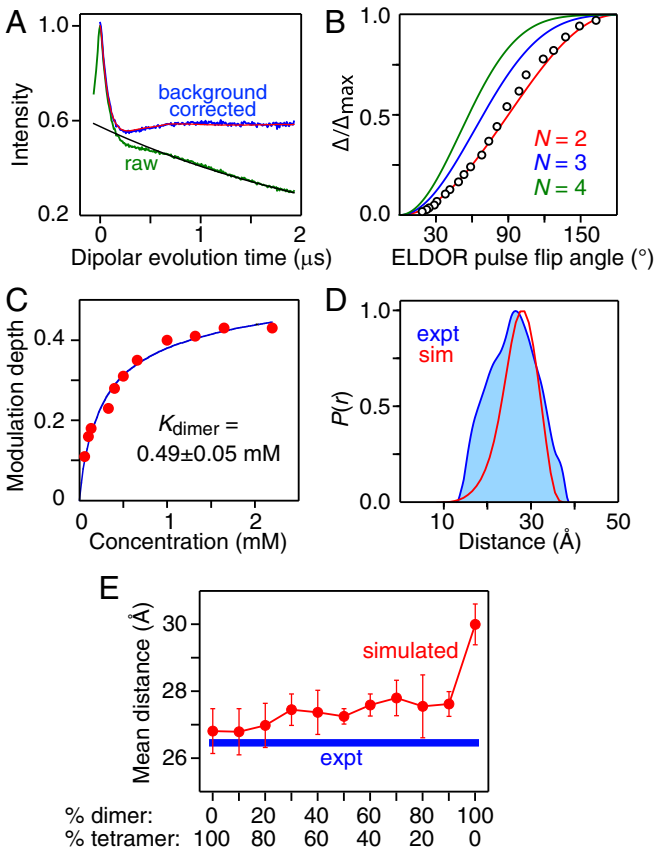

Fig. 10. Validation of the dimeric unit of the PRE-based structures of the htt^{NT}Q$_7$ dimer by Q-band DEER EPR measurements on dimeric GB1-htt^{NT}Q$_7$ (S15C-R1). (*A*) Q-band DEER echo curves (raw data in green; background-corrected curve in blue) recorded at 50 K on 2.2 mM GB1-htt^{NT}Q$_7$ (S15C-R1) in 30% (vol/vol) glycerol. The black line represents the best-fit background represented by an exponential decay. The best-fit echo curve obtained by Tikhonov regularization (40) to generate the distance probability distribution, *P*(*r*), is shown in red, superimposed on the background-corrected experimental curve. (*B*) IM-DEER experiment showing the experimental (circles) normalized modulation depth $\Delta(\theta)/\Delta_{max}$ as a function of ELDOR flip angle θ (where Δ_{max} is the maximum modulation depth obtained at θ = 180°) together with the theoretical dependence for a dimer (red), trimer (blue) and tetramer (green). (*C*) Modulation depth as a function of concentration with the experimental data shown as red circles and the best-fit curve for a monomer–dimer association in blue. (*D*) Comparison of experimental *P*(*r*) distribution between S15C-R1 labels in the GB1-htt^{NT}Q$_7$ dimer obtained by validated Tikhonov regularization (blue) with the average *P*(*r*) distribution predicted from the PRE-based structures with dimer-to-tetramer partitioning ranging from 0%:100% to 90%:10% (red). (*E*) Comparison of the mean distance between spin labels calculated for each ensemble of PRE-based structures (red circles) with the experimental mean distance obtained from DEER (blue thick line).

the tetramer which is three to four orders of magnitude more stable with a dimer–tetramer equilibrium dissociation constant of ~22 μM. The backbone chemical shifts derived for the minor species indicate that residues 3–17 of the htt^{NT} sequence form a contiguous helix in the productive dimer/tetramer with increasing disorder in the polyglutamine tail toward the C terminus (Fig. 7 *A* and *B*); the nonproductive dimer displays only partial helical character within the htt^{NT} sequence (Fig. 7*B*), presumably constituting an ensemble of states with multiple registers of the two subunits. Furthermore, we were able to characterize the structure of the productive dimer/tetramer using intermolecular PRE-measurements (Fig. 7*C*): the dimer is an antiparallel coiled coil, which assembles into a D_2 symmetric tetramer constituted by a dimer of dimers oriented approximately orthogonal to one another (Figs. 8*A* and 9*B*). Packing between the subunits at the dimer (Fig. 9*A*) and dimer of dimers (Fig. 9*B*) interfaces is largely hydrophobic, and the structure of the tetramer is reminiscent of the arrangement of helices in the p53 tetramerization domain (37). Finally, the configuration of the dimer unit was independently validated from two interspin label distances measured by DEER (Fig. 9) and CW (Fig. 10) EPR.

The structure of the htt^{NT}Q$_7$ tetramer explains the dependence of efficient fibril nucleation on poly(Q) length. The formation of a tetramer increases the local concentration of the poly(Q) tract, but for nucleation to occur polyglutamine chains from the different subunits must be able to overlap efficiently and fairly extensively. The distances between the C termini of the helices is ~24 Å within the dimer, and ~20 and ~27 Å between dimers in the tetramer (corresponding to dark red/light red, dark red/light blue, and dark red/dark blue subunits, respectively, in Fig. 8*A*). The end-to-end distance of a disordered 10-residue poly(Q) sequence is predicted to lie between 20 and 25 Å (45), and the longer the poly(Q) tract, the higher the probability of significant interchain glutamine overlap. At pathological poly(Q) lengths (≥36), one can anticipate that nucleation through interchain glutamine contacts could be made between any pair of subunits, as opposed to the two subunit pairs with the shortest separation between htt^{NT} C-termini (that is the dark red/light blue and light red/dark blue subunit pairings in Fig. 8*A*).

Once fibrils are formed, it is uncertain whether stable structures of the htt^{NT} dimer or tetramer persist as solid-state NMR studies have shown that the static β-turn/β-strand fibril core is surrounded by dynamic, molten globule-like htt^{NT} helices (16). Thus, the initial transient oligomerization events described here

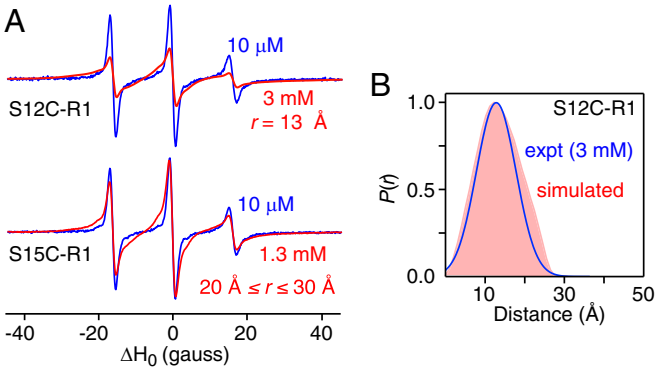

Fig. 11. Validation of the dimeric unit of the PRE-based structures of the htt^{NT}Q$_7$ dimer by X-band CW EPR measurements on dimeric GB1-htt^{NT}Q$_7$ (S12C-R1) in solution. (*A*) Comparison of CW X-band EPR spectra of GB1-htt^{NT}Q$_7$ (S12C-R1) (*Top*) and GB1-htt^{NT}Q$_7$ (S15C-R1) (*Bottom*) at low (blue) and high (red) concentrations. The derivative EPR spectra are normalized to the double integral. At 10 μM, both samples are monomeric; at 3 mM, GB1-htt^{NT}Q$_7$ (S12C-R1) is largely tetrameric (*SI Appendix*, Fig. S14); at 1.2 mM, GB1-htt^{NT}Q$_7$ (S15C-R1) is largely dimeric (Fig. 9 and *SI Appendix*, Fig. S15). The broadening observed at the high concentration of GB1-htt^{NT}Q$_7$ (S12C-R1) is due to strong dipolar coupling between closely spaced S12C-R1 labels within the dimeric unit of the tetramer and corresponds to a distance of 13 Å. (Note the interdimer distances between S12C-R1 labels in the PRE-based structure of the tetramer are too large to cause any significant line broadening of the CW EPR spectrum.) As a control, the high concentration GB1-htt^{NT}Q$_7$ (S15C-R1) sample exhibits only minimal line broadening corresponding to a distance between 20 and 30 Å, fully consistent with the results from DEER EPR (Fig. 10). (*B*) Comparison of the experimental *P*(*r*) distance distribution between S12C-R1 labels (blue) derived from the CW EPR data using the program ShortDistances (27) with the corresponding intradimer average *P*(*r*) distribution calculated from the PRE-based structures with dimer-to-tetramer partitioning ranging from 0%:100% to 90%:10% (light red).

that culminate in the formation of a tetramer of the htt^{NT} sequence constitute a prenucleation trigger or molecular switch that hugely increases the probability of occurrence of intermolecular poly(Q) contacts and hence poly(Q) fibril nucleation. One can speculate that CRISPR-directed mutation within the htt^{NT} sequence designed to prevent productive dimer and tetramer formation may provide a fruitful avenue for preventing or delaying the onset of Huntington's disease.

Experimental Procedures

Full details relating to expression and purification of isotopically labeled samples, nitroxide spin labeling for NMR and EPR studies; NMR, EPR, AUC, and atomic force microscopy measurements; theory and global data fitting of relaxation dispersion and exchange-induced shift data; and PRE-based structure modeling are provided in *SI Appendix*.

Atomic coordinates of the tetramer, as well as experimental restraints, have been deposited in the Protein Data Bank (ID code 6N8C) (47), and backbone assignments for the tetramer have been deposited in the Biological Magnetic Resonance Data Bank (ID code 30545) (48).

ACKNOWLEDGMENTS. We thank Nick Anthis for the design of the huntingtin exon-1 DNA constructs used in this work; Dusty Baber, Dan Garrett, and Jinfa Ying for technical assistance; and Attila Szabo for useful discussions. S.A.K. was supported by a Postdoctoral Research Associate Training Program Fellowship of the National Institute of General Medical Sciences (Fi2GM117609-01). This work was supported by the Intramural Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health (G.M.C.).

1. Orr HT, Zoghbi HY (2007) Trinucleotide repeat disorders. *Annu Rev Neurosci* 30: 575–621.
2. Zuccato C, Valenza M, Cattaneo E (2010) Molecular mechanisms and potential therapeutical targets in Huntington's disease. *Physiol Rev* 90:905–981.
3. Bates GP, et al. (2015) Huntington disease. *Nat Rev Dis Primers* 1:15005.
4. Landles C, et al. (2010) Proteolysis of mutant huntingtin produces an exon 1 fragment that accumulates as an aggregated protein in neuronal nuclei in Huntington disease. *J Biol Chem* 285:8808–8823.
5. Sathasivam K, et al. (2013) Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. *Proc Natl Acad Sci USA* 110:2366–2370.
6. Saudou F, Humbert S (2016) The biology of Huntingtin. *Neuron* 89:910–926.
7. Kar K, Jayaraman M, Sahoo B, Kodali R, Wetzel R (2011) Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent. *Nat Struct Mol Biol* 18:328–336.
8. Ramdzan YM, et al. (2017) Huntingtin inclusions trigger cellular quiescence, deactivate apoptosis, and lead to delayed necrosis. *Cell Rep* 19:919–927.
9. Crick SL, Ruff KM, Garai K, Frieden C, Pappu RV (2013) Unmasking the roles of N- and C-terminal flanking sequences from exon 1 of huntingtin as modulators of polyglutamine aggregation. *Proc Natl Acad Sci USA* 110:20075–20080.
10. Shen K, et al. (2016) Control of the structural landscape and neuronal proteotoxicity of mutant Huntingtin by domains flanking the polyQ tract. *eLife* 5:e18065.
11. Schneider R, et al. (2011) Structural characterization of polyglutamine fibrils by solid-state NMR spectroscopy. *J Mol Biol* 412:121–136.
12. Isas JM, Langen R, Siemer AB (2015) Solid-state nuclear magnetic resonance on the static and dynamic domains of Huntingtin exon-1 fibrils. *Biochemistry* 54:3942–3949.
13. Hoop CL, et al. (2016) Huntingtin exon 1 fibrils feature an interdigitated β-hairpin-based polyglutamine core. *Proc Natl Acad Sci USA* 113:1546–1551.
14. Sivanandam VN, et al. (2011) The aggregation-enhancing huntingtin N-terminus is helical in amyloid fibrils. *J Am Chem Soc* 133:4558–4566.
15. Hoop CL, 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.
16. Lin H-K, et al. (2017) Fibril polymorphism affects immobilized non-amyloid flanking domains of huntingtin exon1 rather than its polyglutamine core. *Nat Commun* 8:15462.
17. Fiumara F, Fioriti L, Kandel ER, Hendrickson WA (2010) Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell* 143:1121–1135.
18. Wetzel R (2012) Physical chemistry of polyglutamine: Intriguing tales of a monotonous sequence. *J Mol Biol* 421:466–490.
19. Jayaraman M, et al. (2012) Slow amyloid nucleation via α-helix-rich oligomeric intermediates in short polyglutamine-containing huntingtin fragments. *J Mol Biol* 415: 881–899.
20. Palmer AG, 3rd (2004) NMR characterization of the dynamics of biomacromolecules. *Chem Rev* 104:3623–3640.
21. Mittermaier A, Kay LE (2006) New tools provide new insights in NMR studies of protein dynamics. *Science* 312:224–228.
22. Palmer AG, 3rd, Massi F (2006) Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem Rev* 106:1700–1719.
23. Vallurupalli P, Bouvignies G, Kay LE (2011) Increasing the exchange time-scale that can be probed by CPMG relaxation dispersion NMR. *J Phys Chem B* 115:14891–14900.
24. Iwahara J, Tang C, Marius Clore G (2007) Practical aspects of ^{1}H transverse paramagnetic relaxation enhancement measurements on macromolecules. *J Magn Reson* 184:185–195.
25. Clore GM, Iwahara J (2009) Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem Rev* 109:4108–4139.
26. Jeschke G (2012) DEER distance measurements on proteins. *Annu Rev Phys Chem* 63: 419–446.
27. Altenbach C, López CJ, Hideg K, Hubbell WL (2015) Exploring structure, dynamics, and topology of nitroxide spin-labeled proteins using continuous-wave electron paramagnetic resonance spectroscopy. *Methods Enzymol* 564:59–100.
28. Aiken CT, et al. (2009) Phosphorylation of threonine 3: Implications for Huntingtin aggregation and neurotoxicity. *J Biol Chem* 284:29427–29436.
29. Anthis NJ, Clore GM (2015) Visualizing transient dark states by NMR spectroscopy. *Q Rev Biophys* 48:35–116.
30. Fawzi NL, Ying J, Ghirlando R, Torchia DA, Clore GM (2011) Atomic-resolution dynamics on the surface of amyloid-β protofibrils probed by solution NMR. *Nature* 480: 268–272.
31. McConnell HM (1958) Reaction rates by nuclear magnetic resonance. *J Chem Phys* 28: 430–431.
32. Rennella E, Sekhar A, Kay LE (2017) Self-assembly of human Profilin-1 detected by Carr–Purcell–Meiboom–Gill nuclear magnetic resonance (CPMG NMR) spectroscopy. *Biochemistry* 56:692–703.
33. Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44: 213–223.
34. Shen Y, Bax A (2015) Protein structural information derived from NMR chemical shift with the neural network program TALOS-N. *Methods Mol Biol* 1260:17–32.
35. Schwieters CD, Kuszewski JJ, Clore GM (2006) Using Xplor-NIH for NMR molecular structure determination. *Prog Nucl Magn Reson Spectrosc* 48:47–62.
36. Schwieters CD, Bermejo GA, Clore GM (2018) Xplor-NIH for molecular structure determination from NMR and other data sources. *Protein Sci* 27:26–40.
37. Clore GM, et al. (1995) Refined solution structure of the oligomerization domain of the tumour suppressor p53. *Nat Struct Biol* 2:321–333.
38. Schmidt T, Ghirlando R, Baber J, Clore GM (2016) Quantitative resolution of monomer-dimer populations by inversion modulated DEER EPR spectroscopy. *ChemPhysChem* 17: 2987–2991.
39. Ceccon A, et al. (2018) Interaction of Huntingtin exon-1 peptides withl lipid-based micellar nanoparticles probed by solution NMR and Q-band pulsed EPR. *J Am Chem Soc* 140:6199–6202.
40. Jeschke G, et al. (2006) DeerAnalysis2006–A comprehensive software package for analyzing pulsed ELDOR data. *Appl Magn Reson* 30:473–498.
41. Arndt JR, Chaibva M, Legleiter J (2015) The emerging role of the first 17 amino acids of huntingtin in Huntington's disease. *Biomol Concepts* 6:33–46.
42. Baias M, et al. (2017) Structure and dynamics of the Huntingtin exon-1 N-terminus: A solution NMR perspective. *J Am Chem Soc* 139:1168–1176.
43. Pandey NK, 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.
44. Bravo-Arredondo JM, et al. (2018) The folding equilibrium of huntingtin exon 1 monomer depends on its polyglutamine tract. *J Biol Chem* 293:19613–19623.
45. Cantor CR, Schimmel PR (1980) *Biophysical Chemistry Part III: The Behavior of Biological Macromolcules* (Freeman, San Francisco).
46. Iwahara J, Schwieters CD, Clore GM (2004) Ensemble approach for NMR structure refinement against ^{1}H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J Am Chem Soc* 126: 5879–5896.
47. Kotler SA, et al. (2018) Structure of the Huttingtin tetramer/dimer mixture determined by paramagnetic NMR. Protein Data Bank. Available at https://www.rcsb.org/structure/6N8C. Deposited November 30, 2018.
48. Kotler SA, et al. (2018) Structure of the Huttingtin tetramer/dimer mixture determined by paramagnetic NMR. Biological Magnetic Resonance Data Bank. Available at http://www.bmrb.wisc.edu/data_library/summary/index.php?bmrbId=30545. Deposited December 6, 2018.