

Structural Model of the Proline-Rich Domain of Huntingtin Exon-1 Fibrils

Alexander S. Falk,¹ José M. Bravo-Arredondo,¹ Jobin Varkey,¹ Sayuri Pacheco,¹ Ralf Langen,¹ and Ansgar B. Siemer^{1,*}

¹Department of Physiology & Neuroscience, Zilkha Neurogenetic Institute, Keck School of Medicine of USC, Los Angeles, California

ABSTRACT Huntington's disease is a heritable neurodegenerative disease that is caused by a CAG expansion in the first exon of the huntingtin gene. This expansion results in an elongated polyglutamine domain that increases the propensity of huntingtin exon-1 to form cross- β fibrils. Although the polyglutamine domain is important for fibril formation, the dynamic, C-terminal proline-rich domain (PRD) of huntingtin exon-1 makes up a large fraction of the fibril surface. Because potential fibril toxicity has to be mediated by interactions of the fibril surface with its cellular environment, we wanted to model the conformational space adopted by the PRD. We ran 800-ns long molecular dynamics simulations of the PRD using an explicit water model optimized for intrinsically disordered proteins. These simulations accurately predicted our previous solid-state NMR data and newly acquired electron paramagnetic resonance double electron-electron resonance distances, lending confidence in their accuracy. The simulations show that the PRD generally forms an imperfect polyproline (polyP) II helical conformation. The two polyP regions within the PRD stay in a polyP II helix for most of the simulation, whereas occasional kinks in the proline-rich linker region cause an overall bend in the PRD structure. The dihedral angles of the glycine at the end of the second polyP region are very variable, effectively decoupling the highly dynamic 12 C-terminal residues from the rest of the PRD.

SIGNIFICANCE Huntington's disease is caused by a polyglutamine expansion in the exon-1 of huntingtin, which results in the formation of fibrillar huntingtin aggregates. Although the polyglutamine domain is the site of the disease-causing mutation, the proline-rich domain of huntingtin exon-1 (HTTex1) is important for fibril toxicity and contains many epitopes of fibril-specific HTTex1 antibodies. Here, we present a structural and dynamic model of the highly dynamic proline-rich domain using a combination of electron paramagnetic resonance, solid-state NMR, and molecular dynamics simulations. This model paves the way for studying known HTTex1 fibril-specific binders and designing new ones.

INTRODUCTION

Huntington's disease (HD) is a fatal neurodegenerative disease for which there is no cure. HD is caused by an expansion of a polyglutamine (polyQ) encoding tract (CAG repeats) in exon 1 of the huntingtin gene (*HTTex1*) beyond 36 repeats (1). This makes HD the most common member of a class of diseases caused by polyQ expansions (2). Post-mortem examination of HD patient brains shows large β -sheet rich deposits of HTTex1 protein, which can be generated by aberrant splicing (3,4). Likewise, HTTex1 can form fibrils in vitro, and these fibrils have been shown to be toxic to cells (5,6).

HTTex1 can be divided into three domains (see Fig. 1): a 17-residue amphiphilic N-terminal (N17) domain, the polyQ tract of variable length, and a C-terminal domain that is rich in prolines (proline-rich domain (PRD)). The PRD contains two 11- and 10-residue polyproline (polyP) tracts that are linked by a proline-rich sequence. The polyQ tract forms the core of HTTex1 fibrils, and an elongated polyQ domain accelerates fibril formation and disease onset (7,8). The N17 domain has been shown to play a major role in fibril formation and adopts a helical structure in fibrils (9–11). The polyQ tract and N17 domain have highly static and intermediate dynamics, respectively (12–14).

In contrast, electron paramagnetic resonance (EPR) and NMR studies showed that the PRD remains in a highly dynamic state even after fibril formation (12–15) and that it has essentially the same conformation in fibrils as in monomers (12). The presence of the PRD is detrimental to fibril formation by polyQ peptides (16–19) and serves

Submitted April 17, 2020, and accepted for publication October 13, 2020.

*Correspondence: asiemer@usc.edu

Editor: Michael Sattler.

<https://doi.org/10.1016/j.bpj.2020.10.010>

© 2020 Biophysical Society.





FIGURE 1 Sequence of HTTex1. The N17 region is highlighted in orange, the polyQ domain in blue, the Pro-rich C-terminus in green, and the His-tag in black. The five DEER distances measured in this study are indicated with colored brackets. The residues that were assigned site specifically using solid-state NMR are highlighted in gray. To see this figure in color, go online.

as a binding site for a number of proteins (20–23). It has also been shown that HTTex1 fibrils that differ in their cellular toxicity differ in the structure and dynamics of the PRD rather than in their N17 domain or polyQ tract (5,24,25). Together, these findings support our previously published bottlebrush model of HTTex1 fibrils in which dynamic PRD bristles form the surface, and the polyQ and N17 regions form the less accessible core (12). Fibril toxicity will always be mediated by the interaction of the fibril surface with its environment. Therefore, our goal is to understand the structure of this fibril surface in detail. Biophysical studies and computer simulations indicated that the PRD adopts an extended polyproline II (PPII) helix (12,13,15,24,26). To test this hypothesis and to refine this model using inter-residue distances, we used a combined EPR, solid-state NMR, and molecular dynamics (MD) simulation-based approach.

The molecular description of intrinsically disordered proteins (IDPs) or intrinsically disordered domains (IDDs) is not a single structure but an ensemble of structures that describes the conformational flexibility of the protein. Both Monte Carlo and MD approaches can be used to generate such an ensemble. The quality of such an ensemble then needs to be either adjusted by selecting a suitable subset of conformers in the case of Monte Carlo simulations (27) or to be verified using experimental data in the case of MD simulations. Accurately reproducing experimental data, such as NMR relaxation rates, using MD simulations of IDPs is still challenging and a focus of active research (28,29). One problem is that most MD force fields have been developed for globular proteins and have a tendency to produce collapsed, globular structures, typically not found in IDPs. Therefore, the choice of suitable force fields and water models is crucial for obtaining MD trajectories that are compatible with experimental data (30).

In the following, we show how double electron-electron resonance (DEER) distances can be used to select a suitable water model, force field, and starting structure for atomistic simulation of the PRD of HTTex1. Our resulting MD simulations not only correctly reproduced DEER distances but also our previously reported NMR relaxation parameters (15). The resulting conformational ensemble shows that the PRD mainly adopts a polyproline II helix, although with a high degree of flexibility and kink in the linker be-

tween the two polyproline tracts and in the residues after the second polyproline tract.

MATERIALS AND METHODS

Protein expression and purification

HTTex1 fusion proteins were expressed, purified, and spin labeled as described previously (13,31). In short, the thioredoxin fusion protein of HTTex1 was recombinantly expressed in a pET32a vector using *Escherichia coli* BL21 (DE3) cells. The double Cys mutants for EPR measurements were first purified using a His60 column (Clontech Laboratories, Mountain View, CA) followed by labeling with 1-oxyl-2,2,5,5-tetramethyl- Δ 3-pyrroline-3-methylmethanethiosulfonate (MTSL) and then purified using a Hi-Trap Q XL anion exchange column (GE Healthcare) via an AKTA fast protein liquid chromatography system (Amersham Biosciences, Little Chalfont, UK). The labeling efficiency of ~99% was verified by comparing the protein concentration of the fusion protein determined via UV spectroscopy and spin concentration of the sample determined via the double integral of continuous-wave EPR spectra (recorded on a Bruker X-band EMX spectrometer (Bruker, Billerica, MA)).

Uniformly ^{13}C - ^{15}N -labeled HTTex1 fibril samples for solid-state NMR experiments were prepared as described previously (15).

EPR spectroscopy

Four-pulse DEER experiments (32) were done to determine the distance between spin labels. The measurements were done on a Bruker ELEXSYS E580 X-band pulse EPR spectrometer equipped with a 3-mm split ring (MS-3) resonator, a continuous-flow cryostat (CF935; Oxford Instruments, Abingdon, UK), and a temperature controller (ITC503S; Oxford Instruments) at a temperature of 78 K. For DEER measurements, 8-ns 90° and 16-ns 180° observe pulses and a 44-ns 180° electron-electron double resonance pulse were used. 20 μL of double spin-labeled samples were adjusted to a final fusion protein concentration of 20 μM , and 10% of glycerol was added as a cryoprotectant. The samples were flash frozen in liquid nitrogen before the measurement. Data were fitted using Tikhonov regularization as implemented in DEER Analysis 2019 (33). The regularization parameters were chosen using the L-curve corner criterion, resulting in regularization parameters of 125.9, 398.1, 125.9, and 31.6 for 63–75, 75–91, 91–102, and 101–114, respectively.

NMR spectroscopy

Solid-state NMR $R_{1\rho}$ rates were measured as described previously (15). In short, a 14.1 T Agilent DD2 solid-state NMR spectrometer with a T3 1.6 mm probe was used. The magic-angle spinning (MAS) frequency was 12 kHz, and the temperature was maintained at 0°C . Hard pulses were done with radio frequency field strengths of 200 and 50 kHz for ^1H and ^{15}N , respectively. $R_{1\rho}$ relaxation dispersion was measured with 6 and 18 kHz ^{15}N spin lock pulses. 2.5 kHz WALTZ ^1H decoupling was used during detection.

MD simulations

Simulations were run in OpenMM using the AMBER ff99SB force field along with the TIP4P-D water model. The PRD was simulated starting from an extended polyproline II helix conformation ($\Phi = -70.00$, $\Psi = 140.00$, $\Omega = 180.00$) (34–36). The starting structure was made using the ProBuilder web server (<https://nova.disfarm.unimi.it/probuilder.htm>). Assuming a 46-residue polyQ domain, residues Q63–P113 were simulated. The N, $\text{C}\alpha$, and CO atoms of Q63 (the last glutamine residue in

the HTTex1 sequence) were constrained to their starting positions by 0.5 kcal/Å restraints to simulate the impact of the PRD being attached to the static fibril core at this position. Two simulations were run for a total of 800 ns each. One with the fragment described above and another simulation with an additional six residue C-terminal His-tag. This was done to be consistent with both the NMR and EPR experiments that were performed with and without the His-tag (15). The chain was orientated along the z axis and centered inside a water box that was $70 \times 70 \times 210$ Å (His-tag) or $100 \times 100 \times 190$ Å (no His-tag). The simulations were run as an NPT (constant number, pressure, and temperature) ensemble with a temperature of 0°C and a pressure of 1 bar to match the conditions in which NMR experiments were performed. All proline residues were simulated starting in the *trans* conformation, and all histidines were simulated in their uncharged form. The system was neutralized via the addition of a single sodium atom. Both simulations were run using 2-fs timesteps, with fixed hydrogen bonds, and frames were taken every 20 ps. The first 200 ns of both simulations was regarded as the equilibration time and not used in the calculation of experimental parameters. The python script used to run the simulation and an OpenMM implementation of the TIP4P-D water model, the starting structures, and all simulation outputs are available upon request.

Calculation of experimental parameters

The program RotamerConvolveMD, which is based on the MDAnalysis python package (37–39), was used to add a set of different MTSL spin label rotamers to every 10th frame of the MD trajectories and to calculate the resulting distance distribution P_{MD} . These calculations used the MTSSL 298 K 2015 rotamer library (38).

R_1 and R_2 relaxation rates were obtained from the simulations using the equations described by Schanda and Ernst (40). First, the correlation function for the NH bond of each residue was computed as follows (41):

$$C(t) = \overline{1.5 * (NH(x) \cdot NH(x+t))} - 0.5.$$

Here, $NH(x)$ and $NH(x+t)$ are the normalized N-H vectors at time x and time $x+t$. The overbar indicates that $C(t)$ is averaged for all possible time points x during the simulation (41). Because there are fewer time points to average when t is larger, the correlation function was calculated for $t = 0$ ns to $t = 400$ ns. Afterwards, the correlation function for each residue was used to fit a model-free, biexponential decay function:

$$C(t) = ae^{-\frac{t}{T_1}} + (1-a)e^{-\frac{t}{T_2}},$$

where a is the relative weight of the two exponential decays with rates k_1 and k_2 . Using this model-free approach and the fact that the order parameter in the dynamic PRD is $S^2 \approx 0$, the spectral density function ($J(\omega)$, i.e., the Fourier transform of the correlation function) can be calculated using the following equation:

$$J(\omega) = \left(\frac{2}{5}\right) \left(\frac{a * k_1}{1 + (\omega k_1)^2} + \frac{(1-a) * k_2}{1 + (\omega k_2)^2} \right),$$

where ω is a frequency. Finally, the spectral density function was used to predict the relaxation rates R_1 and R_2 , using the following equations:

$$R_1 = \left(\frac{\delta_D}{4}\right)^2 (J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I + \omega_S)) + \delta_{CSA}^2 \left(\frac{3}{4}\right) J(\omega_S),$$

$$R_2 = \left(\frac{\delta_D}{4}\right)^2 \left(\frac{4}{3} J(\omega_r) + \left(\frac{2}{3}\right) J(2\omega_r) + \left(\frac{1}{2}\right) J(\omega_I - \omega_S) + 3J(\omega_I) + \left(\frac{3}{2}\right) J(\omega_S) + 3J(\omega_I + \omega_S) \right) + \omega_{CSA}^2 \left(\frac{1}{4} J(\omega_r) + \left(\frac{1}{8}\right) J(2\omega_r) + \left(\frac{3}{8}\right) J(\omega_S) \right),$$

where δ_D is the dipolar coupling between the amide proton and nitrogen, δ_{CSA} is the ^{15}N chemical shift anisotropy, ω_I is the ^1H Larmor frequency, ω_S is the ^{15}N Larmor frequency, and ω_r is the MAS frequency (40). The fit of the correlation function and the calculation of ^{15}N R_1 and R_2 were done using an in-house Mathematica script that is available upon request.

Chemical shifts for each frame of the simulations after 200 ns of equilibration were calculated using the program SHIFTX2 (42), and the resulting shifts were averaged over the entire simulation. The chemical shifts were then converted into secondary chemical shifts by subtracting site-specific random coil chemical shifts calculated using the program POTENCI (43).

Analysis of MD trajectory

The MDAnalysis (0.20) python package (37) (<https://www.mdanalysis.org/>) was used to calculate dihedral angles, C α -C α distances, and K-means clustering of the MD trajectories using in-house python scripts that are available upon request.

RESULTS

Measurement of EPR distances

As a reference data set for our simulations, we determined overall distance distributions within the C-terminus of the HTTex1 fusion protein using DEER EPR. We measured five distances within the C-terminus as indicated in Fig. 1: between 63R1 and 75R1 (where R1 refers to the spin-labeled side chain) in the first polyP stretch (P11); between 75R1 and 91R1 in the proline-rich linker between the two polyP stretches (L17); between 91R1 and 102R1 in the second polyP stretch (P10); between 101R1 and 114R1 in the C-terminal sequence (C12); and between 63R1 and 102R1, as a measure of the extension of the PRD. The DEER data, resulting distance distributions P_{DEER} , and the mode (i.e., the most frequent distance) of these distributions are shown in Fig. 2 (raw data are shown in Fig. S1).

MD simulation of the PRD

Our previous EPR and solid-state NMR data suggested that the PRD structure in fibrils and in the soluble fusion protein is highly similar (12,13). We therefore simulated a monomeric PRD. To compare the simulation to our solid-state NMR relaxation data recorded on HTTex1 fibrils, we fixed the N, CA, and CO atoms of the last glutamine of the polyQ domain (i.e., Q62). The C-terminus was placed in a periodic water-filled box as described in the Materials and Methods. We then ran short initial MD simulations to test several

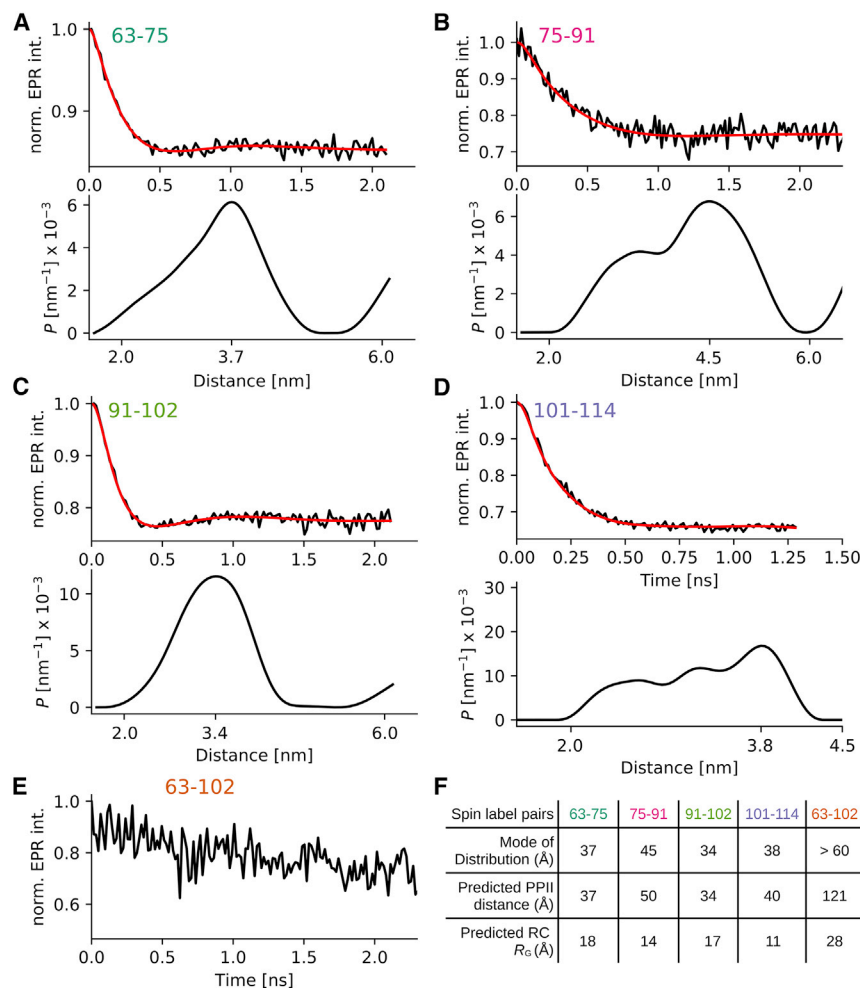


FIGURE 2 Distance distributions in the C-terminus of HTTex1 fusion protein measured via EPR DEER spectroscopy. (A–D) Top panels: baseline corrected DEER data (black) and fit using Tikhonov regularization is shown (red). Bottom panels: distance distribution P_{DEER} corresponding to fit is shown. The mode of the distribution is indicated using a tick mark. (E) Same as (A–D) but because the distance between 63 and 102 was above the detection limit, only the raw data without baseline subtraction and fit are shown. Raw data of the other distances are shown in Fig. S1. (F) Table listing mode of individual distances distributions is shown. For comparison, the theoretical distances for a polyP II helix assuming an increase of 3.1 Å per residue and the theoretical radius of hydration (R_H) for a random coil, are given (R_H was calculated using the formula by Marsh and Forman-Kay (44)). Both distances spanning the polyP regions (P11 and P10) correspond nicely to a PPII distance. The distances between residues 75–91 and 101–114 are significantly shorter than a PPII helix and longer than that expected for a random coil structure. To see this figure in color, go online.

force fields, water models, and starting structures for their ability to reproduce our DEER distances. We found that the starting structure (an extended polyproline II helix) was among the most important parameters to correctly reproduce the DEER distances. This starting structure was also justified by the presence of two long polyP tracts in the PRD and strong proline signals compatible with a PPII helix in our previous NMR spectra (12,15). In addition, the water model was an important factor. An implicit water model and the explicit TIP3P water model with the CHARMM36 force field (45) resulted in collapsed conformations that were inconsistent with our DEER measurements. This aligns with other studies that have shown that this water model is poorly suited for simulating highly extended and dynamic proteins (46). Using the AMBER ff99SB force field (35) for the protein in combination with the explicit TIP4P-D water model (34) developed specifically for intrinsically disordered proteins led to the best fit between the DEER distances (see Table S1). No experimental constraints were used during these simulations besides anchoring the N-terminus of the PRD as described

above. We used this combination of force field and water model together with an extended PPII starting structure for all further simulations. We then ran two separate 800 ns simulations, the first of the HTTex1 C-terminus with a C-terminal His-tag and a second simulation of the C-terminus without a His-tag. The RMSD analysis in Fig. S2 showed that both simulations had achieved equilibrium by 200 ns. All simulation frames following this time point were used for further analysis.

Comparison to DEER distances

To compare the DEER distance distributions, P_{DEER} , to the distance distributions from our MD simulations, we needed to add the corresponding MTSL labels and consider their flexibility (47). We did this by adding a set of different MTSL spin label rotamers to every 10th frame (i.e., every 200 ps) of our simulations using the program RotamerConvolveMD, which is based on the MDAnalysis python package (37–39). This program also calculates a resulting distance distribution, P_{MD} , that can be compared directly

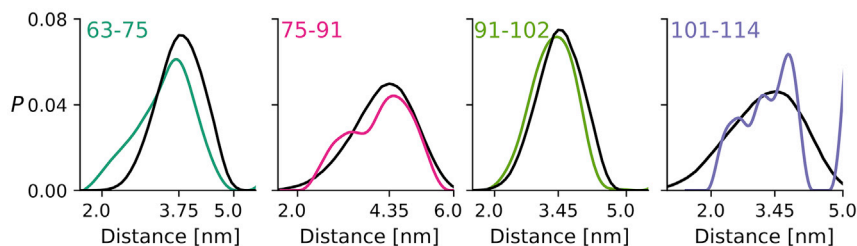


FIGURE 3 DEER distance distributions (P_{DEER}) and MTSL spin label distance distributions calculated from the MD trajectory (P_{MD}) correspond very well. DEER distance distributions between residue pairs as indicated are shown in colors. These correspond to distributions shown in Fig. 2. Distance distributions derived from the MD with His-tag simulation using the RotamerConvolveMD algorithm (P_{MD}) are shown in black. The mode of P_{MD} is indicated on the x axis. To see this figure in color, go online.

to the distance distribution P_{DEER} . As can be seen from Fig. 3, both the mode and the overall shape of the P_{MD} s calculated from the simulation with His-tag are very similar to those of P_{DEER} (the comparison with the simulation without His-tag is shown in Fig. S3).

Comparison to NMR parameters

We recently reported the assignment of the C-terminal residues of HTTex1 fibrils starting from residue G102 (C12 region) using a combination of solution and solid-state NMR techniques. This assignment allowed us to also measure site-specific ^{15}N R_1 and R_2 relaxation rates as well as residual ^1H - ^{15}N dipolar couplings that confirmed the highly dynamic nature of these residues in the context of the fibril (15). Using the traces from our MD simulations, we now calculated the R_1 and R_2 relaxation rates using the theoretical description by Schanda and Ernst (40) and the approach outlined in the Materials and Methods. The comparison of the measured and theoretical relaxation rates is shown in Fig. 4.

The calculated R_1 rates for the C12 region without a His-tag are consistent with the experimental data and, with two exceptions (G102 and E108), within the error margins of the actual rates. For the C-terminus with the His-tag, the calculated R_1 rates match the experimental data at the beginning

of the C12 region but show substantial differences for the residues preceding the His-tag and for the His-tag itself. These differences likely originate from the fact that the simulations were done with a nonprotonated state of the His-tag, whereas the actually His-tag would have been at least partly protonated.

The calculated R_2 rates are generally lower than the R_2 rates measured via $R_{1\rho}$ experiments. Again, the differences are larger for the C12 region with a His-tag than the C12 region without a His-tag, for which experimental and simulated R_2 -values correspond quite well. That the experimental R_2 rates are larger than those calculated from the MD simulations is not surprising considering that experimental R_2 rates usually include contributions other than transverse relaxation. In addition, we showed that there are still residual dipolar ^1H - ^{15}N couplings in the C12 that are not completely averaged out by motion (15). These couplings lead to coherent dephasing and thereby an experimental overestimation of R_2 (40).

The excellent fit between the R_2 and R_1 rates, especially for the C12 region without the His-tag, using simulations of less than 1 μs , indicates that there are no slow dynamics in this region that would not be captured by simulations that are too short. To test this interpretation, we compared $R_{1\rho}$ relaxation rates measured at spin lock fields of 6 and 18 kHz. These relaxation rates should differ (known as

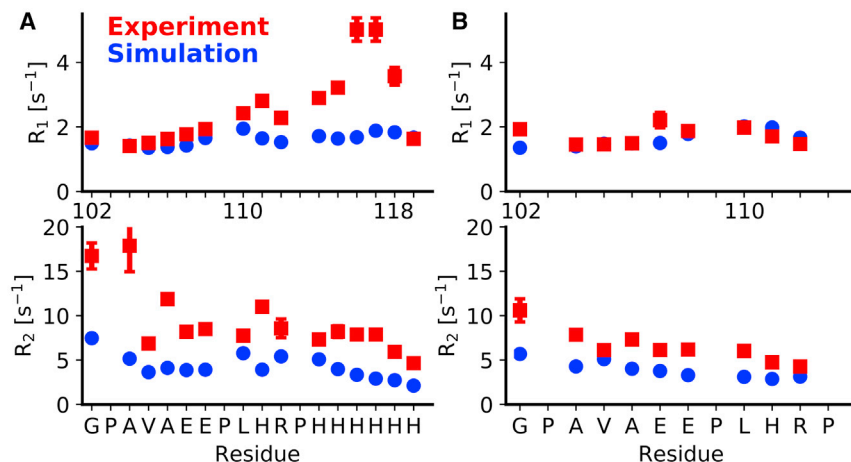


FIGURE 4 Good agreement between experimental and simulated R_1 and R_2 relaxation rates for the C12 region of HTTex1. (A) Shown is a comparison between R_1 and R_2 rates that were measured from fibrils of HTTex1 with a C-terminal His-tag and the equivalent relaxation rates calculated from simulations of the same construct. Proline residues are omitted because of the absence of an amide proton. Error bars indicate standard errors of the fits. (B) Shown is the same comparison as in (A) but between data and simulations of HTTex1 fibrils without a C-terminal His-tag. Experiment and simulations agree better for the HTTex1 fibrils without a His-tag, presumably because of the charge repulsion of the His-tag in our experiments that was not captured by our simulations. In the absence of a His-tag, most R_1 rates are within the margin of error of their calculated counterparts. The experimental

R_2 -values are slightly higher than those calculated from simulations because of residual coherent dephasing that leads to higher apparent R_2 rates. Experimental relaxation rates have been previously reported in (15). To see this figure in color, go online.

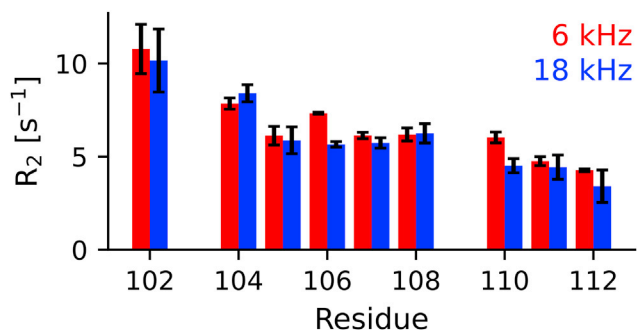


FIGURE 5 Absence of relaxation dispersion in the C12 region of HTTex1 indicates that there are no slow dynamical processes in this region. The effective R_2 rates calculated from $R_{1\rho}$ experiments with a spin lock radio frequency field strength of 6 and 18 kHz are shown as red and blue bars, respectively. Error bars indicate standard errors of the fits. Except for A106 and L110, these measurements are within the margin of error of each other. To see this figure in color, go online.

relaxation dispersion) if slow dynamics are present. As can be seen in Fig. 5, most of the R_2 -values calculated from the $R_{1\rho}$ rates are within the error range, confirming the absence of significant slow dynamics in the C-terminus of HTTex1 fibrils.

The absence of slow processes in the C-terminus of HTTex1 indicates that our simulations captured the conformational space of the PRD quite well. To further confirm this, we calculated site-specific $C\alpha$ chemical shifts from our simulations using the program SHIFTX2 (42). We computed the chemical shifts for each simulation frame after equilibration and calculated their average and SD. The comparison of the secondary $C\alpha$ shifts calculated this way, with our previously published $C\alpha$ shifts (15), is shown in Fig. 6.

Analysis of structural ensemble

The ability of the simulations to reproduce EPR and NMR measurements indicates that they form a representative ensemble of the structural distribution sampled by the PRD. Consequently, we analyzed the results of our simulations to gain additional insights into the structure and behavior of the PRD, with a focus on the simulation with His-tag (figures of the simulation without the His-tag can be found in the Supporting Material).

Visual inspection of the simulations showed that the two polyP stretches remained in relatively stable PPII helices, whereas the L17 region connecting the two polyP stretches and the C12 region were more flexible. Because PPII helices cannot be detected by the DSSP algorithm that is based on hydrogen bond formation (48), we analyzed individual ψ and ϕ angles for all residues postequilibration to confirm this observation. The average ψ and ϕ angles and their SD are shown in Fig. 7. The corresponding data for the simulation without His-tag are shown in Fig. S4. All Pro residues stayed within a canonical PPII helix and showed almost no flexibility in their ϕ angles (49). Interestingly, Pro residues outside or on the edges of the two polyproline stretches displayed more ψ angle flexibility. Almost all non-Pro residues adopted dihedral angles that were between the canonical angles for a β -sheet and a PPII helix. These angles are consistent with our observation that the C-terminus remains in a relatively extended conformation. Two exceptions to these extended dihedral angles were A83 and L86 in the no-His-tag simulation, which had average dihedral angles between those found in a β -sheet and an α -helix. The difference in the dihedral angles of A83 and L86 is one of the few major differences between the two simulations.

Another important exception from generally extended dihedral angles is G102 showing significant flexibility in its dihedral angles as illustrated by the large error bars in Fig. 7. This is not surprising given Gly's nature, but it is worth noting that all residues after G102 show much higher degrees of variation and disorder than the residues preceding G102.

But how flexible is the L17 region; does it allow the PRD to fold back onto itself? To address this question, we plotted $C\alpha$ - $C\alpha$ distances over the course of the simulation with His-tag. As can be seen from Fig. 8, the distance over the L17 region (75–91) is compatible with an extended PPII helix for most of the simulation with clear exceptions in which this region kinks, shortening its overall extension. In contrast, the two PPII stretches (63–75 and 91–102) remain essentially fixed in an extended PPII helical conformation. The C12 region (101–114) is the most flexible part of the PRD. Although it stays relatively extended throughout the simulation, it is generally not in an extended, PPII conformation. The overall extension of the PRD,

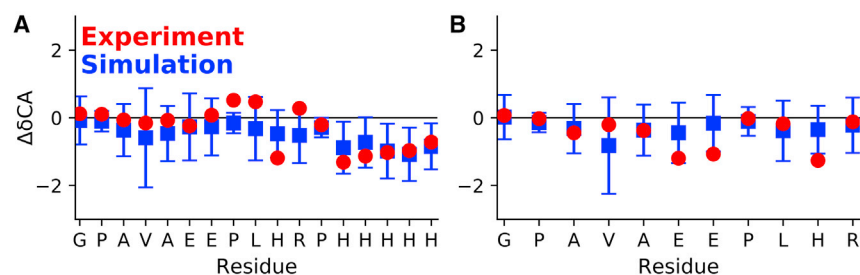


FIGURE 6 Calculated and experimental secondary $C\alpha$ chemical shifts of the PRD C12 region are in good agreement. (A) Average $C\alpha$ secondary chemical shifts of the simulation with a His-tag for the C12 region are shown in blue with error bars corresponding to their SD. The experimental secondary chemical shifts are shown in red. (B) Average $C\alpha$ secondary chemical shift of the simulation and experiments without a His-tag is shown. Error bars indicate standard deviations. To see this figure in color, go online.

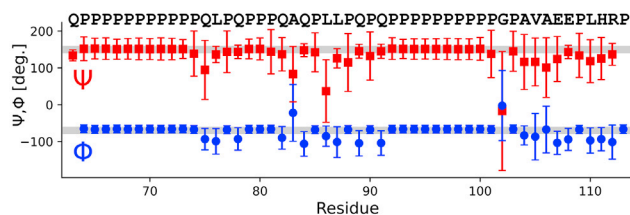


FIGURE 7 The variation in ϕ and ψ angles demonstrates relatively stable PPII helices for the polyP stretches and an increase in disorder for the L17 and C12 region. For the simulation with a His-tag, average per residue ϕ and ψ angles with their SDs are shown in red and blue, respectively. Error bars indicate standard deviations. The ϕ and ψ angles for an idealized PPII helix are indicated with gray bars. The very large error bars for residue G102 shows that the C12 region can rotate relatively freely around this residue. To see this figure in color, go online.

represented by the distance between Q63 and G102, is often correlated with the bending of the L17 region (see *gray boxes* in Fig. 8). The corresponding plots for the simulation without His-tag are shown in Fig. S5.

To get a better sense of the conformational space occupied by the PRD, we clustered the MD trajectory using the K-means algorithm. We determined a suitable number of clusters by dividing the trajectory into 2–20 clusters and calculated the point spread function value (pSF) and SSR/SST ratio (where SSR is the sum of squares regression and SST the total sum of squares) for each of these divisions (see Fig. S6). At a suitable number of clusters, point spread function reaches a local maximum, and the SSR/SST ratio starts to plateau (50). In our case, this was the case at three clusters. The centroids of each of these three clusters together with a schematic of the PRD are shown in Fig. 9. All centroids have extended P11 and P10 regions and a relatively disordered C12 region in common. The conformation of the L17 region determines the overall shape of the domain. Consequently, the PRD is relatively extended in centroid 2 where the L17 domain is extended as well, less extended in centroid 3 where the L17 regions adopts an s-shaped conformation, and significantly shortened in centroid 1 where the L17 is kinked. This analysis further confirms that although the PRD of HTTex1 is predominantly in a PPII helical conformation, it has the ability to kink at the L17 region. The bundles of structures along the trajectory for both simulations shown in Fig. S7 further illustrates the kinked but generally extended nature of the PRD.

DISCUSSION

This study showed that MD simulations using the AMBER ff99SB force field with the TIP4P-D water model led to trajectories for the PRD of HTTex1 that correlates very well with EPR DEER distance distributions and NMR ^{15}N relaxation rates and $C\alpha$ chemical shifts of the C12 region. Overall, the PRD stays relatively extended throughout the simulations with two stable PPII helices, P10 and P11, a

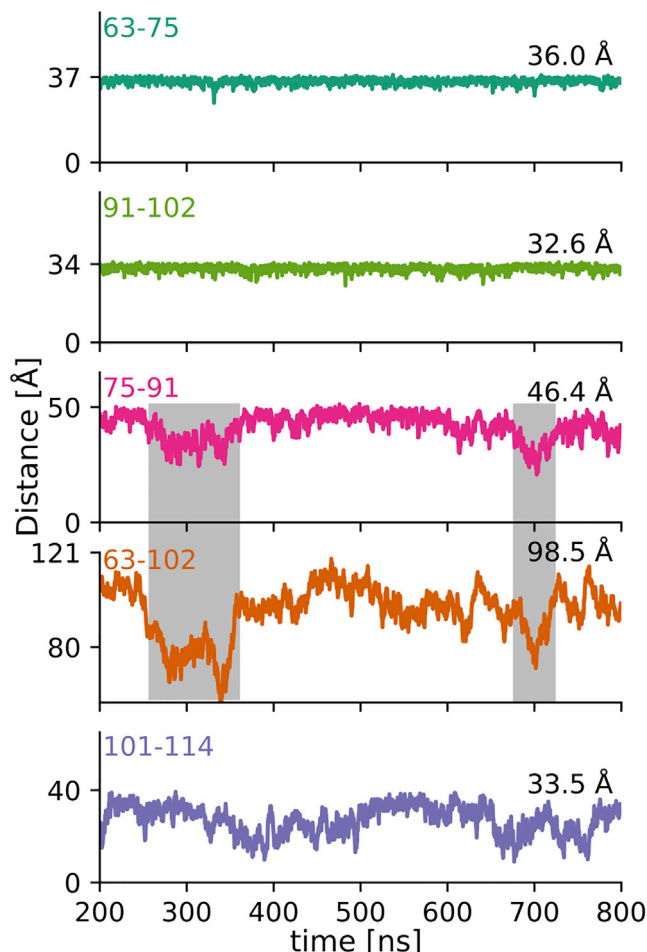


FIGURE 8 The L17 region is responsible for the overall kinks of the PRD. Change in $C\alpha$ - $C\alpha$ distances (as defined in Fig. 1) over time. The upper y axis tick indicates an ideal polyproline II helical distance (see Fig. 2 F); the mode of simulated $C\alpha$ - $C\alpha$ distance distribution is indicated on the right. All distance variations are plotted with comparable y axis scaling. Both polyP stretches stay in an extended PPII conformation for most of the simulation. The L17 region is a less ideal PPII helix. Its deviations from an extended PPII conformation induce an overall kink in the PRD as illustrated by the correlated shortening of distance 63–102 (*gray boxes*). The C12 region is the most flexible and, although still relatively extended, deviates from a PPII helix most of the time. To see this figure in color, go online.

more variable L17 linker region, and a very flexible C-terminal C12 region. Nevertheless, the L17 and, to a lesser extent, the C12 region have average dihedral angles compatible with a PPII helical or β -sheet conformation and are extended for most of the simulation, indicating that these regions are rather imperfect PPII helices than completely disordered. This extended PRD is also compatible with our previous observation that unbundled huntingtin (HTT) fibrils are spaced consistent with fibrils being held apart by extended polyproline bristles (12).

The large distribution of dihedral angles for G102 indicates that this residue may have a role in separating the Pro-rich area from subsequent HTT domains and effectively

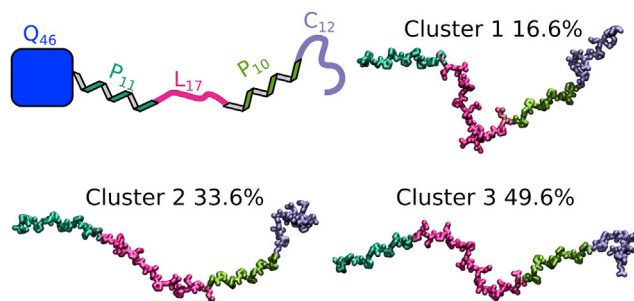


FIGURE 9 Cluster analysis of MD trajectory showing how the conformation of the L17 region defines the overall structure of the PRD. (A) Shown is a schematic of the PRD connected to the polyQ fibril core of HTTex1 fibrils. (B) Centroids of the K-means cluster analysis are shown. The percentages of total frames they represent are indicated. Clusters 2 and 3 are relatively extended, whereas cluster 1, having a strong kink in the L17 region, is less extended. To see this figure in color, go online.

terminates the order imposed by the Pro residues. In addition, the flexibility of G102 explains why the C12 region could be detected in our HSQC spectra in the absence of perdeuteration and at relatively slow MAS frequencies as reported previously (15). G102 allows the residues in the C12 region to rotate relatively freely, resulting in an order parameter that is essentially zero and an almost complete averaging of the ^1H - ^{15}N dipolar couplings that allowed the direct ^1H detection in our NMR experiments. In contrast, the preceding polyP regions could not be detected in the ^1H - ^{15}N HSQC experiment because of the absence of an amide proton in the polyP stretches and the reduced flexibility of the L17 region that was not enough to average the H-N dipolar coupling such that ^1H -detected experiments became feasible under the conditions used. It is interesting to note that the C12 region is evolutionary well conserved, and we speculate that it might serve as a dynamic linker to the well-structured and conserved first HEAT repeat of HTT (51,52).

Our finding that the PRD is mostly extended is compatible with the tadpole model of the HTTex1 monomer by Newcombe and co-workers in which the N17 and polyQ domains are more compact, and the PRD forms the extended tail of a tadpole-like structure (24). In contrast to our simulations, their modeling approach was based on Monte Carlo simulations and an implicit water model optimized for intrinsically disordered proteins as implemented in the ABSINTH program (53). This, and the fact that they assumed the His residues in the C12 region to be protonated, likely explains some of the differences with our results. Namely, the propensity of the C12 region to form an α -helix in their simulation.

For the two polyP regions, the mode of the $\text{C}\alpha$ - $\text{C}\alpha$ distance distribution of our simulation (see Fig. 8) is shorter than an idealized PPII helix but also a bit longer than what was described by Radhakrishnan and co-workers using the ABSINTH algorithm (54). This mode increases after the addition of spin labels using the RotamerConvolveMD algo-

rithm because for both distances (63–75 and 91–102), the labels point into different directions relative to the helix norm. The relatively good fit to the EPR distance distributions suggests that our simulations created a valuable model of the two polyP regions.

In contrast to other amino acid residues, proline can be found in both *trans* and *cis* conformations. The *trans* conformation used in our simulation is dominant. Depending on sequence and conformation, the *cis* conformation can be between 3 and 20% of the proline population (55). Because *trans*-*cis* isomerizations of prolines are, with time constants in the order of minutes, relatively slow (56), they were not part of our simulation and the influence of *cis* conformation was not part of our analysis. Urbanek et al. recently investigated the abundance of proline *cis* conformations in the PRD of HTTex1 (55). They showed that *cis* conformations were present in prolines with nonproline neighbors but were reduced below detection limit inside the P11 region. Similarly, we were not able to detect *cis* proline in our previous NMR study (12).

Our simulations focus on a single PRD because our previous EPR and NMR data showed that the structure of this domain is very similar in the soluble fusion protein and HTTex1 fibrils (12,13). That our simulations reproduce the solid-state NMR data from the C12 region in HTTex1 fibrils further supports this finding, suggesting that this region is not affected by potential PRD-PRD interactions inside the fibril. This seems to be also true for the rest of the PRD. The DEER distances within the PRD of the fusion protein and different fibril types, which we reported previously (25), are very similar.

Our results are consistent with the ability of the PRD to inhibit fibril formation. Because the PRD is dynamic in both soluble HTTex1 and the fibril, it likely counteracts fibril formation by imposing a PPII conformation on the polyQ domain rather than creating an entropic penalty from being placed into the fibril (16,57).

Many HTTex1-specific antibodies bind the PRD. MW7 and 4C9 bind the polyP regions, MW8 binds the C12 region, and PHP1 and PHP2 bind the L17 region (58–60). Interestingly, all of these antibodies are fibril specific and only weakly bind to soluble HTTex1. This work shows that the polyP, L17, and C12 regions not only differ in sequence but also in their degree of dynamics and deviation from a PPII structural motif. Therefore, it is possible that these epitopes are not only distinguished by their amino acid sequence but also by their structural preference. Similarly, we hope that the PRD model presented in this article will help understand how some fibril-specific HTTex1 interactors such as chaperones (61) bind.

CONCLUSIONS

We simulated the PRD of HTTex1 in fibrils using the AMBER ff99SB force field and TIP4P-D water model.

These simulations accurately predicted our EPR and solid-state NMR data, indicating that the PRD does not undergo slow processes that would not be captured by less than 1 μ s of simulation. The PRD adopted a predominantly PPII helical conformation for most of the MD trajectory. The two polyP regions formed stable PPII helices, the L17 region formed an imperfect PPII helix, and the C12 region only loosely maintained the PPII helical conformation. G102, at the beginning of the C12 region, was the most flexible residue, separating the PRD from the following highly conserved regions of HTT. Besides these structural insights, our study shows that modern MD methods in combination with EPR and solid-state NMR can accurately characterize intrinsic disorder in nonsoluble proteins.

SUPPORTING MATERIAL

Supporting Material can be found online at <https://doi.org/10.1016/j.bpj.2020.10.010>.

AUTHOR CONTRIBUTIONS

A.S.F. ran and analyzed MD simulations and co-wrote the article. J.M.B.-A. made EPR samples. J.V. measured EPR data. S.P. performed the cluster analysis of the MD data. R.L. coordinated the EPR work and its interpretation. A.B.S. conceived the study, recorded the NMR data, analyzed MD simulations, and co-wrote the manuscript.

ACKNOWLEDGMENTS

A.B.S. and R.L. would like to acknowledge funding from the National Institutes of Health (R01NS084345, R01GM110521) and the CHDI Foundation (Award A-12640). A.S.F. would like to acknowledge funding from the National Institutes of Health (F31GM120858). J.M.B.-A. would like to acknowledge a University of Southern California-Mexico's National Council of Science and Technology fellowship.

REFERENCES

- Lee, J. M., E. M. Ramos, ..., J. F. Gusella; PREDICT-HD Study of the Huntington Study Group (HSG); REGISTRY Study of the European Huntington's Disease Network; HD-MAPS Study Group; COHORT Study of the HSG. 2012. CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion. *Neurology*. 78:690–695.
- Gatchel, J. R., and H. Y. Zoghbi. 2005. Diseases of unstable repeat expansion: mechanisms and common principles. *Nat. Rev. Genet.* 6:743–755.
- Sathasivam, K., A. Neueder, ..., G. P. Bates. 2013. Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. *Proc. Natl. Acad. Sci. USA*. 110:2366–2370.
- DiFiglia, M., E. Sapp, ..., N. Aronin. 1997. Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science*. 277:1990–1993.
- Nekooki-Machida, Y., M. Kurosawa, ..., M. Tanaka. 2009. Distinct conformations of in vitro and in vivo amyloids of huntingtin-exon1 show different cytotoxicity. *Proc. Natl. Acad. Sci. USA*. 106:9679–9684.
- Pieri, L., K. Madiona, ..., R. Melki. 2012. Fibrillar α -synuclein and huntingtin exon 1 assemblies are toxic to the cells. *Biophys. J.* 102:2894–2905.
- Chen, S., F. A. Ferrone, and R. Wetzel. 2002. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc. Natl. Acad. Sci. USA*. 99:11884–11889.
- Brinkman, R. R., M. M. Mezei, ..., M. R. Hayden. 1997. The likelihood of being affected with Huntington disease by a particular age, for a specific CAG size. *Am. J. Hum. Genet.* 60:1202–1210.
- Sivanandam, V. N., M. Jayaraman, ..., P. C. A. van der Wel. 2011. The aggregation-enhancing huntingtin N-terminus is helical in amyloid fibrils. *J. Am. Chem. Soc.* 133:4558–4566.
- Crick, S. L., K. M. Ruff, ..., R. V. Pappu. 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.
- Jayaraman, M., R. Kodali, ..., R. Wetzel. 2012. Slow amyloid nucleation via α -helix-rich oligomeric intermediates in short polyglutamine-containing huntingtin fragments. *J. Mol. Biol.* 415:881–899.
- Isas, J. M., R. Langen, and A. B. Siemer. 2015. Solid-state nuclear magnetic resonance on the static and dynamic domains of huntingtin exon-1 fibrils. *Biochemistry*. 54:3942–3949.
- Bugg, C. W., J. M. Isas, ..., R. Langen. 2012. Structural features and domain organization of huntingtin fibrils. *J. Biol. Chem.* 287:31739–31746.
- Hoop, C. L., H.-K. Lin, ..., P. C. A. van der Wel. 2014. Polyglutamine amyloid core boundaries and flanking domain dynamics in huntingtin fragment fibrils determined by solid-state nuclear magnetic resonance. *Biochemistry*. 53:6653–6666.
- Caulkins, B. G., S. A. Cervantes, ..., A. B. Siemer. 2018. Dynamics of the proline-rich C-terminus of huntingtin exon-1 fibrils. *J. Phys. Chem. B*. 122:9507–9515.
- Bhattacharyya, A., A. K. Thakur, ..., R. Wetzel. 2006. Oligoproline effects on polyglutamine conformation and aggregation. *J. Mol. Biol.* 355:524–535.
- Dehay, B., and A. Bertolotti. 2006. Critical role of the proline-rich region in Huntingtin for aggregation and cytotoxicity in yeast. *J. Biol. Chem.* 281:35608–35615.
- Darnell, G., J. P. R. O. Orgel, ..., S. C. Meredith. 2007. Flanking polyproline sequences inhibit beta-sheet structure in polyglutamine segments by inducing PPII-like helix structure. *J. Mol. Biol.* 374:688–704.
- Darnell, G. D., J. Derryberry, ..., S. C. Meredith. 2009. Mechanism of cis-inhibition of polyQ fibrillation by polyP: PPII oligomers and the hydrophobic effect. *Biophys. J.* 97:2295–2305.
- Liu, Y. F., R. C. Deth, and D. Devys. 1997. SH3 domain-dependent association of huntingtin with epidermal growth factor receptor signaling complexes. *J. Biol. Chem.* 272:8121–8124.
- Sittler, A., S. Walter, ..., E. E. Wanker. 1998. SH3GL3 associates with the Huntingtin exon 1 protein and promotes the formation of polyglutamine-containing protein aggregates. *Mol. Cell*. 2:427–436.
- Qin, Z.-H., Y. Wang, ..., M. DiFiglia. 2004. Huntingtin bodies sequester vesicle-associated proteins by a polyproline-dependent interaction. *J. Neurosci.* 24:269–281.
- Caron, N. S., C. R. Desmond, ..., R. Truant. 2013. Polyglutamine domain flexibility mediates the proximity between flanking sequences in huntingtin. *Proc. Natl. Acad. Sci. USA*. 110:14610–14615.
- Newcombe, E. A., K. M. Ruff, ..., D. M. Hatters. 2018. Tadpole-like conformations of huntingtin exon 1 are characterized by conformational heterogeneity that persists regardless of polyglutamine length. *J. Mol. Biol.* 430:1442–1458.
- Isas, J. M., N. K. Pandey, ..., A. B. Siemer. 2019. Huntingtin fibrils with different toxicity, structure, and seeding potential can be reversibly interconverted. *bioRxiv* <https://doi.org/10.1101/703769>.
- Warner, J. B., IV, K. M. Ruff, ..., H. A. Lashuel. 2017. Monomeric huntingtin exon 1 has similar overall structural features for wild-type and

- pathological polyglutamine lengths. *J. Am. Chem. Soc.* 139:14456–14469.
27. Ozenne, V., F. Bauer, ..., M. Blackledge. 2012. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics.* 28:1463–1470.
 28. Salvi, N., A. Abyzov, and M. Blackledge. 2016. Multi-timescale dynamics in intrinsically disordered proteins from NMR relaxation and molecular simulation. *J. Phys. Chem. Lett.* 7:2483–2489.
 29. Salvi, N., A. Abyzov, and M. Blackledge. 2017. Atomic resolution conformational dynamics of intrinsically disordered proteins from NMR spin relaxation. *Prog. Nucl. Magn. Reson. Spectrosc.* 102–103:43–60.
 30. Salvi, N., A. Abyzov, and M. Blackledge. 2019. Solvent-dependent segmental dynamics in intrinsically disordered proteins. *Sci. Adv.* 5:eaaax2348.
 31. Isas, J. M., A. Langen, ..., A. B. Siemer. 2017. formation and structure of wild type huntingtin exon-1 fibrils. *Biochemistry.* 56:3579–3586.
 32. Pannier, M., S. Veit, ..., H. W. Spiess. 2000. Dead-time free measurement of dipole-dipole interactions between electron spins. *J. Magn. Reson.* 142:331–340.
 33. Jeschke, G., V. Chechik, ..., H. Jung. 2006. DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. *Appl. Magn. Reson.* 30:473–498.
 34. Piana, S., A. G. Donchev, ..., D. E. Shaw. 2015. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B.* 119:5113–5123.
 35. Hornak, V., R. Abel, ..., C. Simmerling. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins.* 65:712–725.
 36. Eastman, P., J. Swails, ..., V. S. Pande. 2017. OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* 13:e1005659.
 37. Michaud-Agrawal, N., E. J. Denning, ..., O. Beckstein. 2011. MDA-analysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* 32:2319–2327.
 38. Polyhach, Y., E. Bordignon, and G. Jeschke. 2011. Rotamer libraries of spin labelled cysteines for protein studies. *Phys. Chem. Chem. Phys.* 13:2356–2366.
 39. Stelzl, L. S., P. W. Fowler, ..., O. Beckstein. 2014. Flexible gates generate occluded intermediates in the transport cycle of LacY. *J. Mol. Biol.* 426:735–751.
 40. Schanda, P., and M. Ernst. 2016. Studying dynamics by magic-angle spinning solid-state NMR spectroscopy: principles and applications to biomolecules. *Prog. Nucl. Magn. Reson. Spectrosc.* 96:1–46.
 41. Maragakis, P., K. Lindorff-Larsen, ..., D. E. Shaw. 2008. Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins. *J. Phys. Chem. B.* 112:6155–6158.
 42. Han, B., Y. Liu, ..., D. S. Wishart. 2011. SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR.* 50:43–57.
 43. Nielsen, J. T., and F. A. A. Mulder. 2018. POTENCI: prediction of temperature, neighbor and pH-corrected chemical shifts for intrinsically disordered proteins. *J. Biomol. NMR.* 70:141–165.
 44. Marsh, J. A., and J. D. Forman-Kay. 2010. Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* 98:2383–2390.
 45. Huang, J., and A. D. MacKerell, Jr. 2013. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* 34:2135–2145.
 46. Henriques, J., C. Cragnell, and M. Skepö. 2015. Molecular dynamics simulations of intrinsically disordered proteins: force field evaluation and comparison with experiment. *J. Chem. Theory Comput.* 11:3420–3431.
 47. Hatmal, M. M., Y. Li, ..., I. S. Haworth. 2012. Computer modeling of nitroxide spin labels on proteins. *Biopolymers.* 97:35–44.
 48. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.
 49. Adzhubei, A. A., M. J. E. Sternberg, and A. A. Makarov. 2013. Polyproline-II helix in proteins: structure and function. *J. Mol. Biol.* 425:2100–2132.
 50. Shao, J., S. W. Tanner, ..., T. E. Cheatham. 2007. Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *J. Chem. Theory Comput.* 3:2312–2334.
 51. Tartari, M., C. Gissi, ..., E. Cattaneo. 2008. Phylogenetic comparison of huntingtin homologues reveals the appearance of a primitive polyQ in sea urchin. *Mol. Biol. Evol.* 25:330–338.
 52. Guo, Q., Bin Huang, ..., S. Kochanek. 2018. The cryo-electron microscopy structure of huntingtin. *Nature.* 555:117–120.
 53. Vitalis, A., and R. V. Pappu. 2009. ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions. *J. Comput. Chem.* 30:673–699.
 54. Radhakrishnan, A., A. Vitalis, ..., R. V. Pappu. 2012. Improved atomistic Monte Carlo simulations demonstrate that poly-L-proline adopts heterogeneous ensembles of conformations of semi-rigid segments interrupted by kinks. *J. Phys. Chem. B.* 116:6862–6871.
 55. Urbanek, A., M. Popovic, ..., P. Bernadó. 2020. Evidence of the reduced abundance of proline cis conformation in protein poly proline tracts. *J. Am. Chem. Soc.* 142:7976–7986.
 56. Dirnbach, E., D. G. Steel, and A. Gafni. 1999. Proline isomerization is unlikely to be the cause of slow annealing and reactivation during the folding of alkaline phosphatase. *J. Biol. Chem.* 274:4532–4536.
 57. Pandey, N. K., J. M. Isas, ..., R. Langen. 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.
 58. Ko, J., S. Ou, and P. H. Patterson. 2001. New anti-huntingtin monoclonal antibodies: implications for huntingtin conformation and its binding proteins. *Brain Res. Bull.* 56:319–329.
 59. Landles, C., K. Sathasivam, ..., G. P. Bates. 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.
 60. Ko, J., J. M. Isas, ..., A. Khoshnan. 2018. Identification of distinct conformations associated with monomers and fibril assemblies of mutant huntingtin. *Hum. Mol. Genet.* 27:2330–2343.
 61. Scior, A., A. Buntru, ..., J. Kirstein. 2018. Complete suppression of Htt fibrilization and disaggregation of Htt fibrils by a trimeric chaperone complex. *EMBO J.* 37:282–299.