Monomeric Huntingtin Exon 1 Has Similar Overall Structural Features for Wild-Type and Pathological Polyglutamine Lengths

John B. Warner, IV,[†,#] Kiersten M. Ruff,[‡,#] Piau Siong Tan,[§] Edward A. Lemke,[§] Rohit V. Pappu,*[‡] and Hilal A. Lashuel*[†]

[†]Laboratory of Molecular and Chemical Biology of Neurodegeneration, Brain Mind Institute, Station 19, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

[‡]Department of Biomedical Engineering and Center for Biological Systems Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States

[§]Structural and Computational Biology Unit, Cell Biology and Biophysics Unit, European Molecular Biology Laboratory (EMBL), 69117 Heidelberg, Germany

S *Supporting Information*

ABSTRACT: Huntington's disease is caused by expansion of a polyglutamine (polyQ) domain within exon 1 of the huntingtin gene (Httex1). The prevailing hypothesis is that the monomeric Httex1 protein undergoes sharp conformational changes as the polyQ length exceeds a threshold of 36−37 residues. Here, we test this hypothesis by combining novel semi-synthesis strategies with state-of-the-art single-molecule Förster resonance energy transfer measurements on biologically relevant, monomeric Httex1 proteins of five different polyQ lengths. Our results, integrated with atomistic simulations, negate the hypothesis of a sharp, polyQ length-dependent change in the structure of monomeric Httex1. Instead, they support a continuous global compaction with increasing polyQ length that derives from increased prominence of the globular polyQ domain. Importantly, we show that monomeric Httex1 adopts tadpole-like architectures for polyQ lengths below and above the pathological threshold. Our results suggest that higher order homotypic and/or heterotypic interactions within distinct sub-populations of neurons, which are inevitable at finite cellular concentrations, are likely to be the main source of sharp polyQ length dependencies of HD.
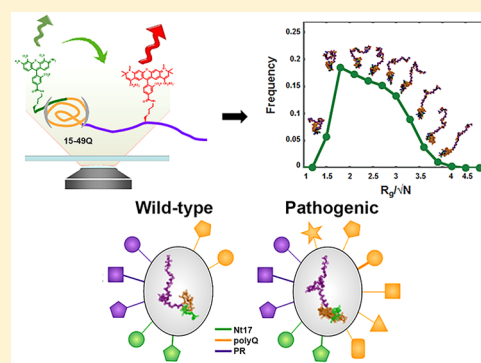
INTRODUCTION

Huntington's disease (HD) is a devastating inherited neurodegenerative disorder that is caused by mutational expansion of a CAG repeat region within the first exon of the huntingtin (Htt) gene.[1] Ages of onset and disease severity are inversely correlated with the length of the CAG repeat expansion. On average, the penetrance and severity at onset increase sharply above a threshold CAG repeat length of 36,[2] although there is considerable variability in the length dependence of the disease phenotype, as quantified in clinical studies.[3]

Recent studies have demonstrated the possibility of CAG repeat-length-dependent aberrant splicing that leads to Htt exon 1 spanning transcripts.[4] When translated, these transcripts yield Htt exon 1 encoded protein fragments, referred to hereafter as Httex1. The sequence architecture of Httex1 is modular. The CAG repeat encodes a central polyglutamine (polyQ) domain. This is flanked N-terminally by a 17-residue amphipathic stretch (Nt17) and C-terminally by a 50-residue proline-rich (PR) domain. N-terminal fragments of the Htt protein, including Httex1, are among the smallest proteins that recapitulate HD pathology in mouse models.[5] These fragments form neuronal intranuclear inclusions and are associated with

the formation of dystrophic neurites in the cortex and striatum in HD.[5] Additionally, Httex1 and N-terminal fragments of Httex1 with expanded polyQ tracts aggregate and lead to toxicity in cell culture models.[6]

The existence of a pathogenic polyQ length threshold for HD has led to the expectation that there should be a sharp conformational change within monomeric Httex1 at and above the pathogenic polyQ length.[7] A direct test of this hypothesis requires atomic-level structural characterization of monomeric Httex1 as a function of polyQ length. These studies have to be performed in the absence of confounding contributions from intermolecular associations. However, detailed structural studies of monomeric forms of monomeric Httex1 are challenging because of the high aggregation propensity and the polyQ-length-dependent insolubility of Httex1,[8] the repetitive nature of the polyQ and PR domains, and the sequence-encoded preference for conformational heterogeneity.[9]
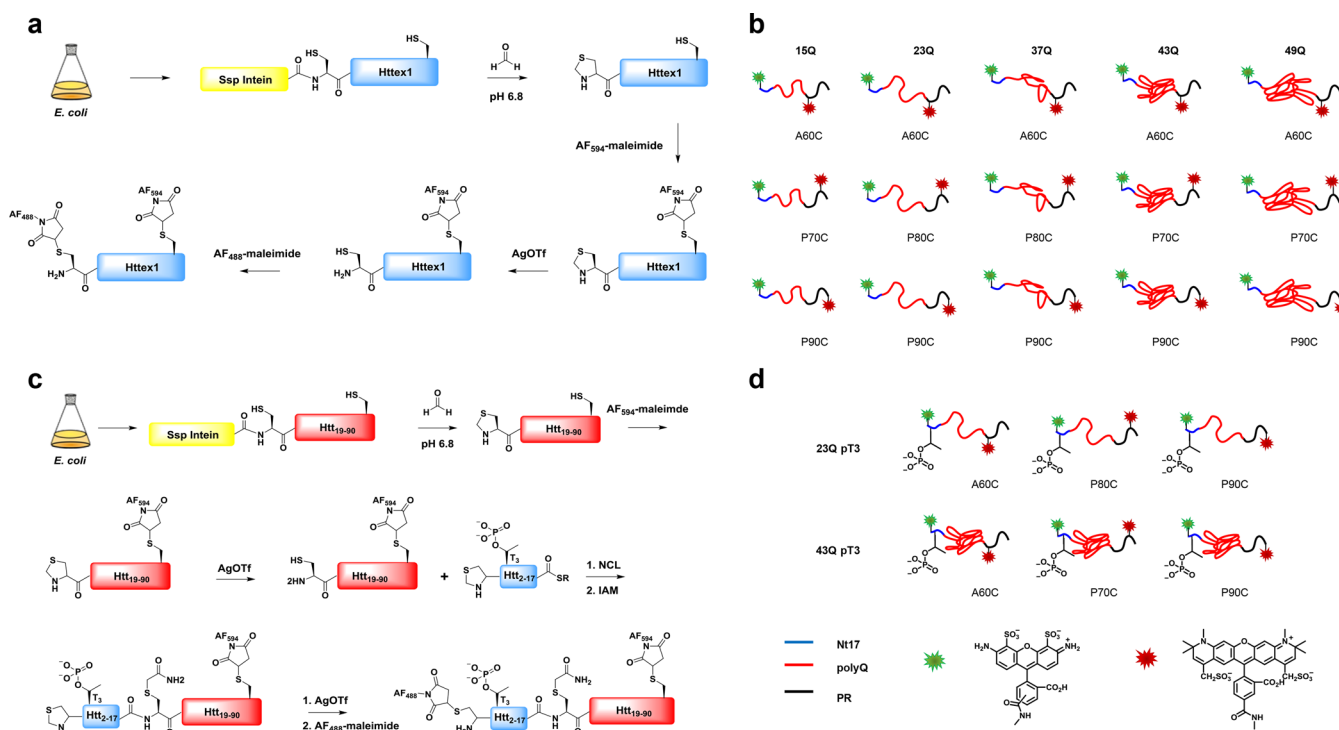
Figure 1. Site-specifically dual-labeled Httex1 library for smFRET measurements. (a) General strategy for sequential labeling of Httex1 at the N-terminal residue and within the C-terminal proline-rich region spanning residues 60−90. Here, Alexa 594 is denoted as AF_{594}, Alexa 488 as AF_{488}, and silver triflate as AgOTf. (b) Unmodified dual-labeled Httex1 constructs labeled with Alexa 488 (green) at the N-terminus and Alexa 594 (red) at the indicated C-terminal position. (c) Semi-synthetic strategy for obtaining dual-labeled Httex1 proteins containing site-specific post-translational modifications within Nt17, e.g., threonine 3 (T3) phosphorylation. (d) pT3-modified dual-labeled Httex1 constructs prepared as described in (c).

Httex1 molecules are highly insoluble and their solubility limits fall below the micromolar range with increasing polyQ length.[8] This poses serious challenges for interpreting data from methods such as nuclear magnetic resonance (NMR) spectroscopy, vibrational spectroscopy, or small-angle X-ray scattering. These methods require protein concentrations that are in the micromolar to millimolar range. Heterogeneous mixtures of monomers, oligomers, and higher-order aggregates inevitably confound interpretations from structural studies and make it difficult to compare results obtained from different techniques and laboratories. Furthermore, the repetitive nature of the polyQ and PR domains can lead to overlapping signals that are difficult to deconvolve. To overcome problems posed by the poor solubility of Httex1, solubilizing sequences (e.g., oligolysine tags) or proteins (e.g., GST, MBP) are usually added to the N- or C-terminal ends of Httex1 proteins and model systems.[7,10] To monitor polyQ-mediated conformational changes and aggregation in cellular models of HD, fluorescent proteins such as GFP and YFP are commonly fused to N- and/or C-terminal ends of Httex1.[7] These protein domains, which are typically larger than 20 kDa, are as large as or larger than the Httex1 construct of interest and can have a significant influence on conformational properties as evidenced by their ability to modulate Httex1 solubility and aggregation mechanisms.[7,11] Even the addition of minimally perturbing solubilizing flanking residues (e.g., Lys_n ($n = 1−8$)) can lead to substantial alterations of the complex aggregation landscape and phase behavior of Httex1 constructs.[8,10]

Bioinformatics predictions, computer simulations,[12] and NMR[13] studies on small fragments suggest that Httex1 molecules are intrinsically disordered. This designation implies that Httex1 molecules are likely to display considerable conformational heterogeneity and lack persistent secondary and tertiary structures. To characterize the conformational ensembles of intrinsically disordered proteins (IDPs) such as Httex1, we need quantitative assessments of intramolecular distances, the amplitudes of conformational fluctuations, and the overall shapes and sizes of molecules. Importantly, such measurements need to be made under conditions where there are no confounding contributions from intermolecular associations.

Here, we report results from investigations that deploy a novel combination of single-molecule Förster resonance energy transfer (smFRET) measurements on site-specifically labeled semisynthetic Httex1 proteins and atomistic computer simulations that are based on the ABSINTH implicit solvation model and force field paradigm.[14] We deployed a recently developed intein-based expression system that enables the generation of *bona fide* Httex1 proteins with polyQ repeats below and above the pathogenic threshold ($Q_{15−49}$).[15] Our investigations quantify the variation of intramolecular distances within Httex1 as a function of polyQ length and provide the first complete structural characterization of monomeric forms of Httex1. The smFRET measurements allow us to characterize the conformational properties of Httex1 constructs in the sub-nanomolar regime, where confounding effects of oligomerization are readily avoided.[16]

Our semi-synthetic strategy allowed us to generate Httex1 with polyQ tracts of five different lengths, viz., $n = 15, 23, 37, 43, 49$. These lengths span the range starting from below and going above the pathological threshold length of 36−37 glutamine residues. We introduced sequential and site-specific donor and acceptor fluorophores to obtain homogeneous dual-labeled Httex1 proteins with consistent localization of the
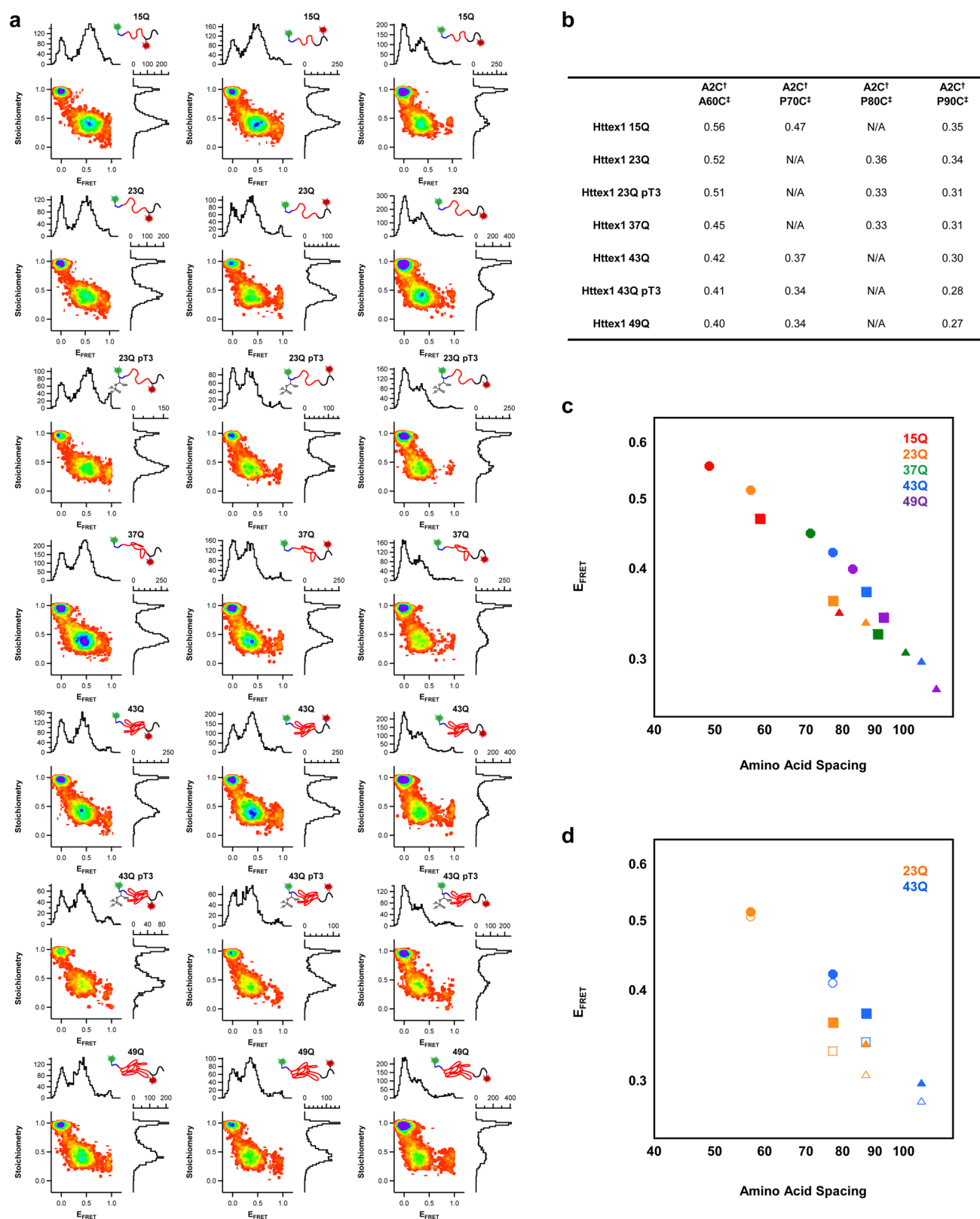
Figure 2. smFRET measurements of Httex1. (a) Two-dimensional E_{FRET} versus S histograms for Httex1 15−49Q with acceptor labeled at position A60C, P70C, P80C, or P90C. (b) $\langle E_{FRET}\rangle$ values calculated from 2D Gaussian fits of E_{FRET} S versus histograms. A2C[†] indicates labeling with Alexa488; P90C[‡] indicates labeling with Alexa594, and NA denotes where no construct is made. (c) Double logarithmic plot of $\langle E_{FRET}\rangle$ versus donor−acceptor amino acid spacing for the unmodified Httex1 constructs. Acceptor label positions are indicated as follows: proximal to the polyQ domain (●), within the PR domain (■), and C-terminal (▲). (d) Double logarithmic plot of $\langle E_{FRET}\rangle$ versus donor−acceptor amino acid spacing for the pT3-modified Httex1 constructs. Unmodified constructs are shown in filled shapes and pT3-modified constructs as open shapes with acceptor positions as indicated previously.

donor and acceptor fluorophores. This enables accurate structural studies of monomeric Httex1 using smFRET measurements. The semi-synthetic strategy also allowed the incorporation of specific post-translational modifications (PTMs) within the Nt17 domain, thus enabling the investigation of the role of PTMs, such as the phosphorylation of Thr 3,[17] in modulating the conformational properties of Httex1.

For each polyQ length, we performed three distinct sets of smFRET experiments to obtain quantitative assessments of the intramolecular distances and the global conformational properties of monomeric Httex1. The measured smFRET efficiencies were combined with a maximum entropy method[18] to reweight ensembles obtained from atomistic simulations. Our approach yields the first-ever detailed atomistic description of the conformational ensembles for Httex1 as a function of polyQ length. We find that monomeric Httex1 adopts "tadpole-like" conformations characterized by a globular head comprising of Nt17 adsorbed on the surface of the polyQ domain and a semi-flexible PR domain that adopts mostly expanded conformations. Additionally, we observed a continuous global compaction of Httex1 as polyQ length increased. This arises from the increased prominence of the globular polyQ domain and does not reflect any special intramolecular interactions among the three domains. Our results negate the hypothesis of a sharp, polyQ length-dependent change in the structure of monomeric Httex1 that emerges from certain classes of computer simulations,[19] although models that invoke sharp structural changes of monomeric Httex1 within oligomers[20] cannot be ruled out. Taken together, our findings provide a structural rationalization for the large variability in the age of onset for a given polyQ repeat length.[2] Importantly, our results suggest that higher order homotypic and/or heterotypic interactions within distinct sub-populations of neurons, as opposed to sharp conformational changes within monomeric Httex1, are likely to be the main source of sharp polyQ length dependencies of HD.

■ RESULTS

Novel Strategies Yield Constructs for smFRET Measurements. smFRET measurements require the generation of fluorescently labeled molecules. This involves the introduction of a pair of cysteine residues and their covalent modification with donor and acceptor fluorophores via maleimide chemistry. These strategies typically result in a heterogeneous mixture of single and dual, albeit randomly, labeled proteins[21] although further preferential labeling can be achieved via kinetic control[22] or through chromatographic separation.[23] We have developed a broadly applicable strategy for sequential, site-specific dual fluorophore protein labeling utilizing cysteine chemistry. We adapted methodologies from peptide science that take advantage of the selective reaction between an N-terminal cysteine and formaldehyde to form a thiazolidine adduct that can be selectively deprotected under mild conditions to allow selective and sequential labeling of cysteine residues.[24] By incorporating this approach into our Ssp-intein-based strategy for producing Httex1 proteins,[15b] we were able to produce site-specific, dual-labeled Httex1 with different polyQ lengths. These Ssp-Httex1 constructs were designed with a fixed cysteine at the N-terminus of Httex1 and a second motile cysteine in the PR domain (Figure 1a). Double cysteine constructs where the N-terminal cysteine residue is thiazolidine protected were obtained by addition of formaldehyde during

the Ssp-Httex1 splicing reaction. Following purification of the N-terminally protected construct, the unprotected C-terminal cysteine was rapidly labeled with an acceptor Alexa 594-maleimide probe. For each polyQ repeat length (15, 23, 37, 43, or 49Q) the acceptor was positioned at position A60C proximal to the polyQ domain, either position P70C or P80C internal to the PR domain, or at the C-terminal P90C residue. Labeling of Httex1 with the acceptor probe was quantitative and site-selective. Deprotection of the N-terminal thiazolidine was then achieved by treatment with silver triflate under mild conditions.[24] The acceptor-labeled Httex1, with a liberated N-terminal cysteine, was then labeled with the donor Alexa 488-maleimide probe. Site-specifically dual-labeled Httex1 constructs were obtained in high purity as determined by sodium dodecyl sulfide−polyacrylamide gel electrophoresis (SDS-PAGE), liquid chromatography mass spectrometry (LC-MS), and C8 reversed-phase ultra-performance liquid chromatography (RP-UPLC) (see Supporting Information, Figures S1−S4 and Table S1).

To be able to investigate the effect of N-terminal PTMs on the structure of Httex1, we also developed a modified semi-synthetic strategy that allows the site-specific introduction of PTMs and sequential labeling of the protein. This strategy was then used to incorporate a phosphorylated threonine residue (pT3) at position 3 into dual-labeled Httex1 constructs (Figure 1c). We focused on T3 for the following reasons: (1) it is the most common N-terminal PTM;[17] (2) the levels of T3 phosphorylation are inversely correlated with the polyQ length repeat;[17] (3) phosphorylation at T3 induces the most pronounced stabilizing effect on the α-helical conformation of the Nt17 domain of Httex1;[25] and (4) the T3-specific kinases have not yet been identified.[26] As with Httex1, the C-terminal 18−90 fragments were expressed from *E. coli* and thiazolidine protected following splicing from the Ssp-Intein. The Nt17 fragment was prepared containing an N-terminal thiazolidine, pT3, and a C-terminal thioester by solid-phase peptide synthesis. Following native chemical ligation (NCL), the ligation site cysteine, C18, was masked by treatment with iodoacetamide to generate a glutamine mimetic rather than desulfurization to alanine. This strategy allows for the site-specific incorporation of single or multiple Nt17 PTMs, thus enabling studies to elucidate the effect of these PTMs on the conformational ensemble of Httex1 at the monomeric and oligomeric levels. Using the recombinant and semi-synthetic strategies described above, we prepared a library of 15 site-specifically dual fluorophore-labeled Httex1 constructs from 15 to 49Q and 6 pT3-modified dual-labeled constructs of 23Q and 43Q (Figure 1b,d) that were suitable for smFRET measurements.

PolyQ Repeat Length Dependence of Httex1 Conformations Obtained Using smFRET. We used smFRET to investigate the effect of polyQ repeat length on intramolecular distances within Httex1. The data were used to generate two-dimensional FRET efficiency (E_{FRET}) versus stoichiometry (S) histograms (Figure 2a).[27] Mean E_{FRET}, $\langle E_{FRET} \rangle$, values were obtained from a two-dimensional Gaussian fit of S versus E_{FRET} histograms (Figure 2b). Given that Httex1 and Httex1-like model systems readily form oligomers or aggregates at micromolar and sub-micromolar concentrations,[8,15b,28] an important and unique advantage of smFRET measurements is that they were performed at sub-nanomolar concentrations thus mitigating the effect of aggregation and allowing for characterization of the intramolecular distances within monomeric

Httex1 as a function of polyQ length. We observed two main populations in the smFRET measurements for dual-labeled Httex1 constructs. These include a population with an S value of 0.4−0.5 and a population with no acceptor population with an S value of 1.0. Donor only populations can arise from dye photophysics and/or incomplete labeling.[16b] Despite rigorous disaggregation,[29] we occasionally observed a third population, even in the picomolar range. Our multi-parameter analysis yielding two-dimensional plots of E_{FRET} and S combined with a burst search algorithm[30] allowed us to separate species within this sub-population, which is most likely due to aggregated species (see Supporting Information, Figure S12). Such a population could emerge from the formation of oligomeric species in the stock solution prior to dilution and additional quenching of the donor Alexa 488. Upon resuspension of the protein in an acetic acid/acetonitrile solvent and serial dilutions into PBS we were able to minimize Httex1 aggregation and significantly reduced the presence of this tertiary population.

The mean smFRET efficiencies were plotted as a function of the amino acid spacing between the donor and acceptor fluorophores for unmodified Httex1 constructs (Figure 2c) and pT3-modified Httex1 constructs (Figure 2d). We observed a consistent trend with regard to $\langle E_{FRET} \rangle$ values versus polyQ lengths. For a particular polyQ length, as the sequence spacing between donor and acceptor FRET pairs increased, the $\langle E_{FRET} \rangle$ became consistently smaller. This trend is preserved upon the introduction of the pT3 modification and we observed a further decrease in the measured $\langle E_{FRET} \rangle$ values from that of the unmodified Httex1. Additionally, for a given dye pair, $\langle E_{FRET} \rangle$ decreased with increasing polyQ length.

All-Atom Simulations Are Used To Convert Mean FRET Efficiencies to Inferences Regarding Httex1 Conformations. For IDPs, the general method to convert measured mean E_{FRET} values to estimates of inter-dye distances, r, requires the assumption of a functional form for the inter-dye distance distribution $P(r)$.[23,31] We do not have *a priori* knowledge of the functional form for $P(r)$ that is applicable for converting $\langle E_{FRET} \rangle$ to estimates of inter-dye distances. Typically, one uses distributions from the Gaussian chain, worm-like chain, or Flory−Fisk models.[32] In the Gaussian chain model $P(r)$ is parametrized in terms of the inter-dye distance r, the number of peptide bonds (n) between dyes, the distance $l = 0.38$ nm between consecutive C_α atoms, and l_p, a free parameter that measures chain stiffness.[33] However, the assumption of a canonical distance distribution function for $P(r)$ is only applicable if we know that the sequence adopts uniformly expanded or compact conformations.[32a] Such models are inapplicable for sequences that are chimeras of distinct types of conformations.[34] Httex1 is likely to fall in this chimeric class of IDPs, as it is composed of a polyQ region that has been previously shown to adopt compact conformations and a semiflexible PR domain with two rod-like polyproline segments.[35] Given that the Gaussian chain model only depends on the effective chain stiffness, quantified by l_p, we can determine l_p for a specific dye pair and use it to extract $\langle E_{FRET} \rangle$ values for the remaining dye pairs. Thus, if the relative error between the measured and calculated $\langle E_{FRET} \rangle$ values were large, then it would suggest that a uniformly scaling model would not describe the protein.[34]

Figure 3 compares the relative errors associated with the calculated $\langle E_{FRET} \rangle$ values extracted using l_p obtained from numerical fits of the Gaussian chain model to the measured A60C $\langle E_{FRET} \rangle$ values. We find that the relative error in the
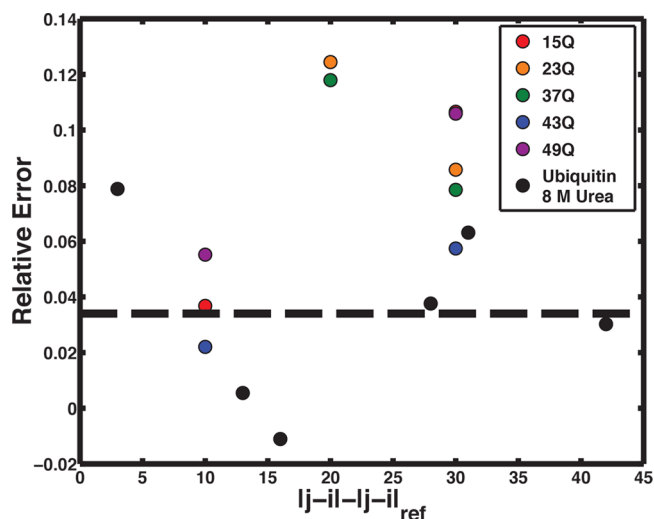

Figure 3. Test of the validity of using the Gaussian chain model to extract distances from FRET efficiencies for Httex1 as compared to denatured ubiquitin.[33] Relative error, $(\langle E_{FRET}^{calc} \rangle - \langle E_{FRET}^{meas} \rangle)/\langle E_{FRET}^{meas} \rangle$, between the measured ($\langle E_{FRET}^{meas} \rangle$) and calculated ($\langle E_{FRET}^{calc} \rangle$) FRET efficiencies as a function of ($|j - i| - |j - i|_{ref}$). Here, j is the position of the C-terminal dye and i is the position of the N-terminal dye. $|j - i|_{ref}$ denotes the number of peptide bonds between dyes for the dye pair used to calculate l_p. For Httex1 constructs, l_p was determined by fitting the measured A60C $\langle E_{FRET} \rangle$ values using the Gaussian chain model. The calculated $\langle E_{FRET} \rangle$ values were determined for the other dye pairs by inserting l_p into the equation for $P(r)$. As a control, the relative error between measured and calculated $\langle E_{FRET} \rangle$ values was calculated for ubiquitin in 8 M urea using $\langle E_{FRET} \rangle$ values from Aznauryan et al.[33] Ubiquitin in 8 M urea (black circles) should follow uniform scaling and thus Gaussian chain models should reasonably approximate the underlying distance distributions for denatured ubiquitin. Here, the K48C-R74C construct was used as the reference construct to calculate l_p. The dashed black line denotes the mean relative error for ubiquitin in 8 M urea.

calculated $\langle E_{FRET} \rangle$ values increases with increasing sequence separation of dyes. The relative errors are as high as 13%. Comparatively, a protein that is uniformly expanded shows a mean relative error that is typically less than 4%.[33] These results suggest that the scaling determined from the A60C dye pair underestimates the distance between dyes for the remaining dye pairs, and this underestimation increases for longer sequence separations. Such a result is consistent with the N-terminus being more compact than the C-terminus of Httex1, as would be expected if the polyQ domain adopts compact conformations, whereas the PR domain adopts expanded conformations. Overall, this analysis shows that a uniform scaling model cannot describe the conformational distributions of monomeric Httex1. Therefore, we combined experimental results with distance distributions extracted from atomistic simulations to obtain refined, atomic level descriptions of the monomeric ensembles of Httex1 as a function of polyQ length.

We performed all atom simulations of Httex1 constructs using the ABSINTH implicit solvation model and force field paradigm.[14] These simulations were performed using unlabeled molecules. However, the smFRET experiments report $\langle E_{FRET} \rangle$ values calculated for constructs comparing efficiencies between an N-terminal donor and a C-terminal acceptor either proximal to the polyQ domain, within the PR domain, or at the C-terminal residue. In order to compare the simulated ensembles to the experimental results, we had to account for the presence
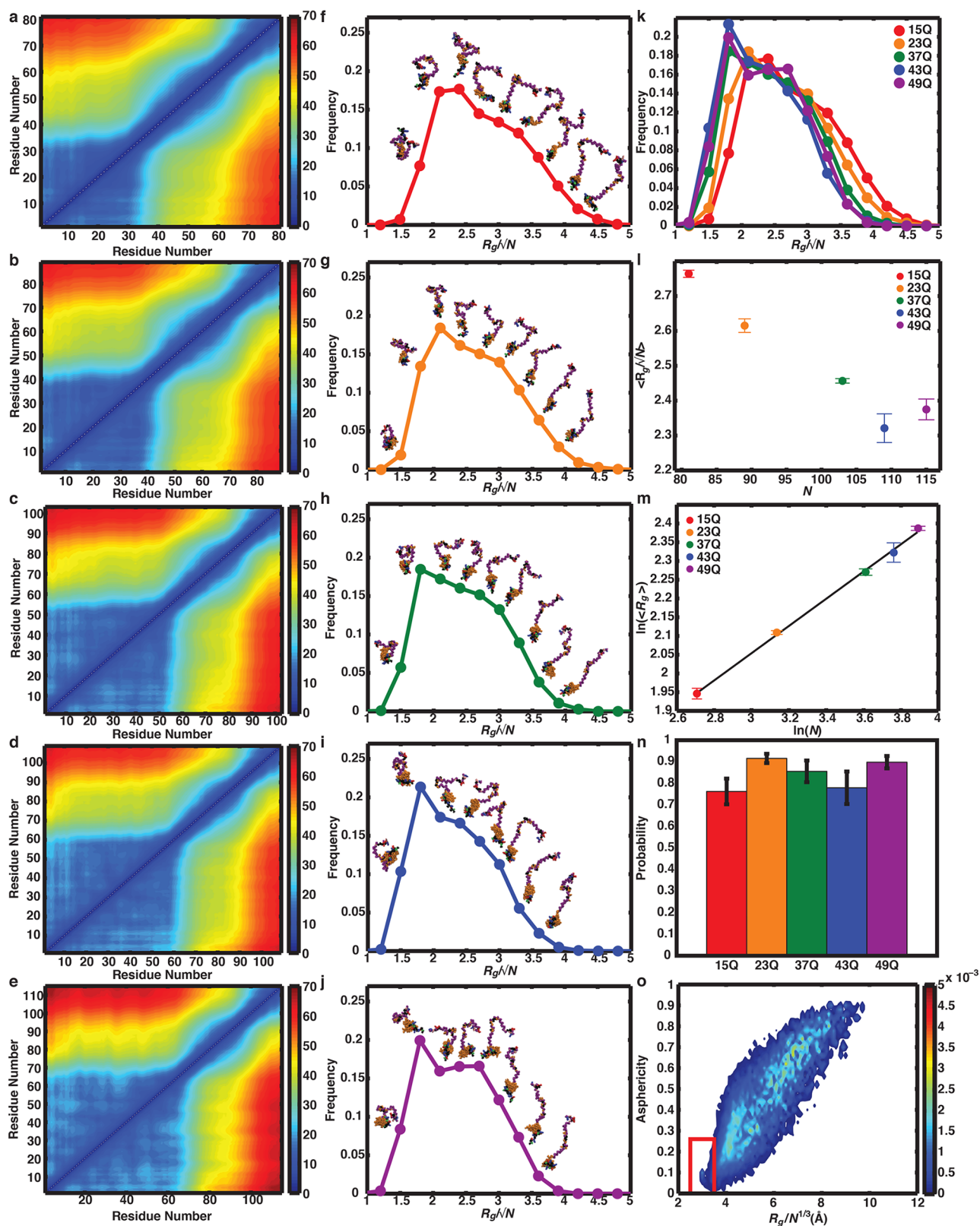
Figure 4. Conformational properties derived from simulated ensembles that match all three smFRET $\langle E_{\text{FRET}} \rangle$ values for a given polyQ length. (a−e) Distance maps quantify the average distance between all pairs of residues (in Å) for 15Q, 23Q, 37Q, 43Q, and 49Q, respectively. The hotter the color, the farther the average distance between a pair of residues. Tadpole-like architectures consisting of an Nt17-polyQ head and a PR domain tail are observed for all Httex1 constructs. (f−j) Normalized R_g distributions for the reweighted conformational ensembles of 15Q, 23Q, 37Q, 43Q, and 49Q, respectively. Here, R_g is normalized by $\sqrt{N}$, where N is the number of residues in the construct. Insets depict highly probable conformations that are consistent with a given $R_g/\sqrt{N}$ value. In these snapshots, glutamine is shown in orange, proline in purple, negatively charged residues in red, positively charged residues in blue, hydrophobic residues in black, non-glutamine polar residues in green, and glycine and histidine in pink. (k) Comparison of normalized R_g distributions for all polyQ lengths. As the polyQ length increases a continuous decrease in the distribution of $R_g/\sqrt{N}$

Figure 4. continued

values is observed. This is a result of the increased presence of a globular polyQ domain and is visually observed from the snapshots in panels f−j. (l) The average $R_g/\sqrt{N}$ as a function of polyQ length. Error bars denote the standard error of the mean calculated over three independent simulations. (m) Scaling of the mean size ($\langle R_g \rangle$) of the polyQ domain as a function of polyQ length. The line shows the best fit to the equation $\ln(\langle R_g \rangle) = \ln(\alpha) + \nu \ln(N)$. Here, $\nu = 0.36$ and $\alpha = 2.62$ Å. Error bars denote the standard error of the mean for three independent simulations. (n) Probability that Nt17-Q_n adopts globular conformations. The probability was calculated from two-dimensional histograms of $R_g/N^{1/3}$ and asphericity, δ. Specifically, the probability was calculated by summing the density within the two-dimensional region defined by 2.5 Å $\leq R_g/N^{1/3} < 3.5$ Å and $0 \leq \delta < 0.26$. The error bars correspond to the standard error of the mean over three independent simulations. (o) Two-dimensional histogram of $R_g/N^{1/3}$ and δ for the polyQ-PR domains of Httex1 49Q. The red rectangle corresponds to the region that corresponds to globular conformations as defined above. For all polyQ lengths the probability of polyQ-PR domains adopting globular conformations is negligible.

of the dyes and their influence on the simulated conformational ensembles. A reasonable, albeit minimalist, assumption is that dyes are fully accessible to the solvent.[34] If this were not the case, then the smFRET and fluorescence polarization anisotropy data would have revealed anomalies such as substantially hindered motions of dyes, which they do not. To account for the presence of fluorescent dyes in each of the three different positions for each polyQ length, we added dyes to the simulated ensembles in a post-processing step (see Methods for details).[34] We assume that solvation shells of radius 5 Å delineate the dyes, and inter-dye distances were calculated between the C19 atoms of Alexa 488 and Alexa 594.

Our goal was to extract atomic level descriptions of conformational ensembles that are concordant with all three experimental $\langle E_{FRET} \rangle$ values for each polyQ length. We achieved this using a maximum entropy reweighting method.[18] The procedure attempts to give all simulated conformations similar weights while minimizing the difference between the experimental and simulated observables. Here, for each polyQ length, the experimental observables were the $\langle E_{FRET} \rangle$ values for the three dye pairs. Using the experimental $\langle E_{FRET} \rangle$ values, rather than converting these mean efficiencies to mean distances, is advantageous because it limits the use of meta-data that depends on assumptions of the underlying experimental distribution.[34] To convert simulated inter-dye distances to FRET efficiencies, we deployed the Förster approximation for each conformation and generated a distribution of FRET efficiencies for each simulation and dye pair. Using the Förster formula, we calculated the conformation-specific FRET efficiency to be

$$E_{FRET} = \left[1 + \left(\frac{r}{R_0} \right)^6 \right]^{-1}$$

Here, r is the conformation and position specific distance between the dyes and R_0 is the Förster radius, which was set to $R_0 = 56$ Å (see methods in Supporting Information) and recent work.[34]

Generating Self-Consistent Conformational Ensembles for Httex1. We analyzed the conformational ensembles obtained at each of the distinct simulation temperatures. We quantified the agreement between calculated $\langle E_{FRET} \rangle$ values from each of the simulated ensembles and the experimentally measured $\langle E_{FRET} \rangle$ values. This procedure involved reweighting the conformations at each of the simulation temperatures to maximize the information theoretic entropy while minimizing the deviation between the calculated and measured $\langle E_{FRET} \rangle$ values. This procedure shows that the extent of reweighting is rather minimal, as quantified by the change in entropy upon reweighting (Supporting Information, Figure S5). These results suggest that the simulations generate ensembles of sufficient

accuracy for pursuing a detailed atomistic description of the conformational preferences of Httex1 as a function of polyQ length. We identified 320 K as the lowest simulation temperature for which the ensembles are most representative of the experimental data. For temperatures above 320 K, ensembles show optimal comparisons with the measured $\langle E_{FRET} \rangle$ values (see Supporting Information, Figure S5). This robustness was preserved for all polyQ lengths examined.

Httex1 Adopts Tadpole-like Conformations with a Globular Nt17-PolyQ Head and Semi-flexible Proline-Rich Tail. Figure 4 summarizes our analysis of various conformational features extracted from the reweighted conformational ensembles for Httex1 as a function of polyQ length. The results are shown for the ensembles obtained at 320 K because this is the lowest temperature that yields an entropy change corresponding to less than a kT change in the simulation energy function upon reweighting (see Methods and Supporting Information, Figure S5). The results presented here do not vary substantially across a broad temperature range spanning from 310 K − 335 K. We calculated the average distances between all pairs of residues from the reweighted ensembles. This provides a quantitative description of the conformational properties across Httex1 constructs as a function of polyQ length. Figure 4a−e shows the results of this analysis for all polyQ repeat lengths. The hotter the color the farther two residues are from each other. The defining features of these distance maps are as follows: (1) the general distance preferences are conserved across polyQ repeat lengths and dye positions; (2) the combination of Nt17 and polyQ domains adopt compact conformations as highlighted by small values for average distances between all pairs of residues within these domains; and (3) the PR domain predominantly adopts extended conformations, although there is a minor, temperature-dependent population characterized mainly by contacts between the flexible linker between polyproline modules of the PR domain and the surface of the polyQ domain (Supporting Information, Figure S6). Overall, these features suggest that Httex1 constructs adopt tadpole-like conformations for all polyQ repeat lengths. The tadpole-like architecture is defined by a globular "head", consisting of Nt17 adsorbed to polyQ, and a semi-flexible "tail", which refers to the PR domain.

Conformational Properties of Httex1 Change Continuously with PolyQ Length. A prevailing hypothesis in the field is of an abrupt conformational change that accompanies an increase in polyQ repeat length beyond the threshold of 36−37 residues. To test whether the smFRET efficiencies are consistent with an abrupt change in conformational properties, we quantified the distributions of normalized radii of gyration for each polyQ repeat length. In order to put all Httex1 constructs on the same scale, the radius of gyration (R_g) distributions, which quantify the size of the conformations in

the simulated ensemble, are normalized by $N^{0.5}$. Here, N is the number of residues in the construct. The results from our analysis are shown in Figure 4f–l. These distributions suggest that Httex1 undergoes a continuous global contraction as the polyQ repeat length increases. This contraction arises from the increased prominence of the globular polyQ domain as the polyQ length increases.

Fluorescence correlation spectroscopy experiments on Gly-$(Gln)_N$-Cys*-Lys$_2$ show that polyQ adopts collapsed conformations.[35a] We asked if this feature is preserved in the context of the native Httex1 constructs. For a uniform collapsed polymer, the ensemble-averaged R_g scales with chain length, N, according to $\langle R_g \rangle = \alpha N^{1/3}$, where $\alpha \approx 3.0$ Å. Figure 4m shows the results of the least-squares regression analysis for $\ln(N)$ versus $\ln(\langle R_g \rangle)$. The parameters for the slope and intercept, obtained from the regression analysis, are found to be 0.36 and 2.62 Å, respectively. This implies that the polyQ domain maintains its intrinsic preference for globular conformations in the context of Httex1. These globular conformations are likely to be more stable as polyQ length increases because the surface-to-volume ratio decreases as $N^{-1/3}$ as N increases. Importantly, unlike recent simulation results, we do not observe a compaction that lead to values below the canonical exponent of 1/3 or any increases in β-sheet contents that were recently reported for polyQ lengths above the pathological threshold.[19] Inasmuch as our simulation results are concordant with and vetted by experimental data, it appears that abrupt conformational transitions are likely to be low likelihood fluctuations that may or may not be enhanced by intermolecular interactions.[36] However, such low likelihood fluctuations are not the defining intrinsic features of the polyQ-length-dependent conformational properties of monomeric Httex1 constructs and are likely to be discernible through the use of biased sampling methods that mimic the effects of intermolecular interactions.[20a]

Previous studies have suggested that Nt17 undergoes a polyQ-mediated expansion that coincides with an adsorption to the polyQ domain for polyQ lengths of greater than ~20.[12a,37] In order to test whether Nt17 adsorbs on the polyQ domain within Httex1 constructs, we constructed two-dimensional histograms of $R_g/N^{1/3}$ and asphericity, δ, calculated over Nt17-Q_n. Here,

$$\delta = 1 - 3 \frac{(\lambda_1 \lambda_2 + \lambda_2 \lambda_3 + \lambda_1 \lambda_3)}{(\lambda_1 + \lambda_2 + \lambda_3)^2}$$

Here, λ_1, λ_2, and λ_3 are the eigenvalues of the conformation-specific gyration tensor.[12c,38] When $\delta \leq 0.25$, conformations are spherical (globular), whereas $\delta \rightarrow 1$ corresponds to rod-like conformations. For globular conformations, which should be observed if Nt17 adsorbs on the polyQ domain, $R_g/N^{1/3}$ should be approximately 3.0 Å. Summation over the density within the two-dimensional region of $2.5 \text{ Å} \leq R_g/N^{1/3} < 3.5 \text{ Å}$ and $0 \leq \delta < 0.26$ quantifies the probability that Nt17 is adsorbed on the polyQ domain. As shown in Figure 4n, all polyQ lengths lead to high degree of adsorption between Nt17 and polyQ with greater than 70% of the conformations being globular for monomeric Httex1 in the absence of oligomerization. We performed a similar analysis over polyQ-PR domains and found that for all polyQ lengths and dye pairs a negligible percentage of the conformations were observed to be globular. An example, two-dimensional histogram of Httex1 49Q is shown in Figure 4o. Most of the density was observed outside the region that corresponds to globular conformations.

DISCUSSION

The prevailing hypothesis in the HD field is that sharp changes in conformational properties of monomeric Httex1 accompany the increase in polyQ length beyond the pathological threshold. We tested this hypothesis using data from our smFRET measurements and computational analysis. Our integrative approach yielded the following insights: Httex1 constructs adopt tadpole-like architectures defined by a globular head comprised of Nt17 adsorbed on the polyQ domain and an extended tail comprised of the PR domain. These results do not support a sharp, polyQ-length-dependent structural change within monomeric Httex1. Instead, they support a continuous global compaction with increasing polyQ length that arises due to the increased prominence of the compact polyQ domain. The cellular concentrations and sub-cellular localization of Httex1 are unknown and need to be measured precisely. Estimates in the literature place the cellular concentrations of Httex1 to be in the nanomolar or sub-nanomolar regime.[39] The sub-nanomolar concentrations used in our experiments are significantly below the critical concentration thresholds that promote aggregation and phase separation *in vitro*[8] and the solubility limits in cells where deleterious phenotypes such as the impairment of proteostasis networks are manifest.[40] By working at sub-nanomolar concentrations, we were able to decouple evidence of a sharp conformational change at the monomer level as the sole reason for the polyQ-length-dependent toxicity threshold observed in HD. Instead we propose that the continuous increase in the surface area of the polyQ domain with increasing polyQ length leads to changes in homotypic and heterotypic interactions and these changes engender the polyQ-length-dependent threshold observed in HD. By studying the conformational properties of monomeric forms of Httex1 we were able to establish that sharp changes observed in Httex1 aggregation or interaction networks as a function of polyQ length are not a result of sharp, polyQ-length-dependent conformational changes at the monomer level. These results imply that in order to understand the polyQ-length-dependent toxicity observed in HD, future biophysical studies should focus on understanding differences in higher order interactions and conformational transitions mediated by intermolecular interactions as a function of polyQ length. This will require a combination of biased sampling methods to construct the appropriate free energy surfaces impacted by conformational fluctuations as well as advanced experimental methods that probe conformations and fluctuations influenced by the interplay between intra- and intermolecular interactions.

A subset of studies, based on the binding of antibodies, also argues against a "structural toxic threshold" model.[41] A continuous increase in binding was observed for two polyQ-targeting antibodies as a function of polyQ length. These results suggest that there is a monotonic increase in the number of surface epitopes rather than a sharp structural change within the Httex1 fusion proteins as the polyQ length increases. Other studies have suggested that the polyQ domain undergoes an increased rigidity transition only above the pathogenic polyQ length threshold.[7] These inferences were based on fluorescence lifetime imaging microscopy FRET experiments conducted in live cells on Httex1 fluorescent protein fusion constructs. As noted earlier, tagged systems generate confounding observations with considerable variability, depending on the tags that are used.
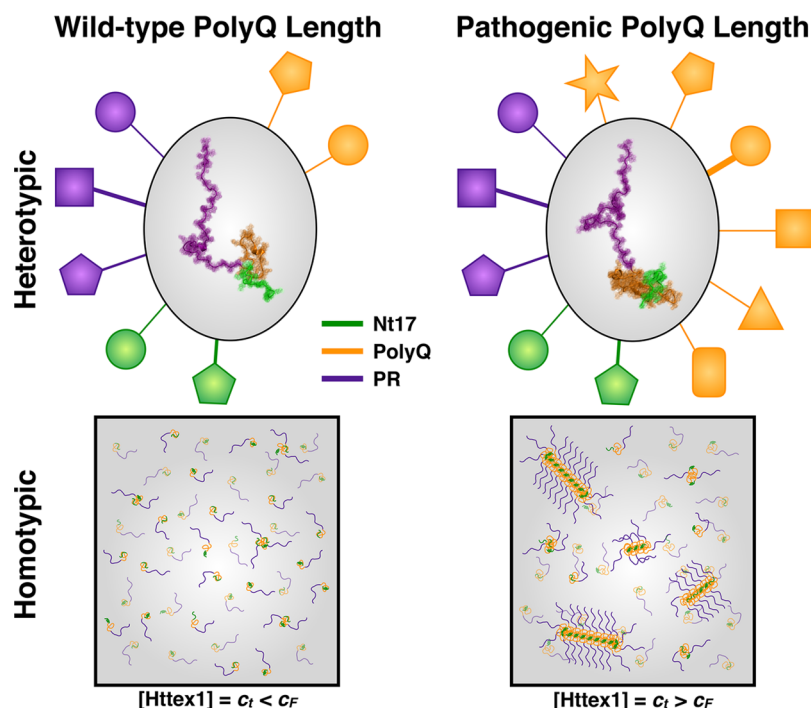
Figure 5. Proposed influences of tadpole-like monomeric Httex1. The top row shows the proposed impact of monomeric Httex1 on heterotypic interactions. Green, orange, and purple symbols and edges depict interactions of monomeric Httex1 through Nt17, polyQ, and the PR domain, respectively. As polyQ length increase, we propose that the number and strengths of heterotypic interactions can increase, vis-à-vis the wild-type, due to the increased prominence of the globular polyQ domain in the tadpole-like architecture of monomeric Httex1. The bottom row shows the proposed impact of polyQ length on homotypic interactions that drive the aggregation and phase separation of Httex1. The total cellular concentration of Httex1 is denoted as c_t. For the wild-type, we propose that $c_t < c_F$, where c_F is the saturation concentration that has to be crossed to drive the formation of insoluble, fibrillar aggregates.[8] Conversely, polyQ expansions lead to a reversal whereby $c_t > c_F$, and hence, depending on the gap between c_F and c_t, there is an increasing driving force for forming large fibrillar aggregates. The tadpole-like structures of monomeric Httex1 determine the overall bottlebrush architecture of the aggregates,[11a] whereas nucleated conformational changes within Httex1 determine the intermolecular interfaces and the strengths of aggregates, including fibrils.[11a,51]

Implications of the Structures of Monomeric Httex1 for Heterotypic Interactions. The surface area of the polyQ globules increases as $N^{2/3}$ with polyQ length N. This increase in polyQ surface area with N should increase the number of surface accessible polyQ sites and enable the emergence of new interactions with proteins in the cellular milieu (Figure 5).[42] Such heterotypic interactions might give rise to sharp changes in cellular phenotypes that influence protein quality control, toxicity, and cell death.[42,43] Even though Htt is ubiquitously expressed, medium spiny and striatal neurons are most susceptible to neurotoxicity and degeneration.[44] This suggests that the growing prominence of the polyQ domain within the tadpole-like structure of monomeric Httex1 might elicit toxic, gain-of-function interactions in specific neuronal sub-types, thus giving rise to the appearance of a sharp pathological transition as a function of polyQ length.[42] The key question is if the tadpole-like architecture is sufficient to engender sharp, polyQ length-dependent gain-of-function heterotypic interactions within cells. A recent study provides preliminary support for this hypothesis, showing that the network of protein–protein interactions, with Httex1 at the hub, changes sharply with polyQ length.[42] Importantly, these findings, in cells, were explained using the central tenets of the tadpole-like architecture presented in this work.

The large variability in ages of onset for a given pathogenic polyQ length suggests that additional factors, including gain of function heterotypic interactions and higher-order homotypic interactions, may be the determinants of HD progression. This idea is consistent with studies that suggest overexpression of proteins housing Q-rich regions bind mutant Htt and suppress cellular toxicity in yeast.[45] This suppression was proposed to be a result of blocking the interactions between more essential proteins and mutant Htt. Wear et al.,[46] showed that there was an enrichment in proteins housing long intrinsically disordered regions (IDRs) associated with mutant Htt aggregates. For two representative binding partners, this interaction was dependent on the presence of the IDR, which may suggest the IDR engages in preferential interactions with expanded polyQ domains. Finally, protein quality-control machineries in striatal neurons, as opposed to cortical neurons, are impaired in response to the expression of mutant Httex1.[44] This suggests that monomeric or soluble forms Httex1 with expanded polyQ tracts might engage deleteriously, albeit in cell-specific ways, with components of the protein quality-control machineries such as the ubiquitin proteasome system and autophagy.

Implications for the Driving Forces for and Mechanisms of Httex1 Aggregation. Inferences from previous *in vitro* studies of aggregation kinetics suggest that the rate of nucleation of β-sheet-rich conformations should increase with increasing polyQ length.[37,47] nificantly high peptide concentrations.[15b,48] Analysis of Httex1 fibril structure by solid-state NMR showed that Httex1 fibrils adopt a β-hairpin-based polyQ core structure, which requires a minimum of 22 glutamine residues.[49] Our smFRET measurements cannot rule out the possibility of increased β-sheet content within the collapsed polyQ domains of monomeric Httex1.[20] However, our

simulation results show negligible secondary structure preferences within polyQ domains, irrespective of polyQ length. The implication is that β-sheet formation is likely to be a rare event that is confronted by high free energy barriers.[20a] The most likely scenario is that the tadpole-like architecture drives the spontaneous formation of lower molecular weight aggregates such as bristled spheres that are characterized by sequestration of the polyQ domains on the interior of the spheres and exposure of the PR domain tails to solvent. Nucleated conformational conversion within these bristled spheres[8,35b] likely promotes the templated formation of high-molecular-weight β-sheet-rich fibrils that have bottlebrush architectures stabilized by polyQ cores and PR domains forming the bristles of the brush.[11a]

From the standpoint of aggregation, the biophysical basis for the pathological length threshold may well be the lowering of the saturation concentrations for forming bristled spheres and bottlebrush fibrils as the polyQ length increases.[8] The tendency for polyQ peptides to form collapsed conformations is consistent with previous results showing that water is a poor solvent for polyQ thus explaining the poor solubility of polyQ peptides in aqueous solutions.[12c,35a] This connection between monomeric collapse and solubility was also observed experimentally by Walters et al., who showed that polyQ peptides that underwent monomeric collapse readily formed soluble aggregates.[35b] Increasing the polyQ length leads to more unfavorable interactions between the surface of the collapsed monomer and the surrounding solvent. The driving force of Httex1 aggregation arises from increased intermolecular interactions that minimize glutamine interactions with the surrounding solvent rather than a sharp polyQ- induced structural rearrangement. Furthermore, this suggests that at physiological concentrations longer polyQ domains will have a more pronounced tendency to aggregate, and this could contribute to the pathogenic polyQ length threshold observed in HD[50] (Figure 5). If the physiological concentration of Httex1 were designated as c_t, then only proteins containing polyQ lengths greater than the pathogenic threshold have a saturation concentration for aggregation greater than c_t. This would lead to a sharp polyQ length threshold for the formation of a heterogeneous set of, potentially toxic, aggregates that are not observed for wild-type polyQ lengths. To test this hypothesis, we need accurate measurements of the physiological concentrations of Httex1, as well as the saturation concentrations for aggregation as a function of polyQ length.

Impact of N- and C-Terminal Flanking Sequences of polyQ. The aggregation of Httex1 is dependent not only on polyQ repeat length but also on the presence of Nt17 and PR domains. For a given polyQ length, the presence of Nt17 increases the drive to form large, linear, insoluble aggregates and decreases the solubility of Httex1 constructs.[8,28,37,52] Two models have been proposed in order to describe how Nt17 modulates Httex1 aggregation.[53] In the proximity model, Nt17 drives Httex1 aggregation by increasing the effective local concentration of polyQ through Nt17-dependent helical bundling.[28,37,54] Sahoo et al., showed that Httex1-like constructs with expanded polyQ tracts readily form tetramers by fluorescent correlation spectroscopy. However, when the Nt17 domain is replaced by di-lysine, only monomers are observed. Although these results do suggest that Nt17 is important for modulating Httex1 aggregation, the lack of oligomerization may be a result of the addition of the di-lysine rather than just the removal of the Nt17 domain. Such a result

is consistent with previous studies, which show that the addition of Lys$_n$ ($n = 1-8$) flanking the polyQ domain can modulate both the degree of collapse within the polyQ domain and the solubility of Httex1-like constructs.[8,35b]

In the domain cross-talk model, Nt17 and polyQ interdomain interactions control the specificity and stability of intermolecular interactions.[53,55] This model suggests that the length of the polyQ domain is the main driver of Httex1 aggregation and Nt17 enhances the formation of linear as opposed to spherical aggregates by providing a surface-adsorbed amphipathic "patch" on polyQ that promotes the formation of linear aggregates by diluting the contacts that lead to spherical aggregates.[8,55] This suggests that as the polyQ repeat length is increased, Nt17, which then makes up a smaller portion of Httex1, should be less effective at modulating polyQ-dependent aggregation. This is supported experimentally by the observation of a decrease in fibril formation rates upon removal of the Nt17 domain.[8,52] Given that the degree of Nt17 adsorption is modulated by interactions between the polyQ domain and uncharged residues of Nt17, increasing the charge within Nt17 is likely to reduce the degree of adsorption between the Nt17 and polyQ domains.[55]

Recent studies have shown that phosphorylating T3, S13, or S16 in Nt17 reduces the driving forces for forming insoluble aggregates.[17,56] However, whereas the phosphorylation of T3 stabilizes Nt17 helicity in isolation, phosphorylation of S13 and S16 destabilizes Nt17 helicity.[25] Together these results suggest that the degree of cross talk between Nt17 and polyQ and/or the charge within Nt17, rather than the degree of intrinsic helicity within Nt17, is likely to be more important as a modulator of aggregation mechanisms. We hypothesize that phosphorylation of Nt17 residues reduces Httex1 aggregation by (1) increasing the charge within Nt17, which may increase the solubility of Httex1 constructs strictly through a charge effect, as well as reduce intermolecular Nt17 interactions and/or (2) reducing the degree of adsorption of Nt17 on the polyQ domain which modulates the types of aggregates that form, as well as reduces the stability that can be gained from intermolecular interactions between the Nt17 and polyQ domains.[55]

In contrast to the effect of Nt17 on Httex1 aggregation, the PR domain, as well as C-terminally truncated versions of the PR domain, increases the solubility and reduces the drive to form fibrils of polyQ-containing constructs.[8,10,57] Our smFRET measurements are consistent with the PR domain being an extended, semi-flexible chain and engaging in relatively few contacts with the Nt17 or polyQ domains. The small degree of conformational coupling between the PR domain and the Nt17 and polyQ domains, when compared to the coupling between the Nt17 and polyQ domains, may explain why the PR domain helps to solubilize polyQ-containing constructs whereas the Nt17 domain decreases the solubility of the same constructs. Beyond the intrinsic solubility of the PR domain, limited interactions with the Nt17 and polyQ domain engender conformations in which an excluded volume tail can restrict the ways in which molecules can come together and may further increase the solubility of polyQ-containing constructs. Furthermore, coarse-grained simulations on Httex1-like constructs suggest that flanking regions that show coil-like properties and limited coupling with the polyQ domain preferentially form spherical aggregates which may kinetically hinder the formation of large, linear aggregates.[55]

■ SUMMARY

The integrative approach deployed here has allowed us to obtain a detailed description of the monomeric forms of Httex1. We propose that as the polyQ length increases, the increased prominence of polyQ domain leads to increased unfavorable interactions with the surrounding solvent. This, in turn, should lead to an increased drive to form higher order homotypic and/or heterotypic interactions through the polyQ domain. As the formation of intermolecular contacts and higher order oligomeric species appears to be at the crux of HD pathophysiology, it will be crucial to isolate and characterize these higher molecular weight species. Furthermore, identification of binding partners that promote or stabilize non-toxic oligomeric Httex1 species will enable advances in understanding the relationship between Httex1 phase behavior and HD pathophysiology.

■ METHODS

Expression and Purification of Httex1 A2C N-Terminal Thiazolidine Double Cysteine Constructs. Expression and purification was performed with modifications as described by Vieweg et al.[15b] Chemo-competent *E. coli* ER2566 cells (NEB) were transformed with resulting vectors pTWIN1-His$_6$-Ssp-Httex1-Q$_N$-A2C-A60/P70/P80/P90C. Isolated single colonies were inoculated in 500 mL lysogeny broth (LB) (100 μg/mL ampicillin) at 37 °C overnight with 180 rpm shaking. The following morning, 12 L of LB (100 μg/mL ampicillin) were mixed with the overnight culture to obtain an OD$_{600}$ of 0.05. Cells were grown at 37 °C until an OD$_{600}$ of 0.1 was reached, the temperature of the incubator was then set to 14 °C. Protein induction was initiated at OD$_{600}$ of 0.3 with 0.4 mM IPTG overnight. Cells were harvested by centrifugation (4 °C, 6238g, 8 min) and cell pellets were kept on ice. Cell pellets were resuspended in 50 mL buffer A (50 mM HEPES, 0.5 M NaCl, pH 8.5) containing 0.3 mM PMSF and 1x CLAP. Cells were lysed on ice by sonication (6 min, pulse on 30s, pulse off 59 s, 70% amplitude) using a vibra cell VCX130 from Sonics. The lysate was cleared by centrifugation (30 min, 4 °C, 27216g). The cleared supernatant was then filtered through 0.45 μm syringe filter membranes and applied to a 5 mL Histrap column (GE Healthcare, 17-5248-02) at a flow rate of 1 mL/min. The column was then washed with 10 column volumes (CV) of buffer B (50 mM HEPES, 0.5 M NaCl) to remove non-specifically bound proteins. The column was then washed with 3 CV of 5% buffer C (50 mM HEPES, 0.5 M NaCl, 0.5 M imidazole). Fusion proteins were then eluted off the Histrap column using a gradient from 4 to 50% buffer C over 50 mL. Elution fractions were analyzed by SDS-PAGE and pooled for splicing. Splicing and in situ N-terminal thiazolidine formation was initiated by addition of 1 mM TCEP and 1 mM formaldehyde and adjusting the pH to 6.8. Splicing was carried out at room temperature (RT) and monitored by SDS-PAGE and analytical C8 reversed-phase ultra-high-performance liquid chromatography (RP-UHPLC). For polyQ repeat lengths >37Q protein was allowed to splice for a maximum of 4 h while for polyQ repeat lengths ≤37Q splicing was performed for 12−16 h. Following splicing, samples were filtered through 0.45 μm syringe filter membranes and injected into a preparative C4 reversed-phase high performance liquid chromatography (RP-HPLC) column (00G-4168-P0-AX, Jupiter C4, 10 μm, 300 Å, 21.2 mm i.d. × 250 mm length) pre-equilibrated with 95% buffer D (water with 0.1% trifluoroacetic acid (TFA)) and 5% buffer E (acetonitrile with 0.1% v/v TFA). Spliced Httex1 constructs were eluted using a gradient of 30−40% buffer E over 40 min. Collected fractions were analyzed by liquid chromatography mass spectrometry (LCMS) using a Thermo Scientific LTQ ion trap mass spectrometer and pooled accordingly for lyophilization. Purity of lyophilized protein was assessed by LCMS using a C3 poroshell 300SB 1.0 × 75 mm, 5 μm column from Agilent (method: 5−95%ACN in 5 min, flow rate of 0.3 mL/min, injection volume of 10 μL). LCMS spectra were deconvoluted with MagTran software v. 1.03b from Amgen.

Httex1 Double Labeling. Purified N-terminally thiazolidine protected Httex1 constructs with C-terminal cysteine residues, 2.0 mg, were disaggregated using trifluoroacetic acid/hexafluoro-isopropanol (TFA/HFIP) (1:1 v/v) as described by O'Nuallain et al.[29a] Protein was resuspended on ice in 1.0 mL labeling buffer (100 mM Tris pH 7.4, 6 M guanidinium HCl (GdHCl)) for constructs with polyQ ≤ 37Q and mutant labeling buffer (100 mM Tris pH 7.4, 6 M GdHCl, 50 mM trehalose, 0.5 M proline) for constructs with polyQ > 37Q. The pH was quickly adjusted to 7.4 as needed followed by the addition of 1.5 equiv of Alexa594-maleimide and incubated on ice for 15 min. The reaction was monitored by LCMS as previously described. Excess Alexa594-maleimide was removed using a PD-10 desalting column equilibrated with thiazolidine deprotection buffer (5% acetic acid, 5% acetonitrile in water with 0.1% TFA). The protein was then diluted to 5.0 mL with thiazolidine deprotection buffer and kept on ice. Thiazolidine deprotection was initiated by addition of 100 equiv of silver triflate for 15−30 min on ice. The reaction was monitored by LCMS. Upon completion of N-terminal deprotection, the reaction was flash frozen in liquid nitrogen and solvent was removed by lyophilization. The protein was resuspended in labeling buffer with the addition of 10 mM TCEP and incubated on ice for 30 min and then precipitated by addition of 14 mL of cold ethanol and stored at −80 °C overnight. Protein was then pelleted by centrifugation (30 min, 4 °C, 5251g), and the supernatant was discarded. The pellet was washed with 5 mL of cold ethanol and collected by centrifugation. Trace solvent was then removed by lyophilization for 1−2 h. Protein was then disaggregated and excess silver was removed by resuspension in TFA/HFIP (1:1 v/v). Insoluble silver salts were removed by centrifugation and the supernatant was carefully removed. The pellet was washed twice with TFA/HFIP (1:1 v/v) and supernatants were combined and dried under a stream of nitrogen. Trace solvent was removed by lyophilization for 1−2 h. Protein was then resuspended in 1.0 mL of either labeling or mutant labeling buffer and labeled with Alexa488-maleimide as described previously for Alexa594-maleimide. Following labeling, excess Alexa488-maleimide was removed by ethanol precipitation as previously described. Prior to final HPLC purification, protein was disaggregated with neat TFA containing a catalytic amount of ammonium iodide to reduce any possible methionine oxidation as described by Christian et al.[58] Following evaporation of TFA, trace solvent was removed by lyophilization. Protein was resuspended in 1.0 mL of either labeling or mutant labeling buffer and directly injected onto a Jupiter 5 μm C4 300 Å 250 × 4.6 mm or Jupiter 5 μm C4 300 Å 250 × 10 mm column. Protein was eluted with a 25−55% gradient of buffer E over 50 min. Collected fractions were analyzed by LCMS for purity and pooled accordingly. Final purity of the doubly labeled protein was performed by SDS-PAGE, C8 UPLC, and LCMS (see Supporting Information, including Figures S1−S4).

Expression and Purification of Htt18−90 Q18C N-Terminal Thiazolidine Double-Cysteine Constructs. Expression and purification was performed as previously described for full length Httex1 constructs. Spliced Htt18−90 constructs were eluted using a gradient of 20−55% buffer E over 50 min. Collected fractions were analyzed by LCMS and pooled accordingly. Final purity was assessed by C8 UPLC.

Semi-synthesis of Dual-Labeled and T3-Phosphorylated Httex1 Proteins. Htt18−90 Q18Thz double cysteine fragments were labeled with Alexa594-maleimide and subsequently N-terminally deprotected as described previously. Following N-terminal deprotection, native chemical ligation (NCL) was performed as described by Chiki et al.[25] Briefly, labeled protein was dissolved in 800 μL of labeling or mutant labeling buffer containing 100 mM TCEP and 50 mM methoxyamine. The pH was increased to ~4.0 and incubated at room temperature for 30 min. Following brief methoxyamine treatment, 50 mM MPAA was added and the pH of the reaction was adjusted to 6.9. NCL was initiated by addition of 3 equiv of Nbz-thioester peptide. The reaction was incubated at room temperature and monitored by LCMS. Upon completion, protein was purified using a C4 semi-prep HPLC with a linear gradient of 10−45% buffer E over 50 min. Fractions were collected and purity was assayed by

LCMS and pooled accordingly. Protein was then dried by lyophilization. Disaggregation was performed using TFA/HFIP (1:1 v/v) as described previously. Iodoacetamide treatment was then performed to mask Q18C as a pseudo-glutamine by resuspending protein in 1.0 mL labeling or mutant labeling buffer and 1 mM freshly prepared iodoacetamide and 1 mM TCEP. The protein was incubated at room temperature for 15 min or on ice for 30 min. The reaction was monitored by LCMS and upon completion desalted into thiazolidine deprotection buffer as described previously. Following, protein was N-terminally deprotected, labeled with Alexa488-maleimide, and HPLC purified as described previously. Final protein purity was characterized by SDS-PAGE, C8 UPLC, and LCMS (see Supporting Information, Figure S8).

Single Fluorophore Httex1 Labeling. Httex1 A2C 15−49Q were expressed and purified as previously described but without the addition of formaldehyde during Intein-Ssp splicing. Protein was labeled with Alexa488-maleimide, donor fluorophore, as described above and excess fluorophore was removed by ethanol precipitation. Donor-only constructs were HPLC purified as previously described. Httex1 A2Thz P90C 15−49Q were prepared as described above. Protein was labeled with Alexa594-maleimide, acceptor fluorophore, as previously described and excess fluorophore was removed by ethanol precipitation. Acceptor only constructs were HPLC purified as previously described. Final protein purity was characterized by SDS-PAGE, C8 UPLC, and LCMS (see Supporting Information, Figures S1−S4).

smFRET Measurements. For all smFRET measurements 5−30 μg portions of protein were weighed out and disaggregated by TFA/HFIP (1:1 v/v) as previously described. Protein samples were resuspended at a target concentration of 1 μM in 20% acetonitrile and 20% acetic acid in water with 0.1% TFA. Aliquots were prepared, flash frozen, and stored at −80 °C. Dual-labeled protein samples were diluted to between 50 and 200 pM in Dubelco's PBS pH 7.4. Initial measurements were made on samples prior to freezing and replicates were collected from −80 °C stored samples. Data were collected using a custom-built multi-parameter single-molecule spectrometer analogous to that previously described.[59] Single-molecule bursts were identified using a burst search and a threshold of 80 photons was subsequently applied over the donor and acceptor channels.[30a] Leakage of donor fluorescence into the acceptor channel was corrected. Mean FRET efficiencies, $\langle E_{\text{FRET}} \rangle$, and stoichiometry, S, were calculated using a custom-written code in IgorPro (Wavemetrics, Lake Oswego, OR).

$$\langle E_{\text{FRET}} \rangle = \frac{I_A}{\gamma I_D + I_A}; \quad S = \frac{\gamma I_D + I_A}{\gamma I_D + I_A + I_A^{\text{dir}}}$$

Measurements were performed in triplicate for all constructs. $\langle E_{\text{FRET}} \rangle$ was calculated as intensity-based FRET efficiencies, where I_D and I_A are donor and acceptor intensities respectively, and γ is a correction factor dependent upon Httex1 donor (Φ_D) and acceptor (Φ_A) quantum yields and the detection efficiencies of the donor channel (η_D) and acceptor channel (η_A) (see Supporting Information, including Figures S9−S11). I_A^{dir} is the intensity from directly excited acceptor molecules using pulse-interleaved excitation with an orange laser.[27b,31] The two-parameter histograms shown in Figure 2a are highly reproducible from one run to the next, and this derives, in part, from the purity of the samples.

Details of All-Atom Simulations. All-atom simulations of Httex1 constructs were performed with the CAMPARI simulation package (http://campari.sourceforge.net) utilizing the ABSINTH implicit solvation model and force field paradigm.[14,60] Simulations were based on the abs3.2_opls.prm parameter set and were combined with temperature replica exchange in order to enhance sampling. The temperature schedule used was $T = [288, 293, 298, 305, 310, 315, 320, 325, 335, 345, 360, 375, 390, 405]$ K. A total of 6.15×10^7 steps were performed for each simulation. Here, a step refers to either a temperature replica exchange swap or a Metropolis Monte Carlo move. The first 10^7 steps were taken as equilibration steps. Observables were collected every 5×10^3 steps during the last 5.15

$\times 10^7$ steps of the simulation to use for further analysis. Simulations were performed in droplets with radii of 150 Å. This radius choice was chosen to ensure against confinement artifacts that arise due to too small a droplet. Excess and neutralizing Na$^+$ and Cl$^-$ ions were modeled explicitly with an excess NaCl concentration of 5 mM. The specific sequences used were ATLEKLMKAFESLKSF-Q$_n$-P$_{11}$-QLPQ-PPPQAQPLLPQPQ-P$_{10}$-GPAVAEEPLHRP, where n = 15, 23, 37, 43, and 49. The N- and C-termini were left uncapped for consistency with the experimental constructs. The sequences were simulated without the double cysteine residues used for dye labeling. Dyes were added to the simulated ensembles post facto as described below.

Addition of Dyes to Simulated Ensembles. In order to add dyes post facto, our in-house program COCOFRET was used. For each dye pair and polyQ length combination, COCOFRET utilizes the atomistic simulation trajectories, dye rotamer libraries, residue positions at which to add the dyes, and the Förster radius, R_0, in order to determine the mean FRET efficiency for each conformation that is consistent with the inclusion of the dye pair. Explicitly, for each conformation, we attempted 100 independent attachments of the Alexa 488 dye at position 2 and the Alexa 594 dye at one of the three C-terminal dye positions. Dye rotamers were randomly chosen from the HandyFRET rotamer libraries and dyes were attached to the protein such that the carbon−sulfur−carbon angle was approximately ideal.[61] A protein + dye conformation was accepted if no steric clashes were observed between the protein and the dye. A steric clash was defined as any protein atom being within the solvation shell of any dye atom. Here, we set the solvation shell of each dye atom to be 5 Å, except for the malemide atoms which were set to 2 Å in order to account for the connectivity of the protein and dye. All-retained protein + Alexa 488 conformations were combined with all-retained protein + Alexa 594 conformations. Conformations of the protein + Alexa 488 + Alexa 594 system were retained if no steric clashes were observed between the dyes. The FRET efficiencies corresponding to these conformations were then calculated using the Förster formula, and the mean and standard error associated with these FRET efficiencies were computed. Distances between dyes were calculated using the positions of the C19 atoms of Alexa 488 and Alexa 594, as defined by the HandyFRET AF488.pdb and AF594m.pdb files (http://karri.anu.edu.au/handy/rl.html), respectively. The mean FRET efficiency was recorded if the standard error was less than 0.005. If this criterion was not met, then the above process was repeated until the standard error was less than 0.005. However, if this criterion was not met after 10 trials then a mean FRET efficiency value was not recorded for the given conformation. Given that our goal was to construct simulated ensembles that are consistent with all three mean FRET efficiencies measured for each polyQ length, only conformations that had mean FRET efficiencies for all three dye pairs were kept for use in the reweighting procedure described next.

Reweighting Simulated Ensembles To Match Mean smFRET Efficiencies. The simulated ensembles were reweighted to match all three mean FRET efficiencies, $\langle E_{\text{FRET}} \rangle$, measured for a given polyQ length using the maximum entropy method of Leung et al.[18] This method maximizes entropy (i.e., tries to give all conformations similar weights) while minimizing the difference between the simulated and experimental $\langle E_{\text{FRET}} \rangle$ efficiencies and yields a unique global solution. Here, the error in the experimental FRET efficiencies was set to be 0.02.[62] In order to determine the simulated temperature that best matches the experimental results, the decrease from maximum entropy (ΔS) was calculated. This calculation is insensitive to the number of conformations used, which is important given that the number of conformations varies for each temperature and polyQ length combination. Here, ΔS is calculated using

$$\Delta S = S(\mathbf{p}^{\text{post}}) - S(\mathbf{p}^{\text{prior}}) = -\sum_{i=1}^{n_c} p_i^{\text{post}} \ln p_i^{\text{post}} - \ln(n_c)$$

Here, $\mathbf{p}^{\text{post}}$ is the posterior vector of weights determined from the maximum entropy method and $\mathbf{p}^{\text{prior}}$ is the vector of equal weights given to each of the n_c conformations. The mean free energy change is given by $kT\Delta S$, where k is the Boltzmann constant and T is the

temperature. Thus, if $\Delta S = -1$, then this is equivalent to adding an auxiliary reweighting term to the potential function that contributes $1kT$ to the overall energy function.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b06659.

> Materials and methods, experimental details, solid-phase peptide synthesis, protein characterization, and supplemental data (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*pappu@wustl.edu
*hilal.lashuel@epfl.ch

ORCID ⓞ

John B. Warner IV: 0000-0001-5562-5620
Kiersten M. Ruff: 0000-0003-3240-1856
Rohit V. Pappu: 0000-0003-2568-1378
Hilal A. Lashuel: 0000-0001-7682-8320

Author Contributions

#J.B.W. and K.M.R. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) MacDonald, M.; et al. *Cell* **1993**, *72* (6), 971−983.
(2) Andrew, S. E.; Goldberg, Y. P.; Kremer, B.; Telenius, H.; Theilmann, J.; Adam, S.; Starr, E.; Squitieri, F.; Lin, B. Y.; Kalchman, M. A.; Graham, R. K.; Hayden, M. R. *Nat. Genet.* **1993**, *4* (4), 398−403.
(3) Walker, F. O. *Lancet* **2007**, *369* (9557), 218−28.
(4) Sathasivam, K.; Neueder, A.; Gipson, T. A.; Landles, C.; Benjamin, A. C.; Bondulich, M. K.; Smith, D. L.; Faull, R. L.; Roos, R. A.; Howland, D.; Detloff, P. J.; Housman, D. E.; Bates, G. P. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (6), 2366−70.
(5) Davies, S. W.; Turmaine, M.; Cozens, B. A.; DiFiglia, M.; Sharp, A. H.; Ross, C. A.; Scherzinger, E.; Wanker, E. E.; Mangiarini, L.; Bates, G. P. *Cell* **1997**, *90* (3), 537−548.
(6) Raspe, M.; Gillis, J.; Krol, H.; Krom, S.; Bosch, K.; van Veen, H.; Reits, E. *J. Cell Sci.* **2009**, *122* (18), 3262−3271.
(7) Caron, N. S.; Desmond, C. R.; Xia, J.; Truant, R. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (36), 14610−14615.
(8) Crick, S. L.; Ruff, K. M.; Garai, K.; Frieden, C.; Pappu, R. V. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (50), 20075−80.
(9) Lyle, N.; Das, R. K.; Pappu, R. V. *J. Chem. Phys.* **2013**, *139* (12), 121907.
(10) Sahoo, B.; Singer, D.; Kodali, R.; Zuchner, T.; Wetzel, R. *Biochemistry* **2014**, *53* (24), 3897−3907.
(11) (a) Bugg, C. W.; Isas, J. M.; Fischer, T.; Patterson, P. H.; Langen, R. *J. Biol. Chem.* **2012**, *287* (38), 31739−31746. (b) Rizzo, M. A.; Springer, G.; Segawa, K.; Zipfel, W. R.; Piston, D. W. *Microsc. Microanal.* **2006**, *12* (3), 238−254.
(12) (a) Williamson, T. E.; Vitalis, A.; Crick, S. L.; Pappu, R. V. *J. Mol. Biol.* **2010**, *396* (5), 1295−309. (b) Lakhani, V. V.; Ding, F.; Dokholyan, N. V. *PLoS Comput. Biol.* **2010**, *6* (4), e1000772.

(c) Vitalis, A.; Wang, X.; Pappu, R. V. *Biophys. J.* **2007**, *93* (6), 1923−1937. (d) Binette, V.; Cote, S.; Mousseau, N. *Biophys. J.* **2016**, *110* (5), 1075−88. (e) Kelley, N. W.; Huang, X. H.; Tam, S.; Spiess, C.; Frydman, J.; Pande, V. S. *J. Mol. Biol.* **2009**, *388* (5), 919−927.
(13) Baias, M.; Smith, P. E. S.; Shen, K.; Joachimiak, L. A.; Żerko, S.; Koźmiński, W.; Frydman, J.; Frydman, L. *J. Am. Chem. Soc.* **2017**, *139* (3), 1168−1176.
(14) Vitalis, A.; Pappu, R. V. *J. Comput. Chem.* **2009**, *30* (5), 673−99.
(15) (a) Ansaloni, A.; Wang, Z.-M.; Jeong, J. S.; Ruggeri, F. S.; Dietler, G.; Lashuel, H. A. *Angew. Chem., Int. Ed.* **2014**, *53* (7), 1928−1933. (b) Vieweg, S.; Ansaloni, A.; Wang, Z.-M.; Warner, J. B.; Lashuel, H. A. *J. Biol. Chem.* **2016**, *291* (23), 12074−86.
(16) (a) Mukhopadhyay, S.; Krishnan, R.; Lemke, E. A.; Lindquist, S.; Deniz, A. A. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (8), 2649−2654. (b) Sisamakis, E.; Valeri, A.; Kalinin, S.; Rothwell, P. J.; Seidel, C. A. M. Accurate Single-Molecule FRET Studies using Multiparameter Fluorescence Detection In *Single Molecule Tools, Part B: Super-Resolution, Particle Tracking, Multiparameter, and Force Based Methods*; Walter, N. G., Ed.; Methods in Enzymology 475; Elsevier Academic Press Inc.: San Diego, 2010; pp 455−514.
(17) Aiken, C. T.; Steffan, J. S.; Guerrero, C. M.; Khashwji, H.; Lukacsovich, T.; Simmons, D.; Purcell, J. M.; Menhaji, K.; Zhu, Y.-Z.; Green, K.; LaFerla, F.; Huang, L.; Thompson, L. M.; Marsh, J. L. *J. Biol. Chem.* **2009**, *284* (43), 29427−29436.
(18) Leung, H. T.; Bignucolo, O.; Aregger, R.; Dames, S. A.; Mazur, A.; Berneche, S.; Grzesiek, S. *J. Chem. Theory Comput.* **2016**, *12* (1), 383−94.
(19) Kang, H.; Vazquez, F. X.; Zhang, L.; Das, P.; Toledo-Sherman, L.; Luan, B.; Levitt, M.; Zhou, R. *J. Am. Chem. Soc.* **2017**, *139* (26), 8820−8827.
(20) (a) Vitalis, A.; Lyle, N.; Pappu, R. V. *Biophys. J.* **2009**, *97* (1), 303−311. (b) Vitalis, A.; Pappu, R. V. *Biophys. Chem.* **2011**, *159* (1), 14−23. (c) Chen, M.; Tsai, M.; Zheng, W.; Wolynes, P. G. *J. Am. Chem. Soc.* **2016**, *138* (46), 15197−15203. (d) Chen, M.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (17), 4406−4411.
(21) Mapa, K.; Sikor, M.; Kudryavtsev, V.; Waegemann, K.; Kalinin, S.; Seidel, C. A. M.; Neupert, W.; Lamb, D. C.; Mokranjac, D. *Mol. Cell* **2010**, *38* (1), 89−100.
(22) Ratner, V.; Kahana, E.; Eichler, M.; Haas, E. *Bioconjugate Chem.* **2002**, *13* (5), 1163−1170.
(23) Schuler, B.; Lipman, E. A.; Eaton, W. A. *Nature* **2002**, *419* (6908), 743−747.
(24) (a) Ericson, M. D.; Rice, K. G. *Biopolymers* **2012**, *98* (6), 510−517. (b) Ericson, M. D.; Rice, K. G. *Tetrahedron Lett.* **2013**, *54* (35), 4746−4748. (c) Luhowy, R. R.; Cieciuch, R. F. W.; Meneghini, F. *Tetrahedron Lett.* **1973**, *14* (15), 1285−1288.
(25) Chiki, A.; DeGuire, S. M.; Ruggeri, F. S.; Sanfelice, D.; Ansaloni, A.; Wang, Z.-M.; Cendrowska, U.; Burai, R.; Vieweg, S.; Pastore, A.; Dietler, G.; Lashuel, H. A. *Angew. Chem., Int. Ed.* **2017**, *56* (19), 5202−5207.
(26) Bustamante, M. B.; Ansaloni, A.; Pedersen, J. F.; Azzollini, L.; Cariulo, C.; Wang, Z.-M.; Petricca, L.; Verani, M.; Puglisi, F.; Park, H.; Lashuel, H.; Caricasole, A. *Biochem. Biophys. Res. Commun.* **2015**, *463* (4), 1317−1322.
(27) (a) Milles, S.; Mercadante, D.; Aramburu, I. V.; Jensen, M. R.; Banterle, N.; Koehler, C.; Tyagi, S.; Clarke, J.; Shammas, S. L.; Blackledge, M.; Grater, F.; Lemke, E. A. *Cell* **2015**, *163* (3), 734−745. (b) Muller, B. K.; Zaychikov, E.; Brauchle, C.; Lamb, D. C. *Biophys. J.* **2005**, *89* (5), 3508−3522.
(28) Sahoo, B.; Arduini, I.; Drombosky, K. W.; Kodali, R.; Sanders, L. H.; Greenamyre, J. T.; Wetzel, R. *PLoS One* **2016**, *11* (6), e0155747.
(29) (a) O'Nuallain, B.; Thakur, A. K.; Williams, A. D.; Bhattacharyya, A. M.; Chen, S.; Thiagarajan, G.; Wetzel, R. Kinetics and thermodynamics of amyloid assembly using a high-performance liquid chromatography-based sedimentation assay. In *Amyloid, Prions, and Other Protein Aggregates, Part C*; Kheterpal, I.; Wetzel, R., Eds.; Methods in Enzymology 413; Academic Press: San Diego, 2006; pp 34−74. (b) Chen, S.; Wetzel, R. *Protein Sci.* **2001**, *10* (4), 887−891.

(30) (a) Eggeling, C.; Berger, S.; Brand, L.; Fries, J. R.; Schaffer, J.; Volkmer, A.; Seidel, C. A. M. *J. Biotechnol.* **2001**, *86* (3), 163–180. (b) Schaffer, J.; Volkmer, A.; Eggeling, C.; Subramaniam, V.; Striker, G.; Seidel, C. A. M. *J. Phys. Chem. A* **1999**, *103* (3), 331–336.

(31) Milles, S.; Lemke, E. A. *Biophys. J.* **2011**, *101* (7), 1710–1719.

(32) (a) O'Brien, E. P.; Morrison, G.; Brooks, B. R.; Thirumalai, D. *J. Chem. Phys.* **2009**, *130* (12), 124903. (b) Hofmann, H.; Soranno, A.; Borgia, A.; Gast, K.; Nettels, D.; Schuler, B. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (40), 16155–16160.

(33) Aznauryan, M.; Delgado, L.; Soranno, A.; Nettels, D.; Huang, J. R.; Labhardt, A. M.; Grzesiek, S.; Schuler, B. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (37), E5389–98.

(34) Fuertes, G.; Banterle, N.; Ruff, K. M.; Chowdhury, A.; Mercadante, D.; Koehler, C.; Kachala, M.; Estrada Girona, G.; Milles, S.; Mishra, A.; Onck, P. R.; Gräter, F.; Esteban-Martín, S.; Pappu, R. V.; Svergun, D. I.; Lemke, E. A. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (31), E6342–E6351.

(35) (a) Crick, S. L.; Jayaraman, M.; Frieden, C.; Wetzel, R.; Pappu, R. V. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (45), 16764–16769. (b) Walters, R. H.; Murphy, R. M. *J. Mol. Biol.* **2009**, *393* (4), 978–992.

(36) Zhang, J.; Muthukumar, M. *J. Chem. Phys.* **2009**, *130* (3), 035102.

(37) Thakur, A. K.; Jayaraman, M.; Mishra, R.; Thakur, M.; Chellgren, V. M.; Byeon, I.-J. L.; Anjum, D. H.; Kodali, R.; Creamer, T. P.; Conway, J. F.; Gronenborn, A. M.; Wetzel, R. *Nat. Struct. Mol. Biol.* **2009**, *16* (4), 380–389.

(38) Theodorou, D. N.; Suter, U. W. *Macromolecules* **1985**, *18* (6), 1206–1214.

(39) (a) Macdonald, D.; Tessari, M. A.; Boogaard, I.; Smith, M.; Pulli, K.; Szynol, A.; Albertus, F.; Lamers, M.; Dijkstra, S.; Kordt, D.; Reindl, W.; Herrmann, F.; McAllister, G.; Fischer, D. F.; Munoz-Sanjuan, I. *PLoS One* **2014**, *9* (5), e96854. (b) Weiss, A.; Abramowski, D.; Bibel, M.; Bodner, R.; Chopra, V.; DiFiglia, M.; Fox, J.; Kegel, K.; Klein, C.; Grueninger, S.; Hersch, S.; Housman, D.; Regulier, E.; Rosas, H. D.; Stefani, M.; Zeitlin, S.; Bilbe, G.; Paganetti, P. *Anal. Biochem.* **2009**, *395* (1), 8–15.

(40) Hipp, M. S.; Patel, C. N.; Bersuker, K.; Riley, B. E.; Kaiser, S. E.; Shaler, T. A.; Brandeis, M.; Kopito, R. R. *J. Cell Biol.* **2012**, *196* (5), 573–87.

(41) Owens, G. E.; New, D. M.; West, A. P., Jr.; Bjorkman, P. J. *J. Mol. Biol.* **2015**, *427* (15), 2507–19.

(42) Newcombe, E. A.; Ruff, K. M.; Sethi, A.; Ormsby, A.; Ramdzan, Y. M.; Fox, A.; Purcell, A. W.; Gooley, P. R.; Pappu, R. V.; Hatters, D. M. *bioRxiv* **2017**, DOI: 10.1101/179663.

(43) Ramdzan, Y. M.; Trubetskov, M. M.; Ormsby, A. R.; Newcombe, E. A.; Sui, X.; Tobin, M. J.; Bongiovanni, M. N.; Gras, S. L.; Dewson, G.; Miller, J. M. L.; Finkbeiner, S.; Moily, N. S.; Niclis, J.; Parish, C. L.; Purcell, A. W.; Baker, M. J.; Wilce, J. A.; Waris, S.; Stojanovski, D.; Bocking, T.; Ang, C. S.; Ascher, D. B.; Reid, G. E.; Hatters, D. M. *Cell Rep.* **2017**, *19* (5), 919–927.

(44) Margulis, J.; Finkbeiner, S. *Front. Cell. Neurosci.* **2014**, *8*, 218.

(45) Kayatekin, C.; Matlack, K. E. S.; Hesse, W. R.; Guan, Y. H.; Chakrabortee, S.; Russ, J.; Wanker, E. E.; Shah, J. V.; Lindquist, S. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (33), 12085–12090.

(46) Wear, M. P.; Kryndushkin, D.; O'Meally, R.; Sonnenberg, J. L.; Cole, R. N.; Shewmaker, F. P. *PLoS One* **2015**, *10* (8), e0136362.

(47) Chen, S. M.; Ferrone, F. A.; Wetzel, R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (18), 11884–11889.

(48) (a) Kar, K.; Jayaraman, M.; Sahoo, B.; Kodali, R.; Wetzel, R. *Nat. Struct. Mol. Biol.* **2011**, *18* (3), 328–336. (b) Scherzinger, E.; Sittler, A.; Schweiger, K.; Heiser, V.; Lurz, R.; Hasenbank, R.; Bates, G. P.; Lehrach, H.; Wanker, E. E. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (8), 4604–4609.

(49) Hoop, C. L.; Lin, H. K.; Kar, K.; Magyarfalvi, G.; Lamley, J. M.; Boatz, J. C.; Mandal, A.; Lewandowski, J. R.; Wetzel, R.; van der Wel, P. C. A. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (6), 1546–1551.

(50) Klein, F. A.; Pastore, A.; Masino, L.; Zeder-Lutz, G.; Nierengarten, H.; Oulad-Abdelghani, M.; Altschuh, D.; Mandel, J. L.; Trottier, Y. *J. Mol. Biol.* **2007**, *371* (1), 235–44.

(51) Sahoo, B.; Drombosky, K. W.; Wetzel, R. *Methods Mol. Biol.* **2016**, *1345*, 67–87.

(52) (a) Shen, K.; Calamini, B.; Fauerbach, J. A.; Ma, B.; Shahmoradian, S. H.; Serrano Lachapel, I. L.; Chiu, W.; Lo, D. C.; Frydman, J. *eLife* **2016**, *5*, e18065. (b) Tam, S.; Spiess, C.; Auyeung, W.; Joachimiak, L.; Chen, B.; Poirier, M. A.; Frydman, J. *Nat. Struct. Mol. Biol.* **2009**, *16* (12), 1279–U98.

(53) Kokona, B.; Rosenthal, Z. P.; Fairman, R. *Biochemistry* **2014**, *53* (43), 6738–46.

(54) Jayaraman, M.; Kodali, R.; Sahoo, B.; Thakur, A. K.; Mayasundari, A.; Mishra, R.; Peterson, C. B.; Wetzel, R. *J. Mol. Biol.* **2012**, *415* (5), 881–99.

(55) Ruff, K. M.; Harmon, T. S.; Pappu, R. V. *J. Chem. Phys.* **2015**, *143* (24), 243123.

(56) (a) Gu, X. F.; Greiner, E. R.; Mishra, R.; Kodali, R.; Osmand, A.; Finkbeiner, S.; Steffan, J. S.; Thompson, L. M.; Wetzel, R.; Yang, X. W. *Neuron* **2009**, *64* (6), 828–840. (b) Mishra, R.; Hoop, C. L.; Kodali, R.; Sahoo, B.; van der Wel, P. C. A.; Wetzel, R. *J. Mol. Biol.* **2012**, *424* (1–2), 1–14. (c) Thompson, L. M.; Aiken, C. T.; Kaltenbach, L. S.; Agrawal, N.; Illes, K.; Khoshnan, A.; Martinez-Vincente, M.; Arrasate, M.; O'Rourke, J. G.; Khashwji, H.; Lukacsovich, T.; Zhu, Y. Z.; Lau, A. L.; Massey, A.; Hayden, M. R.; Zeitlin, S. O.; Finkbeiner, S.; Green, K. N.; LaFerla, F. M.; Bates, G.; Huang, L.; Patterson, P. H.; Lo, D. C.; Cuervo, A. M.; Marsh, J. L.; Steffan, J. S. *J. Cell Biol.* **2009**, *187* (7), 1083–1099.

(57) Bhattacharyya, A.; Thakur, A. K.; Chellgren, V. M.; Thiagarajan, G.; Williams, A. D.; Chellgren, B. W.; Creamer, T. P.; Wetzel, R. *J. Mol. Biol.* **2006**, *355* (3), 524–535.

(58) Hackenberger, C. P. R. *Org. Biomol. Chem.* **2006**, *4* (11), 2291–2295.

(59) Milles, S.; Tyagi, S.; Banterle, N.; Koehler, C.; VanDelinder, V.; Plass, T.; Neal, A. P.; Lemke, E. A. *J. Am. Chem. Soc.* **2012**, *134* (11), 5187–5195.

(60) Radhakrishnan, A.; Vitalis, A.; Mao, A. H.; Steffen, A. T.; Pappu, R. V. *J. Phys. Chem. B* **2012**, *116* (23), 6862–71.

(61) Walczewska-Szewc, K.; Corry, B. *Phys. Chem. Chem. Phys.* **2014**, *16* (24), 12317–12326.

(62) Song, J. H.; Gomes, G. N.; Gradinaru, C. C.; Chan, H. S. *J. Phys. Chem. B* **2015**, *119* (49), 15191–15202.