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, MO 63130, USA

§ Structural and Computational Biology Unit, Cell Biology and Biophysics Unit, European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, 69117 Heidelberg, Germany

Supporting Information

Table of Contents

Materials

pTWIN1 vector (N6951S) and ER2566 *E.coli* (E6901S) competent cells were purchased from New England Biolabs. His$_6$-Ssp-Httex1-Q$_N$ cDNA was synthesized by GeneArt®. Isopropyl-β-D-thiogalactopyranoside (IPTG) was ordered from Applichem (A1008,0025). Phenylmethane-sulfonyl fluoride (PMSF) was purchased from Axonlab (A0999.0005). CLAP protease inhibitor (1000x) was made of 2.5 mg/ml of Leupeptin, Chymostatin, Antipain and Pepstatin A from Applichem (A2183, A2144, A2129, A2205) in DMSO. The PageRuler Prestained Protein Ladder (26617), Alexa488-maleimide (A10254), and Alexa594-maleimide (A10256) were purchased from Thermo Scientific. Silver triflate (88722) was purchased from Alfa Aesar.

Supporting Methods

Cloning and Mutagenesis. His$_6$-Ssp-Httex1-Q$_N$ cDNA was subcloned into the pTWIN1 vector using NdeI/PstI restriction sites by GeneArt®. PCR mutagenesis was performed using PrimeStarMax DNA polymerase to incorporate cysteine mutants into Httex1. Forward and reverse DNA primers were ordered from Microsynth (see **Table S2**). Double cysteine mutants were obtained by sequential rounds of mutagenesis. All clones were confirmed by DNA sequencing using the T7 promoter primer by Microsynth.

Characterization of Alexa 488 Isomers. Dual labeled Httex1 constructs eluted from HPLC purification as two main peaks resulting from the structural connectivity of the Alexa 488-maleimide probe as 5/6′ isomers. Using UHPLC, the two isomers of Alexa 488-maleimide were separated on a C18 column (see **Fig. S7**). Though separable by HPLC, smFRET measurements were performed using a mixture of the constitutional isomers of dual labeled Httex1.

Solid Phase Peptide Synthesis (SPPS) of Htt2-17 A2Thz, pT3, Nbz-Thioeste r. SPPS was performed essentially as described in Ansaloni et al.[1] SPPS was carried out manually using NovaPEG Rink Amide resin LL on a 0.2 mmol scale. An activating solution of 8.7 mL DIPEA (1.0 M final) in 50 mL dimethylformamide (DMF) and 20% piperidine in DMF were made and stored at 4°C. Synthesis of the peptide thioester started with the double coupling of 0.2 mmol of NovaPEG Rink Amide resin with Di-Fmoc-Dbz-OH using 5 equivalents of HATU with 10 equivalents of DIPEA. Following, the Fmoc protecting groups were removed with piperidine. All residues were attached using double couplings with 5 equivalents of Fmoc-amino acid-OH and 10 equivalents of DIPEA at room temp for 30 min each followed by Fmoc removal with piperidine. The Nbz group was activated by three sequential treatments with 5 equivalents of 4-nitrophenylchloroformate in DCM, followed by 0.5 M DIPEA in dichloromethane (DCM). Peptide cleavage was performed using Reagent H: 8.15 mL TFA, 500 mg phenol, 300 μL water, 500 μL thioanisole, 250 μL EDT, 200 μL DMS, and 150 mg ammonia iodide. Cleavage was allowed to proceed for 4 hours at room temperature. Crude peptide was then precipitated dropwise by addition into cold ether, centrifuged, washed twice with cold ether, and dried on the lyophilizer. Crude peptide was then dissolved with 20% buffer E in buffer D and insoluble material was separated by centrifugation and syringe filtration. The thioester peptide was purified on a X-Bridge Prep C18 10 μm OBD 250x19 mm column using a gradient of 30-60% buffer E over 50 min. Fractions were collected and purity was assayed by LCMS and pooled accordingly. Peptide was dried by lyophilization and stored at -20°C until further use. Final characterization was performed by C8 UPLC and LCMS (see **Fig. S8**).

Calculation of Biophysical Parameters of Httex1.

FRET efficiencies and stoichiometry were calculated according to the following equations:

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

where I_D and I_A describe the donor or acceptor intensity excited by donor or acceptor laser, respectively; I_A^{dir} is the intensity from directly excited acceptor molecules by acceptor laser; γ is a correction factor dependent upon quantum yields and detection efficiencies of donor and acceptor. In this work, γ was estimated from leakage and direct excitation corrected FRET efficiencies and stoichiometries, E_{app} and S_{app} (by setting $\gamma = 1$, **Fig. S9**) following a basic procedure developed by Lee et al.[2] and as described in further detail by Fuertes et al.[3]. The pre-assumption by using this method to estimate the γ value is that the quantum yields of both donor and acceptor do not vary across samples. Hence, we extracted lifetime information from smFRET measurements by selectively analyzing donor only and acceptor only populations in E_{FRET} vs S plots, which were found to agree well with control measurement using singly labelled species using ensemble lifetime measurements. From this analysis we derived the quantum yields of the donor (Φ_{AF488}) and acceptor (Φ_{AF594}) of each protein construct. The analysis showed that the quantum yields of both the donor and acceptor are within error between different polyQ length proteins (**Table S3**), a sound observation as labelling chemistry was same for all, and all proteins share common features and experimental conditions were kept constant. We next calculated γ from E_{app} and S_{app}. smFRET experiments were combined from three independent replicates, consisting of protein constructs measured on three independent days. A total of 15 Httex1 constructs, consisting of 5 different polyQ repeat lengths (15-49) with three unique donor-acceptor labeling sites were measured. E_{app} and S_{app} were then determined from a fixed window consisting of all FRET efficiencies and stoichiometry values between 0.3-0.7. A linear fit to a plot of $1/S_{app}$ vs E_{app} yields intercept c and slope m which relates to gamma in the following way[2-3]:

$$\gamma = \frac{c - 1}{c + m - 1}$$

Retrieved E_{app} values spanned a range from 0.19-0.44. The γ value equal to 0.64 was obtained with a good linear correlation $R^2 = 0.75$ (**Fig. S10**). This global γ value was then applied to the Httex1 data sets to obtain real E_{FRET} values for each individual Httex1 construct. Our pulsed-interleaved excitation measurements, which combines the coincidence measurements[4] with measurement of FRET efficiencies allows us to detect potential aggregated species by detecting exceedingly long bursts. The longer and brighter bursts correlate with the appearance of a distinct third population in the S vs. E_{FRET} plot (**Fig. S12**), which is most likely due to aggregated species and is separated well from the main FRET population.

Spectral overlap integral (J) values were calculated (**Fig. S11 and Table S3**), at 1 μM, for each polyQ repeat length from singly labeled Alexa 594 absorption and Alexa 488 emission spectra using:

$$J(\lambda) = \int_{480}^{700} \epsilon_A(\lambda) F_D(\lambda) \lambda^4 d\lambda$$

Where $\epsilon_A(\lambda)$ is the normalized molar absorptivity of the acceptor at wavelength λ and $F_D(\lambda)$ is the normalized fluorescence spectra of the donor. R_0 values were then obtained using the following equation:

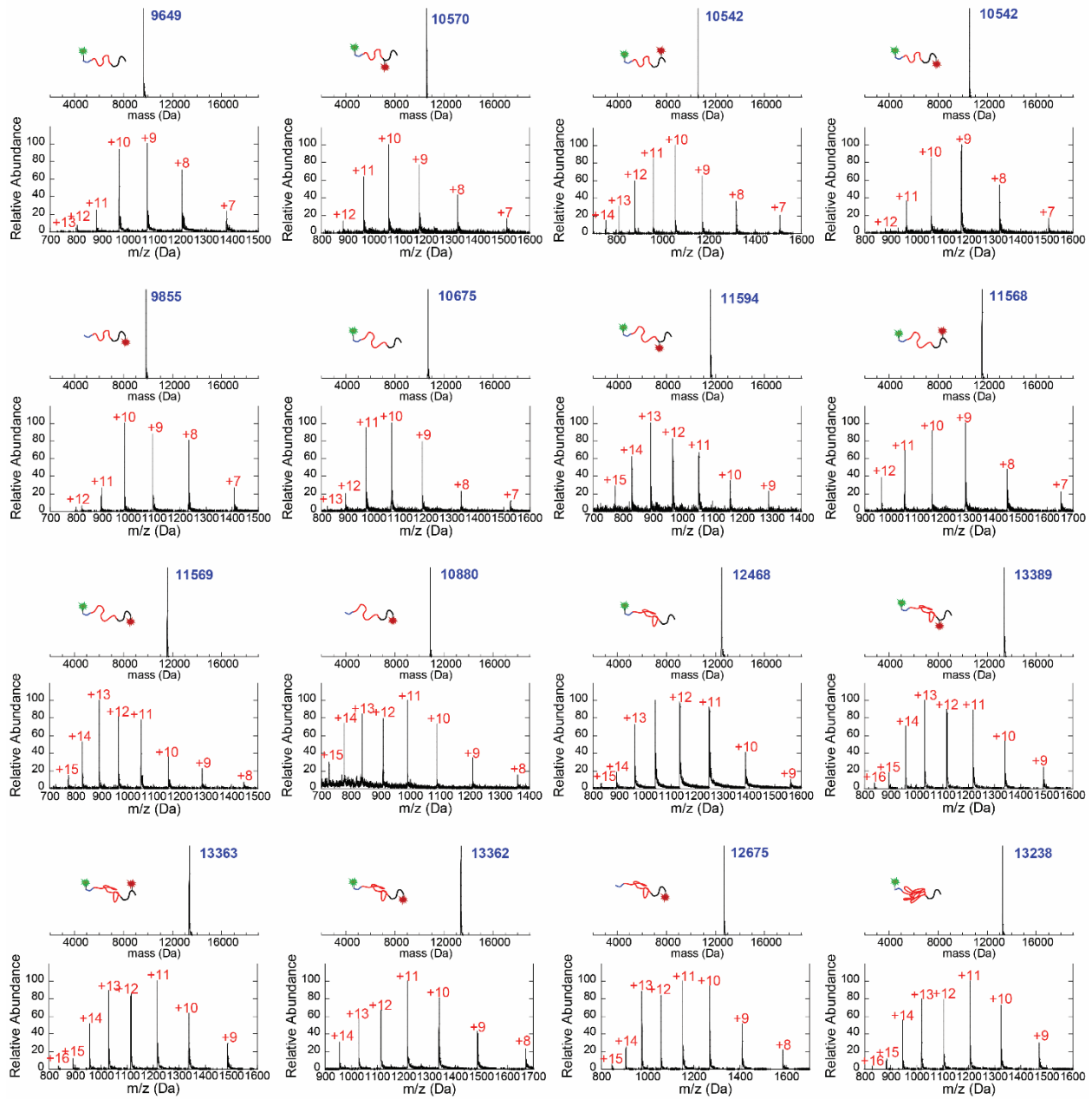$$R_0 = (8.79 x 10^{23} J \kappa^2 \eta^{-4} \Phi_{AF488})^{1/6}$$

Where J and Φ_{AF488} are specific for each polyQ repeat length and κ^2 is set at 2/3 and η is the solvent index of refraction (1.334). An average R_0 of 57±1 Å was obtained for Httex1. This analysis was also performed

globally for all proteins, since all emission and absorption spectra stayed consistent across samples (**Table S3**).
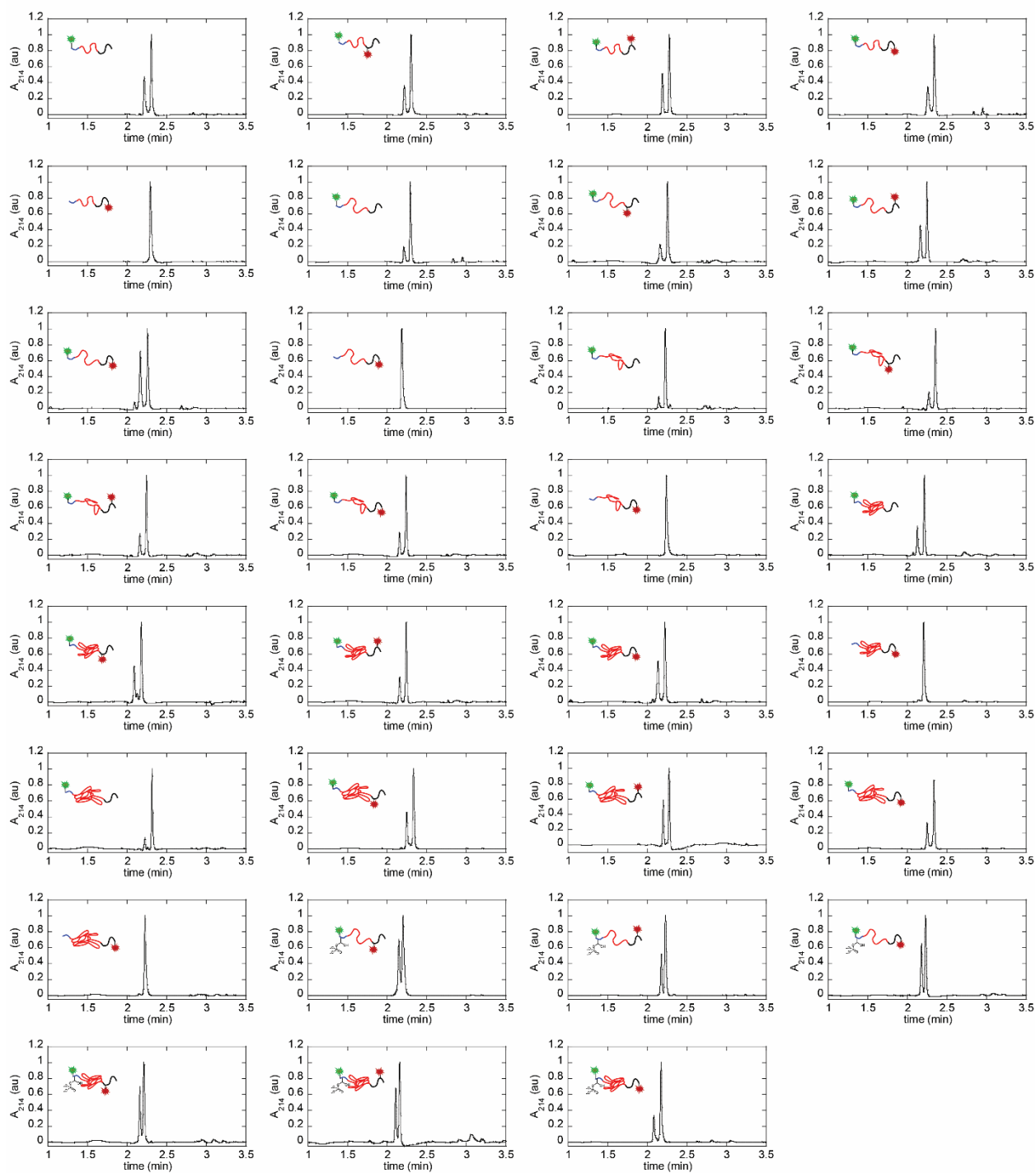
Supplemental Figures:
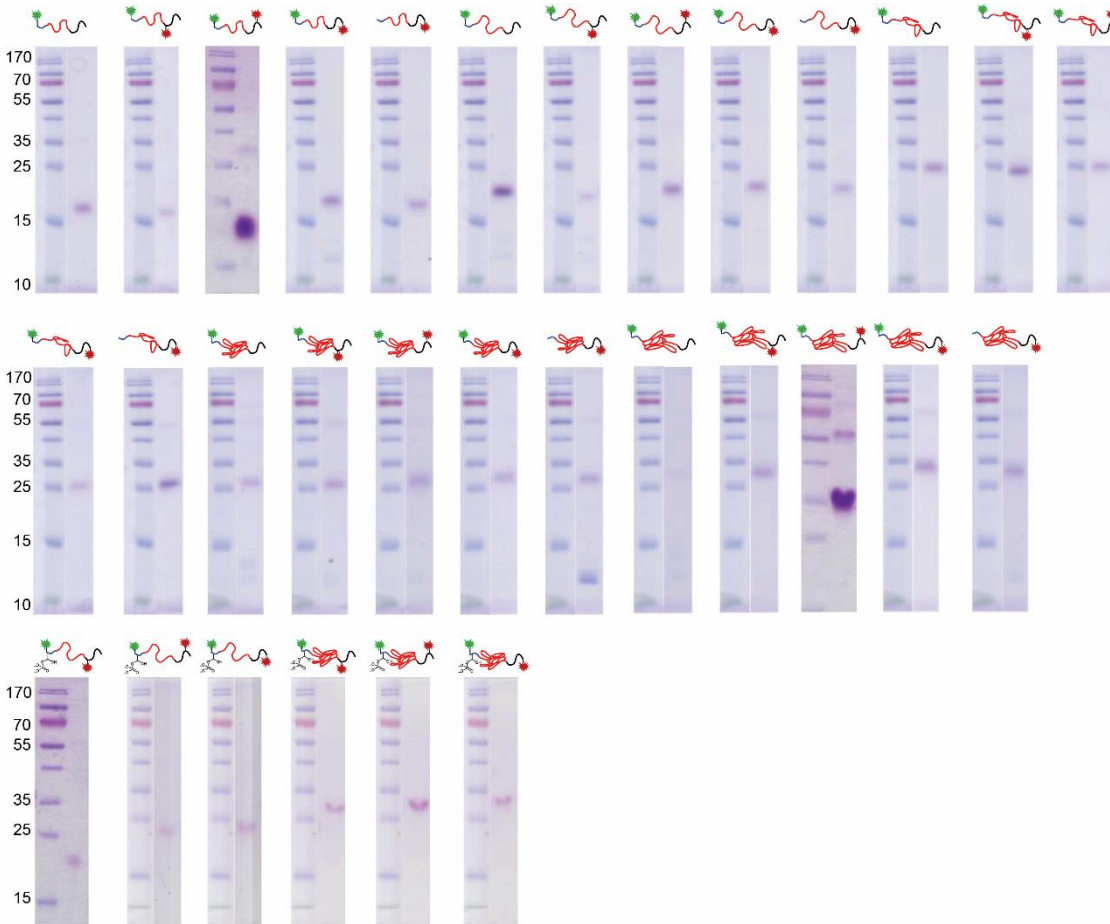
Figure S1 SDS-PAGE analysis of labeled Httex1 proteins. Httex1 15Q A2C[†], Httex1 15Q A2C[†] A60C[‡], Httex1 15Q A2C[†] P70C[‡], Httex1 15Q A2C[†] P90C[‡], Httex1 15Q A2Thz P90C[‡], Httex1 23Q A2C[†], Httex1 23Q A2C[†] A60C[‡], Httex1 23Q A2C[†] P80C[‡], Httex1 23Q A2C[†] P90C[‡], Httex1 23Q A2Thz P90C[‡], Httex1 37Q A2C[†], Httex1 37Q A2C[†] A60C[‡], Httex1 37Q A2C[†] P80C[‡], Httex1 37Q A2C[†] P90C[‡], Httex1 37Q A2Thz P90C[‡], Httex1 43Q A2C[†], Httex1 43Q A2C[†] A60C[‡], Httex1 43Q A2C[†] P70C[‡], Httex1 43Q A2C[†] P90C[‡], Httex1 43Q A2Thz P90C[‡], Httex1 49Q A2C[†], Httex1 49Q A2C[†] A60C[‡], Httex1 49Q A2C[†] P70C[‡], Httex1 49Q A2C[†] P90C[‡], Httex1 49Q A2Thz P90C[‡], Httex1 23Q pT3 A2C[†] A60C[‡], Httex1 23Q pT3 A2C[†] P80C[‡], Httex1 23Q pT3 A2C[†] P90C[‡], Httex1 43Q pT3 A2C[†] A60C[‡], Httex1 43Q pT3 A2C[†] P70C[‡], and Httex1 43Q pT3 A2C[†] P90C[‡]. A2C[†] indicates labeling with Alexa488 and P90C[‡] indicates labeling with Alexa594.

Figure S2 LCMS analysis of labeled Httex1 proteins. Httex1 15Q A2C[†], Httex1 15Q A2C[†] A60C[‡], Httex1 15Q A2C[†] P70C[‡], Httex1 15Q A2C[†] P90C[‡], Httex1 15Q A2Thz P90C[‡], Httex1 23Q A2C[†], Httex1 23Q A2C[†] A60C[‡], Httex1 23Q A2C[†] P80C[‡], Httex1 23Q A2C[†] P90C[‡], Httex1 23Q A2Thz P90C[‡], Httex1 37Q A2C[†], Httex1 37Q A2C[†] A60C[‡], Httex1 37Q A2C[†] P80C[‡], Httex1 37Q A2C[†] P90C[‡], Httex1 37Q A2Thz P90C[‡], and Httex1 43Q A2C[†]. A2C[†] indicates labeling with Alexa488 and P90C[‡] indicates labeling with Alexa594.
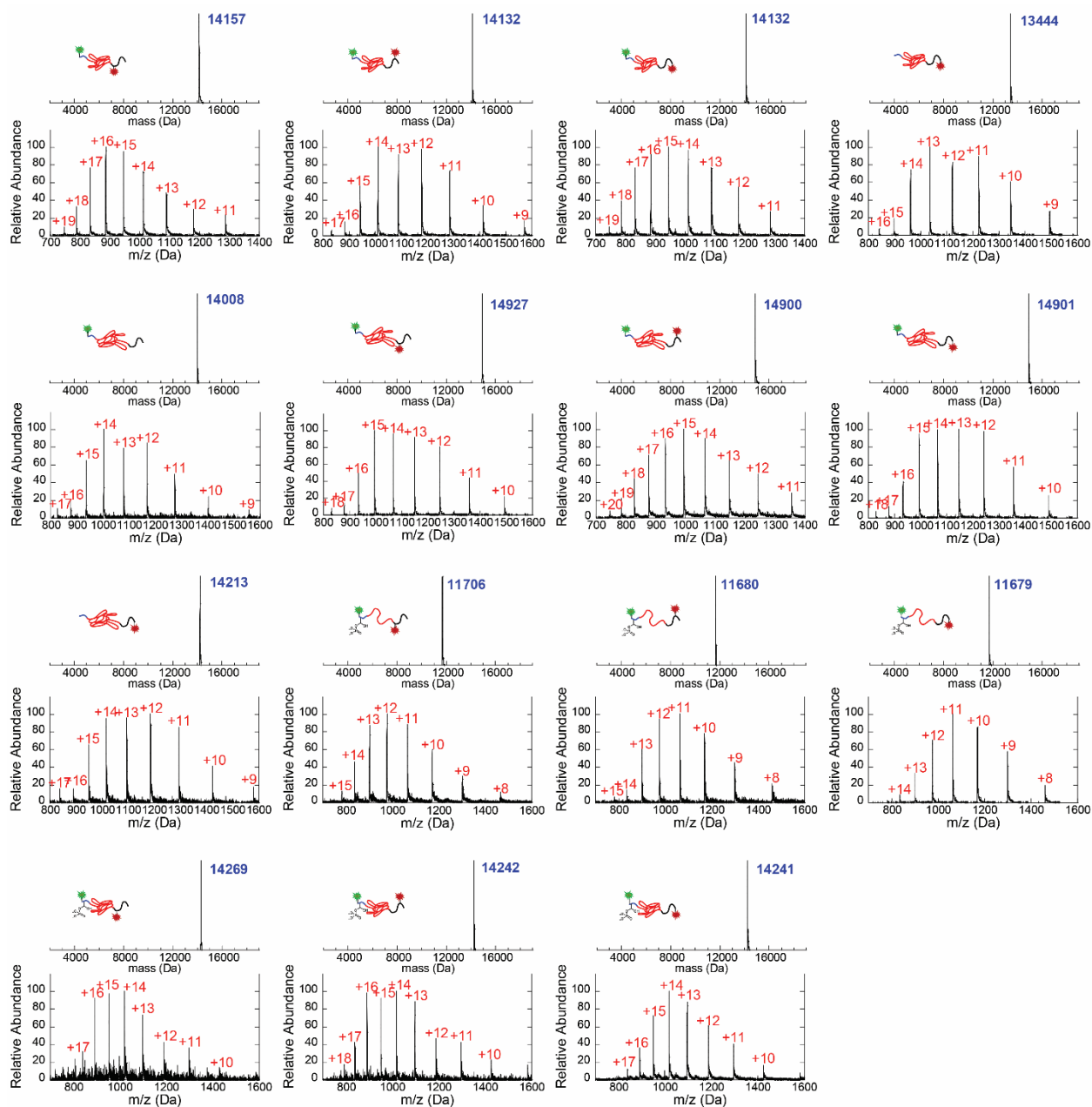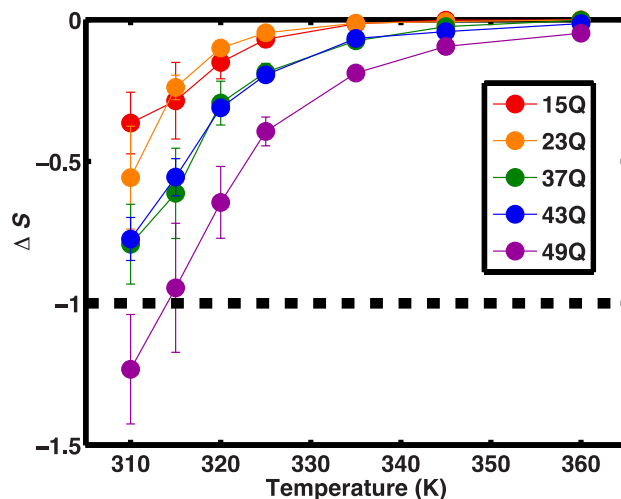
Figure S3 LCMS analysis of labeled Httex1 proteins. Httex1 43Q A2C[†] A60C[‡], Httex1 43Q A2C[†] P70C[‡], Httex1 43Q A2C[†] P90C[‡], Httex1 43Q A2Thz P90C[‡], Httex1 49Q A2C[†], Httex1 49Q A2C[†] A60C[‡], Httex1 49Q A2C[†] P70C[‡], Httex1 49Q A2C[†] P90C[‡], Httex1 49Q A2Thz P90C[‡], Httex1 23Q pT3 A2C[†] A60C[‡], Httex1 23Q pT3 A2C[†] P80C[‡], Httex1 23Q pT3 A2C[†] P90C[‡], Httex1 43Q pT3 A2C[†] A60C[‡], Httex1 43Q pT3 A2C[†] P70C[‡], and Httex1 43Q pT3 A2C[†] P90C[‡]. A2C[†] indicates labeling with Alexa488 and P90C[‡] indicates labeling with Alexa594.

Figure S4 UPLC analysis of labeled Httex1 proteins. Httex1 15Q A2C†, Httex1 15Q A2C† A60C‡, Httex1 15Q A2C† P70C‡, Httex1 15Q A2C† P90C‡, Httex1 15Q A2Thz P90C‡, Httex1 23Q A2C†, Httex1 23Q A2C† A60C‡, Httex1 23Q A2C† P80C‡, Httex1 23Q A2C† P90C‡, Httex1 23Q A2Thz P90C‡, Httex1 37Q A2C†, Httex1 37Q A2C† A60C‡, Httex1 37Q A2C† P80C‡, Httex1 37Q A2C† P90C‡, Httex1 37Q A2Thz P90C‡, Httex1 43Q A2C†, Httex1 43Q A2C† A60C‡, Httex1 43Q A2C† P70C‡, Httex1 43Q A2C† P90C‡, Httex1 43Q A2Thz P90C‡, Httex1 49Q A2C†, Httex1 49Q A2C† A60C‡, Httex1 49Q A2C† P70C‡, Httex1

S8

49Q A2C[†] P90C[‡], Httex1 49Q A2Thz P90C[‡], Httex1 23Q pT3 A2C[†] A60C[‡], Httex1 23Q pT3 A2C[†] P80C[‡], Httex1 23Q pT3 A2C[†] P90C[‡], Httex1 43Q pT3 A2C[†] A60C[‡], Httex1 43Q pT3 A2C[†] P70C[‡], and Httex1 43Q pT3 A2C[†] P90C[‡]. A2C[†] indicates labeling with Alexa488 and P90C[‡] indicates labeling with Alexa594.

Figure S5 Deviation from maximum entropy value ($\Delta S = 0$) for reweighted ensembles. If the simulation results were to generate perfect agreement with the experimentally derived FRET efficiencies, then $\Delta S = 0$. The calculation of ΔS is defined in the Methods section of the main text. We find that ΔS values are robust across the range of temperatures examined for all polyQ lengths. A ΔS value of -1 corresponds to a mean free energy change in the potential function of $1kT$, where k is the Boltzmann constant and T is the temperature. The threshold of $\Delta S = -1$ is used as a cutoff to identify ensembles that undergo minimal changes upon reweighting.
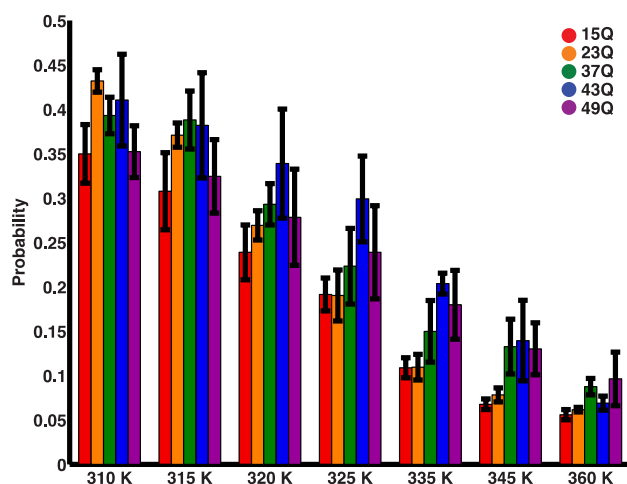
Figure S6 Probability of polyQ-PR domain interactions for reweighted ensembles as a function of polyQ length and temperature. The polyQ and PR domains were defined as interacting if any residue C-terminal to P_{11} stretch within the PR domain was within 8 Å of any residue within the polyQ domain. The interaction between the polyQ and PR domains decreases as temperature is increased.
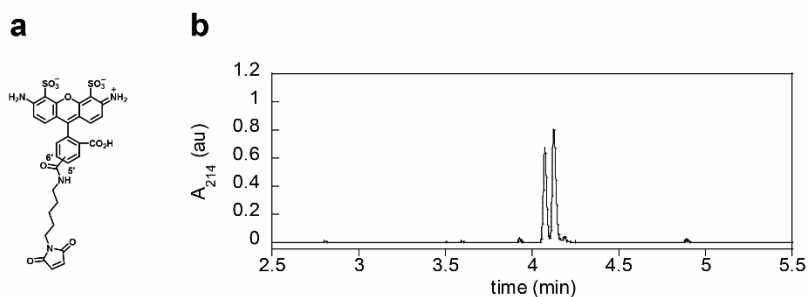
Figure S7 UHPLC analysis of Alexa488-maleimide. a) Structure of Alexa488-C_5-maleimide. b) C18 UHPLC shows separation of 5' and 6'-isomers of Alexa488-maleimide.
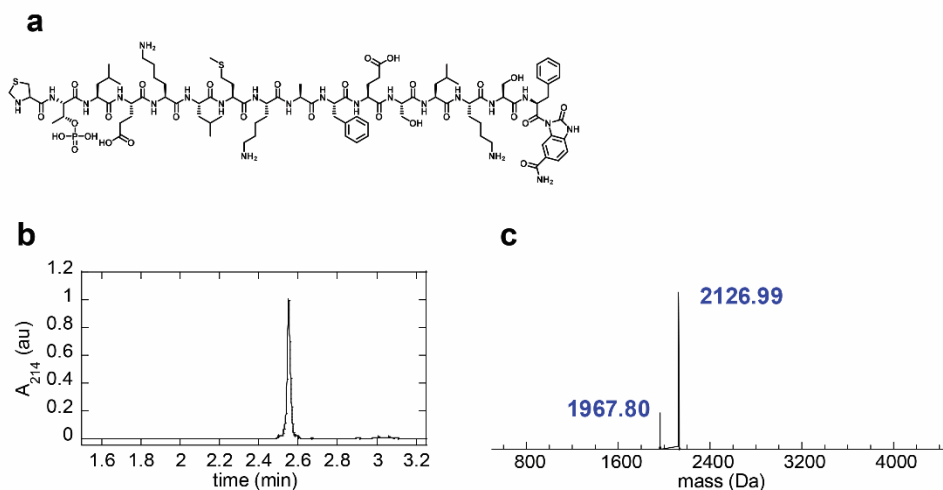
Figure S8 SPPS of Nt17 A2Thz pT3-Nbz. a) Amino acid structure of Httex1 Nt17 A2Thz pT3 Nbz-thioester. b) C8 UHPLC chromatogram of purified thioester peptide following preparative HPLC. c) LCMS ESI chromatogram of purified peptide. Exact mass: 2124.98 Da, observed: 2126.99 Da. Also shown is Nbz hydrolyzed (free carboxylic acid) peptide that co-eluted with the desired peptide (expected mass 1965.93 Da, observed: 1967.80 Da).
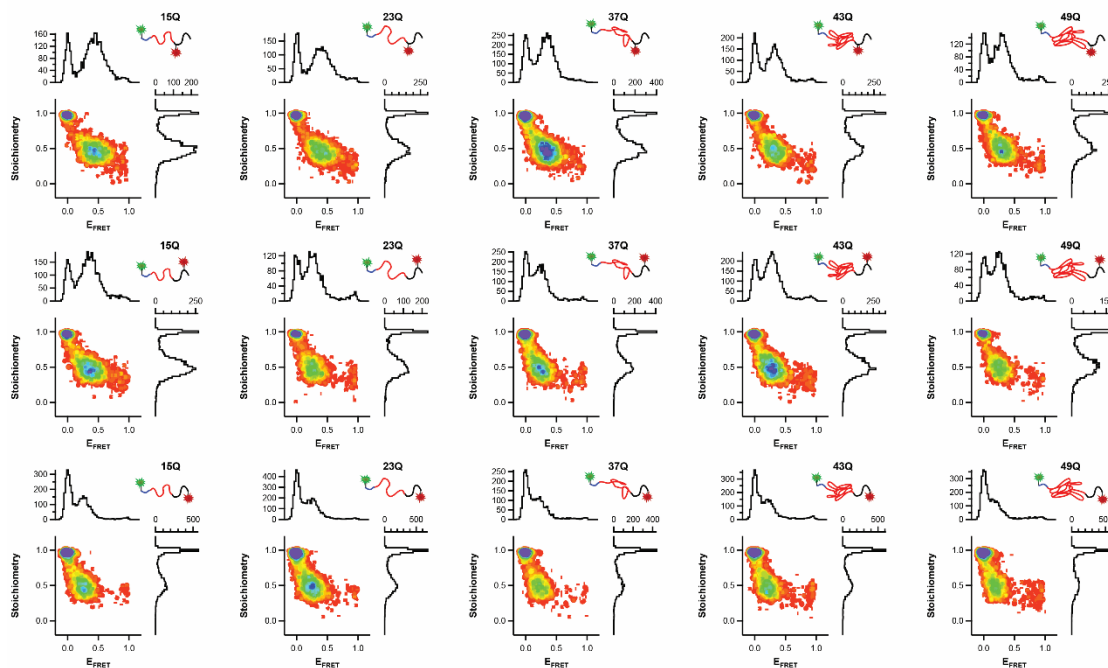
Figure S9 Apparent FRET efficiency versus apparent stoichiometry for calculation of γ. Two-dimensional E_{app} versus S_{app} histograms for Httex1 15-49Q with acceptor labeled at positions A60C, P70C, P80C, or P90C respectively with γ =1. Plots are a sum of three independent replicates for each Httex1 construct (15-49Q) dual labeled at either A60C, P70C, P80C, or P90C.
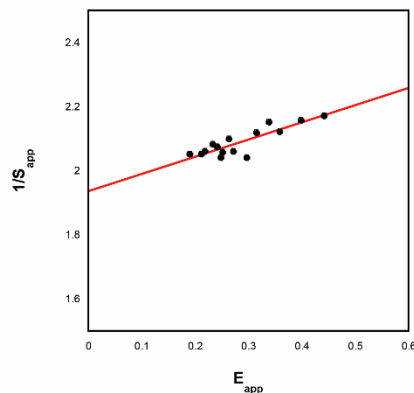
Figure S10 Determination of γ for Httex1. Plot of reciprocal Stoichiometry ($1/S_{app}$) versus FRET efficiencies (E_{app}) for Httex1 constructs. Data points represent a total of 15 Httex1 constructs, consisting of 5 different polyQ repeat lengths (15-49) with three unique donor-acceptor labeling sites. E_{app} and S_{app} were then determined from Fig. S9 using a fixed window consisting of all FRET efficiencies and stoichiometry values between 0.3-0.7. A linear best fit is represented by the red line yielding the equation: $1/S_{app} = 0.5(E_{app}) + 1.9$ with $R^2 = 0.75$.
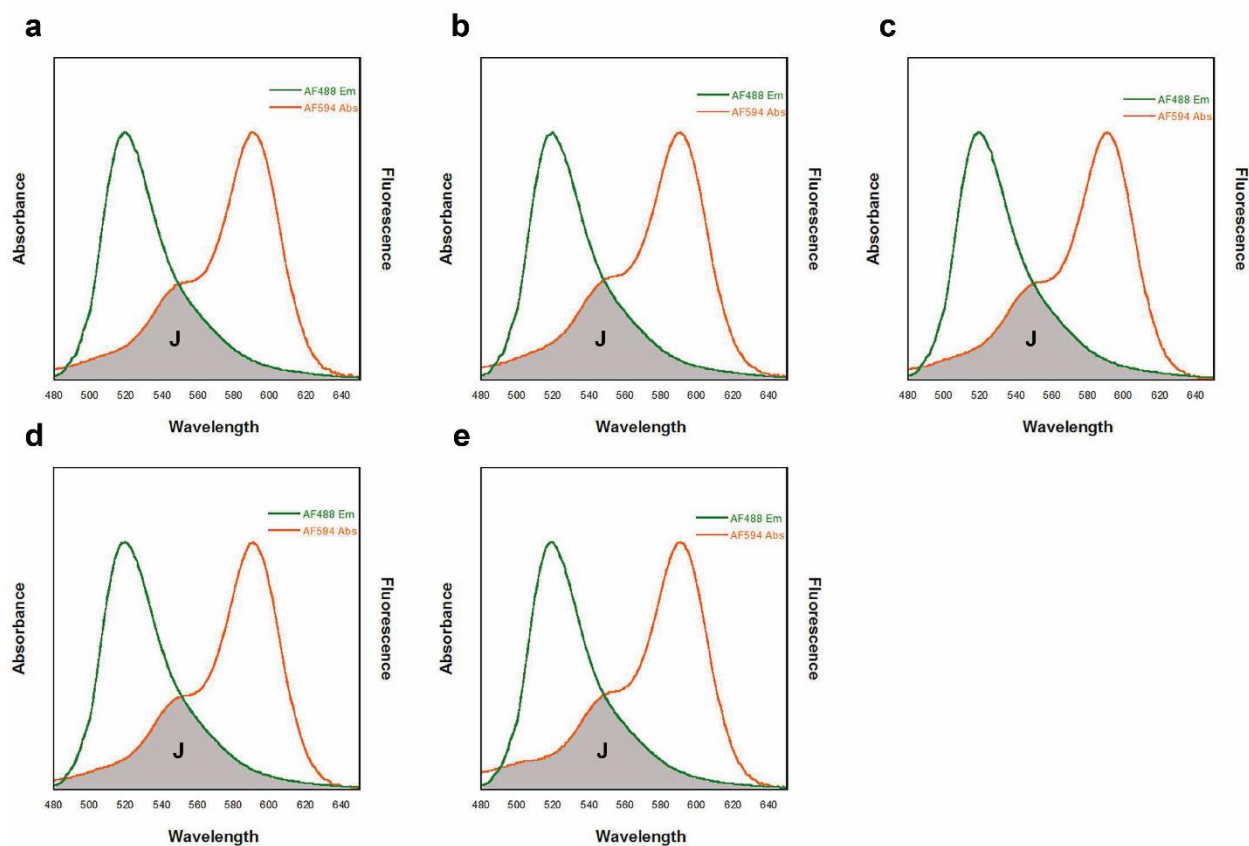
Figure S11 Calculation of Httex1 spectral overlap integral. Plots of Httex1 Alexa488 emission and Alexa594 absorbance versus wavelength. (a) 15Q, (b) 23Q, (c) 37Q, (d) 43Q, and (e) 49Q. Fluorescence and absorption spectra were normalized and the shaded region indicates the area of spectral overlap integral (J).
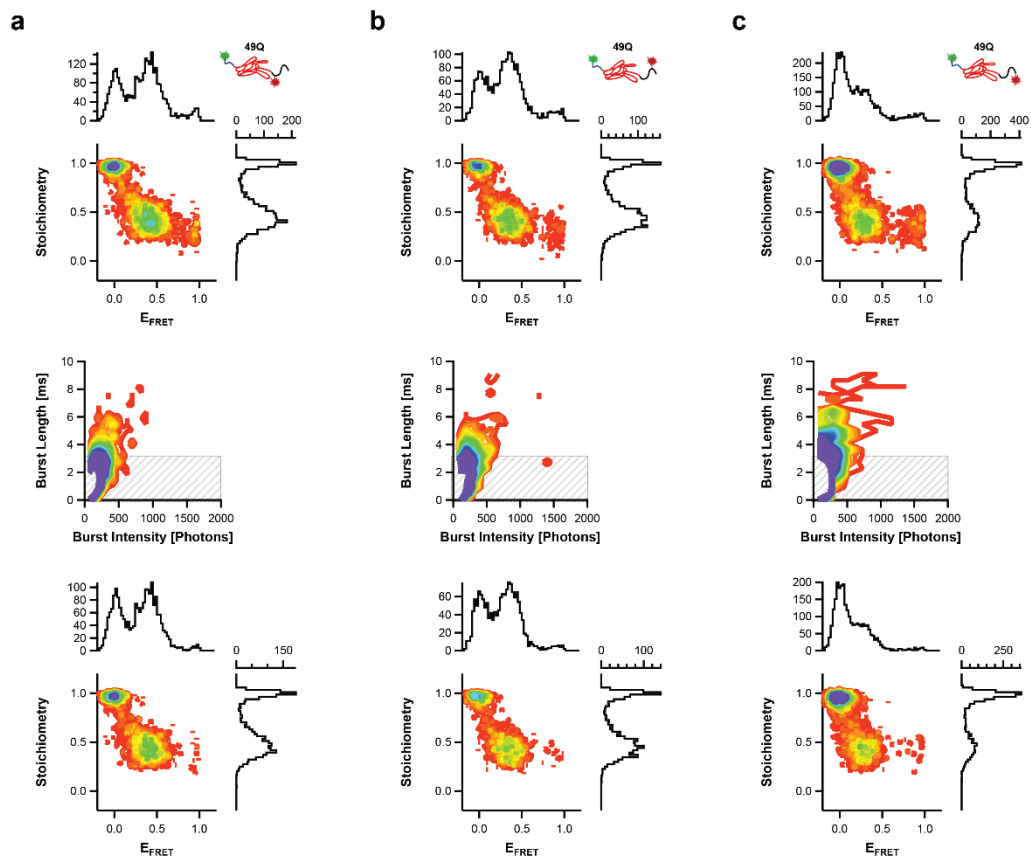
Figure S12 Analysis of oligomeric population in smFRET bursts. (a-c) Httex1 49Q A2C dual labeled at A60C, P70C, and P90C respectively. Shown at top are the stoichiometry versus FRET efficiency histograms with a burst length cutoff of 10 ms. Shown in the middle are the 2D plots of burst length versus burst intensity. The plots shown correspond to a bust length cut off of 10 ms. The gray shaded area corresponds to a burst length cutoff of 3 ms. The corresponding stoichiometry versus FRET efficiency using a 3 ms burst length cutoff is shown on bottom.

Supplemental Tables:

Table S1 Library of single and double labeled Httex1 constructs for smFRET

Labeled Httex1 Constructs	Expected (Da)	Observed (Da)	Yield (μg)
Httex1 15Q A2C[†]	9,659	9,649	483
Httex1 23Q A2C[†]	10,672	10,675	113
Httex1 37Q A2C[†]	12,466	12,468	22
Httex1 43Q A2C[†]	13,235	13,238	134
Httex1 49Q A2C[†]	14,004	14,008	382
Httex1 15Q A2Thz P90C[‡]	9,852	9,855	526
Httex1 23Q A2Thz P90C[‡]	10,877	10,880	152
Httex1 37Q A2Thz P90C[‡]	12,671	12,675	210
Httex1 43Q A2Thz P90C[‡]	13,440	13,444	250
Httex1 49Q A2Thz P90C[‡]	14,209	14,213	40
Httex1 15Q A2C[†] A60C[‡]	10,563	10,570	790
Httex1 15Q A2C[†] P70C[‡]	10,537	10,542	1000
Httex1 15Q A2C[†] P90C[‡]	10,537	10,542	240
Httex1 23Q A2C[†] A60C[‡]	11,588	11,594	450
Httex1 23Q A2C[†] P80C[‡]	11,562	11,568	120
Httex1 23Q A2C[†] P90C[‡]	11,562	11,569	1000
Httex1 37Q A2C[†] A60C[‡]	13,382	13,389	96
Httex1 37Q A2C[†] P80C[‡]	13,356	13,363	500
Httex1 37Q A2C[†] P90C[‡]	13,356	13,362	620
Httex1 43Q A2C[†] A60C[‡]	14,151	14,157	420
Httex1 43Q A2C[†] P70C[‡]	14,125	14,132	216
Httex1 43Q A2C[†] P90C[‡]	14,125	14,132	250
Httex1 49Q A2C[†] A60C[‡]	14,920	14,927	494
Httex1 49Q A2C[†] P70C[‡]	14,894	14,900	670
Httex1 49Q A2C[†] P90C[‡]	14,894	14,901	413
Httex1 23Q A2C[†] pT3 A60C[‡]	11,699	11,706	355
Httex1 23Q A2C[†] pT3 P80C[‡]	11,673	11,680	450
Httex1 23Q A2C[†] pT3 P90C[‡]	11,673	11,679	175
Httex1 43Q A2C[†] pT3 A60C[‡]	14,262	14,269	80
Httex1 43Q A2C[†] pT3 P70C[‡]	14,236	14,242	110
Httex1 43Q A2C[†] pT3 P90C[‡]	14,236	14,241	75

A2C[†] indicates labeling with Alexa488 and P90C[‡] indicates labeling with Alexa594. Expected and observed masses for the Httex1 proteins are indicated as well as final obtained protein yield.

Table S2: DNA primers used for PCR mutagenesis

Primer	Sequence
Httex1 A2C for:	CAAATGATATTATTGTGCACAACTGCACCCTGGAAAAACTGATGAAAG
Httex1 A2C rev:	CTTTCATCAGTTTTTCCAGGGTGCAGTTGTGCACAATAATATCATTTG
Httex1 A60C for:	CAGCCTCCGCCTCAGTGCCAGCCGCTGCTGCCA
Httex1 A60C rev:	TGGCAGCAGCGGCTGGCACTGAGGCGGAGGCTG
Httex1 P70C for:	CTGCCACAGCCTCAGCCATGCCCTCCACCGCC
Httex1 P70C rev:	GGCGGTGGAGGGCATGGCTGAGGCTGTGGCAG
Httex1 P80C for:	TTCTTCTGCAACTGCGCAACCCGGAGGTGGAGGTG
Httex1 P80C rev:	CACCTCCACCTCCGGGTTGCGCAGTTGCAGAAGAA
Httex1 P90C for:	AGAACCGCTGCATCGTTGCTAACTGCAGGAAGGGG
Httex1 P90C rev:	CCCCTTCCTGCAGTTAGCAACGATGCAGCGGTTCT

Table S3 Biophysical properties of Httex1

	Φ_{AF488}	Φ_{AF594}	J (cm^3M^{-1})	R_0 (Å)
15Q	0.85	0.66	2.17×10^{-13}	57
23Q	0.85	0.67	2.28×10^{-13}	57
37Q	0.86	0.66	2.17×10^{-13}	57
43Q	0.86	0.65	2.09×10^{-13}	57
49Q	0.86	0.63	2.16×10^{-13}	57
Average	0.86±0.01	0.65±0.02	$2.17 \pm 0.07 \times 10^{-13}$	57±1

Φ_{AF488} values are the quantum yields for the donor only Httex1 constructs and Φ_{AF594} values are the quantum yields for the acceptor only Httex1 constructs obtained from smFRET experiments. J and R_0 are the spectral overlap integrals and Förster distance of the dye pair for the Httex1 constructs, respectively.

Supplemental References

1. 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.
2. Lee, N. K.; Kapanidis, A. N.; Wang, Y.; Michalet, X.; Mukhopadhyay, J.; Ebright, R. H.; Weiss, S. Biophys J **2005,** 88 (4), 2939-2953.
3. Fuertes, G.; Banterlea, N.; Ruff, K. M.; Chowdhury, A.; Mercadante, D.; Koehler, C.; Kachala, M.; Girona, G. E.; Milles, S.; Mishra, A.; Onck, P. R.; Grater, F.; Esteban-Martin, S.; Pappu, R. V.; Svergun, D. I.; Lemke, E. A. Proc. Natl. Acad. Sci. U.S.A. **2017,** 114 (31), E6342-E6351.
4. Orte, A.; Clarke, R. W.; Klenerman, D. Analytical Chemistry **2008,** 80 (22), 8389-8397.