Supplemental Material

Table S1. Expression vector sequences*

Httex1 25Q - MBP Fusion Protein (bacterial)
MGSSHHHHHH GSSMKIEEGK LVIWINGDKG YNGLAEVGKK FEKDTGIKVT VEHPDKLEEK FPQVAATGDG PDIIFWAHDR FGGYAQSGLL AEITPDKAFQ DKLYPFTWDA VRYNGKLIAY PIAVEALSLI YNKDLLPNPP KTWEEIPALD KELKAKGKSA LMFNLQEPYF TWPLIAADGG YAFKYENGKY DIKDVGVDNA GAKAGLTFLV DLIKNKHMNA DTDYSIAEAA FNKGETAMTI NGPWAWSNID TSKVNYGVTV LPTFKGQPSK PFVGVLSAGI NAASPNKELA KEFLENYLLT DEGLEAVNKD KPLGAVALKS YEEELAKDPR IAATMENAQK GEIMPNIPQM SAFWYAVRTA VINAASGRQT VDEALKDAQT NSSSNNNNNN NNNNLGIEEN LYFQGHMATL EKLMKAFESL KSFQQQQQQQ QQQQQQQQQQ QQQQQQQQPP PPPPPPPPPQ LPQPPPQAQP LLPQPQPPPP PPPPPPGPAV AEEPLHRPKK W
Httex1 46Q - MBP Fusion Protein (bacterial)
MGSSHHHHHH GSSMKIEEGK LVIWINGDKG YNGLAEVGKK FEKDTGIKVT VEHPDKLEEK FPQVAATGDG PDIIFWAHDR FGGYAQSGLL AEITPDKAFQ DKLYPFTWDA VRYNGKLIAY PIAVEALSLI YNKDLLPNPP KTWEEIPALD KELKAKGKSA LMFNLQEPYF TWPLIAADGG YAFKYENGKY DIKDVGVDNA GAKAGLTFLV DLIKNKHMNA DTDYSIAEAA FNKGETAMTI NGPWAWSNID TSKVNYGVTV LPTFKGQPSK PFVGVLSAGI NAASPNKELA KEFLENYLLT DEGLEAVNKD KPLGAVALKS YEEELAKDPR IAATMENAQK GEIMPNIPQM SAFWYAVRTA VINAASGRQT VDEALKDAQT NSSSNNNNNN NNNNLGIEEN LYFQGHMATL EKLMKAFESL KSFQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQP PPPPPPPPPP QLPQPPPQAQ PLLPQPQPPP PPPPPPPGPA VAEEPLHRPK KW
pTREX Httex1 25Q - EmGFP (mammalian)
MATLEKLMKA FESLKSFQQQ QQQQQQQQQQ QQQQQQQQQQ QQPPPPPPPP PPPQLPQPPP QAQPLLPQPQ PPPPPPPPPP GPAVAEEPLH RPGSLVSKGE ELFTGVVPIL VELDGDVNGH KFSVSGEGEG DATYGKLTLK FICTTGKLPV PWPTLVTTLT YGVQCFARYP DHMKQHDFFK SAMPEGYVQE RTIFFKDDGN YKTRAEVKFE GDTLVNRIEL KGIDFKEDGN ILGHKLEYNY NSHKVYITAD KQKNGIKVNF KTRHNIEDGS VQLADHYQQN TPIGDGPVLL PDNHYLSTQS KLSKDPNEKR DHMVLLEFVT AAGITLGMDE LYK
pTREX Httex1 46Q - EmGFP (mammalian)
MATLEKLMKA FESLKSFQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQPPPPPPP PPPPQLPQPP PQAQPLLPQP QPPPPPPPPP PGPAVAEEPL HRPGSLVSKG EELFTGVVPI LVELDGDVNG HKFSVSGEGE GDATYGKLTL KFICTTGKLP VPWPTLVTTL TYGVQCFARY PDHMKQHDFF KSAMPEGYVQ ERTIFFKDDG NYKTRAEVKF EGDTLVNRIE LKGIDFKEDG NILGHKLEYN YNSHKVYITA DKQKNGIKVN FKTRHNIEDG SVQLADHYQQ NTPIGDGPVL LPDNHYLSTQ SKLSKDPNEK RDHMVLLEFV TAAGITLGMD ELYK

pTREX - EmGFP (mammalian)
MLVSKGEELF TGVVPILVEL DGDVNGHKFS VSGEGEGDAT YGKLTLKFIC TTGKLPVPWP TLVTTLTYGV QCFARYPDHM KQHDFFKSAM PEGYVQERTI FFKDDGNYKT RAEVKFEGDT LVNRIELKGI DFKEDGNILG HKLEYNYNSH KVYITADKQK NGIKVNFKTR HNIEDGSVQL ADHYQQNTPI GDGPVLLPDN HYLSTQSKLS KDPNEKRDHM VLLEFVTAAG ITLGMDELYK
pTREX Httex1 25Q - mCherry (mammalian)
MATLEKLMKA FESLKSFQQQ QQQQQQQQQQ QQQQQQQQQQ QQPPPPPPPP PPPQLPQPPP QAQPLLPQPQ PPPPPPPPPP GPAVAEEPLH RPGSVSKGEE DNMAIIKEFM RFKVHMEGSV NGHEFEIEGE GEGRPYEGTQ TAKLKVTKGG PLPFAWDILS PQFMYGSKAY VKHPADIPDY LKLSFPEGFK WERVMNFEDG GVVTVTQDSS LQDGEFIYKV KLRGTNFPSD GPVMQKKTMG WEASSERMYP EDGALKGEIK QRLKLKDGGH YDAEVKTTYK AKKPVQLPGA YNVNIKLDIT SHNEDYTIVE QYERAEGRHS TGGMDELYK
pTREX Httex1 46Q - mCherry (mammalian)
MATLEKLMKA FESLKSFQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQPPPPPPP PPPPQLPQPP PQAQPLLPQP QPPPPPPPPP PGPAVAEEPL HRPGSVSKGE EDNMAIIKEF MRFKVHMEGS VNGHEFEIEG EGEGRPYEGT QTAKLKVTKG GPLPFAWDIL SPQFMYGSKA YVKHPADIPD YLKLSFPEGF KWERVMNFED GGVVTVTQDS SLQDGEFIYK VKLRGTNFPS DGPVMQKKTM GWEASSERMY PEDGALKGEI KQRLKLKDGG HYDAEVKTTY KAKKPVQLPG AYNVNIKLDI TSHNEDYTIV EQYERAEGRH STGGMDELYK
pEGFP-C1 - mCherry (mammalian)
MVSKGEEDNM AIIKEFMRFK VHMEGSVNGH EFEIEGEGEG RPYEGTQTAK LKVTKGGPLP FAWDILSPQF MYGSKAYVKH PADIPDYLKL SFPEGFKWER VMNFEDGGVV TVTQDSSLQD GEFIYKVKLR GTNFPSDGPV MQKKTMGWEA SSERMYPEDG ALKGEIKQRL KLKDGGHYDA EVKTTYKAKK PVQLPGAYNV NIKLDITSHN EDYTIVEQYE RAEGRHSTGG MDELYK
pTREX GFP^{inv} Y66L (mammalian)
MLVSKGEELF TGVVPILVEL DGDVNGHKFS VSGEGEGDAT YGKLTLKFIC TTGKLPVPWP TLVTTLTLGV QCFARYPDHM KQHDFFKSAM PEGYVQERTI FFKDDGNYKT RAEVKFEGDT LVNRIELKGI DFKEDGNILG HKLEYNYNSH KVYITADKQK NGIKVNFKTR HNIEDGSVQL ADHYQQNTPI GDGPVLLPDN HYLSTQSKLS KDPNEKRDHM VLLEFVTAAG ITLGMDELYK

* Sequences are color coded as follows: 6 × Histidine tag; Maltose binding protein; TEV protease recognition sequence; Huntingtin; Fluorescent protein

Table S2. Backbone assignments for Httex1 Q25. * indicates glutamines within polyQ that are sequentially indistinguishable.

Assignment	^{15}N	^{1}H
H02N-H	118.245	8.918
M03N-H	122.906	8.714
A04N-H	126.282	8.590
T05N-H	114.848	8.218
L06N-H	124.242	8.348
K08N-H	120.566	8.334
L09N-H	122.014	8.271
M10N-H	121.389	8.357
K11N-H	121.758	8.260
A12N-H	124.111	8.205
F13N-H	119.419	8.203
E14N-H	121.375	8.238
S15N-H	116.620	8.295
K17N-H	122.483	8.184
S18N-H	116.136	8.260
F19N-H	123.000	8.338
Q20N-H	119.958	8.455
Q21N-H	120.074	8.247
Q22N-H	120.354	8.278
Q23N-H	121.430	8.453
Q24N-H	121.207	8.427
QxN-H*	120.734	8.251
QxN-H*	120.665	8.352
QxN-H*	120.741	8.369
QxN-H*	121.086	8.416
QxN-H*	121.890	8.497
Q43N-H	120.854	8.385
Q44N-H	121.123	8.357
Q56N-H	120.341	8.398
L57N-H	123.383	8.568
Q63N-H	120.843	8.539
A64N-H	126.267	8.512
L67N-H	122.760	8.446
L68N-H	125.129	8.337
Q70N-H	122.245	8.621
Q72N-H	122.079	8.576
G83N-H	109.285	8.344
A85N-H	124.566	8.499
V86N-H	120.350	8.211
A87N-H	128.156	8.469
E88N-H	121.195	8.496
E89N-H	123.695	8.491
L91N-H	126.013	8.507
H92N-H	118.793	8.539
R93N-H	123.802	8.446
K95N-H	122.293	8.471
K96N-H	123.607	8.254
W97N-H	127.594	7.850
W97-Indole	128.853	10.107

	χ^2_{CA}	χ^2_{CB}	χ^2_{CO}	χ^2_{N}	χ^2_{H}
All residues	4.48	0.84	1.21	0.68	0.30
|Z-score|≤1	1.45	0.40	1.06	0.68	0.30
|Z-score|≤1, no C-term	0.81	0.38	1.01	0.53	0.28

Table S3: Quantifies the quality of agreement between the simulation average and experimental data, χ^2_{δ}, for each chemical shift type, δ. Here, $\chi^2_{\delta} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left(\langle\delta_{i,\mathrm{sim}}\rangle-\delta_{i,\mathrm{exp}}\right)^2}{\sigma^2_{i,\delta}}$, where $\langle\delta_{i,\mathrm{sim}}\rangle$ is the average chemical shift for residue *i* from the simulations, $\delta_{i,\mathrm{exp}}$ is the chemical shift for residue *i* from the experiments, and $\sigma_{i,\delta}$ is the combined RMS error from SPARTA+ for chemical shift δ, and from residue *i* across five independent simulations. The combined RMS error was calculated using $\sigma_{i,\delta} = \left(\left(\sigma_{i,\delta,\mathrm{sim}}\right)^2 + \left(\sigma_{\delta,\mathrm{SPARTA+}}\right)^2\right)^{1/2}$. The RMS errors for SPARTA+ are $\sigma_{\mathrm{CA,SPARTA+}}$=0.92 ppm, $\sigma_{\mathrm{CB,SPARTA+}}$=1.13 ppm, $\sigma_{\mathrm{CO,SPARTA+}}$=1.07 ppm, $\sigma_{\mathrm{N,SPARTA+}}$=2.45 ppm, and $\sigma_{\mathrm{H,SPARTA+}}$=0.49 ppm. $\chi^2_{\delta}\leq 1$ implies agreement between the simulation and experimental results within $\sigma_{i,\delta}$. All simulation derived chemical shifts, except CA, are within or approximately within error of the experimental chemical shifts when all the residues are considered (All residues). The agreement between the simulations and the experiments improves for CA, CB, and CO when the experimental chemical shifts that are greater than one standard deviation from the average BMRB chemical shift for a given residue are removed (|Z-score|≤1). Finally, we also removed the residues that show high helicity in the C-terminus of Httex1 in the simulations as a result of using the protonated version of His (|Z-score|≤1, no C-term). Removal of these residues leads to all the simulation derived chemical shifts being within error or approximately within error of the experimental chemical shifts.

Table S4. Summary of proteomics analyses for interactors of soluble Httex1-GFP in Neuro2a cells using label-free quantitation. (*Attached as a separate Microsoft Xcel file*).

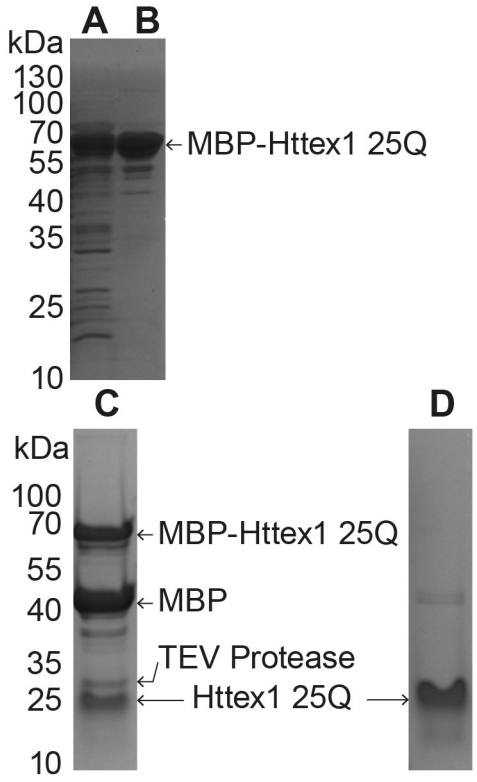

Fig S1. Purification of Httex1 from MBP tagged Httex1 after TEV protease treatment. Coomassie stained gels showing purification steps of Httex1 25Q. (**A**) Purification steps of MBP-Httex1 using a HisTrap column followed by (**B**) an MBPTrap column run on a Tris-Glycine 12% polyacrylamide gel. (**C**) 8-12% Tris-Tricine gel of TEV protease treated fusion protein and (**D**) after filtration of the cleavage reactants via a HisTrap column.

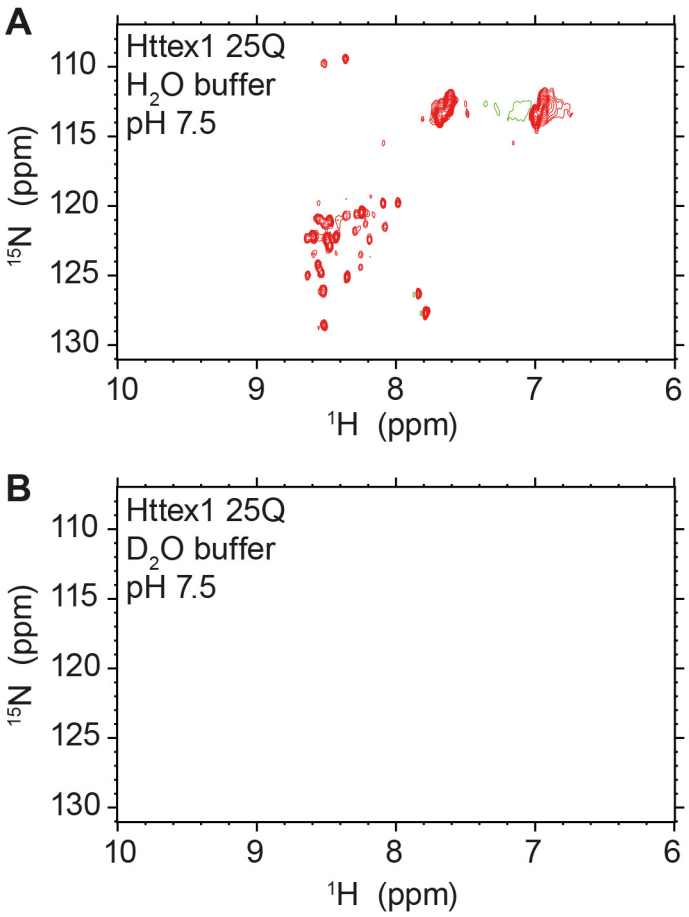

Fig S2. (**A**) Httex1 25Q 2D HSQC taken in low salt buffer (20 mM Tris-HCl, 100 mM NaCl, pH 7.5) at 5 °C and (**B**) in deuterated buffer.

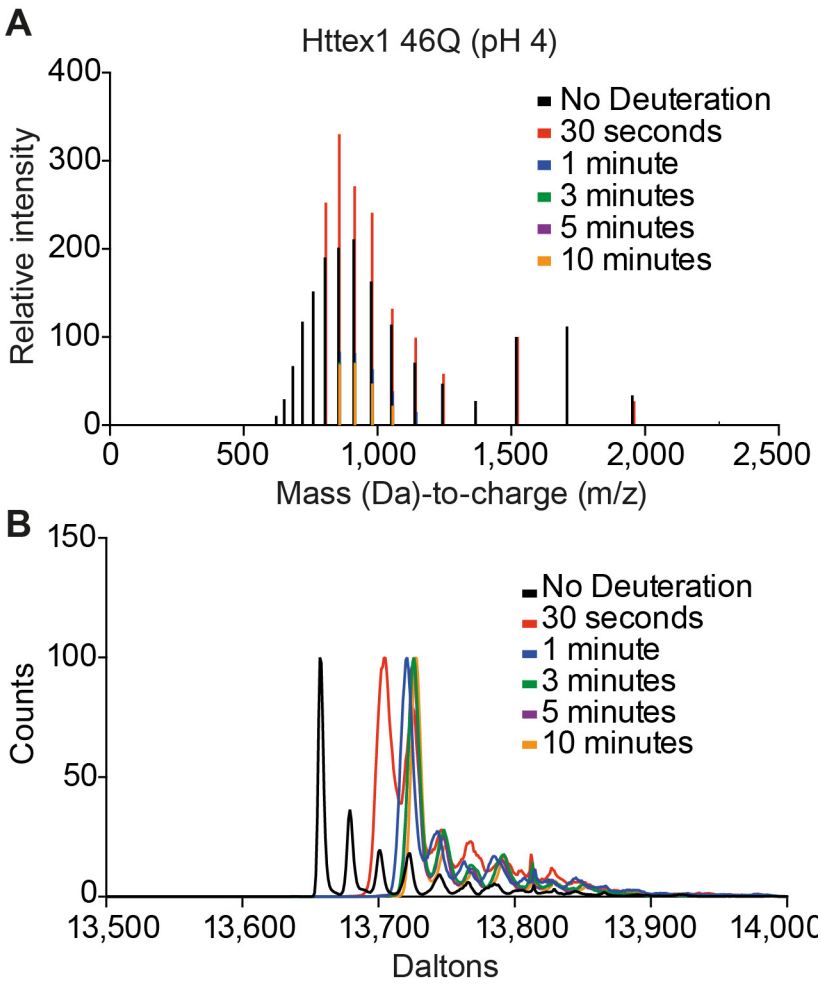

Fig S3. (**A**) Mass spectra of Httex1 46Q in sodium acetate buffer (150 mM, pH 4), following deuteration at different time points. (**B**) The corresponding deconvoluted mass data of the data in panel A.

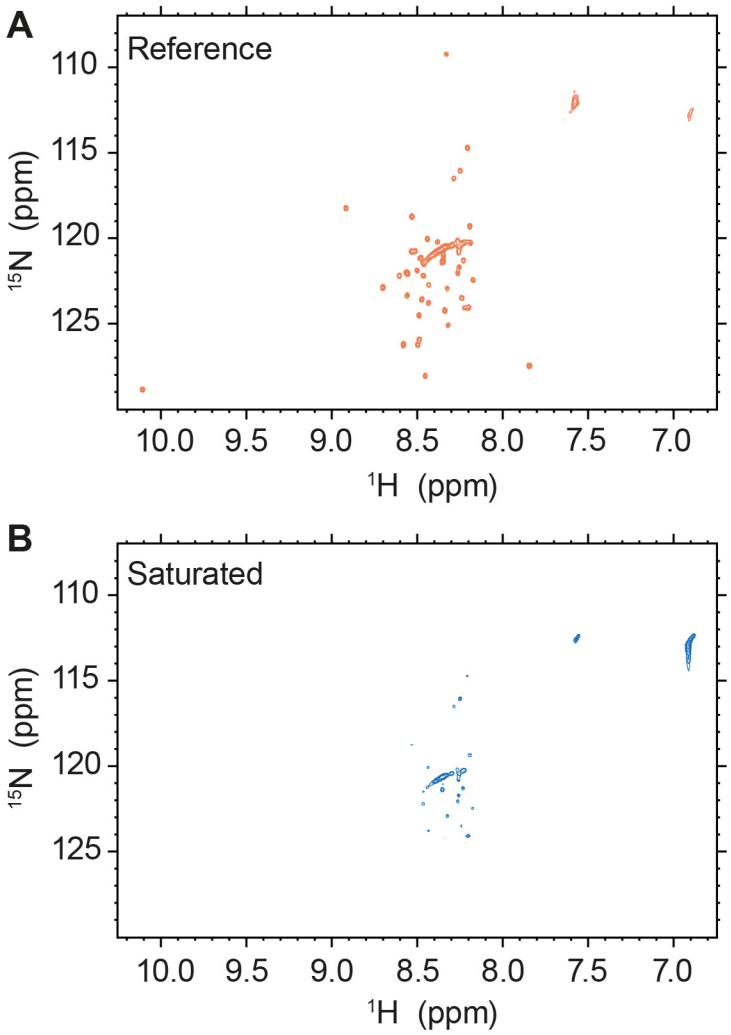

Fig S4. (**A**) ^{15}N{^{1}H}-NOE reference spectra and (**B**) saturated 2D HSQC spectra for Httex1 25Q taken in sodium acetate buffer (150 mM, pH 4) at 5 °C.

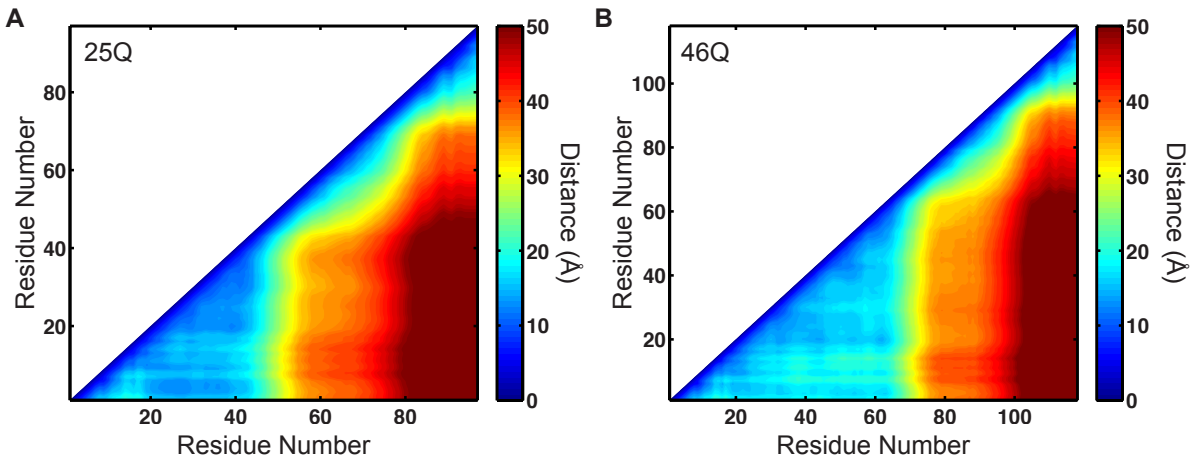

Fig S5. Distance maps from ensembles of Httex1 25Q (A) and 46Q (B). Distance maps quantify the average distance between two residues within the sequence. The shorter the average distance between two residues the cooler the color.

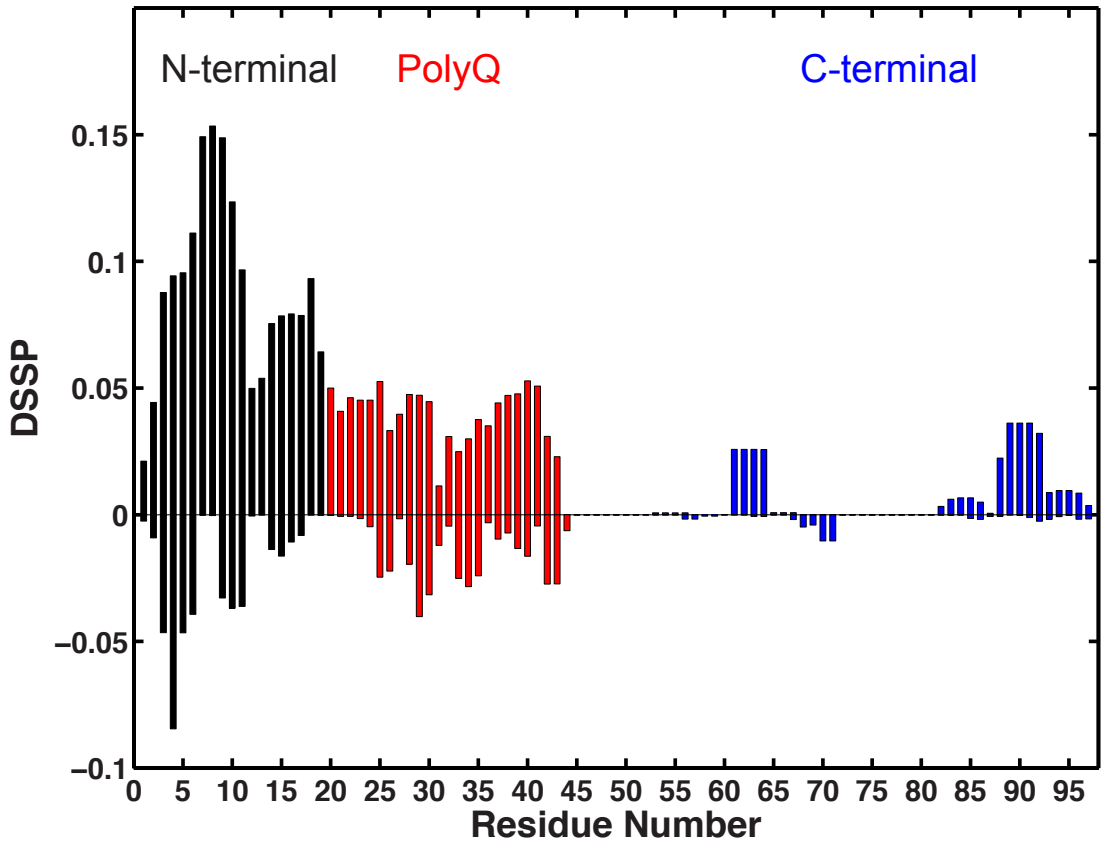

Fig S6. Secondary structure propensities from ensembles of 25Q Httex1 with deprotonated histidine. Probability a residue is found in an α-helical stretch of length ≥4 residues (positive) or an extended stretch of ≥2 residues (negative) as defined by the DSSP algorithm. Domains are colored coded as shown.

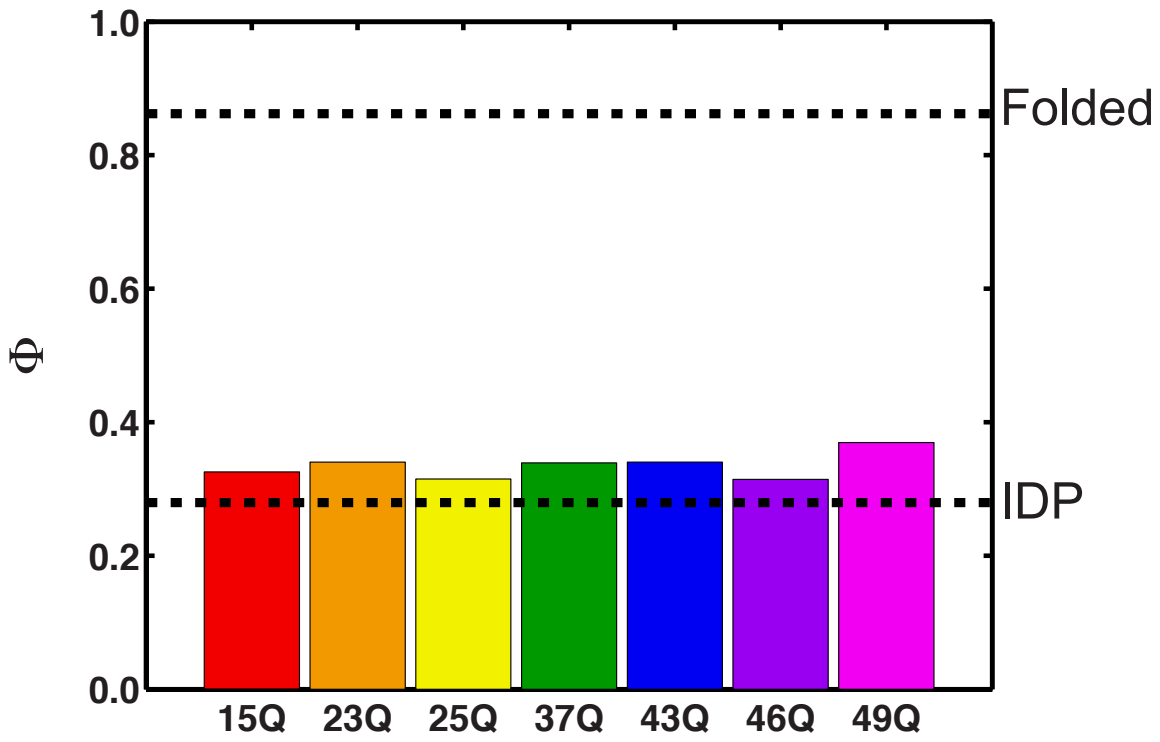

Fig S7. The degree of similarity, Φ, between simulated ensembles calculated over the C-terminal domain of Httex1 for all polyQ lengths. Here, frames in which any residue C-terminal to P_{11} was within 8 Å from any residue within the polyQ domain were filtered out for Httex1 25Q and 46Q. This filtering was used to mimic the observation that in the smFRET reweighted ensembles (15Q, 23Q, 37Q, 43Q, and 49Q) conformations in which the C-terminus interacts with the polyQ domain are down-weighted. Upon filtering, the 25Q and 46Q ensembles are less heterogeneous and more consistent with the other Httex1 constructs (compare with Fig 6c). Low Φ values correspond to high degrees of heterogeneity within the simulated ensembles. The upper dashed line corresponds to the Φ-value associated with a reference folded protein (Ntl9), whereas the lower dashed line corresponds to the Φ-value associated with a reference IDP (Ash1).