A toolkit for the scalable purification of huntingtin from eukaryotic expression systems

Rachel J. Harding[1], Peter Loppnau[1], Suzanne Ackloo[1], Ashley Hutchinson[1], Brittany Hunt[1], Alex Holehouse[2], Alma Seitova[1], Cheryl H. Arrowsmith[1]

[1]Structural Genomics Consortium, University of Toronto, MaRS South Tower, Suite 700, 101 College Street, Toronto, Ontario M5G 1L7, Canada.

[2]Department of Biomedical Engineering and Center for Biological Systems Engineering, Washington University in Saint Louis, Saint Louis, Missouri 63130, USA

Abstract:

The gene mutated in Huntington's disease (HD) patients encodes the 348 kDa huntingtin (HTT) protein. Pathogenic HD CAG-expansion mutations causes a polyglutamine (polyQ) tract at the N-terminus of the HTT protein to expand above a critical threshold of ~35 glutamine residues. The effect of HD mutations on HTT is not well understood, in part due to difficulties in carrying out biochemical studies of this large protein. To facilitate such studies, we have cloned expression constructs for the scalable production of HTT in multiple eukaryotic expression systems. Our set of HTT clones comprises both N and C-terminally FLAG-tagged HTT constructs with polyQ lengths corresponding to those of the general population, HD patients, juvenile HD patients as well as the more extreme polyQ expansions used in HD tissue and animal models. Milligram quantities of HTT may be purified in a two-step purification from either insect or mammalian cells using these constructs, producing pure samples of HTT protein amenable to numerous downstream studies.

Introduction:

Huntington's disease (HD) is a devastating inherited neurodegenerative disorder which causes a range of progressive behavioural, cognitive and physical symptoms. Incidence of HD varies in different parts of the world but HD is thought to affect between 0.42 to 17.2 per 100,000 of the population[1]. There are currently no disease-modifying therapies available for patients[2]. HD is hallmarked by an expansion of a CAG-trinucleotide repeat tract in exon 1 of the HTT gene above a critical threshold of ~35 CAG triplets[3,4], translating to a polyglutamine (polyQ) expansion in the extreme N-terminus of the HTT protein. PolyQ expanded HTT is thought to be responsible for the wide-ranging biochemical dysfunction observed in HD models and patients including proteostasis network impairment[5], transcription dysregulation[6], mitochondrial toxicity[7,8], cellular energy imbalance[9], synaptic dysfunction[10] and axonal transport impairment[8]. Although it is thought that HTT is likely a scaffold protein[11,12], the function of HTT, wildtype or polyQ expanded, is still incompletely understood.

Biochemical investigation of the role of HTT, in either the wildtype or the disease state, is often dependent on obtaining large amounts of pure HTT protein of different polyQ lengths. The HTT protein is 3144 amino acids long (assuming a polyQ stretch of 23 residues), a potentially daunting prospect for expression and purification given its size. Although a number of groups have published tools and methods by which full-length HTT might be expressed and purified from either insect[13–15] or mammalian cells[16]. However, the tools produced and shared with the wider community are often limited by the number of different polyQ lengths, the tag position or their tractability for large scale production. To date, the number of groups reporting experiments with purified HTT protein samples remains limited. Therefore, tools and detailed methods that will enable biochemical and biophysical studies of HTT by a larger number of researchers should accelerate our understanding of the function of this elusive protein.

Towards this end, we have cloned more than 25 pBacMam-based HTT constructs which allow expression of HTT protein through transient transfection of mammalian cells or viral transduction of insect or mammalian cells. Constructs have either N or C-terminal FLAG tags to assist in purification and yields of wildtype and polyglutamine expanded HTT protein using these systems are up to one milligram per litre of suspension culture of either insect or mammalian cells. The protein samples obtained from a simple two-step protocol are highly pure (> 90 % purity) and amenable to numerous downstream analyses and assays. We have analysed and assessed some of the biophysical properties of the huntingtin protein samples produced using this protein production system. This includes mapping post-translational modifications of the protein derived from insect cells, revealing similar modification motifs to mammalian derived samples. We also assessed monodispersity and thermal stability using size-exclusion chromatography in tandem with multi-angle light scattering (SEC-MALS) and differential static light scattering (DSLS) respectively.

Results:

1. Cloning of HTT expression constructs

Ligase-independent cloning (LIC) was used to clone the full-length HTT gene into the baculovirus transfer vector pBacMam2-DiEx-LIC (**Figure 1a**), for expression of proteins in insect cells as well as in mammalian cells. In addition to the sites for LIC, the vector contains a "stuffer" fragment that includes the SacB gene, allowing negative selection on 5 % sucrose, and a truncated VSVG fragment for pseudotyping of the baculovirus. As previously described for other HTT clones, randomly alternating CAG and CAA codons were used to encode the polyQ expansion to help maintain stability and integrity of the DNA sequence through various generations of vector propagation[17].

As a >10 kb gene with multiple repetitive sequence elements, HTT is non-trivial to subclone between different vectors. We first generated N and C-terminally FLAG-tagged pBacMam-DiEx-LIC HTT constructs lacking part of the exon 1 sequence. Using different polyQ length encoding exon 1 PCR generated cDNAs, our LIC cloning protocol generated a variety of different polyQ encoding HTT constructs due to the error-prone nature of the recombination step. By sequencing multiple colonies we identified HTT clones with a variety of polyQ lengths with both N- and C-terminal FLAG tags (**Figure 1b**). This entry vector serves as a valuable reagent to allow future generation of even more polyQ length HTT constructs. Additionally, by using a repetitive mix of codons for the polyQ expansion (CAG CAA CAG CAA CAA)$_n$, we expect improved polyQ stability over generations of plasmid, bacmid and baculovirus propagation compared to repetitive CAG codon tracts[18].

The resulting HTT open reading frames encoded within this series of constructs have either an N-terminal FLAG-octapeptide between the START methionine and the N-terminal methionine of the HTT amino acid sequence, or have the FLAG-octapeptide linked to the extreme C-terminus of HTT via a Gly-Gly-Ser-Gly linker (**Figure 1c**). As subtle changes to the exon 1 amino acid sequence of the HTT protein have been shown to give rise to changes to biophysical properties of the protein[19,20], the C-terminally FLAG-tagged constructs allow expression of a "clean" exon 1 sequence.

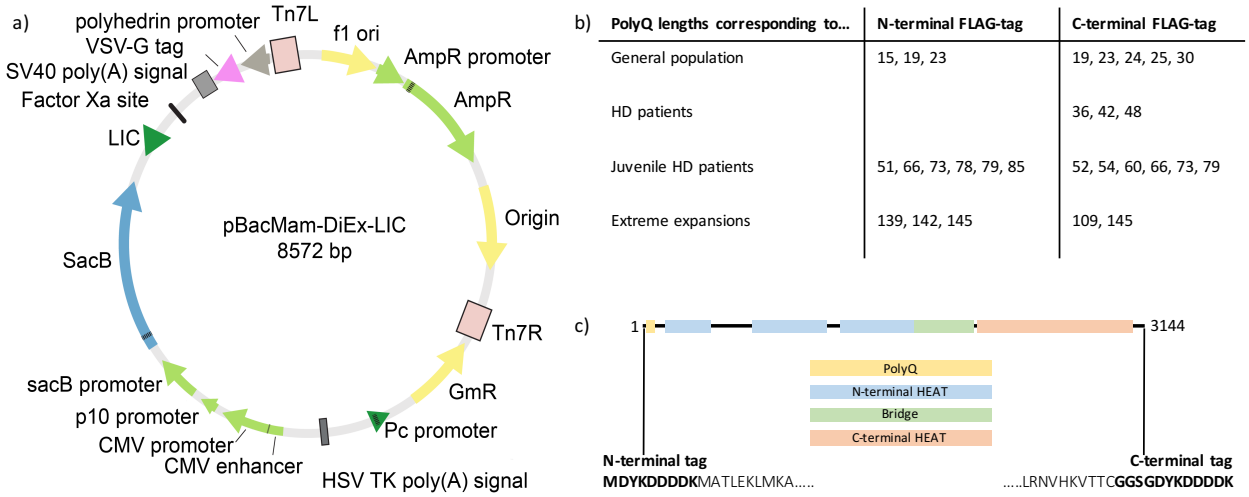

Figure 1. a) pBacMam vector map **b)** 28 HTT expression constructs with different polyQ lengths were generated with either N or C terminal FLAG-tags, **c)** FLAG-tags are appended to either end of the HTT expression construct with minimal additional sequence.

2. Versatile expression of HTT variants in insect cells or mammalian cells

The HTT BacMam expression constructs we have developed allow the expression of HTT protein by three different methods; baculovirus expression in insect cells, transient transfection in mammalian cells or transduction in mammalian cells (**Figure 2**). All three methods allow cell growth in suspension culture permitting rapid scaling of the culture volumes and thus scaling of the protein production as needed.

Irrespective of the expression system, HTT protein could be purified in a 2-step protocol as previously described[14] from cell lysates in a Tris-salt buffer system comprising first a FLAG pull-down step and followed by

size-exclusion chromatography using a Superose6 resin column (**Figure 2a**). Yields of the wildtype (Q23) purified HTT protein samples by the three expression methods can be as high as ~1.6 mg/L production in sf9 insect cell culture, to ~1 mg/L in transiently transfected mammalian cells and then ~0.4 mg/L in transduced mammalian cells when measured after FLAG pull-down. Comparisons of preparations of HTT Q23 samples with either N or C terminal FLAG tag did not show significant difference in yield. On the other hand, the yields of purified huntingtin decreased with increasing polyQ length. This is in part due to low overall expression of the longer polyQ huntingtin proteins as well as a shift in the ratio of the void, aggregated protein peak in For example, in insect cells Q42 yielded ~0.5 mg/L for Q42 in insect cell production whereas Q145 gives yields of <0.1 mg/L in insect cell production. Longer polyQ lengths were also generally found to be more variable in yield between productions. For all constructs in each expression system, the 2-step purification protocol yielded a protein sample which is ~90-95 % pure by Coomassie stained SDS-PAGE analysis (**Figure 2b**).

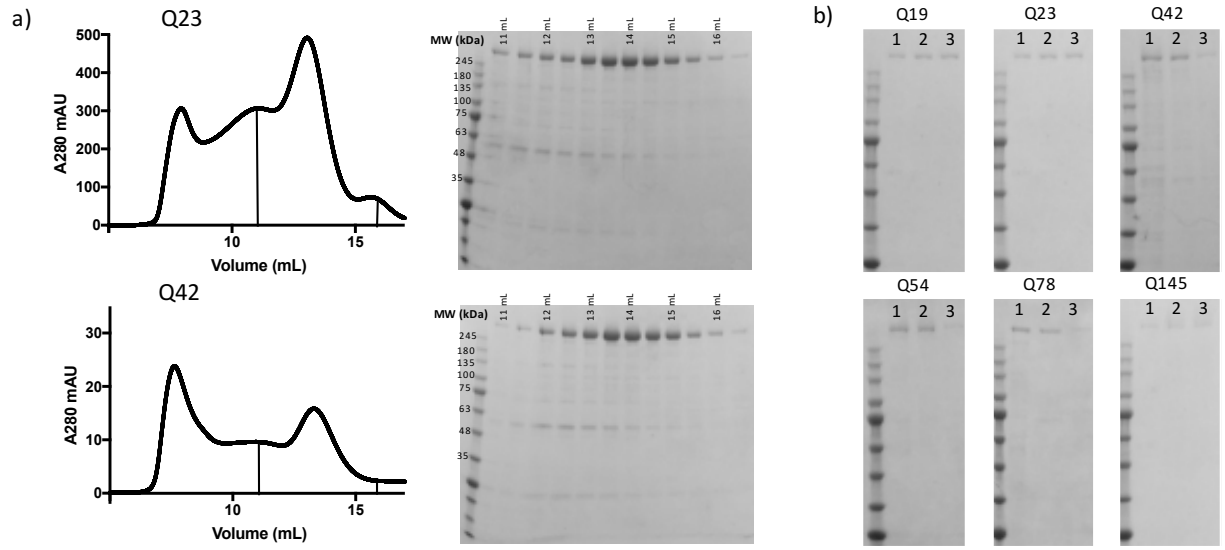

Figure 2. a) Analysis of C-terminally FLAG-tagged samples of insect sf9 cell expressed HTT Q23 and HTT Q42 by size-exclusion chromatography using Superose6 10/300 GL column and Coomassie-stained 4-20 % SDS-PAGE analysis of size-exclusion chromatography fractions (0.4 mL volume) spanning 11-16 mL as marked on the elution profiles, **b)** Coomassie-stained 4-20 % SDS-PAGE analysis of C-terminally FLAG-tagged samples of HTT of different polyQ lengths from lane 1 - baculovirus expression in sf9 insect cells, lane 2 - transient transfection in mammalian EXPI293F cells or lane 3 - transduction in mammalian EXPI293F cells,

3. HTT expressed in insect cells retains phosphorylation patterns observed in HTT from mammalian cells

Post-translational modification of HTT is well described for samples derived from various mammalian cell systems and in some detail for samples derived from HD patient post-mortem brain tissue[21]. However, whether these PTMs are conserved in insect cell produced material is not known. We assessed HTT Q23 insect cell samples for their post-translational modifications for comparison with previously described mammalian derived samples. The purified protein was subjected to bottom-up proteomics. This analysis provided the sequence coverage and an assessment of the landscape of post-translation modifications. In an effort to 'cover' as much of the sequence including regions comprising few lysine (K) and arginine (R) residues, for example, the amino acids 1-80 and amino acids 501-619, we employed five proteases with complementary and non-specific cleavage specificity. These are trypsin, lysargiNase[22], pepsin, wild type α-lytic protease (WaLP) and M190A α-lytic protease (MaLP)[23]. Trypsin and lysargiNase cleave at the C- and N-terminal, respectively, of K and R residues thus yielding complementary (or mirror-image) peptides. WaLP and MaLP preferentially cleave at aliphatic amino acids and are attractive for probing the K- and R-poor regions of the protein.

Good coverage of the huntingtin protein was achieved in a series of digestion and peptide identification mass spectrometry experiments, mapping up to 65 % of the huntingtin amino acid sequence. 10 different

phosphorylation motifs were found in these experiments, 10 of which have been described in the published literature before as well as an additional phosphorylation modifications at S719 and T803.

Disorder analysis of the huntingtin protein sequence with Classification of Intrinsically Disordered Ensemble Regions (CIDER)[24] in combination with analysis of the recently published near-atomic resolution cryo-electron microscopy structure of huntingtin in complex with HAP40[25] (**Figure 3**), reveals that most of the phosphorylation sites are within unstructured regions of the protein structure as described previously[21]. S719 and T803 are both resolved in the structure in the N-term HEAT domain and both are surface residues. T803 is located on the external surface of the N-terminal helical solenoid structure whereas S719, whilst solvent exposed, is on the internal surface of the same solenoidal structure. S719 has been described for HEK293 expressed huntingtin samples and the surface environment nature of both S719 and T803 would suggest that these are plausible phosphorylation sites. This would suggest that the phosphorylation of HTT in insect cells is not entirely aberrant but physiologically relevant for the human protein sequence.

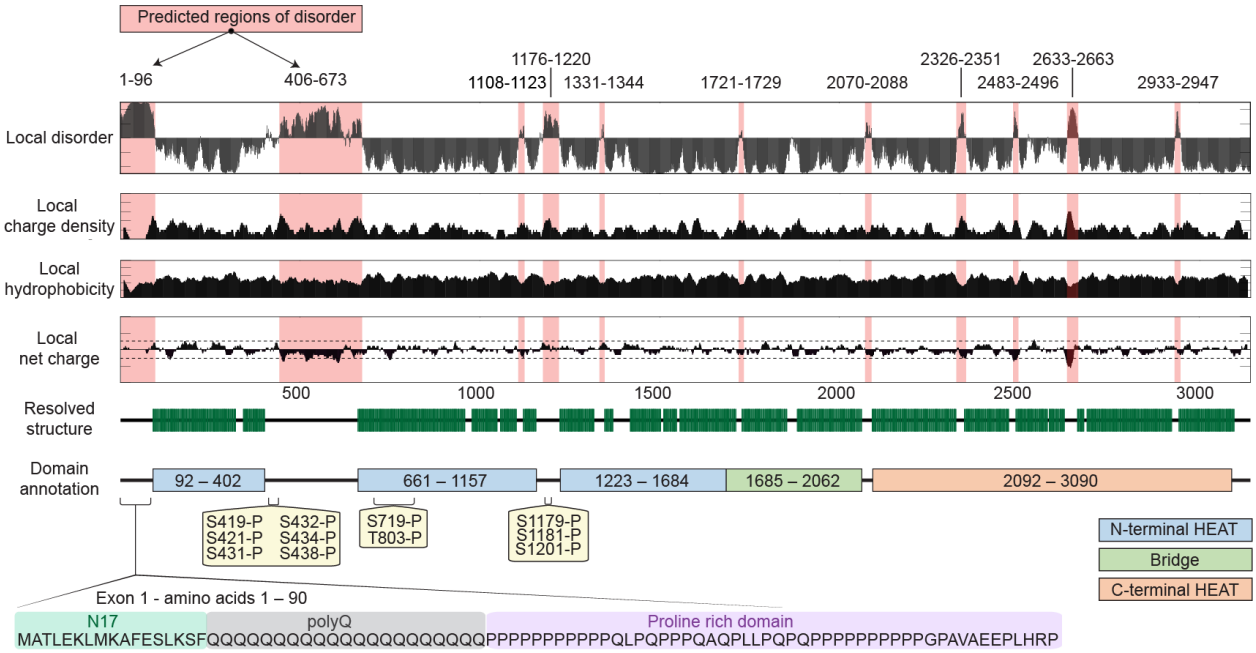

Figure 3. Sequence analysis of Huntingtin complemented by structural data reveals a large modular protein connected by flexible unstructured linkers with numerous short unstructured loops protruding from the ordered domains. Local disorder, charge density, hydrophobicity and net charge are based on sequence analysis, while resolved structure and domain annotation from cryo-electron microscopy[25]. Phosphorylation motifs of insect-cell derived HTT Q23 mapped by peptide mass spectrometry are annotated and are mainly found in unstructured regions of the protein

4. Analysis of huntingtin protein properties by biophysical methods

Gel filtration of HTT protein derived from insect sf9 or mammalian EXPI293F cells using a Superose6 10/300 GL column gives a characteristic elution profile, with a void-aggregate peak followed by peaks previously attributed as being dimer and then monomer species of huntingtin based on column standards[14–16]. The ratio of void-to-monomer peak size seems to increase with increased polyQ length, suggesting that increasing polyQ length promotes self-aggregation. To further understand this tendency for self-aggregation, we pooled monomer fractions of HTT, and concentrated the sample. Rerunning this sample on the column, gives an elution profile with a major peak corresponding to monomeric species, but also has significant amounts of dimer and void-aggregate species. This suggests an equilibrium in which dimer and higher order aggregates accumulate with time (**Figure 4a**). C-terminal FLAG-tagged HTT Q23 samples purified from sf9 cells were also analysed by size-exclusion chromatography in tandem with multi-angle light scattering (SEC-MALS) which allows calculation of the in-solution protein mass. These analyses reveal a monodisperse and monomeric sample at lower concentrations (~ 0.3

mg/mL) but a mixture of HTT monomer and dimer as the concentration of HTT is observed at increased (~ 1 mg/mL) protein concentrations. (**Figure 4b**).

C-terminal FLAG-tagged HTT Q23 and Q42 samples purified from sf9 cells were analysed by differential static light scattering (DSLS) over a temperature gradient from 25-85 ˚C to assess thermal stability. Irrespective of polyQ-length, the HTT samples were stable up to ~50 ˚C with a transition to higher molecular weight aggregates was observed over the course of ~30 ˚C from 50-80 ˚C (**Figure 4c**). This slow transition likely reflects the helical solenoid structure of the HTT protein. Temperature of aggregation (T_{agg}) values[26] for each HTT sample were calculated at ~ 70 ˚C.

The recent cryo-electron microscopy structure of HTT in complex with HAP40 reveals a bi-lobed structure which is stabilised through binding HAP40[25]. Our samples are pure and therefore lack a stabilising partner such as HAP40 which was described as critical for producing a conformationally homogenous HTT sample amenable to cryo-electron microscopy structure determination. Perhaps the mobility of N-HEAT and C-HEAT domains with respect to each other accounts for the broad and overlapping elution peaks observed in gel filtration analyses as well as the tendency for self-association when HTT samples are analysed at higher protein concentrations.

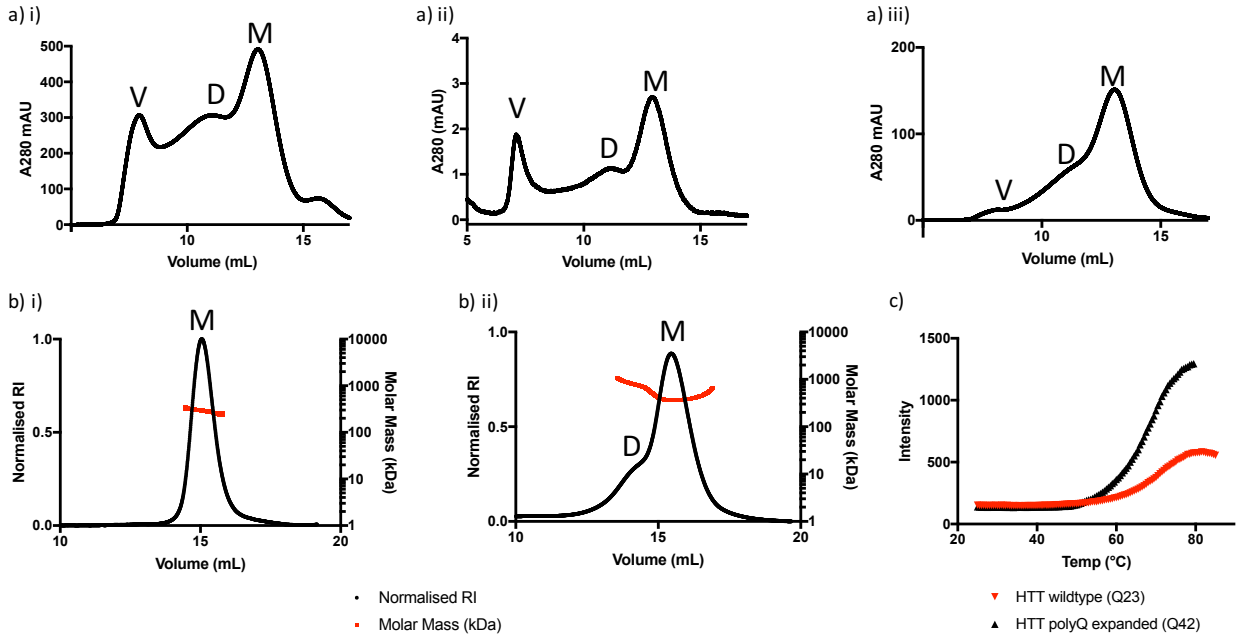

Figure 4. a) Analytical Superose6 10/300 GL gel filtration profiles of C-terminal FLAG-tagged HTT Q23 samples expressed in **i)** insect sf9 cells and **ii)** EXPI293F cells. Concentrating the monomer peak (M) from **i)** and rerunning the sample on the column **iii)** gives a predominantly monomeric sample but dimer (D) and void aggregate (V) are still present **b)** SEC-MALS analysis of HTT at **i)** 0.3 mg/mL and **ii)** 1 mg/mL shows a monodisperse sample of monomeric HTT at lower concentrations and a predominantly monomeric sample at higher concentrations, with some HTT dimer present, **c)** DSLS analysis of the HTT samples shows high thermal stability with a slow transition to an aggregated state, probably reflecting the helical solenoid structure of the HTT protein.

Discussion

We have generated a resource of more than 25 HTT expression constructs which allow the generation of purified HTT samples of different polyQ lengths and tags from three different expression systems. The constructs allow the purification of milligram quantities of wild-type HTT protein from both insect and mammalian cells as well as substantial production of various polyQ expanded HTT species. The precise molecular function of unexpanded huntingtin remains elusive, so it is unclear how the polyQ expansion may alter the huntingtin protein sufficiently to give rise to the wide-ranging biochemical dysfunction observed in HD models and patients. These reagents and the accompanying methods and validation for the production of HTT protein will provide an enabling framework for future research requiring purified HTT of wide ranging polyQ lengths.

Materials and Methods

Cloning of HTT expression constructs

HTT expression constructs were assembled in the mammalian/insect cell vector pBacMam2-DiEx-LIC in two steps. First, entry vectors for N-terminal Flag tagged and C-terminal FLAG-tagged HTT (amino acids 1-3144) were constructed without the polyQ regions, amino acids 7-85. PCR products encoding wildtype HTT were amplified from cDNA (Kazusa Clone FHC15881) using primers N_int_FWD (ttaagaaggagatatactATGGACTACAAAGACGATGACGACAAGATGGCGACCCTGGAAAAGCgctGACCTTAGTCGCTAAcctgca GGAGCCGCTGCACCGACCAAAG)/ N_int_REV (gattggaagtagaggttttaGCAGGTGGTGACCTTGTGGAC) for the N-teminal FLAG-tagged HTT and C_int_FWD (ttaagaaggagatatactATGGCGACCCTGGAAAAGCgctGACCTTAGTCGCTAAcctgcaGGAGCCGCTGCACCGACCAAAG) / C_int_REV (gattggaagtagaggttttaCTTGTCGTCATCGTCTTTGTAGTCaccgcttccaccGCAGGTGGTGACCTTGTGGAC) for the C-terminal FLAG-tagged HTT. All PCR products were inserted using the In-Fusion cloning kit (Clontech) into the pBacMam2-DiEx-LIC which had been linearized with BfuAI. Second, synthetic Poly Q regions were inserted into the intermediate plasmids using the In-Fusion cloning kit. The Poly Q regions were PCR amplified using the primers polyQP_Fwd (ATGGCGACCCTGGAAAAGCTG) / polyQP_Rev (TGGTCGGTGCAGCGGCTCCTC) from template plasmids CH00007 (Q23), CH00008 (Q73), and CH00065 (Q145) (all from CHDI-Coriell Institute biorepository). PCR products were inserted into the intermediate vectors which had been linerarized with Afe1 and Sbf1. Upon screening the assembled HTT expression constructs we found that our cloning method generated a range of polyQ lengths. We selected constructs with poly Q lengths from Q15 to Q145. The HTT coding sequences of intermediate and final expression constructs were confirmed by DNA sequencing.

HTT protein expression

The recombinant transfer vectors pBacMam-DiEx-LIC-HTT were transformed into DH10Bac E. coli cells (Invitrogen, Bac-to-Bac System) to generate recombinant Bacmid DNA. Sf9 cells (Invitrogen) were transfected with Bacmid DNA using jetPRIME® transfection reagent (PolyPlus Transfection, Cat. 89129-924), and recombinant baculovirus particles were recovered. The recombinant virus titre was sequentially amplified from P1 to P3 viral stocks for protein production in the sf9 insect cells and EXPI293F mammalian cells.

Baculovirus expression of HTT in Insect Cells: Sf9 cells at a density of ~4.5 million cells per mL were infected with 8 mL of P3 recombinant baculovirus and grown at 100 rpm and 27°C. HyQ SFX insect serum medium containing 10 µg/mL gentamicin was used as the culture medium. Infected cells were harvested when viability dropped to 80%–85%, normally after ~48 h.

Transduction of HTT in Mammalian Cells: EXPI293F cells (Thermo Fisher, Cat. A14527) were maintained in EXPI293 Expression Medium (Thermo Fisher, Cat. A1435102) in a humidified 8 % CO_2 incubator at 37 °C and 125 rpm. Cells were transduced at a density of 2-3 million cells per mL culture. The transduction used recombinant baculoviruses of HTT constructs generated by transfecting sf9 cells using Transfer vector pBacMam2-DiEx-LIC and JetPRIME® transfection reagent (Cat. 89129-924). The volume of the virus added into the cells was at ratio as 6% of the total volume of the production. Infected cells were harvested after 7-10 days post-transduction depending on cell viability.

Transient Transfection of HTT in Mammalian Cells: EXPI293F cells (Thermo Fisher, Cat. A14527) were maintained in EXPI293 Expression Medium (Thermo Fisher, Cat. A1435102) in a humidified 8 % CO_2 incubator at 37 °C and 125 rpm. Cells were transfected at a density of 2-3 million cells per mL culture. FectoPRO® transfection reagent (VWR, Cat. 116-001) and plasmid pBacMam-DiEx-LIC-HTT DNA were separately diluted in serum-free OptiMEM complexation medium (Thermo Fisher, Cat. 31985062) at 10% of the total production volume in a ratio of 1 µg DNA to 1.2 µL FectoPro to 0.5 µL Booster per mL cell culture. After 5 min of incubation at room temperature, the transfection mixtures were combined and incubated for an additional 20 min. The FectoPRO® transfection reagent-DNA-OptiMEM mixture was then added to cells with an addition of FectoPRO® Booster in a ratio of 1 µg DNA to 1.2 µL FectoPro to 0.5 µL Booster per mL cell culture. The transfected cultures of EXPI293 cells were harvested after 72- 96 h post-transfection dependent cells density and viability.

HTT protein purification

The same protocol was used to purify HTT from insect and mammalian cell culture. Cell cultures were harvested by centrifugation at 4000 rpm, 20 mins, 4 ˚C (Beckman JLA 8.1000), washed in pre-chilled PBS and resuspended in 20

cell paste volumes of preparation buffer (50 mM Tris pH 8, 500 mM NaCl) and stored at -80 ˚C prior to purification. Cell suspensions were thawed and diluted to at least 50 times the cell paste volumes with prechilled buffer and supplemented with 1 mM PMSF, 1 mM benzamidine-HCl and 20 U/mL benzonase. NB: two freeze-thaw cycles of cell suspensions were found to be sufficient for cell lysis. The lysate was clarified by centrifugation at 14,000 rpm, 1 h, 4 ˚C (Beckman JLA 16.2500) and then bound to 0.1 cell paste volumes of anti-FLAG resin (Sigma M2) at 4 ˚C with rocking for 2 hours. Anti-FLAG resin was washed twice with the 100 cell paste volumes of buffer. HTT protein was eluted with 1 cell paste volume of buffer supplemented with 250 µg/mL 3xFLAG peptide (Chempep) run twice over the anti-FLAG resin. Residual HTT protein was washed from the beads with 0.5 cell paste volume of buffer. The sample was spin concentrated with MWCO 100,000. Depending on the protein yield, the sample was run as one or more sample runs on Superose 6 10/300 GL column in 50 mM Tris pH 8, 300 mM NaCl at 0.4 mL/min ensuring no more than 2 mg of protein was applied per run to minimize protein aggregation. Analytical runs of purified HTT samples were run using the same protocol.

HTT mass spectrometry

All data was acquired on an Agilent 1260 capillary HPLC system coupled to an Agilent Q-TOF 6545 mass spectrometer via the Dual Agilent Jetstream ion source.

Intact protein: The purified, recombinant protein was analyzed on a Poroshell 300SB-C3 column (1x75 mm with 5 micron particles) at a flow-rate of 500 microliters/min using a gradient of 5% solvent B (96:4 ACN:H_2O with 1% formic acid) to 95% solvent B in 4 minutes. The column temperature was held at 70 ˚C.

Bottom-up proteomics for sequence coverage and PTM analysis: Proteins were processed according to established protocols[27]. Briefly, proteins were reduced with DTT (10 mM final concentration) for 30 minutes at room temperature, alkylated with iodoacetamide (55 mM final concentration) for 30 minutes at room temperature, and incubated with trypsin (6 µL, 0.2 mg/mL) overnight at 37 °C. The digests acidified to p2 in H, desalted on-column (by diverting the first two minutes to waste), before analysis. Peptides were separated on a C18 Advance BioPeptide column (2.1x150 mm 2.7 micron particles) at a flow rate of 400 microliters/min and an operating pressure of 4,700 psi. Peptides were eluted using a gradient from 100% solvent A (98:2 H_2O:ACN with 1% formic acid) to 50% B (96:4 ACN:H2O with 1% formic acid) in 80 minutes. Mass spectra were acquired from m/z 300–1700 at a rate of 8 spectra per second. The tandem mass spectra were acquired in automated MS/MS mode from m/z 100-1500 with an acquisition rate of 3 spectra per second. The top ten precursors were selected and sorted by abundance only. Collision-induced dissociation was done using all ions at 4*(m/z)/100-1 and -5.

Data analysis: Raw data was processed using PEAKS Studio 8.5 (build 20171002) and the reference complete human proteome FASTA file (Uniprot). Cysteine carbamidomethylation was selected as a fixed modification, and methionine oxidation and N/Q deamidation as variable modifications. A minimum peptide length of five, a maximum of three missed cleavage sites, and a maximum of three labelled amino acids per peptide were employed.

HTT sequence disorder prediction

Disorder prediction was performed using IUPred[28,29] A threshold of 0.5 was used to define disordered or ordered regions, with predicted disordered regions shaded in light red. Further sequence analysis of Huntingtin was performed using the sequence analysis tool local CIDER[24]. Hydrophobicity was calculated using a normalized Kyte-Doolitle scale[30]. Resolved structure and domain annotations based on the solved cryo-EM structure of Huntingtin in complex with HAP40[25].

Differential static light scattering (DSLS)

The thermal stability of HTT Q23 and Q42 samples were analysed by DSLS using StarGazer. HTT proteins at 1 mg/mL in 50 mM Tris-HCl pH 8.0, 300 mM NaCl were heated from 20 °C to 85 °C. Protein aggregation was monitored using a CCD camera. The temperature of aggregation (T_{agg}) was analysed and fitted as described previously[26].

Size Exclusion Chromatography Multi-Angle Light Scattering (SEC-MALS)

The absolute molar masses and mass distributions of purified protein samples of HTT Q23 with C-terminal FLAG tag at 0.3 and 1 mg/mL were determined using SEC-MALS. Samples were injected through a Superose 6 10/300 GL column (GE Healthcare Life Sciences) equilibrated in 50 mM Tris-HCl pH 8.0, 300 mM NaCl followed in-line by a

Dawn Heleos-II light scattering detector (Wyatt Technologies) and a 2414 refractive index detector (Waters). Molecular mass calculations were performed using ASTRA 6.1.1.17 (Wyatt Technologies) assuming a dn/dc value of 0.185 ml/g.

References

1. Kay, C., Hayden, M. R. & Leavitt, B. R. Chapter 3 - Epidemiology of Huntington disease. in *Handbook of Clinical Neurology* (eds. Feigin, A. S. & Anderson, K. E.) **144,** 31–46 (Elsevier, 2017).
2. Ross, C. A. *et al.* Huntington disease: natural history, biomarkers and prospects for therapeutics. *Nat. Rev. Neurol.* **10,** 204–216 (2014).
3. Kay, C., Fisher, E. R. & Hayden, M. R. *Huntington's Disease Chapter 7 (eds Bates, G. P., Tabrizi, S. J. & Jones, L.)*. (Oxford University Press, 2014).
4. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell* **72,** 971–983 (1993).
5. Kim, M. *et al.* Mutant huntingtin expression in clonal striatal cells: dissociation of inclusion formation and neuronal survival by caspase inhibition. *J. Neurosci. Off. J. Soc. Neurosci.* **19,** 964–973 (1999).
6. Seredenina, T. & Luthi-Carter, R. What have we learned from gene expression profiles in Huntington's disease? *Neurobiol. Dis.* **45,** 83–98 (2012).
7. Johri, A., Chandra, A. & Beal, M. F. PGC-1α, mitochondrial dysfunction, and Huntington's disease. *Free Radic. Biol. Med.* **62,** 37–46 (2013).
8. Reddy, P. H. & Shirendeb, U. P. Mutant huntingtin, abnormal mitochondrial dynamics, defective axonal transport of mitochondria, and selective synaptic degeneration in Huntington's disease. *Biochim. Biophys. Acta* **1822,** 101–110 (2012).
9. Ju, T.-C., Lin, Y.-S. & Chern, Y. Energy dysfunction in Huntington's disease: insights from PGC-1α, AMPK, and CKB. *Cell. Mol. Life Sci. CMLS* **69,** 4107–4120 (2012).
10. Nithianantharajah, J. & Hannan, A. J. Dysregulation of synaptic proteins, dendritic spine abnormalities and pathological plasticity of synapses as experience-dependent mediators of cognitive and psychiatric symptoms in Huntington's disease. *Neuroscience* **251,** 66–74 (2013).
11. Zheng, Z. & Diamond, M. I. Huntington disease and the huntingtin protein. *Prog. Mol. Biol. Transl. Sci.* **107,** 189–214 (2012).

12. Rui, Y.-N. *et al.* Huntingtin functions as a scaffold for selective macroautophagy. *Nat. Cell Biol.* **17,** 262–275 (2015).

13. Vijayvargia, R. *et al.* Huntingtin's spherical solenoid structure enables polyglutamine tract-dependent modulation of its structure and function. *eLife* **5,** e11184 (2016).

14. Seong, I. S. *et al.* Huntingtin facilitates polycomb repressive complex 2. *Hum. Mol. Genet.* **19,** 573–583 (2010).

15. Li, W., Serpell, L. C., Carter, W. J., Rubinsztein, D. C. & Huntington, J. A. Expression and Characterization of Full-length Human Huntingtin, an Elongated HEAT Repeat Protein. *J. Biol. Chem.* **281,** 15916–15922 (2006).

16. Huang, B. *et al.* Scalable Production in Human Cells and Biochemical Characterization of Full-Length Normal and Mutant Huntingtin. *PLOS ONE* **10,** e0121055 (2015).

17. Figura, G., Koscianska, E. & Krzyzosiak, W. J. In Vitro Expansion of CAG, CAA, and Mixed CAG/CAA Repeats. *Int. J. Mol. Sci.* **16,** 18741–18751 (2015).

18. Michalik, A., Kazantsev, A. & Van Broeckhoven, C. Method to introduce stable, expanded, polyglutamine-encoding CAG/CAA trinucleotide repeats into CAG repeat-containing genes. *BioTechniques* **31,** 250–252, 254 (2001).

19. Cariulo, C. *et al.* Phosphorylation of huntingtin at residue T3 is decreased in Huntington's disease and modulates mutant huntingtin protein conformation. *Proc. Natl. Acad. Sci. U. S. A.* **114,** E10809–E10818 (2017).

20. Vieweg, S., Ansaloni, A., Wang, Z.-M., Warner, J. B. & Lashuel, H. A. An Intein-based Strategy for the Production of Tag-free Huntingtin Exon 1 Proteins Enables New Insights into the Polyglutamine Dependence of Httex1 Aggregation and Fibril Formation. *J. Biol. Chem.* **291,** 12074–12086 (2016).

21. Ratovitski, T. *et al.* Post-Translational Modifications (PTMs), Identified on Endogenous Huntingtin, Cluster within Proteolytic Domains between HEAT Repeats. *J. Proteome Res.* (2017). doi:10.1021/acs.jproteome.6b00991

22. Huesgen, P. F. *et al.* LysargiNase mirrors trypsin for protein C-terminal and methylation-site identification. *Nat. Methods* **12,** 55–58 (2015).

23. Meyer, J. G. *et al.* Expanding proteome coverage with orthogonal-specificity α-lytic proteases. *Mol. Cell. Proteomics MCP* **13,** 823–835 (2014).

24. Holehouse, A. S., Das, R. K., Ahad, J. N., Richardson, M. O. G. & Pappu, R. V. CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophys. J.* **112,** 16–21 (2017).
25. Guo, Q. *et al.* The cryo-electron microscopy structure of huntingtin. *Nature* (2018). doi:10.1038/nature25502
26. Senisterra, G. A. *et al.* Screening for ligands using a generic and high-throughput light-scattering-based assay. *J. Biomol. Screen.* **11,** 940–948 (2006).
27. Gundry, R. L. *et al.* Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-Up Proteomics Workflow. in *Current Protocols in Molecular Biology* (John Wiley & Sons, Inc., 2001). doi:10.1002/0471142727.mb1025s88
28. Dosztányi, Z., Csizmok, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinforma. Oxf. Engl.* **21,** 3433–3434 (2005).
29. Dosztányi, Z., Csizmók, V., Tompa, P. & Simon, I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol.* **347,** 827–839 (2005).
30. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105–132 (1982).