

# 1 Expanding the Huntington's disease research toolbox; validated subdomain protein 2 constructs for biochemical and structural investigation of huntingtin

3 Matthew G. Alteen<sup>1</sup>, Justin C. Deme<sup>2</sup>, Claudia P. Alvarez<sup>1,#</sup>, Peter Loppnau<sup>1</sup>, Ashley Hutchinson<sup>1</sup>, Alma Seitova<sup>1</sup>,  
4 Renu Chandrasekaran<sup>1</sup>, Eduardo Silva Ramos<sup>3</sup>, Christopher Secker<sup>3</sup>, Mona Alqazzaz<sup>4</sup>, Erich E. Wanker<sup>3</sup>, Susan M.  
5 Lea<sup>2</sup>, Cheryl H Arrowsmith<sup>1,4\*</sup>, Rachel J. Harding<sup>1\*</sup>

6 \*Corresponding Authors

7 <sup>1</sup> Structural Genomics Consortium, University of Toronto, Toronto, Ontario, M5G 1L7, Canada

8 <sup>2</sup> Center for Structural Biology, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702,  
9 USA.

10 <sup>3</sup> Neuroproteomics, Max Delbrück Center for Molecular Medicine, Robert-Rössle-Str. 10, 13125, Berlin,  
11 Germany.

12 <sup>4</sup> Princess Margaret Cancer Centre and Department of Medical Biophysics, University of Toronto, Toronto,  
13 Ontario, Canada

14 #Author has updated affiliation: SCIEX, 71 Four Valley Dr, Vaughan, Ontario, L4K 4V8, Canada

## 15 Abstract

16 Huntington's disease is characterised by CAG expansion in the huntingtin gene above a critical threshold of ~35  
17 repeats, resulting in polyglutamine expansion of the huntingtin protein (HTT). The biological role of wildtype  
18 HTT and the associated mechanisms of disease pathology caused by expanded HTT remain incompletely  
19 understood, in part, due to challenges characterising interactions between HTT and putative binding partners.  
20 Here we describe a biochemical toolkit of rationally designed, high-quality recombinant HTT subdomains; one  
21 spanning the N-terminal HEAT and bridge domains (NTD) and the second spanning the C-terminal HEAT domain  
22 (CTD). Using biophysical methods and cryo-electron microscopy, we show these smaller subdomains are  
23 natively folded and can associate to reconstitute a functional full-length HTT structure capable of forming a  
24 near native-like complex with 40 kDa HTT-associated protein (HAP40). We report biotin-tagged variants of  
25 these subdomains, as well as full-length HTT, that permit immobilisation of each protein for quantitative  
26 biophysical assays without impacting protein quality. We demonstrate the CTD alone can form a stable  
27 complex when co-expressed with HAP40, which can be structurally resolved. The CTD-HAP40 complex binds  
28 the NTD, with a dissociation constant of approximately 10 nM as measured by bio-layer interferometry. We  
29 validate the interaction between the CTD and HAP40 using a luciferase two-hybrid assay and use subdomain  
30 constructs to demonstrate their respective stabilization of HAP40 in cells. These open-source biochemical tools  
31 will enable the wider HD community to study fundamental HTT biology, discover new macromolecular or small-  
32 molecule binding partners and map interaction sites across this very large protein.

## 33 Keywords

34 Huntington disease, neurodegeneration, protein purification, protein structure, cryo-electron microscopy,  
35 protein-protein interaction.

36 \*To whom correspondence should be addressed:

Rachel J. Harding  
Tel.: +1 (416) 946-3795  
Email: Rachel.Harding@utoronto.ca

Cheryl H. Arrowsmith  
Tel.: +1 (416) 946-0881  
Email: Cheryl.Arrowsmith@uhnresearch.ca

## 37 Introduction

38 Huntington's Disease (HD) is an autosomal dominant neurodegenerative disease caused by mutation of the  
39 Huntingtin (HTT) gene, resulting in a CAG repeat expansion above a critical threshold of approximately ~35  
40 repeats (1). HD patients suffer debilitating cognitive, motor and psychiatric symptoms and have a life  
41 expectancy of ~18 years after symptom onset (McColgan and Tabrizi, 2018; Saudou and Humbert, 2016). CAG  
42 expansion results in an expanded polyglutamine (polyQ) tract near the N-terminus of the encoded protein  
43 huntingtin (HTT), resulting in an aberrantly functioning form of the protein (Jimenez-Sanchez et al., 2017;  
44 McColgan and Tabrizi, 2018). HTT has been hypothesised to act as a scaffolding protein, governing numerous  
45 cellular functions including axonal transport (Gunawardena et al., 2003; Vitet et al., 2020), transcriptional  
46 regulation (Ng et al., 2013) and proteostasis regulation by mediating protein-protein interactions (Greco et al.,  
47 2022; Rui et al., 2015). Despite identification of the *HTT* gene almost thirty years ago ("A novel gene containing  
48 a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's  
49 Disease Collaborative Research Group," 1993), there are currently no disease-modifying therapies available, in  
50 part due to limited understanding of HTT function, or how these functions might be modulated by  
51 polyglutamine expansion. There is therefore a pressing need for the development of chemical and biochemical  
52 tools that can be used to illuminate the molecular details of the biological functions of HTT and enable  
53 therapeutic strategies to combat the pathology of expanded HTT.

54 The structure of HTT has been solved by cryo-electron microscopy (Guo et al., 2018; Harding et al., 2021;  
55 Huang et al., 2021a), in complex with 40 kDa huntingtin-associated protein (HAP40) (Peters and Ross, 2001),  
56 forming a ~389 kDa multidomain complex. HTT is primarily composed of HEAT repeats (huntingtin, elongation  
57 factor 3, protein phosphatase 2a, and yeast kinase TOR1) (Andrade et al., 2001) that form a large solenoid-like  
58 structure at the N-terminal region of the protein as well as a more compact C-terminal region. The N-HEAT and  
59 C-HEAT domains are connected by a bridge domain, also composed of HEAT repeats, with HAP40 sandwiched  
60 between these regions (**Figure 1A**). HTT is reported to interact with over 3000 proteins (Culver et al., 2012;  
61 Greco et al., 2022; Ratovitski et al., 2012; Wanker et al., 2019). The similarity of the HTT-HEAT motifs to related  
62 armadillo repeats, which are known to mediate PPIs, may provide a structural basis for HTT interaction  
63 network. Additionally, HTT contains many large, disordered regions, including most of exon 1 which contains  
64 the polyQ region. Importantly, these disordered regions have been shown to contain post-translational  
65 modifications (PTMs) and may mediate interactions with protein binding partners (Ratovitski et al., 2017).  
66 HAP40 remains the only biochemically and structurally validated interaction partner (Guo et al., 2018; Harding  
67 et al., 2021). Interestingly, while HTT has been recombinantly expressed as an apo protein, efforts to  
68 recombinantly express HAP40 on its own have not been successful, suggesting that HTT is required for  
69 expression and stabilization of HAP40 structure (Guo et al., 2018; Harding et al., 2021, 2019). This observation  
70 has been supported by genetic knockdowns of HTT that show a corresponding decrease in HAP40 expression  
71 (Harding et al., 2021; Huang et al., 2021b).

72 We previously reported a toolkit of resources for the scalable production of high-purity HTT and HTT-HAP40  
73 protein samples from various eukaryotic expression systems (Harding et al., 2019). Milligram quantities of  
74 samples can be generated with various Q-lengths as well as an exon 1 deleted ( $\Delta$  exon 1) form of the protein  
75 that are suitable for numerous *in vitro* assays. This expression platform has enabled high-resolution structural  
76 analysis of the HTT-HAP40 complex using cryo-EM, and associated biophysical studies such as small-angle X-ray  
77 scattering, allowing detailed examination of the structural organisation and potential functional roles of this  
78 large multidomain complex (Harding et al., 2021).

79 Nevertheless, the large size of the protein and its associated HAP40 complex presents challenges for many  
80 downstream applications. For example, the analysis of protein-ligand interactions using surface plasmon

## Toolkit of HTT subdomain constructs

81 resonance (SPR), which relies on optical methods to detect mass changes on an immobilised surface, suffer  
82 from decreasing signal-to-noise with increasing mass ratio between ligand and analyte molecules (Mitchell,  
83 2010). Smaller subdomain fragments of HTT could serve as useful biochemical tools for the study of HTT  
84 function and its interaction with other proteins, permitting a wider array of *in vitro* biochemical and biophysical  
85 assays. Additionally, well-defined subdomain fragments could help clarify the relative contributions of each  
86 domain in stabilizing HTT-HAP40 interaction, a potential therapeutic target for HD. HTT fragment constructs  
87 have been reported in the literature prior to the determination of the cryoEM structure, such as a commonly  
88 used fragment encompassing the exon1 region and a portion of the N-HEAT domain (HTT aa. 1-586) (Atwal et  
89 al., 2011; Nath et al., 2015). However, these constructs frequently have start and stop sites in the middle of  
90 HEAT domains or have boundaries in the middle of extensive disordered regions, which likely result in a non-  
91 native protein product. To date, no robust biophysical or structural validation of any such construct has been  
92 published.

93 In this study, we report the cloning, expression, purification and validation of highly pure recombinant HTT  
94 subdomain proteins and the open-source tools required to generate these samples. We also report the  
95 generation of full-length HTT and HTT subdomain constructs with biotinylated tags at both the N-terminus and  
96 C-terminus. We show that constructs encompassing the N-HEAT and bridge domains (NTD, aa. 97-2069) as well  
97 as the C-HEAT domain (CTD, aa. 2095-3138), can be stably expressed in eukaryotic cell culture and display good  
98 biophysical properties. We demonstrate that these subdomains can be combined *in vitro* or co-expressed to  
99 reconstitute the HTT-HAP40 complex, as shown by our cryoEM analysis, indicating that they are folded  
100 properly and can serve as suitable surrogates for structural or biophysical analyses. Additionally, we show that  
101 HAP40 can form a stable complex with the C-HEAT domain in the absence of the N-HEAT and bridge domains,  
102 and that this CTD-HAP40 complex can be resolved by cryoEM. Using these subdomain constructs, we provide  
103 the first kinetic analysis of the HTT-HAP40 interaction, revealing high affinity binding of the two proteins.  
104 Finally, we validate the interaction between the CTD and HAP40 in cells using a luciferase two-hybrid assay and  
105 use subdomain constructs to further probe the HAP40-stabilisation function of HTT in cells. We expect these  
106 subdomains will serve as useful biochemical reagents for further structural investigation of the HTT-HAP40  
107 complex, as well as for screening, characterisation, and validation of other members of the HTT protein-protein  
108 interactome.

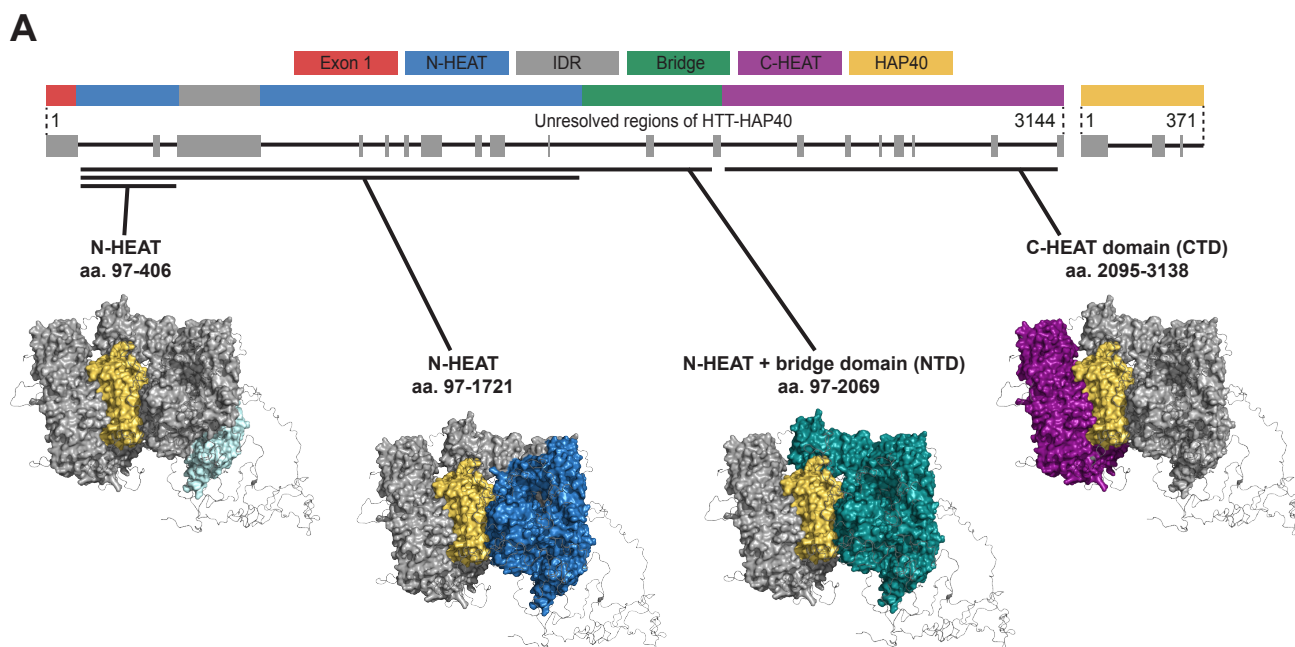
## 109 Results

### 110 *Design and cloning of HTT and HTT subdomain constructs with FLAG and Avi tags*

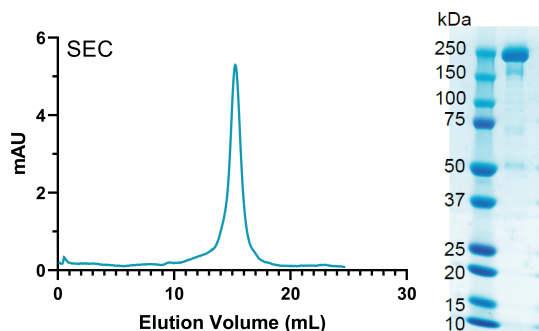
111 Cloning of biotinylated full-length HTT constructs, as well as subdomain constructs, with or without biotin tags,  
112 was performed as previously reported (Harding et al., 2019). Ligase-independent cloning (LIC) was used to  
113 clone gene sequences into a pBMDEL baculovirus transfer vector which can be used for expression in either Sf9  
114 insect cells or mammalian cells. Clones for subdomain constructs were designed by examining the domain  
115 boundaries observed in the cryo-EM model of full-length HTT-HAP40 (**Figure 1A**) and span aa. 97-2069 for the  
116 N-HEAT and bridge domain containing construct, and aa. 2095-3138 for the C-HEAT domain construct. All  
117 clones were designed to incorporate C-terminal FLAG tags for affinity purification. Biotinylated constructs were  
118 cloned to include a 15 amino acid AviTag peptide sequence at either the N- or C-terminus of the protein,  
119 enabling biotinylation by co-expression with BirA biotin ligase and addition of exogenous biotin (Fairhead and  
120 Howarth, 2015).

### 121 *Scalable production of high-purity HTT and HTT subdomain samples*

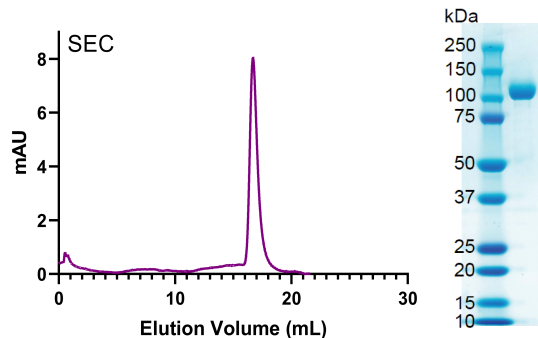
122 After cloning and verifying the gene sequences of the proposed subdomains, as well as the biotinylated full-  
123 length and subdomain constructs, we next aimed to express and purify these constructs using baculovirus-  
124 mediated expression in Sf9 insect cell culture. Samples of HTT obtained from either mammalian or insect cell  
125 expression have resulted in comparable cryoEM structures (Guo et al., 2018; Harding et al., 2021) and our  
126 constructs may be used for production in either system. SDS-PAGE analysis of anti-FLAG purified samples from  
127 small-scale test cultures showed the presence of bands corresponding to the expected molecular weights of  
128 each subdomain. To further characterise each sample, expression volumes were scaled up to 4 L and the  
129 recombinant subdomains were subsequently isolated from lysed cells via FLAG affinity purification and then  
130 further purified by size-exclusion chromatography (SEC). We found that a subdomain construct consisting of  
131 the folded N-terminal region up to the intrinsically disordered domain (IDR, aa. 97-406) was capable of being  
132 expressed and purified by SEC but was found to be poorly stable in biophysical assays (**Figures S1**). Similarly, a  
133 construct consisting of the complete N-terminal solenoid region (aa. 97-1721) showed a non-optimal elution  
134 profile on SEC, suggesting that it was not a suitable protein for *in vitro* assays (**Figure S2**). However, we  
135 discovered that the inclusion of the bridge domain in this construct (aa. 97-2069) caused it to elute as a  
136 monodisperse sample, permitting its isolation at >85% purity as determined by densitometry (**Figure 1B**). The  
137 yield of the N-terminal HEAT and bridge domain (NTD) construct after this two-step purification sequence was  
138 comparable to that of full-length HTT  $\Delta$ exon 1 at ~1.4 mg/L. Additionally, a third construct consisting of the C-  
139 terminal HEAT domain (CTD) could also be purified as a monodisperse sample in high purity (**Figure 1C**), albeit  
140 at lower yields of ~0.3 mg/L. Biotin tagged variants of HTT, HTT-HAP40 and the NTD and CTD constructs  
141 bearing AviTag sequences at either the N- or C-terminus were co-expressed with BirA and purified in the same  
142 manner as constructs bearing only FLAG tags. Size-exclusion chromatography profiles and corresponding yields  
143 of biotinylated proteins were nearly identical to non-biotinylated constructs. The installation of the biotin  
144 moiety was confirmed by streptavidin gel-shift (Fairhead and Howarth, 2015) (**Figure S3**).



**B**



**C**



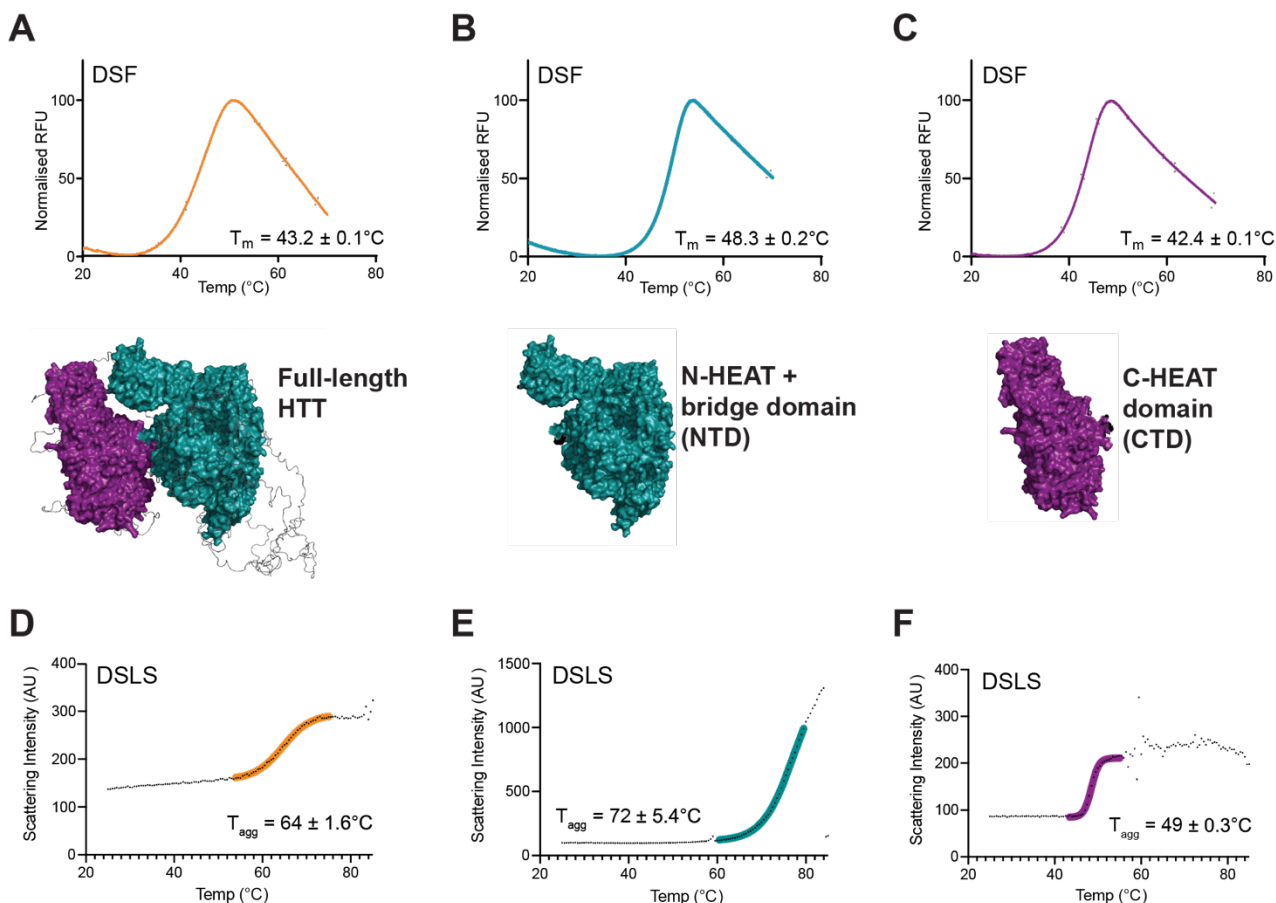
145

146 **Figure 1: NTD and CTD subdomain constructs of HTT can be recombinantly expressed and produced to high**  
147 **purity in milligram quantities, enabling structural, functional and biophysical studies.** A, Schematic  
148 representation of the domain architecture of HTT and regions within the domain boundaries that are  
149 unresolved in cryo-EM models (grey). Constructs analysed within this study are shown on the HTT-HAP40  
150 structural model (Harding et al., 2021). B, Representative SDS-PAGE and size-exclusion chromatogram of  
151 purified NTD. C, Representative SDS-PAGE and size-exclusion chromatogram of purified CTD.

152 *Subdomain constructs show good biophysical properties in multiple orthogonal assays*

153 With purified samples of each subdomain in hand, we next sought to assess their quality and characteristics  
154 through a series of biophysical analyses. We performed differential scanning fluorimetry (DSF) (Niesen et al.,  
155 2007) to determine the thermal stability of each construct under identical buffer, salt and pH conditions as full-  
156 length HTT samples (**Figure 2A-2C**). We found that the NTD, with a calculated melting temperature ( $T_m$ ) of  $48.3$   
157  $\pm 0.2^\circ\text{C}$ , has significantly greater thermal stability relative to full-length apo HTT ( $T_m \sim 43.2 \pm 0.1^\circ\text{C}$ ). Conversely,  
158 we observed a  $T_m$  of  $42.4 \pm 0.1^\circ\text{C}$  for the CTD, suggesting this subdomain is less thermostable than full-length  
159 HTT. Nevertheless, both subdomains showed desirable biophysical characteristics, including a sharp transition  
160 upon heating and high dynamic range of the fluorescence signal between folded and unfolded states (**Figure**  
161 **S4**). We also analysed the protein constructs using differential static light scattering (DSLS) to assess the  
162 aggregation behaviour of each sample upon heating (**Figure 2D-2F**). Mirroring the trend observed by DSF, the

163 NTD showed a significantly higher aggregation temperature ( $T_{agg}$ ) than the CTD. Given the large, multidomain  
164 structure of full-length HTT, it is possible that its  $T_m$  and  $T_{agg}$  values represent an averaging of the values for the  
165 NTD and CTD. Notably, the unfolding of full-length HTT as monitored by DSF occurs over a broader  
166 temperature range, suggesting that the domains may unfold sequentially. Taken together, these data suggest  
167 that the two purified subdomains each comprise a soluble, folded protein that is suitable for use in biophysical  
168 and screening assays.

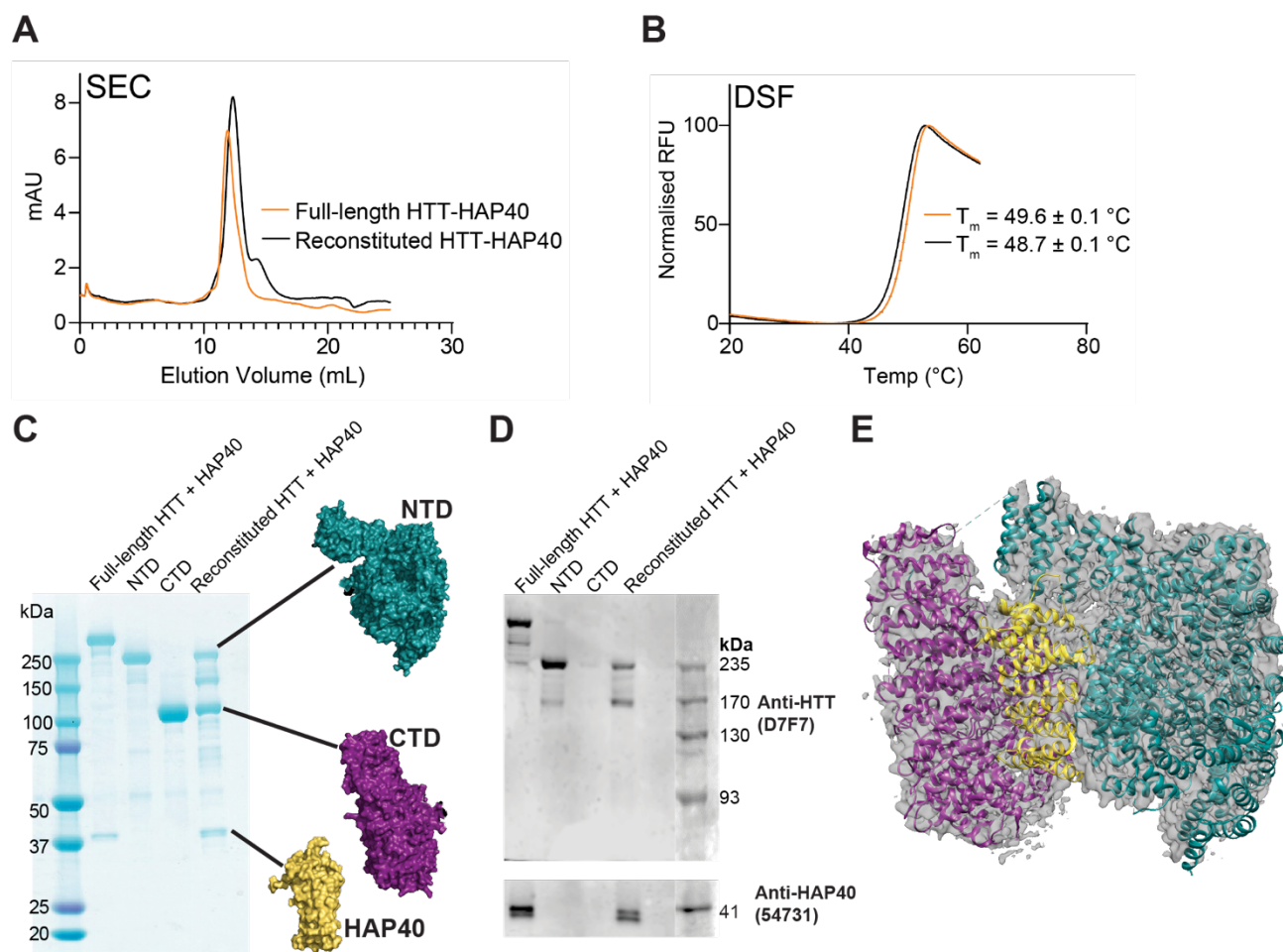


169

170 **Figure 2: NTD and CTD subdomains of HTT have good biophysical properties.** A-C, DSF profiles and calculated  
171 melting temperature ( $T_m$ ) values for full-length HTT (orange) vs NTD (teal) and CTD (magenta). Melting  
172 temperatures were determined from the inflection point of curves obtained by fitting the data to the  
173 Boltzmann sigmoidal function. D-F, DSLS profiles of full-length HTT (orange) vs. NTD (teal) and CTD (magenta)  
174 showing calculated aggregation temperature ( $T_{agg}$ ).

175 *CryoEM reveals the HTT subdomains can stabilise HAP40, forming a functional HTT-HAP40 complex*  
176 Given the results of our biophysical assessment of the NTD and CTD samples, which suggested the proteins  
177 were properly folded, we next examined if they could be used to re-constitute the HTT-HAP40 complex. To test  
178 this, we co-expressed NTD, CTD, and HAP40 in Sf9 cells and attempted to purify the protein as before.  
179 Strikingly, after affinity purification from lysed cells with anti-FLAG resin, we observed protein bands  
180 corresponding with the molecular weight of these three fragments by SDS-PAGE (**Figure S5**). Further  
181 purification of the crude samples by size-exclusion chromatography as before revealed a profile nearly identical  
182 to that obtained with full-length HTT-HAP40, with a major peak eluting at the same volume as expected for the  
183 intact construct (**Figure 3A**). After collecting this peak and performing further analysis by DSF, we observed a  
184 thermal transition profile and  $T_m$  value nearly identical to that of full-length HTT (**Figure 3B, Figure S6**). As

185 further confirmation of the protein bands observed through SDS-PAGE analysis of the purified triplet complex  
186 (**Figure 3C**), we carried out western blotting using antibodies for the N-terminal region of HTT as well as HAP40  
187 (**Figure 3D, Figure S7**). This analysis showed that the 218 kDa band corresponds to the NTD and that the band  
188 at ~40 kDa is HAP40. We also noted the presence of an additional band at ~160 kDa, which persisted despite  
189 several rounds of purification. This impurity also possessed a recognised epitope of the D7F7  $\alpha$ -HTT antibody,  
190 suggesting that it may be a partial fragment of the N-HEAT region that either does not fully express in Sf9 cells  
191 or has been proteolysed by an unknown mechanism. HTT is susceptible to proteolysis within the IDR (30), and  
192 the molecular weight of this fragment suggests this band corresponds to a cleavage product within this region.



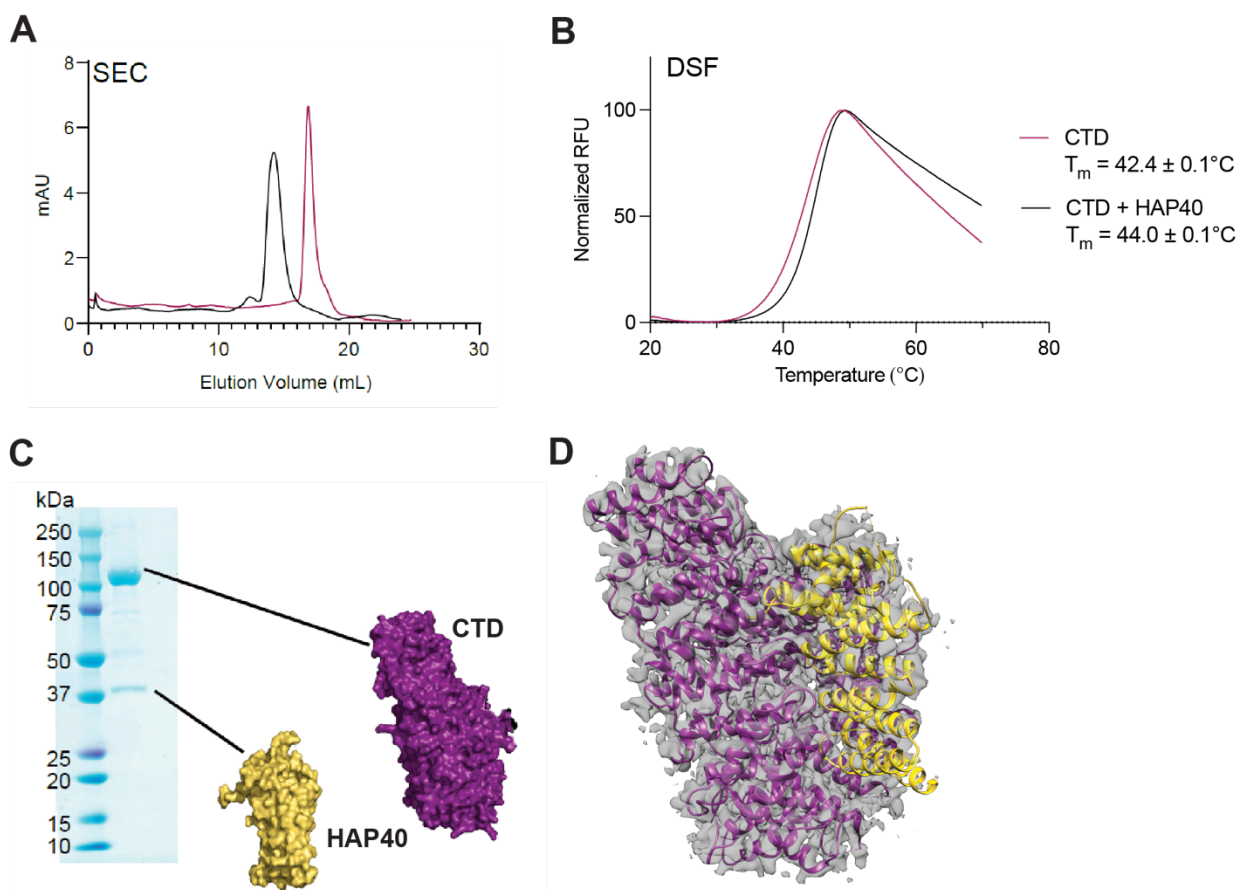
193

194 **Figure 3: HTT subdomains bind HAP40 and can be used to reconstitute the HTT-HAP40 complex.** A, Size  
195 exclusion chromatography profiles of co-expressed NTD, CTD and HAP40 (black trace) after FLAG purification  
196 from Sf9 cells. An elution volume nearly identical to full-length HTT-HAP40 (orange trace) indicates formation  
197 of the trimeric complex. B, Reconstituted and full-length HTT-HAP40 complexes have similar melt profiles as  
198 as determined by DSF. C, SDS-PAGE of full-length HTT (lane 1), NTD (lane 2), CTD (lane 3) and co-expressed  
199 subdomains (lane 4). D, Western blots of full-length HTT-HAP40, NTD, CTD, and reconstituted HTT-HAP40. E,  
200 Overlay of cryo-EM map obtained from reconstituted HTT-HAP40 superimposed on the model generated for  
201 full-length HTT-HAP40 6X9O.

202 To further verify the structural fidelity of the reconstituted complex, we performed cryo-EM of this sample  
203 using similar conditions as performed for the full-length HTT-HAP40 complex (**Table S1**). The 3.3 Å resolution  
204 map obtained from this sample is highly similar to the previously solved atomic resolution model (**Figure 3E**)

205 (Harding et al., 2021). The striking resemblance of this reconstituted complex to the full-length HTT-HAP40  
206 complex indicates the subdomains are capable of properly folding and associating to a degree that enables the  
207 concomitant expression of HAP40 and validates these constructs for the expression of functional and folded  
208 HTT protein subdomains.

209 *The HTT CTD is sufficient to stabilise HAP40, forming a complex which can be resolved by cryoEM*  
210 Having confirmed the robust structural and biophysical integrity of the NTD and CTD constructs, we next asked  
211 whether either the NTD or CTD alone is capable of stabilizing HAP40 when co-expressed. Co-expression  
212 experiments with the NTD and analysis by SDS-PAGE did not show evidence of HAP40 expression (**Figure S8**).  
213 However, after co-expressing HAP40 with the CTD, a band corresponding to HAP40 was present on SDS-PAGE  
214 at the expected molecular weight (**Figure 4A, Figure S9**). Analysis of this sample using analytical size-exclusion  
215 chromatography revealed a significant shift in elution volume, corresponding to an increase in size relative to  
216 the C-HEAT construct (**Figure 4B**). We performed DSF to determine the thermostability of this complex and  
217 found that it possessed a marginally higher melting temperature of approximately 1.6 degrees and a sharper  
218 melting transition curve than the CTD (**Figure 4C, Figure S10**). This suggests that the presence of the HAP40  
219 binding partner is beneficial to the overall stability of the CTD and that the complex is properly folded.



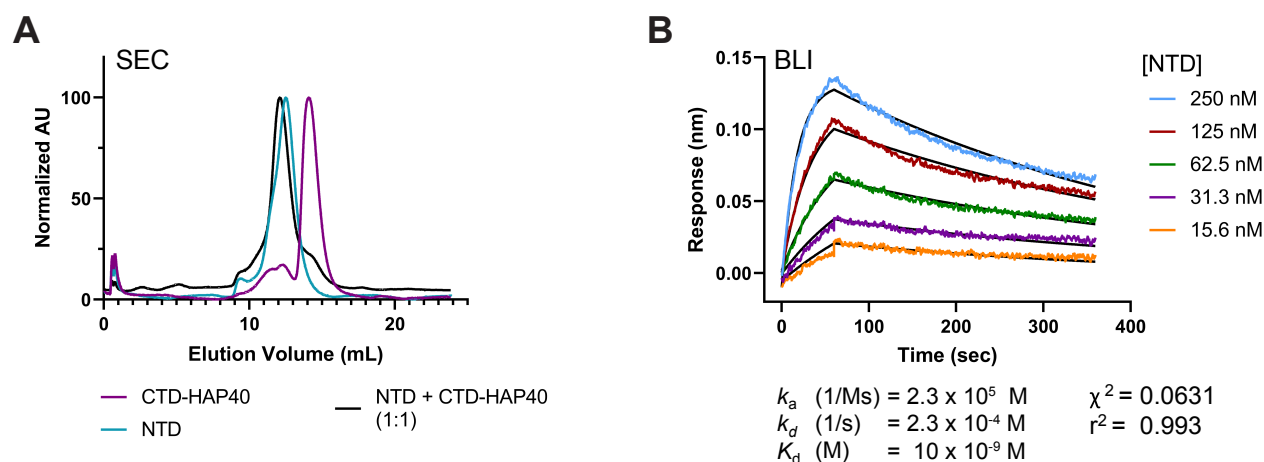
220

221 **Figure 4: HAP40 can be expressed with CTD alone.** A, Overlay of size-exclusion chromatograms of purified co-  
222 expressed CTD and HAP40 (black) vs apo CTD (magenta). B, SDS-PAGE analysis of purified co-expressed C-HEAT  
223 and HAP40, with bands at ~120 kDa and ~40 kDa indicated. C, The CTD-HAP40 complex displays higher  
224 thermostability and a sharper transition upon thermal unfolding as determined by DSF. D, Overlay of the map  
225 of purified co-expressed CTD-HAP40 obtained by cryo-EM (grey mesh) with the cartoon structure of the domains  
226 determined from the full-length cryo-EM model 6X9O.



227 We again studied the structure of this complex by cryo-EM to determine its overall 3D structure relative to full-  
228 length HTT-HAP40 (**Figure 4D, Table S1**). We found that the 3.2 Å resolution map obtained from the CTD-  
229 HAP40 complex fit well into the structural model derived from the full-length complex for these regions of the  
230 protein, with only slight deviations at the termini of the proteins. Overall, these data indicate that the CTD  
231 alone is sufficient to stabilize and permit expression of HAP40.

232 *HTT subdomain constructs allow interrogation of protein-protein interactions by biophysical methods*  
233 With the CTD-HAP40 complex in hand, we proceeded to measure its binding affinity to the NTD *in vitro*. Using  
234 analytical size-exclusion chromatography, we showed that pre-mixing the purified protein subdomains results  
235 in a significant leftward shift in elution volume compared to the elution of CTD-HAP40 alone, suggesting that  
236 the majority of the CTD-HAP40 associates with the NTD (**Figure 5A**). To quantify the interaction, we purified  
237 CTD-HAP40 bearing a C-terminal biotin tag to measure binding kinetics using biolayer interferometry (BLI).  
238 Biotinylated CTD-HAP40 was immobilised onto streptavidin biosensors at a concentration of 1 µg/mL and, after  
239 equilibration, the sensors were incubated with various concentrations of NTD (**Figure 5B, Figure S11**). The  
240 resulting association and dissociation curves revealed a concentration-dependent rate of binding and  
241 dissociation of the two protein subunits. By fitting the curves to a global 1:1 binding model, a dissociation  
242 constant ( $K_d$ ) of  $10 \pm 0.3$  nM was obtained. These data represent the first quantitative *in vitro* assessment of  
243 the strength of HAP40 binding to HTT and support earlier observations that the HTT-HAP40 interaction is  
244 remarkably stable under a wide array of conditions (Harding et al., 2021).

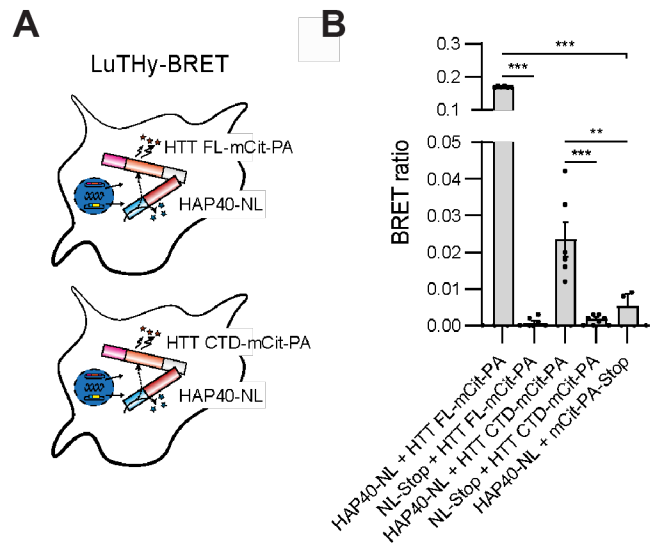


245

246 **Figure 5: Purified co-expressed CTD-HAP40 binds the NTD with high affinity.** A, Overlaid size-exclusion  
247 chromatography profiles of purified CTD-HAP40 (magenta), NTD (teal) and equimolar amounts of pre-mixed  
248 NTD and CTD-HAP40 (black). B, Characterisation of NTD binding affinity to immobilised CTD-HAP using biolayer  
249 interferometry, revealing a  $K_d$  of  $10 \pm 0.3$  nM. Kinetic parameters were determined by fitting the data to a 1:1  
250 binding association and dissociation model (black lines) using GraphPad Prism 9. Three independent  
251 experiments were performed.

252 *Validation of HTT subdomain protein-protein interactions in cells*

253 To further verify that HAP40 can be bound by HTT CTD to form a stable complex under physiological conditions,  
254 we performed a luciferase two-hybrid (LuTHy) assay (Trepte et al., 2018) in HEK293 cells (**Figure 6A**). mCitrine-  
255 Protein A (mCit-PA)-conjugated HTT constructs corresponding to the full-length and CTD sequences, showed  
256 BRET ratios which were significantly increased compared to tag-only controls when co-expressed with HAP40-  
257 NanoLuc luciferase (NL) fusion protein (**Figure 6B**). This confirms that the CTD of HTT is sufficient to bind HAP40  
258 in a cellular context.



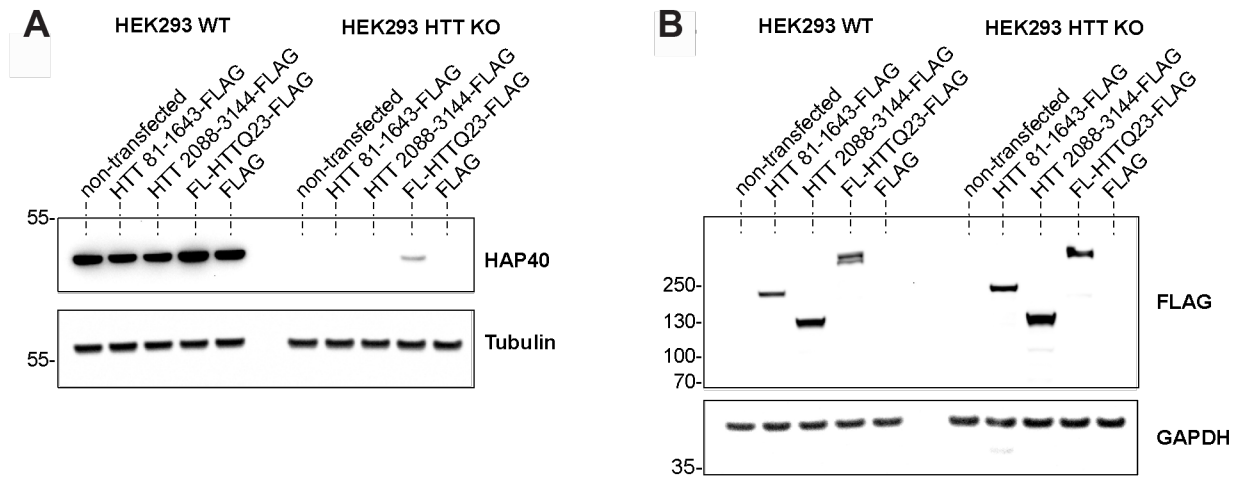
259

260 **Figure 6: LuTHy assay shows interaction of HTT full-length (FL) and its C-terminal domain (CTD) with HAP40 in**  
261 **live cells.** A, Graphical illustration of the LuTHy-BRET assay performed. HTT FL or HTT CTD were expressed as  
262 mCitrine-Protein A (mCit-PA)- and HAP40 as NanoLuc luciferase (NL)-tagged fusion proteins in HEK293 cells.  
263 After expression for 48 h and addition of luciferase substrate, BRET was quantified from live cells. B, BRET  
264 ratios between HAP40-NL and HTT FL-mCit-PA and HTT CTD-mCit-PA. As a control, NL only was co-transfected  
265 with the HTT acceptor constructs, respectively, as well as PA-mCit only with the HAP40-NL donor construct.  
266 HAP40-NL co-expressed with HTT FL-mCit-PA and HTT CTD-mCit-PA showed significantly increased BRET ratios  
267 compared to controls, respectively. Bars represent means  $\pm$  SEM from two independent experiments  
268 performed in triplicate. One-way ANOVA with Tukey's multiple comparisons test, \*\* $p < 0.002$ , \*\*\*  $p < 0.001$ .

### 269 *Investigating HTT cellular functions with subdomain constructs*

270 Next, we sought to scrutinise a defined biomolecular function of HTT using our subdomain tools. HAP40  
271 expression cannot be detected in HTT knockout HEK293 cells (**Figure 7A**), in line with previously published data  
272 which show the levels of HAP40 are dependent on HTT protein levels and that HTT functions to stabilise HAP40  
273 (Harding et al., 2021; Huang et al., 2021b). Thus, by monitoring the protein levels of HAP40 in HTT KO cells in  
274 the presence of exogenous full-length or subdomain constructs of HTT we could observe a rescue of the HAP40  
275 null phenotype in these cells. Overexpression of N-HEAT, CTD or full-length HTT in these cells showed that only  
276 the full-length HTT can restore some degree of HAP40 expression in HTT knock out HEK293 cells. Together with  
277 our LuTHy assay data, this indicates that although HTT CTD can bind HAP40 in a cellular context, it is not  
278 sufficient to restore the HAP40 stabilisation function of HTT.

Toolkit of HTT subdomain constructs



279

280 **Figure 7: Only overexpression of full-length HTT can rescue HAP40 protein levels in HTT-null cells.**

281 A-B, Representative western blot images of total cell lysates from HEK293 wildtype and HEK293 HTT-null lines  
282 after transfection with of various FLAG-tagged constructs for HTT full-length (FL) and truncated species of HTT.

283 A, Nitrocellulose membrane was incubated with rabbit anti-F8A1 antibody, followed by mouse anti-tubulin to

284 detect endogenous HAP40 and tubulin. B, Nitrocellulose membrane was incubated with mouse anti-FLAG

285 antibody, followed by mouse anti-GAPDH to detect overexpressed FLAG-tagged constructs and endogenous

286 GAPDH. Calculated molecular weight for HTT 81-1643 is 175 kDa, for HTT 2088-3144 121 kDa, and for FL

287 HTTQ23 352 kDa. Two independent experiments were performed.

## 288 Discussion

289 The expansion of the HTT polyQ tract has been well-defined as the cause of HD pathobiology. However, the  
290 precise biochemical functions of wildtype HTT, or the changes that result in the presence of expanded HTT,  
291 remain incompletely understood. Given a lack of biochemical tools to interrogate HTT function in cells, genetic  
292 studies have formed the basis of most functional studies, either through knockdown of HTT expression or the  
293 generation of expanded polyQ allelic series in various animal models or cell lines (Menalled et al., 2009; Shen et  
294 al., 2019; Southwell et al., 2013). Additionally, affinity-based proteomic methods coupled with mass  
295 spectrometry have been applied to identify putative protein binders that interact with HTT (Culver et al., 2012;  
296 Ratovitski et al., 2012). Efforts to interrogate HTT function by many biochemical and biophysical assays are  
297 thwarted by the large size of HTT, which renders these assays non-tractable. Smaller HTT subdomains provide a  
298 solution to this problem, but many of the constructs published to date are not designed with knowledge of the  
299 HTT structure in mind, meaning their boundaries may interrupt structural elements or they contain extensive  
300 disordered regions, which likely result in non-native protein products. To date, no robust biophysical or  
301 structural validation of any such construct has been published to date. The toolkit of subdomain constructs  
302 described in this manuscript provides a solution to this problem, through the rational structure-based design of  
303 subdomain constructs and their validation by numerous assays and structural analysis by cryoEM.

304 Because of the suspected role of expanded HTT in triggering HD progression, HTT lowering therapies have been  
305 hypothesised as a potential therapeutic strategy for HD. One proposed pharmacological approach to HTT  
306 lowering is through the use of proteolysis-targeting chimeras (PROTACs) – heterobifunctional ligands that  
307 combine a recognition scaffold for a protein of interest combined with an E3 ligase motif, resulting in ubiquitin-  
308 mediated proteasomal degradation of target proteins (Sakamoto et al., 2001; Schapira et al., 2019). To date,  
309 however, all small-molecule ligands being developed as PROTACs bind mutant aggregated HTT fragments (Hirai  
310 et al., 2022; Tomoshige et al., 2017). As a result, these compounds may suffer from off-target effects because  
311 they do not bind a homogenous or specific binding site on the protein and do not bind soluble HTT that is  
312 present within cells. There is therefore a need for the discovery and validation of novel small-molecule  
313 scaffolds that bind soluble HTT in a reversible and specific manner.

314 To support efforts to validate the complex network of HTT protein-protein interactions, as well as to support  
315 pre-clinical characterisation and optimisation of small-molecule ligands, high quality biochemical tools and  
316 assay methods are needed. The suite of full-length and subdomain constructs we have developed have a range  
317 of potential applications to support fundamental biology and drug discovery efforts. We expect that the  
318 combined use of NTD and CTD constructs will permit approaches to counter-screen novel ligands for selectivity  
319 of binding, as well as to characterise the degree and site of interaction between HTT and potential PPIs.  
320 Furthermore, the smaller size and robust nature of the biotin tags on these constructs enable a variety of  
321 biophysical assays to accurately quantify the affinity of potential HTT binders. Additionally, the smaller  
322 molecular weight of the subdomains and, in some cases, their higher expression yield relative to full-length  
323 HTT, improves their suitability for high-throughput applications.

324 Our efforts to generate truncated HTT subdomains have also provided fundamental insight into the key  
325 structural interactions governing the HTT-HAP40 complex. Our observation that HAP40 can be co-expressed  
326 with the CTD alone, both in vitro, and form a stable, soluble protein complex, shows the importance of this  
327 interface for HAP40 stability. Analysis of the CTD-HAP40 interaction surface shows that this subdomain  
328 accounts for approximately 60% of the total interaction area between HAP40 and full-length HTT, highlighting  
329 the significance of the C-HEAT on binding to HAP40. The generation of this complex permitted quantitative  
330 characterisation of its binding to the NTD, providing additional evidence of the robust nature of HTT-HAP40  
331 affinity. Such analysis has not been possible with full-length HTT to date as HAP40 must be coexpressed with

## Toolkit of HTT subdomain constructs

332 HTT, or, as we demonstrate in this study, the CTD of HTT. Given that no other putative protein binding partners  
333 of HTT have been characterized through a direct binding assay in this manner, the use of the CTD-HAP40 and  
334 NTD association provides an important positive control for prospective PPI analyses.

335 Using our defined and validated subdomain construct boundaries, we were also able to characterise the CTD-  
336 HAP40 interaction in live cells. This revealed that the CTD can bind HAP40 when measured using the BRET-  
337 based LuTHy readout. Interestingly, in a HTT knockout background where HAP40 protein levels are ablated due  
338 to lack of stabilisation by HTT (Harding et al., 2021; Huang et al., 2021b; Xu et al., 2022), we showed that the  
339 full-length, but not the CTD alone, is able to restore HAP40 expression. This suggests that the interdependent  
340 relationship between HTT and HAP40 expression is perhaps not just one of structural stabilisation, as the CTD is  
341 sufficient to form a stable HAP40 complex, but perhaps linked in a more complex co-translational dependent  
342 manner.

343 In summary, we have generated and validated a suite of high-quality biochemical tools to support the  
344 structural and functional characterisation of HTT. We report the design and purification of HTT subdomains  
345 encompassing the N-terminal and C-terminal regions of the protein, as well as biotinylated variants that permit  
346 convenient immobilisation on streptavidin medium. These protein constructs, which are stably expressed and  
347 folded, permit a wide array of biophysical assays to probe the diverse network of HTT protein-protein  
348 interactions and facilitate future screening efforts. Additionally, these subdomains have provided important  
349 information on the nature of the HTT-HAP40 interaction, demonstrating that the C-HEAT domain forms key  
350 interactions with HAP40 and is sufficient to enable the expression of HAP40 in the absence of the NTD, both *in*  
351 *vitro* and in cells. Finally, we used our defined subdomains to probe the HAP40-stabilisation function of HTT,  
352 revealing expression of the full-length HTT protein molecule is required to rescue this function in a knockout  
353 background. We expect that these biochemical tools will find widespread use within the HD community and  
354 will support ongoing efforts to elucidate the biological roles of HTT.

## 355 Methods

### 356 *Cloning of expression constructs*

357 Expression constructs for the NTD (amino acids 97 to 2069) and the CTD (amino acids 2095 to 3138) were  
358 constructed with three different affinity tag arrangements. The first version had a C-terminal Flag tag, the 2nd  
359 version had a N-terminal biotinylation tag / C-terminal Flag tag, and the 3rd version had the biotinylation and  
360 Flag tag on the C-terminal. The biotinylation tag contains the sequence GLNDIFEAQKIEWHE which facilitates  
361 the in vivo conjugation of a single biotin to the lysine residue when the HTT protein is co-expressed with biotin  
362 ligase in Sf9 cells. To assemble each plasmid, the DNA encoding for the HTT gene fragments was PCR amplified  
363 from cDNA (Kazusa clone FHC15881). PCR primers were designed to add coding sequence for N-terminal or C-  
364 terminal biotinylation tags and Flag tags as required. The HTT expression constructs were assembled using the  
365 BD In-Fusion PCR cloning kit in the mammalian/insect cell vector pBMDEL (Addgene plasmid #111751), an  
366 unencumbered vector created for open distribution of these reagents.

367 Two expression constructs were made for the expression of full length (FL) HTT (amino acids 1 to 3144). One  
368 had an N-terminal biotinylation tag and C-terminal flag tag. The other had both the biotinylation tag and Flag  
369 tag on the C-terminal. These plasmids were created through modification of Addgene plasmid #111726, whose  
370 construction we previously described (Harding et al., 2019). Briefly, these modifications were done in two  
371 steps. First, a N-terminal or C-terminal fragment the HTT gene with the required tags was assembled into an  
372 intermediate plasmid. Second, these tagged fragments were transferred into Addgene plasmid #111726 by PCR  
373 amplification followed by insertion into the expression construct using the BD In-Fusion PCR cloning kit.

374 Expression vectors for C-terminal FLAG-tagged constructs were generated by Gateway cloning. To this end, 150  
375 ng of entry vector in a pDONR221 vector for full-length HTTQ23, HTT 81-1643, and HTT 2088-3144 was mixed  
376 with 600 ng of the destination vector pDEST\_gateway-2xFLAG (Addgene plasmid #118372), 1 µl of LR clonase,  
377 and filled up to 10 µl with TE buffer pH 8.0. Reaction was incubated overnight at room temperature.  
378 Afterwards 1 µl of Proteinase kinase K was added and incubated at 37°C for 10 minutes. Reaction was then  
379 transformed into chemically competent Mach1 bacterial cells, plated onto ampicillin agar plates. Clones were  
380 picked, cultured in liquid LB medium, and DNA isolated using Qiaprep Spin Miniprep Kit (Qiagen). Final  
381 constructs were sequenced to validate correct insertion using the following primers;  
382 GAGGTCTATATAAGCAGAGC and AACCATTATAAGCTGCAATAAAC.

383 LuTHy plasmids were generated as described previously (Trepte et al., 2018). In brief, open reading frames  
384 (ORFs) of full-length HTT(Q23) or HTT's CTD were amplified from previously generated expression plasmid  
385 (Addgene plasmid #111723) and resulting attB1 and attB2-flanked PCR products were subcloned into  
386 pDONR221 entry vectors using BP Clonase (Gateway Cloning System, Invitrogen). For HAP40 (F8A1), a cDNA  
387 entry clone was obtained from Source BioSciences (OCAAo5051D1091D). To generate LuTHy expression  
388 plasmids, HAP40 was cloned into a LuTHy donor vector (Addgene plasmid #113447) and the HTT constructs  
389 into a LuTHy acceptor vector (Addgene plasmid #113449) by LR Clonase reactions (Gateway Cloning System,  
390 Invitrogen). All expression plasmids were finally validated by restriction enzyme digest, agarose gel  
391 electrophoresis, and Sanger sequencing.

392 The DNA sequences and expressed protein sequences of plasmids used in this paper are summarized in  
393 **Appendix 1.**

### 394 *Protein expression and purification*

395 Expression of HTT subdomains was performed in Sf9 insect cells as previously described (Harding et al., 2019)  
396 with minor modifications. Briefly, cells were infected with P3 recombinant baculovirus and grown at 37°C until  
397 cell viability reached 80-85%, normally ~72 hours post-infection. For full-length HTT-HAP40, a ratio of 1:1

## Toolkit of HTT subdomain constructs

398 HTT:HAP40 P3 baculovirus was used for infection. For CTD-HAP40, a ratio of 1:3 CTD:HAP40 P3 baculovirus was  
399 used. Biotinylation of AviTag constructs was achieved by co-expression with recombinant BirA in the presence  
400 of  $\mu\text{M}$  biotin. For purification, cells were harvested by centrifugation and lysed with two freeze-thaw cycles.  
401 Lysates were then clarified by centrifugation at 29, 416 x g for 60 minutes. After collecting the supernatant,  
402 proteins were isolated by FLAG affinity chromatography. Crude protein samples from FLAG-eluted fractions  
403 were then pooled and purified by size-exclusion chromatography using a Superose 6 Increase 10/300 GL  
404 column (Cytiva Life Sciences) in buffer containing 20 mM HEPES pH 7.4, 300 mM NaCl, 2.5% (v/v) glycerol, 1  
405 mM TCEP. Peaks corresponding to monodisperse protein were pooled, concentrated and flash frozen and  
406 stored at  $-80^{\circ}\text{C}$  until use. Sample purity was assessed by SDS-PAGE.

### 407 *Differential scanning fluorimetry*

408 Determination of protein thermostability by DSF was performed using a Roche Applied Science Light Cycler 480  
409 II. Samples were prepared in LightCycler 480 white 384-well plates (Roche) at a volume of 20  $\mu\text{L}$  per well using  
410 a final concentration of 0.1 mg/mL protein and 5X Sypro Orange (Invitrogen). All samples were diluted in buffer  
411 containing 20 mM HEPES pH 7.4, 300 mM NaCl, 2.5% (v/v) glycerol, and 1 mM TCEP. Thermal shift assays were  
412 carried out over a temperature range of 20 to  $95^{\circ}\text{C}$  at a ramp rate of  $0.02^{\circ}\text{C}/\text{sec}$ . Fluorescence measurements  
413 were taken using a 465 excitation / 580 emission filter set. Reactions were performed in triplicate and data  
414 were analyzed using GraphPad Prism 9. Melt curves were fitted to a Boltzmann sigmoidal curve and apparent  
415  $T_m$  values were determined by calculating the inflection point of the fitted curves.

### 416 *Differential static light scattering*

417  $T_{\text{agg}}$  values for purified proteins were determined by DLS using a Stargazer instrument (Harbinger Biotech).  
418 Protein samples were diluted to 0.4 mg/mL in buffer containing 20 mM HEPES pH 7.4, 300 mM NaCl, 2.5% (v/v)  
419 glycerol and 1 mM TCEP using a volume of 50  $\mu\text{L}$  per well. Samples were heated from  $20^{\circ}\text{C}$  to  $85^{\circ}\text{C}$  at a rate of  
420  $1^{\circ}\text{C}/\text{min}$  and protein aggregation was monitored by measuring the intensity of scattered light every 30 s with a  
421 CCD camera. Scattering intensity was then plotted and fitted to the Boltzmann equation, and  $T_{\text{agg}}$  values were  
422 determined by measuring the inflection point of each curve.

### 423 *Cryo-EM sample preparation and data acquisition*

424 Full-length HTT-HAP40 or CTD-HAP40 samples were diluted to 0.2 mg/mL in 25 mM HEPES pH 7.4, 300 mM  
425 NaCl, 0.025% w/v CHAPS and 1 mM DTT and adsorbed onto gently glow-discharged suspended monolayer  
426 graphene grids (Graphenea) for 60 s. Grids were then blotted with filter paper for 1 s at 100% humidity,  $4^{\circ}\text{C}$   
427 and frozen in liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific).

428 Data were collected in super-resolution counting mode on a Talos Arctica (Thermo Fisher Scientific) operating  
429 at 200 kV with a BioQuantum imaging filter (Gatan) and K3 direct detection camera (Gatan) at 100,000x  
430 magnification, corresponding to a real pixel size of  $0.81 \text{ \AA}/\text{pixel}$ . Movies were collected at a dose rate of  
431  $22.8 \text{ e}^{-}/\text{\AA}^2/\text{s}$ , exposure time of 2.25 s, resulting in a total dose of  $51.3 \text{ e}^{-}/\text{\AA}^2$  fractionated across 53 frames.

### 432 *Cryo-EM data processing*

433 Movies were processed in real time using the SIMPLE 3.0 pipeline (Caesar et al., 2020), using SIMPLE-unblur for  
434 patched motion correction, SIMPLE-CTFFIND for patched CTF estimation and SIMPLE-picker for particle picking.  
435 Particles were extracted in  $288 \times 288$  pixel boxes, sampling of  $0.81 \text{ \AA}/\text{pixel}$ .

436 For full-length HTT-HAP40, 112,717 particles were generated from 166 movies and subjected to 2D  
437 classification (50 classes,  $160 \text{ \AA}$  mask) within cryoSPARC (Punjani et al., 2017). Particles (64,597) belonging to  
438 the most defined, highest resolution classes were used to generate an ab initio map. This map was lowpass

439 filtered to 30 Å and used as reference for non-uniform refinement within cryoSPARC, yielding a 3.3 Å volume as  
440 assessed by Gold standard Fourier Shell Correlations (FSC) using the 0.143 criterion.

441 For CTD-HAP40, 986,229 particles were generated from 1,461 movies and subjected to initial 2D classification  
442 in SIMPLE 3.0 (200 classes, 140 Å mask). “Good” particles (819,568) were subjected to an additional round of  
443 2D classification within cryoSPARC (200 classes, 140 Å mask). Retained particles (297,261) underwent multi-  
444 class (4 classes) ab initio volume generation within cryoSPARC, producing one sensible class composed of  
445 134,849 particles. The volume corresponding to this class was lowpass filtered to 30 Å and used as reference  
446 for non-uniform refinement within cryoSPARC, generating a 3.2 Å volume as assessed by Gold standard Fourier  
447 Shell Correlations (FSC) using the 0.143 criterion.

#### 448 *Western blotting of recombinant protein samples*

449 General protocols for western blot analyses were performed as previously described (Harding et al., 2021).  
450 Primary antibodies used include anti-HTT D7F7 (Cell Signalling Technology), anti-HAP40 54731 (Novus  
451 Biologicals) and anti-FLAG F1804 (Sigma). Secondary antibodies used are goat-anti-rabbit IgG-IR800 (LI-COR)  
452 and donkey anti-mouse IgG-IR680 (LI-COR). Membranes were visualised using an Odyssey CLx imaging system  
453 (LI-COR).

#### 454 *Bi-layer interferometry*

455 The affinity of the CTD-HAP40 binding interaction to the NTD was determined using an OctetRED96 BLI system  
456 (ForteBio). Experiments were performed at 25°C in 96-well black microplates (Greiner, 655209) with shaking at  
457 1000 rpm. Purified CTD-HAP40 with a C-terminal biotin tag was diluted to 1 µg/mL in buffer containing 100 mM  
458 sodium phosphate pH 7.4, 500 mM NaCl, 0.05% Tween-20, and 0.1% BSA and were loaded onto streptavidin  
459 biosensors (ForteBio, 18-5019) for 300s. Sensors were then washed in buffer for 120 s to establish a baseline  
460 reading. A serial dilution of NTD in identical buffer was prepared in the same plate, and then loaded sensors  
461 were transferred into NTD wells for a 60 s association period to determine  $k_{on}$ . Sensors were then transferred  
462 to wells containing only buffer for 300 s to determine  $k_{off}$ . Binding kinetics were determined using GraphPad  
463 Prism 9 by fitting the data to a 1:1 association then dissociation model. Association and dissociation curves  
464 from all concentrations of NTD were fitted globally to a shared  $k_{on}$  and  $k_{off}$  value. Background signal from  
465 reference sensors without immobilised CTD-HAP40 was subtracted from experimental sensor wells.

#### 466 *Cell culture and transfection*

467 Human embryonic kidney line 293 (HEK293) wildtype and HTT-null cells were grown in Dulbecco's modified  
468 Eagle's medium (ThermoFisher #41965) supplemented with 10% heat-inactivated fetal bovine serum  
469 (ThermoFisher #10500064), and 1% penicillin/streptomycin (ThermoFisher #15140122) at 37 °C, and 5% CO<sub>2</sub>.  
470 Cells were subcultured every three to four days. For overexpression studies, one million cells in a final volume  
471 of 2 ml complete DMEM medium were reversed transfected in 6-well plates with a transfection mix composed  
472 of two µg of DNA in a volume of 200 µl Opti-MEM and 5 µl Fugene transfection reagent (2:5:1 ratio). After 48  
473 hours cells were trypsinized, washed with ice-cold PBS, pelleted and stored at -80°C.

#### 474 *Western blotting of cell lysates*

475 HEK293 cell pellets collected from a 6-well plate were lysed in 30-50 µl HEPES lysis buffer (50 mM HEPES pH  
476 7.0, 150 mM sodium chloride, 10 % glycerol, 1 % NP-40, 20 mM NaF, 1.5 mM MgCl<sub>2</sub>, 1 mM EDTA, 1 mM PMSF,  
477 0.5% sodium deoxycholate, 1x Benzamide, 1x, Roche Complete EDTA-free protease inhibitor cocktail (Merck,  
478 5056189001)) for 30 min on ice. Lysates were centrifuged at 14,000 rpm for 10 min at 4°C and supernatants  
479 collected. Protein concentrations were determined using the Pierce™ BCA assay (Thermo Scientific) and 20 µg  
480 total protein was combined with 50 mM DTT and 1x NuPAGE LDS sample buffer, followed by 5 min at 95°C.  
481 Proteins were separated by SDS-PAGE using a NuPAGE 4-12% Bis-Tris gel and transferred onto nitrocellulose



## Toolkit of HTT subdomain constructs

482 membranes (Cytiva, 10600002). Membranes were blocked for 1 hour in 3% milk in PBS with 0.05% Tween. The  
483 following primary antibodies were applied overnight at 4°C: mouse anti-FLAG (1:1000; Sigma #F3165), rabbit  
484 anti-HAP40 (1:1000; Atlas Antibodies #HPA046960), mouse anti-tubulin (1:80,000; Sigma #T6074), and mouse  
485 anti-GAPDH (1:1000; Santa Cruz #sc-47724). The following secondary antibodies diluted to 1:6000 in 3% milk  
486 and applied for two hours at room temperature: goat anti-Rabbit IgG peroxidase (Sigma #A0545) and goat anti-  
487 mouse IgG peroxidase (Sigma #A0168). Each membrane was incubated with WesternBright™ Quantum  
488 (advanstar, K-12042-D20) solution for two minutes, followed by acquisition of a chemiluminescence image  
489 using an iBright imaging system (ThermoFisher).

### 490 *In-cell interaction validation with LuTHy*

491 LuTHy-BRET assays were performed as described previously (Trepte et al., 2018). In brief, cells were reverse  
492 transfected using linear polyethyleneimine (25 kDa, Polysciences 23966) with LuTHy constructs and cells were  
493 subsequently incubated for 48 h. Previously generated LuTHy control vectors expressing only NL (Addgene  
494 #113442) or PA-mCit (Addgene #113443) were used as background controls. Live cell BRET measurements  
495 were carried out in flat-bottom white 96-well plates (Greiner, 655983) with 24 PPIs per plate (each PPI in  
496 triplicate). Infinite® microplate readers M1000 or M1000Pro (Tecan) were used for the readout with the  
497 following settings: fluorescence of mCitrine recorded at Ex 500 nm/Em 530 nm, luminescence measured using  
498 blue (370–480 nm) and green (520–570 nm) band pass filters with 1,000 ms integration time. BRET ratios were  
499 calculated by dividing the background corrected luminescence intensity at 520-570 nm by the intensity  
500 obtained at 370-480 nm and subsequent donor bleed-through subtraction from NL only expressing wells.

## 501 Acknowledgements

502 This research was supported by CHDI Foundation (RJH, CHA, EEW and CS), the Huntington's Disease Society for  
503 America (ESR), and in part by the intramural research program of the NIH (SML, JCD). The Structural Genomics  
504 Consortium is a registered charity (no: 1097737) that receives funds from Bayer AG, Boehringer Ingelheim,  
505 Bristol Myers Squibb, Genentech, Genome Canada through Ontario Genomics Institute [OGI-196],  
506 EU/EFPIA/OICR/McGill/KTH/Diamond Innovative Medicines Initiative 2 Joint Undertaking [EUbOPEN grant  
507 875510], Janssen, Merck KGaA (aka EMD in Canada and US), Pfizer and Takeda.

## 508 Data Availability

509 Cryo-EM maps can be downloaded at EMDB-28767 and EMDB-28766. All constructs generated in this study  
510 have been deposited in Addgene and are detailed in Appendix 1.

## 511 Contributions

512 RJH, CHA conceptualisation; MGA, RJH, JCD, PL, AH, AS, RC, ESR, CS, MA methodology; MGA, RJH, JCD, ESR, CS,  
513 MA analysis; MGA, RJH, JCD, ESR, CS, MA investigation; MGA, RJH, JCD, ESR, CS, MA manuscript preparation;  
514 MGA, RJH, JCD, ESR, CS, EW, SML, CHA manuscript review and editing.

## 515 Competing Interests

516 The authors declare no competing interests.

## 517 Abbreviations

518 HTT, huntingtin; HAP40, huntingtin associated protein 40 kDa; CTD, C-terminal domain; N-terminal domain;  
519 LuThy, luciferase two-hybrid assay; cryo-EM, cryo-electron microscopy.

## 520 References

- 521 A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease  
522 chromosomes. The Huntington's Disease Collaborative Research Group. 1993. *Cell* **72**:971–983.
- 523 Andrade MA, Petosa C, O'Donoghue SI, Müller CW, Bork P. 2001. Comparison of ARM and HEAT protein  
524 repeats. *J Mol Biol* **309**:1–18. doi:10.1006/jmbi.2001.4624
- 525 Atwal RS, Desmond CR, Caron N, Maiuri T, Xia J, Sipione S, Truant R. 2011. Kinase inhibitors modulate  
526 huntingtin cell localization and toxicity. *Nature Chemical Biology* **7**:453–460.  
527 doi:10.1038/nchembio.582
- 528 Caesar J, Reboul CF, Machello C, Kiesewetter S, Tang ML, Deme JC, Johnson S, Elmlund D, Lea SM, Elmlund H.  
529 2020. SIMPLE 3.0. Stream single-particle cryo-EM analysis in real time. *Journal of Structural Biology: X*  
530 **4**:100040. doi:10.1016/j.yjsbx.2020.100040
- 531 Culver BP, Savas JN, Park SK, Choi JH, Zheng S, Zeitlin SO, Yates JR, Tanese N. 2012. Proteomic analysis of wild-  
532 type and mutant huntingtin-associated proteins in mouse brains identifies unique interactions and  
533 involvement in protein synthesis. *J Biol Chem* **287**:21599–21614. doi:10.1074/jbc.M112.359307
- 534 Fairhead M, Howarth M. 2015. Site-specific biotinylation of purified proteins using BirA. *Methods Mol Biol*  
535 **1266**:171–184. doi:10.1007/978-1-4939-2272-7\_12
- 536 Greco TM, Secker C, Ramos ES, Federspiel JD, Liu J-P, Perez AM, Al-Ramahi I, Cattle JP, Carroll JB, Botas J,  
537 Zeitlin SO, Wanker EE, Cristea IM. 2022. Dynamics of huntingtin protein interactions in the striatum  
538 identifies candidate modifiers of Huntington disease. *Cell Syst* **13**:304–320.e5.  
539 doi:10.1016/j.cels.2022.01.005
- 540 Gunawardena S, Her L-S, Bruschi RG, Laymon RA, Niesman IR, Gordesky-Gold B, Sintasath L, Bonini NM,  
541 Goldstein LSB. 2003. Disruption of Axonal Transport by Loss of Huntingtin or Expression of Pathogenic  
542 PolyQ Proteins in *Drosophila*. *Neuron* **40**:25–40. doi:10.1016/S0896-6273(03)00594-4

Toolkit of HTT subdomain constructs

- 543 Guo Q, Huang B, Cheng J, Seefelder M, Engler T, Pfeifer G, Oeckl P, Otto M, Moser F, Maurer M, Pautsch A,  
544 Baumeister W, Fernández-Busnadiego R, Kochanek S. 2018. The cryo-electron microscopy structure of  
545 huntingtin. *Nature*. doi:10.1038/nature25502
- 546 Harding RJ, Deme JC, Hevler JF, Tamara S, Lemak A, Cattle JP, Szewczyk MM, Begeja N, Goss S, Zuo X, Loppnau  
547 P, Seitova A, Hutchinson A, Fan L, Truant R, Schapira M, Carroll JB, Heck AJR, Lea SM, Arrowsmith CH.  
548 2021. Huntingtin structure is orchestrated by HAP40 and shows a polyglutamine expansion-specific  
549 interaction with exon 1. *Commun Biol* **4**:1–16. doi:10.1038/s42003-021-02895-4
- 550 Harding RJ, Loppnau P, Ackloo S, Lemak A, Hutchinson A, Hunt B, Holehouse AS, Ho JC, Fan L, Toledo-Sherman  
551 L, Seitova A, Arrowsmith CH. 2019. Design and characterization of mutant and wildtype huntingtin  
552 proteins produced from a toolkit of scalable eukaryotic expression systems. *J Biol Chem* **294**:6986–  
553 7001. doi:10.1074/jbc.RA118.007204
- 554 Harding RJ, Tong Y. 2018. Proteostasis in Huntington’s disease: disease mechanisms and therapeutic  
555 opportunities. *Acta Pharmacologica Sinica* **39**:754–769. doi:10.1038/aps.2018.11
- 556 Hirai K, Yamashita H, Tomoshige S, Mishima Y, Niwa T, Ohgane K, Ishii M, Kanamitsu K, Ikemi Y, Nakagawa S,  
557 Taguchi H, Sato S, Hashimoto Y, Ishikawa M. 2022. Conversion of a PROTAC Mutant Huntingtin  
558 Degradator into Small-Molecule Hydrophobic Tags Focusing on Drug-like Properties. *ACS Med Chem Lett*  
559 **13**:396–402. doi:10.1021/acsmchemlett.1c00500
- 560 Huang B, Guo Q, Niedermeier ML, Cheng J, Engler T, Maurer M, Pautsch A, Baumeister W, Stengel F, Kochanek  
561 S, Fernández-Busnadiego R. 2021a. PolyQ expansion does not alter the Huntingtin-HAP40 complex.  
562 *bioRxiv* 2021.02.02.429316. doi:10.1101/2021.02.02.429316
- 563 Huang B, Seefelder M, Buck E, Engler T, Lindenberg KS, Klein F, Landwehrmeyer GB, Kochanek S. 2021b. HAP40  
564 protein levels are huntingtin-dependent and decrease in Huntington disease. *Neurobiol Dis*  
565 **158**:105476. doi:10.1016/j.nbd.2021.105476
- 566 Jimenez-Sanchez M, Licitra F, Underwood BR, Rubinsztein DC. 2017. Huntington’s Disease: Mechanisms of  
567 Pathogenesis and Therapeutic Strategies. *Cold Spring Harb Perspect Med* **7**:a024240.  
568 doi:10.1101/cshperspect.a024240
- 569 McColgan P, Tabrizi SJ. 2018. Huntington’s disease: a clinical review. *Eur J Neurol* **25**:24–34.  
570 doi:10.1111/ene.13413
- 571 Menalled L, El-Khodor BF, Patry M, Suárez-Fariñas M, Orenstein SJ, Zahasky B, Leahy C, Wheeler V, Yang XW,  
572 MacDonald M, Morton AJ, Bates G, Leeds J, Park L, Howland D, Signer E, Tobin A, Brunner D. 2009.  
573 Systematic behavioral evaluation of Huntington’s disease transgenic and knock-in mouse models.  
574 *Neurobiol Dis* **35**:319–336. doi:10.1016/j.nbd.2009.05.007
- 575 Mitchell J. 2010. Small molecule immunosensing using surface plasmon resonance. *Sensors (Basel)* **10**:7323–  
576 7346. doi:10.3390/s100807323
- 577 Nath S, Munsie LN, Truant R. 2015. A huntingtin-mediated fast stress response halting endosomal trafficking is  
578 defective in Huntington’s disease. *Hum Mol Genet* **24**:450–462. doi:10.1093/hmg/ddu460
- 579 Ng CW, Yildirim F, Yap YS, Dalin S, Matthews BJ, Velez PJ, Labadorf A, Housman DE, Fraenkel E. 2013. Extensive  
580 changes in DNA methylation are associated with expression of mutant huntingtin. *Proc Natl Acad Sci U*  
581 *S A* **110**:2354–2359. doi:10.1073/pnas.1221292110
- 582 Niesen FH, Berglund H, Vedadi M. 2007. The use of differential scanning fluorimetry to detect ligand  
583 interactions that promote protein stability. *Nat Protoc* **2**:2212–2221. doi:10.1038/nprot.2007.321
- 584 Peters MF, Ross CA. 2001. Isolation of a 40-kDa Huntingtin-associated Protein. *J Biol Chem* **276**:3188–3194.  
585 doi:10.1074/jbc.M008099200
- 586 Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM  
587 structure determination. *Nat Methods* **14**:290–296. doi:10.1038/nmeth.4169
- 588 Ratovitski T, Chighladze E, Arbez N, Boronina T, Herbrich S, Cole RN, Ross CA. 2012. Huntingtin protein  
589 interactions altered by polyglutamine expansion as determined by quantitative proteomic analysis. *Cell*  
590 *Cycle* **11**:2006–2021. doi:10.4161/cc.20423

Toolkit of HTT subdomain constructs

- 591 Ratovitski T, O’Meally RN, Jiang M, Chaerkady R, Chighladze E, Stewart JC, Wang X, Arbez N, Roby E, Alexandris  
592 A, Duan W, Vijayvargia R, Seong IS, Lavery DJ, Cole RN, Ross CA. 2017. Post-Translational Modifications  
593 (PTMs), Identified on Endogenous Huntingtin, Cluster within Proteolytic Domains between HEAT  
594 Repeats. *J Proteome Res* **16**:2692–2708. doi:10.1021/acs.jproteome.6b00991
- 595 Rui Y-N, Xu Z, Patel B, Chen Z, Chen D, Tito A, David G, Sun Y, Stimming EF, Bellen HJ, Cuervo AM, Zhang S.  
596 2015. Huntingtin functions as a scaffold for selective macroautophagy. *Nat Cell Biol* **17**:262–275.  
597 doi:10.1038/ncb3101
- 598 Sakamoto KM, Kim KB, Kumagai A, Mercurio F, Crews CM, Deshaies RJ. 2001. Protacs: chimeric molecules that  
599 target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. *Proc Natl Acad Sci*  
600 *USA* **98**:8554–8559. doi:10.1073/pnas.141230798
- 601 Saudou F, Humbert S. 2016. The Biology of Huntingtin. *Neuron* **89**:910–926. doi:10.1016/j.neuron.2016.02.003
- 602 Schapira M, Calabrese MF, Bullock AN, Crews CM. 2019. Targeted protein degradation: expanding the toolbox.  
603 *Nat Rev Drug Discov* **18**:949–963. doi:10.1038/s41573-019-0047-y
- 604 Shen M, Wang F, Li M, Sah N, Stockton ME, Tidei JJ, Gao Y, Korabelnikov T, Kannan S, Vevea JD, Chapman ER,  
605 Bhattacharyya A, van Praag H, Zhao X. 2019. Reduced mitochondrial fusion and Huntingtin levels  
606 contribute to impaired dendritic maturation and behavioral deficits in Fmr1-mutant mice. *Nat Neurosci*  
607 **22**:386–400. doi:10.1038/s41593-019-0338-y
- 608 Southwell AL, Warby SC, Carroll JB, Doty CN, Skotte NH, Zhang W, Villanueva EB, Kovalik V, Xie Y, Pouladi MA,  
609 Collins JA, Yang XW, Franciosi S, Hayden MR. 2013. A fully humanized transgenic mouse model of  
610 Huntington disease. *Hum Mol Genet* **22**:18–34. doi:10.1093/hmg/dd3397
- 611 Tomoshige S, Nomura S, Ohgane K, Hashimoto Y, Ishikawa M. 2017. Discovery of Small Molecules that Induce  
612 the Degradation of Huntingtin. *Angew Chem Int Ed Engl* **56**:11530–11533. doi:10.1002/anie.201706529
- 613 Trepte P, Kruse S, Kostova S, Hoffmann S, Buntru A, Tempelmeier A, Secker C, Diez L, Schulz A, Klockmeier K,  
614 Zenkner M, Golusik S, Rau K, Schnoegl S, Garner CC, Wanker EE. 2018. LuTHy: a double-readout  
615 bioluminescence-based two-hybrid technology for quantitative mapping of protein-protein interactions  
616 in mammalian cells. *Mol Syst Biol* **14**:e8071. doi:10.15252/msb.20178071
- 617 Vitet H, Brandt V, Saudou F. 2020. Traffic signaling: new functions of huntingtin and axonal transport in  
618 neurological disease. *Curr Opin Neurobiol* **63**:122–130. doi:10.1016/j.conb.2020.04.001
- 619 Wanker EE, Ast A, Schindler F, Trepte P, Schnoegl S. 2019. The pathobiology of perturbed mutant huntingtin  
620 protein-protein interactions in Huntington’s disease. *J Neurochem* **151**:507–519. doi:10.1111/jnc.14853
- 621 Xu S, Li G, Ye X, Chen D, Chen Z, Xu Z, Daniele M, Tambone S, Ceccacci A, Tomei L, Ye L, Yu Y, Solbach A, Farmer  
622 SM, Stimming EF, McAllister G, Marchionini DM, Zhang S. 2022. HAP40 is a conserved central regulator  
623 of Huntingtin and a potential modulator of Huntington’s disease pathogenesis. *PLoS Genet*  
624 **18**:e1010302. doi:10.1371/journal.pgen.1010302
- 625