

# 1 HAP40 orchestrates huntingtin structure for

## 2 differential interaction with polyglutamine

### 3 expanded exon 1

4 Rachel J. Harding\*<sup>1</sup>, Justin C. Deme<sup>2,3,4</sup>, Johannes F. Hevler<sup>5,6</sup>, Sem Tamara<sup>5,6</sup>, Alexander Lemak<sup>7</sup>,  
5 Jeffrey P. Cantle<sup>8</sup>, Magdalena M. Szewczyk<sup>1</sup>, Xiaobing Zuo<sup>9</sup>, Peter Loppnau<sup>1</sup>, Alma Seitova<sup>1</sup>, Ashley  
6 Hutchinson<sup>1</sup>, Lixin Fan<sup>10</sup>, Matthieu Schapira<sup>1,11</sup>, Jeffrey B. Carroll<sup>8</sup>, Albert J. R. Heck<sup>5,6</sup>, Susan M. Lea<sup>2,3,4</sup>,  
7 Cheryl H. Arrowsmith\*<sup>1,7</sup>

8

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

10 <sup>2</sup> Sir William Dunn School of Pathology, University of Oxford, Oxford, UK

11 <sup>3</sup> Central Oxford Structural Molecular Imaging Centre, University of Oxford, South Parks Road, Oxford,  
12 OX1 3RE

13 <sup>4</sup> Center for Structural Biology, Center for Cancer Research, National Cancer Institute, Frederick, MD  
14 21702, USA.

15 <sup>5</sup> Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and  
16 Utrecht Institute of Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The  
17 Netherlands

18 <sup>6</sup> Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

19 <sup>7</sup> Princess Margaret Cancer Centre and Department of Medical Biophysics, University of Toronto,  
20 Toronto, Ontario M5G 1L7, Canada

21 <sup>8</sup> Behavioral Neuroscience Program, Department of Psychology, Western Washington University,  
22 Bellingham, WA, 98225, United States

23 <sup>9</sup> X-ray Science Division, Argonne National Laboratory, Lemont, Illinois, 60439 USA

24 <sup>10</sup> Basic Science Program, Frederick National Laboratory for Cancer Research, SAXS Core of NCI,  
25 National Institutes of Health, Frederick, Maryland 21701

26 <sup>11</sup> Department of Pharmacology & Toxicology, University of Toronto, Toronto, Ontario M5S 1A8,  
27 Canada

28

29 \*Corresponding authors - [Rachel.Harding@utoronto.ca](mailto:Rachel.Harding@utoronto.ca), [Cheryl.Arrowsmith@uhnresearch.ca](mailto:Cheryl.Arrowsmith@uhnresearch.ca)

30 **Abstract**

31  
32 Huntington's disease results from expansion of a glutamine-coding CAG tract in the huntingtin (HTT)  
33 gene, producing an aberrantly functioning form of HTT. Both wildtype and disease-state HTT form a  
34 hetero-dimer with HAP40 of unknown functional relevance. We demonstrate *in vivo* that HTT and  
35 HAP40 cellular abundance are coupled. Integrating data from a 2.6 Å cryo-electron microscopy  
36 structure, cross-linking mass spectrometry, small-angle X-ray scattering, and modeling, we provide a  
37 near-atomic-level view of HTT, its molecular interaction surfaces and compacted domain architecture,  
38 orchestrated by HAP40. Native mass-spectrometry reveals a remarkably stable hetero-dimer,  
39 potentially explaining the cellular inter-dependence of HTT and HAP40. The polyglutamine tract  
40 containing N-terminal exon 1 region of HTT is dynamic, but shows greater conformational variety in  
41 the mutant than wildtype exon 1. By providing novel insight into the structural consequences of HTT  
42 polyglutamine expansion, our data provide a foundation for future functional and drug discovery  
43 studies targeting Huntington's disease.

## 44 Introduction

45  
46 The autosomal dominant neurodegenerative disorder Huntington's disease (HD) is caused by the  
47 expansion of a CAG repeat tract at the 5' of the *huntingtin* gene above a critical threshold of ~35  
48 repeats<sup>1</sup>. CAG tract expansion corresponds to an expanded polyglutamine tract of the Huntingtin  
49 (HTT) protein which functions aberrantly compared to its unexpanded form<sup>2</sup>. Polyglutamine  
50 expanded HTT is thought to be responsible for disrupting a wide range of cellular processes including  
51 proteostasis<sup>3,4</sup>, transcription<sup>5,6</sup>, mitochondrial function<sup>7</sup>, axonal transport<sup>8</sup> and synaptic function<sup>9</sup>.  
52 HD patients experience a range of physical, cognitive and psychological symptoms and longer repeat  
53 expansions are associated with earlier disease onset<sup>10</sup>. The prognosis for HD patients is poor, with an  
54 average life expectancy of just 18 years from the point of symptom onset and a continuous  
55 deterioration of quality of life through this manifest period. There are currently no disease-modifying  
56 therapies available to HD patients.  
57

58 Huntingtin (HTT) is a 3144 amino acid protein comprised of namesake HEAT (Huntingtin, Elongation  
59 factor 3, protein phosphatase 2A, TOR1) repeats and is hypothesised to function as a scaffold for  
60 larger multi-protein assemblies<sup>11,12</sup>. Many proteomics and interaction studies suggest HTT has an  
61 extensive interactome of hundreds of proteins but the only biophysically and structurally validated  
62 interactor of HTT is the so-called 40-kDa huntingtin-associated protein HAP40<sup>13,14</sup>, an interaction  
63 partner conserved through evolution<sup>15,16</sup>. HAP40 is a TPR domain protein with suggested functions in  
64 endocytosis<sup>17-19</sup>. An earlier 4 Å mid-resolution cryo-electron microscopy (cryo-EM) model of HTT in  
65 complex with HAP40 reveals that the HEAT subdomains of HTT wrap around HAP40 across a large  
66 interaction interface<sup>20</sup>. Biophysical and biochemical analyses comparing purified HTT and HTT-HAP40  
67 samples have revealed that HAP40-bound forms of HTT exhibit reduced aggregation propensity,  
68 greater stability and monodispersity as well as conformational homogeneity<sup>20,21</sup>. Consequently, apo  
69 HTT is a more difficult sample to work with for structural and biophysical characterisation, and several  
70 studies to date have required cross-linking approaches to constrain the HTT molecule to facilitate its  
71 analysis, suggesting HTT-HAP40 interactions may stabilize HTT<sup>22,23</sup>. The biological function of the HTT-  
72 HAP40 complex however, remains elusive, and it is not clear if the function of this complex differs  
73 from apo HTT *in vivo*. It is also not yet understood whether HTT is constitutively bound to HAP40 or  
74 whether apo and HAP40-bound forms of HTT perform different functions in the cell.  
75

76 Current structural information for the full-length HTT molecule sheds little light on the N-terminal  
77 exon 1 region of the protein spanning residues 1-90, which contains the critical polyglutamine and  
78 polyproline tracts. This region of the protein is unresolved in the HTT-HAP40 cryo-EM model (PDBID:  
79 6EZ8; Guo et al., 2018) and therefore the influence of the tract expansion on HTT structure-function  
80 remains unclear. Although many studies have focussed on understanding the effects of polyglutamine  
81 expansion on exon 1 in isolation<sup>24-26</sup>, there is still very little known about this region in the context of  
82 the full-length HTT protein molecule, either in the apo form or in the complex with HAP40. The  
83 intrinsically disordered region (IDR), which spans residues 407-665 is subject to a range of post-  
84 translational modifications, is postulated to be critical in mediating various protein interactions<sup>21,27,28</sup>,  
85 and is also unresolved in the cryo-EM structure. Understanding the function of both wildtype and  
86 expanded forms of HTT is critical as many potential HD treatments currently under clinical  
87 investigation aim to lower HTT expression, using both allele selective or non-selective approaches<sup>29</sup>.  
88 Deeper biological insight into the determinants of cellular HTT protein levels, as well as normal and

89 expanded HTT cellular function would help direct which approaches should be prioritised for long-  
90 term patient therapies.

91  
92 Here, we report *in vivo* studies that show a strong correlation of HTT and HAP40 levels in different  
93 genetic backgrounds, providing evidence for the importance of the HTT-HAP40 complex in a  
94 physiological setting. Combining the power of multiple complementary structural techniques, we  
95 shed light on the missing regions of our high-resolution (2.6 Å) model of HTT-HAP40, including the  
96 biologically critical exon 1 region of HTT and the N-terminal region of HAP40. We demonstrate the  
97 remarkable stability of the HTT-HAP40 complex, potentially explaining *in vivo* codependence of these  
98 two proteins and providing important insight for future drug developments in pursuit of treating HD.

99 **Results**

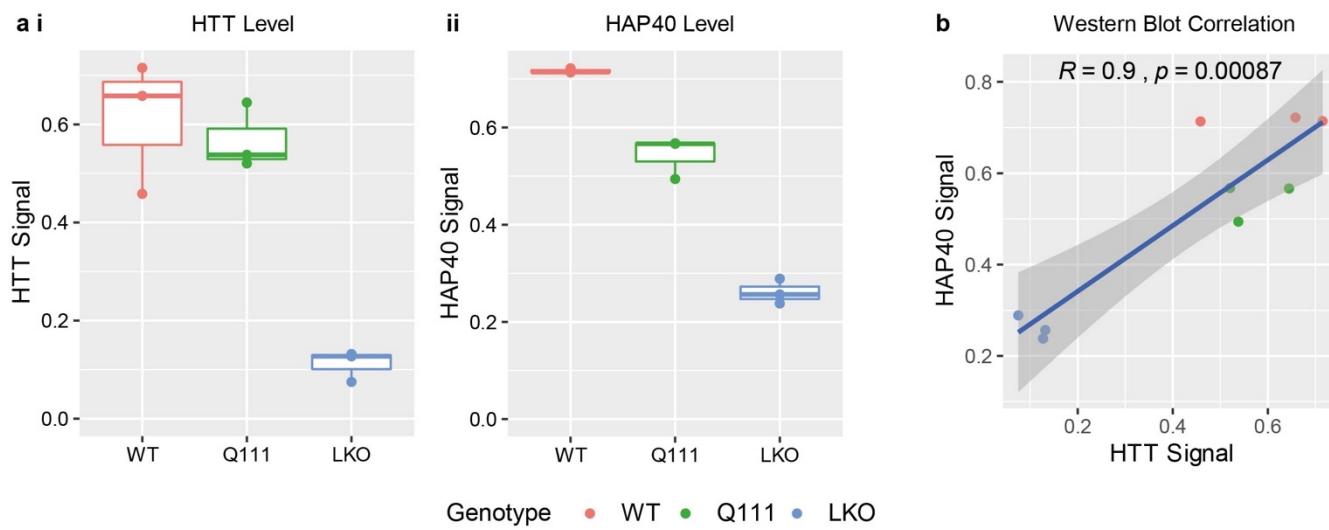
100

101 **HTT and HAP40 protein levels correlate *in vivo*.**

102

103 The Huntingtin-associated protein HAP40 co-evolved with HTT<sup>15</sup> and a HAP40 orthologue has been  
104 identified in many species, including invertebrates<sup>16</sup>. To investigate the *in vivo* relationship and  
105 hypothesised codependency of HTT and HAP40, we analysed the levels of both proteins in liver tissue  
106 from different mouse lines using western blot analysis (**Figure 1**). Comparing wildtype (WT) mice,  
107 Htt<sup>Q111/+</sup> Huntington's knock-in mice<sup>30</sup> which express slightly lower levels of HTT<sup>31</sup>, and hepatocyte-  
108 specific *Htt* knock out mice, a statistically significant correlation was observed for the levels of HTT  
109 and HAP40.

110



111

112

113 **Figure 1. HAP40 levels correlate with the levels of HTT in vivo.**

114

115 a i HTT and ii HAP40 levels were quantified in mouse liver lysates by western blot in wildtype (WT), Htt<sup>Q111/+</sup> and  
116 hepatocyte-specific knockout (LKO) mice. Hepatocytes constitute approximately 80% of liver mass<sup>32</sup> and an approximately  
117 80% reduction in HTT levels is observed in the hepatocyte specific LKO liver tissue as expected. b HTT and HAP40 levels  
correlate in these models with statistical significance.

118 High-resolution structure of HTT-HAP40 Complex

119  
120 HTT-HAP40 was expressed in insect cells and purified as previously described<sup>21</sup>. We determined the  
121 structure of HTT-HAP40 (PDBID: 6X9O) to a nominal resolution of 2.6 Å using cryo-EM (**Figure 2a**,  
122 **Figure 2b** and **Supplementary Figure 1**), improving substantially upon the previously published 4 Å  
123 model (PDBID: 6EZ8; Guo et al., 2018) and two recently deposited models (PDBIDs: 7DXJ [3.6 Å] and  
124 7DKK [4.1 Å]; Huang et al., 2021). Similar to all previous models, flexible regions accounting for ~25%  
125 of the HTT-HAP40 complex, including exon 1 and the IDR, were not resolved in our high-resolution  
126 maps (**Figure 2c**). However, our improved resolution permits more confident positioning of amino  
127 acid side chains of the protein structure resolved in the maps and more precise analysis of the  
128 different features of the structure.

129  
130 The overall structure of the complex is similar to the previously published model (PDBID: 6EZ8) with  
131 an RMSD of 1.9 across the models when superposed. However, key differences exist between the two  
132 models (**Figure 2d**). Two additional C-terminal  $\alpha$ -helices in the HTT C-HEAT domain spanning residues  
133 3105-3137 are resolved in our model (all residue numbering based on HTT NCBI reference  
134 NP\_002102.4 sequence), whereas the resolution of two N-terminal  $\alpha$ -helices of HAP40 spanning  
135 residues 42-82 is lost. The unmodified native HAP40 C-terminus in our model is able to thread into the  
136 centre of the C-HEAT domain (**Figure 2e**). This extended interaction of HAP40 with HTT may be  
137 responsible for a small shift we observe of the C-HEAT domain, which pivots ~5° relative to the  
138 previous model, reducing the interaction interface of HTT-HAP40 from ~5350 Å<sup>2</sup> to ~4700 Å<sup>2</sup>. One  
139 potential reason for this difference is that the C-terminus of HAP40 in our construct is unmodified  
140 whereas Guo and colleagues used a C-terminal Strep-tag in their expression construct which is  
141 unresolved in their model. The differences observed for the HTT and HAP40 interface when  
142 comparing our high-resolution structural model (PDBID: 6X9O) and the previous mid-resolution model  
143 (PDBID: 6EZ8) indicate that the extensive interaction interface is able to accommodate some  
144 variation.

145  
146 Our high-resolution model enables a comprehensive analysis of the surface-charge features of the  
147 HTT-HAP40 complex. The HTT-HAP40 interface is predominantly formed by extensive hydrophobic  
148 interactions between the two proteins (**Figure 2f**). Previous analysis of this interface has also  
149 highlighted a charge-based interaction between the BRIDGE domain of HTT and the C-terminal region  
150 of the HAP40 TPR domain<sup>20</sup>. Interestingly, the N-HEAT domain of HTT has a defined positively  
151 charged tract spanning almost 40 Å in length and 5-10 Å in width formed between two stacked HEAT  
152 repeats in the N-HEAT solenoid (**Figure 2f arrow**). We also conducted an in-depth sequence  
153 conservation analysis of both HTT and HAP40, which we mapped to the high-resolution structure of  
154 the complex. Interestingly this revealed surfaces of the protein on the HAP40-exposed face as highly  
155 conserved, with extended regions of strict conservation partially spanning the C-HEAT domain,  
156 BRIDGE and N-HEAT (**Figure 2g**). However, the opposite face is less conserved, whilst the HTT-HAP40  
157 interface is moderately conserved for both HTT and HAP40. The HTT-HAP40 model was searched for  
158 ligand-able pockets which were assessed for druggability according to various factors, including their  
159 buriedness, hydrophobicity and volume. One of the most promising pockets, which is predicted to be  
160 ligand-able, lies at the HTT-HAP40 interface and is lined by residues from the N-terminal region of the  
161 HAP40 TPR domain as well as the HTT N-HEAT domain (**Figure 2h, Supplementary Table 2**). The high  
162 resolution of our HTT-HAP40 model provides a foundation for virtual screening of such pockets and  
163 other structure-based drug-discovery efforts towards the identification of HTT ligands.

164

165 Our 2.6 Å structure is of sufficient resolution to allow the identification of post-translational  
166 modifications (PTMs). However, no PTMs were observed for any of the resolved residues in the HTT-  
167 HAP40 complex. Native mass spectrometry (MS) analysis, on the other hand, revealed the high purity  
168 of our HTT-HAP40 samples, albeit that a small mass difference (compared to the theoretical mass)  
169 was observed, consistent with the presence of a few PTMs (**Supplementary Figure 2a**). Further  
170 analysis of the HTT-HAP40 complex upon Caspase6 digestion revealed these PTMs to be primarily  
171 phosphorylations (at least two), which could be mapped to the regions spanning 586-2647 and 2647-  
172 3144 of the HTT sequence (**Supplementary Figure 2b, c and d**). Based on the cumulative evidence  
173 from the MS data, these modifications reside within the two flexible portions of HTT not resolved in  
174 our cryo-EM maps. Although many studies have identified numerous different sites and possible  
175 PTMs of the HTT protein<sup>21,27,28,34</sup>, these approaches have so far been qualitative and do not give us a  
176 good understanding of the key proteoforms the Huntington's disease community is studying in either  
177 *in vitro* or *in vivo* models. Our quantitative top- and middle-down MS approaches suggest many post-  
178 translational modifications are in fact only present at very low abundance, at least in our insect cell  
179 expressed samples.

180

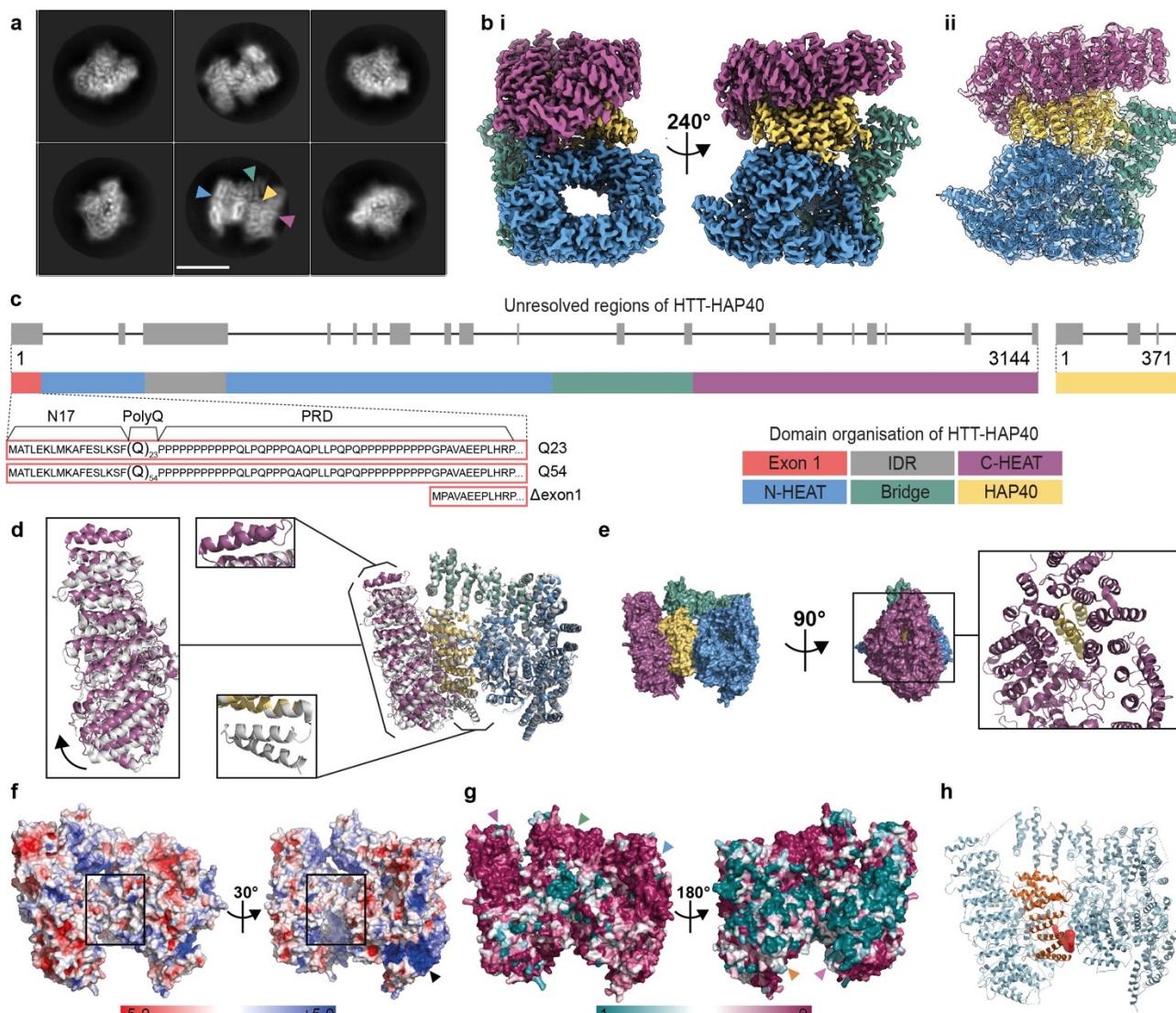
181 We attempted to separately purify HTT and HAP40 for comparison to the complex. As reported by  
182 Guo and colleagues<sup>20</sup>, we were also unable to express recombinant HAP40 alone, although it is  
183 readily expressed in the presence of HTT, a trend that parallels our *in vivo* observations. In the  
184 absence of HAP40, we and others have shown that recombinant HTT self-associates and is  
185 conformationally heterogenous *in vitro*<sup>21,22,34</sup>. Cryo-EM analysis of our apo HTT samples yielded a 12  
186 Å resolution envelope (**Figure 3a and b**). Despite the low resolution of this envelope, it is possible to  
187 identify the N-HEAT domain, with its central cavity, as well as the C-HEAT domain. The HTT portion of  
188 our HTT-HAP40 model can be fitted into this envelope. Comparison of this envelope with the  
189 previously reported apo HTT cryo-EM envelopes that were stabilized by cross-linking (EMD4937 and  
190 EMD10793;<sup>22</sup> shows a less collapsed arrangement of the HTT subdomains. The difference in  
191 resolution between apo HTT and HTT-HAP40 samples observed by cryo-EM analysis emphasizes the  
192 importance of HAP40 in stabilising the HTT protein and constraining the HEAT repeat subdomains into  
193 a more rigid conformation, further supporting the idea that this is a critical interaction for modulating  
194 HTT structure and function.

195

196 Native top-down MS uses gas-phase activation to dissociate protein complexes enabling identification  
197 of complex composition and subunit stoichiometries. The most commonly used activation method  
198 using collisions with neutral gas molecules typically results in dissociation of a non-covalent complex  
199 into constituent subunits. Interestingly, our native top-down MS analysis of the intact HTT-HAP40  
200 complex (**Figure 4a and b**) primarily resulted in backbone fragmentation of HTT, eliminating both N-  
201 and C-terminal fragments (**Figure 4c-g**). Remarkably, the vast majority of concomitantly formed high-  
202 mass dissociation products retained HAP40 (**Figure 4f**), suggesting that the extensive hydrophobic  
203 interaction interface we observe in our high-resolution model keeps the HTT-HAP40 complex  
204 exceptionally stable. Similarly, gas-phase activation of Caspase6-treated HTT-HAP40 revealed that  
205 HAP40 remained intact and bound to HTT even at the highest activation energies, whereas the N- and  
206 C-terminal fragments of HTT produced upon digestion were readily dissociating from the complex  
207 (**Supplementary Figure 2c**).

208

209 The recombinant samples of HTT-HAP40 were found to be highly monodisperse (**Figure 4b**), displaying  
210 optimal biophysical properties (see also **Supplementary Figure 3a**). Systematically screening the  
211 stability of the HTT-HAP40 complex using a differential scanning fluorimetry assay indicates the  
212 complex is highly stable under a broad range of buffer, pH and salt conditions (**Supplementary Figure**  
213 **3b** and **c**). Destabilisation of the complex was only observed at low pH (**Figure 4h**). Similarly, the  
214 interaction between HTT and HAP40 is retained upon mild proteolysis of the complex (**Figure 4i**, all  
215 data in **Supplementary Figure 3d**). Following Caspase-6 treatment, the HTT-HAP40 complex remains  
216 associated under native conditions, although HTT cleavage products are observed under denaturing  
217 conditions<sup>35</sup>. Taken together, our studies reveal the high stability of the HTT-HAP40 complex with  
218 resistance to dissociation by native top-down MS, or proteolytic cleavage in solution. These data  
219 further support the high codependence of HTT and HAP40 protein levels in animal models and  
220 possibly HD patients.

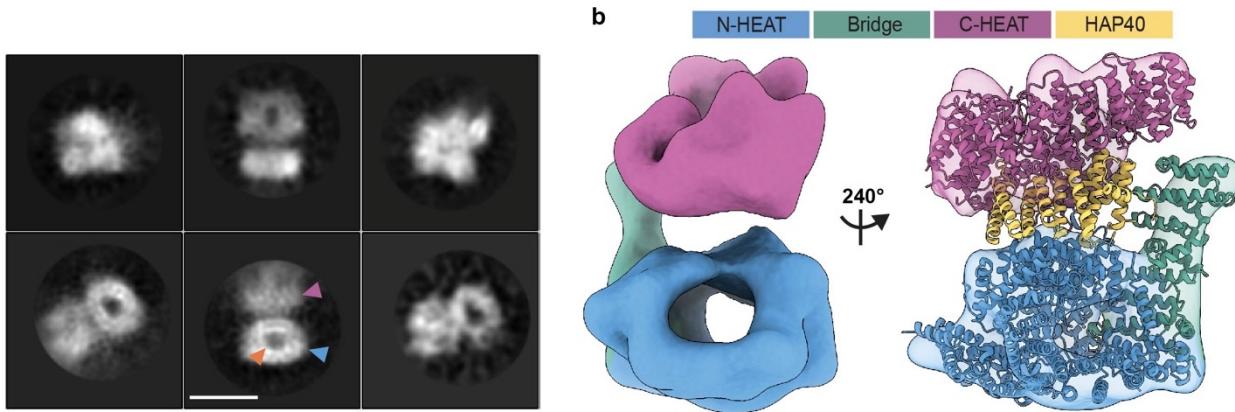


221

222 **Figure 2. HAP40 stabilises the structure of HTT via extensive interactions across all HEAT repeat**  
223 **subdomains.**

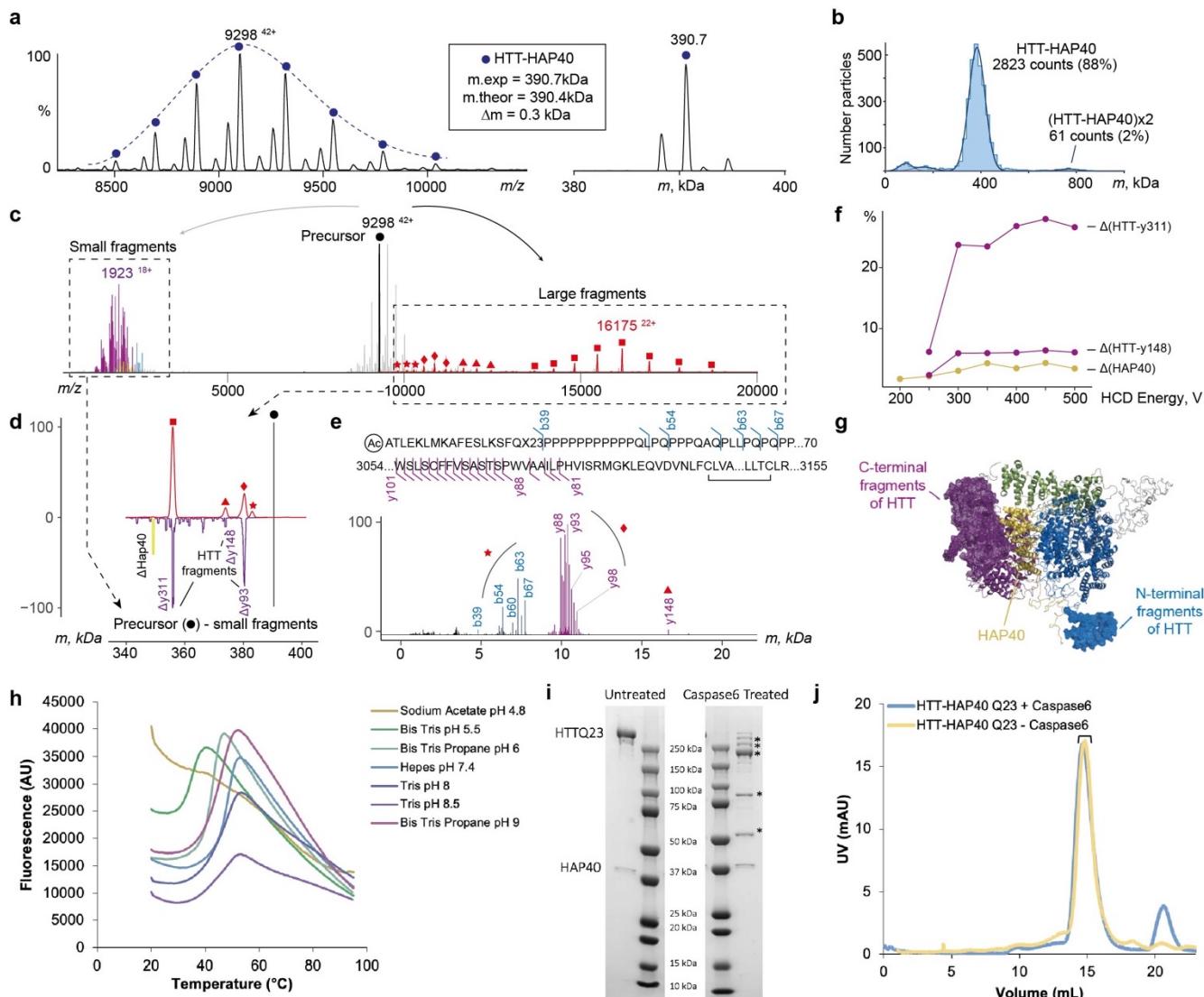
224 a Representative cryo-EM 2D class averages of HTT-HAP40. Scale bar (white) is 90 Å. Blue and purple arrowheads denote  
225 N- and C-HEAT domains of HTT, respectively. Green and yellow arrowheads denote bridge domain of HTT and HAP40,  
226 respectively. b Cryo-EM volume of HTT-HAP40 resolved to 2.6 Å with i HTT N-HEAT in blue, bridge domain in green, C-  
227 HEAT in purple and HAP40 in yellow or ii map shown with HTT-HAP40 modeled in using the same domain colour  
228 convention. c Domain organisation of HTT mapped to linear sequence. Unresolved regions of the structure are in grey and  
229 the three different constructs used in this study are detailed comprising wildtype (23 glutamines; Q23), mutant (54  
230 glutamines; Q54), or HTT with exon 1 partially deleted (Δexon 1; comprising residues 80-3144). d Superposition of our  
231 model (PDBID: 6X9O – same domain colour convention as before) and the previous model (PDBID: 6EZ8 – all grey) with  
232 alignment calculated over N-HEAT and bridge domains. Additional  $\alpha$ -helices observed in either of the models are indicated  
233 with boxes, C-HEAT domain shift is shown with an arrow. e Surface representation of HTT and HAP40 (same domain colour  
234 convention as before) in front and side views, rotated 90°, with additional panel (right) showing same side view of the  
235 complex in cartoon format. f Electrostatic surface representation of HTT with HAP40 removed from the structure.  
236 Positively charged regions are shown in blue, neutral (hydrophobic) regions in white and negatively charged regions in red.  
237 The positively charged tract in the N-HEAT domain is indicated with a black arrowhead. Hydrophobic HTT surface which  
238 binds HAP40, is indicated with hollow black boxes. g Surface representation of HTT-HAP40 complex, coloured according to  
239 Consurf conservation scores: from teal for the least conserved residues (1), to maroon for the most conserved residues  
240 (9). Conserved surfaces for C-HEAT, bridge and N-HEAT domains are indicated with purple, green and blue arrowheads  
241 respectively. Variable N-HEAT and C-HEAT surfaces are indicated with orange and pink arrowheads respectively. h HTT  
242 (pale blue)-HAP40 (orange) complex in cartoon with pocket predicted to be druggable shown as red volume.

243  
244  
245  
246  
247  
248  
249



**Figure 3. HTT HEAT domains are conformationally flexible in the absence of HAP40.**

**a** Representative cryo-EM 2D class averages of HTT Q23. Scale bar (white) shown in the bottom middle panel is 90 Å. Blue arrowhead denotes the N-HEAT domain in which its central cavity (orange arrowhead) is more clearly defined. Purple arrowhead denotes the less well-defined C-HEAT domain, perhaps due to conformational flexibility relative to the N-HEAT.  
**b** Cryo-EM volume of HTT resolved to ~12 Å shown with model of HTT-HAP40 (PDBID: 6X9O) fit to the map. Regions of map and model are displayed with N-HEAT in blue, bridge domain in green, C-HEAT in purple and HAP40 in yellow.



**Figure 4. HTT and HAP40 form a very stable non-covalent complex that withstands dissociation.**

a Raw native (left) and a deconvoluted zero-charged (right) spectrum of the HTT-HAP40 Q23 complex. b Mass profile of HTT-HAP40 complex obtained using mass photometry, showing that the complex is monodisperse. c Composite native top-down mass spectrum of the HTT-HAP40 complex demonstrating large (right of the precursor) and small (left of the precursor) dissociation products produced at the highest activation energy. The data reveal that N- and C-terminal fragments of HTT are eliminated from the HTT-HAP40 complex upon collisional activation, whereas the intact HAP40 remains bound. Small fragment peaks are colored following domain colour convention for the HTT-HAP40 complex. d Mass distribution of the large HTT-HAP40 fragments, mirrored with the mass distribution of precursor mass subtracted the masses of small fragments. e Annotation of small fragments obtained at high-resolution settings and mapping to the sequence of HTT Q23. f Energy-resolved plot of fragment abundances: HTT with HAP40 ejected (yellow), HTT upon release of C-terminal fragment y311 or y148 (purple). g Structure of HTT-HAP40 complex with eliminated regions highlighted and represented as mesh. Colour-coding is in accordance with the domain colour convention for HTT-HAP40. h Assessing HTT-HAP40 Q23 complex stability by measuring transition temperature using DSC in different buffer conditions with 300 mM NaCl. i Caspase6 digestion of HTT-HAP40 Q23 proteins assessed by SDS-PAGE and j analytical gel filtration. Peak fractions from gel filtration run on SDS-PAGE are indicated.

266 Polyglutamine expansion modulates the dynamic sampling of conformational space by exon 1  
267

268 Next, we sought to understand how the disease-causing polyglutamine expansions affect HTT  
269 structure. Our structural, biophysical and biochemical data presented so far focus on wildtype HTT (23  
270 glutamines; Q23) and illustrate the importance of HAP40 in stabilising and orienting the HEAT repeat  
271 subdomains of HTT. However, 25% of the complex is not resolved in the cryo-EM maps, including  
272 many functionally important regions of the protein such as exon 1 (residues 1-90), which harbors the  
273 polyglutamine repeat region, and the IDR (residues 407-665). To further investigate the HTT protein  
274 structure in its entirety and the influence of polyglutamine expansion within exon 1, we repeated the  
275 DSF and proteolysis studies using HTT-HAP40 samples containing either a pathological Huntington's  
276 disease HTT with 54 glutamines (Q54), or an HTT with a partially deleted exon 1 (Δexon 1; comprising  
277 residues 80-3144, missing N17, polyglutamine and proline-rich domain). We found that neither the  
278 Q54 expansion nor the removal of exon 1 had detectable effects on the stability of the HTT-HAP40  
279 complexes compared to the canonical Q23 complex (**Supplementary Figure 3**).  
280

281 To better describe the structure of exon 1 and the effects of the polyglutamine expansion on the HTT-  
282 HAP40 complex, we performed cross-linking mass spectrometry (XL-MS) experiments <sup>36</sup> using the  
283 IMAC-enrichable lysine cross-linker, PhoX <sup>37</sup>. For Q23, Q54 and Δexon1 isoforms of HTT-HAP40, we  
284 mapped approximately 120 cross-links for each sample (**Supplementary Data File 7**). Importantly, the  
285 vast majority of cross-links map to regions unresolved in the cryo-EM maps (**Figure 5a**), thereby  
286 providing valuable restraints for structural modeling of a more complete HTT-HAP40 complex. The  
287 mean distance of cross-links observed for resolved regions of the cryo-EM model was significantly  
288 below the 25 Å distance limit of PhoX in all three datasets (Q23: 7 cross-links – mean distance 13.7 Å;  
289 Q54: 11 cross-links – mean distance 14.8 Å; Δexon 1: 12 cross-links – mean distance 14.9 Å;  
290 **Supplementary Data File 7**). This, together with mass photometry data of cross-linked HTT-HAP40,  
291 indicates that there is a low probability of intermolecular cross-links between HTT molecules, e.g.  
292 from aggregation, being included in our datasets (**Supplementary Figure 4a**).  
293

294 Overall, we obtained very similar cross-link data for the three different HTT-HAP40 constructs (**Figure**  
295 **5b**). However, of particular note are the large number of exon 1 PhoX cross-links in the HTT-HAP40  
296 Q23 and Q54 samples mediated via lysine-6 or lysine-9 within the N-terminal 17 residues (N17 region)  
297 of exon 1. N17 is reported to play key roles for the HTT protein including modulating cellular  
298 localisation, aggregation and toxicity <sup>38-40</sup> and is proposed to interact with distal parts of HTT <sup>41</sup>.  
299

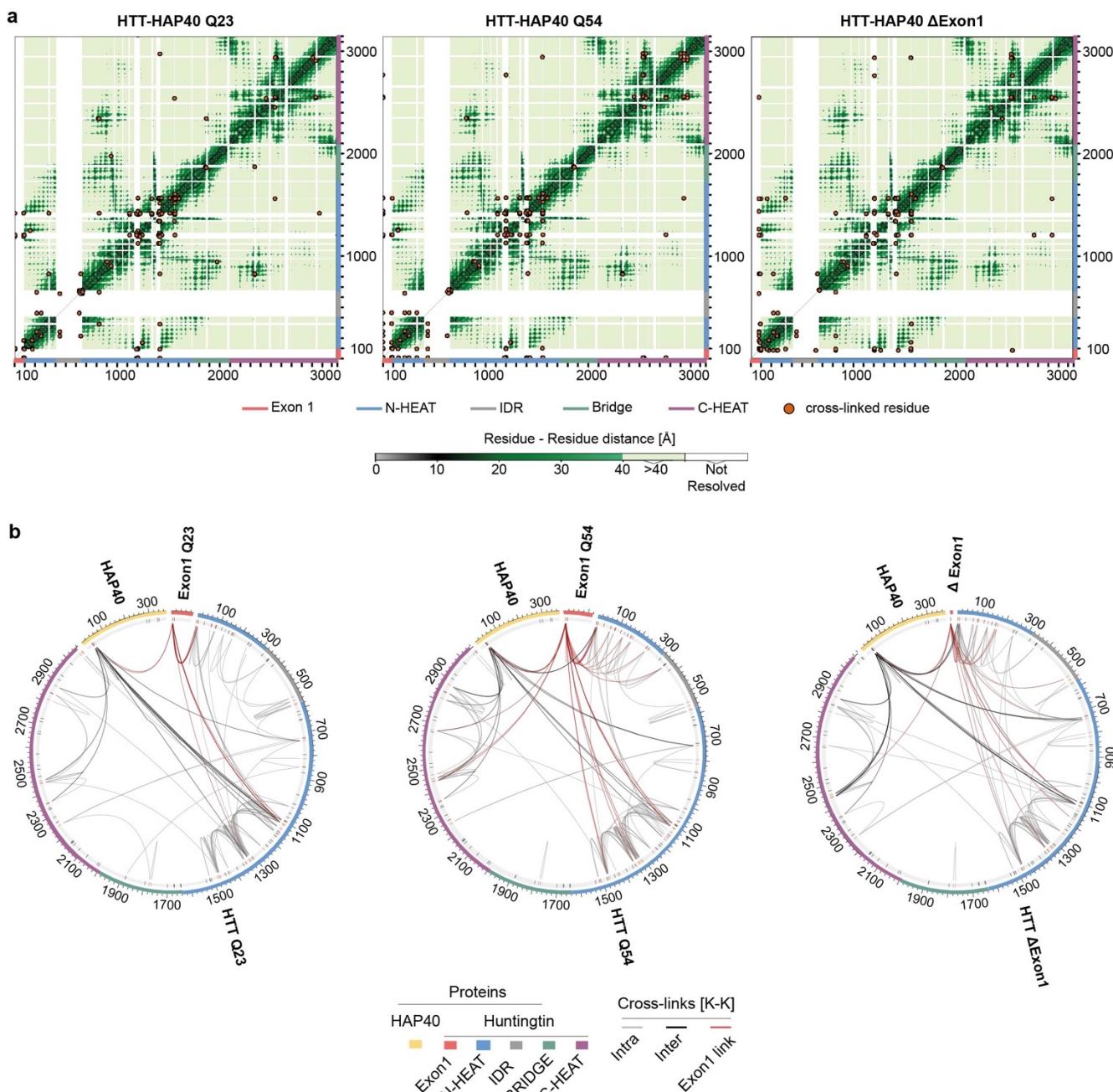
300 For both samples (Q23 and Q54), N17 is found to contact several regions of the N-HEAT domain as  
301 well as the cryo-EM unresolved N-terminal region of HAP40, via lysine-32 and lysine-40. Interestingly,  
302 N17 of Q54 showed additional cross-links to the more distant C-HEAT domain (**Figure 5b**,  
303 **Supplementary Figure 4b**). Finally, the largest uninterrupted stretch of the HTT-HAP40 protein which  
304 is unresolved in the cryo-EM maps is the IDR. However, only a few PhoX cross-links are detected for it,  
305 even though this 258 aa. region harbors 8 lysine residues.  
306

307 Size-exclusion chromatography multi-angle light scattering (SEC-MALS) analysis of this same series of  
308 samples shows no significant difference in mass but does indicate a small shift in the peak for the  
309 elution volume of the HTT-HAP40 Δexon 1 complex compared to Q23 and Q54 complex samples  
310 (**Figure 6a**). Together with the XL-MS data, this suggests that there are subtle structural differences  
311 between the Q23, Q54 and Δexon 1 HTT-HAP40 complexes. To further interpret the cross-linking data

312 in the context of the 3D structure of the HTT-HAP40 complex, we performed SAXS analysis of our  
313 samples to assess any changes to their global structures. We have previously reported SAXS data for  
314 HTT-HAP40 Q23<sup>21</sup>. This revealed that the particle size was significantly larger than the cryo-EM  
315 model, which likely accounts for the ~25% of the protein not resolved in cryo-EM maps and therefore  
316 not modeled in the structure. Similar analysis of the HTT-HAP40 Q54 and HTT-HAP40 Δexon 1 and  
317 comparison with our previous Q23 data shows that polyglutamine expansion or deletion of exon 1 has  
318 only very modest effects on the SAXS profiles (**Figure 6b, c and d**). HTT-HAP40 Q54 is slightly larger  
319 than the HTT-HAP40 Q23 whereas HTT-HAP40 Δexon 1 samples are slightly smaller, as might be  
320 expected, but overall the SAXS determined parameters for the three samples are very similar (**Figure**  
321 **6e**). In line with that, the SAXS-calculated particle envelopes for the three samples are also very  
322 similar in size and shape (**Supplementary Figure 5a**).  
323

324 Next, we modelled the complete structures of HTT-HAP40, including flexible and disordered regions,  
325 integrating our cryo-EM, SAXS and XL-MS data. Coarse-grain modelling molecular dynamics  
326 simulations were performed and an ensemble of models that best fit both the cross-linking and SAXS  
327 data for HTT-HAP40 was calculated for all three variants of the HTT-HAP40 complex (**Supplementary**  
328 **Figure 5b and c**). This modeling approach assumed that the residues with known coordinates in the  
329 cryo-EM model form a quasi-rigid complex, whereas the residues with missing coordinates are  
330 flexible. As expected from our cross-linking results, the conformations adopted by exon 1 in the  
331 ensemble model of Q54 HTT-HAP40 complex are skewed compared to the Q23 ensemble with exon 1  
332 interacting with many more surfaces of the Q54 HTT-HAP40 complex (**Figure 7a**). Mapping our PhoX  
333 exon 1 cross-linked residues for each sample to a representative model from each ensemble reveals  
334 how exon 1 Q23 cross-links are largely constrained to the N-HEAT domain whereas exon 1 Q54 cross-  
335 links are also found on the C-HEAT domain (**Supplementary Figure 4b**). Exon 1 of our HTT-HAP40 Q54  
336 ensemble explores a larger volume of conformational space and this seems to have a knock-on effect  
337 on the conformational space occupied by the IDR (**Figure 7b**). Modeling of our HTT-HAP40 structure  
338 indicates that the exon 1 region of the Q23 HTT is long enough to make cross-links with the C-HEAT  
339 domain, but we do not observe such cross-links in our PhoX datasets (**Supplementary Figure 5d**). This  
340 suggests that the additional cross-links observed for the polyglutamine expanded form of HTT-HAP40  
341 may not be driven solely by the length of the exon 1 region. For all ensembles the IDR is differentially  
342 constrained and occluded from adopting certain conformations depending on the conformational  
343 space occupied by exon 1, suggesting polyglutamine and exon 1-mediated structural changes  
344 propagate to the IDR. For the HTT-HAP40 Q54 model ensemble where exon 1 adopts the most diverse  
345 conformations, the IDR is the most constrained, occupying a more finite space. However, for the HTT-  
346 HAP40 Δexon 1 model ensemble, the IDR is not occluded and so adopts a much wider range of  
347 conformations.  
348

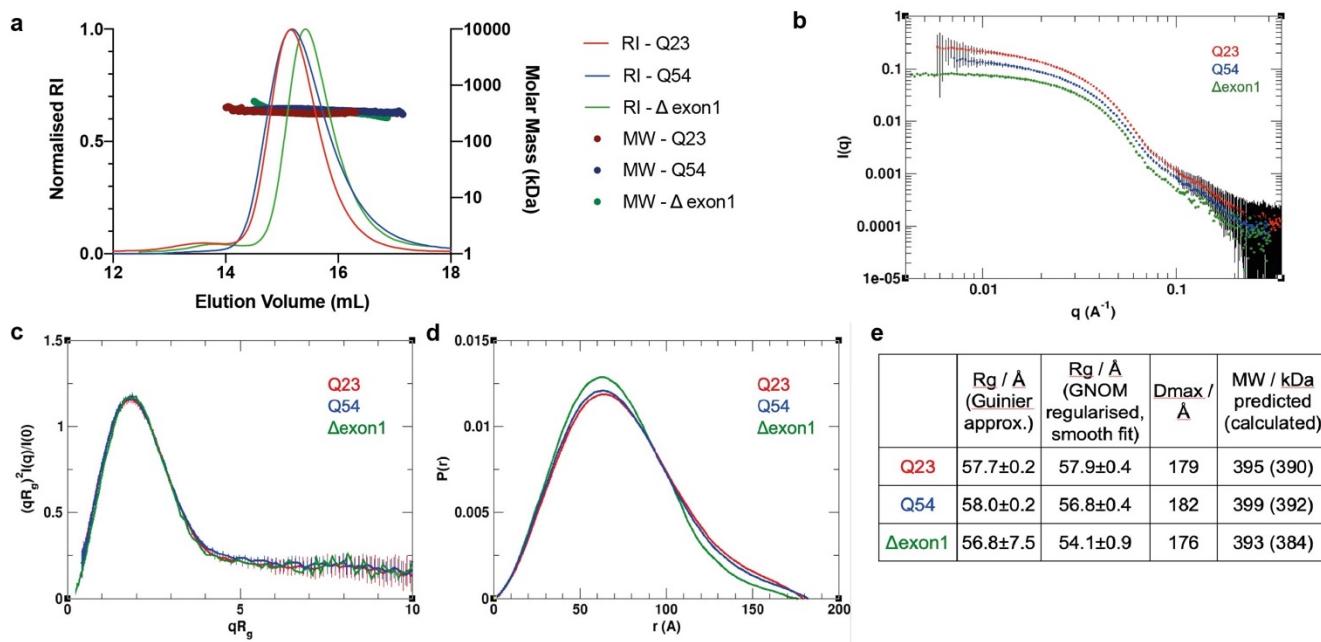
349 Together, our data suggest that whilst polyglutamine expansion does not affect the core HEAT repeat  
350 structure, it does affect the conformational dynamics of not only the exon 1 region but also the IDR.



351

352 **Figure 5. Exon 1 is highly flexible and conformationally dynamic in the context of the full-length**  
353 **protein.**

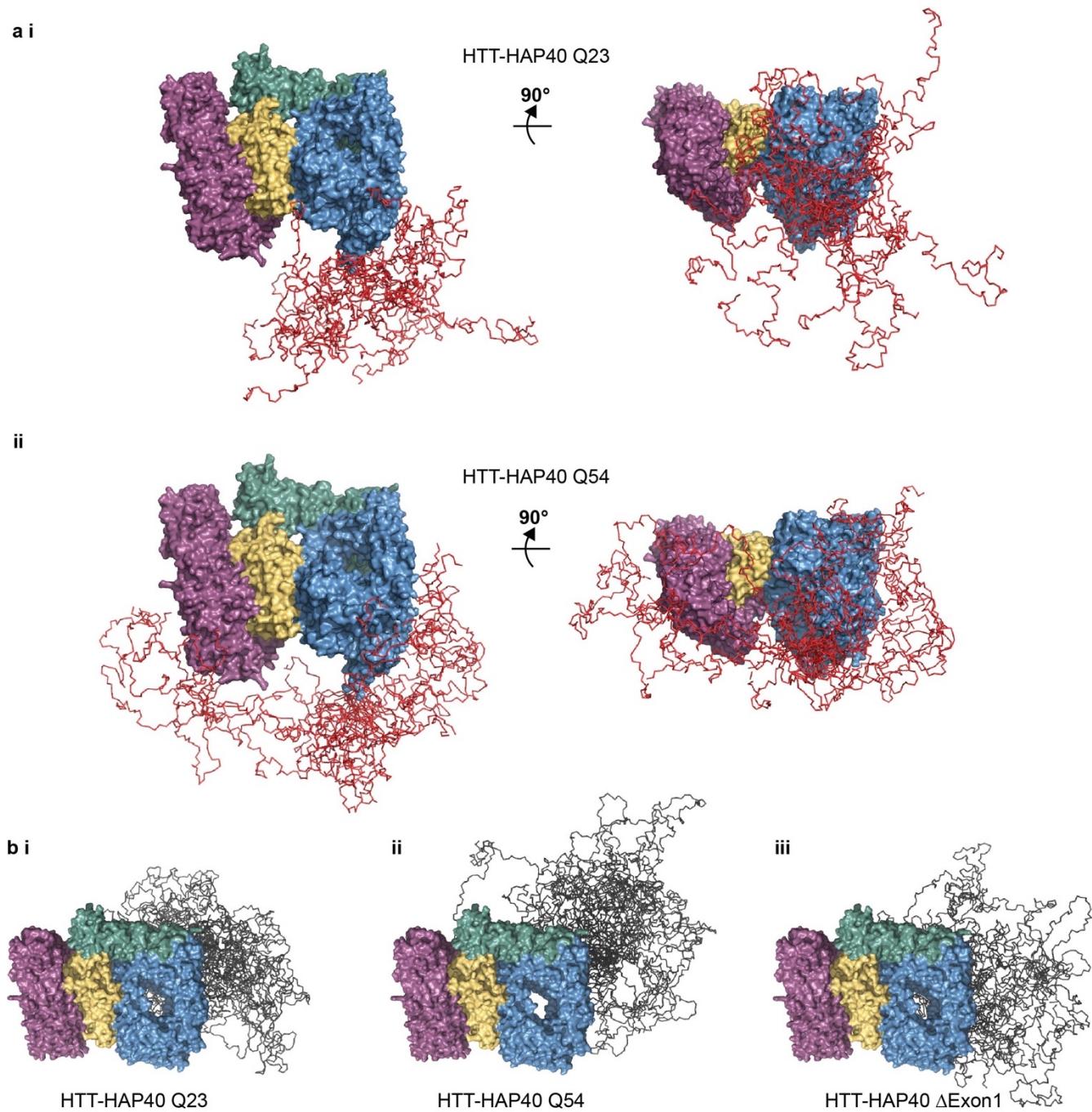
354 **a** Mapping cross-linked sites to the HTT-HAP40 sequence of different samples, with cross-linked residue pairs shown as  
355 orange circles. Intramolecular distances for HTT-HAP40 (PDBID: 6X9O) shown from grey to green as per the coloured scale  
356 bar with unmodelled regions of the protein shown in white. **b** Mapping cross-links to the HTT-HAP40 sequence of different  
357 samples, with exon 1 in red, N-HEAT in blue, bridge domain in green, IDR in grey, C-HEAT in purple and HAP40 in yellow.  
358 Cross-linked lysine residues are indicated in red and unmodified lysine residues are indicated in black on the numbered  
359 sequence. Intermolecular cross-links (HTT-HAP40) are shown in black, intramolecular cross-links (HAP40-HAP40 or HTT-  
360 HTT) are shown in grey and exon 1 cross-links are shown in red. All residues following the exon 1 region of the different  
361 constructs are numbered the same for clarity.



362  
363  
364  
365  
366  
367  
368  
369  
370  
371

**Figure 6. Polyglutamine expansion or deletion of exon 1 has modest effects on the full-length HTT-HAP40 SAXS profile.**

a SEC-MALS analysis of HTT-HAP40 samples Q23 (red), Q54 (blue) and Δexon 1 (green). b Experimental SAXS data. c Rg-based (dimensionless) Kratky plots of experimental SAXS data for HTT-HAP40 Q23 (red), Q54 (blue) and Δexon 1 (green). d Normalized pair distance distribution function P(r) calculated from experimental SAXS data with GNOM for HTT-HAP40 samples. e SAXS parameters for data validation and interpretation including radius of gyration (Rg) calculated using Guinier fit in the q range  $0.015 < q < 0.025 \text{ Å}^{-1}$ , radius of gyration calculated using GNOM, maximum distance between atoms calculated using GNOM, and the molecular mass estimated using SAXSMoW with expected masses from the respective construct sequences shown in the parentheses.



372

373

374

375

376

377

378

379

**Figure 7. Novel insights from integrated model of full-length HTT-HAP40 combining cryo-EM, SAXS and cross-linking mass spectrometry data.**

**a** Ensemble of models for HTT-HAP40 **i** Q23 and **ii** Q54 showing only the residues defined by the cryo-EM model in surface representation (N-HEAT in blue, bridge domain in green, C-HEAT in purple and HAP40 in yellow) and exon 1 simulated residues in ribbon representation (red). **b** Ensemble of models for HTT-HAP40 **i** Q23, **ii** Q54 and **iii** ΔExon 1 showing only the residues defined by the cryo-EM model in surface representation (N-HEAT in blue, bridge domain in green, C-HEAT in purple and HAP40 in yellow) and IDR simulated residues in ribbon representation (grey).

380 Discussion

381

382 We present unprecedented findings for the HTT-HAP40 structure, highlighting the close relationship  
383 between HTT and HAP40 as well as unveiling the effect of the polyglutamine expansion, thereby  
384 contributing to a richer understanding of HTT and its dependence on HAP40.

385

386 HTT is reported to interact with hundreds of different proteins<sup>14</sup> but very few have been validated  
387 and the only interaction partner resolved by structural methods is HAP40. HAP40 is thought to have  
388 coevolved with HTT<sup>15</sup> and orthologues have been identified in species back to flies<sup>16</sup>. The  
389 codependence of HTT and HAP40 is highlighted with our *in vivo* analysis of HTT and HAP40 levels in  
390 mice which shows a strong correlation of the two proteins. It remains to be seen if HTT and HAP40 are  
391 in fact constitutively bound to each other, or if they may exist independently or in complex with other  
392 binding partners. HAP40 plays an important role in stabilising HTT conformation as we have shown  
393 with our biophysical and structural comparison of apo and HAP40-bound HTT samples, but the  
394 molecular mechanisms of how HAP40 functions in endosome transport<sup>17,18</sup> or modulating HTT  
395 toxicity in HD models<sup>16</sup> remains to be determined. Interestingly, despite the exceptional stability of  
396 the HTT-HAP40 interaction, complex integrity was not maintained in our DSF assay at low pH,  
397 conditions similar to that of the local environment of the endosome. The stabilisation of HTT by  
398 HAP40 could be critical for the function of HTT in the stress response to maintain both its structure  
399 and function<sup>42</sup>.

400

401 How polyglutamine expansion of HTT contributes to changes in protein structure-function remains a  
402 critical and unanswered question in HD research. Previously, we have observed that changes in  
403 polyglutamine tract length seem to have minimal effects on the biophysical properties of HTT and  
404 HTT-HAP40 samples<sup>21</sup>. Similarly, in this study, we find no significant differences between our Q23,  
405 Q54 and Δexon 1 HTT-HAP40 samples when assessing monodispersity by mass photometry and native  
406 MS; thermal stability in a systematic buffer screen by DSF or stabilisation by proteolysis experiments.  
407 The structural differences of Q23, Q54 and Δexon 1 HTT-HAP40 samples are not resolved within the  
408 high-resolution cryo-EM maps we calculated. Our experiments using lower resolution structural  
409 methods such as SAXS and mass spectrometry, which do consider the complete protein molecule,  
410 also show modest differences between the samples. One way we might rationalise this observation  
411 with what we know about HD pathology and huntingtin biology in physiological conditions is that our  
412 experimental systems do not capture any subtle, low abundance or slowly occurring differences of  
413 the samples which could be important in HD progression that occurs very slowly, over decades of a  
414 patient's lifetime. Alternatively, it may be that models of HD pathogenesis which posit that large  
415 changes in HTT's globular structure caused by polyglutamine expansion are incorrect.

416

417 Notwithstanding the above caveats, our cross-linking mass spectrometry studies provide some of the  
418 first insight into the structure of the exon 1 portion of the protein in the context of the full-length,  
419 HAP40-bound form of HTT. In both Q23 and Q54 samples, exon 1 appears to be highly dynamic and  
420 able to adopt multiple conformations. We demonstrate clear and novel structural differences  
421 between the unexpanded and expanded forms of exon 1 in the context of the full-length HTT protein  
422 with expanded Q54 forms of exon 1 sampling different conformational space than unexpanded Q23.  
423 This is not just due to the additional length of this form of exon 1, conferring a higher degree of  
424 flexibility and extension to different regions of the protein but perhaps some biophysical consequence

425 of a longer polyglutamine tract. This is the opposite of what has been reported for HTT exon 1 protein  
426 in isolation, where polyglutamine expansion compacts the exon 1 structure<sup>42-44</sup>. Our data suggest  
427 that in the context of the full-length HAP40-bound HTT protein, exon 1 is not compact, but flexible  
428 and conformationally dynamic whilst retaining moderate structural organisation. Our modelling  
429 studies interestingly suggest that the change in exon 1 conformational sampling upon polyglutamine  
430 expansion may have consequent effects on the relative conformations and orientations of the IDR, a  
431 novel insight to HTT structure. Both exon-1 and the IDR have been highlighted as functionally  
432 important regions of HTT, as sites of dynamic PTMs and protease recognition concentrate in these  
433 regions. Our results suggest that structural changes in exon-1 induced by polyglutamine expansion  
434 could influence the accessibility of the IDR to partner proteins which modify residues within the IDR,  
435 despite the relatively rigid intervening regions between them. The flexibility we observe for exon 1 in  
436 both Q23 wildtype and Q54 mutant HTT-HAP40 supports the hypothesis that polyglutamine tracts can  
437 function as sensors, sampling and responding to their local environment<sup>45</sup>.  
438  
439 Overall, our findings show that HTT is stabilised by interaction with HAP40 through an extensive  
440 hydrophobic interface with its distinct HEAT repeat subdomains, creating a highly stable complex.  
441 Expanded and unexpanded exon 1 remains highly dynamic in the context of this complex, sampling a  
442 vast range of conformational space and interacting with different regions of both HTT and HAP40. We  
443 present novel insight into the structural differences of wildtype and mutant HTT, which suggests the  
444 conformational constraints of wildtype and mutant exon 1 are significantly different.

445 **Methods**

446

447 **In vivo HTT-HAP40 levels:**

448 Liver tissue was harvested from *Htt*<sup>Q111/+</sup> mice (JAX:003456) and their WT littermates at 5-6 months of  
449 age. To generate samples with genetic reduction of HTT levels in the liver, mice in which the first exon  
450 of *Htt* is flanked by LoxP sites<sup>46</sup> were crossed with mice expressing CRE recombinase from the *Alb*  
451 promoter (JAX:003574). Liver lysates were prepared for western blotting using non-denaturing lysis  
452 buffer (20mM Tris HCl pH8, 127mM NaCl, 1% NP-40, 2mM EDTA), with 50ug of protein separated  
453 using 3-8% tris-acetate gels (Invitrogen EA0378) and transferred using an iBlot2 transfer system  
454 (Invitrogen IB21001). Probing with antibodies against HTT (Abcam EPR5526; 1:1000) and HAP40  
455 (Novus NBP2-54731; 1:500) was performed with overnight incubation at 4C with gentle shaking,  
456 followed by incubation with near infrared secondary antibodies (Licor 926-68073; 1:10,000). Signal  
457 was normalized to total protein in the lane (Licor 926-11010). Imaging was performed using a Odyssey  
458 imager and signal quantitated using ImageStudio (Licor). All procedures were reviewed and approved  
459 by the animal care and use committee at Western Washington University.

460

461 **Protein expression constructs:**

462 HTT Q23, HTT Q54 and HAP40 constructs used in this study have been previously described<sup>21</sup> and are  
463 available through Addgene with accession numbers 111726, 111727 and 124060 respectively. HTT  
464  $\Delta$ exon 1 clones spanning HTT aa. 80-3144 were also cloned into the pBMDEL vector. A PCR product  
465 encoding HTT from residues P76 to C3140 was amplified from cDNA (Kazusa clone FHC15881) using  
466 primers FWD (ttttaagaaggagatatactatgCCGGCTGTGGCTGAGGAGC) and REV  
467 (gatttggaaaggtagagggttctctgcGCAGGGTGGTGACCTGTGG). PCR products were inserted using the In-Fusion  
468 cloning kit (Clontech) into the pBMDEL that had been linearized with BfuAI. The HTT-coding  
469 sequences of expression constructs were confirmed by DNA sequencing. The sequences were also  
470 confirmed by Addgene where these reagents have been deposited. This clone is available through  
471 Addgene with accession number 162274.

472

473 **Protein expression and purification:**

474 HTT and HTT-HAP40 protein samples were expressed in insect cells and purified using a similar  
475 protocol as previously described<sup>21</sup>. Briefly, Sf9 cells were infected with P3 recombinant baculovirus  
476 and grown until viability dropped to 80–85%, normally after  $\sim$ 72 h post-infection. For HTT–HAP40  
477 complex production, a 1:1 ratio of HTT:HAP40 P3 recombinant baculovirus was used for infection.  
478 Cells were harvested, lysed with freeze-thaw cycles and then clarified by centrifugation. HTT protein  
479 samples were purified by FLAG-affinity chromatography. FLAG eluted samples were bound to Heparin  
480 FF cartridge (GE) and washed with 10 CV 20 mM HEPES pH 7.4, 50 mM KCl, 1 mM TCEP, 2.5 % glycerol  
481 and eluted with a gradient from 50 mM KCl buffer to 1 M KCl buffer over 10 CV. All samples were  
482 purified with a final gel filtration step, using a Superose6 10/300 column in 20 mM HEPES pH 7.4, 300  
483 mM NaCl, 1 mM TCEP, 2.5 % (v/v) glycerol. HTT-HAP40 samples were further purified with an  
484 additional Ni-affinity chromatography step prior to gel filtration. Fractions of the peaks corresponding  
485 to the HTT monomer or HTT-HAP40 heterodimer were pooled, concentrated, aliquoted and flash  
486 frozen prior to use in downstream experiments. Sample purity was assessed by SDS-PAGE. The sample  
487 identities were confirmed by native mass spectrometry (**Figure 5**).

488

489 SDS-PAGE and western blot analysis

490 SDS-PAGE and western blot analysis were performed according to standard protocols. Primary  
491 antibodies used in western blots are anti-HTT EPR5526 (Abcam), anti-HTT D7F7 (Cell Signaling  
492 Technologies) and anti-Flag #F4799 (Sigma). Secondary antibodies used in western blots are goat-anti-  
493 rabbit IgG-IR800 (Licor) and donkey anti-mouse IgG-IR680 (Licor). Membranes were visualized on an  
494 Odyssey® CLx Imaging System (LI-COR).  
495

496 Differential scanning fluorimetry (DSF) analysis of HTT samples

497 HTT samples were diluted in different buffer conditions and incubated at room temperature for 15  
498 minutes before the addition of Sypro Orange (Invitrogen) to a final concentration of 5X. The final  
499 protein concentration was 0.15 mg/mL. Measurements were performed using a Light Cycler 480 II  
500 instrument from Roche Applied Science over the course of 20-95 °C. Temperature scan curves were  
501 fitted to a Boltzmann sigmoid function, and the transition temperature values were obtained from the  
502 midpoint of the transition.  
503

504 Caspase6 proteolysis of HTT protein samples

505 HTT protein samples were mixed with recombinant Caspase6 (Enzo Life Sciences) in a ratio of 100 U  
506 caspase6 to 1 pmol of HTT in 20 mM HEPES pH 7.4, 150 mM NaCl, 1 mM TCEP with a final protein  
507 concentration of ~1 µM. The reaction and control mixture without caspase6 were incubated at room  
508 temperature for 16 hours and then analysed by SDS-PAGE, blue native PAGE and analytical gel  
509 filtration using a Superose6 10/300 column in 20 mM HEPES pH 7.4, 150 mM NaCl, 1 mM TCEP.  
510

511 Cross-linking of HTT-HAP40 samples with PhoX

512 For cross-linking experiments, HTT-HAP40 samples (HTTQ23-HAP40, HTTQ54-HAP40, HTT Δexon 1-  
513 HAP40) were diluted to a protein concentration of 1 mg/1 mL using cross-linking buffer (20 mM Hepes  
514 pH 7.4, 300 mM NaCl, 2.5 % glycerol, 1 mM TCEP). HTT-HAP40 samples were treated with an  
515 optimised concentration of PhoX cross-linker to avoid protein aggregation (**Supplementary Figure 4a**).  
516 After incubation with PhoX (0.5 mM) for 30 min at RT, the reaction was quenched for additional 30  
517 min at RT by the addition of Tris HCl (1 M, pH 7.5) to a final concentration of 50 mM. Protein digestion  
518 was performed in 100 mM Tris-HCl, pH 8.5, 1 % SDC, 5 mM TCEP and 30 mM CAA, with the addition of  
519 Lys-C and Trypsin proteases (1:25 and 1:100 ratio (w/w)) overnight at 37 °C. The reaction was stopped  
520 by addition of TFA to a final concentration of 0.1 % or until pH ~ 2. Next, peptides were desalting using  
521 an Oasis HLB plate, before IMAC enrichment of cross-linked peptides like previously described <sup>37</sup>.  
522

523 LC-MS analysis of cross-linked HTT-HAP40 samples

524 For LC-MS analysis, the samples were re-suspended in 2 % formic acid and analyzed using an  
525 UltiMate™ 3000 RSLCnano System (Thermo Fischer Scientific) coupled on-line to either a Q Exactive  
526 HF-X (Thermo Fischer Scientific), or an Orbitrap Exploris 480 (Thermo Fischer Scientific). Firstly,  
527 peptides were trapped for 5 min in solvent A (0.1 % FA in water), using a 100-µm inner diameter 2-cm  
528 trap column (packed in-house with ReproSil-Pur C18-AQ, 3 µm) prior to separation on an analytical  
529 column (50 cm of length, 75 µM inner diameter; packed in-house with Poroshell 120 EC-C18, 2.7 µm).  
530 Peptides were eluted following a 45 or 55 min gradient from 9-35 % solvent B (80 % ACN, 0.1 % FA),  
531 respectively 9-41 % solvent B. On the Q Exactive HF-X a full scan MS spectra from 375-1600 Da were  
532 acquired in the Orbitrap at a resolution of 60,000 with the AGC target set to 3 x 10<sup>6</sup> and maximum  
533 injection time of 120 ms. For measurements on the Orbitrap Exploris 480, a full scan MS spectra from

534 375-2200 m/z were acquired in the Orbitrap at a resolution of 60,000 with the AGC target set to 2 x  
535 106 and maximum injection time of 25 ms. Only peptides with charged states 3-8 were fragmented,  
536 and dynamic exclusion properties were set to n = 1, for a duration of 10 s (Q Exactive HF-X),  
537 respectively 15 s (Orbitrap Exploris 480). Fragmentation was performed using in a stepped HCD  
538 collision energy mode (27, 30, 33 % Q Exactive HF-X; 20, 28, 36 % Orbitrap Exploris 480) in the ion trap  
539 and acquired in the Orbitrap at a resolution of 30,000 after accumulating a target value of 1 x 105  
540 with an isolation window of 1.4 m/z and maximum injection time of 54 ms (Q Exactive HF-X),  
541 respectively 55 ms Orbitrap Exploris 480.  
542

#### 543 Data analysis of HTT-HAP40 cross-links

544 Raw files for cross-linked HTT-HAP40 samples were analyzed using the XlinkX node <sup>47</sup> in Proteome  
545 Discoverer (PD) software suit 2.5 (Thermo Fischer Scientific), with signal to noise threshold set to 1.4.  
546 Trypsin was set as a digestion enzyme (max. two allowed missed cleavages), the precursor tolerance  
547 set to 10 ppm and the maximum FDR set to 1 %. Additionally, carbamidomethyl modification (Cystein)  
548 was set as fixed modification and acetylation (protein N-terminus) and oxidation (Methionine) were  
549 set as dynamic modifications. Cross-links obtained for respective HTTQ-HAP40 samples were filtered  
550 (only cross-links identified with an XlinkX score > 40 were considered) and further validated using our  
551 recently deposited structure of HTTQ23-HAP40 (PDBID: 6X9O) (EMD-22106). Contact maps and circos  
552 plots were generated in R (<http://www.R-project.org/>) using the circlize <sup>48</sup> and XLmaps <sup>49</sup> packages.  
553

#### 554 Mass photometry

555 Mass photometry analysis was performed on a Refeyn OneMP instrument (Oxford, UK), which was  
556 calibrated using a native marker protein mixture (NativeMark Unstained Protein Standard, Thermo  
557 Scientific). The marker contained proteins in the wide mass range up to 1.2 MDa. Four proteins were  
558 used to generate a standard calibration curve, with following rounded average masses: 66, 146, 480,  
559 and 1048 kDa. The experiments were conducted using glass coverslips, extensively cleaned through  
560 several rounds of washing with Milli-Q water and isopropanol. A set of 4-6 gaskets made of clear  
561 silicone was placed onto the thoroughly dried glass surface to create wells for sample load. Typically,  
562 1  $\mu$ L of HTT samples was applied to 19  $\mu$ L of PBS resulting in a final concentration of  $\sim$  5 nM. Movies  
563 consisting of 6000 final frames were recorded using AcquireMP software at a 100 Hz framerate.  
564 Particle landing events were automatically detected amounting to  $\sim$  3000 per acquisition. The data  
565 was analyzed using DiscoverMP software. Average masses of HTT proteins and HTT-HAP40 complexes  
566 were determined by taking the value at the mode of the normal distribution fitted into the histograms  
567 of particle masses. Finally, probability density function was calculated and drawn over the histogram  
568 to produce the final mass profile. Measurement and analysis of mass photometry data were done for  
569 the following samples: HTT-Q23-HAP40, HTT-Q54-HAP40, and HTT- $\Delta$ exon 1-HAP40.  
570

#### 571 Intact mass and middle-down MS sample preparation

572 Sample preparation: Samples containing HTT-HAP40 complexes were digested using human Caspase6  
573 (Enzo Life Sciences, Farmingdale, USA) by adding 200 U of the enzyme to the 20  $\mu$ g of the protein. The  
574 mixture was stored in PBS for 24 hours. Following the digestion, samples were diluted to the final  
575 concentration of 500 ng/ $\mu$ L with 2 % formic acid. Approximately 2  $\mu$ g of the sample were injected for  
576 a single intact mass LC-MS or middle-down LC-MS/MS experiment.  
577

578 LC-MS(/MS) for intact and middle-down MS

579 Produced peptides of HTT were separated using a Vanquish Flex UHPLC (Thermo Fisher Scientific,  
580 Bremen, Germany) coupled on-line to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo  
581 Fisher Scientific, San Jose, USA) via reversed-phase analytical column (MAbPac, 1 mm × 150 mm,  
582 Thermo Fisher Scientific). The column compartment and preheater were kept at 80°C during the  
583 measurements to ensure efficient unfolding and separation of the analyzed peptides. Analytes were  
584 separated and measured for 22 min at a flow rate of 150 µL/min. Elution was conducted using A  
585 (Milli-Q H<sub>2</sub>O/0.1 % CH<sub>3</sub>O<sub>2</sub>) and B (C<sub>2</sub>H<sub>3</sub>N/0.1 % CH<sub>3</sub>O<sub>2</sub>) mobile phases. In the first minute B was  
586 increased from 10 to 30 %, followed by 30 to 57% B gradient over 14 minutes, 1 min 57 to 95 % B  
587 ramp-up, 95 % B for 1 min, and equilibration of the column at 10 % B for 4 min.

588

589 During data acquisition, Lumos Fusion instrument was set to Intact Protein and Low Pressure mode.  
590 MS1 resolution of 7,500 (determined at 200 m/z and equivalent to 16 ms transient signal length) was  
591 used, which enables optimal detection of protein ions above 30 kDa in mass. Mass range of 500-3,000  
592 m/z, the automatic gain control (AGC) target of 250 %, and a max injection time (IT) of 50 ms were  
593 used for recording of MS1 scans. 2 µscans were averaged in the time domain and recorded for the  
594 7,500 resolution scans during the LC-MS experiment and 5 µscans for when tandem MS (MS/MS) was  
595 performed. MS/MS scans were recorded at a resolution setting of 120,000 (determined at 200 m/z  
596 and equivalent to 16 ms transient signal length), 10,000 % AGC target, 250 ms max IT, and five µscans,  
597 for the single most abundant peak detected in the preceding MS1 scan. The selected ions were mass-  
598 isolated by a quadrupole in a 4 m/z window and accumulated to an estimate of 5e6 ions prior to the  
599 gas-phase activation. Two separate LC-MS/MS runs were recorded per sample with either higher-  
600 energy collisional dissociation (HCD) or electron transfer dissociation (ETD) used for fragmentation.  
601 For ETD following parameters were used: ETD reaction time – 16 ms, max IT of the ETD reagent – 200  
602 ms, and the AGC target of the ETD reagent – 1e6. For HCD, 30 V activation energy was used. MS/MS  
603 scans were acquired with the minimum intensity of the precursor set to 5e4 and the range of 350-  
604 5000 m/z using quadrupole in the high mass isolation mode.

605

606 Data analysis of intact and middle-down MS

607 LC-MS data were deconvoluted with ReSpect algorithm in BioPharma Finder 3.2 (Thermo Fisher  
608 Scientific, San Jose, USA). ReSpect parameters: precursor m/z tolerance – 0.2 Th, target mass – 50  
609 kDa, relative abundance threshold – 0 %, mass range – 3-100 kDa; tolerance – 30 ppm, charge range –  
610 3-100. MS1 and MS2 masses were recalibrated using an external calibrant mixture of intact proteins  
611 (PiercePierce™ Intact Protein Standard Mix, Thermo Scientific) measured before and after each HTT  
612 sample. Iterative sequence adjustments of putative HTT peptides was done until the exact precursor  
613 and fragment masses matched to determine a final set of HTT peptides generated by Caspase6  
614 enzyme. HCD fragments of HTT peptides were used solely to confirm identified sequences.  
615 Phosphorylation was matched as 80 Da variable modification mass, added to the mass of the  
616 identified HTT peptides. Visualization was done in R extended with ggplot2 package.

617

618 Native (top-down) MS sample preparation

619 Samples were stored at -80°C in the buffer containing 20 mM HEPES pH 7.4, 300 mM NaCl, 2.5 % (v/v)  
620 glycerol, 1 mM TCEP. Approximately 40 µg of the HTT-Q23, HTT-Q54, HTT-Δexon 1, and their  
621 respective complexes with Hap40 protein were buffer-exchanged into 150 mM aqueous ammonium  
622 acetate (pH=7.5) by using P-6 Bio-Spin gel filtration columns (Bio-rad, Veenendaal, the Netherlands).

623 The protein's resulting concentration was estimated to be ~2-5  $\mu$ M before native MS analysis. For the  
624 recording of denaturing MS, samples were spiked with formic acid to the final concentration of 2%  
625 right before the MS measurement.

626

627 Native (top-down) data acquisition

628 HTT-containing samples were directly injected into a Q Exactive Ultra-High Mass Range (UHMR)  
629 Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) using in-house pulled and  
630 gold-coated borosilicate capillaries. Following mass spectrometer parameters were used: capillary  
631 voltage – 1.5 kV, positive ion mode, source temperature – 250 °C, S-lens RF level – 200, injection time  
632 – mostly 200 ms, noise level parameter – 3.64. In-source trapping with a desolvation voltage of -100 V  
633 was used to desolvate the proteinaceous ions efficiently. No additional acceleration voltage was used  
634 in the back-end of the instrument. The automatic gain control (AGC) was switched to fixed.  
635 Resolutions of 4,375 and 8,750 (both at  $m/z$  = 200 Th) were used, representing 16 and 32 ms  
636 transient, respectively. Ion guide optics and voltage gradient throughout the instrument were  
637 manually adjusted for optimal transmission and detection of HTT and HTT-HAP40 ions. The higher-  
638 energy collisional dissociation (HCD) cell was filled with Nitrogen, and the trapping gas pressure was  
639 set to 3 or 4 setting value, corresponding to ~2e-10 – 4e-10 mBar for the ultra-high vacuum (UHV)  
640 readout of the instrument. The instrument was calibrated in the  $m/z$  range of interest using a  
641 concentrated aqueous cesium iodide (CsI) solution. Acquisition of the spectra was usually performed  
642 by averaging 100-200  $\mu$ scans in the time domain. Peaks corresponding to the protein complex of  
643 interest were isolated with a 20 Th window for single charge state isolation and a 2000 Th window for  
644 charge-state ensemble isolation. In both cases, isolated HTT-HAP40 ions were investigated for  
645 dissociation using elevated HCD voltages, with direct eV setting varied in the range 1-500 V. For  
646 detection of high- $m/z$  dissociation product ions, mass analyzer detection mode and transmission RF  
647 settings were set to “high  $m/z$ ”. For detection of low- $m/z$  fragment ions, all relevant instrument  
648 settings were set to “low  $m/z$ ”, and the instrument resolution was increased to 140,000 (at  $m/z$  = 200  
649 Th).

650

651 Data Analysis for native (top-down) MS

652 Raw native MS and high- $m/z$  native top-down MS data were processed with UniDec<sup>50</sup> to obtain zero-  
653 charged mass spectra. Native top-down MS data recorded with high resolution (140,000) were  
654 deconvoluted using the Xtract algorithm within FreeStyle software (1.7SP1; Thermo Fisher Scientific).  
655 The resulting zero-charge fragments were matched to the theoretical fragments produced for HTT  
656 and Hap40 using in-house scripts with 5 ppm mass tolerance. Final visualization was performed in R  
657 extended with ggplot2 library.

658

659 Cryo-EM sample preparation and data acquisition

660 HTT was diluted to 0.4 mg/ml in 20 mM HEPES pH 7.5, 300 mM NaCl, 1 mM TCEP and adsorbed to  
661 glow-discharged holey carbon-coated grids (Quantifoil 300 mesh, Au R1.2/1.3) for 10 s. Grids were  
662 then blotted with filter paper for 2 s at 100 % humidity at 4 °C and frozen in liquid ethane using a  
663 Vitrobot Mark IV (Thermo Fisher Scientific).

664

665 HTT-HAP40 was diluted to 0.2 mg/ml in 25 mM HEPES pH 7.4, 300 mM NaCl, 0.025 % w/v CHAPS, 1  
666 mM DTT and adsorbed onto gently glow-discharged suspended monolayer graphene grids

667 (Graphenea) for 60 s. Grids were then blotted with filter paper for 1 s at 100 % humidity, 4 °C and  
668 frozen in liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific).

669

670 Data were collected in counting mode on a Titan Krios G3 (FEI) operating at 300 kV with a  
671 BioQuantum imaging filter (Gatan) and K2 direct detection camera (Gatan) at 165,000x magnification,  
672 pixel size of 0.822 Å. Movies were collected over 32 fractions at a dose rate of 6.0 e-/Å<sup>2</sup>/s, exposure  
673 time of 8 s, resulting in a total dose of 48.0 e-/Å<sup>2</sup>.

674

### 675 Cryo-EM data processing

676 For apo HTT, patched motion correction and dose weighting were performed using MotionCor  
677 implemented in RELION 3.0<sup>51</sup>. Contrast transfer function parameters were estimated using CTFFIND4  
678<sup>52</sup>. Particles were picked in SIMPLE 3.0<sup>53</sup> and processed in RELION 3.0. 669 movies were collected in  
679 total and 108,883 particles extracted. Particles were subjected to one round of reference-free 2D  
680 classification against 100 classes (k = 100) using a soft circular mask of 180 Å in diameter in RELION. A  
681 subset of 25,424 particles were recovered at this stage and subjected to 3D auto-refinement in  
682 RELION using a 40 Å lowpass-filtered map of HTT-HAP40 (EMDB 3984) as initial reference. This  
683 generated a ~12 Å map based on gold-standard Fourier shell correlation curves using the 0.143  
684 criterion as calculated within RELION.

685

686 For HTT-HAP40 (**Supplementary Figure 1**), 15,003 movies were processed in real time using the  
687 SIMPLE 3.0 pipeline, using SIMPLE-unblur for patched motion correction, SIMPLE-CTFFIND for patched  
688 CTF estimation and SIMPLE-picker for particle picking. After initial 2D classification in SIMPLE 3.0 using  
689 the cluster2D\_stream module (k = 500), cleaned particles were imported into RELION and subjected to  
690 reference-free 2D classification (k = 200) using a 180 Å soft circular mask. An *ab initio* map, generated  
691 from a selected subset of particles (372,226), was subsequently lowpass filtered to 40 Å and used as  
692 reference for coarse-sampled (7.5°) 3D classification (k = 4) with a 180 Å soft spherical mask against  
693 the same particle subset. Particles (102,729) belonging to the most defined, highest resolution class  
694 were selected for 3D auto-refinement against its corresponding map, lowpass filtered to 40 Å, using a  
695 soft mask covering the protein which generated a 3.5 Å volume. This map was lowpass filtered to 40  
696 Å and used as initial reference for a multi-step 3D classification (k = 5, 15 iterations at 7.5° followed by  
697 5 iterations at 3.75°), with 180 Å soft spherical mask, against the full cleaned dataset of 2,240,373  
698 particles. Selected particles (647,468) from the highest resolution class were subjected to masked 3D  
699 auto-refinement against its reference map, lowpass filtered to 15 Å, yielding a 3.1 Å volume. CTF  
700 refinement using per-particle defocus plus beamtilt estimation further improved map quality to 3.0 Å.  
701 Bayesian particle polishing followed by an additional round of CTF refinement with per-particle  
702 defocus plus beamtilt estimation on a larger box size (448 x 448) generated a final volume with global  
703 resolution of 2.6 Å as assessed by Gold standard Fourier shell correlations using the 0.143 criterion  
704 within RELION. Map local resolution estimation was calculated within Relion (**Supplementary Figure**  
705 **1**). Additional rounds of 3D classification using either global/local searches or classification only  
706 without alignment did not improve map quality.

707

### 708 Model building and refinement

709 The model for HTT-HAP40 (**Supplementary Table 1**) was generated by rigid body fitting the 4 Å HTT-  
710 HAP40 model<sup>20</sup> (PDBID: 6EZ8) into our globally-sharpened, local resolution filtered 2.6 Å map  
711 followed by multiple rounds of manual real-space refinement using Coot v. 0.95<sup>54</sup> and automated

712 real-space refinement in PHENIX v. 1.18.2-38746<sup>55</sup> using secondary structure, rotamer and  
713 Ramachandran restraints. HTT-HAP40 model was validated using MolProbity<sup>56</sup> within PHENIX. Figures  
714 were prepared using UCSF ChimeraX v.1.1<sup>57</sup> and PyMOL v.2.4.0 (The PyMOL Molecular Graphics  
715 System, v.2.0; Schrödinger).

716

717 **SAXS data collection and analysis**

718 SAXS experiments were performed at beamline 12-ID-B of the Advanced Photon Source (APS) at  
719 Argonne National Laboratory. The energy of the X-ray beam was 13.3 keV (wavelength  $\lambda = 0.9322 \text{ \AA}$ ),  
720 and two setups (small- and wide-angle X-ray scattering) were used simultaneously to cover scattering  
721  $q$  ranges of  $0.006 < q < 2.6 \text{ \AA}^{-1}$ , where  $q = (4\pi/\lambda)\sin\theta$ , and  $2\theta$  is the scattering angle. For HTT-HAP40  
722 Q54, thirty two-dimensional images were recorded for buffer or sample solutions using a flow cell,  
723 with an exposure time of 0.8 s to reduce radiation damage and obtain good statistics. The flow cell is  
724 made of a cylindrical quartz capillary 1.5 mm in diameter and 10  $\mu\text{m}$  wall thickness. Concentration-  
725 series measurements for this sample were carried out at 300 K with concentrations of 0.5, 1.0, and  
726 2.0 mg/ml, in 20 mM HEPES, pH 7.5, 300 mM NaCl, 2.5% (v/v) glycerol, 1 mM TCEP. No radiation  
727 damage was observed as confirmed by the absence of systematic signal changes in sequentially  
728 collected X-ray scattering images. The 2D images were corrected for solid angle of each pixel, and  
729 reduced to 1D scattering profiles using the Matlab software package at the beamlines. The 1D SAXS  
730 profiles were grouped by sample and averaged.

731

732 For HTT-HAP40  $\Delta$ exon 1, data were collected using an in-line FPLC AKTA micro setup with a Superose6  
733 Increase 10/300 GL size exclusion column in 20 mM HEPES, pH 7.5, 300 mM NaCl, 2.5% (v/v) glycerol,  
734 1 mM TCEP. A 150uL sample loop was used and the stock sample concentration was 5 mg/ml. The  
735 sample passed through the FPLC column and was fed to the flow cell for SAXS measurements. The  
736 SAXS data were collected every 2 seconds and the X-ray exposure time was set to 0.75 seconds. Only  
737 the SAXS data collected above the half maximum of the elution peak, about 50-100 frames, were  
738 averaged and for further analysis. Background data were collected before and after the peak (each  
739 100 frames), while data before the peak were found better and used for the background subtraction.

740

741 SAXS data were analyzed with the software package ATSAS 2.8 . The experimental radius of gyration,  
742  $R_g$ , was calculated from data at low  $q$  values using the Guinier approximation. The pair distance  
743 distribution function,  $P(r)$ , the maximum dimension of the protein,  $D_{\max}$ , and  $R_g$  in real space were  
744 calculated with the indirect Fourier transform using the program GNOM<sup>60</sup>. Estimation of the  
745 molecular weight of samples was obtained by both SAXMOW<sup>61,62</sup> and by using volume of correlation,  
746  $V_c$ <sup>63</sup>. The theoretical scattering intensity of the atomic structure model was calculated using FoXS<sup>64</sup>.  
747 *Ab-initio* shape reconstructions (molecular envelopes) were performed using both bead modeling  
748 with DAMMIF<sup>65</sup> and calculating 3D particle electron densities directly from SAXS data with DENSS<sup>66</sup>.

749

750 **Coarse-grained molecular dynamics simulations**

751 We used a Gō-like coarse-grained model of HTT/HAP40 for structural modeling of the complex as it  
752 was described previously<sup>21</sup>. We build two different models that are based on two experimental EM  
753 structures of the complex (PDBIDs: 6EZ8 and 6X9O, respectively). We used experimentally observed  
754 cross-links to improve the sampling of the flexible regions of the model by introducing in the force  
755 field a distance restraint term given by the following potential:

756

757

$$V_{XL}(t) = \sum_{k=1}^{N_{XL}} \delta_{\xi(t)}^k V_l^k; \quad V_l^k = K_{XL}/(1 + e^{-\beta(l_k(t) - l_0)})$$

758 The sum is over all cross-links,  $N_{XL}$  is the number of cross-links;  $l_k$  is the  $C_{\alpha}$ - $C_{\alpha}$  distance for residues  
759 involved in  $k$ th cross-link;  $l_0 = 25$  is the upper bound for PhoX cross-links;  $\beta = 0.5$  is the slope of the  
760 sigmoidal function;  $K_{XL} = 10$  kcal/mol is the force constant;  $\delta_i^k$  is the Kronecker delta; and  $\xi(t)$  is the  
761 random digital number selected from the interval  $[1, N_{XL}]$ . We chose to keep active only about  
762  $N_{XL}/3$  randomly selected restraints, numbers  $\xi(t)$ , that are updated every  $\tau_{XL} = 0.5$  ns during the MD  
763 simulation.

764

765 The goodness-of-fit of an ensemble of structural models of the complex to the SAXS data was  
766 evaluated by comparing an ensemble average profile,  $I_{avg}(q)$ , with the experimental one.  $I_{avg}(q)$  was  
767 calculated either by performing simple averaging of model's theoretical scattering intensities over MD  
768 trajectory or by selecting optimal ensemble using SES method<sup>67</sup>. Theoretical scattering profiles for  
769 each conformation in the MD trajectory were calculated in the  $q$  range  $0 < q < 0.30 \text{ \AA}^{-1}$  using FoXS<sup>64</sup>.  
770

771 **Size-exclusion chromatography multi angle light scattering (SEC-MALS)**

772 The absolute molar masses and mass distributions of purified protein samples of HTT-HAP40 Q23,  
773 HTT-HAP40 Q54 and HTT-HAP40 Δexon 1 at 1 mg/ml were determined using SEC-MALS. Samples were  
774 injected through a Superose 6 10/300 GL column (GE Healthcare) equilibrated in 20 mM HEPES, pH  
775 7.5, 300 mM NaCl, 2.5% (v/v) glycerol, 1 mM TCEP followed in-line by a Dawn Heleos-II light scattering  
776 detector (Wyatt Technologies) and a 2414 refractive index detector (Waters). Molecular mass  
777 calculations were performed using ASTRA 6.1.1.17 (Wyatt Technologies) assuming a dn/dc value of  
778 0.185 ml/g.  
779

780 **In silico analysis of the HTT-HAP40 protein complex structure**

781 HTT-HAP40 models were analysed using Pymol<sup>68</sup> and APBS<sup>69</sup>. For conservation analysis, HTT and  
782 HAP40 orthologues were extracted from Ensembl, parsed to remove low quality or partial sequences  
783 and then aligned using Clustal<sup>70</sup>. Multiple sequence alignments were then analysed using Consurf<sup>71</sup>  
784 and conservation scores mapped to the HTT-HAP40 (PDBID: 6X9O) structure in Pymol. Ligandable  
785 pocket analysis was completed as previously reported<sup>72</sup>. Briefly, HTT-HAP40 model pdb files were  
786 loaded in ICM (Molsoft, San Diego). Proteins were protonated, optimal positions of added polar  
787 hydrogens were generated, correct orientation of side-chain amide groups for glutamine and  
788 asparagine and most favourable histidine isomers were identified. The PocketFinder algorithm  
789 implemented in ICM, which uses a transformation of the Lennard-Jones potential to identify ligand  
790 binding envelopes regardless of the presence of bound ligands, was then applied<sup>73</sup>. Residues with  
791 side-chain heavy atoms within 2.8 Å of the molecular envelope were identified as lining the pocket.

792 Acknowledgements

793

794 We acknowledge the use of the SAXS Core Facility of the Center for Cancer Research (CCR), NCI,  
795 National Institutes of Health. NCI SAXS Core is funded by FNLCR contract HHSN261200800001E and  
796 the intramural research program of the NIH, NCI, CCR. This research used 12-ID-B beamline of the  
797 Advanced Photon Source, a United States Department of Energy (DOE) Office of Science User Facility  
798 operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-  
799 06CH11357.

800

801 This research was supported by the CHDI Foundation (RJH, CHA, JBC), the Huntington Society of  
802 Canada (RJH, CHA), the Wellcome Trust #219477 (SML, JD) and the EU Horizon 2020 program INFRAIA  
803 project Epic-XS Project 823839 (JFH, ST, AJRH). RJH is the recipient of the Huntington's Disease Society  
804 of America Berman Topper Career Development Fellowship.

805 The Structural Genomics Consortium is a registered charity (no: 1097737) that receives funds from  
806 AbbVie, Bayer AG, Boehringer Ingelheim, Genentech, Genome Canada through Ontario Genomics  
807 Institute [OGI-196], the EU and EFPIA through the Innovative Medicines Initiative 2 Joint Undertaking  
808 [EUbOPEN grant 875510], Janssen, Merck KGaA (aka EMD in Canada and US), Pfizer, Takeda and the  
809 Wellcome Trust [106169/ZZ14/Z].

810 Author Contributions

811

812 RJH conceived the project, designed and conducted experiments, analysed and interpreted data,  
813 supervised the project and wrote the manuscript. JD, JFH, ST, AL, JPC, MS and XZ designed and  
814 conducted experiments, analysed and interpreted data and contributed to drafting and editing the  
815 manuscript. MMS, AH, AS and PL conducted experiments and analysed data. AJRH, JBC, CHA. SML and  
816 LF supervised the work, analysed and interpreted data and contributed to drafting and editing the  
817 manuscript.

818

819 The authors declare no competing interests.

820 Materials and correspondence

821

822 All expression constructs are available through Addgene.

823 Cryo-EM maps can be downloaded at EMDB 22106 and model coordinates at PDBID 6X9O.

824

825 All correspondence and requests for materials should be sent to RJH ([Rachel.Harding@utoronto.ca](mailto:Rachel.Harding@utoronto.ca)) or  
826 CHA ([Cheryl.Arrowsmith@uhnresearch.ca](mailto:Cheryl.Arrowsmith@uhnresearch.ca)).

827 References

828

829 1. Donaldson, J., Powell, S., Rickards, N., Holmans, P. & Jones, L. What is the Pathogenic CAG  
830 Expansion Length in Huntington's Disease? *J. Huntingt. Dis.* **10**, 175–202 (2021).

831 2. Saudou, F. & Humbert, S. The Biology of Huntingtin. *Neuron* **89**, 910–926 (2016).

832 3. Harding, R. J. & Tong, Y. Proteostasis in Huntington's disease: disease mechanisms and  
833 therapeutic opportunities. *Acta Pharmacol. Sin.* **39**, 754–769 (2018).

834 4. Koyuncu, S., Fatima, A., Gutierrez-Garcia, R. & Vilchez, D. Proteostasis of Huntingtin in Health  
835 and Disease. *Int. J. Mol. Sci.* **18**, (2017).

836 5. Gao, R. *et al.* Mutant huntingtin impairs PNKP and ATXN3, disrupting DNA repair and  
837 transcription. *eLife* **8**, e42988 (2019).

838 6. Poplawski, G. H. D. *et al.* Injured adult neurons regress to an embryonic transcriptional growth  
839 state. *Nature* **581**, 77–82 (2020).

840 7. Carmo, C., Naia, L., Lopes, C. & Rego, A. C. Mitochondrial Dysfunction in Huntington's Disease.  
841 *Adv. Exp. Med. Biol.* **1049**, 59–83 (2018).

842 8. Vitet, H., Brandt, V. & Saudou, F. Traffic signaling: new functions of huntingtin and axonal  
843 transport in neurological disease. *Curr. Opin. Neurobiol.* **63**, 122–130 (2020).

844 9. Smith-Dijak, A. I., Sepers, M. D. & Raymond, L. A. Alterations in synaptic function and plasticity  
845 in Huntington disease. *J. Neurochem.* **150**, 346–365 (2019).

846 10. McColgan, P. & Tabrizi, S. J. Huntington's disease: a clinical review. *Eur. J. Neurol.* **25**, 24–34  
847 (2018).

848 11. Maiuri, T. *et al.* Huntingtin is a scaffolding protein in the ATM oxidative DNA damage response  
849 complex. *Hum. Mol. Genet.* **26**, 395–406 (2017).

850 12. Rui, Y.-N. *et al.* Huntingtin functions as a scaffold for selective macroautophagy. *Nat. Cell Biol.*  
851 **17**, 262–275 (2015).

852 13. Shirasaki, D. I. *et al.* Network organization of the huntingtin proteomic interactome in  
853 mammalian brain. *Neuron* **75**, 41–57 (2012).

854 14. Wanker, E. E., Ast, A., Schindler, F., Trepte, P. & Schnoegl, S. The pathobiology of perturbed  
855 mutant huntingtin protein-protein interactions in Huntington's disease. *J. Neurochem.* **151**, 507–519  
856 (2019).

857 15. Seefelder, M. *et al.* The evolution of the huntingtin-associated protein 40 (HAP40) in  
858 conjunction with huntingtin. *BMC Evol. Biol.* **20**, 162 (2020).

859 16. Xu, S. *et al.* HAP40 is a conserved central regulator of Huntingtin and a specific modulator of  
860 mutant Huntingtin toxicity. *bioRxiv* 2020.05.27.119552 (2020) doi:10.1101/2020.05.27.119552.

861 17. Pal, A., Severin, F., Lommer, B., Shevchenko, A. & Zerial, M. Huntingtin-HAP40 complex is a  
862 novel Rab5 effector that regulates early endosome motility and is up-regulated in Huntington's  
863 disease. *J. Cell Biol.* **172**, 605–618 (2006).

864 18. Pal, A., Severin, F., Höpfner, S. & Zerial, M. Regulation of endosome dynamics by Rab5 and  
865 Huntingtin-HAP40 effector complex in physiological versus pathological conditions. *Methods Enzymol.*  
866 **438**, 239–257 (2008).

867 19. Peters, M. F. & Ross, C. A. Isolation of a 40-kDa Huntingtin-associated Protein. *J. Biol. Chem.*  
868 **276**, 3188–3194 (2001).

869 20. Guo, Q. *et al.* The cryo-electron microscopy structure of huntingtin. *Nature* (2018)  
870 doi:10.1038/nature25502.

871 21. Harding, R. J. *et al.* Design and characterization of mutant and wild-type huntingtin proteins  
872 produced from a toolkit of scalable eukaryotic expression systems. *J. Biol. Chem.* **jb**c.RA118.007204  
873 (2019) doi:10.1074/jbc.RA118.007204.

874 22. Jung, T. *et al.* The Polyglutamine Expansion at the N-Terminal of Huntingtin Protein Modulates  
875 the Dynamic Configuration and Phosphorylation of the C-Terminal HEAT Domain. *Structure* **28**, 1035–  
876 1050.e8 (2020).

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

879 24. Boatz, J. C. *et al.* Protofilament Structure and Supramolecular Polymorphism of Aggregated  
880 Mutant Huntingtin Exon 1. *J. Mol. Biol.* **432**, 4722–4744 (2020).

881 25. Falk, A. S. *et al.* Structural Model of the Proline-Rich Domain of Huntingtin Exon-1 Fibrils.  
882 *Biophys. J.* **119**, 2019–2028 (2020).

883 26. Matlahov, I. & van der Wel, P. C. Conformational studies of pathogenic expanded  
884 polyglutamine protein deposits from Huntington's disease. *Exp. Biol. Med. Maywood NJ* **244**, 1584–  
885 1595 (2019).

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

889 28. Schilling, B. *et al.* Huntington Phosphorylation Sites Mapped by Mass Spectrometry  
890 MODULATION OF CLEAVAGE AND TOXICITY. *J. Biol. Chem.* **281**, 23686–23697 (2006).

891 29. Tabrizi, S. J., Ghosh, R. & Leavitt, B. R. Huntington Lowering Strategies for Disease Modification  
892 in Huntington's Disease. *Neuron* **101**, 801–819 (2019).

893 30. Wheeler, V. C. *et al.* Length-Dependent Gametic CAG Repeat Instability in the Huntington's  
894 Disease Knock-in Mouse. *Hum. Mol. Genet.* **8**, 115–122 (1999).

895 31. Evers, M. M. *et al.* Making (anti-) sense out of huntingtin levels in Huntington disease. *Mol.*  
896 *Neurodegener.* **10**, 21 (2015).

897 32. Bogdanos, D. P., Gao, B. & Gershwin, M. E. Liver Immunology. in *Comprehensive Physiology*  
898 567–598 (American Cancer Society, 2013). doi:10.1002/cphy.c120011.

899 33. Huang, B. *et al.* PolyQ expansion does not alter the Huntingtin-HAP40 complex. *bioRxiv*  
900 2021.02.02.429316 (2021) doi:10.1101/2021.02.02.429316.

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

903 35. Graham, R. K. *et al.* Cleavage at the Caspase-6 Site Is Required for Neuronal Dysfunction and  
904 Degeneration Due to Mutant Huntingtin. *Cell* **125**, 1179–1191 (2006).

905 36. Liu, F., Lössl, P., Scheltema, R., Viner, R. & Heck, A. J. R. Optimized fragmentation schemes and  
906 data analysis strategies for proteome-wide cross-link identification. *Nat. Commun.* **8**, 15473 (2017).

907 37. Steigenberger, B., Pieters, R. J., Heck, A. J. R. & Scheltema, R. A. PhoX: An IMAC-Enrichable  
908 Cross-Linking Reagent. *ACS Cent. Sci.* **5**, 1514–1522 (2019).

909 38. Gu, X. *et al.* N17 Modifies mutant Huntingtin nuclear pathogenesis and severity of disease in  
910 HD BAC transgenic mice. *Neuron* **85**, 726–741 (2015).

911 39. Jayaraman, M. *et al.* Kinetically competing huntingtin aggregation pathways control amyloid  
912 polymorphism and properties. *Biochemistry* **51**, 2706–2716 (2012).

913 40. Maiuri, T., Woloshansky, T., Xia, J. & Truant, R. The huntingtin N17 domain is a multifunctional  
914 CRM1 and Ran-dependent nuclear and ciliary export signal. *Hum. Mol. Genet.* **22**, 1383–1394 (2013).

915 41. Caron, N. S., Desmond, C. R., Xia, J. & Truant, R. Polyglutamine domain flexibility mediates the  
916 proximity between flanking sequences in huntingtin. *Proc. Natl. Acad. Sci.* **110**, 14610–14615 (2013).

917 42. Nath, S., Munsie, L. N. & Truant, R. A huntingtin-mediated fast stress response halting  
918 endosomal trafficking is defective in Huntington's disease. *Hum. Mol. Genet.* **24**, 450–462 (2015).

919 43. Bravo-Arredondo, J. M. *et al.* The folding equilibrium of huntingtin exon 1 monomer depends  
920 on its polyglutamine tract. *J. Biol. Chem.* **293**, 19613–19623 (2018).

921 44. Newcombe, E. A. *et al.* Tadpole-like Conformations of Huntington Exon 1 Are Characterized by  
922 Conformational Heterogeneity that Persists regardless of Polyglutamine Length. *J. Mol. Biol.* **430**,  
923 1442–1458 (2018).

924 45. Warner, J. B. *et al.* Monomeric Huntingtin Exon 1 Has Similar Overall Structural Features for  
925 Wild-Type and Pathological Polyglutamine Lengths. *J. Am. Chem. Soc.* **139**, 14456–14469 (2017).

926 46. Gerbich, T. M. & Gladfelter, A. S. Moving beyond disease to function: Physiological roles for  
927 polyglutamine-rich sequences in cell decisions. *Curr. Opin. Cell Biol.* **69**, 120–126 (2021).

928 47. Dragatsis, I., Levine, M. S. & Zeitlin, S. Inactivation of Hdh in the brain and testis results in  
929 progressive neurodegeneration and sterility in mice. *Nat. Genet.* **26**, 300–306 (2000).

930 48. Klykov, O. *et al.* Efficient and robust proteome-wide approaches for cross-linking mass  
931 spectrometry. *Nat. Protoc.* **13**, 2964–2990 (2018).

932 49. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize implements and enhances circular  
933 visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).

934 50. Schweppe, D. K., Chavez, J. D. & Bruce, J. E. XLmap: an R package to visualize and score protein  
935 structure models based on sites of protein cross-linking. *Bioinformatics* **32**, 306–308 (2016).

936 51. Marty, M. T. *et al.* Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary  
937 Interactions to Polydisperse Ensembles. *Anal. Chem.* **87**, 4370–4376 (2015).

938 52. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion  
939 correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).

940 53. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron  
941 microscopy micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

942 54. Caesar, J. *et al.* SIMPLE 3.0. Stream single-particle cryo-EM analysis in real time. *J. Struct. Biol.*  
943 **X** **4**, 100040 (2020).

944 55. Brown, A. *et al.* Tools for macromolecular model building and refinement into electron cryo-  
945 microscopy reconstructions. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 136–153 (2015).

946 56. Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta*  
947 *Crystallogr. Sect. Struct. Biol.* **74**, 531–544 (2018).

948 57. Prisant, M. G., Williams, C. J., Chen, V. B., Richardson, J. S. & Richardson, D. C. New tools in  
949 MolProbity validation: CaBLAM for CryoEM backbone, UnDowser to rethink “waters,” and NGL Viewer  
950 to recapture online 3D graphics. *Protein Sci.* **29**, 315–329 (2020).

951 58. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and  
952 developers. *Protein Sci.* **30**, 70–82 (2021).

953 59. Franke, D. *et al.* ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from  
954 macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).

955 60. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSTOL – a Program to Evaluate X-ray Solution  
956 Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28**, 768–773  
957 (1995).

958 61. Fischer, H. *et al.* Determination of the molecular weight of proteins in solution from a single  
959 small-angle X-ray scattering measurement on a relative scale. *J. Appl. Crystallogr.* **43**, 101–109 (2010).

960 62. Piiadov, V., Ares de Araújo, E., Oliveira Neto, M., Craievich, A. F. & Polikarpov, I. SAXSMoW 2.0:  
961 Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data  
962 measured on a relative scale. *Protein Sci. Publ. Protein Soc.* **28**, 454–463 (2019).

963 63. Rambo, R. P. & Tainer, J. A. Accurate assessment of mass, models and resolution by small-  
964 angle scattering. *Nature* **496**, 477–481 (2013).

965 64. Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid computation and  
966 fitting of SAXS profiles. *Nucleic Acids Res.* **38**, W540–W544 (2010).

967 65. Franke, D. & Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in  
968 small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346 (2009).

969 66. Grant, T. D. Ab initio electron density determination directly from solution scattering data.  
970 *Nat. Methods* **15**, 191–193 (2018).

971 67. Berlin, K. *et al.* Recovering a Representative Conformational Ensemble from Underdetermined  
972 Macromolecular Structural Data. *J. Am. Chem. Soc.* **135**, 16595–16609 (2013).

973 68. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).

974 69. Dolinsky, T. J. *et al.* PDB2PQR: expanding and upgrading automated preparation of  
975 biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35**, W522–W525 (2007).

976 70. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids*  
977 *Res.* **47**, W636–W641 (2019).

978 71. Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize  
979 evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).

980 72. Yazdani, S. *et al.* The SARS-CoV-2 replication-transcription complex is a priority target for  
981 broad-spectrum pan-coronavirus drugs. *bioRxiv* 2021.03.23.436637 (2021)  
982 doi:10.1101/2021.03.23.436637.

983 73. An, J., Totrov, M. & Abagyan, R. Pocketome via comprehensive identification and classification  
984 of ligand binding envelopes. *Mol. Cell. Proteomics MCP* **4**, 752–761 (2005).

985