

## Review

---

# Huntingtin and Its Partner Huntingtin-Associated Protein 40: Structural and Functional Considerations in Health and Disease

Manuel Seefelder<sup>a,1</sup>, Fabrice A.C. Klein<sup>b,1</sup>, Bernhard Landwehrmeyer<sup>b</sup>, Rubén Fernández-Busnadiego<sup>c,d</sup> and Stefan Kochanek<sup>a,\*</sup>

<sup>a</sup>*Department of Gene Therapy, Ulm University, Ulm, Germany*

<sup>b</sup>*Department of Neurology, Ulm University, Ulm, Germany*

<sup>c</sup>*Institute of Neuropathology, University Medical Center Göttingen, Göttingen, Germany*

<sup>d</sup>*Cluster of Excellence “Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells” (MBExC), University of Göttingen, Göttingen, Germany*

Pre-press 16 July 2022

**Abstract.** Since the discovery of the mutation causing Huntington's disease (HD) in 1993, it has been debated whether an expanded polyglutamine (polyQ) stretch affects the properties of the huntingtin (HTT) protein and thus contributes to the pathological mechanisms responsible for HD. Here we review the current knowledge about the structure of HTT, alone (apo-HTT) or in a complex with Huntingtin-Associated Protein 40 (HAP40), the influence of polyQ-length variation on apo-HTT and the HTT-HAP40 complex, and the biology of HAP40. Phylogenetic analyses suggest that HAP40 performs essential functions. Highlighting the relevance of its interaction with HTT, HAP40 is one of the most abundant partners copurifying with HTT and is rapidly degraded, when HTT levels are reduced. As the levels of both proteins decrease during disease progression, HAP40 could also be a biomarker for HD. Whether declining HAP40 levels contribute to disease etiology is an open question. Structural studies have shown that the conformation of apo-HTT is less constrained but resembles that adopted in the HTT-HAP40 complex, which is exceptionally stable because of extensive interactions between HAP40 and the three domains of HTT. The complex—and to some extent apo-HTT—resists fragmentation after limited proteolysis. Unresolved regions of apo-HTT, constituting about 25% of the protein, are the main sites of post-translational modifications and likely have major regulatory functions. PolyQ elongation does not substantially alter the structure of HTT, alone or when associated with HAP40. Particularly, polyQ above the disease length threshold does not induce drastic conformational changes in full-length HTT. Therefore, models of HD pathogenesis stating that polyQ expansion drastically alters HTT properties should be reconsidered.

**Keywords:** Huntingtin, HTT-associated protein 40, Huntington's disease, protein conformation, polyglutamine, 3D structure

---

<sup>1</sup>These authors contributed equally to this work.

\*Correspondence to: Prof. Dr. Stefan Kochanek, Department of Gene Therapy, Ulm University, 89081 Ulm, Germany. Tel.: +49 731 500 46 103; E-mail: stefan.kochanek@uni-ulm.de.

Nearly 30 years ago, Huntington's disease (HD) was found to be caused by the amplification of a polymorphic CAG trinucleotide repeat (CAG<sub>n</sub>) encoding a polyglutamine (polyQ) stretch in a gene coding for

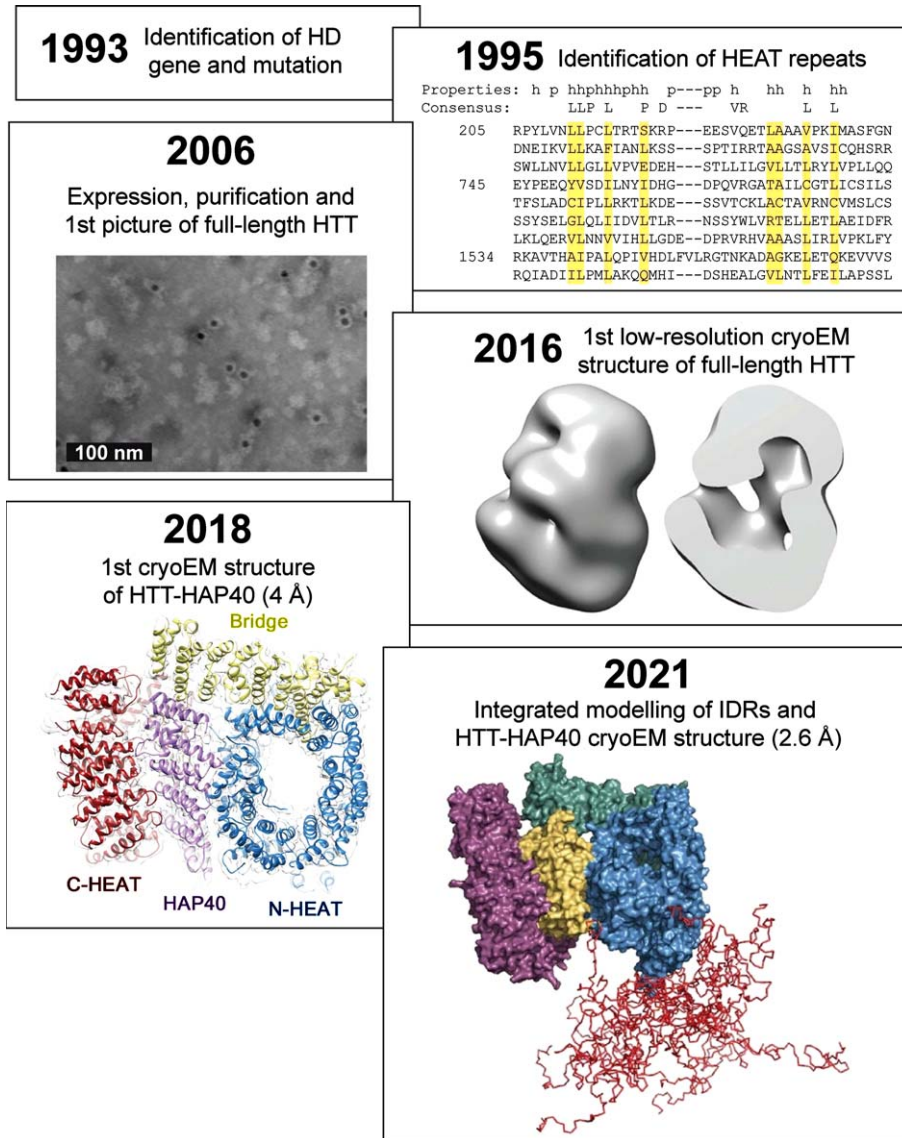


Fig. 1. Timeline of structural studies of HTT. 1993 [1, 2]. 1995, alignment of primary sequence segments of human HTT, adapted from Andrade et al. [25]. Properties: *h* = hydrophobic and *p* = non-hydrophobic. The most conserved hydrophobic positions are framed. Consensus: the most frequent amino acid found in this position is indicated. 2006, adapted from Li et al. [26]. 2016, adapted from Vijayvargia et al. [27]. 2018, adapted from Guo et al. [22]. 2021, adapted from Harding et al. [24].

huntingtin (HTT), a protein of unknown function at the time of discovery [1, 2] (Fig. 1). Besides other mechanisms that may contribute to pathogenicity (e.g., expanded RNA toxicity [3], Repeat Associated Non-AUG translation [4], somatic instability [5, 6], aberrant splicing [7]), the CAG<sub>n</sub> expansion above a pathological length threshold is proposed to provide a gain of toxic function to the mutant huntingtin protein (mHTT) carrying the expanded polyQ [2, 8]. The nature of the toxic role played by expanded polyQ in

the disease's etiology remains a central question in HD.

The amino-terminal (Nter) extremity of HTT (NterHTT) encompasses the polyQ stretch and is the focus of intense scrutiny for several reasons. First, NterHTT fragments accumulate in the tissues of HD patients, animal models and model cell systems, forming inclusion bodies (IBs) that are the pathognomonic hallmark of the disease [9]. Inclusion bodies mainly contain a HTT protein fragment corre-

sponding to the first exon of the gene (HTT<sub>ex1</sub>) and other NterHTT protein cleavage products [10–12], as well as other cellular proteins that are trapped or recruited by these entities [13, 14]. Second, many studies showed, both *in vitro* and *in vivo*, that aggregates generated from NterHTT fragments display amyloid-like and prion-like properties [15–18]. The intrinsic toxicity and propagation mechanisms of amyloid aggregates and other intermediate species generated during the amyloid cascade are intensely studied in amyloid and prion diseases [19]. Third, polyQ expansion in NterHTT was proposed to alter the function of HTT, for instance, by modifying its affinity for interactors [20, 21]. Hence, understanding the influence of expanded polyQ on the structural properties of full-length HTT is of eminent importance.

In this review, we will summarize our current knowledge about the structural properties of full-length wild-type HTT (wtHTT) and mutant HTT (mHTT). We will also highlight why studying the properties and function of the Huntingtin-Associated Protein 40 (HAP40)—whose near-atomic resolution structure was recently solved in a complex with full-length wild-type and mutant HTT [22–24]—is of great interest for the HD research field.

## HTT SEQUENCE GAVE EARLY CLUES AS TO ITS STRUCTURE AND FUNCTION

The first hints into the structure of HTT came from the discovery that, besides extended low complexity regions predicted to be intrinsically disordered, the HTT primary sequence consists mostly of short repeated sequences with strong helical propensity called Huntingtin, elongation factor 3, protein phosphatase 2A, and the yeast kinase TOR1 (HEAT) repeats (Fig. 1). The hydrophobicity of these HEAT repeat sequences suggested a structural role for packing or interaction with partners [25]. HEAT repeats are most likely phylogenetically related to another helical repeat, the Armadillo motif (ARM) [28], both of which belong to a larger family of tandemly repeated helical motifs found in 0.4% of the eukaryotic proteome [29]. HEAT and ARM motifs also share a close structural homology to the so-called epsin N-terminal homology (ENTH) motif, present in a range of unrelated multidomain proteins and primarily serving as protein-protein interaction modules [30]. Antiparallel helical repeats, such as the HEAT and ARM motifs, were probably selected early and

throughout evolution because of their propensity to form solenoid-like super-helical protein structures functioning as scaffolds for protein-protein interactions. This possibility is in strong agreement with the plethora of reported HTT interactors and its proposed function as a hub for the assembly of multi-protein complexes [8, 20, 31, 32].

## BIOCHEMICAL AND LOW-RESOLUTION STRUCTURAL STUDIES OF APO-HTT

The reference HTT protein (NP\_002102.4), used for amino acid numbering throughout the text, carries a polyQ of 23Q, contains 3144 amino acids and is one of the largest proteins in humans (~348 kDa). Although the large size of HTT's cDNA makes developing molecular biology tools and cellular models to express the full-length protein challenging [33, 34], various expression systems were implemented in insect and mammalian cells to produce and purify wild-type and mHTT [26, 34–37]. Valuable resources (HTT and HAP40 expression plasmids) were recently made accessible to the HD community [34]. When purified without an interaction partner, most of the apo-HTT protein is found in the insoluble cell debris or self-associated in high molecular weight soluble aggregates, with only a smaller fraction being isolated as monomer or dimer [26, 34, 35]. Monomeric purified HTT further oligomerizes [34, 35, 38], a tendency that is exacerbated by polyQ expansion [35]. Accordingly, dynamic light-scattering experiments performed on purified monomeric HTT display a polydispersity index suggesting the presence of higher-order HTT species [26]. This tendency to self-associate could be a consequence of the *in vitro* purification of the large and highly concentrated apo-HTT protein isolated without its natural partners. However, a potential physiological role of apo-HTT multimers cannot be excluded.

The propensity of full-length HTT to self-associate after purification can be misleading for structural techniques requiring relatively high protein concentrations, leading to increased aggregation kinetics and, consequently, misinterpretations due to the presence of multimeric species. One of these methods is, for instance, small-angle X-ray scattering (SAXS), which may overestimate the dimensions of HTT in solution in the presence of non-monomeric HTT species, as reported by Harding et al. [34]. In another study, the influence of aggregated HTT species—although observed after purification—was

not assessed [38], the interpretation of SAXS results are thereby subject to caution. Nonetheless, other methods, such as circular dichroism (CD), allowed study of the structural features of HTT at lower concentrations, where limited aggregation would lead to a reduced contribution of multimeric species to the CD spectrum. In agreement with the large proportion of HEAT repeat motifs and disordered regions predicted based on its primary sequence, the CD spectra of HTT indicate a mostly alpha-helical secondary structure [26, 27, 35].

The first glimpse of isolated HTT particles came from pioneering work by Li et al. (Fig. 1), who purified human HTT recombinantly expressed in insect cells and observed particles of dimensions compatible with HTT monomers using negative-staining and transmission electron microscopy (TEM) [26]. Later work using TEM revealed a striking conformational variability of isolated HTT, as around 100 structurally distinguishable structural classes could be identified [37]. Using a mild cross-linking strategy to constrain and prevent HTT multimerization facilitated obtaining its first low-resolution (32Å) 3D-model out of negative-stained TEM micrographs (Fig. 1) [27]. Integrating immunolabelling and cross-linking mass-spectrometry (XL-MS) data, the authors deduced that NterHTT is solvent-exposed and proposed that HTT does not form a linear solenoid but comprises two large domains, which are separated by a hinge and are folding back towards each other. While studying mildly cross-linked HTT particles by cryo-EM, the same overall organization was observed [38]. More recently, native (not crossed-linked) monomeric HTT particles displayed a similar, although less collapsed conformation in cryo-EM [24]. The latter 3D structural model of apo-HTT, obtained at 12Å resolution, displays several features remarkably resembling those observed in near-atomic resolution structures of HTT studied in complex with HAP40 [22–24], which are discussed below.

## NEAR-ATOMIC RESOLUTION STRUCTURAL STUDIES OF HTT-HAP40

HAP40, a 371 amino acids (~39.1 kDa) protein, was initially discovered for its ability to immunoprecipitate with full-length HTT [39]. Although a plethora of HTT interactors were reported over the past decades [8, 20, 21], HAP40 is the only partner that was extensively studied in a complex with HTT using biochemical, biophysical, and near-atomic res-

olution structural techniques. As discussed below, the stable complex formed with HTT places HAP40 in the limelight of HD research, as it may facilitate understanding HTT's function, as well as be relevant to HD pathomechanisms or as a biomarker.

Whereas apo-HTT is polydisperse and oligomerizes [26, 34, 35, 38] and HAP40 could not be over-expressed alone in quantities permitting its biochemical and biophysical characterization [22, 24], the two proteins can be co-purified as a monodisperse and stable HTT-HAP40 (1:1) complex, the structure of which was recently solved at near-atomic resolution by cryo-EM [22–24] (Figs. 1 and 2 and Supplementary Movie 1).

### Overall organization

In agreement with computational predictions, all resolved secondary structure elements of HTT and HAP40 are helical, forming mostly HEAT and other tandem repeats in HTT and tetratricopeptide-like (TPR) helical tandem repeats in HAP40 (Fig. 2A). The regions predicted to be intrinsically disordered (IDR), corresponding to ~25% of the HTT-HAP40 primary sequence, were not resolved. These included, for instance, an NterHTT region (residues 1–96) that encompasses the exon 1 fragment and a large IDR (residues 409–666) in HTT, as well as a central mammalian-specific IDR of HAP40 (residues 217–257) [22, 40] (Fig. 2A). HAP40 comprises a single domain and the resolved portions of HTT form 3 domains: the amino-terminal N-HEAT (residues 97–1,690), the central Bridge (residues 1,691–2,097) and the carboxy-terminal C-HEAT (residues 2,098–3,104) domains [22].

All near-atomic resolution 3D-models of HTT-HAP40 published to date display the same overall structural organization [22–24]. HTT-HAP40 is a globular complex measuring approximately  $120 \times 80 \times 100$  Å, in which the three domains of HTT wrap around and form tight interactions with HAP40. The HAP40-binding occurs within a space observed in apo-HTT, between its subdomains, thereby constraining them into a more rigid conformation, which explains why higher resolutions can be reached in cryo-EM analysis of the HTT-HAP40 complex [24]. Some differences in the HTT-HAP40 interfaces and the relative positioning of the C-HEAT domain of HTT were observed, depending on whether Cter-HAP40 is unaltered, or fused to a peptide-tag. This shows that some variation in the

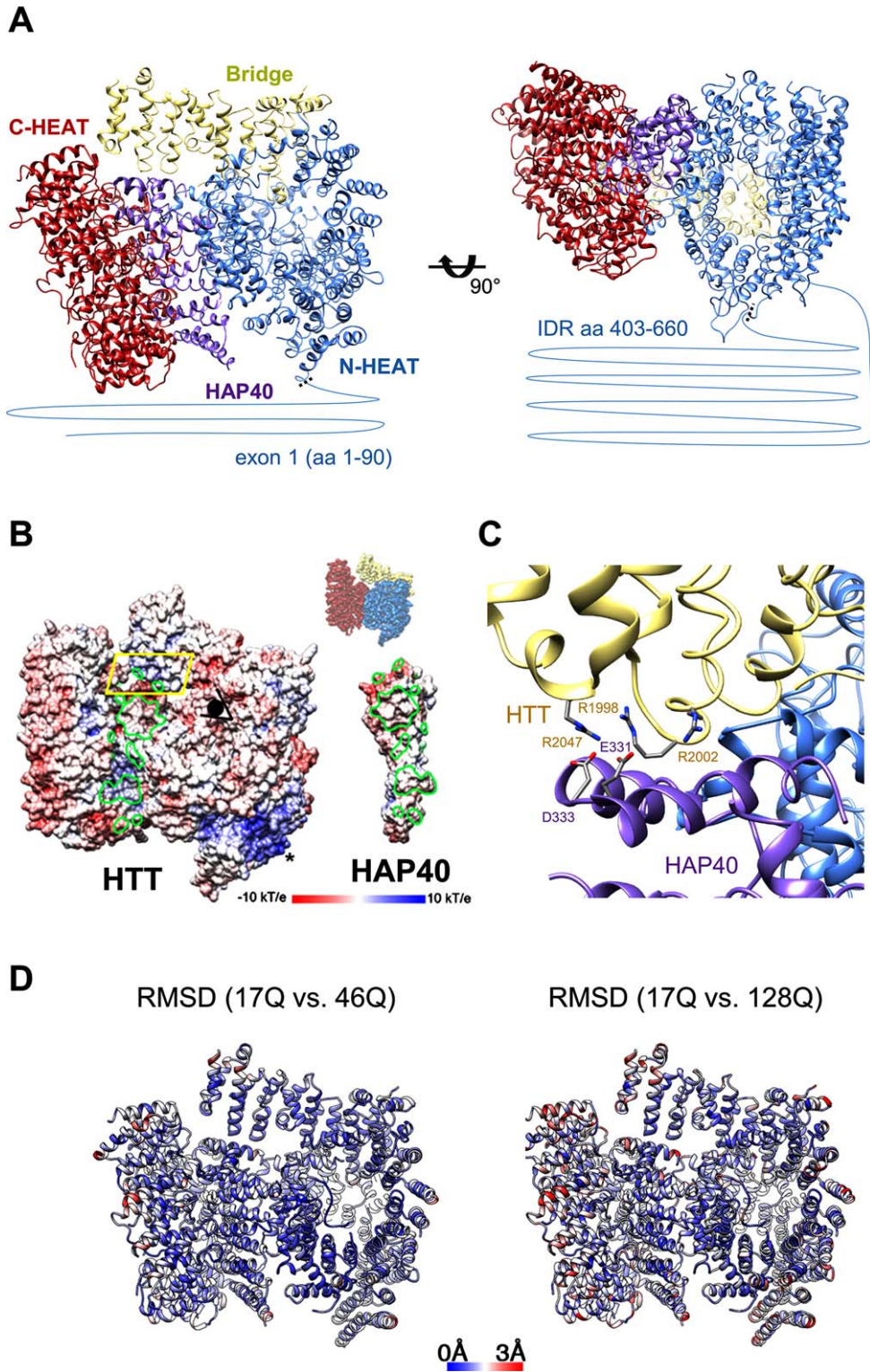


Fig. 2. (Continued)

Fig. 2. Cryo-EM structure of the HTT-HAP40 complex. A) Atomic model of Q17 HTT-HAP40 (PDB 6EZ8; [22]) in ribbon representation. HTT domains are shown as follows: N-HEAT (residues 97-1,690) in blue, Bridge (residues 1,691-2,097) in yellow and C-HEAT (residues 2,098-3,104) in maroon. HAP40 (residues 42-364) is shown in purple. The unstructured regions comprising residues 1-96 of exon 1 fragment and residues 409-666 are shown approximately to scale as lines. B) Electrostatic surface potentials of Q17 HTT and HAP40 (PDB 6EZ8; [22]) in “open book” view. Green contours mark the interaction areas between HTT and HAP40, dominated by hydrophobic interactions. A conserved charge clamp between the HTT Bridge domain and the HAP40 C-terminus is outlined in yellow. An eye indicates the perspective of the charge clamp visualized in C. A positively charged surface on the HTT N-term domain is marked by an asterisk. C) Charge clamp between HTT and HAP40 (PDB 6EZ8; [22]) in ribbon representation. Residues participating in the electrostatic interaction are displayed and shown as sticks. D) Atomic models of Q46 HTT-HAP40 (PDB 7DXJ; [23]; left) and Q128 HTT-HAP40 (PDB 7DXK; [22]; right) in ribbon representation. The colour code shows the root-mean-square deviation with Q17 HTT-HAP40 (PDB 6EZ8; [22]), displaying minimal ( $<3 \text{ \AA}$ ) differences between the models. Numbering in panels A and C correspond to that of reference Q23-HTT protein NP\_002102.4.

relative positioning of the proteins and their interface can be tolerated by the HTT-HAP40 complex without disrupting its organization [24].

### Remarkable structural features

The interface between HTT and HAP40 comprises mostly hydrophobic regions (Fig. 2B) and a conserved charge clamp between the Bridge domain of HTT and the carboxy-terminal (Cter) part of HAP40 [22, 40]. The core residues of this charge clamp, whose configurations are consistent in all published atomic models of the HTT-HAP40 complex to date, are shown in Fig. 2C. As shown by co-immunoprecipitation experiments with wild-type HTT and mutated HAP40 (Glu316Lys-Glu317Lys-Glu331Lys-Asp333Lys), the interaction between HAP40 and HTT is drastically reduced if this conserved charge clamp is mutated [40]. The likely coevolution of HTT and HAP40 [40] allows for a non-strict conservation of their hydrophobic interface (i.e., mutations on one protein can be compensated by mutations on the other protein, while preserving the interface properties). Hence, the residues belonging to the HTT-HAP40 interface are only moderately conserved [24, 40]. Interestingly, while HTT residues exposed on the side opposite to the HAP40 interaction area are poorly conserved, some areas on the HTT side exposed towards HAP40 display a remarkably high conservation level (even higher than that of the HTT-HAP40 complex interface), including a  $40 \times 5\text{--}10 \text{ \AA}$  positively charged surface of unknown function on the N-HEAT domain [22, 24] (Fig. 2B).

Another remarkable feature of the N-HEAT is that it forms a hollow domain, with a main cavity running through it, connected by another lateral cavity. Although the function of these N-HEAT cavities is unknown, their dimensions suggest that relatively large macromolecules could bind there, and Harding et al. speculate that they might accommodate a

dsDNA helix [34]. Other smaller cavities and pockets can also be observed, and a potentially druggable pocket was identified at the N-HEAT/HAP40 interface, which may become of use to target the complex for degradation or to develop specific positron-emission tomography (PET) tracers [24].

The large, disordered regions of HTT-HAP40, which are exposed to the surface of the complex and therefore accessible to cellular proteins and enzymes, are the targets of numerous post-translational modifications (PTMs), including proteolysis [22, 24, 34] and references therein. These PTMs are expected to influence the local structure and regulate activity of the proteins, as shown for instance for the 17 first amino acids region of HTT (N17HTT) [41–43]. Importantly, although plenty of qualitative information has been obtained regarding the position of PTMs in the HTT sequence [22, 24, 34] and references therein, little quantitative information is yet available. Recently published native MS experiments [24] show a low amount of phosphorylation per molecule, based on insect-cells produced HTT. It would be interesting to perform such quantitative analysis to investigate PTMs on endogenous mammalian cell-derived HTT. In addition to the fact that many PTMs are found on IDRs, a low number of PTMs potentially present per protein molecule could thus also contribute to explain why no PTMs are seen to date in cryoEM models.

### The contribution of disordered protein regions to the HTT-HAP40 structure

Two remarkable papers by Harding et al. report the in-depth structural study of apo-HTT and the HTT-HAP40 complex [24, 34]. The part of the complex resolved by cryo-EM provides no information as to the space occupied by disordered regions exposed to the solvent. Accordingly, the observed radius of gyration ( $R_g$ ) of the complex, deduced from SAXS analysis, is much larger than that expected

when considering solely the dimensions of the cryo-EM resolved complex [34]. The authors performed molecular dynamics (MD) simulations to calculate an ensemble of conformations of disordered regions which, when considered together with the cryo-EM model, provide a theoretical scattering curve that fits the experimental SAXS data. These simulations indicate, for instance, that the HTTex1 region and the large IDR extruding from N-HEAT are highly dynamic and flexible, exploring an extended space away from the rigid core of the complex, which makes them easily accessible for post-translational modification (Fig. 1). Accordingly, Guo et al. observed no density for HTTex1 by cryo-EM, even at very low thresholds, which is a sign of high conformational flexibility [22]. Furthermore, XL-MS results showing that single residues from these regions can contact distinct solvent-exposed parts of the complex, the majority of which are other disordered segments, also supported the high flexibility of the disordered sequences [23, 24].

#### *Stability of apo-HTT and the HTT-HAP40 complex*

##### *In vitro and ex vivo stability*

Purified apo-HTT was shown not to be fragmented but rather to remain associated after limited proteolysis, as seen by native-PAGE and size exclusion chromatography [26]. These observations, together with other results obtained from yeast two-hybrid [29] and pull-down [44] experiments, suggest that intra-domain and inter-domain contacts within apo-HTT contribute to stabilizing the protein architecture after proteolytic nicking. These contacts could contribute to apo-HTT adopting a bent conformation [27, 34] resembling that seen in the HTT-HAP40 complex [22, 24]. In the same line, HTT also displayed a good resistance to proteolysis when postmortem brain samples were studied by native electrophoresis [45], although in this *ex vivo* context HTT may have been at least partially in a complex with HAP40. The great resilience of the HTT-HAP40 complex to limited proteolysis was also clearly demonstrated, as the purified core complex remained associated even after multiple cleavage events within HTT [24].

The cryo-EM structure of the HTT-HAP40 complex allows to rationalize why HTT and HTT-HAP40 remain associated after limited proteolysis events, which primarily affects easily accessible and disordered regions. For instance, the major IDR of HTT protrudes at the surface of the complex between

N-HEAT repeats 6 and 7 and constitutes a known hot-spot of HTT proteolysis [8, 46] (Fig. 2A). However, the tight HEAT repeats packing throughout the N-HEAT domain suggested that nicking within this disordered region would not destabilize the domain [22], as later confirmed experimentally [24].

Moreover, the overall organization of the complex, in which HAP40 is tightly packed between and has large contact surfaces with HTT subdomains, also suggested that the core complex should be stable [22]. Differential scanning fluorimetry experiments displaying the great stability of the complex in a wide range of pH and salt concentrations confirmed this [23, 24]. But the most striking illustration of its exceptional stability comes from native mass-spectrometry experiments designed to dissociate non-covalent complexes under harsh conditions, which revealed that the core HTT-HAP40 complex remained associated, while covalent bonds were broken within HTT, resulting in the loss of N- and C-terminal fragments [24].

##### *In vivo stability*

The *in vivo* stability of a given protein or protein complex depends on numerous factors, and cells can take active steps to degrade them efficiently, independent of the *in vitro* resilience of the protein. However, taken together, the observations summarized in the previous section show that the co-evolution of HTT and HAP40 has selected an exceptionally stable complex, suggesting that this stability may be needed for HTT-HAP40 function. This aligns with the findings that HAP40 is degraded rapidly when HTT is knocked down, whereas it is stabilized in the presence of HTT [24, 47].

We could thus speculate: what would be the consequences of limited proteolytic events affecting HTT-HAP40, should the tendency of the complex to remain associated *in vivo* be at least partially comparable to that observed *in vitro*? In this case, the observation of proteolytic HTT fragments, using denaturing techniques, would not necessarily imply that the HTT-HAP40 protein complex is dysfunctional. The removal of disordered regions may even contribute to regulating the activity of the HTT-HAP40 complex through the loss of specific functional features. Removing part of the large N-HEAT and Nter-HAP40 IDRs, for instance, would cause the loss of numerous PTM sites on HTT [22] and references therein) and a putative nuclear-localization signal on HAP40 [39]. Similarly, cleaving off the NterHTT IDR would not destabi-

lize the core complex but would result in the loss of motifs responsible for self-assembly or interaction with lipidic membranes (the N17HTT), for self-assembly (the polyQ stretch) and for interaction with other proteins (the proline-rich domain) [8, 20, 46].

Whether such IDRs, once released from the core complex, could perform physiological functions on their own is an open question. However, in the context of mHTT-HAP40, the proteolytic release of NterHTT fragments carrying the mutant polyQ is relevant to HD pathological mechanism [9, 10, 12, 48].

### **POLYQ-LENGTH INFLUENCE ON APO-HTT AND HTT-HAP40 STRUCTURES**

Besides the fact that multimeric HTT species can affect the outcome of structural and biophysical methods (as mentioned before), an additional hurdle awaits the experimenter who aims to evaluate the influence of polyQ-length variations on HTT structure: the relatively small size of the NterHTT region, which is only about 3 to 6% of the HTT sequence (depending on whether it carries, for instance, 20 or 120 glutamines, respectively). Therefore, the signal retrieved with many techniques (e.g., CD, SAXS, thermal stability, etc.) will be dominated by the other 94 to 97% of the protein, respectively. Measuring subtle effects caused by polyQ-length variations is thus challenging. Despite this inherent difficulty, several studies aimed to evaluate these effects on the structure of HTT. We discuss below the major findings emanating from biochemical, biophysical and MS analyzes applied to apo-HTT or the HTT-HAP40 complex, low-resolution structural studies of apo-HTT, and near-atomic 3D-models of the HTT-HAP40 complex.

Two studies have compared low-resolution EM models of apo-HTT to assess the effect of polyQ length variations on the protein structure. First, it was shown that apo-HTT bearing Q23, Q46, or Q78 all adopt highly similar shapes, and the authors could not conclude whether minor differences observed upon manual superimposition of the maps resulted from polyQ lengthening, or from technical variations and sample heterogeneity [27]. Second, apo-HTT bearing Q23 and Q78 were compared, which led to the conclusion that the polyQ length may strongly influence the relative positioning of the N-HEAT, Bridge, and C-HEAT domains [38]. Although interesting, this finding is surprising, as it does not align with previous results obtained by the same authors [27]. Moreover, hydrogen-deuterium exchange mass-

spectrometry (HDX-MS) revealed no difference [38], whereas some would be expected if the conformation of mHTT was constrained by polyQ-mediated contacts. Also, SAXS analysis displayed a variation of radius of gyration and maximal dimension of only roughly 3% [38], a marginal variation in line with that observed when studying HTT-HAP40 bearing various polyQ lengths, which is expected for samples sharing highly similar structural cores [24]. Furthermore, an important resolution difference existed between the maps of Q23-HTT and Q78-HTT (around 10 Å), and the low-resolution of the Q78-HTT map per se did not allow to visualize the bridge domain properly [38]. For all these reasons, it would be interesting to re-assess the influence of polyQ length variation on the relative positioning of apo-HTT subdomains using more homogenous and potentially higher resolution data.

The CD patterns of apo-HTT bearing various polyQ lengths displayed identical minima (222 and 208 nm) and maximum (195 nm) and were globally similar [27, 35]. This suggested that the polyQ length does not affect the overall secondary structure of HTT, a hypothesis which was later supported by near-atomic resolution models showing that polyQ variation from Q17 to Q128 does not influence the architecture of the resolved, core HTT-HAP40 complex [23] (Fig. 2D). Similarly, polyQ expansion does not change the thermal stability of apo-HTT [27, 34] and the HTT-HAP40 complex [23, 34], showing that polyQ length does not modify the stability of the folded proteins. In the same line, polyQ lengthening does not affect the apparent molecular weight of the HTT-HAP40 complex in solution, as estimated by SAXS or by size exclusion chromatography (SEC) combined with multi-angle light-scattering [24].

The IDRs present on the surface of apo-HTT and HTT-HAP40 represent roughly 25% of the sequence of the complex and are not resolved in the near-atomic resolution cryo-EM 3D-models [22, 24] (Fig. 2A). To assess whether the polyQ length may affect the behavior of these flexible regions, their XL-MS intramolecular cross-linking pattern was analyzed. In a first study comparing Q23-, Q46-, and Q78-apo-HTT proteins, a large majority of intra-HTT cross-links were common to all three or at least two proteins, while a minority of cross-links were unique to either protein [27]. Comparable cross-linking pattern variations were observed when HTT-HAP40 complexes carrying various polyQ lengths were analyzed by XL-MS [23]. Such pattern variations are, however, common even between replicates [23, 24], and addi-



tional cross-linking events scanning the accessible lysine residues (monolinks) on the surface of the complex did not reveal important differences [23]. Hence, polyQ lengthening does not significantly alter the behavior of the disordered regions of HTT or the HTT-HAP40 complex.

In another study comparing the XL-MS patterns of HTT-HAP40 complexes carrying various polyQ lengths, the authors reported that subtle structural differences may nonetheless exist, after observing that the N17 extremity of Q54-HTT displayed two cross-links with the C-HEAT domain, which were not observed in the Q23-HTT case [24]. The complex of HAP40 and HTT with deleted exon 1 fragment (HTT $\Delta$ ex1) displayed a small shift in the SEC elution profile, when compared with the Q23- and Q54-HTT-HAP40 complexes. Furthermore, the authors performed MD simulations, suggesting that the exon 1 fragment of the Q54-HTT-HAP40 complex could explore a larger conformational space than that of Q23-HTT-HAP40, which may constrain the neighboring large N-HEAT-IDR in a more finite space. Conversely, the N-HEAT-IDR appears less constrained and more conformationally flexible in the HTT $\Delta$ ex1-HAP40 context [24]. Hence, polyQ expansion could provoke subtle structural changes by affecting simultaneously the conformational space of exon 1 fragment and that of other IDRs on the HTT-HAP40 surface, but it does not affect the core structure of the HTT-HAP40 complex [23, 24].

#### **CAVEATS TO THE ANALYSIS OF APO-HTT SAMPLES: FROM *IN VITRO* EVIDENCE TO *IN VIVO* INFERENCES**

Among other aspects, the work done with the HTT-HAP40 complex allowed, by comparison, to highlight the difficulty of studying apo-HTT using biophysical and structural methods. As explained before, the HTT-HAP40 complex is rigid, monodisperse, and remains monomeric and stable for extended periods, allowing its reliable characterization [22–24]. In contrast, apo-HTT is flexible, polydisperse and continues aggregating in the test tube, even after purification of apparently monomeric species [26, 34, 35, 38]. For this reason, cross-linking strategies were used to study apo-HTT by cryoEM [27, 38], a process which may constrain the protein in potentially collapsed and unphysiological conformations. This is supported by the fact that the unique 3D model of apo-HTT, obtained with-

out cross-linking, displays a less collapsed structure, whose 3D volume fits remarkably well to the structure of HTT observed in the near-atomic 3D model of HTT-HAP40 [24]. Moreover, despite its low resolution (12 Å), this model allows to visualize the N-HEAT (including its central cavity) and C-HEAT domains [24]. In the same line, given the high propensity of apo-HTT to aggregate, it is difficult to imagine that XL-MS datasets of apo-HTT samples would not be contaminated, to some extent, with intermolecular cross-links. Taken together, these observations should warn the reader that—given the daunting task of trying to tame and make sense of a giant and user-unfriendly protein such as apo-HTT—the conclusions of structural and biophysical studies of apo-HTT should be re-assessed *a posteriori* by comparison with findings obtained with the “gentler giant” HTT-HAP40.

Another caveat, suggested by our understanding of apo-HTT behavior *in vitro*, concerns the use of over-expressed HTT to identify potential physiological partners. We showed that, upon HTT over-expression, the HAP40 protein level also increases, suggesting that part of the over-expressed HTT is “buffered” in a complex with HAP40 [47]. This means, on the one hand, that partners copurifying with HTT may in fact be interactors of the HTT-HAP40 complex. On the other hand, apo-HTT that may well exist in such over-expression systems, would likely display similar properties as seen *in vitro*, i.e., multimerize and aggregate, and interact preferentially with hydrophobic proteins or be targeted by chaperones, raising the question of the relevance of such interactions regarding HTT function. Moreover, numerous partners of HTT were found using only amino-terminal fragments of HTT of various sizes as baits, in yeast two-hybrid or copurification assays [20]. Analysis of these fragments, knowing the structural organization of HTT, strongly suggests that most of them would probably be unstable and aggregate. Great caution should thus be taken as to the relevance of numerous putative partners of HTT.

In addition, we should perhaps also ask, what could be the function of apo-HTT *in vivo*, or even if apo-HTT can exist in a physiological cellular context. Is it even possible that such a large protein, whose domains are exposing large hydrophobic surfaces, remains unbound? Since HTT and HAP40 share large hydrophobic interphases [22], we speculated that a specific function of HAP40 could be to shield hydrophobic patches of HTT, thereby preventing

aggregation [47]. The stable HTT-HAP40 complex could thus be the molecular entity enabling HTT to fulfill its proposed function as hub for multi-protein complexes assembly [8, 20, 31, 32]. Alternatively, HAP40 could also act as a stabilizing reservoir for HTT—or *vice versa*—that once released could fulfill other functions with alternate partners.

We believe that many of these questions could open-up exciting research avenues that could provide answers of great importance to our understanding of HD.

### BIOLOGY OF HAP40 AND ITS RELEVANCE IN HUNTINGTON DISEASE

In humans, HAP40 is encoded by three sequence-identical, X-chromosomally located single-exon genes, the factor VIII intronic transcript genes A1 (F8A1), A2 (F8A2), and A3 (F8A3) [48, 49]. The F8A1 paralog is located inside intron 22 of the blood coagulation factor VIII gene, whereas the other two paralogs are located outside of the F8 gene [49, 50]. Recombination between the intragenic and the two extragenic paralogs accounts for approximately 50% of all hemophilia A cases through F8 gene inactivation [51–53]. High diversity in the F8A gene copy number, genomic location, and gene structure indicates a complex evolution of the F8A gene family [40]. For instance, HAP40 is encoded by mostly X-chromosomally located, single-exon genes in amniotes and multi-exon genes located on autosomal chromosomes in non-amniotes (e.g., *Danio rerio*, *Drosophila melanogaster*, or *Caenorhabditis elegans*). These genes have been duplicated several times during mammalian evolution, as three F8A paralogs exist in humans but only two paralogs in human primates and laurasiatherians [40]. Importantly, most gene duplications that do not confer increased fitness either evolve to functionally inactive pseudogenes or are not fixed in the population and lost [54], suggesting that the F8A gene amplification and fixation confers an evolutionary benefit and that HAP40 probably fulfills essential functions.

Peters & Ross were the first to publish a detergent-resistant interaction between HAP40 and HTT [39]. HAP40 was then identified among the most abundant interactors of HTT in the murine brain [31, 55]. Beyond its abundance, the HTT-HAP40 interaction appears also to be evolutionary conserved, like in zebrafish [40] and fruit fly [56]. The cryo-EM structures of the HTT-HAP40 complex further showed

the unusual interaction properties selected for this complex, involving all three domains of HTT that shield HAP40 through large contact areas [22–24]. The exceptional stability [24] and *in vivo* abundance of the HTT-HAP40 complex [31, 39, 47, 55], the common origin of these proteins at the root of eukaryotes, and their likely coevolution [40] suggest functional relevance of the complex (Fig. 3).

The impact of HD progression on HAP40 mRNA and protein levels was studied in different models. HAP40 mRNA level was mostly unchanged in the striatum, cortex, and liver of HD mice models, with only modest elevations (1.2 to 1.35-fold changes) observed in less than 20% of tested conditions, inconsistently in terms of age, expansion size from Q80 to Q175, or tissue [57]. A recent meta-analysis study conducted on 220 HD patients and 241 healthy controls also concluded that the mRNA levels of HAP40 were not robustly altered [61]. Moreover, two studies reported that downregulating HTT does not affect HAP40 mRNA levels [24, 47]. Concerning the HAP40 protein, one study observed elevated HAP40 amounts in HD patient tissues and murine cells [59]. In contrast, several other studies reported reduced HAP40 protein levels in synaptosomes of HdH140Q/140Q mice [58], HD knock-in mice [57], and primary fibroblasts and lymphocytes from HD patients [47]. Two groups independently reported that HAP40 protein levels directly depended on HTT protein levels [24, 47] (Fig. 3), and a strong positive correlation between HAP40 and HTT levels was observed in all analyzed murine brain tissue and in human lymphocytes and fibroblasts [47]. Furthermore, HTT over-expression increased HAP40 half-life, as shown by cycloheximide- and puromycin-chase assays (Fig. 3) [47]. Hence, current evidence supports the possibility that the amount of HAP40 protein diminishes during HD progression, as do the HTT protein levels (Fig. 3) [62–65], and that HAP40 defines as an obligate partner of HTT (i.e., as a protein whose cellular levels are directly controlled by HTT protein amount, independently of mRNA expression levels, see [47, 62–65]).

Currently, the mechanism by which HTT regulates HAP40 levels is not known. Xu et al. reported that increased HAP40 levels could be observed in HTT knock-out cells after treatment with a proteasome inhibitor [56]. But the degradation mechanisms of HAP40 and of the HTT-HAP40 complex remain largely understudied. Also, other important questions regarding the partnership between HTT and HAP40 are mostly unaddressed. For instance, can the correla-

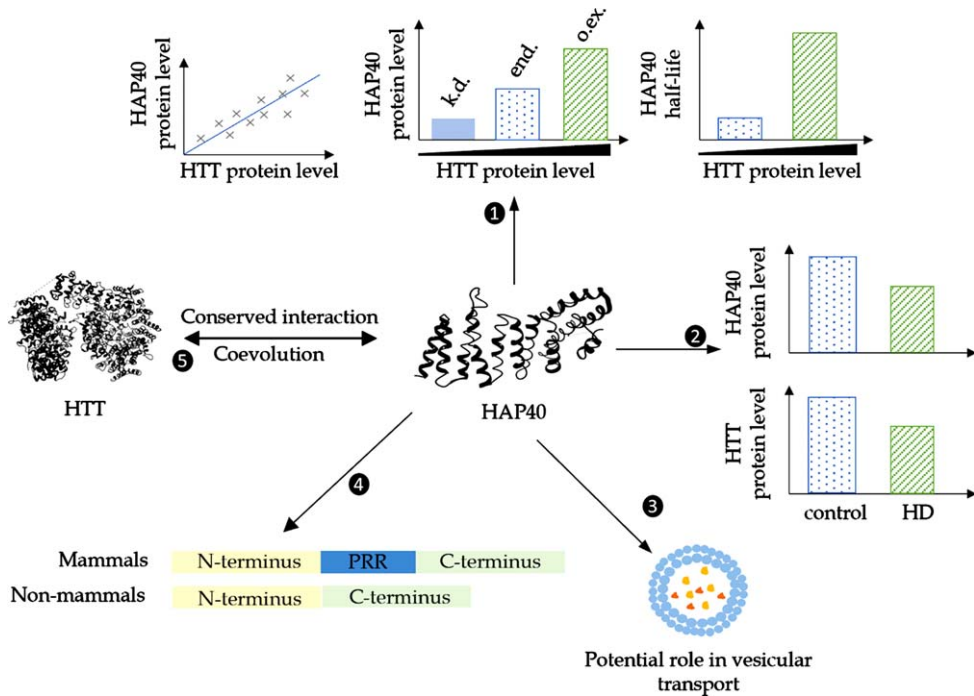


Fig. 3. Current knowledge of HAP40 and its interaction with HTT. (1) HAP40 protein levels directly depend on HTT levels in several tissues [46] and (2) the levels of both proteins are reduced in HD [47, 57, 58]. (3) Previous studies indicated a potential role of HAP40 in vesicular transport through an interaction with the Ras-related protein 5 [59, 60] or the close homology between HAP40 and the N-ethylmaleimide-sensitive factor attachment protein  $\alpha$ ,  $\beta$ , and  $\gamma$  [40]. (4) Further, mammalian HAP40 has a centrally located and mammalian-specific proline-rich region [40]. (5) The conserved interaction between HAP40 and HTT in metazoans and the coevolution between both proteins [40] further corroborates the functional importance of the HTT-HAP40 interaction. k.d., HTT knock-down; end., endogenous HTT expression; o.ex., over-expression of HTT.

tion between HTT and HAP40 levels be generalized to all cells and tissues? Is HAP40 only stabilizing HTT or does it fulfill other functions, as suggested by its accumulation in the nucleus when overexpressed in a HTT knock-down context [39]? Could other proteins replace HAP40 and stabilize HTT in different cellular or functional contexts? In this line, the closest homologues of HAP40 in current protein databases, the soluble N-ethylmaleimide-sensitive factor attachment proteins (SNAP) family, display a 3D structure strikingly resembling that of HAP40 but did not interact with HTT, as seen in co-expression and pull-down experiments [40]. More work is needed to address all these questions.

The evolution of the F8A gene suggesting that HAP40 probably fulfills an essential function [40], the exceptional stability of the HTT-HAP40 complex [24] and the *in vivo* dependence of HAP40 levels on HTT [24, 47, 57, 58] allow speculating on the potential contribution of HAP40 downregulation to the early embryonic lethal phenotype observed in HTT knock-out mice [66–68], and ask whether the

loss of HAP40—due to decreasing HTT levels during HD progression [62–65]—might play a role in the pathophysiology of the disease.

To date, the potential role of HAP40 in HD etiology remains poorly documented. It was reported that over-expressing HAP40 increased the aggregation of HTTex1, interfered with proteasomal degradation [69], and increased mitochondrial fragmentation [70]. However, HAP40 levels appear to be rather downregulated in the HD context [24, 47, 57, 58, 61] and the discrete contacts observed between HAP40 and HTTex1 [24] argue that most likely they would not interact outside the HTT-HAP40 context, suggesting that the increased HTTex1 aggregation may be an indirect effect of HAP40 over-expression. It was also shown that HAP40 and HTT may regulate autophagy [56] and endosome motility through their interaction with the Ras-related protein 5 (Rab5) (Fig. 3) [59, 60]. A mechanistic role of HAP40 in vesicle trafficking is conceivable since HAP40 is the closest homolog of the soluble N-ethylmaleimide-sensitive factor attachment proteins  $\alpha$  (SNAPA),

$\beta$  (SNAPB), and  $\gamma$  (SNAPG) [40], a family of TPR-containing proteins regulating the fusion of synaptic vesicles [71]. However, given the highly diverse scaffolding function of TPR-containing proteins and the long evolutionary time since the divergence between HAP40 and SNAPs [72, 73], it would be important to verify if HAP40 can indeed fulfill a similar function as SNAPA, SNAPB, or SNAPG.

Despite the abundant interaction between HTT and HAP40 [31, 39, 55] and the strong dependency of HAP40 on HTT levels (Fig. 3) [24, 47], which, taken together, suggests that large amounts of HAP40 are constitutively bound to HTT, it remains also important to verify if HAP40 may exist independently or in alternate HTT-free complexes. HAP40 was, for instance, proposed to localize in nuclear substructures independently of HTT [39, 74]. In line with the scaffolding function of TPR-containing proteins [72], HAP40 may indeed function as a hub protein orchestrating the interaction with other proteins. As described for other proline-rich domain-containing proteins [75], the centrally located proline-rich and mammalian-specific region of HAP40 (Fig. 3) [40] may mediate protein-protein interactions and permit regulation of HAP40 functions through post-translational modifications.

Finally, based on the close correlation between HAP40 and HTT levels in lymphocytes and fibroblasts from HD patients and in the striatum and cortex of an HD mouse model [47, 57], it was recently proposed that HAP40 might become a useful surrogate biomarker for HD, instead of the HTT protein itself [47]. This remains to be investigated.

## CONCLUSIONS

In summary, current data suggests that HAP40 is among the most abundant interactors of HTT and likely has essential functions in humans, suggesting a high functional relevance of the HTT-HAP40 complex [22–24, 31, 39, 40, 55]. Addressing the function of HAP40, for which limited information exists, could therefore provide new clues on the pathological mechanism of HD, as well as shed light on yet unknown potential therapeutic targets. Observations from different laboratories support the notion that the amount of HAP40 protein directly depends on HTT levels and that it decreases during HD progression [24, 47, 57, 58], as do HTT levels [62–65]. For these reasons, it would be interesting to study to which

extent decreasing HAP40 levels may contribute to the pathophysiology in HD, and whether HAP40 could be used as a biomarker of HD [47].

Among the hundreds of HTT interactors reported to date [76], HAP40 is the only one that could be extensively studied in a complex with HTT at the structural level. The structure of the HTT-HAP40 complex at near-atomic resolution could be solved because HAP40 locks the otherwise looser conformation of HTT, through interactions mediated by large hydrophobic contacts and a conserved charge clamp [22–24] (Fig. 2B, C). Of interest are the observations that potentially druggable pockets at the HTT-HAP40 interface may be exploited, and that specific features of the structure (conserved surfaces, charged zones, and large cavities in the N-HEAT domain) may have important functional relevance as potential interaction areas with specific macromolecular partners [24, 34].

Of particular importance regarding HD pathophysiology is whether polyQ length variations affect the structure of HTT or the HTT-HAP40 complex. This point was extensively addressed using biochemical, biophysical, and structural approaches. Based on the literature cited above, a large body of evidence suggests that the influence of polyQ on the structure of HTT, alone or in complex with HAP40, is minimal. PolyQ expansion appears to increase the conformational space of the HTTex1 while constraining that of other IDRs on the HTT-HAP40 surface. However, such subtle conformational changes do not affect the overall structure of the complex [23, 24]. Notably, no tangible differences could be observed in near-atomic resolution 3D-models of the core HTT-HAP40 complex [23], and polyQ length variation does not affect the stability [23, 34] or the dimensions [24] of the HTT-HAP40 complex in solution.

The behavior of polyQ stretches, which appear to be highly flexible independently of their length within the HTT-HAP40 complex [23, 24], is actually reminiscent of that observed when HTTex1 [77] or model polyQ sequences [78] were fused to small and stable carrier proteins, where they adopted a random coil conformation independently of the polyQ length and aggregated only if separated from the carrier protein [77, 78]. It appears thus that, when fused to a stable carrier protein, smaller than  $\sim 35$  kDa [76, 77] or as massive as the  $\sim 390$  kDa HTT-HAP40 complex [23, 24], the properties of polyQ of various lengths align with that observed in other truncated HTT constructs, where no abrupt conformational change was observed [79–81]. In the same line, it was

also demonstrated that anti-polyQ antibodies such as 1C2, 3B5H10, and MW1, which were previously thought to recognize a mutant-specific conformation in monomeric polyQ molecules, are binding polyQ peptides as short as Q10, in a linear and extended conformation [77, 78, 82–84].

If polyQ expansion does not drastically modify the structural properties of its carrier protein, how can we then explain the apparent increased or decreased affinities reported between mHTT and numerous partners [20, 21]? Importantly, such aberrant interactions were mostly observed with semi-quantitative techniques (yeast two-hybrid and co-precipitation) often using aggregation-prone HTT fragments (*e.g.* HTTex1) [20, 21]. However, it was shown using quantitative methods that polyQ expansion *per se* does not alter HTTex1 interaction when aggregate-free purified proteins were used, whereas the presence of aggregates could indeed influence the apparent binding affinity [85]. This emphasizes the importance to carefully assess, whether differences observed when comparing interactions with WT-HTT and mHTT are caused by the direct influence of the polyQ, or whether other factors may intervene (*e.g.*, presence of aggregates in biological tissues, proteome variations because of disease context).

In conclusion, our current knowledge regarding the influence of polyQ expansion on the properties of HTT, the HTT-HAP40 complex, and other protein systems indicates that polyQ lengthening does not provoke a drastic conformational change around the disease length threshold. HD pathogenesis models stating that polyQ expansion drastically alters HTT properties should thus be reconsidered.

## ACKNOWLEDGMENTS

M.S. and S.K. were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation – project number 412854449). We thank Dr. Eri Sakata for help with the preparation of Fig. 2 and Dr. Qiang Guo for the movie showing the HTT-HAP40 structure (Supplementary Movie 1). R.F.-B. has received funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through Germany's Excellence Strategy (EXC 2067/1- 390729940).

## CONFLICT OF INTEREST

The authors have no conflict of interest to report.

## SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: <https://dx.doi.org/10.3233/JHD-220543>.

## REFERENCES

- [1] A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. *Cell*. 1993;72(6):971-83.
- [2] Gusella JF, Lee JM, MacDonald ME. Huntington's disease: Nearly four decades of human molecular genetics. *Hum Mol Genet*. 2021;30(R2):R254-63.
- [3] Martí E. RNA toxicity induced by expanded CAG repeats in Huntington's disease. *Brain Pathol*. 2016;26(6):779-86.
- [4] Bañez-Coronel M, Ayhan F, Tarabochia AD, Zu T, Perez BA, Tusi SK, et al. RAN translation in Huntington disease. *Neuron*. 2015;88(4):667-77.
- [5] Monckton DG. The contribution of somatic expansion of the CAG repeat to symptomatic development in Huntington's Disease: A historical perspective. *J Huntingtons Dis*. 2021;10(1):7-33.
- [6] Iyer RR, Pluciennik A. DNA mismatch repair and its role in Huntington's disease. *J Huntingtons Dis*. 2021;10(1):75-94.
- [7] Sathasivam K, Neueder A, Gipson TA, Landles C, Benjamin AC, Bondulich MK, et al. Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. *Proc Natl Acad Sci U S A*. 2013;110(6):2366-70.
- [8] Saudou F, Humbert S. The biology of Huntingtin. *Neuron*. 2016;89(5):910-26.
- [9] DiFiglia M, Sapp E, Chase KO, Davies SW, Bates GP, Vonsattel JP, et al. Aggregation of Huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science*. 1997;277(5334):1990-3.
- [10] Lunkes A, Lindenberg KS, Ben-Haiem L, Weber C, Devys D, Landwehrmeyer GB, et al. Proteases acting on mutant huntingtin generate cleaved products that differentially build up cytoplasmic and nuclear inclusions. *Mol Cell*. 2002;10(2):259-69.
- [11] Landles C, Sathasivam K, Weiss A, Woodman B, Moffitt H, Finkbeiner S, et al. Proteolysis of mutant huntingtin produces an exon 1 fragment that accumulates as an aggregated protein in neuronal nuclei in Huntington disease. *J Biol Chem*. 2010;285(12):8808-23.
- [12] Ratovitski T, Gucek M, Jiang H, Chighladze E, Waldron E, D'Ambola J, et al. Mutant huntingtin N-terminal fragments of specific size mediate aggregation and toxicity in neuronal cells. *J Biol Chem*. 2009;284(16):10855-67.
- [13] Steffan JS, Kazantsev A, Spasic-Boskovic O, Greenwald M, Zhu YZ, Gohler H, et al. The Huntington's disease protein interacts with p53 and CREB-binding protein and represses transcription. *Proc Natl Acad Sci U S A*. 2000;97(12):6763-8.
- [14] Nucifora FC, Sasaki M, Peters MF, Huang H, Cooper JK, Yamada M, et al. Interference by huntingtin and atrophin-1 with cbp-mediated transcription leading to cellular toxicity. *Science*. 2001;291(5512):2423-8.
- [15] Ren PH, Lauckner JE, Kachirskaia I, Heuser JE, Melki R, Kopito RR. Cytoplasmic penetration and persistent infection of mammalian cells by polyglutamine aggregates. *Nat Cell Biol*. 2009;11(2):219-25.

- [16] Cicchetti F, Lacroix S, Cisbani G, Vallières N, Saint-Pierre M, St-Amour I, et al. Mutant huntingtin is present in neuronal grafts in Huntington disease patients. *Ann Neurol*. 2014;76(1):31-42.
- [17] Pearce MMP, Kopito RR. Prion-like characteristics of polyglutamine-containing proteins. *Cold Spring Harb Perspect Med*. 2018;8(2):a024257.
- [18] Bäuerlein FJB, Saha I, Mishra A, Kalemánov M, Martínez-Sánchez A, Klein R, et al. *In Situ* architecture and cellular interactions of PolyQ inclusions. *Cell*. 2017;171(1):179-87.e10.
- [19] Collinge J. Mammalian prions and their wider relevance in neurodegenerative diseases. *Nature*. 2016;539(7628):217-26.
- [20] Harjes P, Wanker EE. The hunt for huntingtin function: Interaction partners tell many different stories. *Trends Biochem Sci*. 2003;28(8):425-33.
- [21] Li SH, Li XJ. Huntingtin-protein interactions and the pathogenesis of Huntington's disease. *Trends Genet*. 2004;20(3):146-54.
- [22] Guo Q, Huang B, Jingdong C, Seefelder M, Engler T, Pfeifer G, et al. The cryo-electron microscopy structure of huntingtin. *Nature*. 2018;555(7694):117-20.
- [23] Huang B, Guo Q, Niedermeier ML, Cheng J, Engler T, Maurer M, et al. Pathological polyQ expansion does not alter the conformation of the Huntingtin-HAP40 complex. *Structure*. 2021;29(8):804-9.e5.
- [24] Harding RJ, Deme JC, Hevler JF, Tamara S, Lemak A, Cattle JP, et al. Huntingtin structure is orchestrated by HAP40 and shows a polyglutamine expansion-specific interaction with exon 1. *Commun Biol*. 2021;4(1):1374.
- [25] Andrade MA, Bork P. HEAT repeats in the Huntington's disease protein. *Nat Genet*. 1995;11(2):115-6.
- [26] Li W, Serpell LC, Carter WJ, Rubinsztein DC, Huntington JA. Expression and characterization of full-length human huntingtin, an elongated HEAT repeat protein. *J Biol Chem*. 2006;281(23):15916-22.
- [27] Vijayvargia R, Eband R, Leitner A, Jung TY, Shin B, Jung R, et al. Huntingtin's spherical solenoid structure enables polyglutamine tract-dependent modulation of its structure and function. *Elife*. 2016;5:e11184.
- [28] Andrade MA, Petosa C, O'Donoghue SI, Müller CW, Bork P. Comparison of ARM and HEAT protein repeats. *J Mol Biol*. 2001;309(1):1-18.
- [29] Palidwor GA, Shcherbinin S, Huska MR, Rasko T, Stelzl U, Arumughan A, et al. Detection of alpha-rod protein repeats using a neural network and application to huntingtin. *PLoS Comput Biol*. 2009;5(3):e1000304.
- [30] Hyman J, Chen H, Di Fiore PP, Camilli Pd, Brunger AT. Epsin 1 undergoes nucleocytoplasmic shuttling and its eps15 interactor NH(2)-terminal homology (ENTH) domain, structurally similar to Armadillo and HEAT repeats, interacts with the transcription factor promyelocytic leukemia Zn(2)+ finger protein (PLZF). *J Cell Biol*. 2000;149(3):537-46.
- [31] Shirasaki DI, Greiner ER, Al-Ramahi I, Gray M, Boontheung P, Geschwind DH, et al. Network organization of the huntingtin proteomic interactome in mammalian brain. *Neuron*. 2012;75(1):41-57.
- [32] Kaltenbach LS, Romero E, Becklin RR, Chettier R, Bell R, Phansalkar A, et al. Huntingtin interacting proteins are genetic modifiers of neurodegeneration. *PLoS Genet*. 2007;3(5):e82.
- [33] Dong X, Zong S, Witting A, Lindenberg KS, Kochanek S, Huang B. Adenovirus vector-based *in vitro* neuronal cell model for Huntington's disease with human disease-like differential aggregation and degeneration. *J Gene Med*. 2012;14(7):468-81.
- [34] Harding RJ, Loppnau P, Ackloo S, Lemak A, Hutchinson A, Hunt B, et al. Design and characterization of mutant and wildtype huntingtin proteins produced from a toolkit of scalable eukaryotic expression systems. *J Biol Chem*. 2019;294(17):6986-7001.
- [35] Huang B, Lucas T, Kueppers C, Dong X, Krause M, Bepperling A, et al. Scalable production in human cells and biochemical characterization of full-length normal and mutant huntingtin. *PLoS One*. 2015;10(3):e0121055.
- [36] Pace JB, Huang NN, Séguin JP, Esquina C, Olin E, Zhu G, et al. Efficient and scalable production of full-length human Huntingtin variants in mammalian cells using a transient expression system. *J Vis Exp*. 2021(178):doi: 10.3791/63190.
- [37] Seong IS, Woda JM, Song JJ, Lloret A, Abeyathne PD, Woo CJ, et al. Huntingtin facilitates polycomb repressive complex 2. *Hum Mol Genet*. 2009;19(4):573-83.
- [38] Jung T, Shin B, Tamo G, Kim H, Vijayvargia R, Leitner A, et al. The polyglutamine expansion at the N-Terminal of Huntingtin protein modulates the dynamic configuration and phosphorylation of the C-Terminal HEAT domain. *Structure*. 2020;28(9):1035-50.e8.
- [39] Peters MF, Ross CA. Isolation of a 40-kDa Huntingtin-associated protein. *J Biol Chem*. 2001;276(5):3188-94.
- [40] Seefelder M, Alva V, Huang B, Engler T, Baumeister W, Guo Q, et al. The evolution of the huntingtin-associated protein 40 (HAP40) in conjunction with huntingtin. *BMC Evol Biol*. 2020;20(1):162.
- [41] Rockabrand E, Slepko N, Pantalone A, Nukala VN, Kazantsev A, Marsh JL, et al. The first 17 amino acids of Huntingtin modulate its sub-cellular localization, aggregation and effects on calcium homeostasis. *Hum Mol Genet*. 2007;16(1):61-77.
- [42] Atwal RS, Xia J, Pinchev D, Taylor J, Eband RM, Truant R. Huntingtin has a membrane association signal that can modulate huntingtin aggregation, nuclear entry and toxicity. *Hum Mol Genet*. 2007;16(21):2600-15.
- [43] Atwal RS, Desmond CR, Caron N, Maiuri T, Xia J, Sipione S. Kinase inhibitors modulate huntingtin cell localization and toxicity. *Nat Chem Biol*. 2011;29(7):453-60.
- [44] El-Daher MT, Hangen E, Bruyère J, Poizat G, Al-Ramahi I, Pardo R, et al. Huntingtin proteolysis releases non-polyQ fragments that cause toxicity through dynamin 1 dysregulation. *EMBO J*. 2015;34(17):2255-71.
- [45] Sapp E, Valencia A, Li X, Aronin N, Kegel KB, Vonsattel J-P, et al. Native mutant huntingtin in human brain: Evidence for prevalence of full-length monomer. *J Biol Chem*. 2012;287(16):13487-99.
- [46] Wetzel R, Mishra R. Order, disorder, and conformational flux in the structural biology of Huntington's Disease. In: Bates G, Tabrizi S, Jones L, editors. *Huntington's disease*. 4th edition ed. Oxford; New York: Oxford University Press; 2014.
- [47] Huang B, Seefelder M, Buck E, Engler T, Lindenberg KS, Klein F, et al. HAP40 protein levels are huntingtin-dependent and decreased in Huntington disease. *Neurobiol Dis*. 2021;158:105476.
- [48] Zuccato C, Cattaneo E. Normal function of huntingtin. In: Bates G, Tabrizi S, Jones L, editors. *Huntington's disease*. 4th edition ed. Oxford; New York: Oxford University Press; 2014. pp. 243-73.

- [49] Levinson B, Kenwrick S, Lakich D, Hammonds G, Gitschier J. A transcribed gene in an intron of the human factor VIII gene. *Genomics*. 1990;7(1):1-11.
- [50] Levinson B, Kenwrick S, Gamel P, Fisher K, Gitschier J. Evidence for a third transcript from the human factor VIII gene. *Genomics*. 1992;14(3):585-9.
- [51] Naylor JA, Buck D, Green P, Williamson H, Bentley D, Giannelli F. Investigation of the factor VIII intron 22 repeated region (int22h) and the associated inversion junctions. *Hum Mol Genet*. 1995;4(7):1217-24.
- [52] Bagnall RD, Ayres KL, Green PM, Giannelli F. Gene conversion and evolution of Xq28 duplicons involved in recurring inversions causing severe hemophilia A. *Genome Res*. 2005;15(2):214-23.
- [53] De Brasi CD, Bowen DJ. Molecular characteristics of the intron 22 homologs of the coagulation factor VIII gene: An update. *J Thromb Haemost*. 2008;6(10):1822-4.
- [54] Zhang J. Evolution by gene duplication: An update. *Trends Ecol Evol*. 2003;18(6):292-8.
- [55] Sap KA, Guler AT, Bury A, Dekkers D, Demmers JAA, Reits EA. Identification of full-length wild-type and mutant huntingtin interacting proteins by crosslinking immunoprecipitation in mice brain cortex. *J Huntingtons Dis*. 2021;10(3):335-347.
- [56] Xu S, Li G, Ye X, Chen D, Chen Z, Xu Z, et al. HAP40 is a conserved central regulator of Huntingtin and a specific modulator of mutant Huntingtin toxicity. *bioRxiv*. 2020; <https://doi.org/10.1101/2020.05.27.119552>.
- [57] Langfelder P, Cantle JP, Chatzopoulou D, Wang N, Gao F, Al-Ramahi I, et al. Integrated genomics and proteomics define huntingtin CAG length-dependent networks in mice. *Nat Neurosci*. 2016;19(4):623-33.
- [58] Valencia A, Sapp E, Kimm JS, McClory H, Ansong KA, Yohrling G, et al. Striatal synaptosomes from Hdh140Q/140Q knock-in mice have altered protein levels, novel sites of methionine oxidation, and excess glutamate release after stimulation. *J Huntingtons Dis*. 2013;2(4):459-75.
- [59] Pal A, Severin F, Höpfner S, Zerial M. Regulation of endosome dynamics by Rab5 and Huntingtin-HAP40 effector complex in physiological versus pathological conditions. *Methods Enzymol*. 2008;438:239-57.
- [60] Pal A, Severin F, Lommer B, Shevchenko A, Zerial M. Huntingtin-HAP40 complex is a novel Rab5 effector that regulates early endosome motility and is up-regulated in Huntington's disease. *J Cell Biol*. 2006;172(4):605-18.
- [61] Seefelder M, Kochanek S. A meta-analysis of transcriptomic profiles of Huntington's disease patients. *PLoS One*. 2021;16(6):e0253037.
- [62] Gutekunst CA, Levey AI, Heilman CJ, Whaley WL, Yi H, Nash NR, et al. Identification and localization of huntingtin in brain and human lymphoblastoid cell lines with anti-fusion protein antibodies. *Proc Natl Acad Sci U S A*. 1995;92(19):8710-4.
- [63] Persichetti F, Carlee L, Faber PW, McNeil SM, Ambrose CM, Srinidhi J, et al. Differential expression of normal and mutant Huntington's disease gene alleles. *Neurobiol Dis*. 1996;3(3):183-90.
- [64] Tsvetkov AS, Arrasate M, Barmada S, Ando DM, Sharma P, Shaby BA, et al. Proteostasis of polyglutamine varies among neurons and predicts neurodegeneration. *Nat Chem Biol*. 2013;9(9):586-92.
- [65] Wu P, Lu MX, Cui XT, Yang HQ, Yu SL, Zhu JB, et al. A high-throughput-compatible assay to measure the degradation of endogenous Huntingtin proteins. *Acta Pharmacol Sin*. 2016;37(10):1307-14.
- [66] Nasir J, Floresco SB, O'Kusky JR, Diewert VM, Richman JM, Zeisler J, et al. Targeted disruption of the Huntington's disease gene results in embryonic lethality and behavioral and morphological changes in heterozygotes. *Cell*. 1995;81(5):811-23.
- [67] Zeitlin S, Liu JP, Chapman DL, Papaioannou VE, Efstratiadis A. Increased apoptosis and early embryonic lethality in mice nullizygous for the Huntington's disease gene homologue. *Nat Genet*. 1995;11(2):155-63.
- [68] Barnes GT, Duyao MP, Ambrose CM, McNeil S, Persichetti F, Srinidhi J, et al. Mouse Huntington's disease gene homolog (Hdh). *Somat Cell Mol Genet*. 1994;20(2):87-97.
- [69] Huang ZN, Her LS. The ubiquitin receptor ADRM1 modulates HAP40-induced proteasome activity. *Mol Neurobiol*. 2017;54(9):7382-400.
- [70] Huang ZN, Chung HM, Fang SC, Her LS. Adhesion regulating molecule 1 mediates HAP40 overexpression-induced mitochondrial defects. *Int J Biol Sci*. 2017;13(11):1420-37.
- [71] Clary DO, Griff IC, Rothman JE. SNAPS, a family of NSF attachment proteins involved in intracellular membrane fusion in animals and yeast. *Cell*. 1990;61(4):709-21.
- [72] Perez-Riba A, Itzhaki LS. The tetratricopeptide-repeat motif is a versatile platform that enables diverse modes of molecular recognition. *Curr Opin Struct Biol*. 2019;54:43-9.
- [73] D'Andrea LD, Regan L. TPR proteins: The versatile helix. *Trends Biochem Sci*. 2003;28(12):655-62.
- [74] Milman P, Woulfe J. Novel variant of neuronal intranuclear rodlet immunoreactive for 40 kDa huntingtin associated protein and ubiquitin in the mouse brain. *J Comp Neurol*. 2013;521(16):3832-46.
- [75] Kay BK, Williamson MP, Sudol M. The importance of being proline: The interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J*. 2000;14(2):231-41.
- [76] Alanis-Lobato G, Andrade-Navarro MA, Schaefer MH. HIPPIE v2.0: Enhancing meaningfulness and reliability of protein-protein interaction networks. *Nucleic Acids Res*. 2017;45(D1):D408-14.
- [77] Bennett MJ, Huey-Tubman KE, Herr AB, West AP, Ross SA, Bjorkman PJ. A linear lattice model for polyglutamine in CAG-expansion diseases. *Proc Natl Acad Sci U S A*. 2002;99(18):11634-9.
- [78] Klein FAC, Pastore A, Masino L, Zeder-Lutz G, Nierengarten H, Oulad-Abdelghani M, et al. Pathogenic and non-pathogenic polyglutamine tracts have similar structural properties: Towards a length-dependent toxicity gradient. *J Mol Biol*. 2007;371(1):235-44.
- [79] Bravo-Arredondo JM, Kegulian NC, Schmidt T, Pandey NK, Situ AJ, Ulmer TS, et al. The folding equilibrium of huntingtin exon 1 monomer depends on its polyglutamine tract. *J Biol Chem*. 2018;293(51):19613-23.
- [80] Newcombe EA, Ruff KM, Sethi A, Ormsby AR, Ramdzan YM, Fox A, et al. Tadpole-like conformations of Huntingtin Exon 1 are characterized by conformational heterogeneity that persists regardless of polyglutamine length. *J Mol Biol*. 2018;430(10):1442-58.
- [81] Warner JB, Ruff KM, Tan PS, Lemke EA, Pappu RV, Lashuel HA. Monomeric Huntingtin Exon 1 has similar overall structural features for wild-type and pathological polyglutamine lengths. *J Am Chem Soc*. 2017;139(41):14456-69.
- [82] Li P, Huey-Tubman KE, Gao T, Li X, West AP, Bennett MJ, et al. The structure of a polyQ-anti-polyQ complex reveals

- binding according to a linear lattice model. *Nat Struct Mol Biol.* 2007;14(5):381-7.
- [83] Klein FAC, Zeder-Lutz G, Cousido-Siah A, Mitschler A, Katz A, Eberling P, et al. Linear and extended: A common polyglutamine conformation recognized by the three antibodies MW1, 1C2 and 3B5H10. *Hum Mol Genet.* 2013;22(20):4215-23.
- [84] Owens GE, New DM, West AP, Bjorkman PJ. Anti-PolyQ antibodies recognize a short PolyQ stretch in both normal and mutant Huntingtin Exon 1. *J Mol Biol.* 2015;427(15):2507-19.
- [85] Davranche A, Aviolat H, Zeder-Lutz G, Busso D, Altschuh D, Trottier Y, et al. Huntingtin affinity for partners is not changed by polyglutamine length: Aggregation itself triggers aberrant interactions. *Hum Mol Genet.* 2011;20(14):2795-806.