



HAL
open science

The structure of pathogenic huntingtin exon-1 defines the bases of its aggregation propensity

Carlos A Elena-Real, Amin Sagar, Annika Urbanek, Matija Popovic, Anna Morató, Alejandro Estaña, Aurélie Fournet, Xamuel L Lund, Zhen-Dan Shi, Luca Costa, et al.

► To cite this version:

Carlos A Elena-Real, Amin Sagar, Annika Urbanek, Matija Popovic, Anna Morató, et al. The structure of pathogenic huntingtin exon-1 defines the bases of its aggregation propensity. 2022. hal-03846359

HAL Id: hal-03846359

<https://hal.science/hal-03846359>

Preprint submitted on 10 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1

2 **The structure of pathogenic huntingtin exon-1 defines the bases of its** 3 **aggregation propensity**

4

5

6 Carlos A. Elena-Real¹, Amin Sagar¹, Annika Urbanek¹, Matija Popovic¹, Anna Morató¹, Alejandro
7 Estaña^{1,2}, Aurélie Fournet¹, Xamuel L. Lund^{1,3}, Zhen-Dan Shi⁴, Luca Costa¹, Aurélien Thureau⁵,
8 Frédéric Allemand¹, Rolf E. Swenson⁴, Pierre-Emmanuel Milhiet¹, Alessandro Barducci¹, Juan
9 Cortés², Davy Sinnaeve^{6,7}, Nathalie Sibille¹, Pau Bernadó^{1,*}

10

11 1- Centre de Biologie Structurale (CBS), INSERM, CNRS, Université de Montpellier. 29, rue de Navacelles,
12 34090 Montpellier, France.

13 2- LAAS-CNRS, Université de Toulouse, CNRS, 31400 Toulouse, France.

14 3- Institut Laue Langevin, 38000 Grenoble, France.

15 4- The Chemistry and Synthesis Center, National Heart, Lung, and Blood Institute, National Institutes of
16 Health. Rockville, Maryland, USA.

17 5- Swing Beamline, Synchrotron SOLEIL, 91192 Gif sur Yvette, France

18 6- Univ. Lille, INSERM, Institut Pasteur de Lille, CHU Lille, U1167 - RID-AGE - Risk Factors and
19 Molecular Determinants of Aging-Related Diseases, 59000 Lille, France.

20 7- CNRS, ERL9002 - Integrative Structural Biology, 59000 Lille, France.

21

22 **Corresponding author:** Pau Bernadó (pau.bernado@cbs.cnrs.fr)

23

24

25 **Abstract**

26 Huntington's Disease is a neurodegenerative disorder caused by a CAG expansion of the first exon
27 of the *HTT* gene, resulting in an extended poly-glutamine (poly-Q) tract in the N-terminus of the
28 protein huntingtin (httex1). The structural changes occurring to the poly-Q when increasing its
29 length remain poorly understood mainly due to its intrinsic flexibility and the strong compositional
30 bias of the protein. The systematic application of site-specific isotopic labeling has enabled residue-
31 specific NMR investigations of the poly-Q tract of pathogenic httex1 variants with 46 and 66
32 consecutive glutamines. The integrative analysis of the data reveals that the poly-Q tract adopts
33 long α -helical conformations stabilized by glutamine side-chain to backbone hydrogen bonds. ^{19}F -
34 NMR of site-specifically incorporated fluoro-glutamines and molecular dynamics simulations
35 demonstrate that the mechanism propagating α -helical conformations towards the poly-Q from the
36 upstream N17 domain is independent of the poly-Q track length. Aggregation and atomic force
37 microscopy experiments show that the presence of long and persistent α -helices in the poly-Q tract
38 is a stronger signature in defining the aggregation kinetics and the structure of the resulting fibrils
39 than the number of glutamines. The ensemble of our observations provides a structural perspective
40 of the pathogenicity of expanded httex1 and paves the way to a deeper understanding of poly-Q
41 related diseases.

42

43 **Introduction**

44 Among the nine neurodegenerative disorders caused by expansions of polyglutamine (poly-Q)
45 tracts, Huntington's Disease (HD) stands out due to its prevalence and devastating effects¹. HD is
46 triggered by an abnormal expansion of the poly-Q tract located in exon1 (httex1) of the 348-kDa
47 huntingtin, a ubiquitous protein involved in multiple pathways^{2,3}. In its non-pathogenic form, the
48 httex1 poly-Q tract is comprised of 17-20 glutamines⁴; however, when the number of consecutive
49 glutamines exceeds the pathogenic threshold of 35, it results in aggregation-prone mutants. Indeed,
50 fragments of mutant httex1 can be found forming large cytoplasmic and nuclear aggregates within
51 neurons of the striatum, a well-known hallmark of HD^{5,6}. The presence of such aggregates, the
52 neuronal degeneration, the age of onset and disease severity, all correlate with the length of the
53 expanded poly-Q tract⁷. Notably, the mutant httex1 fragment alone suffices to reproduce HD
54 symptoms in mice⁸. Unfortunately, no effective treatment is currently available, mainly because of
55 the lack of knowledge regarding the molecular mechanisms underlying the disease⁹.

56 Amyloidogenic aggregates were initially proposed to be the toxic species in HD, acting by
57 sequestering essential cellular proteins and components¹⁰. However, they have also been related to
58 neuronal survival in certain cell types, suggesting that aggregates may be protective¹¹⁻¹⁴. It has also
59 been proposed that toxicity is caused by soluble species of httex1, which are characterized by a high
60 polymorphism comprising monomers, dimers, tetramers and small oligomers^{15,16}. It has been shown
61 that N17, the 17 residue-long fragment preceding poly-Q (Fig. 1a), is the pivotal element triggering
62 httex1 self-recognition and enhancing aggregation¹⁷⁻¹⁹. However, the structural mechanism by
63 which N17 propagates aggregation and cytotoxicity of expanded httex1 remains to be defined.

64 Two models have been suggested to connect the pathological threshold and toxicity, the 'toxic
65 structure' and the 'linear lattice' models²⁰. While the 'toxic structure' model proposes the
66 appearance of a distinct toxic conformation when the poly-Q tract is expanded beyond the
67 pathological threshold^{21,22}; the 'linear lattice' model suggests that poly-Q tracts are inherently toxic
68 and that their toxicity systematically increases with the homorepeat length^{23,24}. Intriguingly, both
69 models have been supported by antibody recognition experiments in different studies. Furthermore,
70 the absence of sharp changes in single-molecule Förster resonance energy transfer (smFRET),
71 circular dichroism and electron paramagnetic resonance (EPR) experiments around the pathological
72 threshold has been argued to substantiate the 'linear lattice' model²⁵⁻²⁷. High-resolution structures
73 of non-pathogenic and pathogenic httex1 variants are required to evaluate the changes occurring
74 upon poly-Q expansion, discriminate between both models and finally define the bases of
75 httex1 cytotoxicity.

76 Until recently, the detailed high-resolution structural characterization of soluble httex1, especially
77 of those with pathogenic poly-Q length, has been hampered by the intrinsic properties of the
78 protein, namely the highly flexible nature, its aggregation propensity and the strong compositional
79 bias. X-ray diffraction or electron microscopy, normally used to study folded proteins, cannot be
80 applied to probe disordered proteins such as httex1. Furthermore, the presence of low complexity
81 regions, such as the long glutamine and proline homorepeats, results in important signal overlap
82 when nuclear magnetic resonance (NMR) is used, hampering the collection of high-resolution
83 information²⁸. Despite these difficulties, several high-resolution NMR studies of non-pathogenic
84 httex1 and N-terminal fragments have been reported^{19,26,27,29,30}. However, due to the severe overlap,
85 only assignments of the first and last glutamines of the poly-Q tract could be achieved.

86 To circumvent the NMR signal overlap, our lab recently developed a site-specific isotopic labeling
87 (SSIL) strategy that combines cell-free protein expression and non-sense suppression, enabling the
88 investigation of homorepeats in a residue-specific manner^{31,32}. When applying this methodology to
89 a non-pathogenic version of httex1 containing a 16-residue-long poly-Q tract (H16), it was shown
90 that the protein was enriched in helical conformations whose length and stability were defined by
91 flanking regions³³. While the helical propensity was propagated from N17 to the poly-Q through a
92 hydrogen-bond network, it was blocked by the helix-breaking effect caused by the proline-rich
93 region (PRR) that follows the poly-Q tract (Fig. 1a). The structural effects imposed by N17 and
94 PRR may explain the positive and negative regulation of httex1 aggregation by both poly-Q
95 flanking regions, respectively^{34,35}. Whether these structural mechanisms also govern the
96 conformational landscape of pathogenic httex1 remains to be discerned.

97 In the present study, we applied SSIL to a pathogenic form of httex1 containing a 46 residue-long
98 poly-Q tract (H46) to unambiguously assign sixteen of these glutamines spread along the tract and
99 probed the structure and dynamics of the homorepeat. The integration of the NMR information and
100 small-angle X-ray scattering (SAXS) data provided the structural description of H46 as an ensemble
101 of elongated, partially helical conformations, whose propagation and stability mechanisms were
102 deciphered by ¹⁹F-NMR and molecular dynamics (MD) simulations. All together, our observations
103 provides a detailed structural perspective of the ‘linear lattice’ toxicity model, demonstrating that
104 the presence of long, persistent, aggregation-prone α -helices is concomitant to the expansion of the
105 poly-Q tract beyond the pathological threshold.

106
107
108
109

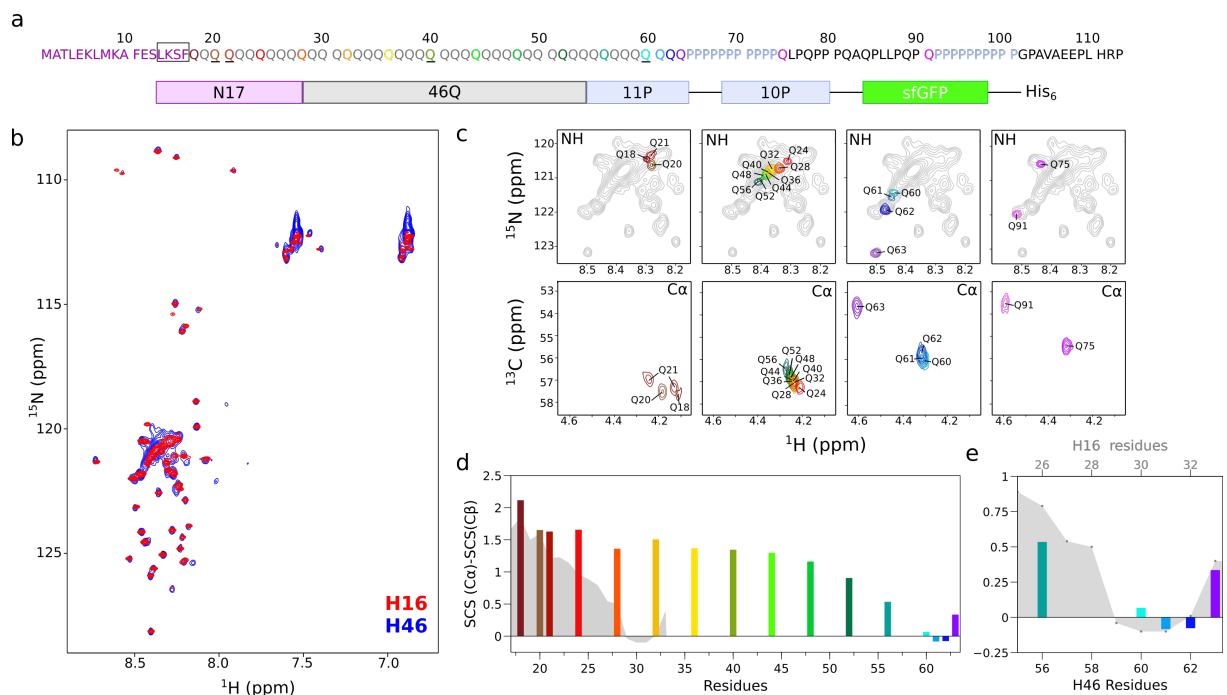
110 **Results**

111 **Pathogenic and non-pathogenic httex1 forms present similar structural features**

112 In order to study a pathogenic form of httex1, a construct comprising the N17 domain, a 46-residue-
113 long poly-Q tract and the PRR was fused to superfolder GFP (sfGFP) (Fig. 1a). A fully ¹⁵N-labeled
114 H46 sample was prepared and a ¹⁵N-HSQC NMR spectrum was recorded in order to evaluate its
115 general spectroscopic features. Similarly to H16³³, the spectrum revealed that while peaks from N17
116 and the PRR were well dispersed, the peaks corresponding to glutamine residues remained in a
117 large poorly resolved density (Fig. 1b). As expected, the increased molecular weight led to a
118 general peak broadening, which was particularly striking in the region of the expanded poly-Q tract.
119 The reduced stability of H46 at high concentration precluded the use of traditional 3D-NMR
120 experiments for the frequency assignment. Thus, to confirm the similarities found between the
121 pathogenic and non-pathogenic httex1 constructs, selectively labeled samples of H46 were prepared
122 (¹⁵N-Ala and ¹⁵N-Lys; ¹⁵N-Gly, ¹⁵N-Ser and ¹⁵N-Arg; ¹⁵N-Leu and ¹⁵N-Glu; and ¹⁵N-Phe), therefore
123 reducing the overlap of some N17 and PRR signals (Fig. S1a). The selective labeling showed that
124 the vast majority of peaks corresponding to N17 and PRR residues nicely overlap for both H46 and
125 H16. Interestingly, F17 peaks in H16 and H46 displayed different chemical shifts. Altogether, our
126 observations suggest that both httex1 forms share similar structural features outside of the poly-Q
127 tract, although some perturbations in the N17/poly-Q boundary are induced upon the extension of
128 the homorepeat.

129 In order to obtain high-resolution information on the poly-Q tract of H46, the SSIL approach using
130 previously optimized protocols³⁶ was used to specifically study 16 of the 46 glutamines of the H46
131 poly-Q tract, as well as two glutamines within the PRR (Fig. 1a). ¹⁵N-HSQC spectra of these
132 samples revealed that the glutamines adjacent to N17 (Q18, Q20 and Q21) adopted the lowest
133 chemical shift values of the poly-Q region without any specific trend, while the following
134 glutamines (Q24-Q56) exhibited steadily increasing ¹H and ¹⁵N chemical shifts. This last
135 observation points towards a gradual structural change along the homorepeat (Fig. 1c, upper
136 panels). Finally, the last glutamines of the tract (Q61-Q63) were found at the highest chemical
137 shifts, with Q63, affected by the adjacent poly-P tract, displaying an isolated peak outside of the
138 poly-Q density. The signals corresponding to Q75 and Q91 in the PRR did not show any specific
139 trend. The same features were observed when monitoring the C α -H α correlations of the same SSIL
140 samples (Fig. 1c, lower panels). Notably, the Q21 SSIL sample presented two C α peaks with
141 similar intensity, suggesting the presence of two slowly interconverting conformations. While one
142 of the peaks appeared close to these of Q18 and Q20, the second one was shifted to the same
143 spectral region of the central glutamines of the homorepeat.

144 The secondary chemical shift (SCS) analysis of the glutamines using a neighbor-corrected random
145 coil database³⁷ showed that the poly-Q tract was highly enriched in α -helical conformations, in line
146 with previous observations for non-pathological constructs^{26,29,33} (Fig. 1d). However, this
147 propensity was not homogeneous; the helicity reached its maximum at Q18 and presented a steady
148 plateau over 30 residues, until around Q48, in which the helical propensity was maintained. Only
149 for the C-terminal part of the H46 poly-Q a smooth decrease of helicity was observed, reaching
150 negative SCS values for the last glutamines of the tract. When comparing this SCS analysis with
151 that previously reported for H16³³, an increase in helicity for the N-terminal part of the H46 poly-Q
152 tract was observed. This phenomenon also explains the shift of the NH, $N\epsilon H_2$ and $C\alpha H\alpha$ NMR
153 signals of Q18 and Q21 towards more helical positions in H46 when compared with H16 (Fig.
154 S1b). Interestingly, the two $C\alpha$ peak observed for Q21 presented SCS values of 1.27 and 1.63 ppm,
155 corresponding to two conformations with different helical content.



156

157 **Figure 1. NMR analyses of H46 and comparison with H16.** (a) Primary sequence of H46 and scheme of the sfGFP-
158 fused construct used in this study. The color code identifies the site-specifically labeled glutamines throughout the
159 study. Underlined glutamines indicate the positions in which 2S,4R-fluoroglutamine (4F-Gln) was introduced for ¹⁹F-
160 NMR experiments. The box indicates the section mutated in the LKGG-H46 and LLLF-H46 mutants. (b) Overlay of
161 the ¹⁵N-HSQC spectra of fully labeled H46 (blue) with the previously reported for H16 (red)³³. (c) Zoom of the ¹⁵N-
162 HSQC (upper panels) and ¹³C-HSQC (lower panels) with individually colored SSIL spectra showing the poly-Q NH
163 and C α regions for different glutamine clusters (Q18-Q21; Q24-Q56; Q60-Q63; and PRR glutamines). In the upper
164 panels, the ¹⁵N-HSQC of the fully labeled ¹⁵N-H46 sample is shown in gray. (d) Secondary chemical shift (SCS)
165 analysis of H46 poly-Q using experimental C α and C β chemical shifts and a neighbor-corrected random-coil library³⁷.
166 The SCS analysis of H16 is shown in gray³³. (e) Comparison of the SCS values of glutamines flanking the PRR in H16
167 (gray area) and H46 (colored bars) when aligning both sequences from the C-terminus of the poly-Q tract.
168

169

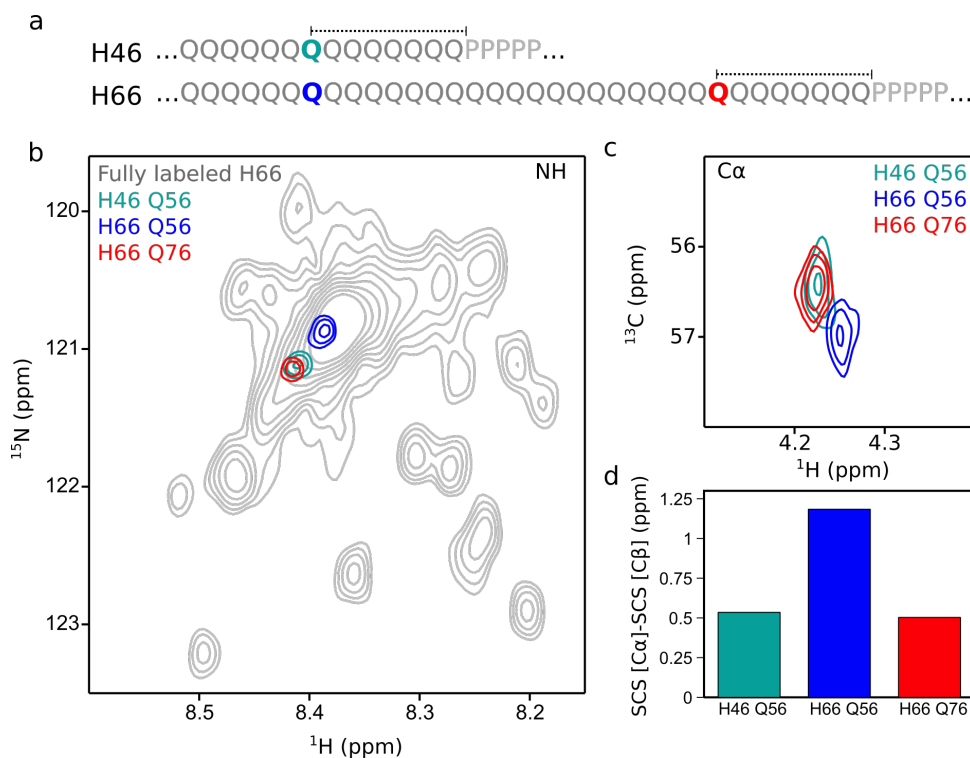
169 The extent of the structural effects induced by the PRR was analyzed by aligning the SCS values of
170 H16 and H46 from the C-terminus of the poly-Q tract (Fig. 1e). Interestingly, while the last four

171 glutamines of the tract (Q30-Q33 in H16 and Q60-Q63 in H46) showed the same conformational
172 trends, this similarity was reduced for more distant glutamines, such as Q56. Notice that this residue
173 is closer to the helix-promoting N17 in H16 (Q26) than in H46.

174

175 **H66 substantiates the persistence of long α -helical conformations in long poly-Q tracts**

176 The above results suggest that the distance to the PRR is the only parameter defining the length of
177 α -helical conformations in httex1 and that helical conformations encompass larger sections of the
178 homorepeat when the length of the poly-Q tract is increased. In order to validate this hypothesis, we
179 studied an httex1 construct with 66 consecutive glutamines (H66). The ^{15}N -HSQC of H66 produced
180 by cell-free expression nicely overlapped with that of H46, indicating that no structural changes
181 occur in httex1 upon incorporating twenty additional glutamines (Fig. S2a). The glutamine region
182 of H66 and H46 spectra displayed the same elongated shape and their maximum intensities were
183 centered at the same proton and nitrogen frequencies. Then, using our standard protocols, we
184 applied the SSIL strategy to two glutamines (Q56 and Q76) of H66 to probe their helical content.
185 Note that these two residues are located at equivalent positions to residue Q56 in H46 when
186 aligning both sequences from the N- and C-termini, respectively (Fig. 2a). The NH peak of H66-
187 Q76 appeared in a position equivalent to H46-Q56, whereas H66-Q56 had a lower chemical shift,
188 suggesting that this glutamine adopted a more helical structure than the other two residues (Fig. 2b).
189 This observation was further confirmed by monitoring the $C\alpha$ - $H\alpha$ correlations for the three residues
190 (Fig. 2c). Indeed, while the $C\alpha$ - $H\alpha$ peaks for H46-Q56 and H66-Q76 overlapped, H66-Q56 was
191 shifted towards more helical conformations. The helical propensity of these residues was quantified
192 with the SCS analysis, indicating an enhanced helicity for H66-Q56 (1.18 ppm) when compared
193 with H66-Q76 (0.50 ppm) and H46-Q56 (0.53 ppm) (Fig. 3c). Interestingly, the SCS value for H66-
194 Q56 was very similar to those observed for the plateau in H46, indicating that the additional twenty
195 glutamines in H66 adopt helical conformations (Fig. S2b). These results evidenced that the extent
196 of the α -helix-breaking capacity of the PRR is the same in all httex1 forms, allowing the α -helical
197 propensity to be propagated through a larger number of glutamines when the length of the
198 homorepeat is increased.



199

200 **Figure 2. NMR analyses of H66 and comparison with H46.** (a) Partial sequences of H46 and H66 indicating the
201 positions of Q56 in H66 (green) and H46 (blue), and Q76 in H66 (red). This color code identifies the individual
202 glutamines throughout the figure. Dashed lines highlight the equal distance of Q56 in H46 and Q76 in H66 to the PRR.
203 Zoom of the ^{15}N -HSQC (b) and ^{13}C -HSQC (c) with individually superimposed colored SSIL spectra showing the poly-
204 Q NH and $\text{C}\alpha$ regions, respectively. (d) Secondary chemical shift (SCS) analysis of H66-Q56, H66-Q76 and H46-Q56,
205 using experimental $\text{C}\alpha$ and $\text{C}\beta$ chemical shifts and a neighbor-corrected random-coil library.
206

207 SAXS indicates that H46 is an elongated particle in solution with a large degree of flexibility

208 Small angle X-ray scattering (SAXS) was applied to derive the overall size of H46 in solution. To
209 this end, we applied size-exclusion chromatography (SEC) coupled with SAXS, which allowed us
210 to isolate the H46 peak from potential aggregates and partially digested sfGFP (Fig. S3). The
211 analysis of the resulting profile (Fig. 3a and Table S2) indicates that H46 is a monomeric particle
212 with a radius of gyration, R_g , of $41.5 \pm 0.5 \text{ \AA}$ according to Guinier analysis and a maximum
213 dimension, D_{max} , of 148 \AA , calculated using pair-wise distance distribution, $p(r)$, analysis (Fig.
214 S3c). Although the Kratky representation displayed a peak indicating the presence of a folded
215 protein, which was attributed to the sfGFP, the intensity did not completely return to base line,
216 suggesting that the httex1 part of H46 exhibited a high level of flexibility (Fig. 3b). The smooth
217 decrease of the $p(r)$ when approaching the D_{max} value and the departure in the maximum of the
218 Kratky plot from the standard values of a globular protein substantiate the flexibility and the overall
219 extendedness of the protein.

220 To further confirm that the overall extendedness is inherent to httex1, we performed an equivalent
221 SAXS analysis for H16 (FigS. 3a, S3 and Table S2). Not surprisingly, the resulting SAXS profile of
222 H16 indicated that the protein is a smaller particle ($R_g=32.9 \pm 0.2 \text{ \AA}$ and $D_{max} = 126 \text{ \AA}$) in solution

223 than H46 (Table S2). Importantly, H16 retained SAXS features corresponding to an extended and
224 flexible particle observed for H46.

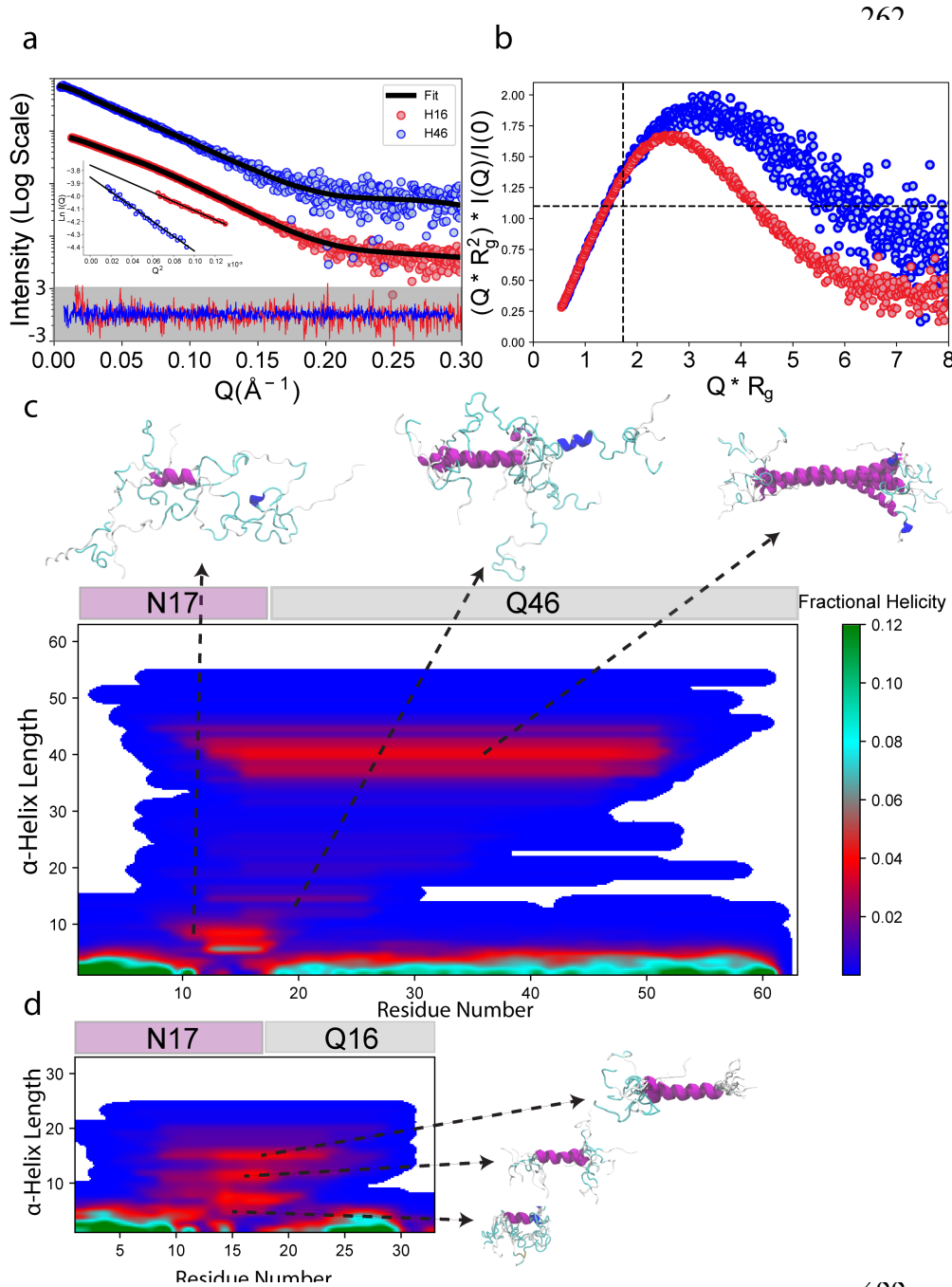
225

226 **H46 consists of a mixture of α -helical conformations of different lengths**

227 The $C\alpha$ chemical shifts and the SAXS data measured for H46 were integrated to derive a structural
228 model of the protein. Similar to our previous study of the non-pathogenic H16, two families of
229 ensembles were generated to capture the conformational influence of both flanking regions using a
230 conformational sampling method for intrinsically disordered proteins (see Methods section for
231 details)³⁸. For the first family (N \rightarrow C ensembles), starting with the ¹⁰AFESLKS¹⁶ region of N17 as
232 partially structured, multiple ensembles of 5,000 conformations were built by successively
233 including F17 and an increasing number of glutamines in the poly-Q tract (from Q18 to Q63) as
234 partially structured, while the rest of the chain was considered to be fully disordered. An equivalent
235 strategy was followed for the second family of ensembles (N \leftarrow C ensembles) for which glutamines
236 (from Q63 to Q18) were successively considered as partially structured starting from the poly-P
237 tract. Note that using this building strategy, secondary structural propensities are naturally
238 propagated due to the neighboring effects. For the resulting ensembles of each family and after
239 building the side chains with the program SCWRL4³⁹, averaged $C\alpha$ chemical shifts were computed
240 with SPARTA⁴⁰. For each family, the relative populations of these ensembles were optimized
241 using the experimental $C\alpha$ chemical shifts with Q55 as the boundary position (Fig. S4). Then, an
242 NMR-compatible ensemble was built by randomly taking conformation from the selected
243 ensembles with the appropriate populations (see Methods for details). This ensemble was further
244 refined by integrating the SAXS data with the ensemble optimization method (EOM)^{41,42}. For this,
245 we included the crystallographic structure of sfGFP (PDBID: 3LVA) and the C-terminal His-tag to
246 the individual conformations of the NMR-optimized ensemble to generate the pool of
247 conformations used by EOM (see Methods). It must be noted that the H46 ensemble had an overall
248 size ($R_g = 42.7 \text{ \AA}$) very similar to the experimentally determined one ($R_g = 41.5 \pm 0.5 \text{ \AA}$),
249 highlighting the quality of our starting model despite using local NMR information for the
250 refinement. Sub-ensembles selected with EOM yielded an excellent fit to the experimental profile
251 ($\chi^2 = 0.2$) (Fig. 3a). The resulting R_g distribution was quite broad and very similar to that of the
252 initial ensemble, indicating that H46 is a highly flexible particle in solution and slightly more
253 extended than the ensemble derived only using chemical shifts (Fig. S4).

254 The structural analysis of the NMR and SAXS compatible ensemble was performed with SS-map⁴³,
255 which displays the length, the residues involved and the population of the α -helices present in the
256 ensemble in a comprehensible manner. This analysis substantiated the presence of a mixture of
257 multiple helical conformations encompassing different sections of the H46 poly-Q tract (Fig. 3c).

258 These α -helices were initiated in the last residues of the N17 domain (14 LKSF 17) and propagated
 259 along the tract. Interestingly, an enrichment of long α -helices encompassing around 40 residues and
 260 reaching up to Q52 was observed. The presence of these long stable helical conformations explains
 261 the steady plateau observed in the SCS analysis (Fig. 1d).



289 **Figure 3. A structural model of pathogenic and non-pathogenic httex1 from the synergistic integration of NMR**
 290 **and SAXS data. (a)** The SAXS intensity profiles for H16 (red) and H46 (blue) along with the theoretical profiles of
 291 EOM selected sub-ensembles (black lines). The residuals from EOM fitting are shown at the bottom. The inset shows
 292 the Guinier plots with linear fits as black lines. **(b)** The normalized Kratky plots for H16 (red) and H46 (blue)
 293 displaying a shift from the values expected for globular proteins (shown as black dashed lines) on both X and Y axes.
 294 **(c)** SS-maps calculated from the conformations selected during 100 cycles of EOM for H46 (top) and H16 (bottom).
 295 The population of the different α -helix lengths is shown according to the color code on the right. Some representative
 296 conformations with different lengths of helices are also shown.
 297

298 A similar structural refinement was performed for H16, using the previously reported NMR-refined
299 ensemble³³ and the SAXS data (Fig. 3a). The EOM fit yielded an excellent agreement with the
300 experimental curve ($\chi^2 = 1.12$) and, as in the case of H46, the resulting sub-ensemble was slightly
301 more elongated than that obtained using only the NMR information. We observed that H16 also
302 consisted of a mixture of α -helical structures of different lengths, with prevalence for these
303 encompassing a large fraction of the homorepeat (Fig. 3d).

304

305 **Side chain to backbone hydrogen bonds trigger and stabilize helical conformations in H46**

306 It has been shown in non-pathogenic H16 and the androgen receptor that a network of hydrogen
307 bonds involving backbone and side chains is at the origin of the helical propagation from the N-
308 flanking region to the poly-Q tract^{33,44}. Here, we investigated whether this effect is conserved in
309 pathogenic httex1. To this end, we monitored the C β , H β , C γ , H γ and NH ϵ chemical shifts, which
310 could be unambiguously assigned through SSIL samples (Fig. 4a and S5). The two diastereotopic
311 H β glutamine protons displayed resolved responses, with the difference in their chemical shifts
312 increasing along the sequence. The maximum difference was observed for the last glutamines of the
313 tract (Q62 and Q63) and those within the PRR (Q75 and Q91), while Q18 displayed the smallest
314 one, featuring virtually degenerate H β chemical shifts (Fig. 4a, Fig. S5a). These observations
315 indicate a correlation between the chemical shift difference and the level of disorder. Interestingly,
316 Q21 exhibited three peaks that we attributed to the equilibrium between two conformations in slow
317 exchange at the NMR timescale. One of the two conformations of Q21 displayed the spectroscopic
318 features observed for Q18 (degenerate H β chemical shifts) and the other one those observed for the
319 rest of the glutamines (distinct H β chemical shifts). This observation is in agreement with the two
320 C α peaks observed for Q21 (Fig. 1c), which were attributed to two conformations with different
321 helical content.

322 The H γ signals followed an inverse trend than that observed for H β . The last glutamines of the tract
323 and those located in the PRR presented a single correlation, corresponding to degenerate
324 diastereotopic H γ chemical shifts, as expected for a flexible glutamine side chain (Fig. 4a and S5b).
325 Conversely, the majority of glutamines of the homorepeat (up to Q48) displayed a small but
326 measurable difference between the H γ chemical shifts, suggesting a transient rigidification of the
327 side chain. Again, in line with the H β signals, Q18, Q20 and Q21 were exceptions to this behavior.
328 While Q18 exhibited a single C γ -H γ peak in the ¹³C-HSQC, Q21 displayed three peaks,
329 substantiating the equilibrium between the two previously eluded conformational states. Although
330 Q20 only exhibited two C γ -H γ peaks, their difference in intensity suggested a similar situation as

331 for Q21, but with the C γ -H γ peak of the conformational state with degenerate H γ chemical shifts
332 overlapping with one of the C γ -H γ peaks of the other state.

333 These results confirm previous observations for a non-pathological version of httex1 and the
334 androgen receptor^{33,44}, demonstrating the presence of $i \rightarrow i+4$ bifurcated hydrogen bonds structurally
335 connecting the first residues of the poly-Q tract with the upstream flanking region (Fig. 4b).
336 Furthermore, our data indicate that these hydrogen bonds are present, although to a lower extent,
337 along the homorepeat, incorporating an additional mechanism for structural stabilization.

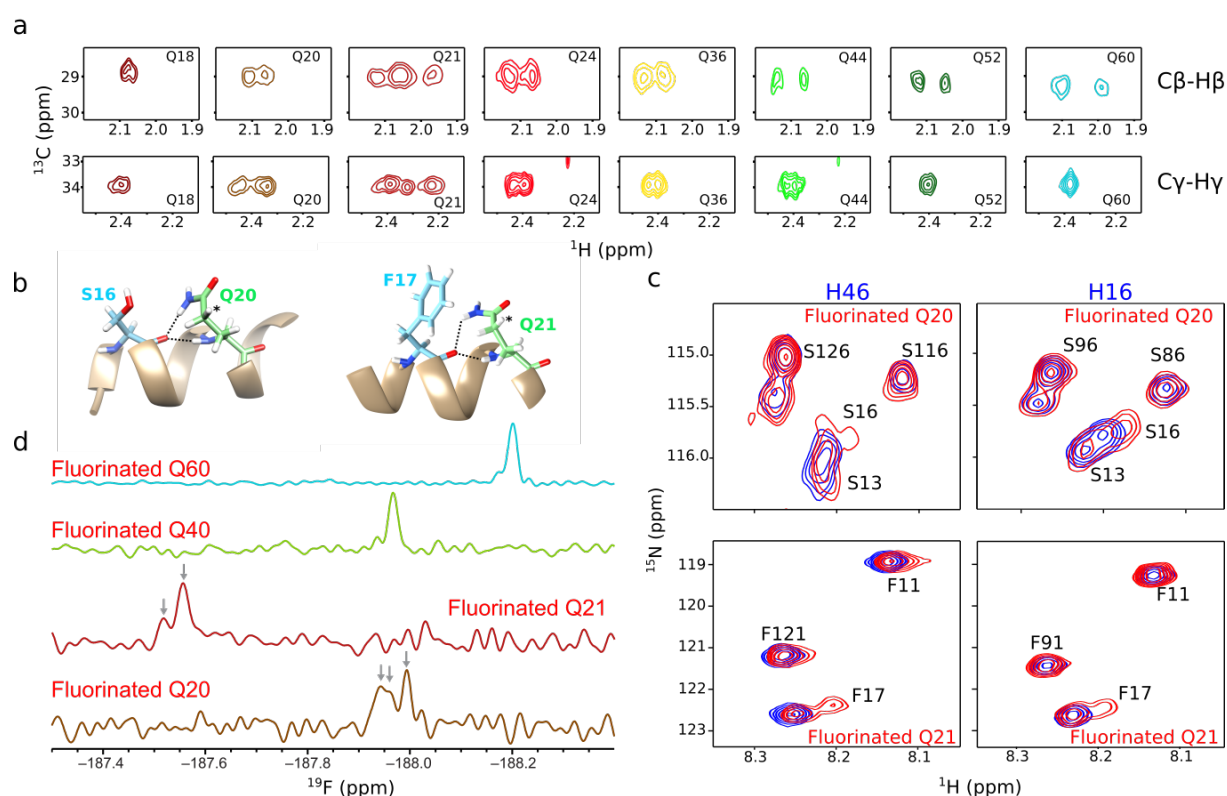
338

339 **2S,4R-Fluoroglutamine (4F-Gln) as a new probe to study the structure and dynamics of** 340 **httex1**

341 In order to confirm the different conformational behavior of glutamine side chains along the poly-Q
342 tract and benefiting from the SSIL methodology, a fluorinated glutamine was site-specifically
343 incorporated at different positions of H46 (Fig. 1a). Consequently, we synthesized with high yield
344 and stereospecificity 2S,4R-fluoroglutamine (4F-Gln), in which a fluorine atom replaced a
345 hydrogen atom on C γ (Fig. S6)⁴⁵. This 4F-Gln was successfully loaded onto the tRNA_{CUA} using the
346 yeast glutamyl-tRNA synthetase with similar yields as for canonical glutamine (Fig. S7),
347 suggesting that the fluorine atom is not changing the structural and electronic properties of
348 glutamine, enabling the recognition by the enzyme³¹.

349 4F-Gln was used in two sets of experiments. First, we used 4F-Gln to substantiate the participation
350 of Q20 and Q21 in the hydrogen bond network that propagates helicity in the homorepeat. To this
351 end, we incorporated 4F-Gln in positions Q20 or Q21 in two H46 samples that were also
352 isotopically labeled with ¹⁵N-Ser or ¹⁵N-Phe, respectively (Fig. 4c left panels and Fig. S8). When
353 comparing their ¹⁵N-HSQC spectra with those of non-fluorinated H46 (Fig. 4c, left panels),
354 substantial changes could be observed in S16 and F17 upon Q20 or Q21 fluorination, respectively.
355 The presence of a fluorinated glutamine in position 20 produced a slight chemical shift change of
356 S16, while fluorination of Q21 induced a stronger effect on F17, leading to the appearance of a
357 second peak. Furthermore, no chemical shift changes were observed for the other serines and
358 phenylalanines of H46, highlighting the specificity of the interaction. Importantly, similar
359 observations were made when 4F-Gln was incorporated in the same positions in H16, in samples
360 simultaneously labeled with ¹⁵N-Ser and ¹⁵N-Phe (Fig. 4c right panels and Fig. S8). Indeed, when
361 Q20 was fluorinated in the non-pathogenic form, not only S16 was affected, but F17 was also
362 perturbed (Fig. S8). Fluorination of Q21 again resulted in the appearance of a second F17 signal
363 and slightly affected S16. These data underline the structural coupling between Q20 and S16 as
364 well as Q21 and F17, in both pathogenic and non-pathogenic forms of httex1.

365 In a second set of experiments, we incorporated 4F-Gln in four H46 positions strategically located
 366 at the beginning (Q20 and Q21), the middle (Q40) and the end (Q60) of the poly-Q tract to be
 367 monitored by 1D ^{19}F -NMR (Fig. 1a). Note that the ^{19}F chemical shift is exquisitely sensitive to
 368 small differences in the electronic surroundings of the fluorine nucleus, and thus an excellent
 369 reporter on biomolecular structure and dynamics^{46,47}. Strong differences in ^{19}F chemical shifts were
 370 observed for the four positions despite the homogeneity of the amino acid sequence (Fig. 4b). This
 371 demonstrates important structural changes along the poly-Q tract, most likely linked to the amount
 372 of helical content. The chemical shift of Q21 was particularly high, most probably due to its
 373 proximity to the ring currents exerted by F17 in the hydrogen-bonded form (Fig 4b and 4d).



374
 375 **Figure 4. NMR analysis of H46 side chains.** (a) C β -H β and C γ -H γ regions of the ^{13}C -HSQC spectra of selected
 376 glutamines within the poly-Q. The spectra of Q60 display the standard behavior of disordered glutamines with non-
 377 degenerate and degenerate diastereotopic protons for C β -H β and C γ -H γ , respectively. (b) Structural model of the N17-
 378 poly-Q coupling showing bifurcated H-bonds between S16 and Q20 (left) and F17 and Q21 (right). The asterisk
 379 indicates the proton substituted with a fluorine atom when 4F-Gln was incorporated. (c) Zoom of ^1H - ^{15}N HSQC spectra
 380 of H46 (left panels) and H16 (right panels) samples labeled with either ^{15}N -Ser or ^{15}N -Phe. In red, spectra of samples
 381 with fluorinated glutamine at position Q20 (upper panels) or Q21 (lower panels). Non-fluorinated samples are colored
 382 in blue. (d) 1D ^{19}F spectra of H46 samples with fluorinated glutamines at positions Q20, Q21, Q40 or Q60.

383
 384 Interestingly, multiple ^{19}F -NMR resonances were observed for Q20 and Q21. For Q20, three
 385 distinct responses were measured, including two signals with very close chemical shifts that could
 386 only just be resolved. In line with the C γ -H γ peaks, it can be concluded that the Q20 side chain
 387 adopts at least two different conformations in a slow exchange regime. Given the sensitivity of the

388 ^{19}F chemical shift, the two closely resonating signals can be explained by the fluorine sensing the
389 different conformational states of neighboring glutamines. For Q21, a weak second ^{19}F signal could
390 be detected, in agreement with the two populations previously identified in the $\text{C}\alpha\text{-H}\alpha$, $\text{C}\beta\text{-H}\beta$ and
391 $\text{C}\gamma\text{-H}\gamma$ correlations. Altogether, the site-specific incorporation of 4F-Gln substantiates the structural
392 link of N17 with the poly-Q tract, the distinct glutamine side-chain conformational preference along
393 the poly-Q tract, and the presence of multiple conformations in the homorepeat.

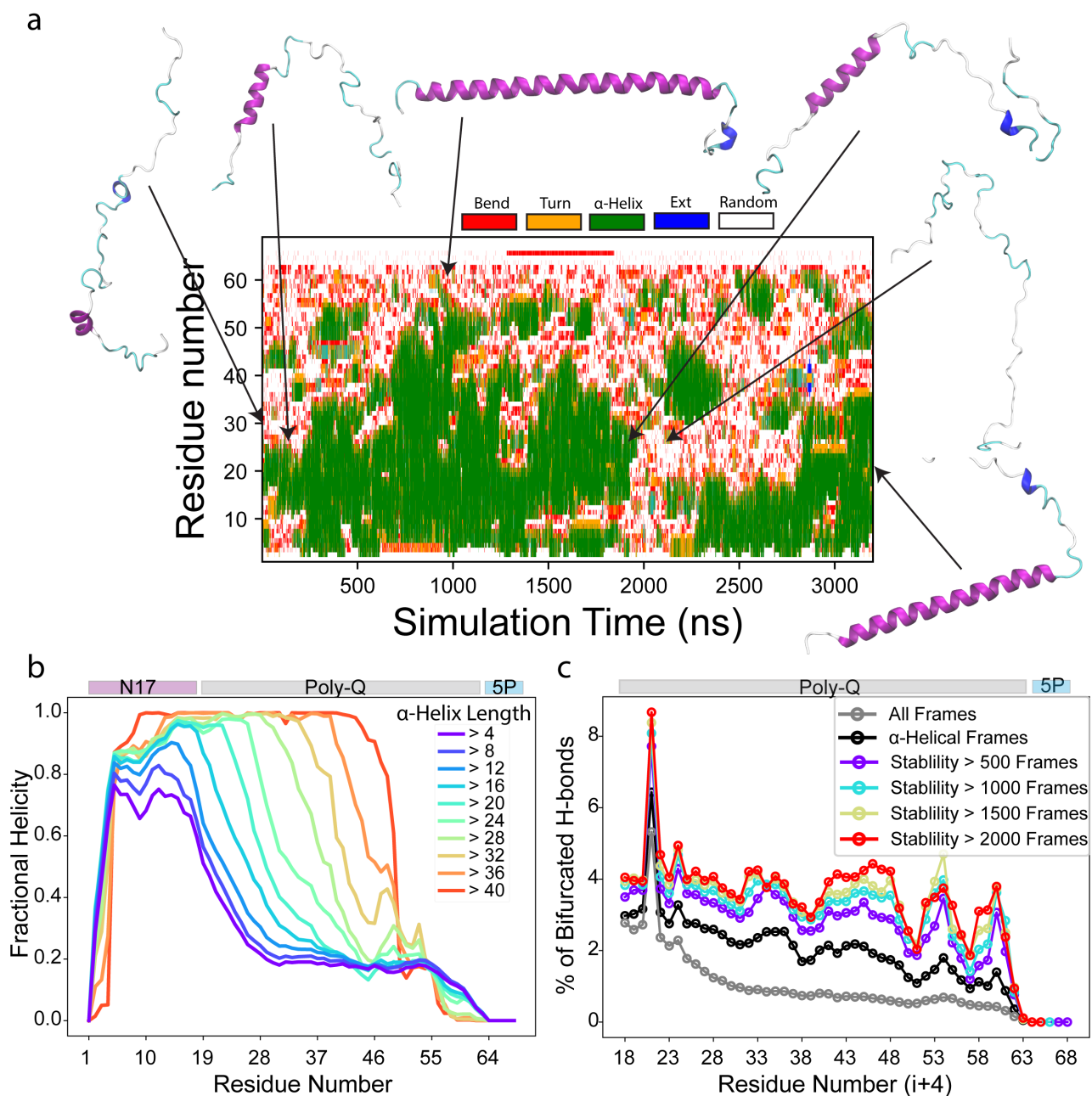
394

395 **Molecular dynamics simulations provides insights into the helix propagation and stabilization** 396 **mechanisms in httex1**

397 We performed Gaussian accelerated molecular dynamics (GaMD)⁴⁸ simulations to understand the
398 secondary structure propensities of httex1 and get further insights into the role of N17 in inducing
399 helicity. Previous MD studies of httex1 have yielded vastly different results ranging from
400 completely α -helical to highly β -strand rich structures, most likely due to the use of non-adapted
401 force-fields⁴⁹. To simulate a httex1 fragment encompassing N17, the polyQ tract and five prolines,
402 we used the recently developed ff99SBws-STQ force-field, which has been refined to describe the
403 secondary structure propensities of low-complexity proteins⁵⁰. In all the 8 independent MD
404 simulations, with an aggregated time of ≈ 20 μs , we observed that the poly-Q tract sampled a wide
405 conformational landscape and mainly adopted α -helical and disordered conformations, while N17
406 presented a higher helical propensity (Fig. 5a and S9a). Interestingly, several helical folding and
407 melting events could be observed along these trajectories. For some of the simulations, we observed
408 the formation of long α -helices that spanned almost the whole length of N17 and poly-Q tract and
409 were only absent in the residues close to the poly-P region (Fig. 5a). In line with the directionality
410 observed experimentally, the process of α -helix melting systematically occurred from C to N, as
411 observed in the 1600-1800 ns period in Fig. 5a, while α -helices in httex1 preferentially grew from N
412 to C (Fig. S10). The spontaneous formation of short α -helical conformations unconnected with N17
413 was also observed along the trajectories (*e.g.*, frames 2000-2300 in Fig. 5a), although they dissolved
414 relatively fast. This suggests a small inherent α -helical propensity of poly-Q tracts.

415 The weighted average per-residue fractional helicity of all the simulations indicated that the poly-Q
416 tract adopted higher α -helical propensity when closer to N17 (Fig. S9b). Moving further from N17,
417 the helicity decreased until Q30 from where it remained flat until Q55 to finally decrease when
418 approaching the poly-P. Importantly, this behavior was qualitatively very similar to the
419 experimentally determined α -helix propensity of httex1 (Fig. 1d and 1e). This suggests that our
420 simulations captured the structural mechanisms present in httex1, although the relative population
421 of α -helix in the poly-Q tract was underestimated. When analyzing the frames with α -helices with

422 increasing length (from 4 to 40), we observed a gradual increase in fractional helicity from N- to C-
 423 terminus, substantiating the structural propagation from N17 towards the poly-Q (Fig. 5b).



424

425 **Figure 5. Insights into the conformational landscape of httex1 with MD simulations.** (a) The per-residue secondary
 426 structure plot for one of the GaMD trajectories showing a transition from an almost completely random-coil
 427 conformation to a long α -helix and back to random coil. (b) The reweighted fractional helicity calculated using frames
 428 with increasing minimum α -helix length ranging from 4 to 40. (c) The percentage of frames with bifurcated hydrogen
 429 bonds for each residue of the poly-Q tract using all frames (gray), only the frames where the fragment ($i \rightarrow i+4$) was in
 430 helical conformation (black) and the segments of the trajectory where the fragment ($i \rightarrow i+4$) formed a stable α -helix
 431 for an increasing number of frames ranging from 500 to 2000.

432

433 We analyzed the trajectories for the existence of $i \rightarrow i+4$ bifurcated hydrogen bonds. Interestingly,
 434 they were found throughout the poly-Q, although they were more abundant in the beginning of the

435 tract where the hydrogen bonds were formed with the residues of N17 region and their number
436 slowly decreased along the tract (Fig. 5c). Residue Q21 presented the highest propensity to form
437 bifurcated hydrogen bonds with F17. Taken together, the trend of bifurcated hydrogen bonds agrees
438 with the results of the appearance of doublets in $C\gamma$ - $H\gamma$ correlations and the results from specifically
439 fluorinated glutamines probing the initial residues of the poly-Q (Fig. 4). Then, we analyzed the
440 correlation between bifurcate hydrogen bonds and α -helix stability. Not surprisingly, the percentage
441 of these hydrogen bonds was higher in frames where the segment ($i \rightarrow i+4$) adopted a helical
442 conformation. Importantly, the population consistently increased with the stability of the helix,
443 suggesting that bifurcated hydrogen bonds stabilize α -helical conformations in httex1 (Fig. 5c).

444

445 **The structure of the poly-Q governs aggregation kinetics and fibril structure of httex1**

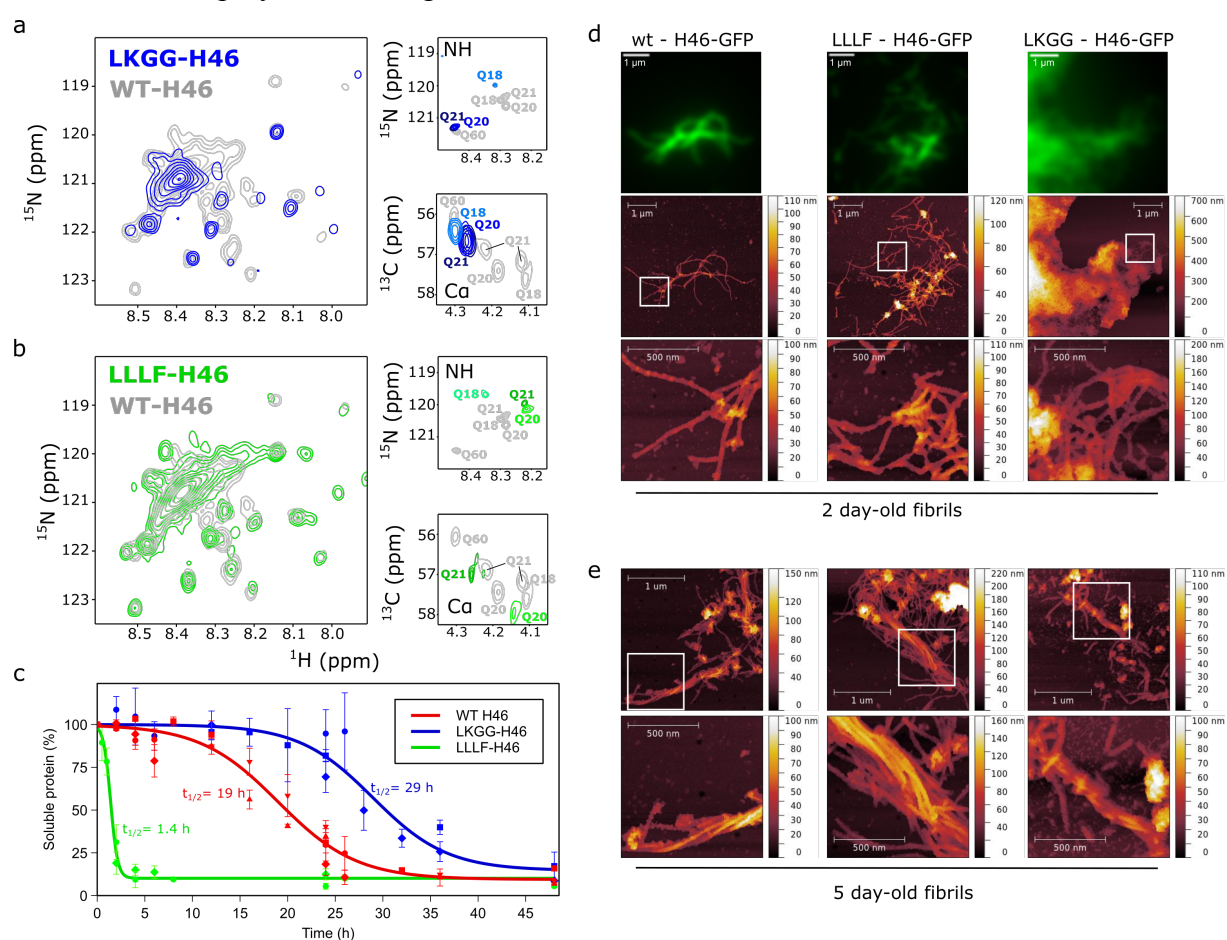
446 The hydrogen bond network connecting N17 and the poly-Q tract enabled the design of point
447 mutants in H16 altering the level of structure in the homorepeat while preserving its length³³. Its
448 application to H46 allowed the interrogation of the relative relevance of the helical content and the
449 poly-Q tract length for the aggregation propensity in a pathogenic variant of httex1. With this aim,
450 two N17 mutants of H46 were produced (Fig. 1a). First, by substituting ¹⁶SF¹⁷ by ¹⁶GG¹⁷ (LKGG-
451 H46), we hamper the H-bond network connecting both domains. Second, when ¹⁵KS¹⁶ were
452 replaced by ¹⁵LL¹⁶ (LLLH-H46), the hydrogen bond network was strengthened by incorporating two
453 additional large hydrophobic residues. Uniformly labeled and SSIL samples of these mutants were
454 produced and analyzed by NMR. On one hand, LKGG-H46 glutamine ¹⁵N-HSQC signals collapsed
455 in a broad, high-intensity, downfield-shifted peak, proving a substantial loss of helicity in
456 comparison with the wild-type H46 (Fig. 6a). Interestingly, this broad peak did not overlap with the
457 positions corresponding to fully unstructured glutamines, which were shifted further downfield.
458 This indicates that poly-Q, even when disconnected from the flanking region, contains a small
459 intrinsic propensity for helical conformations, in agreement with our MD simulations and previous
460 studies³⁵. On the other hand, the ¹⁵N-HSQC spectrum of fully labeled LLLH-H46 displayed a more
461 dispersed density of glutamine peaks and an additional upfield density, pointing to a helicity
462 increase of the poly-Q tract (Fig. 6b). The detailed analysis of NH, NεH₂ and Cα signals of Q18,
463 Q20 and Q21 SSIL samples from both mutants confirmed the decrease in structuration of LKGG-
464 H46 and the increase in helicity in LLLH-H46, in comparison with the wild-type form (H46) (Fig
465 6a,b and Fig. S11). Importantly, Q21 in LLLH-H46 displayed signatures of two conformations as
466 also observed for the wild-type. Interestingly, this residue exhibited two NHε correlation peaks,
467 suggesting the formation of a stronger bifurcated hydrogen bond than in the wild-type.

468 Unfortunately, a single $C\alpha$ peak, corresponding to the less helical conformation, was observed for
469 Q21 in this mutant, suggesting an unfavorable exchange regime for the NMR detection.

470 Once the structural features of the H46 mutants were confirmed, their propensities to aggregate
471 were measured. For this, 15 μ M samples of H46, LKGG-H46 and LLLF-H46 were incubated at
472 37°C and the soluble fractions of the proteins were analyzed by SDS-PAGE over a period of 48 h
473 (Fig. 6c and Fig. S12). Several essays were performed to better sample the monitored period. H46
474 presented a moderate aggregation propensity, with a systematic decrease of the soluble fraction
475 from the first hours of incubation that almost disappeared after 48 h. When the soluble fraction was
476 fitted to a reverse sigmoid function an aggregation half-time, $t_{1/2}$, of 19 h was obtained. A much
477 stronger aggregation propensity was observed for the LLLF-H46 mutant, for which a $t_{1/2}$ of only 1.4
478 h was derived. These observations indicate that increasing the α -helical stability strongly enhances
479 the aggregation propensity of httex1. When the same experiment was performed with the LKGG-
480 H46 mutant, which has a more disordered poly-Q tract similar to pure poly-Q peptides⁵¹, the first
481 signs of aggregation occur after 20 h of incubation and a $t_{1/2}$ of 29 h was obtained. The presence of
482 sfGFP fused to the C-terminus of the httex1 variants, which certainly slows down aggregation,
483 makes the quantitative comparison with previous studies difficult⁵². However, the relative
484 aggregation propensity of the three httex1 forms can be unambiguously obtained.

485 The morphology of the aggregates for H46 and the two mutants after 48 and 120 h were
486 investigated by correlative atomic force microscopy (AFM) – total internal reflection fluorescence
487 (TIRF) (Fig. 6d,e and S13a,b). The intrinsic fluorescence of the proteins allowed the easier
488 localization of the aggregates on the silica surface and indicated that no proteolytic activity
489 preceded the aggregation process of the samples. The inspection of H46 micrographs measured
490 after 48 h of incubation revealed the presence of typical amyloid structures with interconnected
491 fibrils, often presenting a length larger than 1 μ m (Fig. 6d, upper panels, left row). Those fibrils
492 exhibited a heterogeneous morphology, with notable variations in width and height, similarly to
493 those recently described^{53,54}. Interestingly, LLLF-H46 aggregates displayed similar features,
494 although the presence of fibrils was substantially more abundant (Fig. 6d, middle panels). Indeed,
495 large aggregates containing long fibrils were easily found in the sample (Fig S13a). Again, width
496 variations were often observed along the fibrils. The analysis of LKGG-H46 preparation revealed a
497 different behavior. Considerably larger and more heterogeneous aggregates, presenting often well-
498 defined limits, were observed for this mutant. Images indicated the presence of very short fibrils,
499 although relatively long isolated fibrils with a similar morphology than the wild-type and the LLLF-
500 H46 were found in the boundaries of the heterogeneous aggregates.

501 To verify whether the morphology was maintained after longer maturation times, 5-day-old fibrils
 502 were also imaged (Fig. 6e and Fig. S13b). Long fibrils were found in wild-type and LLLF-H46
 503 preparations and, interestingly, elongated bundle structures involving several paired filaments were
 504 observed. However, these coiled structures were seldom found in LKGG-H46 aggregates and, when
 505 observed, they were less ordered (Fig. 6e). Indeed, fibrils around LKGG-H46 bundles were
 506 fragmented and small aggregates could be detected along the whole sample, suggesting reduced
 507 fibril stability. Altogether, aggregation and AFM experiments demonstrate that the structural
 508 properties of httex1 exert a stronger influence on the aggregation propensity and the final form of
 509 the fibrils than the poly-Q tract length.



510

511 **Figure 6. Structural and fibrillation analyses of H46 N17 mutants.** (a, b) Overlay of the glutamine region of the
 512 ¹⁵N-HSQC of fully labeled H46 (gray) with the N17 mutants LKGG-H46 (blue) and LLLF-H46 (green), as well as
 513 zooms of the NH and Ca regions of the corresponding Q18, Q20 and Q21 SSIL samples. (c) Time course of
 514 aggregation for wild-type, LKGG- and LLLF-H46 (15 μM) at 37 °C. Each data point corresponds to the mean and
 515 associated standard deviation calculated from three replicates. Symbols represent different independent experiments.
 516 Half-time (t_{1/2}) values calculated for each H46 species are indicated according to the color code shown in the legend. (d)
 517 Fluorescence microscopy (upper panels) and AFM (middle and lower panels) images of 2-day-old fibrils of wild-type,
 518 LLLF- and LKGG-H46. Each fluorescence image corresponds to the average of 150 pictures. (e) AFM images of 5-
 519 day-old fibrils of three H46 species. White squares indicate the zoom region displayed in the panels below.

520

521

522

523 Discussion

524 Huntington's disease is the most notorious example of the poly-Q-related diseases and findings
525 connecting structure and disease unveiled for huntingtin are most probably applicable to the whole
526 family of pathologies¹. In this study, we have characterized at high resolution a pathogenic form of
527 httex1 containing 46 consecutive glutamines, a system that is out of the reach for traditional
528 structural biology approaches. We demonstrate that SSIL can be systematically applied to highly
529 repetitive sequences, such as H66, and the only limitations arise from the capacity to purify the
530 protein and preserve its stability in solution.

531 The systematic application of SSIL to H46 demonstrates that this protein retains the α -helical
532 conformation previously observed for non-pathogenic versions of httex1^{26,29,33}. Using circular
533 dichroism, α -helical propensity in pathogenic httex1 had been previously identified for httex1 with
534 up to 55 glutamines²⁷. Moreover, it was shown that the helical content increased concomitantly with
535 the length of the tract. However, the low-resolution of circular dichroism hampered the analysis of
536 the extent and stability of helical conformations. One of the most striking observations of our study
537 is the fairly flat plateau of measured secondary chemical shifts along a large fraction of the poly-Q
538 tract (Fig. 1d). These chemical shifts are consistent with the coexistence of multiple partially
539 formed α -helices of different lengths, spanning almost the complete poly-Q tract. However, these
540 helices are not equally populated and long α -helices are prevalent according to our synergistic
541 analysis of NMR and SAXS data (Fig. 3). This behavior evidences that the helices, which are
542 formed through a nucleation process triggered by the interaction between the N17 and the first
543 glutamines of the tract, are cooperatively propagated along the homorepeat. A faster decrease in the
544 SCS values would be expected if all helical lengths would display a similar thermodynamic
545 stability. The comparison of the NMR observables for H16 with those of H46 also substantiates the
546 helical stabilization with the poly-Q length. For instance, despite both proteins having the same
547 sequence context, NH, C α and N ϵ H₂ chemical shifts for initial glutamines of the tract are
548 systematically shifted towards more helical conformations in H46 than in H16 (Fig. S1). Moreover,
549 experiments performed on H66 demonstrate that α -helical conformations are maintained for long
550 poly-Q tracts and the distance of individual glutamines to the PRR defines their helical propensity
551 (Fig. 2).

552 From a structural perspective, the α -helical stability found in httex1 could arise from the formation
553 of bifurcate hydrogen bonds all along the homorepeat. Indeed, up to Q48, the diastereotopic H γ
554 protons have non-degenerate chemical shifts, although the chemical shift difference decreases along
555 the repeat (Fig. S5). An opposite trend is observed for the diastereotopic H β protons, which present
556 increasing chemical shift differences when reaching the less structured glutamines of H46. These

557 spectroscopic features have been associated with the formation of bifurcated hydrogen bonds and
558 the concomitant rigidification of glutamine side chains^{33,44}. Our MD simulations indicated that the
559 percentage of bifurcated hydrogen bonds is correlated with the stability of α -helical conformations.
560 The ensemble of these observations suggests that glutamine side chains are actively involved in the
561 stabilization of the helical conformation of the poly-Q tract and that the strength of this mechanism
562 declines when approaching the PRR.

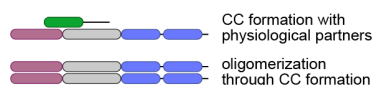
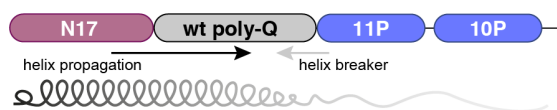
563 Partial structuration in the poly-Q implies that individual glutamines co-exist in (at least) two
564 different conformational states. A similar conclusion was reached by EPR experiments after
565 incorporating stable radicals in several positions of a pathogenic httex1²⁷. In that study, two
566 dynamic regimes were identified in N17 and poly-Q tract residues, but not in glutamines in the
567 PRR. However, structural details of this conformational fluctuation could not be unveiled. Several
568 NMR signatures collected for H46 using SSIL samples, which can be rationalized through the MD
569 simulations, define the structural bases of this conformational equilibrium. These include, among
570 others, the two $C\alpha$ - $H\alpha$ peaks for Q21 (Fig. 1c), the two sets of $C\beta$ - $H\beta$ and $C\gamma$ - $H\gamma$ peaks for Q20 and
571 Q21 (Fig. 4a), and the multiple ¹⁹F-NMR frequencies also detected for these two residues.
572 Remarkably, the co-existence of a conformational equilibrium is also manifested in most of the
573 other glutamines of the tract, which exhibit the previously eluded non-degenerate $C\gamma$ - $H\gamma$ peaks. Our
574 observations demonstrate that httex1 fluctuates between a rigid α -helical conformation, which is
575 stabilized with bifurcate hydrogen bonds, and a more disordered state. Furthermore, our data
576 suggest that the dynamic regime also changes in conjunction with the stability of the helical
577 conformation. While the first glutamines of the tract, which have the highest helical propensity,
578 exhibit a slow exchange on the NMR frequency timescale, a fast exchange regime is observed for
579 the other helical glutamines. This asymmetric behavior suggests that helix unwinding is initiated in
580 the proximity of the PRR and progresses towards the N-terminus, substantiating the previously
581 proposed protective role exerted by the PRR^{34,35,55,56}.

582 The synergistic combination of NMR and SAXS data has enabled the elucidation of an ensemble
583 model of H46⁵⁷, showing that the presence of long α -helices determines the overall shape of httex1.
584 The refined ensemble indicates that H46 is a flexible elongated particle in solution and that the
585 overall size is correlated with the length of the poly-Q tract (Fig. 3c). Importantly, our structural
586 model is in contradiction to previously reported compact models of httex1^{25,26,49,58}. In these models,
587 compactness is driven by extensive fuzzy contacts between N17 and the poly-Q, a situation that is
588 not compatible with our data. Despite the overall extendedness of the httex1 structure derived here,
589 our ensemble description requires that a large fraction of the protein ensemble remains disordered.
590 This disorder explains the fluorescence transfer efficiency observed in smFRET measurements²⁵

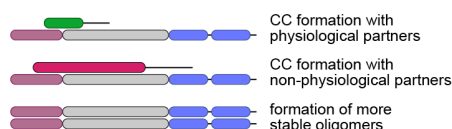
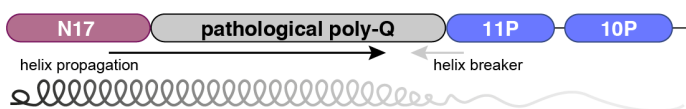
591 and the lack of permanent hydrogen bonds observed in NMR hydrogen deuterium exchange (HDX)
592 experiments²⁶.

593

non-pathological httex1



pathological httex1



594

595 **Figure 7. Scheme illustrating the structural influences within non-pathogenic and pathogenic httex1 and their**
596 **respective modes of interaction. (Top, non-pathogenic httex1)** The poly-Q tract of non-pathogenic httex1
597 experiences opposing structural effects from the N17 (α -helix propagation; black arrow) and the PRR (helix breaking;
598 gray arrow). The gradually shaded helix represents the decrease in helical propensity. Due to its helical conformation,
599 non-pathogenic httex1 can form coiled coil (CCs) with physiological partners or with other httex1 molecules, resulting
600 in oligomers. **(Bottom, pathogenic httex1)** Pathogenic httex1 experiences the same structural effects, however, helix
601 propagation outweighs the helix-breaking effect coming from the PRR and enhances the formation of CCs. Pathogenic
602 httex1 can still interact with its physiological partners *via* CC formation, but it can also interact with non-physiological
603 partners and form more stable oligomers, which eventually drive to fibrils.

604

605 The similarity of the mechanisms defining the structure of non-pathogenic and pathogenic forms of
606 httex1 validates 'linear lattice' as the model accounting for the existence of the pathological
607 threshold^{20,22–24}. According to this model, the expansion of the poly-Q induces a systematic increase
608 of the toxicity of httex1 and, beyond the pathological threshold, triggers cytotoxicity and neuronal
609 aggregation. Importantly, our study provides a structural perspective for this model. Our results
610 demonstrate that the poly-Q extension is associated with an increase in the length and stability of
611 the coexisting helical conformations. This is relevant as our aggregation experiments
612 unambiguously show that the α -helical content and not the homorepeat length is the key factor
613 promoting aggregation. Although N17 has been demonstrated to be the aggregation-triggering
614 domain^{18,30,34}, httex1 becomes aggregation-prone when N17 is structurally coupled to the poly-Q
615 tract. This suggests that when structurally uncoupled, such as in the LKGG-H46 mutant, partially
616 helical N17 is still able to oligomerize, but the resulting oligomers are less stable and aggregation
617 propensity is notably reduced. Conversely, strengthening the structural coupling between both
618 domains stabilizes the helical content of the protein and accelerates aggregation, as observed for
619 LLLF-H46. In line with these observations, previous aggregation experiments with httex1
620 analogues modifying the poly-Q tract structure highlighted the relevance of the homorepeat
621 conformational preferences in defining the aggregation propensity^{15,59}. In addition to modified
622 aggregation kinetics, we have observed that mutations in the flanking regions also result in fibers

623 with distinct morphologies, with the LKGG-H46 mutant forming shorter amyloids and exhibiting
624 less capacity to associate to form bundles. Polymorphism in httex1 aggregates has been previously
625 observed *in vivo* and *in vitro* when modifying the experimental conditions or when deleting
626 flanking regions^{15,52,60-62}. This polymorphism has been associated to the contribution of N17 and
627 the PRR to the fibril packaging^{62,63}. Conversely, the amyloid core is consistently formed by
628 antiparallel poly-Q stretches forming β -sheet monolayers connected through interdigitated
629 glutamine side chains^{63,64}.

630 In light of our observations, we can speculate about the structural bases of the mechanisms leading
631 to the pathology (Fig. 7). The helical propensity in the poly-Q tract may facilitate intermolecular
632 assemblies through coiled coil (CC) interactions. Indeed, more than 60% of the described httex1
633 partners have, or are predicted to have, CCs⁶⁵. Furthermore, CC formation has also been suggested
634 to be an important step for the oligomerization and subsequent aggregation of httex1^{17,19}. In the
635 non-pathological scenario, the short length and relatively low stability of the poly-Q helical
636 conformations precisely define the selectivity of partners recognized by httex1. Moreover, resulting
637 httex1 oligomers exhibit reduced stability. When the number of glutamines exceeds the pathological
638 threshold, there is a concomitant population increase of long α -helical conformations. These long
639 poly-Q helices could still interact with their physiological partners, although most probably with
640 different thermodynamic properties compared to the non-pathological scenario. However, they
641 could also associate with other non-physiological partners, sequester them and perturb crucial
642 signaling and metabolic pathways, inducing the symptoms associated to HD². In terms of the
643 fibrillation capacity, longer helices can form more stable oligomers through CC interactions. These
644 oligomers can eventually nucleate the formation of the amyloidogenic fibrils^{17,66} or being the source
645 of cytotoxicity by sequestering crucial cellular components⁶⁷.

646 Altogether, this study shows that the expansion of the poly-Q tract in pathogenic httex1 is
647 associated with an increase in the length and stability of α -helical conformations, which are the
648 main driving force for the enhanced aggregation propensity. Our results provide a high-resolution
649 structural perspective of the pathological threshold in HD that goes beyond the length of the poly-Q
650 tract by underlining the associated conformational preferences. The generalization of these
651 observations to the other poly-Q related diseases remains to be unveiled. However, the possibility to
652 explore their associated proteins at the residue level and independently of the length of the
653 homorepeat paves the way to a detailed structural understanding of the origin of these pathologies.

654 **Materials and Methods**

655 **Huntingtin exon1 constructs**

656 All plasmids were prepared as previously described³³. Synthetic genes of wild-type huntingtin
657 exon1 with 16, 46 and 66 consecutive glutamines (H16, H46 and H66 respectively) or H46 and H66
658 carrying the amber codon (TAG) instead of the glutamine codon, *e.g.* Q18 (H46Q18), were ordered
659 from GeneArt. Following this strategy, 16 amber mutants for H46 and two for H66 were purchased:
660 Q18, Q20, Q21, etc. Synthetic genes of the structural H46 mutants (LKGG-H46 and LLLF-H46)
661 and their corresponding amber codon mutants (Q18, Q20, Q21) were also ordered from GeneArt.
662 All genes were cloned into pIVEX 2.3d, giving rise to pIVEX-httex1-3C-sfGFP-His₆ and mutants.
663 The sequence of all plasmids was confirmed by sequencing by GENEWIZ.

664

665 **Synthesis of 2S,4R-fluoroglutamine**

666 The synthesis of the 2S,4R-fluoroglutamine ((2S, 4R)-2,5-Diamino-4-fluoro-5-oxopentanoic acid),
667 4F-Gln, was performed as detailed by Qu *et al.*⁴⁵. The purity and the enantiomeric excess (98%)
668 were evaluated by ¹H- and ¹⁹F-NMR (see Fig. S6 in Supplementary information).

669

670 **Preparation and aminoacylation of suppressor tRNA_{CUA}**

671 A tRNA_{CUA}/tRNA synthetase pair based on the Gln2 tRNA and glutamine ligase GLN4 from
672 *Saccharomyces cerevisiae* was prepared in house as previously described³². Briefly, the artificial
673 suppressor tRNA_{CUA} was transcribed *in vitro* and purified by phenol-chloroform extraction. Prior to
674 use, the suppressor tRNA_{CUA} was refolded in 100 mM HEPES-KOH pH 7.5, 10 mM KCl at 70°C
675 for 5 min and a final concentration of 5 mM MgCl₂ was added just before the reaction was placed
676 on ice. The refolded tRNA_{CUA} was then aminoacylated with [¹⁵N,¹³C]-glutamine (CortecNet) in a
677 standard aminoacylation reaction: 20 μM tRNA_{CUA}, 0.5 μM GLN4, 0.1 mM [¹⁵N, ¹³C]-Gln (or 0.5
678 mM of 4F-Gln) in 100 mM HEPES-KOH pH 7.5, 10 mM KCl, 20 mM MgCl₂, 1 mM DTT and
679 10 mM ATP³⁷. After incubation at 37°C for 1 hour GLN4 was removed by addition of glutathione
680 beads and loaded suppressor tRNA_{CUA} was precipitated with 300 mM sodium acetate pH 5.2 and
681 2.5 volumes of 96% EtOH at -80°C and stored as dry pellets at -20°C. Successful loading was
682 confirmed by urea-PAGE (6.5% acrylamide 19:1, 8 M urea, 100 mM sodium acetate pH 5.2).

683

684 **Standard cell-free expression conditions**

685 Lysate was prepared as previously described³³ and based on the *Escherichia coli* strain BL21 Star
686 (DE3)::RF1-CBD₃, a gift from Gottfried Otting (Australian National University, Canberra,
687 Australia)⁶⁸. Cell-free protein expression was performed in batch mode as described by Apponyi *et*
688 *al.*⁶⁹. The standard reaction mixture consisted of the following components: 55 mM HEPES-KOH

689 (pH 7.5), 1.2 mM ATP, 0.8 mM each of CTP, GTP and UTP, 1.7 mM DTT, 0.175 mg/mL *E. coli*
690 total tRNA mixture (from strain MRE600), 0.64 mM cAMP, 27.5 mM ammonium acetate, 68 μ M
691 1-5-formyl-5,6,7,8-tetrahydrofolic acid (folinic acid), 1 mM of each of the 20 amino acids, 80 mM
692 creatine phosphate (CP), 250 μ g/mL creatine kinase (CK), plasmid (16 μ g/mL) and 22.5% (v/v)
693 S30 extract. The concentrations of magnesium acetate (5-20 mM) and potassium glutamate (60-
694 200 mM) were adjusted for each new batch of S30 extract. A titration of both compounds was
695 performed to obtain the maximum yield.

696

697 **Preparation of NMR samples**

698 Samples for NMR studies were produced in cell-free at 5-15 mL scale and incubated at 23°C and
699 450 rpm in a thermomixer for 4 h. Uniformly labeled NMR samples were obtained by substituting
700 the standard amino acid mix with 3 mg/mL [^{15}N , ^{13}C]-labeled ISOGRO⁴⁰ (an algal extract lacking
701 four amino acids: Asn, Cys, Gln and Trp) and additionally supplying [^{15}N , ^{13}C]-labeled Asn, Cys
702 and Trp (1 mM each) and 4 mM Gln. H46 samples in which only certain amino acids were
703 selectively labeled (Ala and Lys; Gly, Ser and Arg; Leu and Glu; and Phe) were prepared by
704 substituting the respective amino acids for the [^{15}N , ^{13}C]-labeled ones. To enable the labeling of
705 glutamates, potassium glutamate was substituted with potassium acetate, which was optimized by
706 testing a range of concentrations³⁶. To produce site-specifically labeled samples, the standard
707 reaction mixture was slightly modified. Instead of adding 1 mM of each amino acid, proline and
708 glutamine were substituted by deuterated versions (Eurisotop) and used at 2 or 4 mM, respectively.
709 10 μ M of [^{15}N , ^{13}C]-Gln or 4F-Gln suppressor tRNA_{CUA} were added for suppressed samples. The
710 same procedures were used for the preparation of H66 samples.

711

712 **Expression of H16 in *E. coli***

713 *Escherichia coli* BL21 (DE3) transformed with H16 construct was grown in LB medium
714 supplemented with 50 μ g/mL kanamycin at 37°C under stirring. When an OD_{600nm} 0.7 was reached,
715 the culture was induced using 1 mM IPTG and grown for 24 hours at 23°C. The cell pellet was
716 collected by centrifugation at 5,000 x g for 15 minutes at 4°C and resuspended in 10 mL buffer A
717 (50 mM Tris, 1000 mM NaCl, pH 8.5), supplemented with cComplete EDTA free protease inhibitor
718 tablet (Roche), per 1 L of expression volume. Cells were lysed by sonication at 35% for 2 minutes
719 with on-off cycles and cell debris was pelleted by centrifugation at 20,000 xg for 30 minutes at 4°C.
720 *E. coli* H16 was used for SAXS experiments.

721

722 **Protein purification**

723 The cell-free reactions were diluted 5-10 fold with buffer A (50 mM Tris-HCl pH 7.5, 500 mM
724 NaCl, 5 mM imidazole) before incubating it 1 h with 1.5 mL of Ni-resin (cOmplete™ His-Tag
725 Purification Resin). The matrix was packed by gravity-flow and washed with buffer B (50 mM
726 Tris-HCl pH 7.5, 1000 mM NaCl, 5 mM imidazole) and the target protein was eluted with buffer C
727 (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 250 mM imidazole). For SAXS measurements, this
728 affinity chromatography step was carried out in an AKTA pure System (GE Healthcare) with a 5
729 mL Histrap® Excel column. Elution fractions were checked under UV light and fluorescent
730 fractions were pooled, protease inhibitors were added (cOmplete EDTA-free protease inhibitor
731 cocktail, Sigma Aldrich) and the sample was dialyzed against NMR buffer (20 mM BisTris-HCl
732 pH 6.5, 150 mM NaCl) at 4°C using SpectraPor 4 MWCO 12-14 kDa dialysis tubing (Spectrum
733 Labs). Dialyzed protein was then concentrated with 10 kDa MWCO Vivaspin centrifugal
734 concentrators (3,500 xg, 4°C) (Sartorius). Protein concentrations were determined by means of
735 fluorescence using an sfGFP calibration curve. Final NMR sample concentrations ranged from 4 to
736 15 µM. Protein integrity was analyzed by SDS-PAGE.

737 For both aggregation and SAXS experiments, an additional size-exclusion chromatography step,
738 using a Superdex S200 10/300 column, was carried out. For aggregation measurements, this step
739 was performed in aggregation buffer (50 mM sodium phosphate, pH 7.5, 150 mM NaCl) and for
740 SAXS measurements in NMR buffer.

741

742 **NMR experiments and data analysis**

743 All NMR samples contained final concentrations of 10% D₂O and 0.5 mM 4,4-dimethyl-4-
744 silapentane-1-sulfonic acid (DSS). ¹⁵N and ¹³C-HSQC experiments, in order to determine amide
745 (¹H_N and ¹⁵N) and aliphatic (¹H_{aliphatic} and ¹³C_{aliphatic}) chemical shifts, were performed at 293 K on a
746 Bruker Avance III spectrometer equipped with a cryogenic triple resonance probe and Z gradient
747 coil, operating at a ¹H frequency of 700 MHz or 800 MHz. Spectra acquisition parameters were set
748 up depending on the sample concentration and the magnet strength. All spectra were processed with
749 TopSpin v3.5 (Bruker Biospin) and analyzed using CCPN-Analysis software⁷⁰. Chemical shifts
750 were referenced with respect to the H₂O signal relative to DSS using the ¹H/X frequency ratio of the
751 zero point according to Markley *et al.*⁷¹.

752 Random coil chemical shifts were predicted using POTENCI, a pH, temperature and neighbor
753 corrected IDP library (<http://nmr.chem.rug.nl/potenci/>)³⁷. Secondary chemical shifts (SCS) were
754 obtained by subtracting the predicted value from the experimental one (SCS=δ_{exp}-δ_{pred}). For better
755 reliability of the results regarding possible referencing errors, we used the combined C_α and C_β
756 secondary chemical shifts (SCS(C_α)-SCS(C_β)).

757

758 **¹⁹F-NMR experiments**

759 All NMR samples were first concentrated up to a ca. 200 μ l volume using Vivaspin centrifugal
760 concentrators (Sartorius) with a 5 kDa cutoff at 4°C. 0.1 μ l of a trimethylsilylpropanoic acid
761 (TMSP) solution for chemical shift referencing and 10 μ l of D₂O were added before NMR
762 measurement. All ¹⁹F NMR experiments were performed on a Bruker Avance III HD spectrometer
763 operating at a ¹H and ¹⁹F frequencies of 600.13 MHz and 564.69 MHz, respectively, equipped with
764 a CP-QCI-F cryoprobe with ¹⁹F cryo-detection. All ¹⁹F 1D experiments were performed at 293.0 K
765 with ¹H decoupling during acquisition using waltz16 composite pulse decoupling. An acquisition
766 time of 0.58 s, spectral width of 100.6 s and relaxation delay of 1.0 s was used for all samples,
767 except for the sample fluorinated at Q20, where an acquisition time of 0.29 s and a relaxation delay
768 of 0.5 s was used. Concatenated 1D ¹⁹F spectra of 128 transients each were acquired in order to
769 monitor any spectral changes over time. Signal averaging was then performed up to a time point
770 before significant spectral changes over time could be detected. Final number of transients varied
771 were 77824, 49152, 39680 and 53760 for samples fluorinated at Q20, Q21, Q40 and Q60,
772 respectively. ¹⁹F spectra were referenced to the ¹H signal of TMSP using the unified chemical shift
773 scale.

774

775 **Model building and chemical shift ensemble optimization**

776 Ensemble models for the two families capturing the conformational influences of the flanking
777 regions, N \rightarrow C and N \leftarrow C, were constructed with the algorithm described in reference³⁸, which uses
778 a curated database of three-residue fragments extracted from high-resolution protein structures. The
779 model building strategy consecutively appends residues, which are considered to be either fully
780 disordered or partially structured. For fully disordered residues, amino acid specific ϕ/ψ angles
781 defining the residue conformation are randomly selected from the database, disregarding their
782 sequence context. For partially structured residues, the nature and the conformation of the flanking
783 residues are taken into account when selecting the conformation of the incorporated residue. Steric
784 clashes are tested at each step, and a backtracking strategy is applied to solve possible conflicts (see
785 detailed explanation of the algorithm in reference³⁸).

786 Two families of ensembles were built. For the first family (N \rightarrow C ensembles), starting with the
787 ¹⁰AFESLKS¹⁶ region of N17 as partially structured, multiple ensembles of 5,000 conformations
788 were built by successively including an increasing number of glutamines in the poly-Q tract (from
789 F17 to Q63) as partially structured, while the rest of the chain was considered to be fully disordered.
790 An equivalent strategy was followed for the second family of ensembles (N \leftarrow C ensembles) for
791 which glutamines were considered successively as partially structured from the poly-P tract (from
792 Q63 to Q18). Note that in the partially structured building strategy secondary structural elements

793 are propagated due to the conformational influence of neighboring residues. Two tripeptide
794 databases were used to generate the conformational ensemble models. Both were constructed from
795 the protein domains in the SCOP (Structural Classification of Proteins)^{72,73} repository filtered to
796 95% sequence identity. An “unfiltered” tripeptide database was built disregarding secondary
797 structure content, and “coil” database included tripeptides not participating in α -helices or β -strands.
798 For the N \rightarrow C ensembles, the best results were obtained when using the “unfiltered” and “coil”
799 databases to sample the partially-structured and the fully disordered sections, respectively. For the
800 N \leftarrow C ensembles, the “coil” database yielded the best results. For the resulting 47 ensembles of
801 each family, and after building the side chains with the program SCWRL4³⁹, averaged C α chemical
802 shifts were computed with SPARTA+⁴⁰, and used for the optimization. The optimized ensemble
803 model of H46 was built by reweighting the populations of the pre-computed ensembles, minimizing
804 the error with respect to the experimental C α CSs. In order to capture the influence of the flanking
805 regions, glutamines within the tract were divided into two groups: those influenced by N17 and
806 those influenced by the poly-P tract, whose chemical shifts were fitted with the N \rightarrow C and N \leftarrow C
807 ensembles, respectively. The limit between both families was systematically explored by computing
808 the agreement between the experimental and optimized CSs through a χ^2 value. An optimal
809 description of the complete CS profile was obtained when Q55 was chosen as the last residue
810 structurally connected with N17. Finally, an ensemble of 11,000 conformations was built using the
811 optimized weights and it was used to derive secondary structure population using SS-map⁴³ and to
812 analyze the SAXS data.

813

814 **SAXS data measurement and analysis**

815 The SAXS data for H16 were collected at the SWING beamline at the SOLEIL synchrotron,
816 France, equipped with an Eiger 4M detector with a sample-to-detector distance of 1.5 m⁷⁴. The data
817 for H46 were collected at EMBL-bioSAXS-P12 beamline at PETRAIII, Hamburg, Germany
818 equipped with a Pilatus 6M detector with a sample-to-detector distance of 3 m⁷⁵. The parameters
819 used for SAXS data collection are given in Table S2. All the data were collected in SEC-SAXS
820 mode with an in-line Superdex 200 Increase 10/300 GL column (GE Healthcare). Both proteins
821 were concentrated to 8 mg/mL and centrifuged at 20,000 x g immediately before injecting the
822 protein onto the column. 80 μ L of the sample were injected into the column and the flow rate was
823 maintained at 0.5 mL/min. The initial data processing steps including masking and azimuthal
824 averaging were performed using the program FOXTROT⁷⁶ for H16 and SASFLOW pipeline⁷⁵ for
825 H46. The resulting 1D profiles were analyzed using CHROMIXS⁷⁷ from ATSAS suite to select the
826 frames corresponding to sample and buffer and perform buffer subtraction. The final buffer
827 subtracted and averaged SAXS profiles were analyzed using ATSAS 2.8 software package⁷⁸,

828 including AUTORG for calculating the radius of gyration and calculation of extrapolated value of
829 radius of gyration (R_g), GNOM⁷⁹ for calculation of pairwise distance distribution profiles and
830 DATBAYES⁸⁰ for calculation of molecular weight by Bayesian estimate from four approaches. The
831 ensemble optimization approach (EOM) was used to select sub-ensembles that collectively describe
832 the SAXS data. The program RanCH was first used to join each of the conformations of H16 and
833 H46 (generated as described above) to the modelled structure of sfGFP and the hexahistidine tag
834 used for purification. GAJOE was then used to find a sub-ensemble from this pool which
835 collectively describes the SAXS data^{41,42}. The graphical representations were generated using the
836 program VMD⁸¹.

837

838 **Molecular dynamics simulations**

839 We performed Gaussian Accelerated Molecular Dynamics (GaMD)⁴⁸ simulations to explore the
840 conformational landscape and the secondary structure propensities of a fragment of httex1
841 consisting of N17, 46 glutamines and 5 prolines. We used ff03ws-STQ⁵⁰ force field, which is
842 adapted to proteins with low-complexity sequences (obtained from https://bitbucket.org/jeetain/all-atom_ff_refinements/src/master/). We chose an extended conformation built using the protocol
843 described earlier in the model building section as the starting structure and prepared the simulation
844 system using tools available with GROMACS 2020.5. This included addition of hydrogens,
845 solvation and addition of ions (Na^+ and Cl^-) to neutralize the system and set the final salt
846 concentration to 0.15 M. Thereafter, the system was converted to AMBER format using ParmEd
847 tool in AMBER20. At this stage, hydrogen mass repartitioning was also done to allow a time step of
848 4 fs. We used periodic boundary conditions and restrained the bonds containing hydrogen atoms
849 using SHAKE algorithm. Particle Mesh Ewald summation (PME) was used to calculate
850 electrostatic interaction with a cut-off of 9 Å on the long-range interactions. The system was energy
851 minimized for 5000 steps. This was followed by an NVT (constant number, volume and
852 temperature) equilibration for 5 ns and further equilibration in NPT (constant number, pressure and
853 temperature) for 10ns. For all the simulations, the temperature was maintained at 293 K and for
854 NPT simulations the pressure was maintained at 1 atm. This was followed by a GaMD equilibration
855 stage which consisted of a classical MD simulation of 40 ns during which the potential statistics to
856 calculate GaMD parameters were collected followed by a 120 ns long equilibration during which
857 the boost was added and updated. Finally, 8 independent simulations with an aggregate simulation
858 time of ~ 20 μs were launched using the boost parameters obtained in the equilibration stage. The
859 simulations were run in “dual-boost” mode and the reference energy was set to the lower bound
860 ($E=V_{\text{max}}$). The average and standard deviation of the potential energy were calculated every 2ns
861 (number of steps ≈ 4 *system size).

863

864 **Aggregation experiments**

865 Time-dependent aggregation of H46 variants was followed with SDS-PAGE analysis as previously
866 described, with minor modifications⁵². 15 μ M wild-type-, LKGG- and LLLF-H46 samples,
867 prepared in aggregation buffer (50 mM sodium phosphate, pH 7.5, 150 mM NaCl), were incubated
868 at 37°C for 48 h, without shaking. 10 μ L-aliquots were extracted at different time intervals,
869 immediately mixed with denaturing buffer (125 mM Tris-HCl, pH 6.8, 20% glycerol, 4% SDS, 200
870 mM dithiothreitol, 0.05% bromophenol blue), incubated for 10 min at 95°C, and frozen at -20 °C
871 until analysis on Bolt™ 4–12% Bis-Tris Plus gels (Invitrogen). The gels were washed in water,
872 stained with Instant Blue Coomassie Protein Stain (Abcam), and visualized using a Gel Doc™ Ez
873 Imager (Bio-Rad). The amount of SDS-soluble species trapped in the stacking gel was quantified
874 using the Image Lab 5.1 software. The percentages of soluble protein were referenced to time 0
875 values, and plotted against time. The plots were fitted using GraphPad Prism Software. For each
876 protein variant, at least two independent experiments, with three replicates at each time point, were
877 recorded.

878

879 **Atomic force microscopy (AFM) and total internal reflection fluorescence (TIRF)**

880 Correlative AFM-TIRF microscopy was developed in-house⁸². AFM images were acquired using a
881 Nanowizard 4 (JPK Instruments, Bruker) mounted on a Zeiss inverted optical microscope and
882 equipped with a Vortis-SPM control unit. A custom-made TIRF microscope was coupled to the
883 AFM using a LX 488-50 OBIS laser source (Coherent). We used an oil immersion objective with a
884 1.4 numerical aperture (Plan-Apochromat 100x, Zeiss). Fluorescence was collected with an
885 EmCCD iXon Ultra897 (Andor) camera. The setup includes a 1.5x telescope to obtain a final
886 imaging magnification of 150-fold, corresponding to a camera pixel size of 81.3 nm. An ET800sp
887 short pass filter (Chroma) was used in the emission optical path to filter out the light source of the
888 AFM optical beam deflection system. The excitation laser wavelength was centered at 488 nm and
889 the power was measured before the objective with a PM100 energy meter (purchased from
890 Thorlabs) and was optimized in all the experiments in the range of 1-5 μ W. Fluorescence images
891 were acquired using an ET525/50 nm (Chroma) emission filter and an acousto-optic tuneable filter
892 (AOTFnc-400.650-TN, AA opto-electronics) to modulate the laser intensity. Fluorescence images
893 were obtained by averaging 150 individual images, each acquired over 50 ms as exposure time.

894 AFM images were collected in liquid environment (Dulbecco's PBS named D-PBS) using the
895 quantitative-imaging (QI) mode. Each image was acquired with 256x256 lines/pixels and the
896 following scan size: 5 μ m \times 5 μ m, 2.5 μ m \times 2.5 μ m and 1 μ m \times 1 μ m. Typical force *versus* distance
897 curves were recorded with a tip approach speed ranging from 10 μ m/s to 30 μ m/s and an oscillation

898 amplitude (Z length) of 100 nm or 150 nm, adjusted depending on the height of the aggregates. The
899 maximal force exerted in each pixel was set to 100-150 pN and optimized during the image
900 acquisition. We used MSNL-D and MSNL-E (Bruker) AFM probes with resonances in liquid of ≈ 2
901 kHz and 10 kHz, respectively, and nominal spring constants of 0.03 N/m and 0.1 N/m. MSNL
902 cantilevers have a sharp tip radius (≈ 2 nm), which is ideal for high-resolution imaging. The inverse
903 optical lever sensitivity was calibrated with the acquisition of a force *versus* distance curve on the
904 glass coverslip whereas the cantilever stiffness was calibrated using thermal method⁸³.

905 Samples for correlative AFM and TIRF were prepared on circular glass coverslips (2.5 cm, 165 μ m
906 thick, purchased from Marienfeld). Coverslips were cleaned with a 15 min cycle of sonication with
907 ultrasounds in 1M KOH, rinsed 20 times with deionized water and finally with a second cycle of
908 sonication in deionized water. Fibrils were then deposited on the clean glass coverslips and let dry
909 before being immersed in D-PBS for AFM imaging.

910

911 **Acknowledgements**

912 The authors thank Gottfried Otting for providing the BL21 (DE3) Star::RF1-CBD3 strain. This
913 work was supported by the European Research Council under the European Union's H2020
914 Framework Programme (2014-2020) / ERC Grant agreement n° [648030] awarded to PB, ANR-17-
915 CE11-0022-01 awarded to NS, and Labex EpiGenMed, an « Investissements d'avenir » program
916 (ANR-10-LABX-12-01). The CBS is a member of France-BioImaging (FBI) and the French
917 Infrastructure for Integrated Structural Biology (FRISBI), 2 national infrastructures supported by
918 the French National Research Agency (ANR-10-INBS-04-01 and ANR-10-INBS-05, respectively).
919 AU is supported by a grant from the Fondation pour la Recherche Médicale (SPF20150934061). DS
920 acknowledge a grant from the Métropole Européenne de Lille (PUSHUP). Géraldine Levy,
921 Université de Lille, is thanked for help with sample preparation and the ¹⁹F-NMR experiments. This
922 work benefited from the HPC resources of the CALMIP supercomputing center under the
923 allocations 2016-P16032 and 2021-P21043. The 600 MHz spectrometer for ¹⁹F NMR
924 measurements is funded by the Nord Region Council, CNRS, Institut Pasteur de Lille, the European
925 Community (ERDF), the French Ministry of Research and the Université de Lille and by the CTRL
926 CPER cofunded by the European Union with the European Regional Development Fund (ERDF),
927 by the Hauts-de-France Regional Council (contract n°17003781), Métropole Européenne de Lille
928 (contract n°2016_ESR_05), and French State (contract n°2017-R3-CTRL-Phase1). The authors
929 thank the SWING beamline at SOLEIL synchrotron, Saint-Aubin, France (proposal 20181386) and
930 P12 beamline at PETRAIII, Hamburg, Germany for beamtime allocation to the project and
931 assistance during data collection.

932

933 **Author Contributions**

934 P.B. conceived the project. C.A.E.R, A.U., A.S., A.B., D.S., N.S. and P.B. designed experiments.

935 C.A.E.R, A.U., A.S., M.P., A.M., A.E., A.F., X.L.L., Z.D.S., L.C., A.T., F.A. and D.S. performed

936 experiments. R.E.S., P.E.M., A.B., J.C., N.S. and P.B. supervised experiments. C.A.E.R., A.U.,

937 A.S. and P.B. wrote the manuscript with the help of all the co-authors.

938

939 **Competing Interest**

940 The authors declare no competing interests

941

942 **Data Availability Statement**

943 The datasets generated during and/or analysed during the current study are available from the

944 corresponding author on reasonable request.

945

946

947 **References**

948

949 (1) Orr, H. T. Beyond the Qs in the Polyglutamine Diseases. *Genes Dev.* **2001**, *15* (8), 925–932.
950 <https://doi.org/10.1101/gad.888401>.

951 (2) Walker, F. O. Huntington's Disease. *Lancet* **2007**, *369* (9557), 218–228.
952 [https://doi.org/10.1016/S0140-6736\(07\)60111-1](https://doi.org/10.1016/S0140-6736(07)60111-1).

953 (3) Saudou, F.; Humbert, S. The Biology of Huntingtin. *Neuron* **2016**, *89* (5), 910–926.
954 <https://doi.org/10.1016/j.neuron.2016.02.003>.

955 (4) Kremer, B.; Goldberg, P.; Andrew, S. E.; Theilmann, J.; Telenius, H.; Zeisler, J.; Squitieri,
956 F.; Lin, B.; Bassett, A.; Almqvist, E.; Bird, T. D.; Hayden, M. R. A Worldwide Study of the
957 Huntington's Disease Mutation: The Sensitivity and Specificity of Measuring CAG Repeats.
958 *N. Engl. J. Med.* **1994**, *330* (20), 1401–1406.
959 <https://doi.org/10.1056/NEJM199405193302001>.

960 (5) Benn, C. L.; Landles, C.; Li, H.; Strand, A. D.; Woodman, B.; Sathasivam, K.; Li, S.-H.;
961 Ghazi-Noori, S.; Hockly, E.; Faruque, S. M. N. N.; Cha, J.-H. J.; Sharpe, P. T.; Olson, J. M.;
962 Li, X.-J.; Bates, G. P. Contribution of Nuclear and Extranuclear PolyQ to Neurological
963 Phenotypes in Mouse Models of Huntington's Disease. *Hum. Mol. Genet.* **2005**, *14* (20),
964 3065–3078. <https://doi.org/10.1093/hmg/ddi340>.

965 (6) Graham, R. K.; Deng, Y.; Slow, E. J.; Haigh, B.; Bissada, N.; Lu, G.; Pearson, J.; Shehadeh,
966 J.; Bertram, L.; Murphy, Z.; Warby, S. C.; Doty, C. N.; Roy, S.; Wellington, C. L.; Leavitt,
967 B. R.; Raymond, L. A.; Nicholson, D. W.; Hayden, M. R. Cleavage at the Caspase-6 Site Is
968 Required for Neuronal Dysfunction and Degeneration Due to Mutant Huntingtin. *Cell* **2006**,
969 *125* (6), 1179–1191. <https://doi.org/10.1016/J.CELL.2006.04.026>.

970 (7) Zuccato, C.; Valenza, M.; Cattaneo, E. Molecular Mechanisms and Potential Therapeutical
971 Targets in Huntington's Disease. *Physiol. Rev.* **2010**, *90* (3), 905–981.
972 <https://doi.org/10.1152/physrev.00041.2009>.

973 (8) Mangiarini, L.; Sathasivam, K.; Seller, M.; Cozens, B.; Harper, A.; Hetherington, C.;
974 Lawton, M.; Trotter, Y.; Lehrach, H.; Davies, S. W.; Bates, G. P. Exon 1 of the HD Gene
975 with an Expanded CAG Repeat Is Sufficient to Cause a Progressive Neurological Phenotype
976 in Transgenic Mice. *Cell* **1996**, *87* (3), 493–506. [https://doi.org/10.1016/S0092-
977 8674\(00\)81369-0](https://doi.org/10.1016/S0092-8674(00)81369-0).

978 (9) Pan, L.; Feigin, A. Huntington's Disease: New Frontiers in Therapeutics. *Curr. Neurol.*
979 *Neurosci. Rep.* **2021**, *21* (3), 10. <https://doi.org/10.1007/s11910-021-01093-3>.

980 (10) Davies, S. W.; Turmaine, M.; Cozens, B. A.; DiFiglia, M.; Sharp, A. H.; Ross, C. A.;
981 Scherzinger, E.; Wanker, E. E.; Mangiarini, L.; Bates, G. P. Formation of Neuronal
982 Intranuclear Inclusions Underlies the Neurological Dysfunction in Mice Transgenic for the
983 HD Mutation. *Cell* **1997**, *90* (3), 537–548. [https://doi.org/https://doi.org/10.1016/S0092-
984 8674\(00\)80513-9](https://doi.org/https://doi.org/10.1016/S0092-8674(00)80513-9).

- 985 (11) Miller, J.; Arrasate, M.; Shaby, B. A.; Mitra, S.; Masliah, E.; Finkbeiner, S. Quantitative
986 Relationships between Huntingtin Levels, Polyglutamine Length, Inclusion Body Formation,
987 and Neuronal Death Provide Novel Insight into Huntington's Disease Molecular
988 Pathogenesis. *J. Neurosci.* **2010**, *30* (31), 10541 LP – 10550.
989 <https://doi.org/10.1523/JNEUROSCI.0146-10.2010>.
- 990 (12) Ramdzan, Y. M.; Trubetskov, M. M.; Ormsby, A. R.; Newcombe, E. A.; Sui, X.; Tobin, M.
991 J.; Bongiovanni, M. N.; Gras, S. L.; Dewson, G.; Miller, J. M. L.; Finkbeiner, S.; Moily, N.
992 S.; Niclis, J.; Parish, C. L.; Purcell, A. W.; Baker, M. J.; Wilce, J. A.; Waris, S.; Stojanovski,
993 D.; Böcking, T.; Ang, C.-S.; Ascher, D. B.; Reid, G. E.; Hatters, D. M. Huntingtin Inclusions
994 Trigger Cellular Quiescence, Deactivate Apoptosis, and Lead to Delayed Necrosis. *Cell Rep.*
995 **2017**, *19* (5), 919–927. <https://doi.org/10.1016/j.celrep.2017.04.029>.
- 996 (13) Kim, Y. E.; Hosp, F.; Frottin, F.; Ge, H.; Mann, M.; Hayer-Hartl, M.; Hartl, F. U. Soluble
997 Oligomers of PolyQ-Expanded Huntingtin Target a Multiplicity of Key Cellular Factors.
998 *Mol. Cell* **2016**, *63* (6), 951–964.
999 <https://doi.org/https://doi.org/10.1016/j.molcel.2016.07.022>.
- 1000 (14) Arrasate, M.; Mitra, S.; Schweitzer, E. S.; Segal, M. R.; Finkbeiner, S. Inclusion Body
1001 Formation Reduces Levels of Mutant Huntingtin and the Risk of Neuronal Death. *Nature*
1002 **2004**, *431* (7010), 805–810. <https://doi.org/10.1038/nature02998>.
- 1003 (15) Wetzel, R. Exploding the Repeat Length Paradigm While Exploring Amyloid Toxicity in
1004 Huntington's Disease. *Acc. Chem. Res.* **2020**, *53* (10), 2347–2357.
1005 <https://doi.org/10.1021/acs.accounts.0c00450>.
- 1006 (16) Sahoo, B.; Arduini, I.; Drombosky, K. W.; Kodali, R.; Sanders, L. H.; Greenamyre, J. T.;
1007 Wetzel, R. Folding Landscape of Mutant Huntingtin Exon1: Diffusible Multimers,
1008 Oligomers and Fibrils, and No Detectable Monomer. *PLoS One* **2016**, *11* (6), e0155747.
1009 <https://doi.org/10.1371/journal.pone.0155747>.
- 1010 (17) Ceccon, A.; Tugarinov, V.; Ghirlando, R.; Clore, G. M. Abrogation of Prenucleation,
1011 Transient Oligomerization of the Huntingtin Exon 1 Protein by Human Profilin I. *Proc. Natl.*
1012 *Acad. Sci.* **2020**, *117* (11), 5844–5852. <https://doi.org/10.1073/PNAS.1922264117>.
- 1013 (18) Kotler, S. A.; Tugarinov, V.; Schmidt, T.; Ceccon, A.; Libich, D. S.; Ghirlando, R.;
1014 Schwieters, C. D.; Clore, G. M. Probing Initial Transient Oligomerization Events Facilitating
1015 Huntingtin Fibril Nucleation at Atomic Resolution by Relaxation-Based NMR. *Proc. Natl.*
1016 *Acad. Sci. U. S. A.* **2019**, *116* (9), 3562–3571. <https://doi.org/10.1073/pnas.1821216116>.
- 1017 (19) Thakur, A. K.; Jayaraman, M.; Mishra, R.; Thakur, M.; Chellgren, V. M.; L Byeon, I.-J.;
1018 Anjum, D. H.; Kodali, R.; Creamer, T. P.; Conway, J. F.; M Gronenborn, A.; Wetzel, R.
1019 Polyglutamine Disruption of the Huntingtin Exon 1 N Terminus Triggers a Complex
1020 Aggregation Mechanism. *Nat. Struct. & Mol. Biol.* **2009**, *16*, 380.
- 1021 (20) Feng, X.; Luo, S.; Lu, B. Conformation Polymorphism of Polyglutamine Proteins. *Trends*
1022 *Biochem. Sci.* **2018**, *43* (6), 424–435. <https://doi.org/10.1016/j.tibs.2018.03.002>.

- 1023 (21) Caron, N. S.; Desmond, C. R.; Xia, J.; Truant, R. Polyglutamine Domain Flexibility
1024 Mediates the Proximity between Flanking Sequences in Huntingtin. *Proc. Natl. Acad. Sci.*
1025 **2013**, *110* (36), 14610 LP – 14615.
- 1026 (22) Nucifora, L. G.; Burke, K. A.; Feng, X.; Arbez, N.; Zhu, S.; Miller, J.; Yang, G.; Ratovitski,
1027 T.; Delannoy, M.; Muchowski, P. J.; Finkbeiner, S.; Legleiter, J.; Ross, C. A.; Poirier, M. A.
1028 Identification of Novel Potentially Toxic Oligomers Formed in Vitro from Mammalian-
1029 Derived Expanded Huntingtin Exon-1 Protein. *J. Biol. Chem.* **2012**, *287* (19), 16017–16028.
1030 <https://doi.org/10.1074/jbc.M111.252577>.
- 1031 (23) Li, P.; Huey-Tubman, K. E.; Gao, T.; Li, X.; West, A. P.; Bennett, M. J.; Bjorkman, P. J. The
1032 Structure of a PolyQ–Anti-PolyQ Complex Reveals Binding According to a Linear Lattice
1033 Model. *Nat. Struct. Mol. Biol.* **2007**, *14* (5), 381–387. <https://doi.org/10.1038/nsmb1234>.
- 1034 (24) Owens, G. E.; New, D. M.; West, A. P.; Bjorkman, P. J. Anti-PolyQ Antibodies Recognize a
1035 Short PolyQ Stretch in Both Normal and Mutant Huntingtin Exon 1. *J. Mol. Biol.* **2015**, *427*
1036 (15), 2507–2519. <https://doi.org/10.1016/J.JMB.2015.05.023>.
- 1037 (25) Warner, J. B.; Ruff, K. M.; Tan, P. S.; Lemke, E. A.; Pappu, R. V.; Lashuel, H. A.
1038 Monomeric Huntingtin Exon 1 Has Similar Overall Structural Features for Wild-Type and
1039 Pathological Polyglutamine Lengths. *J. Am. Chem. Soc.* **2017**, *139* (41), 14456–14469.
1040 <https://doi.org/10.1021/jacs.7b06659>.
- 1041 (26) Newcombe, E. A.; Ruff, K. M.; Sethi, A.; Ormsby, A. R.; Ramdzan, Y. M.; Fox, A.; Purcell,
1042 A. W.; Gooley, P. R.; Pappu, R. V.; Hatters, D. M. Tadpole-like Conformations of
1043 Huntingtin Exon 1 Are Characterized by Conformational Heterogeneity That Persists
1044 Regardless of Polyglutamine Length. *J. Mol. Biol.* **2018**, *430* (10), 1442–1458.
1045 <https://doi.org/10.1016/j.jmb.2018.03.031>.
- 1046 (27) Bravo-Arredondo, J. M.; Kegulian, N. C.; Schmidt, T.; Pandey, N. K.; Situ, A. J.; Ulmer, T.
1047 S.; Langen, R. The Folding Equilibrium of Huntingtin Exon-1 Monomer Depends on Its
1048 Polyglutamine Tract. *J. Biol. Chem.* **2018**, jbc.RA118.004808.
1049 <https://doi.org/10.1074/jbc.RA118.004808>.
- 1050 (28) Mier, P.; Paladin, L.; Tamana, S.; Petrosian, S.; Hajdu-Soltész, B.; Urbanek, A.; Gruca, A.;
1051 Plewczynski, D.; Grynberg, M.; Bernadó, P.; Gáspári, Z.; Ouzounis, C. A.; Promponas, V. J.;
1052 Kajava, A. V.; Hancock, J. M.; Tosatto, S. C. E.; Dosztanyi, Z.; Andrade-Navarro, M. A.
1053 Disentangling the Complexity of Low Complexity Proteins. *Brief. Bioinform.* **2020**, *21* (2),
1054 458–472. <https://doi.org/10.1093/bib/bbz007>.
- 1055 (29) Baias, M.; Smith, P. E. S.; Shen, K.; Joachimiak, L. A.; Zerko, S.; Koźmiński, W.; Frydman,
1056 J.; Frydman, L. Structure and Dynamics of the Huntingtin Exon-1 N-Terminus: A Solution
1057 NMR Perspective. *J. Am. Chem. Soc.* **2017**, *139* (3), 1168–1176.
1058 <https://doi.org/10.1021/jacs.6b10893>.
- 1059 (30) Cecon, A.; Schmidt, T.; Tugarinov, V.; Kotler, S. A.; Schwieters, C. D.; Clore, G. M.
1060 Interaction of Huntingtin Exon-1 Peptides with Lipid-Based Micellar Nanoparticles Probed
1061 by Solution NMR and Q-Band Pulsed EPR. *J. Am. Chem. Soc.* **2018**, *140* (20), 6199–6202.
1062 <https://doi.org/10.1021/jacs.8b02619>.

- 1063 (31) Urbanek, A.; Elena-Real, C. A.; Popovic, M.; Morató, A.; Fournet, A.; Allemand, F.;
1064 Delbecq, S.; Sibille, N.; Bernadó, P. Site-Specific Isotopic Labeling (SSIL): Access to High-
1065 Resolution Structural and Dynamic Information in Low-Complexity Proteins. *ChemBioChem*
1066 **2020**, *21* (6), 769–775. <https://doi.org/10.1002/cbic.201900583>.
- 1067 (32) Urbanek, A.; Morató, A.; Allemand, F.; Delaforge, E.; Fournet, A.; Popovic, M.; Delbecq,
1068 S.; Sibille, N.; Bernadó, P. A General Strategy to Access Structural Information at Atomic
1069 Resolution in Polyglutamine Homorepeats. *Angew. Chem. Int. Ed. Engl.* **2018**, *57* (14),
1070 3598–3601. <https://doi.org/10.1002/anie.201711530>.
- 1071 (33) Urbanek, A.; Popovic, M.; Morató, A.; Estaña, A.; Elena-Real, C. A.; Mier, P.; Fournet, A.;
1072 Allemand, F.; Delbecq, S.; Andrade-Navarro, M. A.; Cortés, J.; Sibille, N.; Bernadó, P.
1073 Flanking Regions Determine the Structure of the Poly-Glutamine in Huntingtin through
1074 Mechanisms Common among Glutamine-Rich Human Proteins. *Structure* **2020**, *28* (7), 733-
1075 746.e5. <https://doi.org/https://doi.org/10.1016/j.str.2020.04.008>.
- 1076 (34) Shen, K.; Calamini, B.; Fauerbach, J. A.; Ma, B.; Shahmoradian, S. H.; Serrano Lachapel, I.
1077 L.; Chiu, W.; Lo, D. C.; Frydman, J. Control of the Structural Landscape and Neuronal
1078 Proteotoxicity of Mutant Huntingtin by Domains Flanking the PolyQ Tract. *Elife* **2016**, *5*,
1079 e18065. <https://doi.org/10.7554/eLife.18065>.
- 1080 (35) Bhattacharyya, A.; Thakur, A. K.; Chellgren, V. M.; Thiagarajan, G.; Williams, A. D.;
1081 Chellgren, B. W.; Creamer, T. P.; Wetzel, R. Oligoproline Effects on Polyglutamine
1082 Conformation and Aggregation. *J. Mol. Biol.* **2006**, *355* (3), 524–535.
1083 <https://doi.org/https://doi.org/10.1016/j.jmb.2005.10.053>.
- 1084 (36) Morató, A.; Elena-Real, C. A.; Popovic, M.; Fournet, A.; Zhang, K.; Allemand, F.; Sibille,
1085 N.; Urbanek, A.; Bernadó, P. Robust Cell-Free Expression of Sub-Pathological and
1086 Pathological Huntingtin Exon-1 for Nmr Studies. General Approaches for the Isotopic
1087 Labeling of Low-Complexity Proteins. *Biomolecules* **2020**, *10* (10).
1088 <https://doi.org/10.3390/biom10101458>.
- 1089 (37) Nielsen, J. T.; Mulder, F. A. A. POTENCI: Prediction of Temperature, Neighbor and PH-
1090 Corrected Chemical Shifts for Intrinsically Disordered Proteins. *J. Biomol. NMR* **2018**, *70*
1091 (3), 141–165. <https://doi.org/10.1007/s10858-018-0166-5>.
- 1092 (38) Estaña, A.; Sibille, N.; Delaforge, E.; Vaisset, M.; Cortés, J.; Bernadó, P. Realistic Ensemble
1093 Models of Intrinsically Disordered Proteins Using a Structure-Encoding Coil Database.
1094 *Structure* **2019**, *27* (2), 381–391. <https://doi.org/10.1016/j.str.2018.10.016>.
- 1095 (39) Krivov, G. G.; Shapovalov, M. V.; Dunbrack Jr., R. L. Improved Prediction of Protein Side-
1096 Chain Conformations with SCWRL4. *Proteins Struct. Funct. Bioinforma.* **2009**, *77* (4), 778–
1097 795. <https://doi.org/https://doi.org/10.1002/prot.22488>.
- 1098 (40) Shen, Y.; Bax, A. SPARTA+: A Modest Improvement in Empirical NMR Chemical Shift
1099 Prediction by Means of an Artificial Neural Network. *J. Biomol. NMR* **2010**, *48* (1), 13–22.
1100 <https://doi.org/10.1007/s10858-010-9433-9>.

- 1101 (41) Tria, G.; Mertens, H. D. T.; Kachala, M.; Svergun, D. I. Advanced Ensemble Modelling of
1102 Flexible Macromolecules Using X-Ray Solution Scattering. *IUCrJ* **2015**, *2* (Pt 2), 207–217.
1103 <https://doi.org/10.1107/S205225251500202X>.
- 1104 (42) Bernadó, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. Structural
1105 Characterization of Flexible Proteins Using Small-Angle X-Ray Scattering. *J. Am. Chem.*
1106 *Soc.* **2007**, *129* (17), 5656–5664. <https://doi.org/10.1021/ja069124n>.
- 1107 (43) Iglesias, J.; Sanchez-Martínez, M.; Crehuet, R. SS-Map: Visualizing Cooperative Secondary
1108 Structure Elements in Protein Ensembles. *Intrinsically Disord. Proteins* **2013**, *1* (1), e25323.
1109 <https://doi.org/10.4161/idp.25323>.
- 1110 (44) Escobedo, A.; Topal, B.; Kunze, M. B. A.; Aranda, J.; Chiesa, G.; Mungianu, D.; Bernardo-
1111 Seisedos, G.; Eftekhazadeh, B.; Gairí, M.; Pierattelli, R.; Felli, I. C.; Diercks, T.; Millet,
1112 O.; García, J.; Orozco, M.; Crehuet, R.; Lindorff-Larsen, K.; Salvatella, X. Side Chain to
1113 Main Chain Hydrogen Bonds Stabilize a Polyglutamine Helix in a Transcription Factor. *Nat.*
1114 *Commun.* **2019**, *10* (1), 2034. <https://doi.org/10.1038/s41467-019-09923-2>.
- 1115 (45) Qu, W.; Zha, Z.; Ploessl, K.; Lieberman, B. P.; Zhu, L.; Wise, D. R.; B. Thompson, C.;
1116 Kung, H. F. Synthesis of Optically Pure 4-Fluoro-Glutamines as Potential Metabolic Imaging
1117 Agents for Tumors. *J. Am. Chem. Soc.* **2011**, *133* (4), 1122–1133.
1118 <https://doi.org/10.1021/ja109203d>.
- 1119 (46) Gimenez, D.; Phelan, A.; Murphy, C. D.; Cobb, S. L. 19F NMR as a Tool in Chemical
1120 Biology. *Beilstein J. Org. Chem.* **2021**, *17*, 293–318.
- 1121 (47) Kitevski-LeBlanc, J. L.; Prosser, R. S. Current Applications of 19F NMR to Studies of
1122 Protein Structure and Dynamics. *Prog. Nucl. Magn. Reson. Spectrosc.* **2012**, *62*, 1–33.
1123 <https://doi.org/https://doi.org/10.1016/j.pnmrs.2011.06.003>.
- 1124 (48) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics:
1125 Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.*
1126 **2015**, *11* (8), 3584–3595. <https://doi.org/10.1021/acs.jctc.5b00436>.
- 1127 (49) Moldovean, S. N.; Chiş, V. Molecular Dynamics Simulations Applied to Structural and
1128 Dynamical Transitions of the Huntingtin Protein: A Review. *ACS Chem. Neurosci.* **2020**, *11*
1129 (2), 105–120. <https://doi.org/10.1021/acschemneuro.9b00561>.
- 1130 (50) Tang, W. S.; Fawzi, N. L.; Mittal, J. Refining All-Atom Protein Force Fields for Polar-Rich,
1131 Prion-like, Low-Complexity Intrinsically Disordered Proteins. *J. Phys. Chem. B* **2020**, *124*
1132 (43), 9505–9512. <https://doi.org/10.1021/acs.jpcc.0c07545>.
- 1133 (51) Wetzel, R. Physical Chemistry of Polyglutamine: Intriguing Tales of a Monotonous
1134 Sequence. *J. Mol. Biol.* **2012**, *421* (4–5), 466–490.
1135 <https://doi.org/10.1016/j.jmb.2012.01.030>.
- 1136 (52) Monsellier, E.; Redeker, V.; Ruiz-Arlandis, G.; Bousset, L.; Melki, R. Molecular Interaction
1137 between the Chaperone Hsc70 and the N-Terminal Flank of Huntingtin Exon 1 Modulates
1138 Aggregation. *J. Biol. Chem.* **2015**, *290* (5), 2560–2576.
1139 <https://doi.org/10.1074/jbc.M114.603332>.

- 1140 (53) Galaz-Montoya, J. G.; Shahmoradian, S. H.; Shen, K.; Frydman, J.; Chiu, W. Cryo-Electron
1141 Tomography Provides Topological Insights into Mutant Huntingtin Exon 1 and PolyQ
1142 Aggregates. *Commun. Biol.* **2021**, *4* (1), 849. <https://doi.org/10.1038/s42003-021-02360-2>.
- 1143 (54) Ruggeri, F. S.; Vieweg, S.; Cendrowska, U.; Longo, G.; Chiki, A.; Lashuel, H. A.; Dietler,
1144 G. Nanoscale Studies Link Amyloid Maturity with Polyglutamine Diseases Onset. *Sci. Rep.*
1145 **2016**, *6*, 31155. <https://doi.org/10.1038/srep31155>.
- 1146 (55) Mier, P.; Elena-Real, C.; Urbanek, A.; Bernadó, P.; Andrade-Navarro, M. A. The Importance
1147 of Definitions in the Study of PolyQ Regions: A Tale of Thresholds, Impurities and
1148 Sequence Context. *Comput. Struct. Biotechnol. J.* **2020**, *18*.
1149 <https://doi.org/10.1016/j.csbj.2020.01.012>.
- 1150 (56) Dehay, B.; Bertolotti, A. Critical Role of the Proline-Rich Region in Huntingtin for
1151 Aggregation and Cytotoxicity in Yeast*. *J. Biol. Chem.* **2006**, *281* (47), 35608–35615.
1152 <https://doi.org/https://doi.org/10.1074/jbc.M605558200>.
- 1153 (57) Cordeiro, T. N.; Herranz-Trillo, F.; Urbanek, A.; Estaña, A.; Cortés, J.; Sibille, N.; Bernadó,
1154 P. Small-Angle Scattering Studies of Intrinsically Disordered Proteins and Their Complexes.
1155 *Curr. Opin. Struct. Biol.* **2017**, *42*, 15–23.
1156 <https://doi.org/https://doi.org/10.1016/j.sbi.2016.10.011>.
- 1157 (58) Kang, H.; Vázquez, F. X.; Zhang, L.; Das, P.; Toledo-Sherman, L.; Luan, B.; Levitt, M.;
1158 Zhou, R. Emerging β -Sheet Rich Conformations in Supercompact Huntingtin Exon-1 Mutant
1159 Structures. *J. Am. Chem. Soc.* **2017**, *139* (26), 8820–8827.
1160 <https://doi.org/10.1021/jacs.7b00838>.
- 1161 (59) Drombosky, K. W.; Rode, S.; Kodali, R.; Jacob, T. C.; Palladino, M. J.; Wetzel, R.
1162 Mutational Analysis Implicates the Amyloid Fibril as the Toxic Entity in Huntington’s
1163 Disease. *Neurobiol. Dis.* **2018**, *120*, 126–138.
1164 <https://doi.org/https://doi.org/10.1016/j.nbd.2018.08.019>.
- 1165 (60) Sathasivam, K.; Lane, A.; Legleiter, J.; Warley, A.; Woodman, B.; Finkbeiner, S.; Paganetti,
1166 P.; Muchowski, P. J.; Wilson, S.; Bates, G. P. Identical Oligomeric and Fibrillar Structures
1167 Captured from the Brains of R6/2 and Knock-in Mouse Models of Huntington’s Disease.
1168 *Hum. Mol. Genet.* **2010**, *19* (1), 65–78. <https://doi.org/10.1093/hmg/ddp467>.
- 1169 (61) Gruber, A.; Hornburg, D.; Antonin, M.; Krahmer, N.; Collado, J.; Schaffer, M.; Zubaite, G.;
1170 Lüchtenborg, C.; Sachsenheimer, T.; Brügger, B.; Mann, M.; Baumeister, W.; Hartl, F. U.;
1171 Hipp, M. S.; Fernández-Busnadiego, R. Molecular and Structural Architecture of PolyQ
1172 Aggregates in Yeast. *Proc. Natl. Acad. Sci.* **2018**, *115* (15), E3446 LP-E3453.
1173 <https://doi.org/10.1073/pnas.1717978115>.
- 1174 (62) Mario Isas, J.; Pandey, N. K.; Xu, H.; Teranishi, K.; Okada, A. K.; Fultz, E. K.; Rawat, A.;
1175 Applebaum, A.; Meier, F.; Chen, J.; Langen, R.; Siemer, A. B. Huntingtin Fibrils with
1176 Different Toxicity, Structure, and Seeding Potential Can Be Interconverted. *Nat. Commun.*
1177 **2021**, *12* (1), 4272. <https://doi.org/10.1038/s41467-021-24411-2>.

- 1178 (63) Lin, H.-K.; Boatz, J. C.; Krabbendam, I. E.; Kodali, R.; Hou, Z.; Wetzel, R.; Dolga, A. M.;
1179 Poirier, M. A.; van der Wel, P. C. A. Fibril Polymorphism Affects Immobilized Non-
1180 Amyloid Flanking Domains of Huntingtin Exon1 Rather than Its Polyglutamine Core. *Nat.*
1181 *Commun.* **2017**, *8* (1), 15462. <https://doi.org/10.1038/ncomms15462>.
- 1182 (64) Hoop, C. L.; Lin, H.-K.; Kar, K.; Magyarfalvi, G.; Lamley, J. M.; Boatz, J. C.; Mandal, A.;
1183 Lewandowski, J. R.; Wetzel, R.; Wel, P. C. A. van der. Huntingtin Exon 1 Fibrils Feature an
1184 Interdigitated β -Hairpin-Based Polyglutamine Core. *Proc. Natl. Acad. Sci.* **2016**, *113* (6),
1185 1546–1551. <https://doi.org/10.1073/PNAS.1521933113>.
- 1186 (65) Fiumara, F.; Fioriti, L.; Kandel, E. R.; Hendrickson, W. A. Essential Role of Coiled Coils for
1187 Aggregation and Activity of Q/N-Rich Prions and PolyQ Proteins. *Cell* **2010**, *143* (7), 1121–
1188 1135. <https://doi.org/https://doi.org/10.1016/j.cell.2010.11.042>.
- 1189 (66) Jayaraman, M.; Mishra, R.; Kodali, R.; Thakur, A. K.; Koharudin, L. M. I.; Gronenborn, A.
1190 M.; Wetzel, R. Kinetically Competing Huntingtin Aggregation Pathways Control Amyloid
1191 Polymorphism and Properties. *Biochemistry* **2012**, *51* (13), 2706–2716.
1192 <https://doi.org/10.1021/bi3000929>.
- 1193 (67) Nekooki-Machida, Y.; Kurosawa, M.; Nukina, N.; Ito, K.; Oda, T.; Tanaka, M. Distinct
1194 Conformations of in Vitro and in Vivo Amyloids of Huntingtin-Exon1 Show Different
1195 Cytotoxicity. *Proc. Natl. Acad. Sci.* **2009**, *106* (24), 9679 LP – 9684.
1196 <https://doi.org/10.1073/pnas.0812083106>.
- 1197 (68) Loscha, K. V.; Herlt, A. J.; Qi, R.; Huber, T.; Ozawa, K.; Otting, G. Multiple-Site Labeling of
1198 Proteins with Unnatural Amino Acids. *Angew. Chem. Int. Ed. Engl.* **2012**, *51* (9), 2243–
1199 2246. <https://doi.org/10.1002/anie.201108275>.
- 1200 (69) Apponyi, M. A.; Ozawa, K.; Dixon, N. E.; Otting, G. Cell-Free Protein Synthesis for
1201 Analysis by NMR Spectroscopy. In *Structural Proteomics: High-Throughput Methods*;
1202 Kobe, B., Guss, M., Huber, T., Eds.; Humana Press: Totowa, NJ, 2008; pp 257–268.
1203 https://doi.org/10.1007/978-1-60327-058-8_16.
- 1204 (70) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E.
1205 L.; Markley, J. L.; Ionides, J.; Laue, E. D. The CCPN Data Model for NMR Spectroscopy:
1206 Development of a Software Pipeline. *Proteins* **2005**, *59* (4), 687–696.
1207 <https://doi.org/10.1002/prot.20449>.
- 1208 (71) Markley, J. L.; Bax, A.; Arata, Y.; Hilbers, C. W.; Kaptein, R.; Sykes, B. D.; Wright, P. E.;
1209 Wüthrich, K. Recommendations for the Presentation of NMR Structures of Proteins and
1210 Nucleic Acids. *J. Mol. Biol.* **1998**, *280* (5), 933–952. <https://doi.org/10.1006/jmbi.1998.1852>.
- 1211 (72) Andreeva, A.; Kulesha, E.; Gough, J.; Murzin, A. G. The SCOP Database in 2020: Expanded
1212 Classification of Representative Family and Superfamily Domains of Known Protein
1213 Structures. *Nucleic Acids Res.* **2020**, *48* (D1), D376–D382.
1214 <https://doi.org/10.1093/nar/gkz1064>.

- 1215 (73) Andreeva, A.; Howorth, D.; Chothia, C.; Kulesha, E.; Murzin, A. G. SCOP2 Prototype: A
1216 New Approach to Protein Structure Mining. *Nucleic Acids Res.* **2014**, *42* (D1), D310–D314.
1217 <https://doi.org/10.1093/nar/gkt2124>.
- 1218 (74) Thureau, A.; Roblin, P.; Pérez, J. BioSAXS on the SWING beamline at Synchrotron
1219 SOLEIL. *J. Appl. Cryst.* **2021**, *54*. <https://doi.org/10.1107/S1600576721008736>.
- 1220 (75) Blanchet, C. E.; Spilotos, A.; Schwemmer, F.; Graewert, M. A.; Kikhney, A.; Jeffries, C.
1221 M.; Franke, D.; Mark, D.; Zengerle, R.; Cipriani, F.; Fiedler, S.; Roessle, M.; Svergun, D. I.
1222 Versatile Sample Environments and Automation for Biological Solution X-Ray Scattering
1223 Experiments at the P12 Beamline (PETRA III, DESY). *J. Appl. Crystallogr.* **2015**, *48* (Pt 2),
1224 431–443. <https://doi.org/10.1107/S160057671500254X>.
- 1225 (76) Girardot, R.; Viguier, G.; Pérez, J.; Ounsy, M. FOXTROT: A Java-Based Application to
1226 Reduce and Analyse SAXS and WAXS Piles of 2D Data at Synchrotron SOLEIL. In
1227 *canSAS-VIII*; 2015.
- 1228 (77) Panjkovich, A.; Svergun, D. I. CHROMIXS: Automatic and Interactive Analysis of
1229 Chromatography-Coupled Small-Angle X-Ray Scattering Data. *Bioinformatics* **2018**, *34*
1230 (11), 1944–1946. <https://doi.org/10.1093/bioinformatics/btx846>.
- 1231 (78) Franke, D.; Petoukhov, M. V.; Konarev, P. V.; Panjkovich, A.; Tuukkanen, A.; Mertens, H.
1232 D. T.; Kikhney, A. G.; Hajizadeh, N. R.; Franklin, J. M.; Jeffries, C. M.; Svergun, D. I.
1233 ATSAS 2.8: A Comprehensive Data Analysis Suite for Small-Angle Scattering from
1234 Macromolecular Solutions. *J. Appl. Crystallogr.* **2017**, *50*, 1212–1225.
1235 <https://doi.org/10.1107/S1600576717007786>.
- 1236 (79) Svergun, D. I. Determination of the Regularization Parameter in Indirect-Transform Methods
1237 Using Perceptual Criteria. *J. Appl. Crystallogr.* **1992**, *25* (pt 4), 495–503.
1238 <https://doi.org/10.1107/S0021889892001663>.
- 1239 (80) Hajizadeh, N. R.; Franke, D.; Jeffries, C. M.; Svergun, D. I. Consensus Bayesian Assessment
1240 of Protein Molecular Mass from Solution X-Ray Scattering Data. *Sci. Rep.* **2018**, *8* (1).
1241 <https://doi.org/10.1038/s41598-018-25355-2>.
- 1242 (81) Humphrey, W.; Dalke, A.; Schulten, K. {VMD} -- {V}isual {M}olecular {D}ynamics. *J.*
1243 *Mol. Graph.* **1996**, *14*, 33–38.
- 1244 (82) Dahmane, S.; Doucet, C.; Le Gall, A.; Chamontin, C.; Dosset, P.; Murcy, F.; Fernandez, L.;
1245 Salas, D.; Rubinstein, E.; Mougél, M.; Nollmann, M.; Milhiet, P.-E. Nanoscale Organization
1246 of Tetraspanins during HIV-1 Budding by Correlative DSTORM/AFM. *Nanoscale* **2019**, *11*
1247 (13), 6036–6044. <https://doi.org/10.1039/C8NR07269H>.
- 1248 (83) Proksch, R.; Schäffer, T. E.; Cleveland, J. P.; Callahan, R. C.; Viani, M. B. Finite Optical
1249 Spot Size and Position Corrections in Thermal Spring Constant Calibration. *Nanotechnology*
1250 **2004**, *15* (9), 1344–1350. <https://doi.org/10.1088/0957-4484/15/9/039>.
1251
1252