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, Christine Doucet, Xamuel L Lund, Zhen-Dan Shi, et al.

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

Carlos A. Elena-Real[1], Amin Sagar[1], Annika Urbanek[1], Matija Popovic[1], Anna Morató[1], Alejandro Estaña[1,2], Aurélie Fournet[1], Christine Doucet[1], Xamuel L. Lund[1,3], Zhen-Dan Shi[4], Luca Costa[1], Aurélien Thureau[5], Frédéric Allemand[1], Rolf E. Swenson[4], Pierre-Emmanuel Milhiet[1], Ramon Crehuet[6], Alessandro Barducci[1], Juan Cortés[2], Davy Sinnaeve[7,8], Nathalie Sibille[1], Pau Bernadó[1,*]

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

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

3- Institut Laue Langevin, 38000 Grenoble, France.

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

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

6- Institute for Advanced Chemistry of Catalonia (IQAC), CSIC, E-08034 Barcelona, Spain.

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

8- CNRS, EMR9002 - Integrative Structural Biology, 59000 Lille, France.

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

Cite this article

Elena-Real, C.A., Sagar, A., Urbanek, A. *et al.* The structure of pathogenic huntingtin exon 1 defines the bases of its aggregation propensity. *Nat Struct Mol Biol* **30**, 309–320 (2023). https://doi.org/10.1038/s41594-023-00920-0

Abstract

Huntington's Disease is a neurodegenerative disorder caused by a CAG expansion in the first exon of the *HTT* gene, resulting in an extended poly-glutamine (poly-Q) tract in huntingtin (httex1). The structural changes occurring to the poly-Q when increasing its length remain poorly understood due to its intrinsic flexibility and the strong compositional bias. The systematic application of site-specific isotopic labeling has enabled residue-specific NMR investigations of the poly-Q tract of pathogenic httex1 variants with 46 and 66 consecutive glutamines. Integrative data analysis reveals that the poly-Q tract adopts long α-helical conformations propagated and stabilized by glutamine side chain to backbone hydrogen bonds. We show that α-helical stability is a stronger signature in defining aggregation kinetics and the structure of the resulting fibrils than the number of glutamines. Our observations provide a structural perspective of the pathogenicity of expanded httex1 and pave the way to a deeper understanding of poly-Q-related diseases.

Introduction

Among the nine neurodegenerative disorders caused by expansions of polyglutamine (poly-Q) tracts, Huntington's Disease (HD) stands out due to its prevalence and devastating effects[1]. HD is triggered by an abnormal expansion of the poly-Q tract located in exon1 (httex1) of the 348-kDa huntingtin, a ubiquitous protein involved in multiple pathways[2,3]. In its non-pathogenic form, the httex1 poly-Q tract is comprised of 17-20 glutamines[4]; however, when the number of consecutive glutamines exceeds the pathogenic threshold of 35, it results in aggregation-prone mutants. Indeed, fragments of mutant httex1 are found forming large cytoplasmic and nuclear aggregates within neurons of the striatum, a well-known hallmark of HD[5,6]. The presence of such aggregates, the neuronal degeneration, the age of onset and disease severity correlate with the length of the poly-Q tract[7]. Notably, the mutant httex1 fragment alone suffices to reproduce HD symptoms in mice[8].

Two models have been suggested to connect the pathological threshold and toxicity, the 'toxic structure' and the 'linear lattice' models[20]. While the 'toxic structure' model proposes the appearance of a distinct toxic conformation when the poly-Q tract is expanded beyond the pathological threshold[21,22], the 'linear lattice' model suggests that poly-Q tracts are inherently toxic and that their toxicity systematically increases with the homorepeat length[23,24]. Intriguingly, both models have been supported by antibody recognition experiments in different studies. Furthermore, the absence of sharp changes in single-molecule Förster resonance energy transfer (smFRET), circular dichroism (CD) and electron paramagnetic resonance (EPR) experiments around the pathological threshold have been argued to substantiate the 'linear lattice' model[25–27]. High-resolution structures of non-pathogenic and pathogenic httex1 variants are required to evaluate the changes occurring upon poly-Q expansion, discriminate between both models and define the bases of httex1 cytotoxicity.

Until recently, the detailed high-resolution structural characterization of soluble httex1, especially of those with pathogenic poly-Q lengths, has been hampered by the intrinsic properties of the protein, namely the highly flexible nature and the strong compositional bias. While X-ray diffraction or electron microscopy cannot be applied to probe disordered proteins such as httex1, the presence of low complexity regions results in important signal overlap when nuclear magnetic resonance (NMR) is used[28]. Despite these difficulties, several NMR studies of non-pathogenic httex1 and N-terminal fragments have been reported, but only assignments of the first and last glutamines of the poly-Q could be achieved[19,26,27,29,30].

Our lab recently developed a site-specific isotopic labeling (SSIL) strategy that combines cell-free protein expression and non-sense suppression, enabling the investigation of homorepeats in a residue-specific manner[31,32]. By applying this methodology to a non-pathogenic version of httex1

containing a 16-residue-long poly-Q tract (H16), it was shown that the protein was enriched in helical conformations whose length and stability were defined by flanking regions[33]. While the helical propensity was propagated from N17 to the poly-Q through a hydrogen bond network, it was blocked by the helix-breaking effect caused by the proline-rich region (PRR) that follows the poly-Q tract (Fig. 1a). The structural effects imposed by N17 and PRR may explain the positive and negative regulation of httex1 aggregation by both poly-Q flanking regions, respectively[34,35]. Whether these structural mechanisms also govern the conformational landscape of pathogenic httex1 remains to be discerned.

Here, we applied SSIL to a pathogenic httex1 containing a 46 residue-long poly-Q tract (H46) to unambiguously assign 16 of these glutamines spread along the tract and probed the structure and dynamics of the homorepeat. The integration of the NMR and small-angle X-ray scattering (SAXS) data provided the structural description of H46 as an ensemble of elongated, partially helical conformations, whose propagation and stability mechanisms were deciphered by ^{19}F-NMR and molecular dynamics (MD) simulations. Altogether, our observations provide a detailed structural perspective of the 'linear lattice' model, demonstrating that the presence of long, persistent, aggregation-prone α-helices is concomitant to the expansion of the poly-Q tract and strongly influences aggregation kinetics and the structure of the resulting fibrils.

Results

Pathogenic and non-pathogenic httex1 are structurally similar

In order to study a pathogenic form of httex1, a construct comprising the N17 domain, a 46-residue-long poly-Q tract and the PRR was fused to superfolder GFP (sfGFP) (Fig. 1a). A fully ^{15}N-labeled H46 sample was prepared and a ^{15}N-HSQC spectrum was recorded. Similarly to H16[33], which also contained a fused sfGFP, the spectrum revealed that, while peaks from N17 and the PRR were well dispersed, glutamine frequencies remained poorly resolved (Fig. 1b).

The reduced stability of H46 precluded the use of traditional 3D-NMR experiments. To confirm the similarities found between the pathogenic and non-pathogenic httex1, selectively labeled samples of H46 were prepared (^{15}N-Ala and ^{15}N-Lys; ^{15}N-Gly, ^{15}N-Ser and ^{15}N-Arg; ^{15}N-Leu and ^{15}N-Glu; and ^{15}N-Phe) (Extended Data Fig. 1a). The selective labeling showed that the vast majority of peaks corresponding to N17 and PRR residues nicely overlap for H46 and H16. Interestingly, F17 peaks in H16 and H46 displayed different chemical shifts (CS). Altogether, our observations suggest that both httex1 forms share similar structural features, although perturbations in the N17/poly-Q boundary are induced upon expansion.

The SSIL approach was used to specifically study 16 of the 46 glutamines of the H46 poly-Q tract, as well as two glutamines within the PRR[36] (Fig. 1a and Supplementary Table 1). ^{15}N-HSQC spectra of these samples revealed that the glutamines adjacent to N17 (Q18, Q20 and Q21) adopted the lowest CS values of the poly-Q without any specific trend, while the following glutamines (Q24-Q56) exhibited steadily increasing ^{1}H and ^{15}N CSs, pointing towards gradual structural changes along the homorepeat (Fig. 1c). The last glutamines of the tract (Q61-Q63) were found at the highest CSs, with Q63, displaying an isolated peak like Q75 and Q91 in the PRR. The same features were observed when monitoring the Cα-Hα correlations of the same SSIL samples (Fig. 1c). Notably, Q21 presented two Cα peaks with similar intensity, suggesting the presence of two slowly interconverting conformations.

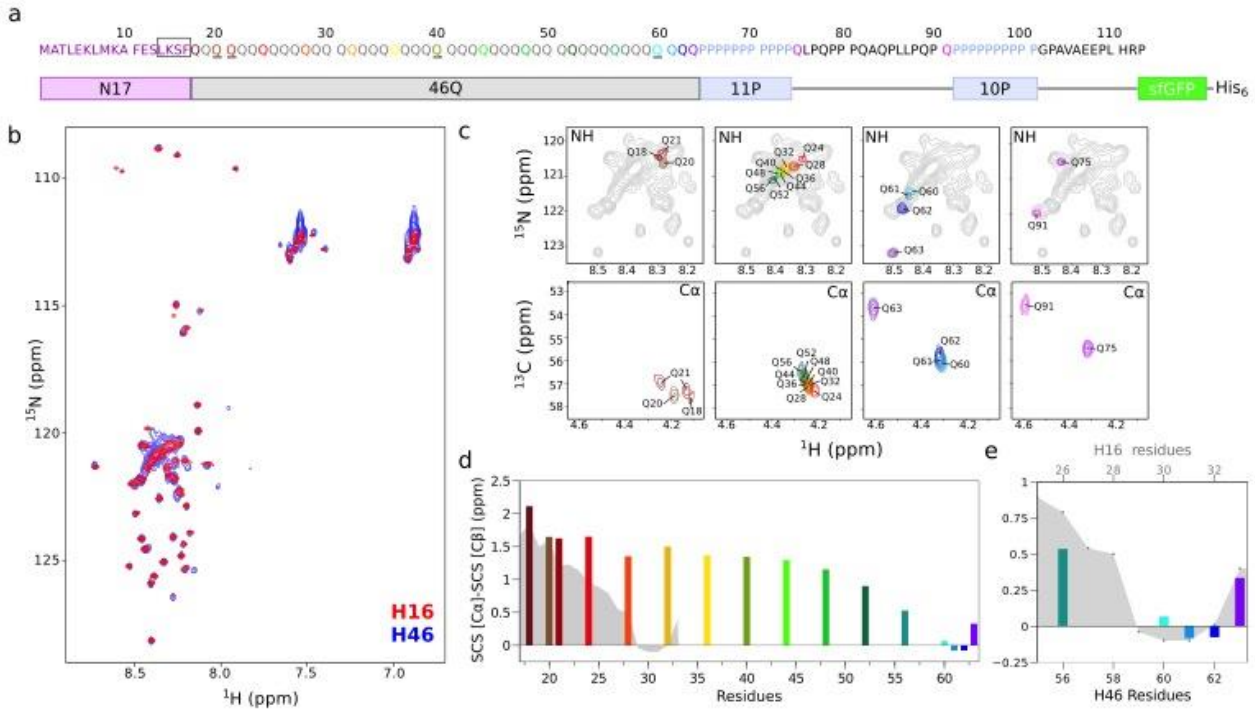

Figure 1. NMR analyses of H46 and comparison with H16. (**a**) Primary sequence of H46 and scheme of the sfGFP-fused construct used in this study. The color code identifies the site-specifically labeled glutamines throughout the study. Underlined glutamines indicate the positions in which 2S,4R-fluoroglutamine (4F-Gln) was introduced for ^{19}F-NMR experiments. The box indicates the section mutated in the LKGG-H46 and LLLF-H46 mutants. (**b**) Overlay of the ^{15}N-HSQC spectra of fully labeled H46 (blue) with the previously reported for H16 (red)[33]. (**c**) Zoom of the ^{15}N-HSQC (upper panels) and ^{13}C-HSQC (lower panels) with individually colored SSIL spectra showing the poly-Q NH and Cα regions for different glutamine clusters (Q18-Q21; Q24-Q56; Q60-Q63; and PRR glutamines). In the upper panels, the ^{15}N-HSQC of the fully labeled ^{15}N-H46 sample is shown in gray. (**d**) Secondary chemical shift (SCS) analysis of H46 poly-Q using experimental Cα and Cβ chemical shifts and a neighbor-corrected random-coil library[37]. The SCS analysis of H16 is shown in gray[33]. (**e**) Comparison of the SCS values of glutamines flanking the PRR in H16 (gray area) and H46 (colored bars) when aligning both sequences from the C-terminus of the poly-Q tract.

The secondary chemical shift (SCS) analysis of the glutamines using a neighbor-corrected random coil database[37] showed that the poly-Q was highly enriched in α-helical conformations, in line with previous observations for non-pathological constructs[26,29,33] (Fig. 1d). However, this propensity was not homogeneous; the helicity reached its maximum at Q18 and presented a steady helical plateau over 30 residues, until around Q48. In the C-terminal part of the poly-Q, a smooth decrease of helicity was observed, reaching negative SCS values for the last glutamines. When comparing this SCS analysis with that for H16[33], an increase in helicity for the N-terminal part of the H46 poly-Q tract was observed. This phenomenon also explains the shift of the NH, $N\varepsilon H_2$ and CαHα NMR signals of Q18 and Q21 towards more helical positions in H46 (Extended Data Fig. 1b). Interestingly, the two Cα peaks of Q21 presented SCS values of 1.27 and 1.63 ppm, corresponding to two conformations with different helical content.

The extent of the structural effects induced by the PRR was analyzed by aligning the SCS values of H16 and H46 from the C-terminus (Fig. 1e). While the last four glutamines of the tract (Q30-Q33 in H16 and Q60-Q63 in H46) showed the same conformational trends, this similarity was reduced for more distant glutamines (Q56).

H66 substantiates the persistence of long α-helices

The above results suggest that the distance to the PRR is key in defining the length of α-helical conformations. In order to validate this hypothesis, we studied an httex1 construct with 66 consecutive glutamines (H66), whose ^{15}N-HSQC nicely overlapped with that of H46 (Extended Data Fig. 2a). We applied the SSIL strategy to two glutamines (Q56 and Q76) of H66 to probe their helical content. Note that these two residues are located at equivalent positions to residue Q56 in H46 when aligning both sequences from the N- and C-termini, respectively (Fig. 2a). NH and Cα-Hα correlations of H66-Q76 appeared in equivalent positions to these of H46-Q56, but not H66-Q56 (Fig. 2b,c). The helical propensity of these residues was quantified with the Cα SCS analysis, indicating an enhanced helicity for H66-Q56 (1.18 ppm) when compared with H66-Q76 (0.50 ppm) and H46-Q56 (0.53 ppm) (Fig. 2d). Interestingly, the SCS value for H66-Q56 was very similar to those observed for the plateau in H46, indicating that the additional twenty glutamines in H66 adopt helical conformations (Extended Data Fig. 2b). These results evidenced that the extent of the α-helix-breaking capacity of the PRR is the same in all httex1 forms, allowing the α-helical propensity to be propagated through a larger number of glutamines when the length of the homorepeat is increased.

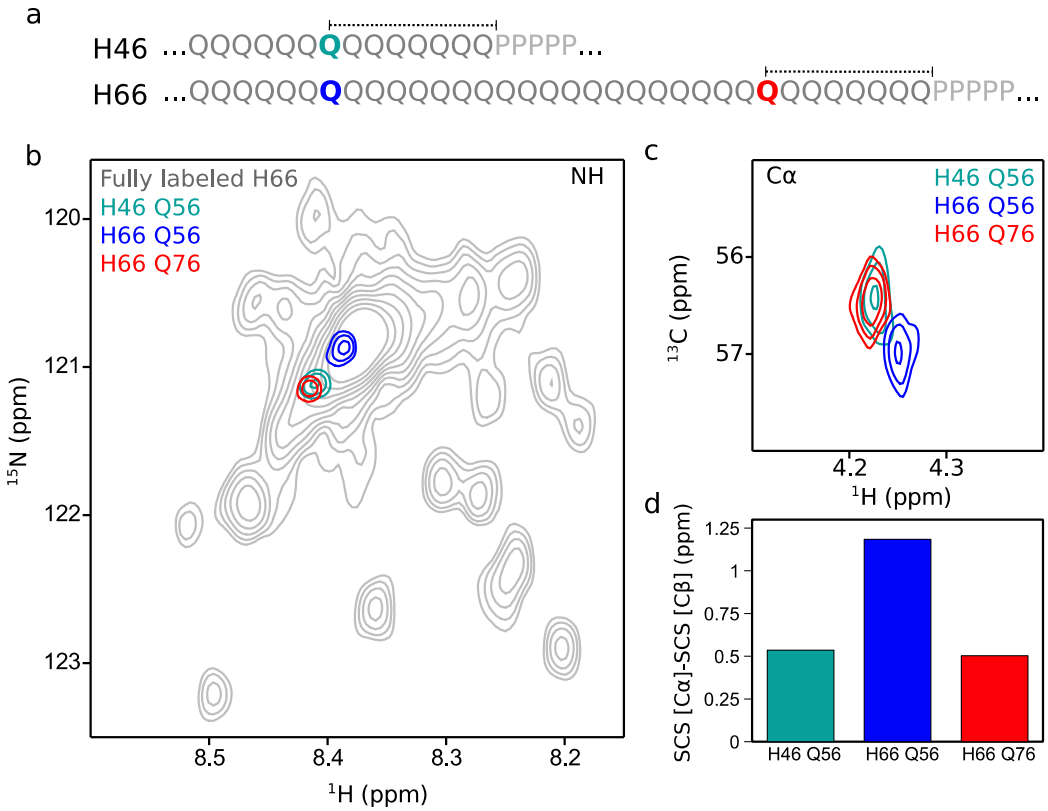

Figure 2. NMR analyses of H66 and comparison with H46. (a) Partial sequences of H46 and H66 indicating the positions of Q56 in H66 (blue) and H46 (green), and Q76 in H66 (red). This color code identifies the individual glutamines throughout the figure. Dashed lines highlight the equal distance of Q56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

H46 is a flexible elongated particle in solution

Size-exclusion chromatography coupled small angle X-ray scattering (SEC-SAXS) was applied to derive the overall size of H46 and H16 in solution (Extended Data Fig. 3). The analysis of the resulting profiles indicated that both constructs were flexible monomeric particles, exhibiting an increase in size with the length of the poly-Q (Fig. 3a and Supplementary Table 2). A detailed analysis is provided in Supplemental Note 1.

H46 consists of a mixture of α-helical conformations

The Cα CSs and the SAXS data measured for H46 were integrated to derive a structural model of the protein. Note that the derivation of an ensemble model for a disordered system is an ill-posed problem[38], which is further enhanced for httex1 given the sparseness of the NMR data and the low resolution of SAXS. In order to reduce the degeneracy of the resulting models, we applied a sampling strategy that captured the conformational influence of both flanking regions and biased

the starting ensemble with our previous knowledge of httex1 that was subsequently refined with the Cα CSs[39] (see Methods for details and Fig. S1). After modeling the sfGFP and the C-terminal His-tag to the individual conformations, the NMR-optimized ensemble was further refined by integrating the SAXS data with the ensemble optimization method (EOM)[42,43]. Sub-ensembles selected with EOM yielded an excellent fit to the experimental profile (χ^2=0.2) (Fig. 3a). The resulting R_g distribution was broad, indicating that H46 is a highly flexible particle in solution and slightly more extended than the CS-derived ensemble (Extended Data Fig. 3).

The structural analysis of the resulting ensemble[44], substantiated the presence of a mixture of multiple helical conformations encompassing different sections of the H46 poly-Q tract (Fig. 3c). These α-helices originated at the last residues of the N17 domain (^{14}LKSF^{17}) and propagated along the tract. Interestingly, an enrichment of long α-helices encompassing around 40 residues and reaching up to Q52 was observed. The presence of these long stable helical conformations explains the steady plateau observed in the SCS analysis (Fig. 1d).

A similar structural refinement was performed for H16, using the previously reported NMR-refined ensemble[33] and the SAXS data. The EOM fit yielded an excellent agreement with the experimental curve (χ^2=1.12). We observed that H16 also consisted of a mixture of α-helical structures of different lengths, with prevalence for those encompassing a large fraction of the homorepeat (Fig. 3d).

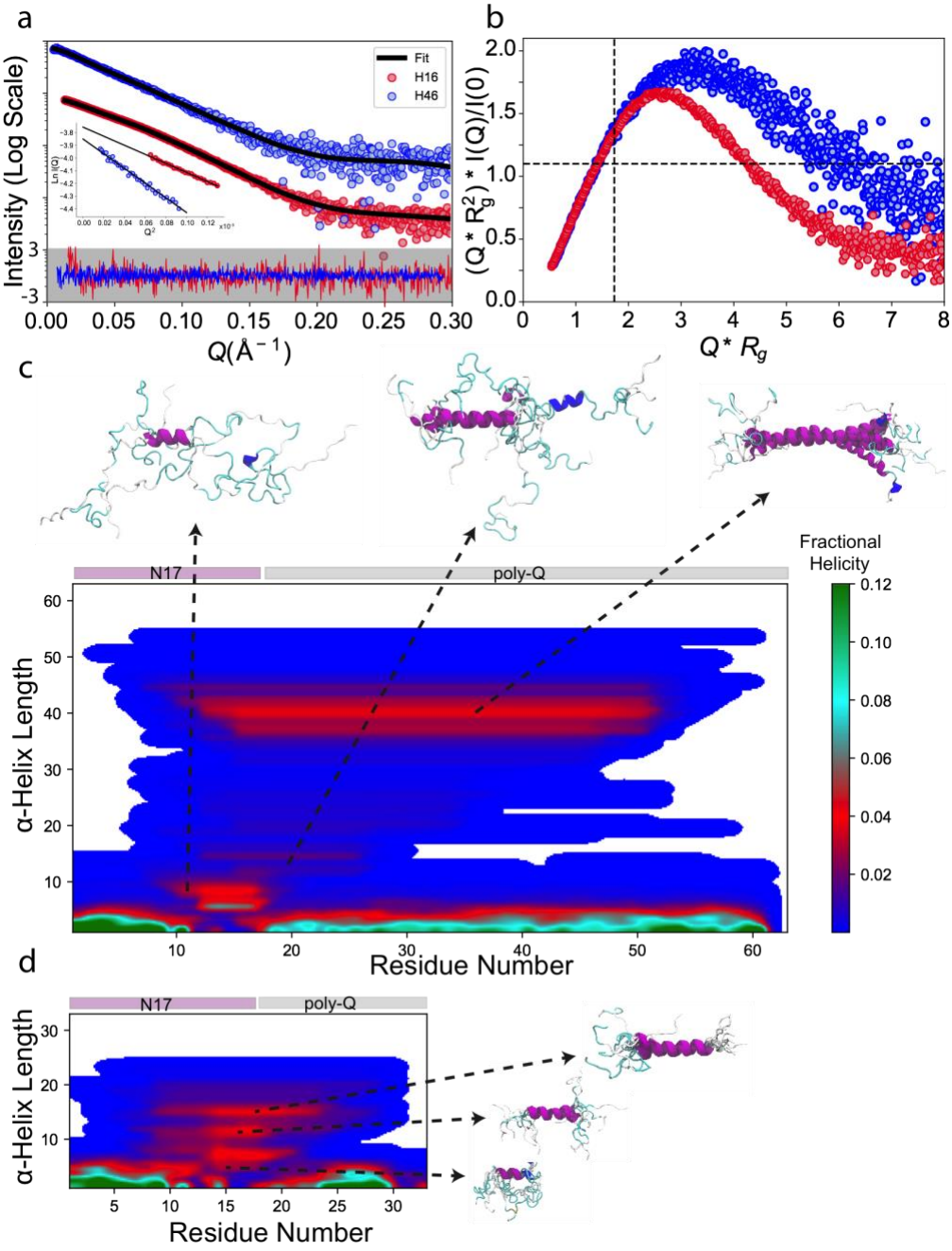

Figure 3. A structural model of pathogenic and non-pathogenic httex1 from the synergistic integration of NMR and SAXS data. (a) The SAXS intensity profiles for H16 (red) and H46 (blue) along with the theoretical profiles of EOM selected sub-ensembles (black lines). The residuals from EOM fitting are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

Bifurcated hydrogen bonds trigger and stabilize α-helices

It has been shown in non-pathogenic H16 and the androgen receptor that a network of hydrogen bonds involving glutamine backbone and side chains, the so-called bifurcated hydrogen bonds, is at the origin of the helical propagation from N-flanking regions to the poly-Q[33,45]. Here, we investigated whether this effect is conserved in pathogenic httex1 by monitoring the unambiguously assigned Cβ, Hβ, Cγ, Hγ and NHε CSs (Fig. 4a, Supplementary Table 1 and Extended Data Fig. 4). The diastereotopic Hβ glutamine protons displayed resolved responses, with increasing CS difference along the sequence. The maximum was observed for the last glutamines of the tract (Q62 and Q63) and those within the PRR (Q75 and Q91), while Q18 displayed the smallest one with virtually degenerate Hβ CSs, indicating a correlation between the CS difference and disorder. Interestingly, Q21 exhibited three peaks that we attributed to the equilibrium between two conformations in slow exchange. One of them displayed degenerate Hβ, as in Q18, while the other exhibited two Hβ frequencies. This observation is in agreement with the two Cα peaks observed for Q21 (Fig. 1c).

Hγ signals followed an inverse trend than Hβ ones. The last glutamines of the tract and those located in the PRR presented a single correlation, corresponding to degenerate diastereotopic Hγ CSs, as expected for a flexible glutamine side chain (Fig. 4a and Extended Data Fig. 4b). Conversely, the majority of glutamines of the homorepeat (up to Q48) displayed a measurable difference between the Hγ CSs, suggesting a transient rigidification of the side chain that slows down conformational averaging. Again, Q18, Q20 and Q21 were exceptions to this behavior. While Q18 exhibited a single Hγ peak, Q21 displayed three, substantiating the equilibrium between the two previously eluded conformational states. Although Q20 only exhibited two Hγ peaks, their difference in intensity suggested a similar scenario as for Q21. In addition to demonstrating the presence of $i \rightarrow i+4$ bifurcated hydrogen bonds structurally connecting the first residues of the poly-Q tract with N17[33,45] (Fig. 4b), our data indicate that these hydrogen bonds are present, although to a lower extent, along the homorepeat, incorporating an additional mechanism for structural stabilization.

Fluoro-glutamine, a new probe for protein structural studies

To exploit the power of ^{19}F as structural probe, we synthesized with high yield and stereospecificity 2S,4R-fluoroglutamine (4F-Gln), in which a fluorine atom replaced a hydrogen atom on Cγ (Fig. S2)[46]. 4F-Gln was successfully loaded onto the $tRNA_{CUA}$ using the yeast glutaminyl-tRNA synthetase with similar yields as for canonical glutamine[31] (Extended Data Fig. 5) and site-specifically incorporated at different positions of H46 poly-Q (Fig. 1a).

4F-Gln was used in two sets of experiments. First, we used 4F-Gln to substantiate the participation of Q20 and Q21 in the hydrogen bond network that propagates helicity in the homorepeat. To this end, we incorporated 4F-Gln in positions Q20 or Q21 in two H46 samples that were also isotopically labeled with ^{15}N-Ser or ^{15}N-Phe, respectively (Fig. 4b,c). When comparing their ^{15}N-HSQC spectra with those of non-fluorinated H46, substantial changes could be observed in S16 and F17, respectively. The presence of a fluorinated glutamine in position 20 produced a slight CS change of S16, while fluorination of Q21 induced appearance of a second peak on F17. Importantly, similar observations were made when 4F-Gln was incorporated in the same positions in H16, in samples simultaneously labeled with ^{15}N-Ser and ^{15}N-Phe (Fig. 4c and S3). Indeed, when Q20 was fluorinated in H16, both S16 and F17 were perturbed. Fluorination of Q21 resulted in the appearance of a second F17 signal and slightly affected S16. These data underline the proximity between Q20 and S16 as well as Q21 and F17 in pathogenic and non-pathogenic httex1.

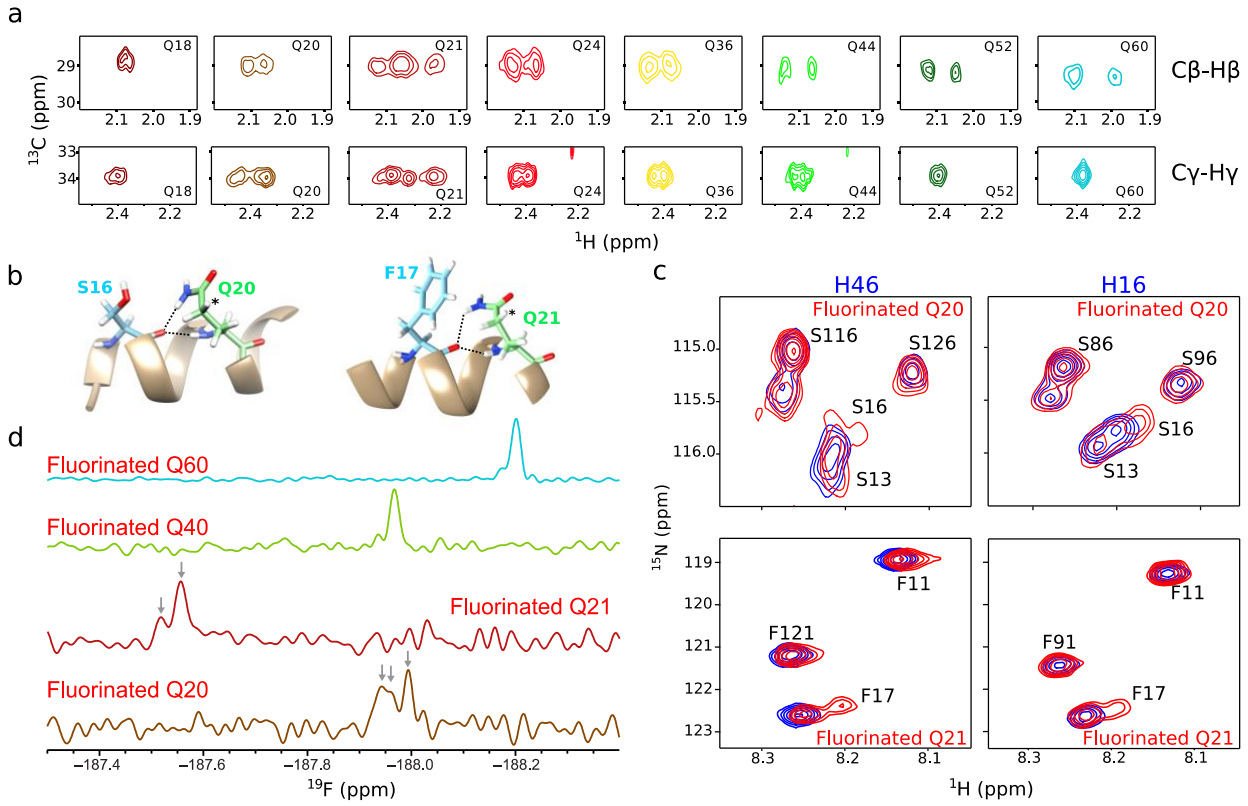

Figure 4. NMR analysis of H46 side chains. (**a**) Cβ-Hβ and Cγ-Hγ regions of the ^{13}C-HSQC spectra of selected glutamines within the poly-Q. The spectra of Q60 display the standard behavior of disordered glutamines with non-degenerate and degenerate diastereotopic protons for Cβ-Hβ and Cγ-Hγ, respectively.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

In a second set of experiments, we incorporated 4F-Gln in four H46 positions located at the beginning (Q20 and Q21), the middle (Q40) and the end (Q60) of the poly-Q to be monitored by 1D-^{19}F-NMR, which is an excellent reporter of biomolecular structure and dynamics[47,48] (Fig. 1a). Strong differences in ^{19}F-CSs were observed for the four positions (Fig. 4d). This demonstrates important structural differences along the poly-Q. The Q21^{19}F-CS was particularly high, due to its proximity to the ring currents exerted by F17 in the hydrogen-bonded form (see below) (Fig 4b and 4d). Interestingly, multiple ^{19}F-NMR resonances were observed for Q20 and Q21. For Q20, three distinct responses were measured, including two signals with very similar CSs. In line with the Cγ-Hγ peaks, this indicates that the Q20 side chain adopts at least two different conformations in a slow exchange regime. For Q21, a weak second ^{19}F signal could be detected, in agreement with the two populations previously identified for its Cα-Hα, Cβ-Hβ and Cγ-Hγ correlations.

Helix propagation and stabilization, a simulation perspective

We performed Gaussian accelerated molecular dynamics (GaMD)[49] simulations to understand the secondary structure propensities of httex1 and the role of N17 in propagating helicity. To simulate an httex1 fragment encompassing N17, the poly-Q tract and five prolines, we used the ff99SBws-STQ force-field, which has been refined to simulate low-complexity proteins[51]. For the eight independent MD simulations, with an aggregated time of ≈20μs, the poly-Q adopted α-helical and disordered conformations, while N17 presented a higher helical propensity (Fig. 5a and S4a). Interestingly, several α-helix folding/melting events were observed. For some of the simulations, α-helices spanned almost the whole length of N17 and the poly-Q and were only absent close to the poly-P (Fig. 5a). In line with the directionality observed experimentally, the process of α-helix melting systematically occurred from C-to-N, as observed around 1600-1800ns, while α-helices in httex1 preferentially grew from N-to-C (Extended Data Fig. 6). The spontaneous formation of short α-helices disconnected from N17 was also observed (*e.g.,* frames 2000-2300ns), suggesting a small inherent α-helical propensity in poly-Q.

The largest fractional helicity was observed for glutamines close to N17, it decreased until Q30 from where it remained flat until Q55 to finally vanish when approaching the poly-P (Fig. S4b). This behavior was qualitatively similar to that observed experimentally in H46, suggesting that our simulations captured the structural mechanisms present in httex1. When analyzing the frames with α-helices with increasing length (from 4 to 40), we observed a gradual increase in fractional helicity from N-to-C, substantiating the structural propagation from N17 (Fig. 5b).

Interestingly, *i→i+4* bifurcated hydrogen bonds were found throughout the poly-Q, although they were more abundant in the beginning of the tract, with Q21 presenting the highest population, and

their number slowly decreased for inner glutamines (Fig. 5c). Importantly, this trend agrees with the presence of Cγ-Hγ doublets (Fig. 4). Then, we analyzed the correlation between bifurcate hydrogen bonds and α-helix stability. Not surprisingly, the percentage of these hydrogen bonds was higher in frames where the segment (*i*→*i*+*4*) adopted a helical conformation. Importantly, the population consistently increased with the stability of the helix, suggesting that bifurcated hydrogen bonds stabilize α-helical conformations (Fig. 5c).

Quantum chemistry calculations of GaMD snapshots demonstrated that the high CS observed for the ^{19}F signal when introducing F-Gln in position Q21 originated by its proximity to the F17 aromatic ring (see Supplemental Note 2 and Extended Data Fig. 7).

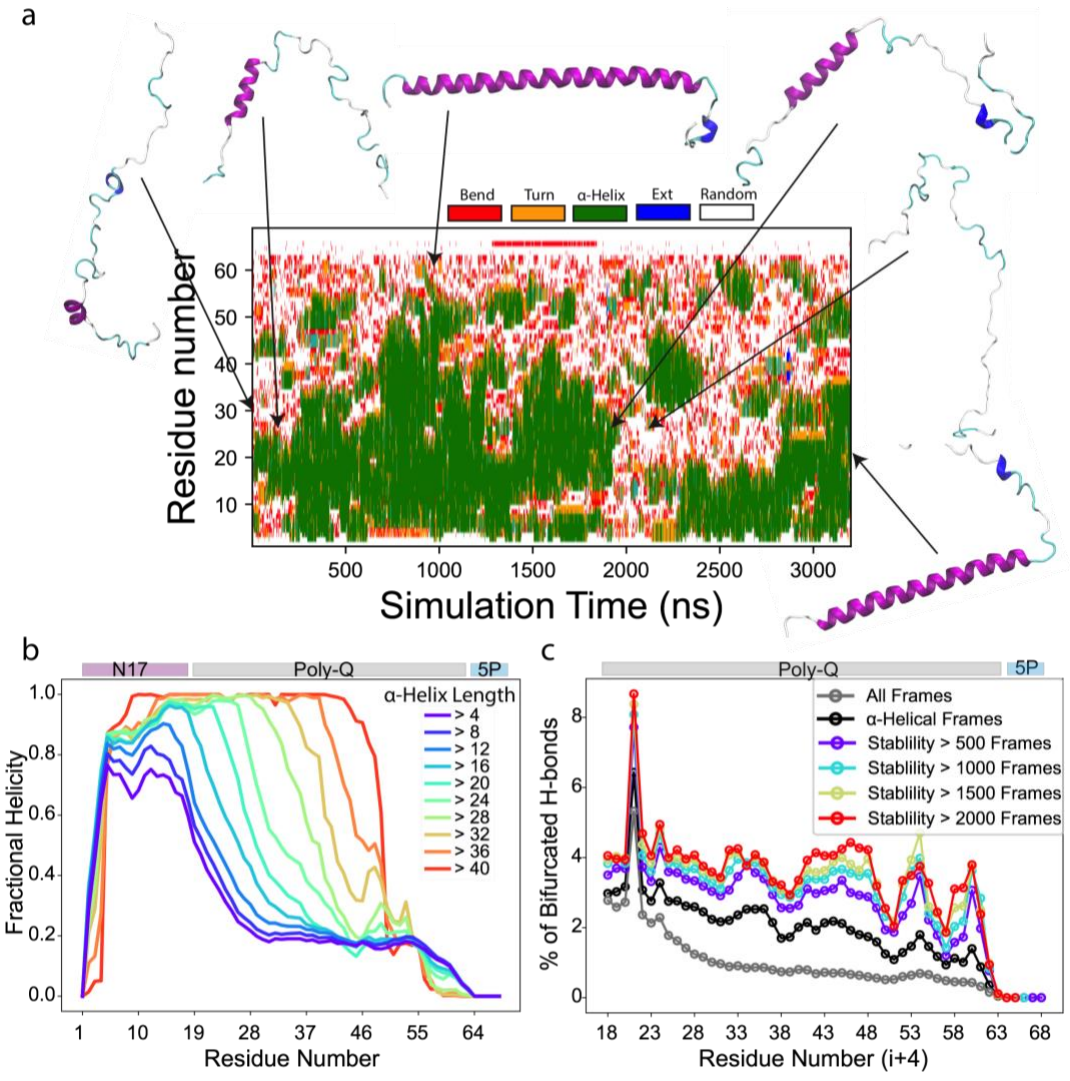

Figure 5. Insights into the conformational landscape of httex1 with MD simulations. (a) The per-residue secondary structure plot for one of the GaMD trajectories showing a transition from an almost completely random-coil conformation to a long α-helix and back to random coil. **(b)** The reweighted fractional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

Role of poly-Q structure in httex1 aggregation

We investigated the relevance of the helical content for the aggregation propensity of H46 with two N17 mutants (Fig. 1a). First, by substituting ^{16}SF^{17} by ^{16}GG^{17} (LKGG-H46), we hampered the hydrogen-bond network connecting both domains. Second, when replacing ^{15}KS^{16} by ^{15}LL^{16} (LLLF-H46), the network was strengthened by incorporating two large hydrophobic residues. Using uniformly and site-specifically labeled samples, we validated the expected structural effects exerted by these mutations (Extended Data Fig. 8). 15µM samples of H46, LKGG-H46 and LLLF-H46 were incubated at 37°C and the soluble fractions were analyzed by SDS-PAGE over a period of 48h (Fig. 6a and S5). H46 presented a moderate aggregation propensity (half-time, $t_{1/2}$, of 19h), exhibiting a decrease of the soluble fraction from the first hours of incubation that almost disappeared after 48h. A much stronger aggregation propensity was observed for LLLF-H46 ($t_{1/2}$=1.4h), while the first signs of aggregation for LKGG-H46 occurred after 20h of incubation ($t_{1/2}$=29h). The presence of sfGFP, which certainly slows down aggregation, hampers the quantitative comparison with previous studies[53]. However, the relative aggregation propensity of the three H46 forms could be unambiguously obtained, indicating that α-helical stability strongly accelerates the aggregation propensity.

The morphology of the aggregates for H46 and the two mutants after 48 and 120h was investigated by correlative atomic force microscopy (AFM) – total internal reflection fluorescence (TIRF) (Fig. 6b,c and Extended Data Fig. 9). The intrinsic fluorescence of the proteins allowed the easier localization of the aggregates on the silica surface and indicated that no proteolytic activity preceded aggregation. The inspection of H46 micrographs measured after 48h of incubation revealed the presence of typical amyloid structures with interconnected fibrils, exhibiting a heterogeneous morphology, with notable variations in width and height, similarly to those recently described[54,55]. Interestingly, LLLF-H46 aggregates displayed similar features, although the presence of fibrils was substantially more abundant. LKGG-H46 exhibited a different behavior with considerably larger and more heterogeneous aggregates with well-defined limits. However, relatively long isolated fibrils with a similar morphology to wild-type and LLLF-H46 were found in the boundaries of the heterogeneous aggregates. After five days of incubation, elongated bundle structures involving several paired filaments were observed for H46 and LLLF-H46 (Fig. 6c and Extended Data Fig. 9b). However, these coiled structures were seldom found in LKGG-H46 aggregates and, when observed, they were less ordered and fragmented, suggesting reduced fibril stability (Fig. 6c).

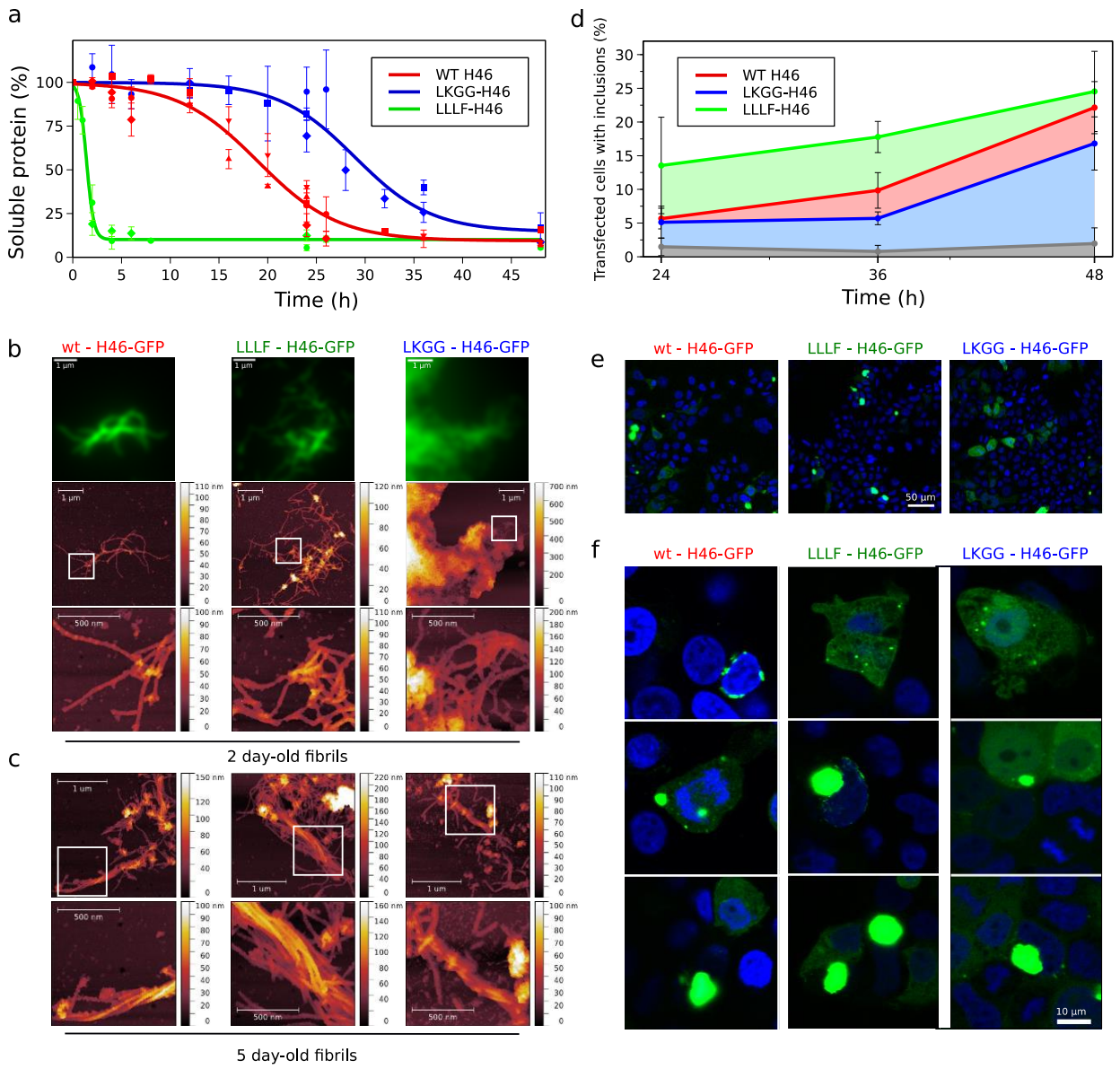

Figure 6. Effect of the α-helical stability on the aggregation properties of huntingtin. (a) Time course of aggregation for wild-type (red), LLLF- (green) and LKGG-H46 (blue) (15 µM) at 37 °C. Each data point corresponds to the mean and associated standard deviation calculated from three independent replicates. Symbols represent different independent experiments. Half-time ($t_{1/2}$) values calculated for each H46 species are indicated according to the color code shown in the legend. **(b)** Representative fluorescence microscopy (upper panels) and AFM (middle and lower panels) images of 2-day-old fibrils of wild-type, LLLF- and LKGG-H46. Each fluorescence image corresponds to the average of 150 frames. **(c)** AFM images of 5-day-old fibrils of three H46 species. White squares indicate the zoom region displayed in the panels below. At least 6 fluorescence and AFM images for each H46 variant and time point were extracted from two large fields (8.13µm x 8.13µm). **(d)** Time-dependent percentage of transfected HEK 293 cells exhibiting cytoplasmic inclusions. Displayed points correspond to the average value from three independent experiments (n=3) with a large (from 63 to 977) number of transfected cells and the error bars present the standard deviations. **(e)** Three representative confocal fields recorded for (from left to right) H46, LLLF-H46 and LKGG-H46 36h hours after transfection. Nuclei were stained with DAPI (blue) and sfGFP fused proteins are observed in green. The scale bar corresponds to 50 µm. The number of confocal fields measured for each variant and time point varied from n=24 (LLLF-H46 at 48h) to n=60 (H46 at 24h and LKGG-H46 at 24h). Note that more fields were recorded for shorter times to compensate the reduced number of transfected cells. **(f)** Representative confocal images of HEK 293 cells overexpressing (from left to right) H46, LLLF-H46 and LKGG-H46. The scale bar corresponds to 10 µm. While some transfected cells display small aggregates, all the H46 constructs can form large inclusions. The number of cells monitored was large, spanning from n=533 (H46 at 48h) to n=1350 (LLLF-H46 at 36h).

Next, we investigated the aggregation properties of H46 variants in a cellular context by transfecting them in HEK 293 cells, a widely used model to study httex1 intracellular inclusions[56–58]. Using confocal microscopy, we monitored the time-dependent formation of inclusions, indicating that all three variants can form large cytoplasmic aggregates in the proximity of the nucleus, although with different propensities (Fig. 6d-f). Already 24h and 36h post-transfection, LLLF-H46 presented a higher percentage of HEK cells with inclusions than the wild-type and LKGG-H46 (Fig. 6d). Conversely, when non-pathogenic H16 was transfected, the presence of inclusions was minimal and did not vary with time (Fig. S6a)[57,58]. The cytoplasmic inclusions for the three variants spanned a broad range of sizes and their average increased with time (Fig. S6b). However, those from LKGG-H46 were significantly smaller than those for wild-type and LLLF-H46 in the first 36h after transfection (Fig. S6c). These observations suggest that the stability of poly-Q α-helical structure modulates the aggregation kinetics of huntingtin in cells, but it does not compromise the capacity of the protein to form large inclusions sequestering other proteins, organelles and lipids[34,58,59].

Discussion

The systematic application of SSIL to H46 demonstrates that this protein retains its helicity as previously observed for non-pathogenic versions[26,29,33]. Using CD, α-helical propensity had been previously identified for httex1 with up to 55 glutamines and showed that the helical content increased concomitantly with the length of the tract [27]. One of the most striking observations of our study is the fairly flat plateau of positive SCS along a large fraction of the poly-Q (Fig. 1d). These SCS are consistent with the coexistence of multiple partially formed α-helices of different lengths, spanning almost the complete poly-Q. However, long α-helices are prevalent according to our structural analysis, evidencing that helices are cooperatively propagated along the homorepeat (Fig. 3). The comparison of the NMR observables for H16 with those of H46 also substantiates the helical stabilization with the poly-Q length. For instance, despite both proteins having the same sequence context, several CSs of initial glutamines are systematically shifted towards more helical conformations in H46 (Extended Data Fig. 1). Moreover, experiments performed on H66 demonstrate that α-helices are maintained for long poly-Q tracts and the distance of individual glutamines to the PRR defines their helicity (Fig. 2).

From a structural perspective, the α-helical stability could arise from the formation of bifurcate hydrogen bonds all along the homorepeat. The spectroscopic features exhibited by Hβ and Hγ have been associated with the formation of bifurcated hydrogen bonds and the concomitant rigidification

of glutamine side chains[33,45]. Our MD simulations indicated that the percentage of bifurcated hydrogen bonds is correlated with the stability of α-helical conformations. The ensemble of these observations suggests that glutamine side chains are actively involved in the stabilization of the helical conformation of the poly-Q tract and that the strength of this mechanism declines when approaching the PRR.

Partial structuration in the poly-Q implies that individual glutamines co-exist in (at least) two different conformations. A similar conclusion was reached by EPR experiments, where two dynamic regimes were identified in N17 and poly-Q tract residues, but not in glutamines in the PRR[27]. However, structural details of this conformational fluctuation could not be unveiled. Several NMR signatures, which can be rationalized through the MD simulations, define the structural bases of this equilibrium. These include, among others, the two Cα-Hα peaks for Q21 (Fig. 1c), the two sets of Cβ-Hβ and Cγ-Hγ peaks for Q20 and Q21 (Fig. 4a), and the multiple ^{19}F-NMR frequencies detected for these two residues. The co-existence of this equilibrium is also manifested in other glutamines of the tract with non-degenerate Cγ-Hγ peaks. Our observations demonstrate that httex1 fluctuates between a rigid α-helical conformation stabilized by bifurcate hydrogen bonds and a more disordered state. Furthermore, our data suggest that the dynamic regime also changes in conjunction with the helical stability. While the first glutamines of the tract exhibit a slow exchange on the NMR timescale, a fast exchange regime is observed for the others. This asymmetric behavior suggests that helix unwinding is initiated in the proximity of the PRR and progresses towards the N-terminus.

The refined ensemble indicates that H46 is a flexible elongated particle in solution and that the overall size is correlated with the length of the poly-Q (Fig. 3c). Importantly, our structural model is in contradiction to previously reported compact models of httex1 stabilized by extensive fuzzy contacts between N17 and poly-Q[25,26,50,63]. Despite the overall extendedness of httex1, our description requires a large degree of disorder, explaining the fluorescence transfer efficiency observed in smFRET[25] and the lack of permanent hydrogen bonds in NMR hydrogen deuterium exchange (HDX) experiments, which have been argued to support the compact model[26].

The similarity of the mechanisms defining the structure of non-pathogenic and pathogenic forms of httex1 validates the 'linear lattice' model[20,22–24]. Our results demonstrate that poly-Q extension is only associated with an increase in the length and stability of helical conformations. This is relevant as our aggregation experiments unambiguously show that the α-helical content is a key factor promoting aggregation. Although N17 has been demonstrated to be the aggregation-triggering domain[18,30,34], httex1 only becomes aggregation-prone when N17 is structurally coupled to the poly-Q tract. This suggests that when structurally uncoupled, such as in LKGG-H46, partially helical

N17 is still able to oligomerize, but the resulting oligomers are less stable and aggregation propensity is reduced. Conversely, strengthening the structural coupling between both domains stabilizes the helical content of the protein and accelerates aggregation, as observed for LLLF-H46. In line with these observations, aggregation experiments with httex1 analogues modifying the poly-Q tract structure highlighted the relevance of the homorepeat structure in defining the aggregation propensity[15,64]. In addition to modified aggregation kinetics, we have observed that mutations in N17 also yield fibers with distinct morphologies. Polymorphism in httex1 aggregates has been observed *in vivo* and *in vitro* when modifying the experimental conditions or when deleting flanking regions[15,53,65–67,68].

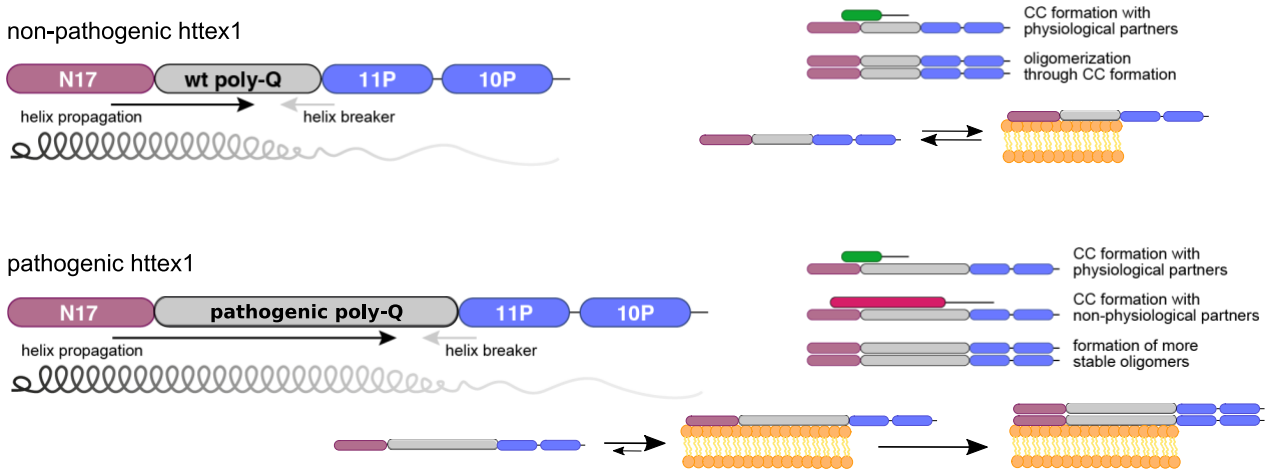

Figure 7. Scheme illustrating the structural influences within non-pathogenic and pathogenic httex1 and their respective modes of interaction. (**Top, non-pathogenic httex1**) The poly-Q tract of non-pathogenic httex1 experiences opposing structural effects from the N17 (α-helix propagation; black arrow)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

In light of our observations, we can speculate about the structural bases of the mechanisms leading to the pathology (Fig. 7). The helical propensity in the poly-Q tract may facilitate intra- and intermolecular assemblies through coiled coil (CC) interactions[17,19,70]. In the non-pathological scenario, the low-stability poly-Q helices define partner selectivity and oligomers exhibit reduced affinity. Beyond the pathological threshold, there is a concomitant population increase of long α-helices, which could still interact with their physiological partners, although most probably with different thermodynamic properties. However, they could also associate with other non-physiological partners and perturb crucial signalling and metabolic pathways. Longer helices can

form more stable oligomers that nucleate amyloidogenic fibrils[17,71] or induce cytotoxicity[72]. Biological membranes can also enhance httex1 local concentration and trigger aggregation[73–75,30,76,77,76,78] (see suppemental discussion).

This study shows that the expansion of the poly-Q in httex1 is associated with an increase in the length and stability of α-helical conformations, which are the main driving forces for aggregation. Our results provide a novel perspective of the pathological threshold in HD that goes beyond the length of the poly-Q by underlining the relevance of conformational preferences. The generalization of these observations to the other poly-Q-related diseases remains to be unveiled. However, the possibility to explore their associated proteins at the residue level and independently of the poly-Q length paves the way to a structural understanding of these pathologies.

Acknowledgements

The authors thank Gottfried Otting (Australian National University, Canberra, Australia) for providing the BL21 (DE3) Star::RF1-CBD3 strain. This work was supported by the European Research Council under the European Union's H2020 Framework Programme (2014-2020) / ERC Grant agreement n° [648030] and Labex EpiGenMed, an « Investissements d'avenir » program (ANR-10-LABX-12-01) awarded to PB, ANR-17-CE11-0022-01 awarded to NS, ANR-19-PI3A-0004 awarded to JC. The CBS is a member of France-BioImaging (FBI) and the French Infrastructure for Integrated Structural Biology (FRISBI), 2 national infrastructures supported by the French National Research Agency (ANR-10-INBS-04-01 and ANR-10-INBS-05, respectively). AU is supported by a grant from the Fondation pour la Recherche Médicale (SPF20150934061). DS acknowledge a grant from the Métropole Européenne de Lille (PUSHUP). Géraldine Levy (Université de Lille) is thanked for help with sample preparation and the ^{19}F-NMR experiments. This work benefited from the HPC resources of CSUC and the CALMIP supercomputing center under the allocations 2016-P16032 and 2021-P21043. The 600 MHz spectrometer for ^{19}F-NMR measurements is funded by the Nord Region Council, CNRS, Institut Pasteur de Lille, the European Community (ERDF), the French Ministry of Research and the Université de Lille and by the CTRL CPER cofunded by the European Union with the European Regional Development Fund (ERDF), by the Hauts-de-France Regional Council (contract n°17003781), Métropole Européenne de Lille (contract n°2016_ESR_05), and French State (contract n°2017-R3-CTRL-Phase1). The authors thank the SWING beamline at SOLEIL synchrotron, Saint-Aubin, France (proposal 20181386) and P12 beamline at PETRAIII, Hamburg, Germany for beamtime allocation to the project and assistance during data collection.

Author Contributions

P.B. conceived the project. C.A.E.R, A.U., A.S., C.D., R.C., A.B., D.S., N.S. and P.B. designed experiments. C.A.E.R, A.U., A.S., M.P., A.M., A.E., A.F., C.D., X.L.L., Z.D.S., L.C., A.T., F.A., R.C. and D.S. performed experiments. R.E.S., P.E.M., A.B., J.C., N.S. and P.B. supervised experiments. C.A.E.R., A.U., A.S. and P.B. wrote the manuscript with the help of all the co-authors.

Competing Interest

The authors declare no competing interests

References

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

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

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

(4) Kremer, B.; Goldberg, P.; Andrew, S. E.; Theilmann, J.; Telenius, H.; Zeisler, J.; Squitieri, F.; Lin, B.; Bassett, A.; Almqvist, E.; et al. A Worldwide Study of the Huntington's Disease Mutation: The Sensitivity and Specificity of Measuring CAG Repeats. *N. Engl. J. Med.* **1994**, *330* (20), 1401–1406. https://doi.org/10.1056/NEJM199405193302001.

(5) Benn, C. L.; Landles, C.; Li, H.; Strand, A. D.; Woodman, B.; Sathasivam, K.; Li, S.-H.; Ghazi-Noori, S.; Hockly, E.; Faruque, S. M. N. N.; et al. Contribution of Nuclear and Extranuclear PolyQ to Neurological Phenotypes in Mouse Models of Huntington's Disease. *Hum. Mol. Genet.* **2005**, *14* (20), 3065–3078. https://doi.org/10.1093/hmg/ddi340.

(6) Graham, R. K.; Deng, Y.; Slow, E. J.; Haigh, B.; Bissada, N.; Lu, G.; Pearson, J.; Shehadeh, J.; Bertram, L.; Murphy, Z.; et al. Cleavage at the Caspase-6 Site Is Required for Neuronal Dysfunction and Degeneration Due to Mutant Huntingtin. *Cell* **2006**, *125* (6), 1179–1191. https://doi.org/10.1016/J.CELL.2006.04.026.

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

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

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

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

(11) Miller, J.; Arrasate, M.; Shaby, B. A.; Mitra, S.; Masliah, E.; Finkbeiner, S. Quantitative

Relationships between Huntingtin Levels, Polyglutamine Length, Inclusion Body Formation, and Neuronal Death Provide Novel Insight into Huntington's Disease Molecular Pathogenesis. *J. Neurosci.* **2010**, *30* (31), 10541 LP – 10550. https://doi.org/10.1523/JNEUROSCI.0146-10.2010.

(12) Ramdzan, Y. M.; Trubetskov, M. M.; Ormsby, A. R.; Newcombe, E. A.; Sui, X.; Tobin, M. J.; Bongiovanni, M. N.; Gras, S. L.; Dewson, G.; Miller, J. M. L.; et al. Huntingtin Inclusions Trigger Cellular Quiescence, Deactivate Apoptosis, and Lead to Delayed Necrosis. *Cell Rep.* **2017**, *19* (5), 919–927. https://doi.org/10.1016/j.celrep.2017.04.029.

(13) Kim, Y. E.; Hosp, F.; Frottin, F.; Ge, H.; Mann, M.; Hayer-Hartl, M.; Hartl, F. U. Soluble Oligomers of PolyQ-Expanded Huntingtin Target a Multiplicity of Key Cellular Factors. *Mol. Cell* **2016**, *63* (6), 951–964. https://doi.org/https://doi.org/10.1016/j.molcel.2016.07.022.

(14) Arrasate, M.; Mitra, S.; Schweitzer, E. S.; Segal, M. R.; Finkbeiner, S. Inclusion Body Formation Reduces Levels of Mutant Huntingtin and the Risk of Neuronal Death. *Nature* **2004**, *431* (7010), 805–810. https://doi.org/10.1038/nature02998.

(15) Wetzel, R. Exploding the Repeat Length Paradigm While Exploring Amyloid Toxicity in Huntington's Disease. *Acc. Chem. Res.* **2020**, *53* (10), 2347–2357. https://doi.org/10.1021/acs.accounts.0c00450.

(16) Sahoo, B.; Arduini, I.; Drombosky, K. W.; Kodali, R.; Sanders, L. H.; Greenamyre, J. T.; Wetzel, R. Folding Landscape of Mutant Huntingtin Exon1: Diffusible Multimers, Oligomers and Fibrils, and No Detectable Monomer. *PLoS One* **2016**, *11* (6), e0155747. https://doi.org/10.1371/journal.pone.0155747.

(17) Ceccon, A.; Tugarinov, V.; Ghirlando, R.; Clore, G. M. Abrogation of Prenucleation, Transient Oligomerization of the Huntingtin Exon 1 Protein by Human Profilin I. *Proc. Natl. Acad. Sci.* **2020**, *117* (11), 5844–5852. https://doi.org/10.1073/PNAS.1922264117.

(18) Kotler, S. A.; Tugarinov, V.; Schmidt, T.; Ceccon, A.; Libich, D. S.; Ghirlando, R.; Schwieters, C. D.; Clore, G. M. Probing Initial Transient Oligomerization Events Facilitating Huntingtin Fibril Nucleation at Atomic Resolution by Relaxation-Based NMR. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (9), 3562–3571. https://doi.org/10.1073/pnas.1821216116.

(19) Thakur, A. K.; Jayaraman, M.; Mishra, R.; Thakur, M.; Chellgren, V. M.; L Byeon, I.-J.; Anjum, D. H.; Kodali, R.; Creamer, T. P.; Conway, J. F.; et al. Polyglutamine Disruption of the Huntingtin Exon 1 N Terminus Triggers a Complex Aggregation Mechanism. *Nat. Struct. Mol. Biol.* **2009**, *16*, 380.

(20) Feng, X.; Luo, S.; Lu, B. Conformation Polymorphism of Polyglutamine Proteins. *Trends Biochem. Sci.* **2018**, *43* (6), 424–435. https://doi.org/10.1016/j.tibs.2018.03.002.

(21) Caron, N. S.; Desmond, C. R.; Xia, J.; Truant, R. Polyglutamine Domain Flexibility Mediates the Proximity between Flanking Sequences in Huntingtin. *Proc. Natl. Acad. Sci.* **2013**, *110* (36), 14610 LP – 14615.

(22) Nucifora, L. G.; Burke, K. A.; Feng, X.; Arbez, N.; Zhu, S.; Miller, J.; Yang, G.; Ratovitski, T.; Delannoy, M.; Muchowski, P. J.; et al. Identification of Novel Potentially Toxic Oligomers Formed in Vitro from Mammalian-Derived Expanded Huntingtin Exon-1 Protein. *J. Biol. Chem.* **2012**, *287* (19), 16017–16028. https://doi.org/10.1074/jbc.M111.252577.

(23) Li, P.; Huey-Tubman, K. E.; Gao, T.; Li, X.; West, A. P.; Bennett, M. J.; Bjorkman, P. J. The Structure of a PolyQ–Anti-PolyQ Complex Reveals Binding According to a Linear Lattice Model. *Nat. Struct. Mol. Biol.* **2007**, *14* (5), 381–387. https://doi.org/10.1038/nsmb1234.

(24) Owens, G. E.; New, D. M.; West, A. P.; Bjorkman, P. J. Anti-PolyQ Antibodies Recognize a Short PolyQ Stretch in Both Normal and Mutant Huntingtin Exon 1. *J. Mol. Biol.* **2015**, *427* (15), 2507–2519. https://doi.org/10.1016/J.JMB.2015.05.023.

(25) Warner, J. B.; Ruff, K. M.; Tan, P. S.; Lemke, E. A.; Pappu, R. V; Lashuel, H. A. Monomeric Huntingtin Exon 1 Has Similar Overall Structural Features for Wild-Type and Pathological Polyglutamine Lengths. *J. Am. Chem. Soc.* **2017**, *139* (41), 14456–14469. https://doi.org/10.1021/jacs.7b06659.

(26) Newcombe, E. A.; Ruff, K. M.; Sethi, A.; Ormsby, A. R.; Ramdzan, Y. M.; Fox, A.; Purcell, A. W.; Gooley, P. R.; Pappu, R. V.; Hatters, D. M. Tadpole-like Conformations of Huntingtin Exon 1 Are Characterized by Conformational Heterogeneity That Persists Regardless of Polyglutamine Length. *J. Mol. Biol.* **2018**, *430* (10), 1442–1458. https://doi.org/10.1016/j.jmb.2018.03.031.

(27) Bravo-Arredondo, J. M.; Kegulian, N. C.; Schmidt, T.; Pandey, N. K.; Situ, A. J.; Ulmer, T. S.; Langen, R. The Folding Equilibrium of Huntingtin Exon 1 Monomer Depends on Its Polyglutamine Tract. *J. Biol. Chem.* **2018**, *293* (51), 19613–19623. https://doi.org/10.1074/jbc.RA118.004808.

(28) Mier, P.; Paladin, L.; Tamana, S.; Petrosian, S.; Hajdu-Soltész, B.; Urbanek, A.; Gruca, A.; Plewczynski, D.; Grynberg, M.; Bernadó, P.; et al. Disentangling the Complexity of Low Complexity Proteins. *Brief. Bioinform.* **2020**, *21* (2), 458–472. https://doi.org/10.1093/bib/bbz007.

(29) Baias, M.; Smith, P. E. S.; Shen, K.; Joachimiak, L. A.; Zerko, S.; Koźmiński, W.; Frydman, J.; Frydman, L. Structure and Dynamics of the Huntingtin Exon-1 N-Terminus: A Solution NMR Perspective. *J. Am. Chem. Soc.* **2017**, *139* (3), 1168–1176. https://doi.org/10.1021/jacs.6b10893.

(30) Ceccon, A.; Schmidt, T.; Tugarinov, V.; Kotler, S. A.; Schwieters, C. D.; Clore, G. M.

Interaction of Huntingtin Exon-1 Peptides with Lipid-Based Micellar Nanoparticles Probed by Solution NMR and Q-Band Pulsed EPR. *J. Am. Chem. Soc.* **2018**, *140* (20), 6199–6202. https://doi.org/10.1021/jacs.8b02619.

(31) Urbanek, A.; Elena-Real, C. A.; Popovic, M.; Morató, A.; Fournet, A.; Allemand, F.; Delbecq, S.; Sibille, N.; Bernadó, P. Site-Specific Isotopic Labeling (SSIL): Access to High-Resolution Structural and Dynamic Information in Low-Complexity Proteins. *ChemBioChem* **2020**, *21* (6). https://doi.org/10.1002/cbic.201900583.

(32) Urbanek, A.; Morató, A.; Allemand, F.; Delaforge, E.; Fournet, A.; Popovic, M.; Delbecq, S.; Sibille, N.; Bernadó, P. A General Strategy to Access Structural Information at Atomic Resolution in Polyglutamine Homorepeats. *Angew. Chem. Int. Ed. Engl.* **2018**, *57* (14), 3598–3601. https://doi.org/10.1002/anie.201711530.

(33) Urbanek, A.; Popovic, M.; Morató, A.; Estaña, A.; Elena-Real, C. A.; Mier, P.; Fournet, A.; Allemand, F.; Delbecq, S.; Andrade-Navarro, M. A.; et al. Flanking Regions Determine the Structure of the Poly-Glutamine in Huntingtin through Mechanisms Common among Glutamine-Rich Human Proteins. *Structure* **2020**, *28* (7), 733-746.e5. https://doi.org/https://doi.org/10.1016/j.str.2020.04.008.

(34) Shen, K.; Calamini, B.; Fauerbach, J. A.; Ma, B.; Shahmoradian, S. H.; Serrano Lachapel, I. L.; Chiu, W.; Lo, D. C.; Frydman, J. Control of the Structural Landscape and Neuronal Proteotoxicity of Mutant Huntingtin by Domains Flanking the PolyQ Tract. *Elife* **2016**, *5*, e18065. https://doi.org/10.7554/eLife.18065.

(35) Bhattacharyya, A.; Thakur, A. K.; Chellgren, V. M.; Thiagarajan, G.; Williams, A. D.; Chellgren, B. W.; Creamer, T. P.; Wetzel, R. Oligoproline Effects on Polyglutamine Conformation and Aggregation. *J. Mol. Biol.* **2006**, *355* (3), 524–535. https://doi.org/https://doi.org/10.1016/j.jmb.2005.10.053.

(36) Morató, A.; Elena-Real, C. A.; Popovic, M.; Fournet, A.; Zhang, K.; Allemand, F.; Sibille, N.; Urbanek, A.; Bernadó, P. Robust Cell-Free Expression of Sub-Pathological and Pathological Huntingtin Exon-1 for Nmr Studies. General Approaches for the Isotopic Labeling of Low-Complexity Proteins. *Biomolecules* **2020**, *10* (10). https://doi.org/10.3390/biom10101458.

(37) Nielsen, J. T.; Mulder, F. A. A. POTENCI: Prediction of Temperature, Neighbor and PH-Corrected Chemical Shifts for Intrinsically Disordered Proteins. *J. Biomol. NMR* **2018**, *70* (3), 141–165. https://doi.org/10.1007/s10858-018-0166-5.

(38) Milles, S.; Salvi, N.; Blackledge, M.; Jensen, M. R. Characterization of Intrinsically Disordered Proteins and Their Dynamic Complexes: From in Vitro to Cell-like Environments. *Prog. Nucl. Magn. Reson. Spectrosc.* **2018**, *109*, 79–100.

https://doi.org/https://doi.org/10.1016/j.pnmrs.2018.07.001.

(39) Estaña, A.; Sibille, N.; Delaforge, E.; Vaisset, M.; Cortés, J.; Bernadó, P. Realistic Ensemble Models of Intrinsically Disordered Proteins Using a Structure-Encoding Coil Database. *Structure* **2019**, *27* (2), 381–391. https://doi.org/10.1016/j.str.2018.10.016.

(40) Krivov, G. G.; Shapovalov, M. V; Dunbrack Jr., R. L. Improved Prediction of Protein Side-Chain Conformations with SCWRL4. *Proteins Struct. Funct. Bioinforma.* **2009**, *77* (4), 778–795. https://doi.org/https://doi.org/10.1002/prot.22488.

(41) Shen, Y.; Bax, A. SPARTA+: A Modest Improvement in Empirical NMR Chemical Shift Prediction by Means of an Artificial Neural Network. *J. Biomol. NMR* **2010**, *48* (1), 13–22. https://doi.org/10.1007/s10858-010-9433-9.

(42) Tria, G.; Mertens, H. D. T.; Kachala, M.; Svergun, D. I. Advanced Ensemble Modelling of Flexible Macromolecules Using X-Ray Solution Scattering. *IUCrJ* **2015**, *2* (Pt 2), 207–217. https://doi.org/10.1107/S205225251500202X.

(43) Bernadó, P.; Mylonas, E.; Petoukhov, M. V; Blackledge, M.; Svergun, D. I. Structural Characterization of Flexible Proteins Using Small-Angle X-Ray Scattering. *J. Am. Chem. Soc.* **2007**, *129* (17), 5656–5664. https://doi.org/10.1021/ja069124n.

(44) Iglesias, J.; Sanchez-Martínez, M.; Crehuet, R. SS-Map: Visualizing Cooperative Secondary Structure Elements in Protein Ensembles. *Intrinsically Disord. Proteins* **2013**, *1* (1), e25323. https://doi.org/10.4161/idp.25323.

(45) Escobedo, A.; Topal, B.; Kunze, M. B. A.; Aranda, J.; Chiesa, G.; Mungianu, D.; Bernardo-Seisdedos, G.; Eftekharzadeh, B.; Gairí, M.; Pierattelli, R.; et al. Side Chain to Main Chain Hydrogen Bonds Stabilize a Polyglutamine Helix in a Transcription Factor. *Nat. Commun.* **2019**, *10* (1), 2034. https://doi.org/10.1038/s41467-019-09923-2.

(46) Qu, W.; Zha, Z.; Ploessl, K.; Lieberman, B. P.; Zhu, L.; Wise, D. R.; B. Thompson, C.; Kung, H. F. Synthesis of Optically Pure 4-Fluoro-Glutamines as Potential Metabolic Imaging Agents for Tumors. *J. Am. Chem. Soc.* **2011**, *133* (4), 1122–1133. https://doi.org/10.1021/ja109203d.

(47) Gimenez, D.; Phelan, A.; Murphy, C. D.; Cobb, S. L. 19F NMR as a Tool in Chemical Biology. *Beilstein J. Org. Chem.* **2021**, *17*, 293–318.

(48) Kitevski-LeBlanc, J. L.; Prosser, R. S. Current Applications of 19F NMR to Studies of Protein Structure and Dynamics. *Prog. Nucl. Magn. Reson. Spectrosc.* **2012**, *62*, 1–33. https://doi.org/https://doi.org/10.1016/j.pnmrs.2011.06.003.

(49) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11* (8), 3584–3595. https://doi.org/10.1021/acs.jctc.5b00436.

(50) Moldovean, S. N.; Chiş, V. Molecular Dynamics Simulations Applied to Structural and Dynamical Transitions of the Huntingtin Protein: A Review. *ACS Chem. Neurosci.* **2020**, *11* (2), 105–120. https://doi.org/10.1021/acschemneuro.9b00561.

(51) Tang, W. S.; Fawzi, N. L.; Mittal, J. Refining All-Atom Protein Force Fields for Polar-Rich, Prion-like, Low-Complexity Intrinsically Disordered Proteins. *J. Phys. Chem. B* **2020**, *124* (43), 9505–9512. https://doi.org/10.1021/acs.jpcb.0c07545.

(52) Wetzel, R. Physical Chemistry of Polyglutamine: Intriguing Tales of a Monotonous Sequence. *J. Mol. Biol.* **2012**, *421* (4–5), 466–490. https://doi.org/10.1016/j.jmb.2012.01.030.

(53) Monsellier, E.; Redeker, V.; Ruiz-Arlandis, G.; Bousset, L.; Melki, R. Molecular Interaction between the Chaperone Hsc70 and the N-Terminal Flank of Huntingtin Exon 1 Modulates Aggregation. *J. Biol. Chem.* **2015**, *290* (5), 2560–2576. https://doi.org/10.1074/jbc.M114.603332.

(54) Galaz-Montoya, J. G.; Shahmoradian, S. H.; Shen, K.; Frydman, J.; Chiu, W. Cryo-Electron Tomography Provides Topological Insights into Mutant Huntingtin Exon 1 and PolyQ Aggregates. *Commun. Biol.* **2021**, *4* (1), 849. https://doi.org/10.1038/s42003-021-02360-2.

(55) Ruggeri, F. S.; Vieweg, S.; Cendrowska, U.; Longo, G.; Chiki, A.; Lashuel, H. A.; Dietler, G. Nanoscale Studies Link Amyloid Maturity with Polyglutamine Diseases Onset. *Sci. Rep.* **2016**, *6*, 31155. https://doi.org/10.1038/srep31155.

(56) Ormsby, A. R.; Cox, D.; Daly, J.; Priest, D.; Hinde, E.; Hatters, D. M. Nascent Mutant Huntingtin Exon 1 Chains Do Not Stall on Ribosomes during Translation but Aggregates Do Recruit Machinery Involved in Ribosome Quality Control and RNA. *PLoS One* **2020**, *15* (7), e0233583.

(57) Vieweg, S.; Mahul-Mellier, A.-L.; Ruggeri, F. S.; Riguet, N.; DeGuire, S. M.; Chiki, A.; Cendrowska, U.; Dietler, G.; Lashuel, H. A. The Nt17 Domain and Its Helical Conformation Regulate the Aggregation, Cellular Properties and Neurotoxicity of Mutant Huntingtin Exon 1. *J. Mol. Biol.* **2021**, *433* (21), 167222. https://doi.org/https://doi.org/10.1016/j.jmb.2021.167222.

(58) Riguet, N.; Mahul-Mellier, A.-L.; Maharjan, N.; Burtscher, J.; Croisier, M.; Knott, G.; Hastings, J.; Patin, A.; Reiterer, V.; Farhan, H.; et al. Nuclear and Cytoplasmic Huntingtin Inclusions Exhibit Distinct Biochemical Composition, Interactome and Ultrastructural Properties. *Nat. Commun.* **2021**, *12* (1), 6579. https://doi.org/10.1038/s41467-021-26684-z.

(59) Bäuerlein, F. J. B.; Saha, I.; Mishra, A.; Kalemanov, M.; Martínez-Sánchez, A.; Klein, R.; Dudanova, I.; Hipp, M. S.; Hartl, F. U.; Baumeister, W.; et al. In Situ Architecture and Cellular Interactions of PolyQ Inclusions. *Cell* **2017**, *171* (1), 179-187.e10.

https://doi.org/10.1016/j.cell.2017.08.009.

(60) Mier, P.; Elena-Real, C.; Urbanek, A.; Bernadó, P.; Andrade-Navarro, M. A. The Importance of Definitions in the Study of PolyQ Regions: A Tale of Thresholds, Impurities and Sequence Context. *Comput. Struct. Biotechnol. J.* **2020**, *18*. https://doi.org/10.1016/j.csbj.2020.01.012.

(61) Dehay, B.; Bertolotti, A. Critical Role of the Proline-Rich Region in Huntingtin for Aggregation and Cytotoxicity in Yeast*. *J. Biol. Chem.* **2006**, *281* (47), 35608–35615. https://doi.org/https://doi.org/10.1074/jbc.M605558200.

(62) Cordeiro, T. N.; Herranz-Trillo, F.; Urbanek, A.; Estaña, A.; Cortés, J.; Sibille, N.; Bernadó, P. Small-Angle Scattering Studies of Intrinsically Disordered Proteins and Their Complexes. *Curr. Opin. Struct. Biol.* **2017**, *42*, 15–23. https://doi.org/https://doi.org/10.1016/j.sbi.2016.10.011.

(63) Kang, H.; Vázquez, F. X.; Zhang, L.; Das, P.; Toledo-Sherman, L.; Luan, B.; Levitt, M.; Zhou, R. Emerging β-Sheet Rich Conformations in Supercompact Huntingtin Exon-1 Mutant Structures. *J. Am. Chem. Soc.* **2017**, *139* (26), 8820–8827. https://doi.org/10.1021/jacs.7b00838.

(64) Drombosky, K. W.; Rode, S.; Kodali, R.; Jacob, T. C.; Palladino, M. J.; Wetzel, R. Mutational Analysis Implicates the Amyloid Fibril as the Toxic Entity in Huntington's Disease. *Neurobiol. Dis.* **2018**, *120*, 126–138. https://doi.org/https://doi.org/10.1016/j.nbd.2018.08.019.

(65) Sathasivam, K.; Lane, A.; Legleiter, J.; Warley, A.; Woodman, B.; Finkbeiner, S.; Paganetti, P.; Muchowski, P. J.; Wilson, S.; Bates, G. P. Identical Oligomeric and Fibrillar Structures Captured from the Brains of R6/2 and Knock-in Mouse Models of Huntington's Disease. *Hum. Mol. Genet.* **2010**, *19* (1), 65–78. https://doi.org/10.1093/hmg/ddp467.

(66) Gruber, A.; Hornburg, D.; Antonin, M.; Krahmer, N.; Collado, J.; Schaffer, M.; Zubaite, G.; Lüchtenborg, C.; Sachsenheimer, T.; Brügger, B.; et al. Molecular and Structural Architecture of PolyQ Aggregates in Yeast. *Proc. Natl. Acad. Sci.* **2018**, *115* (15), E3446 LP-E3453. https://doi.org/10.1073/pnas.1717978115.

(67) Mario Isas, J.; Pandey, N. K.; Xu, H.; Teranishi, K.; Okada, A. K.; Fultz, E. K.; Rawat, A.; Applebaum, A.; Meier, F.; Chen, J.; et al. Huntingtin Fibrils with Different Toxicity, Structure, and Seeding Potential Can Be Interconverted. *Nat. Commun.* **2021**, *12* (1), 4272. https://doi.org/10.1038/s41467-021-24411-2.

(68) Lin, H.-K.; Boatz, J. C.; Krabbendam, I. E.; Kodali, R.; Hou, Z.; Wetzel, R.; Dolga, A. M.; Poirier, M. A.; van der Wel, P. C. A. Fibril Polymorphism Affects Immobilized Non-Amyloid Flanking Domains of Huntingtin Exon1 Rather than Its Polyglutamine Core. *Nat. Commun.* **2017**, *8* (1), 15462. https://doi.org/10.1038/ncomms15462.

(69) Hoop, C. L.; Lin, H.-K.; Kar, K.; Magyarfalvi, G.; Lamley, J. M.; Boatz, J. C.; Mandal, A.; Lewandowski, J. R.; Wetzel, R.; Wel, P. C. A. van der. Huntingtin Exon 1 Fibrils Feature an Interdigitated β-Hairpin–Based Polyglutamine Core. *Proc. Natl. Acad. Sci.* **2016**, *113* (6), 1546–1551. https://doi.org/10.1073/PNAS.1521933113.

(70) Fiumara, F.; Fioriti, L.; Kandel, E. R.; Hendrickson, W. A. Essential Role of Coiled Coils for Aggregation and Activity of Q/N-Rich Prions and PolyQ Proteins. *Cell* **2010**, *143* (7), 1121–1135. https://doi.org/https://doi.org/10.1016/j.cell.2010.11.042.

(71) Jayaraman, M.; Mishra, R.; Kodali, R.; Thakur, A. K.; Koharudin, L. M. I.; Gronenborn, A. M.; Wetzel, R. Kinetically Competing Huntingtin Aggregation Pathways Control Amyloid Polymorphism and Properties. *Biochemistry* **2012**, *51* (13), 2706–2716. https://doi.org/10.1021/bi3000929.

(72) Nekooki-Machida, Y.; Kurosawa, M.; Nukina, N.; Ito, K.; Oda, T.; Tanaka, M. Distinct Conformations of in Vitro and in Vivo Amyloids of Huntingtin-Exon1 Show Different Cytotoxicity. *Proc. Natl. Acad. Sci.* **2009**, *106* (24), 9679 LP – 9684. https://doi.org/10.1073/pnas.0812083106.

(73) Rockabrand, E.; Slepko, N.; Pantalone, A.; Nukala, V. N.; Kazantsev, A.; Marsh, J. L.; Sullivan, P. G.; Steffan, J. S.; Sensi, S. L.; Thompson, L. M. The First 17 Amino Acids of Huntingtin Modulate Its Sub-Cellular Localization, Aggregation and Effects on Calcium Homeostasis. *Hum. Mol. Genet.* **2007**, *16* (1), 61–77. https://doi.org/10.1093/hmg/ddl440.

(74) Atwal, R. S.; Truant, R. A Stress Sensitive ER Membrane-Association Domain in Huntingtin Protein Defines a Potential Role for Huntingtin in the Regulation of Autophagy. *Autophagy* **2008**, *4* (1), 91–93. https://doi.org/10.4161/auto.5201.

(75) Atwal, R. S.; Xia, J.; Pinchev, D.; Taylor, J.; Epand, R. M.; Truant, R. Huntingtin Has a Membrane Association Signal That Can Modulate Huntingtin Aggregation, Nuclear Entry and Toxicity. *Hum. Mol. Genet.* **2007**, *16* (21), 2600–2615. https://doi.org/10.1093/hmg/ddm217.

(76) Michalek, M.; Salnikov, E. S.; Werten, S.; Bechinger, B. Membrane Interactions of the Amphipathic Amino Terminus of Huntingtin. *Biochemistry* **2013**, *52* (5), 847–858. https://doi.org/10.1021/bi301325q.

(77) Marquette, A.; Aisenbrey, C.; Bechinger, B. Membrane Interactions Accelerate the Self-Aggregation of Huntingtin Exon 1 Fragments in a Polyglutamine Length-Dependent Manner. *International Journal of Molecular Sciences* . 2021. https://doi.org/10.3390/ijms22136725.

(78) Michalek, M.; Salnikov, E. S.; Bechinger, B. Membrane Interactions of the Amphipathic Amino Terminus of Huntingtin. *Biophys. J.* **2013**, *105* (3), 699–710. https://doi.org/10.1016/j.bpj.2013.06.030.

Materials and Methods

Huntingtin exon1 constructs

Synthetic genes of wild-type and mutated huntingtin exon1 with 46 and 66 consecutive glutamines (H46 and H66 respectively) and H46 and H66 carrying the amber codon (TAG) instead of the glutamine codon, *e.g*. Q18 (H46Q18), were ordered from GeneArt. All genes were cloned into pIVEX 2.3d as previously described[33], giving rise to pIVEX-httex1-3C-sfGFP-His_6. The sequence of all plasmids was confirmed by sequencing by GENEWIZ. Primers used for cloning are provided in Supplementary Table 3.

Synthesis of 2S,4R-fluoroglutamine

The synthesis of the 2S,4R-fluoroglutamine ((2S, 4R)-2,5-Diamino-4-fluoro-5-oxopentanoic acid), 4F-Gln, was performed as detailed by Qu *et al*.[46]. The purity and the enantiomeric excess (98%) were evaluated by ^{1}H- and ^{19}F-NMR (see Fig. S2 in Supplementary information).

Preparation and aminoacylation of suppressor $tRNA_{CUA}$

The $tRNA_{CUA}$ was aminoacylated with and tRNA synthetase based on the glutamine ligase GLN4 from *Saccharomyces cerevisiae* as previously described[32].

Standard cell-free expression conditions

Lysate was prepared as previously described[33] and based on the *Escherichia coli* strain BL21 Star (DE3)::RF1-CBD_3, a gift from Gottfried Otting (Australian National University, Canberra, Australia)[79]. Cell-free protein expression was performed in batch mode as described by Apponyi *et al*.[80] and detailed in[33]. The concentrations of magnesium acetate (5-20 mM) and potassium glutamate (60-200 mM) were adjusted for each new batch of S30 extract. A titration of both compounds was performed to obtain the maximum yield. A list of reagents used for the lysate production and cell-free reactions is provided is Supplementary Table 4.

Preparation of NMR samples

NMR samples were produced as previously described[33]. Uniformly labeled NMR samples were produced with 3 mg/mL of [^{15}N,^{13}C]-labeled ISOGRO additionally supplemented with [^{15}N,^{13}C]-labeled Asn, Cys and Trp (1 mM each) and 4 mM of Gln. H46 samples in which only certain amino acids were selectively labeled (Ala and Lys; Gly, Ser and Arg; Leu and Glu; and Phe) were prepared by substituting the respective amino acids for the [^{15}N,^{13}C]-labeled ones. To enable the labeling of glutamates, potassium glutamate was substituted with potassium acetate, which was optimized by

testing a range of concentrations[36]. To produce site-specifically labeled samples, the standard reaction mixture was slightly modified by introducing deuterated proline (2 mM) and glutamine 4 mM (Eurisotop). 10 µM of [^{15}N, ^{13}C]-Gln or 4F-Gln loaded $tRNA_{CUA}$ was added for suppressed samples. The same procedures were used for the preparation of H66 samples.

Expression of H16 in *E. coli*

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

Protein purification

The httex1 produced with cell-free reactions were purified as previously described[33]. For NMR experiments, samples were dialyzed against NMR buffer (20 mM BisTris-HCl pH 6.5, 150 mM NaCl) at 4°C using SpectraPor 4 MWCO 12-14 kDa dialysis tubing (Spectrum Labs). Dialyzed protein was then concentrated with 10 kDa MWCO Vivaspin centrifugal concentrators (3,500 xg, 4°C) (Sartorius). For SAXS measurements, the affinity chromatography step was carried out in an AKTA pure System (GE Healthcare) with a 5 mL Histrap® Excel column. For both aggregation and SAXS experiments, an additional size-exclusion chromatography step, using a Superdex S200 10/300 column, was carried out. For aggregation measurements, this step was performed in aggregation buffer (50 mM sodium phosphate, pH 7.5, 150 mM NaCl) and for SAXS measurements in NMR buffer. Protein concentrations were determined by means of fluorescence using an sfGFP calibration curve. Final NMR sample concentrations ranged from 4 to 15 µM. Protein integrity was analyzed by SDS-PAGE.

NMR experiments and data analysis

NMR sample preparation and chemical shift analyses were performed as previously described[33]. ^{15}N- and ^{13}C-HSQC experiments were performed at 293 K on a Bruker Avance III spectrometer equipped with a cryogenic triple resonance probe and Z gradient coil, operating at a ^{1}H frequency of 700 MHz or 800 MHz. All spectra were processed with TopSpin v4.0.5 (Bruker Biospin) and

analyzed with CCPN-Analysis software[81]. Chemical shifts were referenced with respect to the H_2O signal relative to DSS using the $^1H/X$ frequency ratio of the zero point according to Markley *et al.*[82]. Random coil chemical shifts were predicted using POTENCI (http://nmr.chem.rug.nl/potenci/)[37]. Secondary chemical shifts (SCS) were obtained by subtracting the predicted value from the experimental one ($SCS=\delta_{exp}-\delta_{pred}$). The combined C_α and C_β secondary chemical shifts ($SCS(C_\alpha)-SCS(C_\beta)$) were used to improve reliability.

^{19}F-NMR experiments

All NMR samples were first concentrated up to a ca. 200 µl volume using Vivaspin centrifugal concentrators (Sartorius) with a 5 kDa cutoff at 4°C. 0.1 µl of a trimethylsilylpropanoic acid (TMSP) solution for chemical shift referencing and 10 µl of D_2O were added before NMR measurement. All ^{19}F NMR experiments were performed on a Bruker Avance III HD spectrometer operating at a ^{1}H and ^{19}F frequencies of 600.13 MHz and 564.69 MHz, respectively, equipped with a CP-QCI-F cryoprobe with ^{19}F cryo-detection. All ^{19}F 1D experiments were performed at 293.0 K with ^{1}H decoupling during acquisition using waltz16 composite pulse decoupling. An acquisition time of 0.58 s, spectral width of 100.6 s and relaxation delay of 1.0 s was used for all samples, except for the sample fluorinated at Q20, where an acquisition time of 0.29 s and a relaxation delay of 0.5 s was used. Concatenated 1D ^{19}F spectra of 128 transients each were acquired in order to monitor any spectral changes over time. Signal averaging was then performed up to a time point before significant spectral changes over time could be detected. Final number of transients varied were 77824, 49152, 39680 and 53760 for samples fluorinated at Q20, Q21, Q40 and Q60, respectively. ^{19}F spectra were referenced to the ^{1}H signal of TMSP using the unified chemical shift scale.

Model building and chemical shift ensemble optimization

Ensemble models for the two families capturing the conformational influences of the flanking regions, N→C and N←C, were constructed with the algorithm described in reference[39], which uses a curated database of three-residue fragments extracted from high-resolution protein structures. The model building strategy consecutively appends residues, which are considered to be either fully disordered or partially structured. For fully disordered residues, amino acid specific ϕ/ψ angles defining the residue conformation are randomly selected from the database, disregarding their sequence context. For partially structured residues, the nature and the conformation of the flanking residues are taken into account when selecting the conformation of the incorporated residue. Steric clashes are tested at each step, and a backtracking strategy is applied to solve possible conflicts (see detailed explanation of the algorithm in reference[39]).

Two families of ensembles were built. For the first family (N→C ensembles), starting with the ^{10}AFESLKS^{16} region of N17 as partially structured, multiple ensembles of 5,000 conformations were built by successively including an increasing number of glutamines in the poly-Q tract (from F17 to Q63) as partially structured, while the rest of the chain was considered to be fully disordered. An equivalent strategy was followed for the second family of ensembles (N←C ensembles) for which glutamines were considered successively as partially structured from the poly-P tract (from Q63 to Q18). Note that in the partially structured building strategy secondary structural elements are propagated due to the conformational influence of neighboring residues. Two tripeptide databases were used to generate the conformational ensemble models. Both were constructed from the protein domains in the SCOP (Structural Classification of Proteins)[83,84] repository filtered to 95% sequence identity. An "unfiltered" tripeptide database was built disregarding secondary structure content, and a "coil" database included tripeptides not participating in α-helices or β-strands. For the N→C ensembles, the best results were obtained when using the "unfiltered" and "coil" databases to sample the partially structured and the fully disordered sections, respectively. For the N←C ensembles, the "coil" database yielded the best results. For the resulting 47 ensembles of each family, and after building the side chains with the program SCWRL4[40], averaged Cα chemical shifts were computed with SPARTA+[41], and used for the optimization. The optimized ensemble model of H46 was built by reweighting the populations of the pre-computed ensembles, minimizing the error with respect to the experimental Cα CSs. In order to capture the influence of the flanking regions, glutamines within the tract were divided into two groups: those influenced by N17 and those influenced by the poly-P tract, whose chemical shifts were fitted with the N→C and N←C ensembles, respectively. The limit between both families was systematically explored by computing the agreement between the experimental and optimized CSs through a $\chi 2$ value. An optimal description of the complete CS profile was obtained when Q55 was chosen as the last residue structurally connected with N17. Finally, an ensemble of 11,000 conformations was built using the optimized weights and it was used to derive secondary structure population using SS-map[44] and to analyze the SAXS data.

SAXS data measurement and analysis

The SAXS data for H16 were collected at the SWING beamline at the SOLEIL synchrotron, France, equipped with an Eiger 4M detector with a sample-to-detector distance of 1.5 m[85]. The data for H46 were collected at EMBL-bioSAXS-P12 beamline at PETRAIII, Hamburg, Germany equipped with a Pilatus 6M detector with a sample-to-detector distance of 3 m[86]. The parameters used for SAXS data collection are given in Supplementary Table 2. All the data were collected in SEC-SAXS mode with an in-line Superdex 200 Increase 10/300 GL column (GE Healthcare). Both

proteins were concentrated to 8 mg/mL and centrifuged at 20,000 x g immediately before injecting the protein onto the column. 80 µL of the sample were injected into the column and the flow rate was maintained at 0.5 mL/min. The initial data processing steps including masking and azimuthal averaging were performed using the program FOXTROT[87] for H16 and SASFLOW pipeline[86] for H46. The resulting 1D profiles were analyzed using CHROMIXS[88] from ATSAS suite to select the frames corresponding to sample and buffer and perform buffer subtraction. The final buffer subtracted and averaged SAXS profiles were analyzed using ATSAS 2.8 software package[89], including Primus3.3, AUTORG for calculating the radius of gyration and calculation of extrapolated value of radius of gyration (R_g), GNOM[90] for calculation of pairwise distance distribution profiles and DATBAYES[91] for calculation of molecular weight by Bayesian estimate from four approaches. The ensemble optimization approach (EOM 2.1) was used to select sub-ensembles that collectively describe the SAXS data. The program RanCH was first used to join each of the conformations of H16 and H46 (generated as described above) to the modelled structure of sfGFP (PDBID: 3LVA) and the hexahistidine tag used for purification. GAJOE was then used to find a sub-ensemble from this pool which collectively describes the SAXS data[42,43]. The graphical representations were generated using the program VMD[92].

Molecular dynamics simulations

We performed Gaussian Accelerated Molecular Dynamics (GaMD)[49] simulations to explore the conformational landscape and the secondary structure propensities of a fragment of httex1 consisting of N17, 46 glutamines and 5 prolines. We used the ff03ws-STQ[51] force field, which is 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 described earlier in the model building section as the starting structure and prepared the simulation system using tools available with GROMACS 2020.5[93]. This included addition of hydrogens, solvation with ~20,000 TIP4P/2005s water molecules[94] in an octahedral periodic box and addition of ions (Na^+ and Cl^-) to neutralize the system and set the final salt concentration to 0.15 M. Thereafter, the system was converted to AMBER format using ParmEd tool[95] in AMBER20[96]. At this stage, hydrogen mass repartitioning was also done to allow a time step of 4 fs. We used periodic boundary conditions and constrained the bonds containing hydrogen atoms using SHAKE algorithm. Particle Mesh Ewald summation (PME) was used to calculate electrostatic interaction with a cut-off of 9 Å on the long-range interactions. The system was energy minimized for 5000 steps. This was followed by an NVT (T=293K) equilibration for 5 ns and further equilibration in NPT (P=1 atm, T=293K) ensemble for 10 ns. For all the simulations, the temperature was maintained at 293K using a Langevin thermostat and for NPT simulations the pressure was

maintained at 1 atm using a Monte Carlo barostat[97]. This was followed by a GaMD equilibration stage which consisted of a classical MD simulation of 40 ns during which the potential statistics to calculate GaMD parameters were collected followed by a 120 ns long equilibration during which the boost was added and updated. Finally, eight independent simulations with an aggregate simulation time of ~ 20 μs were launched using the boost parameters obtained in the equilibration stage. The simulations were run in "dual-boost" mode and the reference energy was set to the lower bound (E=Vmax). The average and standard deviation of the potential energy were calculated every 2ns (number of steps ≈ 4*system size).

Quantum chemistry calculations of ^{19}F chemical shifts

^{19}F chemical shielding calculations were performed with the Orca 5.0.2 software[98,99] in a reduced system prepared with pDynamo[100]. The structures used for these calculations were randomly collected from the GaMD trajectory. Concretely, 50 structures presenting either a bifurcate hydrogen bond between S16 and Q21 or between F17 and Q21 were selected. Furthermore, 100 conformations with a α-helix in the S16-Q21 stretch but not presenting bifurcate hydrogen bonds, and 100 additional ones not presenting any secondary structure (random coil) in this stretch were also selected for the calculations. Note that for the α-helical and the random coil scenarios, half of the structures were used to compute ^{19}F chemical shielding for the fluorine in Q20 and the other 50 were used for the fluorine in Q21. For each of the structures we selected residues from S16 to Q22 and the limiting backbone atoms to retain the peptide bond. Hydrogen atoms were added to saturate the excised covalent bonds. The fluorine atom was added by substituting the corresponding Hγ atom in Q20 or Q21 at a bond distance of 1.35Å. Note that no structural optimization was performed previous to the chemical shielding calculations. The final system used for the chemical shielding calculation consisted of 143 atoms. The TPSS functional[101], which has been shown to provide precise chemical shieldings[102,103], was used for the calculations. The D3BJ dispersion correction[104,105] was used together with the pcSseg-2 basis set[106], which has been optimized for the calculation of chemical shielding constants. Solvation effects were included using the CPCM implicit solvent model[107]. According to Fedorov and Krivdin[108], ^{19}F chemical shifts are reported in parts per million (ppm) relative to the calculated absolute shielding of CCl_3F in chloroform (σ = 149.7 ppm) using the same level of theory. We optimized the geometry of CCl_3F but, for consistency with the other calculations, we kept the C-F distance fixed to 1.35 Å. The effect of this constrain was minor, as the calculated chemical shielding after optimizing this distance was 149.0 ppm.

Aggregation experiments

Time-dependent aggregation of H46 variants was followed with SDS-PAGE analysis as previously described, with minor modifications[53]. 15 µM wild-type-, LKGG- and LLLF-H46 samples, prepared in aggregation buffer (50 mM sodium phosphate, pH 7.5, 150 mM NaCl), were incubated at 37°C for 48 h, without shaking. 10 µL-aliquots were extracted at different time intervals, immediately mixed with denaturing buffer (125 mM Tris-HCl, pH 6.8, 20% glycerol, 4% SDS, 200 mM dithiothreitol, 0.05% bromophenol blue), incubated for 10 min at 95°C, and frozen at -20 °C until analysis on Bolt™ 4–12% Bis-Tris Plus gels (Invitrogen). The gels were washed in water, stained with Instant Blue Coomassie Protein Stain (Abcam), and visualized using a Gel Doc™ Ez Imager (Bio-Rad). The amount of SDS-soluble species trapped in the stacking gel was quantified using the Image Lab 5.1 software. The percentages of soluble protein were referenced to time 0 values, and plotted against time. The plots were fitted using GraphPad Prism 7.0a. For each protein variant, at least two independent experiments, with three replicates at each time point, were recorded.

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

Correlative AFM-TIRF microscopy was developed in-house[109]. AFM images were acquired using a Nanowizard 4 (JPK Instruments, Bruker) mounted on a Zeiss inverted optical microscope and equipped with a Vortis-SPM control unit. A custom-made TIRF microscope was coupled to the AFM using a LX 488-50 OBIS laser source (Coherent). We used an oil immersion objective with a 1.4 numerical aperture (Plan-Apochromat 100x, Zeiss). Fluorescence was collected with an EmCCD iXon Ultra897 (Andor) camera. The setup includes a 1.5x telescope to obtain a final imaging magnification of 150-fold, corresponding to a camera pixel size of 81.3 nm. An ET800sp short pass filter (Chroma) was used in the emission optical path to filter out the light source of the AFM optical beam deflection system. The excitation laser wavelength was centered at 488 nm and the power was measured before the objective with a PM100 energy meter (purchased from Thorlabs) and was optimized in all the experiments in the range of 1-5 µW. Fluorescence images were acquired using an ET525/50 nm (Chroma) emission filter and an acousto-optic tunable filter (AOTFnc-400.650-TN, AA opto-electronics) to modulate the laser intensity. Fluorescence images were obtained by averaging 150 individual images, each acquired over 50 ms as exposure time.

AFM images were collected in liquid environment (Dulbecco's PBS named D-PBS) using the quantitative-imaging (QI) mode. Each image was acquired with 256x256 lines/pixels and the following scan size: 5 µm × 5 µm, 2.5 µm × 2.5 µm and 1 µm × 1 µm. Typical force *versus* distance curves were recorded with a tip approach speed ranging from 10 µm/s to 30 µm/s and an oscillation amplitude (Z length) of 100 nm or 150 nm, adjusted depending on the height of the aggregates. The maximal force exerted in each pixel was set to 100-150 pN and optimized during the image

acquisition. We used MSNL-D and MSNL-E (Bruker) AFM probes with resonances in liquid of ≈ 2 kHz and 10 kHz, respectively, and nominal spring constants of 0.03 N/m and 0.1 N/m. MSNL cantilevers have a sharp tip radius (≈ 2 nm), which is ideal for high-resolution imaging. The inverse optical lever sensitivity was calibrated with the acquisition of a force *versus* distance curve on the glass coverslip whereas the cantilever stiffness was calibrated using thermal method[110].
Samples for correlative AFM and TIRF were prepared on circular glass coverslips (2.5 cm, 165 µm thick, purchased from Marienfeld). Coverslips were cleaned with a 15 min cycle of sonication with ultrasounds in 1 M KOH, rinsed 20 times with deionized water and finally with a second cycle of sonication in deionized water. Fibrils were then deposited on the clean glass coverslips and let dry before being immersed in D-PBS for AFM imaging. AFM Images were analyzed with the program Gwyddion 2.44.

Cell culture and transfection

The four genes (H16, H46 and the two mutants LLLF-H46 and LKGG-H46), fused with the sfGFP at the C-terminus, were amplified from corresponding pIVEX vector constructs with primers provided as additional data. The amplified genes were cloned in a pcDNA 3.1/Myc-His (-) B vector, digested by BamH I and Xho I, with In-Fusion\ (Takara Bio USA, Inc.) cloned and transformed into the Stellar *E. coli* HST08 strain. Clones were characterized by sequencing (GENEWIZ) and the product from the DNA maxi-preparation (NucleoBond\ Xtra Maxi Macherey-Nagel) was used for cell transfections.

HEK 293 cells were grown in DMEM Glutamax (Gibco) complemented with 10% fetal calf serum. For transient expression, cells were plated on glass coverslips as follows: for each construct, three 12mm coverslips were placed in a 35mm well and coated with poly-ornithine for 30min. Cells were then seeded at a density of $4x10^5$ cells per well. The next day, cells were transfected with 1µg of plasmid and 2 µL of lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Medium was changed 4 hours after transfection. For each time point of the expression kinetics (24h, 36h and 48h after transfection), a coverslip was rinsed in PBS and fixed with 4% PFA for 5 minutes at room temperature. Coverslips were then rinsed twice in PBS, stained with Hoechst (1 µg/mL), mounted on glass slides with a drop of Vectashield (Vector Laboratories) and sealed with nail polish.

Confocal and wide-field microscopy

Samples were imaged with a Leica SP8 confocal microscope with an oil immersion objective (63x or 40x). As the fluorescence intensity in the protein aggregates was very high, a compromise had to be found to detect transfected cells with dilute protein expression, while not saturating the detector. An argon laser set to a power of 0.04% was used to excite sfGFP. Gain and laser power were kept constant for all acquisitions. For each condition, at least 15 images were acquired, containing in total at least 100 transfected cells. Images were then processed with Fiji (ImageJ2, v2.3.0) and the percentages of transfected cells with intracellular inclusions were counted with an in-house Python script. Results of three independent experiments were reported.

To quantitatively analyze the size of huntingtin inclusions accumulated in cells, we used an upright microscope (Zeiss Axio Imager A1) equipped with a Zeiss Axiocam camera, a 50x/0.55 LD objective and a Lumencor Light Engine as a light source. Acquisition was performed with the Zen software, with 8 ms acquisition time. At least 40 images per condition were analyzed with a custom Python script. Images were denoised and subjected to a first order flattening to correct for inhomogeneous illumination. A threshold was then applied to all images and inclusions were automatically segmented. Areas of the aggregates were then plotted for each variant and time point. Scipy, numpy, skimage and matplotlib packages were used for the analysis.

Data Availability Statement

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. The accession codes for the SAXS data (SASDB) are SASDQR8 (H16) SASDQS8 (H46) and the ensembles have bee deposited in the PEDB under the accession codes PED00223 (H16) and PED00224 (H46). 3D structure of sfGFP was downloaded from the PDB (PDBID: 3LVA).

Code Availability

In-house script to analyse fluorescence images and disordered chain building program will be made available from the corresponding author on reasonable request.

References from Methods

(79) Loscha, K. V; Herlt, A. J.; Qi, R.; Huber, T.; Ozawa, K.; Otting, G. Multiple-Site Labeling of Proteins with Unnatural Amino Acids. *Angew. Chem. Int. Ed. Engl.* **2012**, *51* (9), 2243–2246. https://doi.org/10.1002/anie.201108275.

(80) Apponyi, M. A.; Ozawa, K.; Dixon, N. E.; Otting, G. Cell-Free Protein Synthesis for Analysis by NMR Spectroscopy. In *Structural Proteomics: High-Throughput Methods*; Kobe, B., Guss, M., Huber, T., Eds.; Humana Press: Totowa, NJ, 2008; pp 257–268. https://doi.org/10.1007/978-1-60327-058-8_16.

(81) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins* **2005**, *59* (4), 687–696. https://doi.org/10.1002/prot.20449.

(82) Markley, J. L.; Bax, A.; Arata, Y.; Hilbers, C. W.; Kaptein, R.; Sykes, B. D.; Wright, P. E.; Wüthrich, K. Recommendations for the Presentation of NMR Structures of Proteins and Nucleic Acids. *J. Mol. Biol.* **1998**, *280* (5), 933–952. https://doi.org/10.1006/jmbi.1998.1852.

(83) Andreeva, A.; Howorth, D.; Chothia, C.; Kulesha, E.; Murzin, A. G. SCOP2 Prototype: A New Approach to Protein Structure Mining. *Nucleic Acids Res.* **2014**, *42* (D1), D310–D314. https://doi.org/10.1093/nar/gkt1242.

(84) Andreeva, A.; Kulesha, E.; Gough, J.; Murzin, A. G. The SCOP Database in 2020: Expanded Classification of Representative Family and Superfamily Domains of Known Protein Structures. *Nucleic Acids Res.* **2020**, *48* (D1), D376–D382. https://doi.org/10.1093/nar/gkz1064.

(85) Thureau, A.; Roblin, P.; Pérez, J. BioSAXS on the SWING Beamline at Synchrotron SOLEIL. *J. Appl. Crystallogr.* **2021**, *54* (6). https://doi.org/10.1107/S1600576721008736.

(86) Blanchet, C. E.; Spilotros, A.; Schwemmer, F.; Graewert, M. A.; Kikhney, A.; Jeffries, C. M.; Franke, D.; Mark, D.; Zengerle, R.; Cipriani, F.; et al. Versatile Sample Environments and Automation for Biological Solution X-Ray Scattering Experiments at the P12 Beamline (PETRA III, DESY). *J. Appl. Crystallogr.* **2015**, *48* (Pt 2), 431–443. https://doi.org/10.1107/S160057671500254X.

(87) Girardot, R.; Viguier, G.; Pérez, J.; Ounsy, M. FOXTROT: A Java-Based Application to Reduce and Analyse SAXS and WAXS Piles of 2D Data at Synchrotron SOLEIL. In *canSAS-VIII*; 2015.

(88) Panjkovich, A.; Svergun, D. I. CHROMIXS: Automatic and Interactive Analysis of Chromatography-Coupled Small-Angle X-Ray Scattering Data. *Bioinformatics* **2018**, *34* (11), 1944–1946. https://doi.org/10.1093/bioinformatics/btx846.

(89) Franke, D.; Petoukhov, M. V.; Konarev, P. V.; Panjkovich, A.; Tuukkanen, A.; Mertens, H. D. T.; Kikhney, A. G.; Hajizadeh, N. R.; Franklin, J. M.; Jeffries, C. M.; et al. ATSAS 2.8: A Comprehensive Data Analysis Suite for Small-Angle Scattering from Macromolecular Solutions. *J. Appl. Crystallogr.* **2017**, *50*, 1212–1225. https://doi.org/10.1107/S1600576717007786.

(90) Svergun, D. I. Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *J. Appl. Crystallogr.* **1992**, *25* (pt 4), 495–503. https://doi.org/10.1107/S0021889892001663.

(91) Hajizadeh, N. R.; Franke, D.; Jeffries, C. M.; Svergun, D. I. Consensus Bayesian Assessment of Protein Molecular Mass from Solution X-Ray Scattering Data. *Sci. Rep.* **2018**, *8* (1). https://doi.org/10.1038/s41598-018-25355-2.

(92) Humphrey, W.; Dalke, A.; Schulten, K. {VMD} -- {V}isual {M}olecular {D}ynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(93) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/https://doi.org/10.1016/j.softx.2015.06.001.

(94) Best, R. B.; Zheng, W.; Mittal, J. Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10* (11), 5113–5124. https://doi.org/10.1021/ct500569b.

(95) Shirts, M. R.; Klein, C.; Swails, J. M.; Yin, J.; Gilson, M. K.; Mobley, D. L.; Case, D. A.; Zhong, E. D. Lessons Learned from Comparing Molecular Dynamics Engines on the SAMPL5 Dataset. *J. Comput. Aided. Mol. Des.* **2017**, *31* (1), 147–161. https://doi.org/10.1007/s10822-016-9977-1.

(96) Case DA, Betz RM, Botello-Smith W, Cerutti DS, Cheatham TE, D. T. AMBER 2016. University of California, San Francisco 2016.

(97) Chow, K.-H.; Ferguson, D. M. Isothermal-Isobaric Molecular Dynamics Simulations with Monte Carlo Volume Sampling. *Comput. Phys. Commun.* **1995**, *91* (1), 283–289. https://doi.org/https://doi.org/10.1016/0010-4655(95)00059-O.

(98) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA Quantum Chemistry Program Package. *J. Chem. Phys.* **2020**, *152* (22), 224108. https://doi.org/10.1063/5.0004608.

(99) Neese, F. Software Update: The ORCA Program System—Version 5.0. *WIREs Comput. Mol. Sci.* **2022**, *n/a* (n/a), e1606. https://doi.org/https://doi.org/10.1002/wcms.1606.

(100) Field, M. J. The PDynamo Program for Molecular Simulations Using Hybrid Quantum Chemical and Molecular Mechanical Potentials. *J. Chem. Theory Comput.* **2008**, *4* (7), 1151–1161. https://doi.org/10.1021/ct800092p.

(101) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91* (14), 146401. https://doi.org/10.1103/PhysRevLett.91.146401.

(102) Stoychev, G. L.; Auer, A. A.; Izsák, R.; Neese, F. Self-Consistent Field Calculation of Nuclear Magnetic Resonance Chemical Shielding Constants Using Gauge-Including Atomic Orbitals and Approximate Two-Electron Integrals. *J. Chem. Theory Comput.* **2018**, *14* (2), 619–637. https://doi.org/10.1021/acs.jctc.7b01006.

(103) Schattenberg, C. J.; Kaupp, M. Extended Benchmark Set of Main-Group Nuclear Shielding Constants and NMR Chemical Shifts and Its Use to Evaluate Modern DFT Methods. *J. Chem. Theory Comput.* **2021**, *17* (12), 7602–7621. https://doi.org/10.1021/acs.jctc.1c00919.

(104) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32* (7), 1456–1465. https://doi.org/10.1002/jcc.21759.

(105) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15), 154104. https://doi.org/10.1063/1.3382344.

(106) Jensen, F. Segmented Contracted Basis Sets Optimized for Nuclear Magnetic Shielding. *J. Chem. Theory Comput.* **2015**, *11* (1), 132–138. https://doi.org/10.1021/ct5009526.

(107) Cammi, R.; Mennucci, B.; Tomasi, J. Fast Evaluation of Geometries and Properties of Excited Molecules in Solution: A Tamm-Dancoff Model with Application to 4-Dimethylaminobenzonitrile. *J. Phys. Chem. A* **2000**, *104* (23), 5631–5637. https://doi.org/10.1021/jp000156l.

(108) Fedorov, S. V; Krivdin, L. B. Computational Protocols for the 19F NMR Chemical Shifts. Part 1: Methodological Aspects. *J. Fluor. Chem.* **2020**, *238*, 109625.

(109) Dahmane, S.; Doucet, C.; Le Gall, A.; Chamontin, C.; Dosset, P.; Murcy, F.; Fernandez, L.; Salas, D.; Rubinstein, E.; Mougel, M.; et al. Nanoscale Organization of Tetraspanins during HIV-1 Budding by Correlative DSTORM/AFM. *Nanoscale* **2019**, *11* (13), 6036–6044. https://doi.org/10.1039/C8NR07269H.

(110) Proksch, R.; Schäffer, T. E.; Cleveland, J. P.; Callahan, R. C.; Viani, M. B. Finite Optical Spot Size and Position Corrections in Thermal Spring Constant Calibration. *Nanotechnology* **2004**, *15* (9), 1344–1350. https://doi.org/10.1088/0957-4484/15/9/039.