Multi-site-specific isotopic labeling accelerates high-resolution structural investigations of pathogenic huntingtin exon-1

Carlos Elena-Real, Annika Urbanek, Xamuel Lund, Anna Morató, Amin Sagar, Aurélie Fournet, Alejandro Estaña, Tracy Bellande, Frédéric Allemand, Juan Cortés, et al.

Multi-Site Specific Isotopic Labelling Accelerates High-Resolution Structural Investigations of Pathogenic Huntingtin Exon-1

Carlos A. Elena-Real[1], Annika Urbanek[1], Xamuel L. Lund[1,2], Anna Morató[1], Amin Sagar[1], Aurélie Fournet[1], Alejandro Estaña[1,3], Tracy Bellande[4], Frédéric Allemand[1], Juan Cortés[3], Nathalie Sibille[1], Ronald Melki[4] and Pau Bernadó[1,*]

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

[2]Institut Laue Langevin, 38000-Grenoble, France

[3]LAAS-CNRS, Université de Toulouse, CNRS, 31400, France

[4]Institut François Jacob, Molecular Imaging Center (MIRCen), Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA) and Laboratory of Neurodegenerative Diseases, Centre National de la Recherche Scientifique (CNRS), Université Paris-Saclay. CEA-Fontenay-aux-Roses Bâtiment 61. 18, route du Panorama. 92265 Fontenay-aux-Roses cedex, France.

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

Summary

Huntington's disease neurodegeneration occurs when the number of consecutive glutamines in the huntingtin Exon-1 (HTTExon1) exceeds a pathological threshold of 35. The sequence homogeneity of HTTExon1 reduces the NMR signal dispersion spectra, hampering its structural characterization. By simultaneously introducing three isotopically labelled glutamines in a site-specific manner in multiple concatenated samples, eighteen glutamines of a pathogenic HTTExon1 with 36 glutamines were unambiguously assigned. Chemical shifts analysis indicates the α-helical persistence in the homorepeat and the absence of an emerging toxic conformation around the pathological threshold. Using the same type of samples, the recognition mechanism of Hsc70 molecular chaperone has been investigated, indicating that it binds to the N17 region of HTTExon1, inducing the partial unfolding of the poly-Q. The proposed strategy facilitates high-resolution structural and functional studies in low-complexity regions.

Keywords: Huntingtin, Huntington's disease, Nuclear Magnetic Resonance, Poly-Q, tRNA suppression, isotopic labelling.

Introduction

Huntington's Disease (HD), a deadly neurodegenerative pathology, is triggered by the expansion of the poly-glutamine (poly-Q) tract located within the first exon of the huntingtin protein (HTTExon1)[1,2]. HD manifests when the number of consecutive glutamines in this tract exceeds 35, the so-called pathological threshold. Expanded HTTExon1 forms cytoplasmic and nuclear inclusions in neurons of the striatum, which trigger multiple neurological and behavioural symptoms. Neuronal degeneration, age of onset and disease severity all correlate with the poly-Q tract length[1,2]. In addition to HD, eight other degenerative diseases have been connected to poly-Q expansion within distinct proteins[3–5]. Although they share several features, variations in their pathological threshold have been observed.

Describing the structural bases of HTTExon1 function and the mechanisms leading to cytotoxicity remains an enormous challenge. Its intrinsic flexibility leaves Nuclear Magnetic Resonance (NMR) as the only adapted high-resolution structural biology technique that can be applied for its characterization. However, the strong bias in the amino acid composition of HTTExon1, including the poly-Q and two poly-proline (poly-P) stretches (Fig. 1A), causes a high degeneracy of the resulting chemical shifts, hampering the application of standard NMR strategies even when high magnetic fields are used[6]. Despite these difficulties, several high-resolution NMR studies of non-pathogenic HTTExon1 and N-terminal fragments of the protein have been reported[7–11]. However, the complete assignment of the homorepeat has only been achieved for a HTTExon1 N-terminal fragment with 17 glutamines[7], while for the other studies only glutamine residues located at the poly-Q stretch flanks could be unambiguously assigned.

In order to circumvent the limitations of traditional approaches when addressing Low Complexity Regions (LCRs), our group has developed a novel chemical biology technique allowing the site-specific isotopic labelling (SSIL) of glutamines and prolines[12,13]. In SSIL, a single [^{15}N and ^{13}C] isotopically labelled amino acid is incorporated into the protein using an externally loaded nonsense tRNA ($tRNA_{CUA}$)[14–16]. Protein synthesis is performed using an *Escherichia coli*-based Cell-Free (CF) reaction in the presence of a plasmid in which an amber stop codon (TAG) is placed at the position of interest. Moreover, CF enables the control of the isotopic labelling scheme applied to the rest of the protein[17,18]. SSIL has been proven to be very useful for the unambiguous NMR assignment of the poly-Q tracts in sub-pathogenic and pathogenic HTTExon1 containing 16 (H16) and 46 (H46) consecutive

glutamine residues, respectively[19,20]. Frequencies for backbone and side-chain glutamine atoms could be identified and structurally interpreted. Despite the success of SSIL in disentangling the complexity of crowded NMR spectra, this approach remains laborious, requiring a large number of samples, one per residue to be assigned, and highly demanding in terms of NMR time. Moreover, analyses requiring probing several sites of the protein, such as dynamic or interaction studies are severely hampered by the need for producing multiple samples. Here, we report our efforts to make the assignment and interaction studies of LCRs more efficient. The proposed methodology, which we name multi-SSIL (m-SSIL), is based on the simultaneous labelling of three different positions of the protein. We demonstrate that the careful design of polypeptide variants permits a faster unambiguous NMR assignment and facilitates the quantitative interpretation of protein-protein interaction NMR experiments. The power of the approach has been validated using HTTExon1 with a 36 glutamine long stretch fused to super-folded GFP (H36) (Fig. 1B). The poly-Q stretch length of H36, close to the pathological threshold of 35, makes this HTTExon1 form relevant to identify putative emerging toxic conformations. The interaction of the chaperone Hsc70 with H36 has been subsequently investigated in a site-specific manner, revealing its mechanism to reduce the aggregation propensity of HTTExon1.

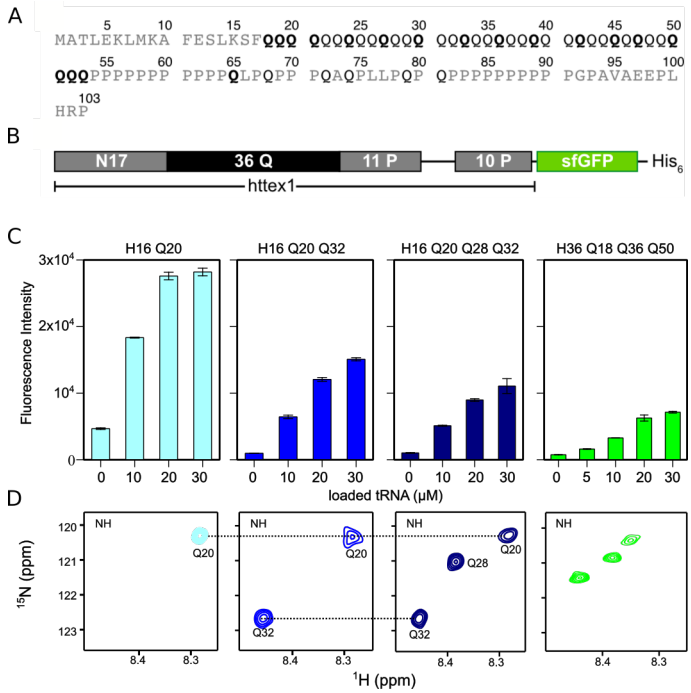

Figure 1. Primary structure and schematic representation of HTTExon1 with 36 consecutive glutamines (H36) and validation of the multi-SSIL strategy. ***(A)*** *Primary structure of H36. The glutamines labelled in this study are shown in black.* ***(B)*** *Schematic representation of the HTTExon1 used in this study, comprising the N-terminal domain N17, a poly-Q stretch and the proline-rich region (PRR) fused to superfolder GFP (sfGFP) and a C-terminal hexa-histidine tag.* ***(C)*** *Fluorescence for the titration of CF reactions with increasing concentrations of [^{15}N,^{13}C]-Gln-tRNA$_{CUA}$ for the expression of single (Q20, cyan), double (Q20-Q32, blue) and triple (Q20-Q28-Q32, dark blue) suppression constructs of H16 and a triple suppression mutant of H36 (Q18-Q36-Q50, green).* ***(D)*** *Zoom of the ^{15}N-HSQC spectra of SSIL constructs described in panel (C), showing the backbone NH signals. Dotted lines indicate the position of Q20 and Q32 in the nitrogen dimension through the spectra.*

Results and Discussion

Validation of multi-Site Specific Isotopic Labelling

The efficient introduction of several chemically modified amino acids has been previously demonstrated *in vivo* and in CF in the presence of the corresponding amino-acyl tRNA synthetase (aaRS)[21–24]. Note that, in order to avoid the incorporation of non-isotopically labelled amino acids present in the reaction mixture, the enzyme cannot be added to our CF reactions. We first evaluated the feasibility of triple-suppression under CF expression conditions. Starting with H16, we gradually increased the number of amber codons in the target gene from one (Q20), to two (Q20-Q32) and three (Q20-Q28-Q32). Monitoring the fluorescence intensity, the CF reactions were titrated with increasing amounts of Gln-

$tRNA_{CUA}$ (Fig. 1C). While all constructs showed a Gln-$tRNA_{CUA}$ concentration dependent increase in fluorescence intensity, a systematic decrease (≈30-40%) in the yield was observed when increasing the number of stop codons. This reduction was slightly higher than those observed in previous studies where the aaRS was present in the bacterial expression[22,24]. Importantly, the H36 triple mutant (Q18-Q36-Q50) followed the same trends as H16 (Fig. 1C). Taking into account the results of the titration experiments, large-scale productions (5 mL of CF) with 20 μM of Gln-$tRNA_{CUA}$ were implemented to produce NMR samples of the same constructs. The recorded ^{15}N-HSQC spectra of our single and multiple suppression mutants are shown in Figure 1D. For all H16 and H36 suppression mutants, the expected number of peaks was detected and the signals were well dispersed, overcoming the severe signal overlap observed for the homogeneously labelled protein (Fig. S1). Furthermore, the frequencies for Q20 and Q32 of H16 appeared in the same position in the respective spectra.

NMR assignment and structural investigation of H36

With the above-described proof-of-principle, the following strategy was designed to speed up the NMR assignment of the H36 poly-Q stretch. Nine triple suppression mutants were concatenated in a way that each construct shared one of its labelled positions with another one. When overlaying two concatenated spectra, the overlapped resonances unambiguously identified the common residue. Thus, in each spectrum, two of the peaks could be straightforwardly connected with two other spectra, while the third one remained unconnected, making its assignment also unambiguous (Fig. 2). In this way, the same number of residues could be assigned using half of the samples, reducing production costs and the required NMR time. In order to design the suppression mutants, we followed a strategy consisting in dividing the poly-Q in three sections, N-terminal, central and C-terminal, and selecting one residue from each region in each mutant. This approach, which seems the most generally applicable, exploits the expected different chemical environment provided by the flanking regions of the homorepeat. Note that, in the presence of partial structuration, an enhancement of the chemical shift dispersion is expected, making our strategy less sensitive to the specific design of the mutants. Following this strategy, atomic frequencies for 18 glutamines of H36 (17 from the poly-Q tract and Q65) were assigned. The capacity of this strategy to assign the H36 poly-Q tract is exemplified for NH, Cα-Hα and NεHε in Figs. 2 and S2. While this approach worked well for NH and Cα-Hα, some overlap was observed for NεHε correlations. Concretely, Q52-Q33-Q19 only displayed two pairs of NεHε peaks, one

corresponding to Q52 and the other to Q19/Q33. A distinct scenario was observed for the Cβ-Hβ and Cγ-Hγ correlations, where multiple peaks appeared in a very narrow range of frequencies, rendering the assignment by concatenating experiments unfeasible (Fig. S3).

With the successful assignment of the H36 poly-Q tract, a Secondary Chemical Shift (SCS) analysis was performed by comparing experimental and neighbour-corrected random coil Cα chemical shifts[25] (Fig. 3A). Note that Cβ data could not be used due to the reduced dispersion of our triple mutant spectra (Fig. S3). The SCS analysis revealed that H36 presents an important α-helical propensity for those glutamines close to N17 that smoothly decreases when approaching the PRR (Fig. 3A). Interestingly, H36 presents conformational trends that lie between those previously found for H16[19] and H46[20]. The SCS values measured for the last glutamines of the H36 poly-Q stretch (Q50 to Q53), which report on the extent of the conformational influence of the PRR, were equivalent to those measured for H16 and H46 (Fig. S4). This observation strongly indicated that the helix-breaking capacity exerted by the PRR was not influenced by the poly-Q stretch length.

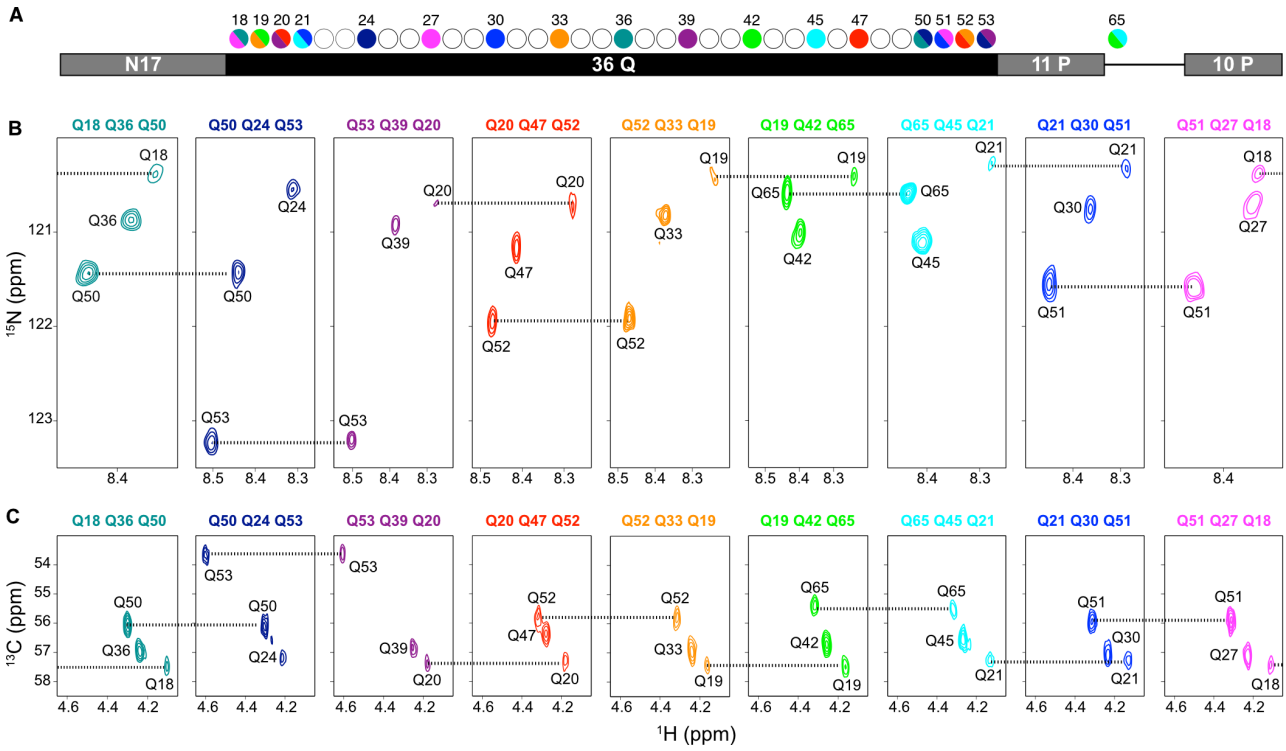

Figure 2. NMR assignment of H36 with m-SSIL samples. ***(A)*** *Scheme of H36 where the circles on top indicate the residues that were probed by NMR. The colours of the filled circles indicate in which of the spectra the residue can be found. Zooms of the* ^{15}*N-HSQCs* ***(B)*** *and* ^{13}*C-HSQCs* ***(C)*** *of the triple-labelled H36 samples. The identity of the triplet is given at the top of each panel. The dashed lines indicate peaks with equivalent frequencies that are*

present in two concatenated spectra. The third peak in each spectrum is unique and can be assigned by exclusion.

The experimental Cα chemical shifts were used to derive a structural ensemble model for H36. To this aim and using a previously described methodology based on three-residue fragments extracted from high-resolution structures[26], we built two families of sub-ensembles in which the amount of structuration was systematically increased from residue F17 to Q53 (N→C ensembles) and from Q53 to Q18 (N←C ensembles) (see Supplementary information for details). Note that this building strategy propagates the conformational preferences of both flanking regions towards the poly-Q, emphasizing the previously described structural asymmetry of HTTExon1[19]. After computing the theoretical average Cα chemical shifts for each one of the sub-ensembles with SPARTA+[27], the relative population of each one of the sub-ensembles was optimized using the experimental Cα chemical shifts. Six sub-ensembles of the N→C family, which propagate helicity from N17 towards the poly-Q tract, were enough to obtain an excellent agreement with the experimental data (Fig. 3B). The selected sub-ensembles combined α-helical segments of different lengths with disordered sections. Interestingly, the most populated sub-ensemble (69.7%) presented a fully disordered poly-Q tract. A quantitative structural picture of the optimized ensemble was obtained using the SS-map representation (Fig. 3C)[28]. This analysis indicated that the overall helicity observed in H36 was consistent with multiple partially helical conformations encompassing the N17 and different sections of the poly-Q. Interestingly, long helices incorporating more than 30 glutamines were slightly overpopulated, suggesting a cooperative stabilization of this structural element with the number of glutamines. This observation is in line with previous estimations based on the thermodynamic analyses of melting temperatures for HTTExon1 with an increasing number of glutamines[10].

We next analysed whether N17 exerted a conformational influence in H36 poly-Q stretch. Others and we have previously showed that the α-helical N17 structure spreads towards the poly-Q tract through a network of bifurcate hydrogen bonds[19,20,29]. This structural property manifests in H16 and H46 by a Cβ-Hβ singlet for Q18 and Q19, and two Cγ-Hγ frequencies for Q20 and Q21. Unfortunately, these spectroscopic features could not be straightforwardly identified for H36 triple suppression mutants due to the severe overlap. In order to disentangle this complexity, we overlaid the Cβ-Hβ and Cγ-Hγ spectra of m-SSIL H36 samples with those measured for the corresponding residues in H46 (Fig. S3B). An excellent agreement between both sets of peaks was observed, demonstrating that the structural features defining

the poly-Q structure in H46 are reproduced in H36. Concretely, this comparison substantiated the presence of *i*->*i*+4 bifurcate hydrogen bonds between the N17 and the first residues of the poly-Q stretch and the presence of two conformational states in slow exchange. All together, our NMR analyses proved the cumulative α-helical propensity upon poly-Q length increase and confirmed the absence of sharp conformational changes when surpassing the pathological threshold.

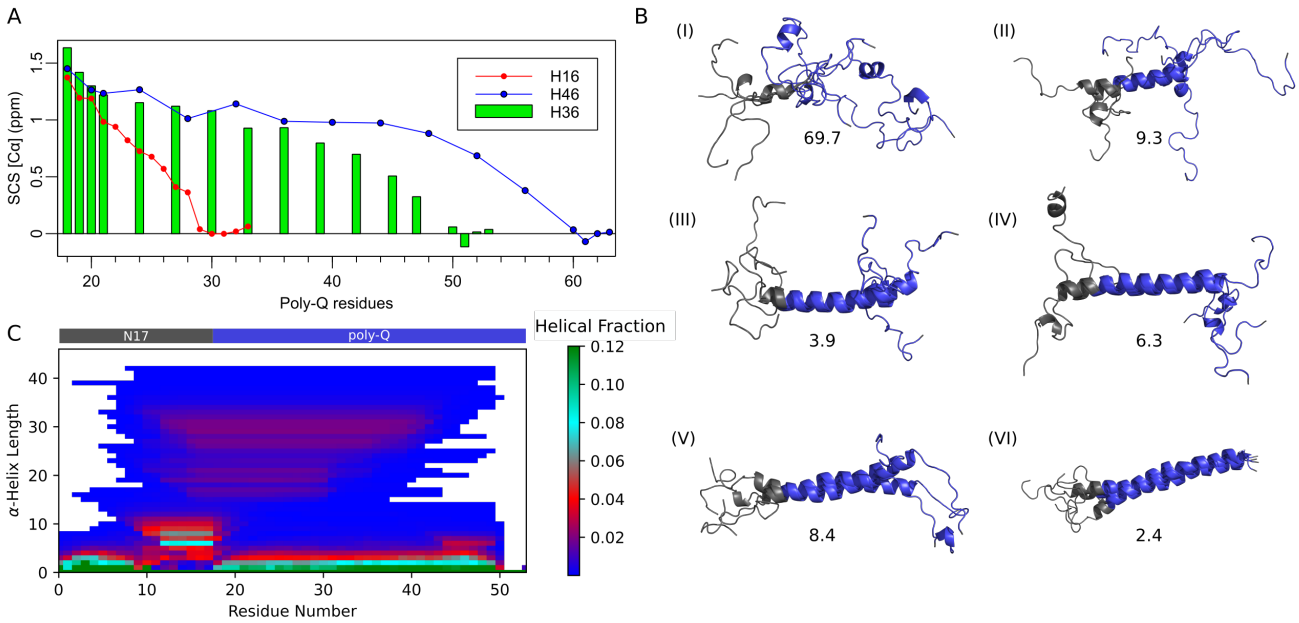

Figure 3. Structural analyses of H36. ***(A)*** *Cα Secondary Chemical Shift (SCS) profile of H36 (green bars) in comparison with these obtained for H16 (red line with dots) and H46 (blue line with dots). Due to the severe overlap of the Cβ signals, only Cα data was used in the SCS calculation.* ***(B)*** *Five representative structures and populations of the six N→C family ensembles selected to fit the experimental Cα chemical shifts. N17 and the poly-Q tract are displayed in grey and blue, respectively.* ***(C)*** *SS-map computed for the final ensemble where the helical length, the residues involved and the population are represented.*

Recognition mechanism of huntingtin by chaperone Hsc70

We used m-SSIL samples to investigate the recognition mechanism of H36 by the constitutive molecular chaperone Hsc70, which belongs to the heat shock protein 70 (Hsp70) family. Multiple studies have revealed that molecular chaperones modulate and reverse the aggregation of proteins bearing poly-Q stretches both *in vitro* and *in vivo*[30–35]. We previously demonstrated that human Hsc70 affects monomeric HTTExon1 aggregation in a poly-Q independent manner by binding N17[36]. Hsc70 slowed down the aggregation kinetics but did not affect the amount of fibrillar HTTExon1 at steady state, suggesting that it transiently sequesters monomeric HTTExon1. We further identified the interacting interfaces in Hsc70

and N17 through cross-linking studies and mass spectrometry. However, how the binding of Hsc70 to N17 affects the structure of the poly-Q remains unknown. To decipher the binding mechanism of Hsc70 to HTTExon1, we titrated two triple mutants of H36 with increasing amounts of Hsc70 (Fig. 4A and S6). In addition to the three mutated sites, all residues with the exception of glutamine and glutamic acid were ^{15}N-labelled in these samples. Peak intensities measured for three titration points of the two samples indicated a highly heterogeneous behaviour (Fig. 4B and S7). Residues located within N17 exhibited an important intensity decrease already at the lowest H36 to Hsc70 ratio (1:0.5) and many of them disappeared at the highest ratio (1:1.5). This behaviour is illustrated for K6, M8 and F11 in Fig. 4A. Conversely, residues located at the PRR appeared almost unaffected throughout the titration as occurs when Hsc70 acts in conjunction with DNAJB1[34]. Importantly, a small signal intensity decrease was observed for some residues of the linker connecting the HTTExon-1 and the sfGFP (data not shown). This interaction, which is probably driven by the hydrophobic nature of the linker, strongly suggests that the fusion protein does not screen the interaction of Hsc70 with the C-terminal region of HTTExon-1. Conversely to the PRR, position dependent intensities were observed for the six glutamines from the poly-Q stretch. The resonance intensities from those close to N17 (Q18 and Q21) strongly decreased, especially in the 1:1.5 ratio. The resonances of the other glutamines (Q30, Q36, Q50 and Q51) were systematically less affected by the addition of Hsc70. The gradual effect from N- to C-termini suggested that Hsc70 specifically interacts with and affects the conformation of N17. Upon binding N17, the molecular chaperone perturbs the hydrodynamic behaviour of the rest of HTTExon1 in a position dependent manner, even in the absence of direct interaction.

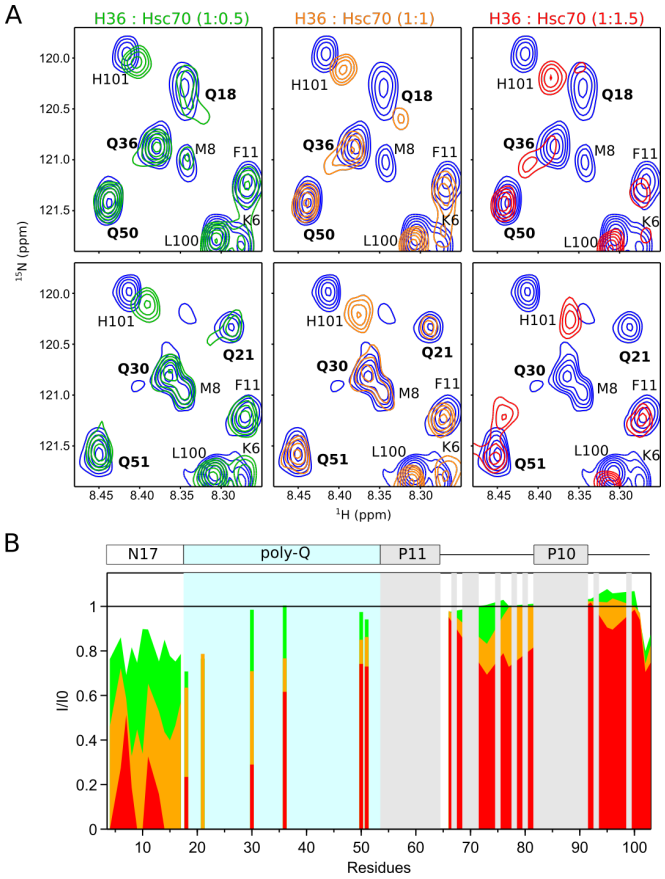

Figure 4. *Interaction between H36 and Hsc70.* ***(A)*** *Zoom of the ^{15}N-HSQC spectra showing the H36 poly-Q region upon the addition of increasing amounts of Hsc70. Two H36 triple mutants were used to monitor the interaction (Q18-Q36-Q50 in upper panels, and Q21-Q30-Q51 in lower panels).* ***(B)*** *Averaged intensity ratios of H36 triple mutant peaks with increasing amounts of Hsc70. Green, orange and red correspond to H36:Hsc70 of 1:0.5, 1:1 and 1:1.5 ratios, respectively. Light blue and grey indicate the poly-Q region and prolines, respectively.*

No systematic chemical shift changes were observed for the vast majority of the N17 and PRR resonances throughout the titration experiment. Residue H101, most probably due to the sensitivity of histidine residues to small pH changes, represented an exception to this behaviour (Fig. 4A). A different scenario was observed for some of the monitored glutamines. For Q18, a minor second signal appeared at 1:0.5 ratio that became the only one observed at 1:1, before vanishing at 1:1.5 (Fig. 4A). This peak behaviour was not observed in the neighbouring Q21, which experienced a systematic intensity decrease with increasing amounts of Hsc70. Similarly to Q18, the appearance of a second signal with a concomitant intensity decrease of the initial one was observed for Q36. When increasing the Hsc70 ratio to 1:2, the second signal was the only one observed for this residue (Fig. S6). The appearance of a second signal was also detected for Q30, although it was only observed at a 1:1.5 ratio.

Interestingly, the smaller frequencies of the two emerging signals for Q30 and Q36 indicated a reduced helical content in the poly-Q stretch of H36 upon Hsc70 binding. These observations suggest that the interaction of Hsc70 with N17 unfolds the poly-Q tract most probably breaking the hydrogen bond network connecting N17 with the first glutamine residues of the stretch[19,29]. Recently, the length and stability of α-helical conformations in HTTExon1 poly-Q stretch have been reported to play a key role in HTTExon1 aggregation kinetics[20]. Thus, our observations suggests that the Hsc70-mediated HTTExon1 aggregation slowdown arises from the α-helix melting of the poly-Q stretch, which becomes less prone to self-interact and trigger aggregation.

Concluding Remarks

In summary, we have developed a novel chemical biology tool that facilitates the high-resolution characterization of LCRs, a family of proteins that have remained out of the reach for structural biology approaches[37]. We expect that m-SSIL will unveil the structural bases of relevant biological phenomena linked to LCRs. At present, only non-sense suppression for glutamine and proline has been developed. However, the extension of the SSIL strategy to other amino acids will enable structural and functional studies of other systems, such as the emerging phenomenon of LCR-driven liquid-liquid phase separation[38,39], alanine expansion diseases[40] and large macromolecular machines[41].

Acknowledgements

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 and by the EU Joint Program on Neurodegenerative disease Research and ANR (Trans-PathND, ANR-17-JPND-0002-02) awarded to RM. The CBS is a member of France-BioImaging (FBI) and the French Infrastructure for Integrated Structural Biology (FRISBI), ANR-10-INBS-04-01 and ANR-10-INBS-05, respectively.

Author Contributions

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

Declaration of interest

The authors declare no conflict of interest.

References

1. Saudou, F. & Humbert, S. (2016). The Biology of Huntingtin. *Neuron*. **89**, 910–926.
2. Walker, F. O. (2007). Huntington's Disease. *Lancet*. **369**, 218–228.
3. Silva, A., de Almeida, A. V. & Macedo-Ribeiro, S. (2018). Polyglutamine Expansion Diseases: More than Simple Repeats. *J. Struct. Biol.* **201**, 139–154.
4. Orr, H. T. (2001). Beyond the Qs in the Polyglutamine Diseases. *Genes Dev.* **15**, 925–932.
5. Orr, H. T. & Zoghbi, H. Y. (2007). Trinucleotide Repeat Disorders. *Annu. Rev. Neurosci.* **30**, 575–621.
6. Urbanek, A., Elena-Real, C. A., Popovic, M., Morató, A., Fournet, A., Allemand, F., Delbecq, S., Sibille, N. & Bernadó, P. (2020). Site-Specific Isotopic Labeling (SSIL): Access to High-Resolution Structural and Dynamic Information in Low-Complexity Proteins. *ChemBioChem*. **21**, 769–775.
7. Baias, M., Smith, P. E. S., Shen, K., Joachimiak, L. A., Zerko, S., Koźmiński, W., Frydman, J. & Frydman, L. (2017). Structure and Dynamics of the Huntingtin Exon-1 N-Terminus: A Solution NMR Perspective. *J. Am. Chem. Soc.* **139**, 1168–1176.
8. 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. (2018). Tadpole-like Conformations of Huntingtin Exon 1 Are Characterized by Conformational Heterogeneity That Persists Regardless of Polyglutamine Length. *J. Mol. Biol.* **430**, 1442–1458.
9. 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., M Gronenborn, A. & Wetzel, R. (2009). Polyglutamine Disruption of the Huntingtin Exon 1 N Terminus Triggers a Complex Aggregation Mechanism. *Nat. Struct. &Amp; Mol. Biol.* **16**, 380.
10. Bravo-Arredondo, J. M., Kegulian, N. C., Schmidt, T., Pandey, N. K., Situ, A. J., Ulmer, T. S. & Langen, R. (2018). The Folding Equilibrium of Huntingtin Exon-1 Monomer Depends on Its

Polyglutamine Tract. *J. Biol. Chem.* jbc.RA118.004808.

11. Ceccon, A., Schmidt, T., Tugarinov, V., Kotler, S. A., Schwieters, C. D. & Clore, G. M. (2018). Interaction of Huntingtin Exon-1 Peptides with Lipid-Based Micellar Nanoparticles Probed by Solution NMR and Q-Band Pulsed EPR. *J. Am. Chem. Soc.* **140**, 6199–6202.
12. Urbanek, A., Morató, A., Allemand, F., Delaforge, E., Fournet, A., Popovic, M., Delbecq, S., Sibille, N. & Bernadó, P. (2018). A General Strategy to Access Structural Information at Atomic Resolution in Polyglutamine Homorepeats. *Angew. Chem. Int. Ed. Engl.* **57**, 3598–3601.
13. Urbanek, A., Popovic, M., Elena-Real, C. A., Morató, A., Estaña, A., Fournet, A., Allemand, F., Gil, A. M., Cativiela, C., Cortés, J., Jiménez, A. I., Sibille, N. & Bernadó, P. (2020). Evidence of the Reduced Abundance of Proline Cis Conformation in Protein Poly Proline Tracts. *J. Am. Chem. Soc.* **142**, 7976–7986.
14. Ellman, J. A., Volkman, B. F., Mendel, D., Schulz, P. G. & Wemmer, D. E. (1992). Site-Specific Isotopic Labeling of Proteins for NMR Studies. *J. Am. Chem. Soc.* **114**, 7959–7961.
15. Peuker, S., Andersson, H., Gustavsson, E., Maiti, K. S., Kania, R., Karim, A., Niebling, S., Pedersen, A., Erdelyi, M. & Westenhoff, S. (2016). Efficient Isotope Editing of Proteins for Site-Directed Vibrational Spectroscopy. *J. Am. Chem. Soc.* **138**, 2312–2318.
16. Yabuki, T., Kigawa, T., Dohmae, N., Takio, K., Terada, T., Ito, Y., Laue, E. D., Cooper, J. A., Kainosho, M. & Yokoyama, S. (1998). Dual Amino Acid-Selective and Site-Directed Stable-Isotope Labeling of the Human c-Ha-Ras Protein by Cell-Free Synthesis. *J. Biomol. NMR*. **11**, 295–306.
17. Ozawa, K., Wu, P. S. C., Dixon, N. E. & Otting, G. (2006). N-Labelled Proteins by Cell-Free Protein Synthesis. Strategies for High-Throughput NMR Studies of Proteins and Protein-Ligand Complexes. *FEBS J.* **273**, 4154–4159.
18. Morató, A., Elena-Real, C. A., Popovic, M., Fournet, A., Zhang, K., Allemand, F., Sibille, N., Urbanek, A. & Bernadó, P. (2020). 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*. **10**.
19. Urbanek, A., Popovic, M., Morató, A., Estaña, A., Elena-Real, C. A., Mier, P., Fournet, A., Allemand, F., Delbecq, S., Andrade-Navarro, M. A., Cortés, J., Sibille, N. & Bernadó, P. (2020). Flanking Regions Determine the Structure of the Poly-Glutamine in Huntingtin through Mechanisms Common among Glutamine-Rich Human Proteins. *Structure*. **28**, 733-746.e5.
20. Elena-Real, C. A., Sagar, A., Urbanek, A., Popovic, M., Morató, A., Estaña, A., Fournet, A., Doucet, C., Lund, X. L., Shi, Z.-D., Costa, L., Thureau, A., Allemand, F., Swenson, R. E., Milhiet,

P.-E., Crehuet, R., Barducci, A., Cortés, J., Sinnaeve, D., Sibille, N. & Bernadó, P. (2023). The Structure of Pathogenic Huntingtin Exon 1 Defines the Bases of Its Aggregation Propensity. *Nat. Struct. Mol. Biol., in press.*

21. Loscha, K. V, Herlt, A. J., Qi, R., Huber, T., Ozawa, K. & Otting, G. (2012). Multiple-Site Labeling of Proteins with Unnatural Amino Acids. *Angew. Chem. Int. Ed. Engl.* **51**, 2243–2246.
22. Martin, R. W., Des Soye, B. J., Kwon, Y., Kay, J., Davis, R. G., Thomas, P. M., Majewska, N. I., Chen, C. X., Marcum, R. D., Weiss, M. G., Stoddart, A. E., Amiram, M., Ranji Charna, A. K., Patel, J. R., Isaacs, F. J., Kelleher, N. L., Hong, S. H. & Jewett, M. C. (2018). Cell-Free Protein Synthesis from Genomically Recoded Bacteria Enables Multisite Incorporation of Noncanonical Amino Acids. *Nat. Commun.* **9**, 1203.
23. Orton, H. W., Qianzhu, H., Abdelkader, E. H., Habel, E. I., Tan, Y. J., Frkic, R. L., Jackson, C. J., Huber, T. & Otting, G. (2021). Through-Space Scalar 19F–19F Couplings between Fluorinated Noncanonical Amino Acids for the Detection of Specific Contacts in Proteins. *J. Am. Chem. Soc.* **143**, 19587–19598.
24. Liu, Y., Davis, R. G., Thomas, P. M., Kelleher, N. L. & Jewett, M. C. (2021). In Vitro-Constructed Ribosomes Enable Multi-Site Incorporation of Noncanonical Amino Acids into Proteins. *Biochemistry*. **60**, 161–169.
25. Nielsen, J. T. & Mulder, F. A. A. (2018). POTENCI: Prediction of Temperature, Neighbor and PH-Corrected Chemical Shifts for Intrinsically Disordered Proteins. *J. Biomol. NMR*. **70**, 141–165.
26. Estaña, A., Sibille, N., Delaforge, E., Vaisset, M., Cortés, J. & Bernadó, P. (2019). Realistic Ensemble Models of Intrinsically Disordered Proteins Using a Structure-Encoding Coil Database. *Structure*. **27**, 381–391.
27. Shen, Y. & Bax, A. (2010). SPARTA+: A Modest Improvement in Empirical NMR Chemical Shift Prediction by Means of an Artificial Neural Network. *J. Biomol. NMR*. **48**, 13–22.
28. Iglesias, J., Sanchez-Martínez, M. & Crehuet, R. (2013). SS-Map: Visualizing Cooperative Secondary Structure Elements in Protein Ensembles. *Intrinsically Disord. Proteins*. **1**, e25323.
29. Escobedo, A., Topal, B., Kunze, M. B. A., Aranda, J., Chiesa, G., Mungianu, D., Bernardo-Seisdedos, G., Eftekharzadeh, B., Gairí, M., Pierattelli, R., Felli, I. C., Diercks, T., Millet, O., García, J., Orozco, M., Crehuet, R., Lindorff-Larsen, K. & Salvatella, X. (2019). Side Chain to Main Chain Hydrogen Bonds Stabilize a Polyglutamine Helix in a Transcription Factor. *Nat. Commun.* **10**, 2034.
30. Hay, D. G., Sathasivam, K., Tobaben, S., Stahl, B., Marber, M., Mestril, R., Mahal, A., Smith, D. L., Woodman, B. & Bates, G. P. (2004). Progressive Decrease in Chaperone Protein Levels in a

Mouse Model of Huntington's Disease and Induction of Stress Proteins as a Therapeutic Approach. *Hum. Mol. Genet.* **13**, 1389–1405.

31. Scior, A., Buntru, A., Arnsburg, K., Ast, A., Iburg, M., Juenemann, K., Pigazzini, M. L., Mlody, B., Puchkov, D., Priller, J., Wanker, E. E., Prigione, A. & Kirstein, J. (2018). Complete Suppression of Htt Fibrilization and Disaggregation of Htt Fibrils by a Trimeric Chaperone Complex. *EMBO J.* **37**, 282–299.
32. Kuiper, E. F. E., de Mattos, E. P., Jardim, L. B., Kampinga, H. H. & Bergink, S. (2017). Chaperones in Polyglutamine Aggregation: Beyond the Q-Stretch. *Front. Neurosci.* **11**, 145.
33. Chan, H. Y., Warrick, J. M., Gray-Board, G. L., Paulson, H. L. & Bonini, N. M. (2000). Mechanisms of Chaperone Suppression of Polyglutamine Disease: Selectivity, Synergy and Modulation of Protein Solubility in Drosophila. *Hum. Mol. Genet.* **9**, 2811–2820.
34. Ayala Mariscal, S. M., Pigazzini, M. L., Richter, Y., Özel, M., Grothaus, I. L., Protze, J., Ziege, K., Kulke, M., ElBediwi, M., Vermaas, J. V, Colombi Ciacchi, L., Köppen, S., Liu, F. & Kirstein, J. (2022). Identification of a HTT-Specific Binding Motif in DNAJB1 Essential for Suppression and Disaggregation of HTT. *Nat. Commun.* **13**, 4692.
35. Tam, S., Spiess, C., Auyeung, W., Joachimiak, L., Chen, B., Poirier, M. A. & Frydman, J. (2009). The Chaperonin TRiC Blocks a Huntingtin Sequence Element That Promotes the Conformational Switch to Aggregation. *Nat. Struct. Mol. Biol.* **16**, 1279–1285.
36. Monsellier, E., Redeker, V., Ruiz-Arlandis, G., Bousset, L. & Melki, R. (2015). Molecular Interaction between the Chaperone Hsc70 and the N-Terminal Flank of Huntingtin Exon 1 Modulates Aggregation. *J. Biol. Chem.* **290**, 2560–2576.
37. Mier, P., Paladin, L., Tamana, S., Petrosian, S., Hajdu-Soltész, B., Urbanek, A., Gruca, A., Plewczynski, D., Grynberg, M., Bernadó, P., Gáspári, Z., Ouzounis, C. A., Promponas, V. J., Kajava, A. V, Hancock, J. M., Tosatto, S. C. E., Dosztanyi, Z. & Andrade-Navarro, M. A. (2020). Disentangling the Complexity of Low Complexity Proteins. *Brief. Bioinform.* **21**, 458–472.
38. Bremer, A., Farag, M., Borcherds, W. M., Peran, I., Martin, E. W., Pappu, R. V & Mittag, T. (2022). Deciphering How Naturally Occurring Sequence Features Impact the Phase Behaviours of Disordered Prion-like Domains. *Nat. Chem.* **14**, 196–207.
39. Franzmann, T. M. & Alberti, S. (2019). Prion-like Low-Complexity Sequences: Key Regulators of Protein Solubility and Phase Behavior. *J. Biol. Chem.* **294**, 7128–7136.
40. Amiel, J., Trochet, D., Clément-Ziza, M., Munnich, A. & Lyonnet, S. (2004). Polyalanine Expansions in Human. *Hum. Mol. Genet.* **13**, R235–R243.
41. Gauto, D. F., Estrozi, L. F., Schwieters, C. D., Effantin, G., Macek, P., Sounier, R., Sivertsen, A.

C., Schmidt, E., Kerfah, R., Mas, G., Colletier, J.-P., Güntert, P., Favier, A., Schoehn, G., Schanda, P. & Boisbouvier, J. (2019). Integrated NMR and Cryo-EM Atomic-Resolution Structure Determination of a Half-Megadalton Enzyme Complex. *Nat. Commun.* **10**, 2697.