

# Sequence Context Influences the Structure and Aggregation Behavior of a PolyQ Tract

B. Eftekharzadeh<sup>□,1</sup> A. Piai<sup>□,2</sup> G. Chiesa,<sup>1</sup> D. Mungianu,<sup>1</sup> J. García,<sup>1</sup> R. Pierattelli,<sup>2</sup> I. C. Felli,<sup>2</sup> and X. Salvatella<sup>1,3,\*</sup>

<sup>1</sup>Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldori Reixac 10, 08028 Barcelona, Spain

<sup>2</sup>CERM and Department of Chemistry "Ugo Schiff", University of Florence  
Via Luigi Sacconi 6, 50019 Sesto Fiorentino (Florence), Italy

<sup>3</sup>ICREA, Barcelona, Spain.

\*to whom correspondence should be addressed  
e-mail: [xavier.salvatella@irbbarcelona.org](mailto:xavier.salvatella@irbbarcelona.org)  
tel: +34 934020459

## Abstract

Expansions of polyglutamine (polyQ) tracts in nine different proteins cause a family of neurodegenerative disorders called polyQ diseases. Since polyQ tracts are potential therapeutic targets for these pathologies there is great interest in characterizing the conformations that they adopt and in understanding how their aggregation behavior is influenced by the sequences flanking them. We used solution NMR to study at single-residue resolution a 156-residue proteolytic fragment of the androgen receptor that contains a polyQ tract associated with the disease called *spinobulbar muscular atrophy*, also known as *Kennedy disease*. Our findings indicate that a Leu-rich region preceding the polyQ tract causes it to become  $\alpha$ -helical and appears to protect the protein against aggregation, which represents a new mechanism by which sequence context can minimize the deleterious properties of these repetitive regions. Our results have implications for drug discovery for polyQ diseases because they suggest that the residues flanking these repetitive sequences may represent viable therapeutic targets.

## Introduction

A group of nine disorders termed *polyglutamine (polyQ) diseases* occur as a result of the expansion of polymorphic polyQ tracts in proteins that are otherwise unrelated (1). The variable length of such tracts is due to the propensity of the CAG and GTC codon repeats that codify for them to form non-B-DNA structures that cause slippage during DNA replication (2). As a consequence of these expansions, encoded proteins with expanded polyQ tracts are formed, which often misfold, oligomerize and aggregate to form fibrillar species resembling amyloid fibrils that may play a role in the onset of polyQ diseases (3–5).

Characterizing the structural and dynamical properties of polyQ tracts is crucial for understanding the molecular basis of polyQ diseases and for developing therapeutic strategies to inhibit their aggregation (6). However, these regions of low sequence complexity are challenging targets for conventional methods of structural biology. This is due, in addition to their low solubility, to their high propensity to be intrinsically disordered, that in general hinders crystallization, and to the highly repetitive nature of their amino acid sequence, that can hamper their investigation by NMR. A number of pioneering studies have characterized the structural and dynamic properties of polyQ tracts, but to date it has not been possible to report on the conformation of one such a tract in its native context and without fusing it to a solubilizing moiety (7, 8).

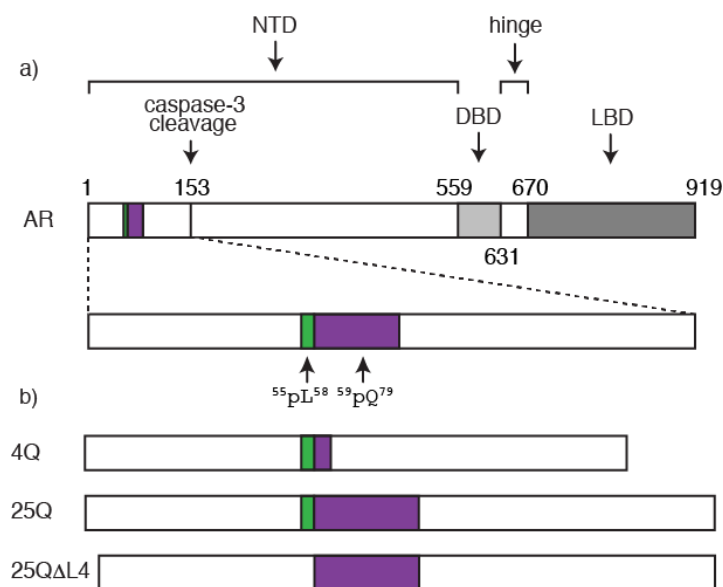


Figure 1. Structural organization of AR. a) Position of the transactivation (NTD), DNA binding (DBD) and ligand binding (LBD) domains of AR with an indication of the positions of the <sup>55</sup>LLLL<sup>58</sup> motif, shown in green, of the polymorphic polyQ tract, shown in purple, and of the caspase 3 cleavage site. b) AR constructs used in this work (see also Fig. S1).

We focus here on a polyQ tract occurring in the transactivation domain of the androgen receptor (AR) which plays a key role in the onset of *spinal bulbar muscular atrophy* (SBMA), a rare hereditary neuromuscular polyQ disease also known as *Kennedy disease* (9). AR is a nuclear receptor activated by androgens such as dihydrotestosterone that regulates the expression of the male phenotype (10). The

polymorphic polyQ tract in AR starts at position 59 and can be between 14 and 34 residues long in healthy individuals (Fig. 1a). Sizes over 37 residues are associated with SBMA and the length of the polyQ tract anticorrelates with the age of onset.

The polyQ tract of AR is involved in the formation of the aggregates associated with SBMA (11). However, neither its structure nor its role in the mechanism of AR aggregation are known and this acts as an important hurdle for the development of therapeutic approaches for this disease based on preventing or decreasing the rate of aggregation. To remedy this and, especially, to better understand how sequence context can influence the structural properties of polyQ tracts we used solution NMR to investigate the conformation of an N-terminal fragment of AR found in the aggregates associated with SBMA. This 156-residue fragment is the product of proteolytic cleavage by caspase 3 and plays a key role in the progression of the disease (Fig. 1a) (9).

## **Materials and Methods**

### Protein expression and purification

The genes codifying for 4Q, 25Q and 25QΔL (Figs. 1b and S1) were purchased from GeneArt and cloned in a pDEST-HisMBP vector (Addgene). Expression of the resulting genes led to fusion proteins containing a His<sub>6</sub> tag and a maltose binding protein (MBP) moiety that were used, respectively, to purify and increase the solubility of the proteins. <sup>15</sup>N labeled protein expression was carried in Rosetta *E. coli* cells grown in MOPS at 37°C until the value of OD<sub>600</sub> was 0.7 and induced with 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4 hours at 28°C. Cells were harvested by centrifugation and re-suspended in core buffer (50 mM sodium phosphate, 500 mM NaCl, 5% (v/v) Glycerol, 1 mM 2-Mercapto-ethanol, pH 8.0). A HisTrap HP 5 ml column (GE Healthcare) was used to purify the proteins, which were eluted by an imidazole gradient (final composition: 500 mM imidazole, 50 mM sodium phosphate, 500 mM NaCl, 5% Glycerol, 1 mM 2-mercapto-ethanol, pH 8.0), followed by a size exclusion chromatography step carried out in a Superdex HighLoad S200 26/60 column (GE Healthcare) equilibrated in a buffer with the following composition: 500 mM NaCl, 20 mM sodium phosphate, 5% (v/v) Glycerol, 1 mM DTT, pH 7.5. The pure proteins were then incubated with His<sub>6</sub>-tagged TEV protease for 16 hours at 4°C by dialysis against a buffer containing 20 mM sodium phosphate, 100 mM NaCl and 0.5 mM EDTA (Ethylenediaminetetraacetic acid), pH 8.0. The product of the proteolytic cleavage was purified by Ni<sup>2+</sup> affinity chromatography, employing a buffer containing 8 M urea (500 mM imidazole, 50 mM sodium phosphate, 100 mM NaCl, 8 M urea, pH 8.0) to prevent the aggregation of the cleaved AR in the column. Finally, the cleaved proteins were stored at -80°C.

### NMR sample preparation

Except for the experiments presented in fig. 4b the protein solutions stored at -80°C in 8M urea were thawed and dialyzed (1:1000) for 16 hours at 4°C against a buffer containing 20 mM sodium phosphate and 1 mM tris(2-carboxyethyl)phosphine (TCEP) at pH 7.4. Finally, 10% (v/v) D<sub>2</sub>O and 0.015 mM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) were added to the samples; the concentration of these NMR samples was 400 μM. The experiments presented in fig. 4b were carried out

with samples prepared with the procedure used for the CD and DLS samples (see below); the concentrations of these samples were 43  $\mu\text{M}$  (25Q $\Delta$ L4) and 50  $\mu\text{M}$  (25Q).

### DLS and CD sample preparation

The protocol was adapted from the one developed by Linse group for A $\beta$  peptide kinetic measurements (12 and Chiesa *et al* in preparation). After purification, the lyophilized protein was dissolved in a buffer containing 20 mM sodium phosphate, 100 mM NaCl, 6 M Guanidine Thiocyanate, pH 7.4 and 5 mM TCEP until complete reduction of the protein, as monitored by HPLC measurement with a C18 column. The solution was passed through a PD-10 column (GE Healthcare) equilibrated in 20 mM sodium phosphate, pH 7.4. The most concentrated fraction was then purified with a Superdex 75 10/300 size exclusion chromatography column and the fractions containing exclusively the monomeric protein were centrifuged at 386,000 g for 1 hour at 4°C, using an Optima MAX preparative ultracentrifuge (Beckman). Only the upper  $\frac{3}{4}$  of the centrifuged solutions were used for the assays.

### Dynamic light scattering (DLS)

Before the measurements, samples were centrifuged by using a tabletop centrifuge at 4°C for 10 minutes at 13,000 rpm. Measurements were taken with a Malvern Zetasizer Nano S equipped with a He-Ne of 633 nm wavelength laser. For each experiment, 20 repetitions of 20 seconds were recorded. Three measurements were performed at each time point. The concentration of the DLS samples was 20  $\mu\text{M}$  and the experiments were performed at 310K.

### Circular Dichroism (CD) spectroscopy

Far-UV CD measurements were performed on a JASCO 815 spectropolarimeter using a 0.01 cm length cuvette. Freshly prepares sample were diluted to the desired concentration with 20 mM sodium phosphate buffer (pH 7.4). The spectra were acquired at 0.2 nm resolution with a scan rate of 50 nm/min. For each sample, 10 spectra were collected and averaged, after subtraction the blank. The concentrations of the CD samples were 130  $\mu\text{M}$  for 25Q, 120 $\mu\text{M}$  for 4Q and 140 $\mu\text{M}$  for 25Q $\Delta$ L4 and the experiments were performed at 310K.

### NMR experiments

0.4 mM samples of  $^{13}\text{C}$ ,  $^{15}\text{N}$  double-labeled 4Q and 25Q in 20 mM sodium phosphate, 1 mM TCEP, pH 7.4 were prepared. 10 % (v/v) D $_2$ O was added for the lock. Identical samples, but exclusively enriched in  $^{15}\text{N}$ , were used to acquire  $^{15}\text{N}$  relaxation experiments.  $^{13}\text{C}$ -detected and  $^1\text{H}$ -detected NMR experiments for sequence-specific resonance assignment were acquired at 16.4 T on a Bruker Avance spectrometer operating at 700.06 MHz  $^1\text{H}$ , 176.03 MHz  $^{13}\text{C}$  and 70.94 MHz  $^{15}\text{N}$  frequencies, equipped with a cryogenically cooled probehead optimized for  $^{13}\text{C}$ -direct detection (TXO), and at 22.3 T on a Bruker Avance III spectrometer operating at 950.20 MHz  $^1\text{H}$ , 238.93 MHz  $^{13}\text{C}$  and 96.28 MHz  $^{15}\text{N}$  frequencies, equipped with a cryogenically cooled probehead (TCI).  $^{15}\text{N}$  relaxation experiments were performed at 16.4 T Bruker Avance spectrometer operating at 700.13 MHz  $^1\text{H}$ , 176.05 MHz  $^{13}\text{C}$  and 70.94 MHz  $^{15}\text{N}$  frequencies, equipped with a cryogenically cooled probehead (TXI), by

measuring  $^{15}\text{N}$  backbone longitudinal ( $R_1$ ) and transverse ( $R_2$ ) relaxation rates and the heteronuclear  $^{15}\text{N}\{^{15}\text{H}\}$  NOEs. The experiments presented in figure 4b were acquired at 18.8 T on a Bruker Digital Avance spectrometer operating at 800 MHz  $^1\text{H}$ , 201.20 MHz  $^{13}\text{C}$  and 81.08 MHz  $^{15}\text{N}$  frequencies, equipped with a cryogenically cooled probehead (TCI), with  $^{15}\text{N}$ -enriched samples. All the experiments were collected at 278 K.

A data set consisting of a combination of  $^{13}\text{C}$ -detected (4D HCBCACON (13), 4D HCBCANCO (13), 4D (HCA)CON(CA)CON (14) and 4D (HN)CON(CA)CON) (14) and  $^1\text{H}$ -detected (3D TROSY HNCO (15), 4D TROSY (H)NCO(CA)NNH (16) and 4D TROSY HN(COCA)NNH) (16) NMR experiments was used to achieve the full sequence-specific assignment of 4Q. Instead,  $^{13}\text{C}$ -detected 4D HCBCACON (13) and 4D (HN)CON(CA)CON (14) experiments, and  $^1\text{H}$ -detected 3D TROSY HNCO (15), 3D TROSY HN(CA)CO (17) and 4D TROSY HN(COCA)NNH (16) experiments were acquired to obtain the complete resonances assignment of 25Q. All the experiments were performed using on-grid non-uniform sampling (NUS). The o “Poisson disk” sampling scheme was chosen to generate the time schedules with the RSPack program (18). The assignments of  $\text{H}^{\text{N}}$ ,  $\text{H}^{\alpha}$ ,  $\text{H}^{\beta}$ ,  $\text{C}'$ ,  $\text{C}^{\alpha}$ ,  $\text{C}^{\beta}$  and N resonances of 4Q and 25Q are reported in the BMRB ([www.bmrb.wisc.edu](http://www.bmrb.wisc.edu)) (19), entries 25606 and 25607. The parameters used for the acquisition of the experiments are reported in Tables S1-S4.

### NMR data processing and analysis

NUS NMR data were converted with *NMRPipe* (20) and then processed using *ToASTD* (21) and *reduced* (22, 23) programs. Uniformly sampled NMR data were processed with *TopSpin*. *Sparky* (24) and *CcpNMR Analysis* (25) were used to visualize the spectra and analyze the  $^{15}\text{N}$  relaxation data, respectively. The secondary structure propensity from the heteronuclear chemical shifts was determined for 4Q and 25Q by using the *neighbor corrected structural propensity calculator* (ncSPC) method (26), available online at <http://nmr.chem.rug.nl/ncSPC/>. The Tamiola, Acar and Mulder random coil chemical shift library was chosen for the analyses (27) using ncSPC. The secondary structure propensities were also calculated by using the  $\delta 2\text{D}$  method (28), available online at <http://www-mvsoftware.ch.cam.ac.uk/>.

## **Results and Discussion**

In order to investigate how the properties of the polyQ tract depend on its length we used constructs containing 4 and 25 Gln residues (4Q and 25Q, Figs. 1b and S1). Constructs featuring longer polyQ tracts could not be used in the study because they aggregate too fast with respect to the time required to perform the NMR experiments. The disordered nature of the protein and the presence of a 25 residue-long polyQ tract required an experimental strategy based on the use of recently developed 4D  $^{13}\text{C}$ -detected NMR experiments that lead to the complete sequence specific assignment of the fragment (13, 14). Heteronuclei were exploited to take advantage of their improved chemical shift dispersion, compared to that of protons, and of their reduced sensitivity to exchange processes (29). The extensive cross-peak overlap in the spectra was overcome by acquiring high-dimensional experiments and by

exploiting non-uniform sampling (NUS) to achieve high spectral resolution in the indirect dimensions. To our knowledge, a full characterization at atomic resolution of a polyQ tract under native conditions, as the one reported here, is unprecedented.

As shown in Figure 2, in which we present the central regions of the 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC and 2D CON-IPAP spectra of 4Q and 25Q, the spectral differences between these two constructs are limited to the resonances of the polyQ tract and of the 4 Leu residues immediately preceding it *i.e.* the  $^{55}\text{LLLL}^{58}$  motif. An interesting feature of the spectra of 25Q is the presence of pseudo-diagonals, a characteristic fingerprint, consisting of the cross peaks corresponding to the residues of the polyQ tract, which have  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}'$  chemical shifts that correlate with their position in the sequence of the fragment (Fig. S2 in the supplemental material).

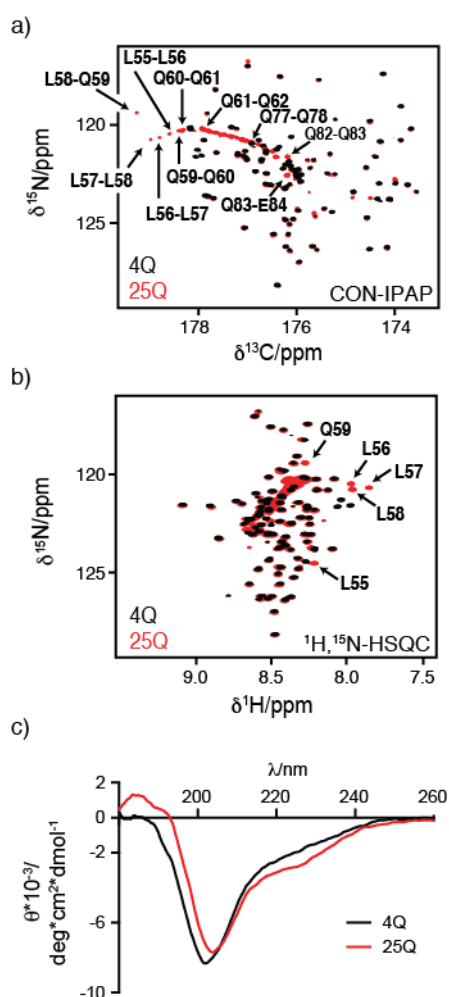


Figure 2. Comparison of the NMR (a,b) and CD (c) spectra of 4Q and 25Q. Central region of (a) the CON-IPAP spectrum and (b) the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of 400  $\mu\text{M}$  4Q (black) and 25Q (red) at 278 K with an indication, in panel b, of the resonances which experience the largest chemical shift variations upon increasing the length of the polyQ tract. A close up of the CON-IPAP spectrum of 25Q with the full assignment of the polyQ tract is provided as supplemental material (Fig. S2). c) CD spectrum of 120  $\mu\text{M}$  4Q and 130  $\mu\text{M}$  25Q at 310 K.

We analyzed the NMR signals chemical shifts to derive the conformational properties of 4Q and 25Q (Fig. 3) in terms of secondary structural propensities by using the algorithms ncSPC (27, 30) and  $\delta$ 2D (28). The values obtained with both methods indicate the absence of persistent secondary structure in the polypeptide except for the polyQ tract and for the 4 Leu residues preceding it, which show some  $\alpha$ -helical propensity, and a dramatic increase in helical propensity of the polyQ tract upon expansion from 4 to 25 residues. Although similar observations have been reported (31) this is a surprising result considering that polyQ tracts are in general disordered or, in the fibrillar state, in an extended conformation (8, 32). The fact that the helical propensity of the polyQ tract increases with its length was confirmed by circular dichroism (CD) analysis (Fig. 2c) and  $^{15}\text{N}$  relaxation measurements (Figs. S3). Interestingly, the helicity was found to be most pronounced at the beginning of the polyQ tract and to decrease gradually towards its end.

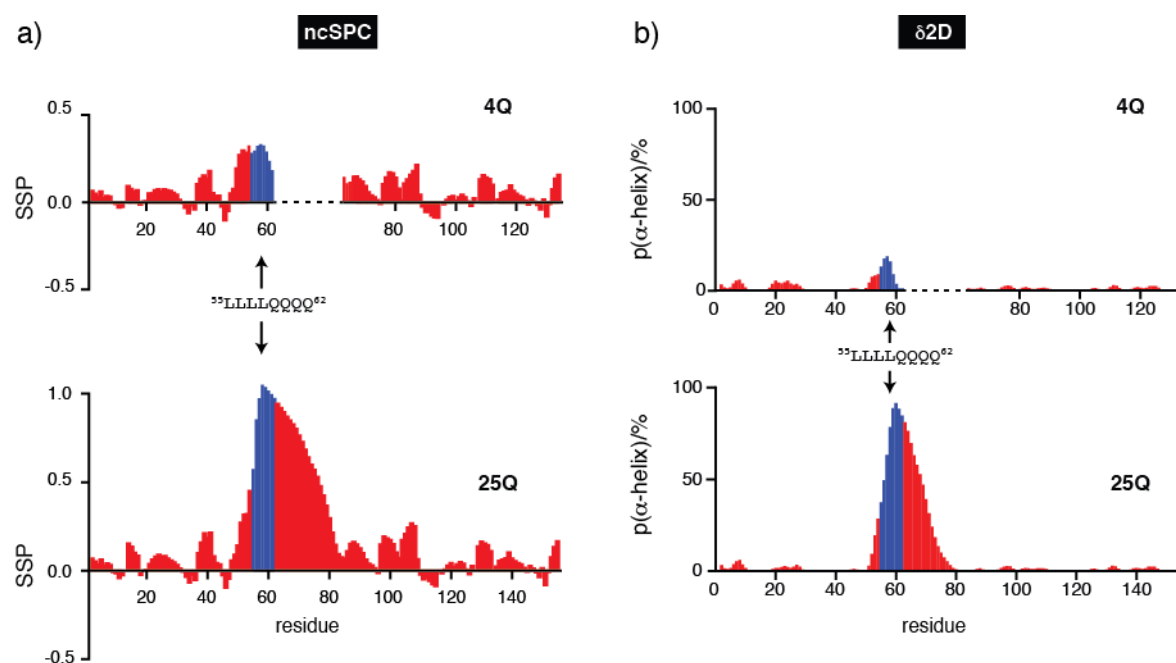


Figure 3. Secondary structure propensity of 4Q and 25Q obtained by using ncSPC (a) and population of  $\alpha$ -helix obtained by using  $\delta$ 2D (b). To facilitate the comparison, values for residues of 4Q which follow the polyQ tract are shifted to the right by 21 units. Values for residues 55 to 62, corresponding to the  $^{55}\text{LLLL}^{58}$  motif and the first 4 Gln of the polyQ tract, are shown in blue to highlight the variation of the structural properties of the protein due to the different length of the polyQ tract.

The fact that the residues of the  $^{55}\text{LLLL}^{58}$  motif, which precedes the polyQ tract, undergo very substantial chemical shift changes upon expansion of the latter (Fig. 2 and Fig. S3) indicates a certain degree of cooperativity in the conformational transition caused by expansion of the polyQ tract and suggests that they induce the formation of the  $\alpha$ -helix. To verify this hypothesis, we analyzed by CD and NMR a mutant of 25Q in which the motif  $^{55}\text{LLLL}^{58}$  had been removed (25Q $\Delta$ L4, Fig. 1b and S1). The results obtained indicate a loss of  $\alpha$ -helical propensity (Fig. 4a), and a



decrease in the chemical shift dispersion in the resonances corresponding to the polyQ tract (Fig. 4b). Both observations confirm the hypothesis that the motif <sup>55</sup>LLLL<sup>58</sup> induces a helical conformation in the polyQ tract and that in its absence the tract adopts a conformation equivalent to that of synthetic polyQ peptides (11).

To investigate the effect of the 4 Leu residues on the kinetics of aggregation of the AR fragment, we compared the temporal evolution of 20  $\mu$ M solutions of 4Q, 25Q and 25Q $\Delta$ L4 by dynamic light scattering (DLS), a technique that is well-suited to characterize the early stages of protein aggregation (Fig. 4c). The results show that the particle size ( $D_{hz}$ ) increases much faster and reaches higher values for 25Q $\Delta$ L4 than for 25Q (Fig. 4c), indicating that the presence of the motif <sup>55</sup>LLLL<sup>58</sup> modifies the aggregation behavior of the protein. 4Q, as expected (33), was found to hardly aggregate in the timescale probed in this experiment. The effect of the <sup>55</sup>LLLL<sup>58</sup> motif on the aggregation properties of the AR fragment is conceptually related to that obtained for huntingtin, the protein bearing the polyQ tract involved in Huntington's disease. In that case a polyproline flanking region was found to decrease the aggregation rate by altering, albeit in a different way, the structural properties of the polyQ tract (34).

It is important to emphasize that the mechanism by which the helical motif <sup>55</sup>LLLL<sup>58</sup> alters the aggregation behavior of the AR fragment is fundamentally different from the mechanism by which previously described helical flanking regions modulate the aggregation tendencies of polyQ tracts. The N-terminal 17-residue helical region of huntingtin, for example, greatly accelerates the rate of aggregation of this protein not by modifying the structural properties of the polyQ tract but, rather, by establishing intermolecular interactions, with the same region of other monomers, that increase the local concentration of polyQ tracts and therefore facilitate their oligomerization (35); a similar behavior has been observed also for other proteins bearing polyQ tracts (36).

The high resolution investigation of a protein bearing a polyQ tract achieved here by using a recently developed NMR strategy in combination with complementary techniques (CD, DLS) has revealed a new mechanism by which protein flanking regions influence the aggregation behavior of such tracts. Our findings reveal how a Leu-rich motif flanking a polyQ tract at its N-terminus can induce helical structure in the latter and in turn modify the mechanism of aggregation by adding a helix unfolding step to it. This is so because the intra-molecular hydrogen bonds that stabilize the compact  $\alpha$ -helical secondary structure must break before the cross- $\beta$  structure, rather elongated and stabilized by intermolecular hydrogen bonds, can form (37).

These results also show that the influence of flanking regions on the secondary structure of polyQ tracts spans a limited range, in our case *ca* 20 residues (Fig. 3) and that it may cease to be effective for the longer tracts which cause polyQ diseases (11, 38). In this scenario the potential protective role played by the induction of the helical structure would only be efficient for polyQ tracts of limited length, contributing to explaining why the protein becomes toxic only when the number of Gln residues exceeds a certain threshold. Sequence context would in this way be a contributing factor to the disease and protein-specific nature of the threshold that triggers protein aggregation and the onset of polyQ disease.

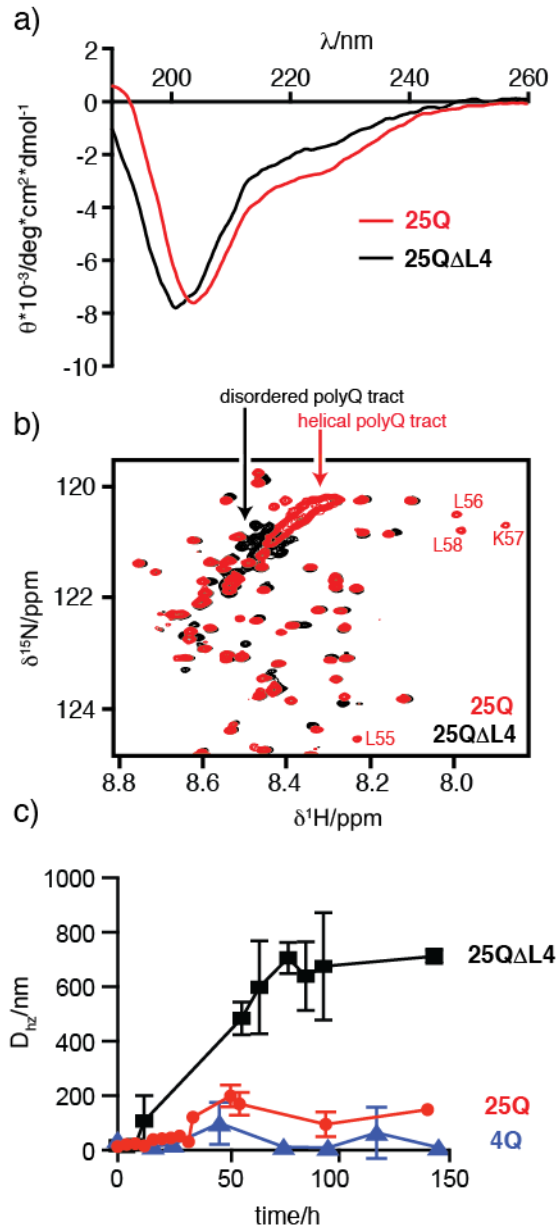


Figure 4. Characterization of the structural and aggregation properties of 25Q (red) and 25Q $\Delta$ L4 (black). a) Comparison of the CD spectra of 130  $\mu$ M 25Q and 140  $\mu$ M 25Q $\Delta$ L4 at 310K. b) Comparison of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR spectra of 50  $\mu$ M 25Q and 43  $\mu$ M 25Q $\Delta$ L4 at 278K. The cross-peaks belonging to the four Leu of 25Q immediately preceding the polyQ tract, not present in 25Q $\Delta$ L4, are indicated. The absence of helical propensity in 25Q $\Delta$ L4, which reduces the chemical shift dispersion of the resonances corresponding to the polyQ tract, is highlighted. c) Temporal evolution of the mean particle size of 4Q (blue), 25Q and 25Q $\Delta$ L4 at 20  $\mu$ M and 310K.

## Conclusion

Our findings provide new insights into the structural properties of polyQ tracts and offer a compelling example of how flanking regions can alter the aggregation behavior and, therefore, the formation of potentially toxic fibrillar species by proteins harboring these peculiar sequences. Our increased understanding of how sequence context influences the properties of polyQ tracts opens up new avenues for the development of therapeutic strategies for polyQ diseases based on targeting residues in the flanking regions rather than the polyQ tract themselves.

## Author Contributions

R.P., I.C.F. and X.S. conceived the project and designed the experiments; B.E., A.P., G.C. D.M. and J.G. performed the experiments and analyzed the data; all authors contributed to writing the paper.

## Acknowledgments

This work was funded by BioNMR (EC contract no. 261863), IRB, ICREA, Obra Social “la Caixa” (B.E.), AGAUR (G.C.), MICINN (CTQ2009-08850 to X.S.), MINECO (BIO2012-31043 to X.S.), Marató de TV3 (102030 to X.S.) and the ERC (CONCERT, 648201 to X.S.). D.M. was a recipient of an Erasmus placement fellowship from the University of Florence (Italy).

## References

1. Zoghbi, H.Y., and H.T. Orr. 2000. Glutamine repeats and neurodegeneration. *Annu. Rev. Neurosci.* 23: 217–47.
2. Mirkin, S.M. 2007. Expandable DNA repeats and human disease. *Nature.* 447: 932–940.
3. Liu, K.-Y., Y.-C. Shyu, B. a. Barbaro, Y.-T. Lin, Y. Chern, L.M. Thompson, C.-K. James Shen, and J.L. Marsh. 2014. Disruption of the nuclear membrane by perinuclear inclusions of mutant huntingtin causes cell-cycle re-entry and striatal cell death in mouse and cell models of Huntington’s disease. *Hum. Mol. Genet.* 24: 1602–1616.
4. Bates, G.P., R. Dorsey, J.F. Gusella, M.R. Hayden, C. Kay, B.R. Leavitt, M. Nance, C. a. Ross, R.I. Scahill, R. Wetzel, E.J. Wild, and S.J. Tabrizi. 2015. Huntington disease. *Nat. Rev. Dis. Prim.* : 15005.
5. Sahl, S.J., L.E. Weiss, W.C. Duim, J. Frydman, and W.E. Moerner. 2012. Cellular inclusion bodies of mutant huntingtin exon 1 obscure small fibrillar aggregate species. *Sci. Rep.* 2: 895.
6. Nalavade, R., N. Griesche, D.P. Ryan, S. Hildebrand, and S. Krauss. 2013. Mechanisms of RNA-induced toxicity in CAG repeat disorders. *Cell Death Dis.* 4: e752.
7. Li, P., K.E. Huey-Tubman, T. Gao, X. Li, A.P. West Jr., M.J. Bennett, and P.J. Bjorkman. 2007. The structure of a polyQ-anti-polyQ complex reveals binding according to a linear lattice model. *Nat Struct Mol Biol.* 14: 381–387.
8. Masino, L., G. Kelly, K. Leonard, Y. Trottier, and A. Pastore. 2002. Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins. *FEBS*

- Lett. 513: 267–272.
9. Merry, D.E., Y. Kobayashi, C.K. Bailey, a. a. Taye, and K.H. Fischbeck. 1998. Cleavage, aggregation and toxicity of the expanded androgen receptor in spinal and bulbar muscular atrophy. *Hum. Mol. Genet.* 7: 693–701.
  10. Gelmann, E.P. 2002. Molecular Biology of the Androgen Receptor. *J. Clin. Oncol.* 20: 3001–3015.
  11. Chen, S., V. Berthelie, W. Yang, and R. Wetzel. 2001. Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity. *J. Mol. Biol.* 311: 173–82.
  12. Hellstrand, E., B. Boland, D.M. Walsh, and S. Linse. 2010. Amyloid  $\beta$ -protein aggregation produces highly reproducible kinetic data and occurs by a two-phase process. *ACS Chem. Neurosci.* 1: 13–18.
  13. Bermel, W., I. Bertini, I.C. Felli, L. Gonnelli, W. Koźmiński, A. Piai, R. Pierattelli, and J. Stanek. 2012. Speeding up sequence specific assignment of IDPs. *J. Biomol. NMR.* 53: 293–301.
  14. Bermel, W., I.C. Felli, L. Gonnelli, W. Koźmiński, A. Piai, R. Pierattelli, and A. Zawadzka-Kazimierczuk. 2013. High-dimensionality  $^{13}\text{C}$  direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. *J. Biomol. NMR.* 57: 353–361.
  15. Kay, L.E., and A. Bax. 1990. New Methods for the Measurement of NH- & H Constants in “N-Labeled Proteins Coupling. *J. Magn. Reson.* 86: 110–126.
  16. Piai, A., T. Hošek, L. Gonnelli, A. Zawadzka-Kazimierczuk, W. Koźmiński, B. Brutscher, W. Bermel, R. Pierattelli, and I.C. Felli. 2014. “CON-CON” assignment strategy for highly flexible intrinsically disordered proteins. *J. Biomol. NMR.* 60: 209–18.
  17. Clubb, R.T., V. Thanabal, and G. Wagner. 1992. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue  $^1\text{HN}$ ,  $^{15}\text{N}$ , and  $^{13}\text{C}'$  chemical shifts in  $^{15}\text{N}$ -  $^{13}\text{C}$ -labelled proteins. *J. Magn. Reson.* 97: 213–217.
  18. Kazimierczuk, K., A. Zawadzka, and W. Koźmiński. 2008. Optimization of random time domain sampling in multidimensional NMR. *J. Magn. Reson.* 192: 123–130.
  19. Ulrich, E.L., H. Akutsu, J.F. Doreleijers, Y. Harano, Y.E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C.F. Schulte, D.E. Tolmie, R. Kent Wenger, H. Yao, and J.L. Markley. 2008. BioMagResBank. *Nucleic Acids Res.* 36: D402–8.
  20. Delaglio, F., S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* 6: 277–293.
  21. Kazimierczuk, K., A. Zawadzka, W. Koźmiński, and I. Zhukov. 2006. Random sampling of evolution time space and Fourier transform processing. *J. Biomol. NMR.* 36: 157–68.
  22. Kazimierczuk, K., A. Zawadzka-Kazimierczuk, and W. Koźmiński. 2010. Non-uniform frequency domain for optimal exploitation of non-uniform sampling. *J. Magn. Reson.* 205: 286–92.
  23. Kazimierczuk, K., A. Zawadzka, and W. Koźmiński. 2009. Narrow peaks and high dimensionalities: exploiting the advantages of random sampling. *J. Magn. Reson.* 197: 219–28.
  24. Kneller, D.G., and I.D. Kuntz. 1993. UCSF Sparky: An NMR Display, Annotation and Assignment Tool. *J. Cell. Biochem.* 53 (S17C): 254.

25. Vranken, W.F., W. Boucher, T.J. Stevens, R.H. Fogh, A. Pajon, M. Llinas, E.L. Ulrich, J.L. Markley, J. Ionides, and E.D. Laue. 2005. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*. 59: 687–96.
26. Tamiola, K., B. Acar, and F. a a Mulder. 2010. Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J. Am. Chem. Soc.* 132: 18000–18003.
27. Tamiola, K., and F. a a Mulder. 2012. Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. *Biochem. Soc. Trans.* 40: 1014–20.
28. Camilloni, C., A. De Simone, W.F. Vranken, and M. Vendruscolo. 2012. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry*. 51: 2224–2231.
29. Gil, S., T. Hošek, Z. Solyom, R. Kümmerle, B. Brutscher, R. Pierattelli, and I.C. Felli. 2013. NMR spectroscopic studies of intrinsically disordered proteins at near-physiological conditions. *Angew. Chem. Int. Ed. Engl.* 52: 11808–12.
30. Marsh, J. a, V.K. Singh, Z. Jia, and J.D. Forman-Kay. 2006. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci.* 15: 2795–2804.
31. Davies, P., K. Watt, S.M. Kelly, C. Clark, N.C. Price, and I.J. McEwan. 2008. Consequences of poly-glutamine repeat length for the conformation and folding of the androgen receptor amino-terminal domain. *J. Mol. Endocrinol.* 41: 301–314.
32. Buchanan, L.E., J.K. Carr, A.M. Fluit, A.J. Hoganson, S.D. Moran, J.J. de Pablo, J.L. Skinner, and M.T. Zanni. 2014. Structural motif of polyglutamine amyloid fibrils discerned with mixed-isotope infrared spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* 111: 5796–5801.
33. Chen, S., F.A. Ferrone, and R. Wetzel. 2002. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc. Natl. Acad. Sci. U. S. A.* 99: 11884–11889.
34. Bhattacharyya, A., A.K. Thakur, V.M. Chellgren, G. Thiagarajan, A.D. Williams, B.W. Chellgren, T.P. Creamer, and R. Wetzel. 2006. Oligoproline effects on polyglutamine conformation and aggregation. *J. Mol. Biol.* 355: 524–35.
35. Thakur, A.K., M. Jayaraman, R. Mishra, M. Thakur, V.M. Chellgren, I.-J.L. Byeon, D.H. Anjum, R. Kodali, T.P. Creamer, J.F. Conway, A.M. Gronenborn, and R. Wetzel. 2009. Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism. *Nat. Struct. Mol. Biol.* 16: 380–389.
36. Ellisdon, A.M., B. Thomas, and S.P. Bottomley. 2006. The two-stage pathway of ataxin-3 fibrillogenesis involves a polyglutamine-independent step. *J. Biol. Chem.* 281: 16888–16896.
37. Calloni, G., C. Lendel, S. Campioni, S. Giannini, A. Gliozzi, A. Relini, M. Vendruscolo, C.M. Dobson, X. Salvatella, and F. Chiti. 2008. Structure and dynamics of a partially folded protein are decoupled from its mechanism of aggregation. *J. Am. Chem. Soc.* 130: 13040–50.
38. Scherzinger, E., A. Sittler, K. Schweiger, V. Heiser, R. Lurz, R. Hasenbank, G.P. Bates, H. Lehrach, and E.E. Wanker. 1999. Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology. *Proc. Natl. Acad. Sci. U. S. A.*

96: 4604–9.