

RESEARCH ARTICLE

# The Protein Structure Context of PolyQ Regions

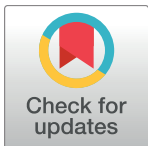
Franziska Totzeck<sup>1</sup>, Miguel A. Andrade-Navarro<sup>1,2</sup>, Pablo Mier<sup>1,2\*</sup>

**1** Faculty of Biology, Johannes Gutenberg University Mainz, Gresemundweg 2, Mainz, Germany, **2** Institute of Molecular Biology, Ackermannweg 4, Mainz, Germany

\* [munoz@uni-mainz.de](mailto:munoz@uni-mainz.de)

## Abstract

Proteins containing glutamine repeats (polyQ) are known to be structurally unstable. Abnormal expansion of polyQ in some proteins exceeding a certain threshold leads to neurodegenerative disease, a symptom of which are protein aggregates. This has led to extensive research of the structure of polyQ stretches. However, the accumulation of contradictory results suggests that protein context might be of importance. Here we aimed to evaluate the structural context of polyQ regions in proteins by analysing the secondary structure of polyQ proteins and their homologs. The results revealed that the secondary structure in polyQ vicinity is predominantly random coil or helix. Importantly, the regions surrounding the polyQ are often not solved in 3D structures. In the few cases where the point of insertion of the polyQ was mapped to a full protein, we observed that these are always located in the surface of the protein. The findings support the hypothesis that polyQ might serve to extend coiled coils at their C-terminus in highly disordered regions involved in protein-protein interactions.



## OPEN ACCESS

**Citation:** Totzeck F, Andrade-Navarro MA, Mier P (2017) The Protein Structure Context of PolyQ Regions. PLoS ONE 12(1): e0170801. doi:10.1371/journal.pone.0170801

**Editor:** Patrick van der Wel, University of Pittsburgh School of Medicine, UNITED STATES

**Received:** September 9, 2016

**Accepted:** January 11, 2017

**Published:** January 26, 2017

**Copyright:** © 2017 Totzeck et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information file.

**Funding:** The authors received no specific funding for this work.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

Homopeptide repeats are consecutive stretches of the same amino acid in protein sequences. They are surprisingly common in proteins, and it has been suggested that they form unstructured stretches within a protein and may serve a function in protein-protein interaction (PPI) [1, 2]. Polyglutamine (polyQ) in particular is one of the most common homopeptide repeats in eukaryotic proteomes [1, 3]. It can be found in a variety of protein families which do not appear to be related [4].

Although there is an increasing amount of studies indicating the function of normal polyQ [5–8], most works study the effects of its abnormal extension, caused in human sequences by CAG trinucleotide expansion. There are several diseases known to be caused by such expansion of polyQ; most infamously Huntington’s disease [9]. All polyQ diseases involve neural degeneration, the onset and severity of which depends on the length of the expanded polyQ. Wild type huntingtin, for example, has a stretch of 23 glutamines; its expansion to less than 34 glutamines is not pathogenic, while a stretch of more than 36 glutamines results in neural degeneration [10, 11]. PolyQ proteins are known to be involved in PPIs [5], which may lead to

aggregation when the polyQ is abnormally expanded. Indeed, polyQ diseases are characterised by aggregates of the respective protein containing an expanded polyQ tract within neuronal cells; however, their specific role in the disease is not quite clear. While these aggregates, as found in inclusion bodies, were originally thought to be the cause of polyQ diseases, later research suggested that they might serve a protective role and merely be a symptom of the disease rather than the cause [12].

Determination of the structure of polyQ regions has proven to be extremely difficult as they appeared to be very unstable. Most experimental studies seem to suggest a predominant random coil conformation [13–15]; however, there is also evidence for  $\beta$ -sheet as well as helical structure [16, 17]. Regarding the structure of aggregates, the current consensus is that these contain a high amount of  $\beta$ -sheets, possibly formed by a gradual change of conformation of polyQ proteins [18, 19]. While small polyQ-containing peptides have also been found to aggregate, it appears that expanded polyQ aggregates far more rapidly [13]. A subsequent recruitment of proteins with small polyQ tracts has also been reported [14].

Regarding the structure of the polyQ itself there is no consensus. It has been proposed that the structure of polyQ of both pathogenic and non-pathogenic length is largely the same [13], while some studies found a slight change in the overall secondary structure content [20], and others even reported a sharp change from an extended monomeric conformation to a collapsed state [21].

Recent studies have addressed the possible effect of the protein context on polyQ. It has been noted that many proteins containing polyQ also contain coiled coils, which facilitate protein interaction and, in case of polyQ expansion, participate in aggregation [22]. It was noted that polyQ appears often at the C-terminus of coiled coil regions and this was taken as an indication that polyQ could serve to increase the length of the coiled coil to modulate their interactions [5, 22]. We also noted that proteins that interact with polyQ proteins often contain coiled coils too and later we found evidence that their interaction with an expanded polyQ construct promotes its aggregation [23].

Other studies have addressed the possible effect of adjacent homorepeats on polyQ regions. polyP regions, which can often be found in the vicinity (C-terminal) of polyQ [5], were found to suppress aggregation of peptides containing a pathogenic polyQ stretch [24]. On the contrary, polyA regions can trigger polyQ aggregation through coiled coil formation [25].

Finally, additional studies concentrated on the domain context and provided evidence that indeed surrounding domains appear to play a role in polyQ aggregation [26–31]. However, an evaluation of the structural context of polyQ is as yet missing, although it is quite conceivable that secondary structure in the neighbourhood of polyQ might put constraints on its conformation.

This work focuses on the analysis of the natural protein structure context of polyQ, using the structures that have been determined so far of both polyQ proteins and their homologs for an analysis of the secondary structure around it.

## Methods

### Definition of polyQ

PolyQ was defined as a consecutive stretch of amino acids that contains at least eight glutamines per ten residues (a maximum of two mismatches in a minimum of ten residues); however, other definitions for polyQ were also tested to check the dependence of the results. PolyQ thresholds are written as a fraction, where the denominator indicates the number of amino acids and the numerator the minimum number of glutamines found in it (e.g. 8/10 means a threshold of at least eight glutamines per ten amino acids).

## Dataset

The dataset was obtained using FastaHerder2 mode 4 [32] to look for clusters of similar protein sequences of comparable length containing at least one protein with polyQ (in FastaHerder2 defined as 8/10) and at least one protein with associated PDB annotation. Clusters were generated from SwissProt proteins (548,454 proteins, release 2015\_05), and the proteins within them are at least 53% identical. Theoretical protein structures were not taken into account, and clusters for which only theoretical models were available were dismissed. This produced a total of 178 clusters, 74 of which consisted only of one protein. All in all the dataset contained 926 proteins.

Protein sequences were downloaded from the UniProt database (<http://www.uniprot.org>) [33]. Clusters containing multiple proteins were aligned using Clustal Omega [34] on the UniProt web server [33] with default parameters. All structure files were obtained from the RCSB PDB (<http://www.rcsb.org/pdb/home/home.do>) [35]. Pictures of protein structures were produced using UCSF Chimera 1.10.2 [36].

## Script for analysis

The script used to analyse the data was written in Java programming language. Secondary structure information was extracted from the PDB files with the aid of BALL 1.4 [37], for which a small C++ script was written and called from the main Java script. The analysed data, the main Java script and the helper script in C++ can be found in our web server (<http://cbdm-01.zdv.uni-mainz.de/~munoz/polyq/>).

## Obtaining secondary structure information for a protein

Information extracted from the PDB files was gathered for every residue of every chain by retrieving its index number, amino acid type and secondary structure. The secondary structure contained the information whether a given residue of the chain had the geometric properties of a residue within a helix, sheet or random coil. Subsequently, a matching of the amino acid sequence of the chain and the respective protein was attempted; at first by using the index. However, many PDB structures contain information from several proteins, making it necessary to compare also the sequence identity of both chain and respective part of the protein. Furthermore, many proteins are in multiple PDB files, which often contain multiple chains of the same sequence. In order to obtain one coherent structure string, for every position either the secondary structure with the best resolution was used, or, if resolutions were the same but structures differed, an identical likelihood for both conformations was assumed for that position.

Most polyQ proteins in the dataset have no associated structure information. However, protein structure is more conserved than protein sequence. The structure of related proteins is very likely to be similar to that of the protein of interest itself. For some polyQ proteins, more than one of the related proteins had structure information for a given position. If this was the case, all of the possible conformations at that position were taken into account and normalised for the total amount of secondary structure information at that particular position (e.g. if two proteins had a helical conformation and one a random coil, the structure was counted as 0.66 helical and 0.33 random coil). If the polyQ protein happened to have own secondary structure information for a given position, none of the structures of related proteins were taken into account.

## Normalisation of the data

There was a varying amount of polyQ proteins per cluster; furthermore, some proteins contained more than one polyQ. In this analysis, the data were normalised for the number of

polyQ, regardless of the number of actual proteins. For example, if a cluster had two polyQ proteins, one with one polyQ and the other with four polyQs, each of these polyQs was taken into account individually and added up; subsequently, the sum for each position of the secondary structure information (helix, sheet, random coil or no information) was divided by five.

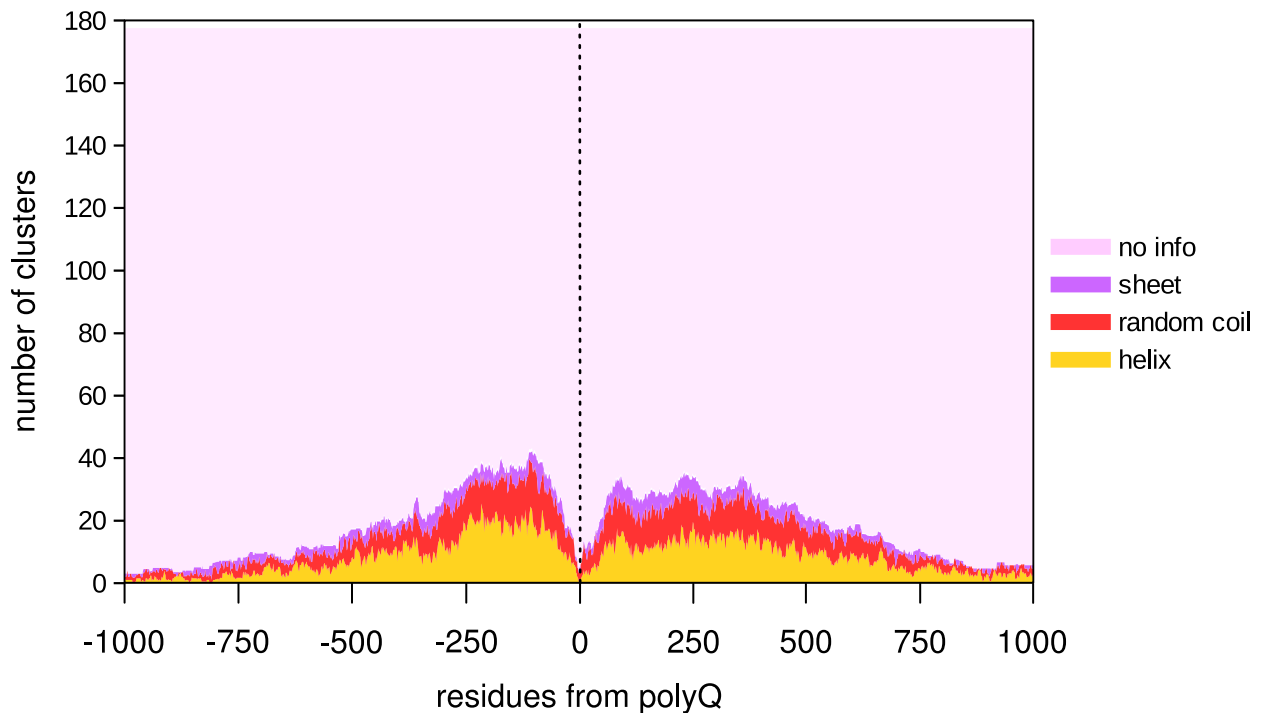
## Results

A set of 178 clusters, consisting of polyQ proteins and their homologs (at least a 53% identity between proteins in a cluster, see [Methods](#) for details), was analysed for secondary structure context in the vicinity of polyQ regions. Homologs were taken into account in order to increase the amount of secondary structure information, which was categorised into helix, sheet or random coil. The clusters in the dataset were of varying size, from one to 243 proteins. A total of 282 proteins out of the 926 in the dataset contain a polyQ with at least eight glutamines per ten amino acids (an 8/10 polyQ). Most of them only contain one polyQ, but there are a few that contain a higher number, up to six.

### The protein structure context of polyQ

Generally, structure information is available only for fragments of proteins and not for complete proteins, firstly because usually only parts of proteins are studied, but also because some regions might be disordered and will not adopt a single structure that could be resolved. The latter effect greatly influenced our observations since for most of the clusters we used there was no structural information available in the closer proximity of the polyQ ([Fig 1](#)).

Despite the fact that all clusters have some structural information, a feature for which they were selected, only a maximum of around 40 clusters have overlapping information for any given residue in relation to the polyQ position. In any case, taking into account the structure's

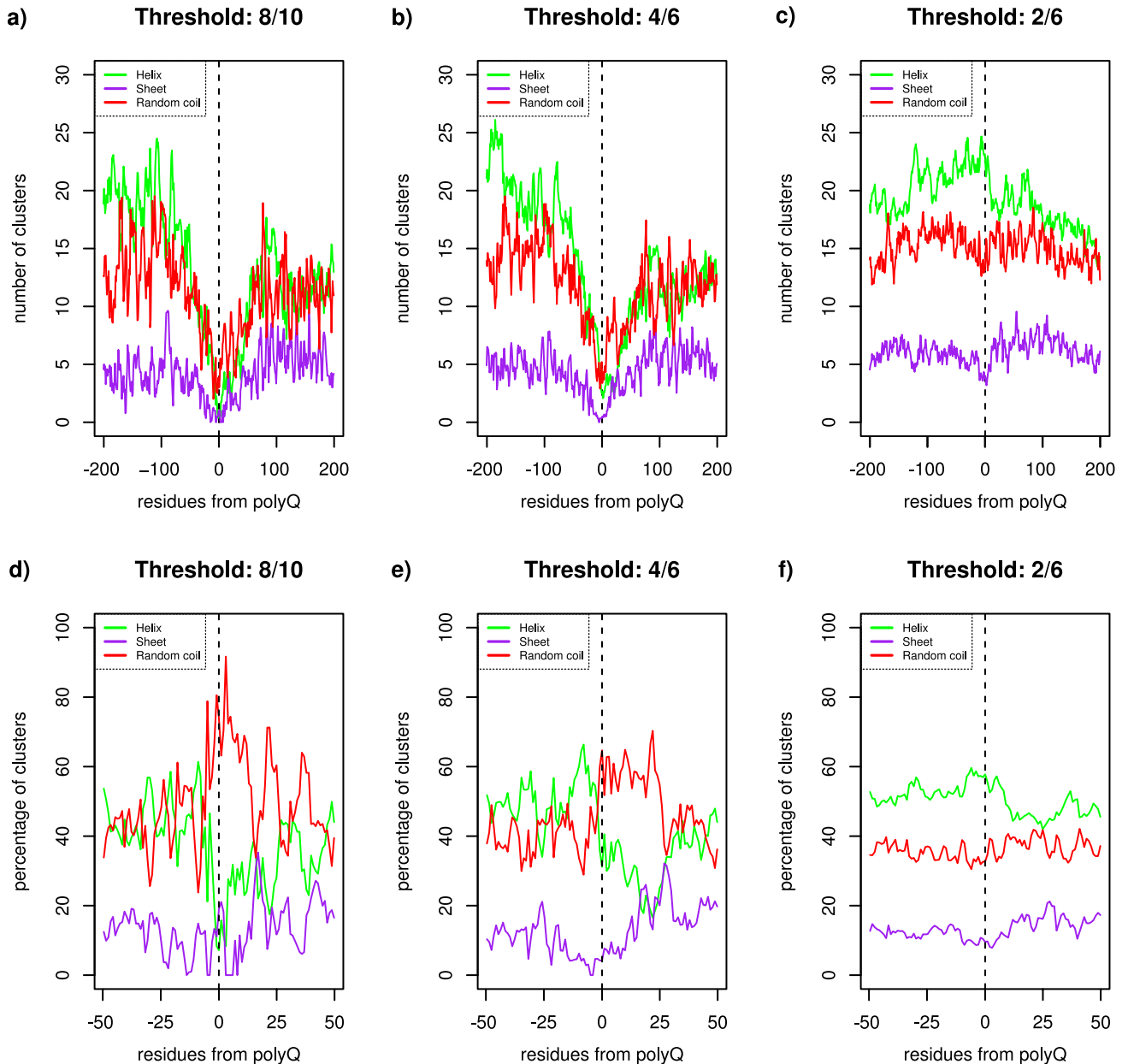


**Fig 1. Structural information per residue surrounding the polyQ stretch (8/10).** Each residue could either have no structural information (“no info”, pink), or have the information: sheet (purple), random coil (red) or helix (yellow). The drop of secondary structure information at the sides of the graph only reflects the size distribution of the fragments used.

doi:10.1371/journal.pone.0170801.g001

results, these show a prevalence of helical structure as opposed to random coils and  $\beta$ -sheet (range -500:500).

Next, we analysed the distribution of structures for various thresholds for polyQ definition (Fig 2). In the -200:200 range of the polyQ, there is an overall prevalence of both helix and random coil conformation; sheet conformation is particularly rare close and N-terminal to the polyQ (Fig 2A). Considering the immediate proximity of the polyQ (range -25:25), the N-terminal region is more likely to be in helical structure than the C-terminal region. This can be better seen when the data are normalised by the number of clusters with structural information for each position: while the ratio of helical structure is roughly between 40% and 60% in the closest



**Fig 2. Structure context of polyQ regions using different thresholds.** Number (a, b, c) or percentage (d, e, f) of clusters with a certain structural conformation per residue surrounding the polyQ stretch, using a 8/10 (a and d), 4/6 (b and e), or 2/6 (c and f) polyQ threshold.

doi:10.1371/journal.pone.0170801.g002

positions to the polyQ origin (range -25:5), it is clearly lower in the closest C-terminal positions (range 5:25), roughly between 20% and 40% (Fig 2D). The results suggest that helical structures are preferably N-terminal to the polyQ middle position, while random coils are preferably C-terminal to it. Since position 0 is taken as the middle of the polyQ, the distribution of structures also shows that the secondary structure most likely to overlap with the middle of a polyQ is a random coil or a helix, if there is a solved structure for the region at all. In this respect, a very strong observation is the lack of solved structure close to the 0 (Fig 2A and 2B).

Results may differ depending on the threshold used to consider a polyQ. To characterise in detail the impact a polyQ region may have on the protein structure, several different polyQ definitions were also used. Using a 6/8 polyQ, the amount of available structural data increases slightly, while the ratio of structures in the vicinity of the polyQ still looks largely the same as that for 8/10 (data not shown). This suggests that a 6/8 polyQ is behaving the same way in relation to the rest of the protein and its structural context as an 8/10 polyQ.

A similar result is achieved when the definition of polyQ is lowered down to 4/6 (Fig 2B and 2E). The absence of sheet conformation close and N-terminal to the polyQ remains, but in the longer range it becomes more apparent that there is more sheet content in the 0:200 than in the -200:0 range, a bias that we cannot explain. Regarding the distribution of helix and random coil in the -25:25 range of polyQ, now, almost 40% of the structure overlapping with polyQ is a helix (Fig 2E). The valley of low helix conformation near polyQ is still present but shifted about 25 residues towards the C-terminus of the polyQ. The polyQ structural context differs considerably when using a definition of polyQ as weak as 2/6. The secondary structures are almost evenly distributed around the polyQ, independently of the distance to it (Fig 2F). The available structural information is not reduced in the proximity of the polyQ region (no valley close to position 0) (Fig 2C), which suggests that a 2/6 threshold should not be considered as a polyQ, as the structure around it is not disrupted by its presence.

## Places of polyQ insertion in experimental protein structures

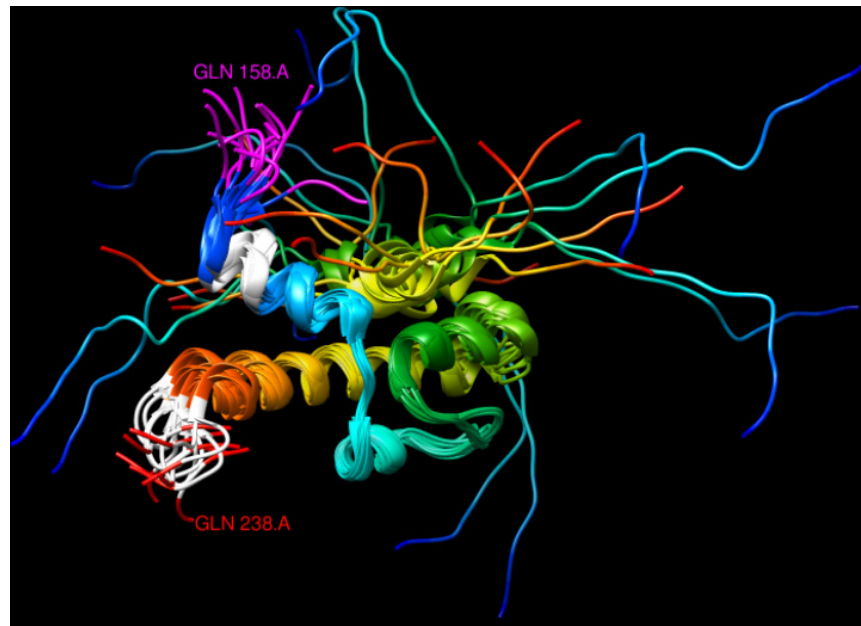
While, as shown above, polyQ regions are generally not part of solved structures, there were a few cases where the structure of a polyQ region was available. Among glutamine stretches associated with helices, both N- and C-terminal locations could be found. In the mouse protein OTX2 (UniProt:P80206), for example, a 7/7 polyQ is at the C-terminus of a helix (PDB:2DMS), while in the human CREB-binding protein (UniProt:Q92793) a 5/5 polyQ is present at the N-terminus of a helix (PDB:1ZOQ). Interestingly, in the structure of the yeast protein Gal11 (UniProt:P19659, PDB:2LPB) both cases happen, as one 4/4 polyQ is present N-terminal to a helix and another 6/8 polyQ is C-terminal to another helix (Fig 3, in white).

Although the majority of structures in the dataset were from protein fragments, a few of them were nearly complete proteins, like the structures of the proteins WHY1 from *Solanum tuberosum* (PDB:1L3A), WDR5 from *Rattus norvegicus* (PDB:4QQE), glycinin G1 from the soybean (PDB:1FXZ) and SEC23 from *Saccharomyces cerevisiae* (PDB:1M2O). In all four of these structures, the part of the structure closest to the polyQ region (or its place of insertion, when considering homologous proteins) is located at the outside of the protein (data not shown). Such exposed position of the polyQ in relation to the global structure of a protein supports its involvement in protein-protein interactions (PPIs), as polyQ needs to be placed on the outside of the folded protein to have a role in PPI.

## Discussion

We previously reported the co-occurrence of coiled coils and polyQ by sequence analysis of proteins containing polyQ [5]. There, these findings were complemented with the analysis of





**Fig 3. NMR structure of yeast protein Gal11 (PDB:2LPB), residues 158–238.** One 4/4 and another 6/8 polyQ regions are coloured in white (top and bottom, respectively), while the last amino acids of a 12/12 polyQ are coloured in pink. The protein structure is shown with a ribbon representation and non polyQ regions were coloured from blue to red according to the sequence position from N- to C-terminal, respectively. The overlapping structures represent an ensemble of models from the nuclear magnetic resonance (NMR) spectra of the protein in solution.

doi:10.1371/journal.pone.0170801.g003

the protein interaction networks surrounding polyQ proteins (e.g. finding that polyQ proteins tend to interact with other polyQ proteins). Here, we used a different set of sequences, which for the most part consists of proteins of known 3D structure without polyQ but homologous to polyQ proteins. This allowed us to have a focus on the study of the structure surrounding polyQ while at the same time avoiding the problem of lack of 3D structures of polyQ regions.

We found that polyQ is generally located in a helix / random coil context where the helix is preferably N-terminal and the random coil preferably C-terminal to the polyQ middle position. This pattern of secondary structure distribution in relation to the polyQ position was also detected for lower thresholds of polyQ, like 4/6. This suggests that polyQ function is linked to a helix / random coil structure context, and that even short stretches of repeats can serve this function.

The few examples of polyQ stretches in experimental structures are all located at helix terminals. However, these could be found C- as well as N-terminal to a helix. Possibly the function of polyQ is improved if it is positioned towards the C-terminus of a helix but still takes place even at the N-terminus.

Lately, it was proposed that polyQ might extend coiled coils. A recent study found that polyQ proteins often contain coiled coils and that aggregation of polyQ proteins can be promoted by enhancing coiled coil propensity. The authors suggested that the polyQ might enhance protein-protein interaction through this coiled coil extension [22]. The finding of co-occurrence of polyQ and coiled coils was later supported by our own work [5]. Here, relying on protein sequences of known structure most often lacking the polyQ but homologous to the region surrounding a polyQ in another protein, we have shown that the point of insertion of polyQ is N-terminal to a region enriched in alpha-helix. While theoretical coiled-coil predictions might be unreliable for polyQ proteins, collectively previous work seems to suggest a

relation between polyQ and coiled-coils; here, we suggest that polyQ is often inserted right after regions with alpha-helical structure, and that, in our opinion, would be consistent with their role in extending such type of structure. We expect that experimentally verified 3D structures will eventually confirm this association. For example, a helical polyQ structure is found at the C-terminus of a coiled coil in the solved structure of the N-terminus of huntingtin [38].

Interestingly, in every structure we found where polyQ was included, the polyQ was located towards the outside of the protein, which further supports a general function for polyQ in protein-protein interaction.

PolyQ proteins are known to be often unstructured and as such difficult to crystallize [16]. Our results indicate that this is not only due to polyQ itself, but also possibly due to the place where polyQ is inserted in evolution since the lack of solved structures affects not just the polyQ protein but also the homologs of polyQ proteins that lack the polyQ (Fig 1; S1 Fig). Interestingly, the structural information reaches levels similar to the background when using a 2/6 threshold for polyQ, which suggests that indeed it should not be considered as a polyQ region.

Though the findings of our study are compelling, the scarcity of the data rather impacts their reliability. While there seems to be a clear tendency for a helix / random coil context surrounding polyQ, one should use the present findings to investigate further. In particular the suggestion that lower thresholds of polyQ definition show a similar structure context as higher ones might help defining functional polyQ regions, which should contribute to our understanding of the functions and interactions of many proteins.

## Supporting Information

**S1 Fig. Structural information per residue surrounding the place where the polyQ (8/10) is inserted in evolution, only taking into account proteins without polyQ.** Each residue could either have no structural information (“no info”, pink), or have the information: sheet (purple), random coil (red) or helix (yellow). The drop of secondary structure information at the sides of the graph only reflects the size distribution of the fragments used. (PDF)

## Acknowledgments

The authors thank all members of the CBDM group for helpful and constructive discussions. We gratefully acknowledge the support of the COST Association (Cost Action BM1405).

## Author Contributions

**Conceptualization:** PM MAAN.

**Data curation:** FT.

**Formal analysis:** FT.

**Funding acquisition:** MAAN.

**Investigation:** FT.

**Methodology:** FT.

**Project administration:** PM MAAN.

**Resources:** MAAN.



**Software:** FT.

**Supervision:** PM MAAN.

**Validation:** FT PM MAAN.

**Visualization:** PM.

**Writing – original draft:** FT PM.

**Writing – review & editing:** PM MAAN.

## References

1. Faux NG, Bottomley SP, Lesk AM, Irving JA, Morrison JR, de la Banda MG, et al. Functional insights from the distribution and role of homopeptide repeat-containing proteins. *Genome research*. 2005; 15(4):537–51. doi: [10.1101/gr.3096505](https://doi.org/10.1101/gr.3096505) PMID: [15805494](https://pubmed.ncbi.nlm.nih.gov/15805494/)
2. Pelassa I, Fiumara F. Differential Occurrence of Interactions and Interaction Domains in Proteins Containing Homopolymeric Amino Acid Repeats. *Front Genet*. 2016; 6:345.
3. Oma Y, Kino Y, Toriumi K, Sasagawa N, Ishiura S. Interactions between homopolymeric amino acids (HPAAs). *Protein science: a publication of the Protein Society*. 2007; 16(10):2195–204.
4. Hands SL, Wyttenbach A. Neurotoxic protein oligomerisation associated with polyglutamine diseases. *Acta neuropathologica*. 2010; 120(4):419–37. doi: [10.1007/s00401-010-0703-0](https://doi.org/10.1007/s00401-010-0703-0) PMID: [20514488](https://pubmed.ncbi.nlm.nih.gov/20514488/)
5. Schaefer MH, Wanker EE, Andrade-Navarro MA. Evolution and function of CAG/polyglutamine repeats in protein-protein interaction networks. *Nucleic acids research*. 2012; 40(10):4273–87. doi: [10.1093/nar/gks011](https://doi.org/10.1093/nar/gks011) PMID: [22287626](https://pubmed.ncbi.nlm.nih.gov/22287626/)
6. Albrecht M, Golatta M, Wullner U, Lengauer T. Structural and functional analysis of ataxin-2 and ataxin-3. *European journal of biochemistry / FEBS*. 2004; 271(15):3155–70.
7. Schulte J, Littleton JT. The biological function of the Huntingtin protein and its relevance to Huntington's Disease pathology. *Current trends in neurology*. 2011; 5:65–78. PMID: [22180703](https://pubmed.ncbi.nlm.nih.gov/22180703/)
8. Cattaneo E, Zuccato C, Tartari M. Normal huntingtin function: an alternative approach to Huntington's disease. *Nature reviews Neuroscience*. 2005; 6(12):919–30. doi: [10.1038/nrn1806](https://doi.org/10.1038/nrn1806) PMID: [16288298](https://pubmed.ncbi.nlm.nih.gov/16288298/)
9. Orr HT, Zoghbi HY. Trinucleotide repeat disorders. *Annual review of neuroscience*. 2007; 30:575–621. doi: [10.1146/annurev.neuro.29.051605.113042](https://doi.org/10.1146/annurev.neuro.29.051605.113042) PMID: [17417937](https://pubmed.ncbi.nlm.nih.gov/17417937/)
10. Zoghbi HY, Orr HT. Glutamine repeats and neurodegeneration. *Annual review of neuroscience*. 2000; 23:217–47. doi: [10.1146/annurev.neuro.23.1.217](https://doi.org/10.1146/annurev.neuro.23.1.217) PMID: [10845064](https://pubmed.ncbi.nlm.nih.gov/10845064/)
11. Hands S, Sinadinos C, Wyttenbach A. Polyglutamine gene function and dysfunction in the ageing brain. *Biochimica et biophysica acta*. 2008; 1779(8):507–21. doi: [10.1016/j.bbaggm.2008.05.008](https://doi.org/10.1016/j.bbaggm.2008.05.008) PMID: [18582603](https://pubmed.ncbi.nlm.nih.gov/18582603/)
12. Arrasate M, Mitra S, Schweitzer ES, Segal MR, Finkbeiner S. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature*. 2004; 431(7010):805–10. doi: [10.1038/nature02998](https://doi.org/10.1038/nature02998) PMID: [15483602](https://pubmed.ncbi.nlm.nih.gov/15483602/)
13. Klein FA, Pastore A, Masino L, Zeder-Lutz G, Nierengarten H, Oulad-Abdelghani M, et al. Pathogenic and non-pathogenic polyglutamine tracts have similar structural properties: towards a length-dependent toxicity gradient. *Journal of molecular biology*. 2007; 371(1):235–44. doi: [10.1016/j.jmb.2007.05.028](https://doi.org/10.1016/j.jmb.2007.05.028) PMID: [17560603](https://pubmed.ncbi.nlm.nih.gov/17560603/)
14. Chen S, Berthelie V, Yang W, Wetzel R. Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity. *Journal of molecular biology*. 2001; 311(1):173–82. doi: [10.1006/jmbi.2001.4850](https://doi.org/10.1006/jmbi.2001.4850) PMID: [11469866](https://pubmed.ncbi.nlm.nih.gov/11469866/)
15. Robertson AL, Horne J, Ellisdon AM, Thomas B, Scanlon MJ, Bottomley SP. The structural impact of a polyglutamine tract is location-dependent. *Biophysical journal*. 2008; 95(12):5922–30. doi: [10.1529/biophysj.108.138487](https://doi.org/10.1529/biophysj.108.138487) PMID: [18849414](https://pubmed.ncbi.nlm.nih.gov/18849414/)
16. Masino L. Polyglutamine and neurodegeneration: structural aspects. *Protein and peptide letters*. 2004; 11(3):239–48. PMID: [15182225](https://pubmed.ncbi.nlm.nih.gov/15182225/)
17. Papaleo E, Invernizzi G. Conformational diseases: structural studies of aggregation of polyglutamine proteins. *Current computer-aided drug design*. 2011; 7(1):23–43. PMID: [20807186](https://pubmed.ncbi.nlm.nih.gov/20807186/)
18. Ross CA, Poirier MA, Wanker EE, Amzel M. Polyglutamine fibrillogenesis: the pathway unfolds. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100(1):1–3. doi: [10.1073/pnas.0237018100](https://doi.org/10.1073/pnas.0237018100) PMID: [12509507](https://pubmed.ncbi.nlm.nih.gov/12509507/)

19. Poirier MA, Jiang H, Ross CA. A structure-based analysis of huntingtin mutant polyglutamine aggregation and toxicity: evidence for a compact beta-sheet structure. *Human molecular genetics*. 2005; 14(6):765–74. doi: [10.1093/hmg/ddi071](https://doi.org/10.1093/hmg/ddi071) PMID: [15689354](https://pubmed.ncbi.nlm.nih.gov/15689354/)
20. Davies P, Watt K, Kelly SM, Clark C, Price NC, McEwan IJ. Consequences of poly-glutamine repeat length for the conformation and folding of the androgen receptor amino-terminal domain. *Journal of molecular endocrinology*. 2008; 41(5):301–14. doi: [10.1677/JME-08-0042](https://doi.org/10.1677/JME-08-0042) PMID: [18762554](https://pubmed.ncbi.nlm.nih.gov/18762554/)
21. Walters RH, Murphy RM. Examining polyglutamine peptide length: a connection between collapsed conformations and increased aggregation. *Journal of molecular biology*. 2009; 393(4):978–92. doi: [10.1016/j.jmb.2009.08.034](https://doi.org/10.1016/j.jmb.2009.08.034) PMID: [19699209](https://pubmed.ncbi.nlm.nih.gov/19699209/)
22. Fiumara F, Fioriti L, Kandel ER, Hendrickson WA. Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell*. 2010; 143(7):1121–35. doi: [10.1016/j.cell.2010.11.042](https://doi.org/10.1016/j.cell.2010.11.042) PMID: [21183075](https://pubmed.ncbi.nlm.nih.gov/21183075/)
23. Petrakis S, Schaefer MH, Wanker EE, Andrade-Navarro MA. Aggregation of polyQ-extended proteins is promoted by interaction with their natural coiled-coil partners. *BioEssays: news and reviews in molecular, cellular and developmental biology*. 2013; 35(6):503–7.
24. Bhattacharyya A, Thakur AK, Chellgren VM, Thiagarajan G, Williams AD, Chellgren BW, et al. Oligoproline effects on polyglutamine conformation and aggregation. *Journal of molecular biology*. 2006; 355(3):524–35. doi: [10.1016/j.jmb.2005.10.053](https://doi.org/10.1016/j.jmb.2005.10.053) PMID: [16321399](https://pubmed.ncbi.nlm.nih.gov/16321399/)
25. Pelassa I, Cora D, Cesano F, Monje FJ, Montarolo PG, Fiumara F. Association of polyalanine coiled coils mediates expansion disease-related protein aggregation and dysfunction. *Hum Mol Gen*. 2014; 23(13):3402–20.
26. de Chiara C, Menon RP, Dal Piaz F, Calder L, Pastore A. Polyglutamine is not all: the functional role of the AXH domain in the ataxin-1 protein. *Journal of molecular biology*. 2005; 354(4):883–93. doi: [10.1016/j.jmb.2005.09.083](https://doi.org/10.1016/j.jmb.2005.09.083) PMID: [16277991](https://pubmed.ncbi.nlm.nih.gov/16277991/)
27. Saunders HM, Gilis D, Rooman M, Dehouck Y, Robertson AL, Bottomley SP. Flanking domain stability modulates the aggregation kinetics of a polyglutamine disease protein. *Protein science: a publication of the Protein Society*. 2011; 20(10):1675–81.
28. Sahoo B, Arduini I, Drombosky KW, Kodali R, Sanders LH, Greenamyre, Wetzel R. Folding landscape of mutant huntingtin exon1: diffusible multimers, oligomers and fibrils, and no detectable monomer. *PLoS One*. 2016; 11(6).
29. Ruff KM, Khan SJ, Pappu RV. A coarse-grained model for polyglutamine aggregation modulated by amphipatic flanking sequences. *Biophys J*. 2014; 107(5):1226–35. doi: [10.1016/j.bpj.2014.07.019](https://doi.org/10.1016/j.bpj.2014.07.019) PMID: [25185558](https://pubmed.ncbi.nlm.nih.gov/25185558/)
30. Kokona B, Rosenthal ZP, Fairman R. Role of the coiled-coil structural motif in polyglutamine aggregation. *Biochemistry*. 2014; 53(43):6738–46. doi: [10.1021/bi500449a](https://doi.org/10.1021/bi500449a) PMID: [25310851](https://pubmed.ncbi.nlm.nih.gov/25310851/)
31. Kokona B, Johnson KA, Fairman R. Effect of helical flanking sequences on the morphology of polyglutamine-containing fibrils. *Biochemistry*. 2014; 53(43):6747–53. doi: [10.1021/bi501066q](https://doi.org/10.1021/bi501066q) PMID: [25207433](https://pubmed.ncbi.nlm.nih.gov/25207433/)
32. Mier P, Andrade-Navarro MA. FastHerder2: Four Ways to Research Protein Function and Evolution with Clustering and Clustered Databases. *Journal of computational biology: a journal of computational molecular cell biology*. 2016; 23(4):270–8.
33. UniProt C. UniProt: a hub for protein information. *Nucleic acids research*. 2015; 43(Database issue): D204–12. doi: [10.1093/nar/gku989](https://doi.org/10.1093/nar/gku989) PMID: [25348405](https://pubmed.ncbi.nlm.nih.gov/25348405/)
34. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*. 2011; 7:539. doi: [10.1038/msb.2011.75](https://doi.org/10.1038/msb.2011.75) PMID: [21988835](https://pubmed.ncbi.nlm.nih.gov/21988835/)
35. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic acids research*. 2000; 28(1):235–42. PMID: [10592235](https://pubmed.ncbi.nlm.nih.gov/10592235/)
36. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*. 2004; 25(13):1605–12. doi: [10.1002/jcc.20084](https://doi.org/10.1002/jcc.20084) PMID: [15264254](https://pubmed.ncbi.nlm.nih.gov/15264254/)
37. Hildebrandt A, Dehof AK, Rurainski A, Bertsch A, Schumann M, Toussaint NC, et al. BALL—biochemical algorithms library 1.3. *BMC bioinformatics*. 2010; 11:531. doi: [10.1186/1471-2105-11-531](https://doi.org/10.1186/1471-2105-11-531) PMID: [20973958](https://pubmed.ncbi.nlm.nih.gov/20973958/)
38. Kim MW, Chelliah Y, Kim SW, Otwinowski Z, Bezprozvanny I. Secondary structure of Huntingtin amino-terminal region. *Structure*. 2009; 17(9):1205–12. doi: [10.1016/j.str.2009.08.002](https://doi.org/10.1016/j.str.2009.08.002) PMID: [19748341](https://pubmed.ncbi.nlm.nih.gov/19748341/)