

## RESEARCH ARTICLE

# An Exploration of the Universe of Polyglutamine Structures

Àngel Gómez-Sicilia<sup>1,2</sup>, Mateusz Sikora<sup>3</sup>, Marek Cieplak<sup>4\*</sup>, Mariano Carrión-Vázquez<sup>1,2</sup>

**1** Instituto Cajal/CSIC, Madrid, Spain, **2** Instituto Madrileño de Estudios Avanzados en Nanociencia (IMDEA-Nanociencia), Madrid, Spain, **3** Institute of Science and Technology Austria, Klosterneuburg, Austria, **4** Instytut Fizyki PAN, Warsaw, Poland

\* [mc@ifpan.edu.pl](mailto:mc@ifpan.edu.pl)



## Abstract

Deposits of misfolded proteins in the human brain are associated with the development of many neurodegenerative diseases. Recent studies show that these proteins have common traits even at the monomer level. Among them, a polyglutamine region that is present in huntingtin is known to exhibit a correlation between the length of the chain and the severity as well as the earliness of the onset of Huntington disease. Here, we apply bias exchange molecular dynamics to generate structures of polyglutamine expansions of several lengths and characterize the resulting independent conformations. We compare the properties of these conformations to those of the standard proteins, as well as to other homopolymeric tracts. We find that, similar to the previously studied polyvaline chains, the set of possible transient folds is much broader than the set of known-to-date folds, although the conformations have different structures. We show that the mechanical stability is not related to any simple geometrical characteristics of the structures. We demonstrate that long polyglutamine expansions result in higher mechanical stability than the shorter ones. They also have a longer life span and are substantially more prone to form knotted structures. The knotted region has an average length of 35 residues, similar to the typical threshold for most polyglutamine-related diseases. Similarly, changes in shape and mechanical stability appear once the total length of the peptide exceeds this threshold of 35 glutamine residues. We suggest that knotted conformers may also harm the cellular machinery and thus lead to disease.

## OPEN ACCESS

**Citation:** Gómez-Sicilia À, Sikora M, Cieplak M, Carrión-Vázquez M (2015) An Exploration of the Universe of Polyglutamine Structures. *PLoS Comput Biol* 11(10): e1004541. doi:10.1371/journal.pcbi.1004541

**Editor:** Guanghong Wei, Fudan University, CHINA

**Received:** April 30, 2015

**Accepted:** September 8, 2015

**Published:** October 23, 2015

**Copyright:** © 2015 Gómez-Sicilia et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://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:** For the studied sets of Q<sub>n</sub> with n = 16, 20, 25, 30, 33, 38, 40, 60 and 80, as well as for V60, the independent conformers can be found in [www.ifpan.edu.pl/~cieplak/POLYQ](http://www.ifpan.edu.pl/~cieplak/POLYQ).

**Funding:** We acknowledge the support by the EU Joint Programme in Neurodegenerative Diseases (JPND AC14/00037) project. The project is supported through the following funding organisations under the aegis of JPND—[www.jpnd.eu](http://www.jpnd.eu): Ireland, HRB; Poland, National Science Centre; and Spain, ISCIII. MCV acknowledges a grant by the Spanish Ministry of Economy (MINECO SAF2013-49179-C2-1-R). ÀGS was recipient of JAE-Pre and EMBO Short Term fellowships. The funders had no role in study design,

## Author Summary

Misfolding and aggregation of several proteins are known to be related to neurodegenerative diseases. Among them, polyglutamine expansions are known to be responsible for at least 9 diseases, including Huntington. Nonetheless, the structural properties of these intrinsically disordered proteins are difficult to study using classical techniques because of their rapid fluctuations that result in high conformational polymorphism. Here, we use molecular dynamics simulations to study polyglutamines of different chain lengths, starting with short non-pathogenic ones, and study the independent structures they are able to

data collection and analysis, decision to publish, or preparation of the manuscript.

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

form. We characterize all structures by their geometrical properties, connectivity, putative mechanical stability and residence time (life span). Similar to the findings of a previous study with polyvalines, only some of the conformers are similar to those found in natural globular proteins. Moreover, we find structures that contain knots in both polyglutamine and polyvaline 60-mers, although the former contains many more knotted conformers than the latter. We suggest that these knotted conformers may impair the cell machinery for degradation and eventually lead to toxicity.

## Introduction

Less than two thousands protein folds have been identified in nature [1, 2], indicating that similar folds can be adopted by large numbers of sequences. These folds have been characterized and classified in the CATH database [3]. Recently, Cossio *et al.* [4] considered a single sequence—polyvaline (polyV) 60-mer (denoted here as  $V_{60}$ )—and generated, through all-atom simulations, an exhaustive database with 30 063 conformations. Interestingly, only a small fraction of the  $V_{60}$  conformations turned out to be CATH-like in that they had at least one similar structure in the CATH database. The similarity was assessed by a TM-score being higher than 45%. The score is obtained through an algorithm for protein comparison based on secondary structure alignment [5]. Thus they explored, in their own words [4], the universe of protein structures beyond the Protein Data Bank. They argued that there must be an evolutionary principle that favors shorter loops and directs the evolution to a certain spot in the universe of possible conformations.

Long polyV chains do not exist in nature. However, many proteins in eukaryotic cells contain homopolymeric tracts, defined as repetitions of the same residue. In particular, upon inspection of the revised human proteome stored in UniProt Knowledge Database [6], we found that 18.9% of the human proteome involves homopolymeric tracts of size 5 or greater, while the probability of one happening by chance is  $6 \cdot 10^{-6}$ . Among these, the longest chains have been found for polyserine (polyS, 58 repeats, in the TNRC18 protein, with a random probability of  $4 \cdot 10^{-76}$ ) and polyglutamine (polyQ, 40 repeats, in FOXP2 protein, random probability of  $9 \cdot 10^{-53}$ ).

PolyQ chains are known to be responsible for several brain disorders, including Huntington disease (HD). HD is caused by a protein in the human brain known as huntingtin (HTT)—of, as yet, not fully elucidated function. HTT is known to be highly involved in development [7], and is thought to be related to gene expression regulation [8] and to anchoring or transport of vesicles [9]. A HTT mutant with an expansion of polyQ that exceeds the threshold of about 35-mer was linked to the disease [10]. Even though polyQ tracts have been extensively studied [11–13], the molecular physiopathology behind the connection between sequence length and disease remains elusive.

Another example of disease-related homopolymeric tracts is polyalanine (polyA), occurring in transcription factors. Expansions of the polyA tracts beyond certain thresholds (*e.g.* 19) have been recognized as the cause of congenital malformation syndromes, skeletal dysplasia and nervous system anomalies [14, 15]. The strong evolutionary conservation of the polyA tracts suggests the existence of critical structural or functional constraints [14]. It should be noted that in human proteins the polyA tracts are short compared to those of polyQ [15].

Here, we focus on polyQ chains of various lengths,  $Q_n$ , where  $n$  goes from 16 to 80. The case of  $n = 62$  was the subject of a recent single-molecule force spectroscopy study [16] that revealed a large conformational polymorphism (monitored as a spectrum of different breaking points

and characteristic force-peak heights, up to 800 pN). The questions we ask are as follows: 1) can we explain this conformational polymorphism? 2) can polyQ tracts generate non-CATH-like conformers? 3) what are the structural and mechanical properties of the polyQ structures? In order to answer them, we follow a bias exchange molecular dynamics approach (BEMD) [17] also used by Cossio *et al.* [4]—one of the meta dynamics approaches—and explore the structural and dynamical properties of  $Q_n$  with a particular focus on  $Q_{20}$  and  $Q_{60}$ , representative examples below and above the HD's pathological threshold.

We take two perspectives in our analysis: 1) making comparisons of  $Q_{60}$  to  $V_{60}$  and to the similar-sized proteins from the CATH database; 2) investigating the changes in the physical properties of the conformations corresponding to  $Q_n$  as one varies  $n$ . The dynamical properties can be conveniently captured by their mechanical stability, as characterized by the characteristic force,  $F_{\max}$ , needed to unravel a structure by pulling by its termini at a constant speed,  $v_p$ . This part of our studies makes use of a structure-based coarse-grained model [18, 19] to access the regime of near-experimental speeds and to deal with the large statistics. It should be noted that typical fluctuation times of these structures are much smaller than those needed to fully unravel them [20]. Thus, the results on  $F_{\max}$  are merely indicators of the putative mechanical stability of each specific conformer that do not take the intrinsic evolution of the disordered protein into account.

We find that relatively large mechanical stability may arise not only from structures with large secondary structure content (SS, measured as the percentages of residues belonging to  $\alpha$ -helices,  $\beta$ -strands and hydrogen-bonded turns) but also from those with SS of about 30%. Interestingly, we also find spontaneous generation of knotted structures for  $n = 60$ , which tend to be of a size of 36 residues, about HD's threshold. This is a novel feature in neurotoxic proteins that needs further investigation.

## Methods

### Generation and selection of structures

Our BEMD [17] simulations were carried out using the GROMACS molecular dynamics package [21] and the PLUMED extension [22]. The force field used is AMBER99SB [23] and the implicit solvent model is the generalized Born surface area method [24]. The same force field has been used before in folding simulations with explicit solvent [17, 25], but implicit solvent is preferred in order to efficiently explore the energy landscape [4]. Structures were initialized randomly using the MODELLER software [26]: 10 off-template models were done for each protein; the models that contained knots were discarded and the remaining ones were minimized through up to 1000 steps of the steepest descent method followed by up to 4000 steps of the conjugate gradient algorithm [27]. The system which acquired the smallest potential energy after the two minimization stages was chosen for further studies.

In order to generate a variety of  $Q_n$  structures, we applied the BEMD method with six replicas, each with a different secondary structure bias: the first one with no bias; the next three with a preference to  $\alpha$ -helix in the first, second and last third of the chain sequence; the fifth with a preference to anti-parallel  $\beta$ -strands and the last one with parallel  $\beta$ -strands. The method is explained in detail in the [S1 Text](#).

We first obtained a number of conformations with a varying secondary structure content. To select the structures of interest, we followed the three-sieve protocol used in [4], described in the [S1 Text](#), that yields structures with SS > 30 % which are temporally and structurally independent. From a 2  $\mu$ s simulation of  $Q_{60}$ , 246 independent conformers were obtained from 953 time clusters. For  $Q_{20}$ , a 0.66  $\mu$ s simulation resulted into 491 independent conformers out of 517 time clusters. Interestingly, half the simulation time for the short peptide yielded twice

as many independent conformations as the longer one, which indicates higher polymorphism and faster dynamics in  $Q_{20}$ . The procedure led to the emergence of some knotted structures even though there were no knots present in the initial homology-derived conformers.

After clustering, all independent structures underwent a minimization process of 10 000 steepest descent steps or until the maximum force between a pair of atoms was smaller than 0.25 J/(mol nm) so that the structure in the closest energy minimum is obtained. In this process, some of the residues may form or break contacts, thus changing their secondary structure content slightly. Therefore, even though the structures were selected with  $SS \geq 30\%$  before the first clustering, some of the final structures may have a smaller  $SS$  content.

The structures of  $V_{60}$  were taken from ref. [4]. Their 50  $\mu$ s simulation yielded 30 063 time clusters. We have applied the structural clustering to this set and obtained 7076 independent conformers—as these were not available. All the independent structures generated are provided in the [S1 Text](#).

Three previous works have explored the aggregation properties of  $Q_n$  by computer simulations. In the first of them, the focus is on the temporal stability of the structures and on the evaluation of their amyloidogenesis and fibrillation capabilities [28]. The second study explores the landscape of possible conformations by simplifying the structure of glutamine and generating a model that efficiently samples many conformations [29]. Finally, the third one applies replica exchange molecular dynamics to explore the dimerization of polyQ [30]. The three works show structures such as steric zippers and a  $\beta$ -helix, which have been found in our sampling among the strongest  $Q_{60}$  and  $Q_{20}$  conformations shown in [S1](#) and [S2](#) Figs. Furthermore, rod-like conformers with close to 100%  $\alpha$ -helical content were also described in the aforementioned works and have likewise been found in this sampling. Our simulations did not find the mainly  $\beta$  conformations suggested in Ref. [30] because we consider monomers instead of dimers.

## Descriptors of the structures

Our structural analysis deals with several descriptors. One is the radius of gyration,  $R_g$ , which characterizes the linear size of the molecule. Another is the  $w$  parameter which describes the shape [31, 32], which is defined through the diagonalization of the tensor of inertia and by making combinations of the three main radii such that a near-zero  $w$  corresponds to a globular shape, a positive  $w$  to an elongated one, and a negative  $w$  to a flattened object. The third descriptor is the  $SS$  parameter, which is determined by using the DSSP procedure [33]. This parameter is a sum of several ingredients: the  $\alpha$ -helical content, the  $\beta$ -content (strands and bridges), and the hydrogen-bonded turn content.

The next two descriptors are  $F_{\max}$  and  $\langle z \rangle$ —the average coordination number. The former relates to the dynamics directly, while the latter relates to it indirectly since  $z$  measures the number of residues a given residue interacts with. These interactions are of two kinds: through the peptide bond with the two nearest residues along the sequence and through contact interactions with residues which are not sequential neighbors. The contacts play a dynamical role in coarse-grained structure-based models but they can also be used as descriptors in all-atom models. The specific definition of the contacts we use is based on enlarged van der Waals spheres associated with the heavy atoms [18, 19] and the radii of the spheres are given in ref. [34]: a contact between two residues exists if there is at least one pair of heavy atoms with overlapping spheres.

In the structure-based model, we assign Lennard-Jones potentials of depth  $\epsilon$  to these contacts (the potential minimum is at the distance between the  $C_\alpha$  atoms in the reference structure) so that larger values of  $\langle z \rangle$  are expected to correspond to more stable structures. Maxwell demonstrated [35] that large three-dimensional systems of particles with pairwise interactions

are stable provided the  $\langle z \rangle$  is bigger than 6. In particular, this finding has been shown to be consistent with the behavior of virus capsids [36]. In our case, the structures are much smaller than the capsids and, therefore, the threshold value of  $\langle z \rangle$  is reduced, as explained in the [S1 Text](#). Furthermore, the systems we study are also endowed with the local backbone stiffness—a four-particle interaction [37]—which favors the local chirality of the reference state. Thus non-zero values of  $F_{\max}$  for structures with  $\langle z \rangle$  smaller than 6 are allowed.

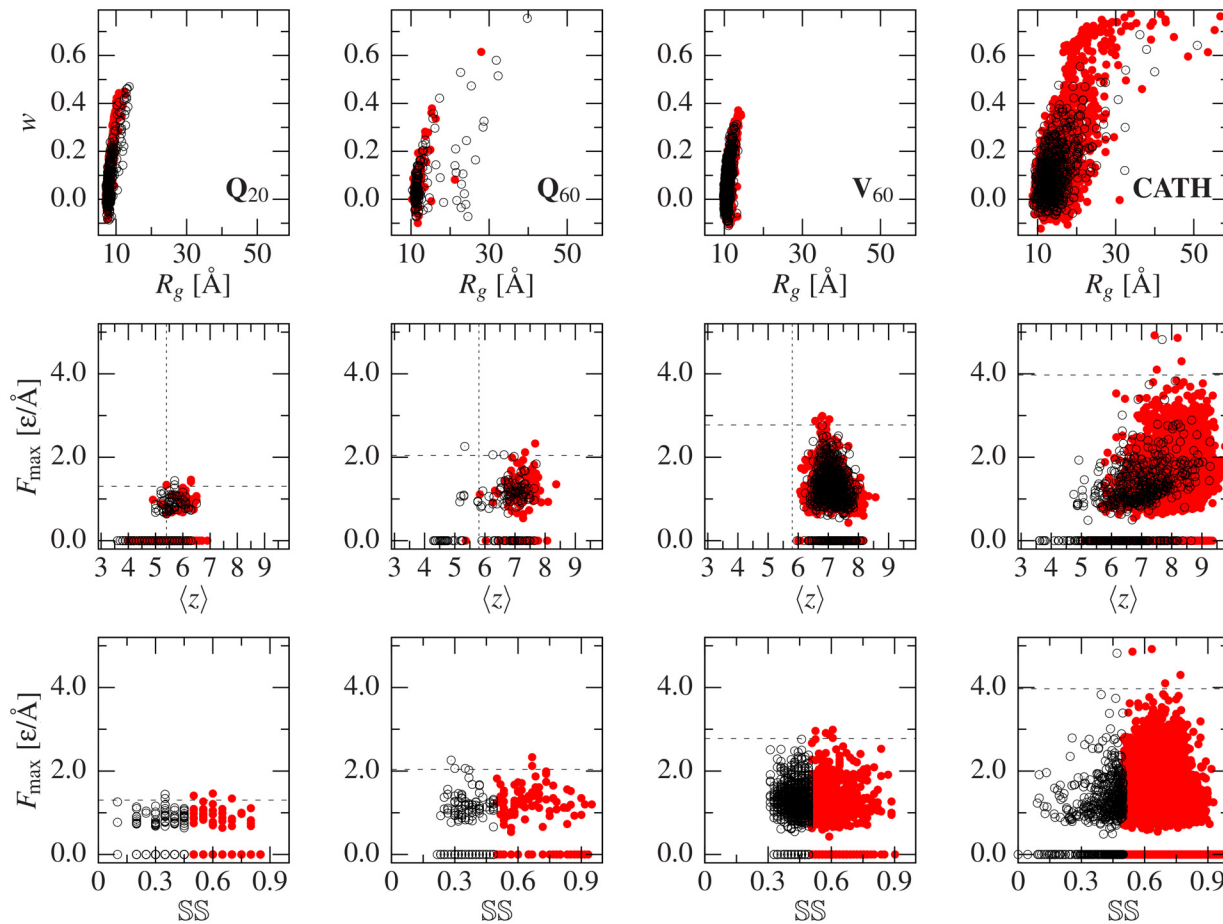
A stretching force ( $F$ ) vs. displacement ( $d$ ) curve may include articulated force peaks—which exceed the thermal noise level of about  $0.1 \text{ } \epsilon/\text{\AA}$ —before the  $F$  raises indefinitely due to stretching of the peptide bonds. The calculations are done at the temperature of  $0.3 \text{ } \epsilon/k_B$ , where  $k_B$  denotes the Boltzmann constant. The number of peaks is denoted by  $n_p$ .  $F_{\max}$  is defined as the height of the largest peak. If none exists then  $F_{\max}$  is defined to be zero, even though it could take any value below the baseline of the curve (around  $0.4 \text{ } \epsilon/\text{\AA}$ ). We simulate stretching at  $v_p$  of  $5 \cdot 10^{-3} \text{ } \text{\AA}/\tau$ , where  $\tau$  is of order 1 ns. Experimental  $v_p$ 's are typically lower, e.g.  $4 \cdot 10^{-6} \text{ } \text{\AA}/\text{ns}$  in ref. [16]. We have calibrated [19]  $\epsilon$  to be 110 pN  $\text{\AA}$  (with a 25% error bar) by comparing theoretical and experimental values of  $F_{\max}$  in 38 proteins (the theoretical results involved extrapolation to the experimental  $v_p$ 's). The temperature in the coarse-grained simulations is in the vicinity of the room temperature.

Finally, one can consider the contact order ( $CO$ ) as yet another structural descriptor. It is related to the number of contacts as well as the average distance along the sequence between the contacting residues, as defined in Ref. [38] as  $CO = \frac{1}{L \cdot N} \sum_k S_k$ , where  $S_k$  is the distance between the residues that form contact  $k$ ,  $L$  is the number of residues in the protein and  $N$  the number of contacts. There is a question whether  $CO$  correlates with the folding time or not (see [39] and [40] for the arguments for and against it), so one may also inquire whether  $F_{\max}$  correlates with  $CO$ . It seems unlikely that the free energy barrier to mechanical unfolding is of the same nature as the one for folding (or thermal unfolding) [41] but this does not preclude a correlation with  $CO$ . However, we do not find the correlation to be valid (see the [S1 Text](#)) which is consistent with the fact that the green fluorescent protein (PDB code 1GFL) has a bigger  $F_{\max}$  than the I27 domain of titin (the PDB code 1TIT), 2.7 vs. 2.1  $\epsilon/\text{\AA}$  [19], whereas its  $CO$  is smaller, 0.22 vs. 0.36.

## Results

### Properties of $Q_{60}$ and $Q_{20}$

We first consider the  $Q_{60}$  set, so that one can compare with  $V_{60}$  from [4] and with the experimental results on  $Q_{62}$  in [16]. [S1 Fig](#) shows structures corresponding to the top five values of  $F_{\max}$ . Similar figures for other sets studied are shown in [S2](#) and [S3 Figs](#). The values range between 2.1 and 2.3  $\epsilon/\text{\AA}$  (approximately between 230 and 250 pN) which is of the order of what has been found—about 200 pN—for the I27 domain of titin [42, 43] at smaller  $v_p$ 's. The fact that these values are much smaller than the ones found experimentally in [16] can be attributed to a small statistics, since the experimental systems yielded high force only with low probability ( $p(F_{\max} > 200 \text{ pN}) = 7 \pm 6\%$ ). The figure also shows the corresponding  $F - d$  patterns together with distances at which particular contacts break down (the distance in the contact exceeds the reference distance by 50%) for the last time. The contacts are labeled by the sequential distances  $|i - j|$  between residues  $i$  and  $j$ . The number of force peaks varies between 1 and 4, corresponding to several substructures forming in each conformer. The third column of panels in [S1 Fig](#) provides the values of the relevant descriptors.  $R_g$  is seen to range between 11.05 and 14.20  $\text{\AA}$  and the values of  $w$  indicate that the fifth structure is elongated while the other four are nearly globular. The most stable structure of the five shown (the top row of panels) corresponds to the largest  $\langle z \rangle$  (7.67), and  $\text{SS}$  (66.7%)—the secondary structure is, in this case, exclusively of the  $\beta$  type.



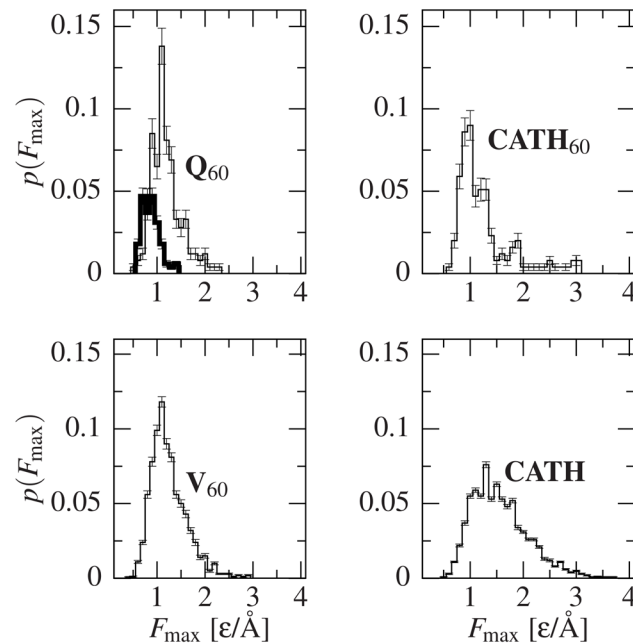
**Fig 1. Scatter plot relating the specified variables for four different sets, from left to right, Q<sub>20</sub>, Q<sub>60</sub>, V<sub>60</sub> and CATH.** The empty black points represent the conformers with less than 50% secondary structure content, while the filled red dots represent the more structured conformers. The vertical dotted lines in the middle panels mark the simply stiff limits of stability for each case (see the S1 Text). The conformers to the left of this line are more volatile. The horizontal dashed lines in the middle and bottom panels mark off the top five conformers with respect to the value of  $F_{\max}$ .

doi:10.1371/journal.pcbi.1004541.g001

Interestingly, the second most robust structure, as judged by the value of  $F_{\max}$ , has very low  $\langle z \rangle$  (5.33) and SS (28.3%).

This system was compared to Q<sub>20</sub>, which is unrelated to disease and is close to Q<sub>19</sub>, which was studied experimentally [16]. The two left columns of Fig 1 refer to all structures in sets Q<sub>20</sub> and Q<sub>60</sub>. In particular, the first row represents the geometries obtained on the  $R_g$ — $w$  plane. The convention we use is that we represent the data corresponding to structures with SS of at least 50% and those with lower SS by filled and open symbols, respectively. The structures from Q<sub>60</sub> are seen to overlap with the region taken by Q<sub>20</sub> but they also extend to much larger  $R_g$  and to bigger  $w$ . The largest value of  $R_g$  corresponds to a low SS, while high  $w$  can be achieved with any value of SS.

The two bottom rows of Fig 1 show scatter plots that compare the values of  $F_{\max}$  in Q<sub>60</sub> to those in Q<sub>20</sub> when represented as a function of  $\langle z \rangle$  and SS. S4 Fig provides a continuation in which  $F_{\max}$  is plotted against the  $\alpha$ -,  $\beta$ -, and turns ( $\tau$ ) content. It is clear that, for a given  $n$ , the mechanical stability is not related in any simple manner to either  $\langle z \rangle$ , CO or SS content. This is because typical high-force motifs include  $\beta$ -structured regions, while high  $\langle z \rangle$  and SS can be achieved with  $\alpha$ -structure and hydrogen-bonded turns [43]. S4 Fig shows that mechanical stability has no direct correlation to  $\alpha$ -content or hydrogen-bonded turns ( $\tau$ ), and while most of



**Fig 2. Distributions of  $F_{\max}$  for the studied species.** The top left panel shows the distribution for  $Q_{20}$  in a thick line. The conformations with no force peaks are not plotted in the histograms but contribute to normalization. The amount of such non-mechanostable conformers is  $(79 \pm 2)\%$  for  $Q_{20}$ ,  $(34 \pm 3)\%$  for  $Q_{60}$ ,  $(16.5 \pm 0.2)\%$  for  $V_{60}$ , and  $(47 \pm 3)\%$  and  $(20.2 \pm 0.5)\%$  for  $CATH_{60}$  and  $CATH$ , respectively. The errors were computed using a bootstrapping method and the size of the error bar indicates the standard deviation.

doi:10.1371/journal.pcbi.1004541.g002

the high  $\beta$ -content conformers lead to high forces, these can also be observed in cases with no  $\beta$ -content. Similar results are shown in the top panels of [S5 Fig](#) for the CO.

Furthermore, in the top left panel of [Fig 2](#) there is a comparison of the distributions of  $F_{\max}$  for  $Q_{60}$  and  $Q_{20}$ . We observe that although our BEMD simulations bias the chain towards the acquisition of  $\beta$ SS, many conformers do not produce any articulated force peaks above the 10% noise level ( $F_{\max} = 0$ ). In particular,  $Q_{20}$  presents  $(79 \pm 2)\%$  of this kind of conformers, while  $Q_{60}$  only contains  $(34 \pm 3)\%$  of them. This result is consistent with the experimental data [[16](#)], where no force peaks were detected in  $Q_{19}$ , while some were found in  $Q_{62}$ . Remarkably, although the diversity in mechanical stability for  $Q_{20}$  is smaller than for  $Q_{60}$ , the frequency of independent structure generation is greater in the former (see [S6 Fig](#)), so its conformational polymorphism should be higher. The volatility of each conformer, as assessed by  $\langle z \rangle$  lower than their threshold, also agrees with this result, with  $(49 \pm 2)\%$  volatile conformers in  $Q_{20}$  vs.  $(13 \pm 2)\%$  in  $Q_{60}$ .

Taken together, these results show that  $F_{\max}$  is inherently different in  $Q_{20}$  and  $Q_{60}$  sets, even when  $\beta$ SS is similar. This further points to  $F_{\max}$  not being related to hydrogen-bonded turns,  $\alpha$  helices and even  $\beta$ -strand content and  $\beta$ SS, and also neither to CO nor to  $\langle z \rangle$ .

### Comparisons of $Q_{60}$ to the remaining sets of structures

[Fig 2](#) shows the normalized distributions of  $F_{\max}$  within the  $Q_{60}$ ,  $Q_{20}$ ,  $V_{60}$ ,  $CATH_{60}$ , and  $CATH$  sets. The distributions do not show the peak at  $F_{\max} = 0$ , but its value is shown in the caption and contributes to the normalization.  $CATH_{60}$  is defined as structures containing between 57 and 63 residues and it contains 256 proteins. In order to exclude short peptides and most multidomain proteins we take  $CATH$  to represent all those proteins in the  $CATH$  database that are 40 to 250 residue long. This set comprises 5403 structures.

The characteristic forces grow on moving from  $Q_{20}$  to  $Q_{60}$ . However, in all sets but  $Q_{20}$ , the most probable  $F_{\max}$  is about the same,  $1.2 \text{ e}/\text{\AA}$ , but the shapes of the distributions differ. The distributions are comparably broad for CATH and  $CATH_{60}$  and comparably narrower for  $Q_{60}$  and  $V_{60}$ , indicating the role of the stronger compositional homogeneity in the latter two sets. The rougher look of the distribution for  $Q_{60}$  is likely due to the one order of magnitude smaller statistics. Furthermore, it should be noted that, among the systems of about 60 residues,  $CATH_{60}$  leads to the biggest number of situations with no force peaks,  $(47 \pm 3)\%$ ; and  $V_{60}$  to the smallest,  $(16.5 \pm 0.2)\%$ .

Despite the similarity of the distribution of the forces between  $Q_{60}$  and  $V_{60}$ , the geometrical character of structures in the two sets are distinct. Fig 1 shows that  $V_{60}$  conformers are more compact and less elongated than  $Q_{60}$  or  $CATH_{60}$ . Furthermore, this figure indicates that size and shape of a chain need not be correlated.

Fig 1 further shows that most of the structures in the 60-sized sets, and also for CATH, come with  $\langle z \rangle$  between 5.5 and 8.5. However, the largest values of  $F_{\max}$  arise for  $\langle z \rangle$  between 6.3 and 8.1 and many large  $\langle z \rangle$  structures come with average or even small forces, including  $F_{\max}$  of 0. Similarly, Fig 1 also demonstrates that large  $\text{SS}$  may come with low or zero forces and large  $F_{\max}$  may arise when  $\text{SS}$  is at its lower range. Furthermore, the scattering of the data points in the  $F_{\max}$ -CO plane shown in Fig. [38] also points in the direction of statistical independence. This observation further proves that there is no correlation between  $F_{\max}$  and  $\langle z \rangle$ , CO or  $\text{SS}$ . Interestingly, none of the independent  $V_{60}$  structures obtained from [4] have  $\langle z \rangle$  below the volatility threshold, while  $(5.5 \pm 1.4)\%$  of the ones in  $CATH_{60}$  and  $(2.3 \pm 0.2)\%$  of CATH do. Our comparison reinforces the remarkable conclusion that  $F_{\max}$  is unrelated to CO,  $\text{SS}$  or  $\langle z \rangle$  and extends it to general globular proteins instead of being a property specific for polyQ. This is further discussed in the S1 Text, where the independence of  $F_{\max}$  with the rest of parameters is proved.

## Life span of the structures

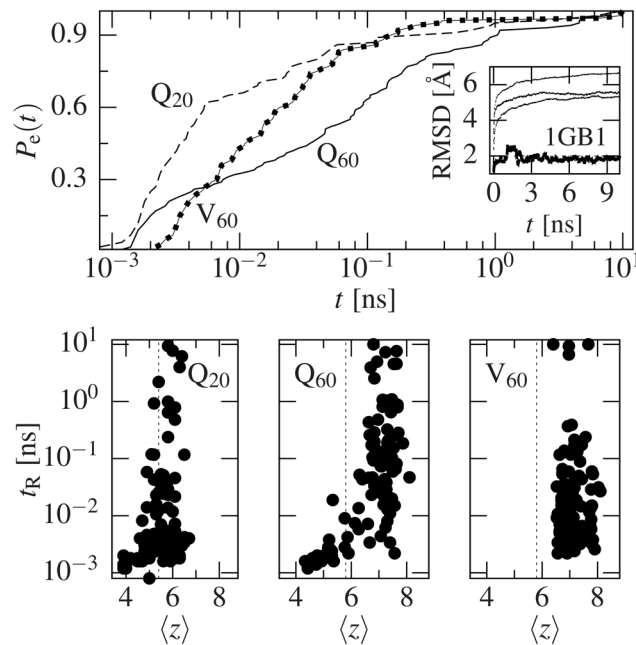
In order to test whether the coordination number is actually related to temporal stability, we performed 10 ns free-dynamics simulations on 100 structures chosen randomly from each set:  $Q_{20}$ ,  $Q_{60}$  and  $V_{60}$ . We have studied the time dependence of RMSD relative to the initial structure and the last time that it fluctuated below  $2 \text{ \AA}$  was recorded for each conformer as its time of residence ( $t_R$ ). Similarly, we define the escape probability ( $P_e(t)$ ) as the probability of leaving the initial conformation before time  $t$ .

Fig 3 shows the results of this study. The top panel shows that  $Q_{60}$  conformers last longer than  $Q_{20}$  in a specific state, while the average escape probability of  $V_{60}$  initially is lower but soon rises much faster than the other two sets. For completeness, we run the same study on three regular proteins: Trp-cage (PDB code 1L2Y, 20 residues long), an immunoglobulin binding domain of protein G (1GB1, 56 residues) and the 27th immunoglobulin domain of human cardiac titin (1TIT, 89 residues). All of them remained in the same conformation for longer than 10 ns: their RMSD was never higher than  $2 \text{ \AA}$ .

The bottom panels of Fig 3 show scatter plots of  $t_R$  vs.  $\langle z \rangle$ . This figure shows that  $\langle z \rangle$  is not only unrelated to  $F_{\max}$ , but also to the temporal stability of the conformers (as measured by  $t_R$ ) in the cases where  $\langle z \rangle$  is above the simply stiff limit. In the case of conformers with  $\langle z \rangle$  below this limit, however,  $t_R$  is always below 1 ns, reinforcing Maxwell's theory on frame stiffness [35].

Interestingly, both theoretical and especially experimental pulling experiments are typically done at  $v_p$ 's such that the time the protein is being pulled is far longer than 10 ns. In particular, the pulling simulations performed in this work take  $\approx 50 \mu\text{s}$  to completely extend a protein with 60 residues, while experiments such as the ones performed in [16] take around 60 ms to





**Fig 3. Time evolution of the studied structures.** For each set in  $Q_{60}$ ,  $Q_{20}$  and  $V_{60}$ , 100 randomly chosen structures have been placed under a free-dynamics evolution for 10 ns. After that, the RMSD has been studied and the last time when it fluctuates above 2 Å is recorded as the residence time ( $t_R$ ). The top graph shows the escape probability ( $P_e(t)$ ), defined as the probability of having left the initial state of a conformer at time  $t$ . We can see how  $Q_{20}$  fluctuates out of the initial structure much faster than  $Q_{60}$ , while  $V_{60}$  starts more slowly but rapidly outruns both  $Q_{60}$  and  $Q_{20}$ . The inset shows the average evolution of the RMSD for the three sets compared to an example of a similar-sized globular protein, an immunoglobulin binding domain of protein G (PDB code 1GB1, 56 residues). The latter lasts for longer than 10 ns fluctuating around 2 Å, while the other three rapidly evolve out of the initial structure. The bottom graphs show scatter plots of  $\langle z \rangle$  vs.  $t_R$ . No simple relation can be established between these two quantities above the stiff limit (dashed vertical lines), while below it residence times never exceed 1 ns.

doi:10.1371/journal.pcbi.1004541.g003

accomplish the same task. This leads to question whether the force peaks present in the experimental traces really relate to the initial conformers or have actually been formed while the molecule was being pulled.

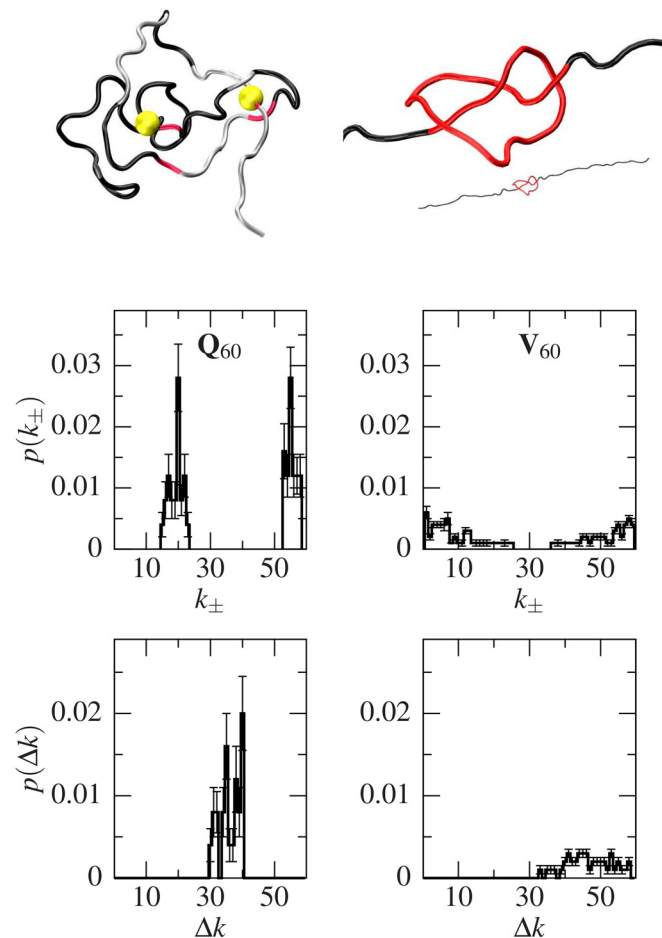
Therefore, one must look at  $F_{max}$  carefully since it has different meaning in this kind of simulation than in experiments such as those in [16]: Here, mechanical stability is associated directly with a conformer, since simulations are based on the initial contact map. On the other hand, in experiments, molecules are subjected to fluctuations with a characteristic time of 1 ns and the  $F-d$  curves carry information not only about the initial conformer but also about the stretching-unrelated intrinsic shape transformations that the protein may undergo. All in all, we observe that disordered proteins such as polyglutamines are not long lasting when compared to structured globular ones, and that mechanical stabilities need to be looked at in the context of how they were measured, either referred to the initial conformer if done through structure-based modelling, or including bond formation during the stretching if performed experimentally.

### Structures with knots

Even though the starting structures were not knotted, our BEMD simulation yielded some knotted conformers. In particular,  $(9.3 \pm 1.8)\%$  of the independent  $Q_{60}$  conformers have a knot, while  $Q_{20}$  include no knotted conformers. Moreover, only  $(3.6 \pm 0.5)\%$  of the  $V_{60}$  structures contain a knot, and none of the CATH structures have one. All knots generated in  $V_{60}$

are trefoil ( $3_1$ ), while  $Q_{60}$  also contains one three-twist ( $5_2$ ). Upon stretching, only  $(13 \pm 7)\%$  of the  $Q_{60}$  knotted structures untie, while  $(45 \pm 6)\%$  of the  $V_{60}$  ones do. As shown in ref. [44], tightening of knots may be associated with force peaks. Both for  $Q_{60}$  and  $V_{60}$ , knot tightening yields  $F_{\max}$  from 0.9 to  $2.4 \text{ e}/\text{\AA}$ .

Fig 4 shows an example of a  $3_1$  knotted conformer found in  $Q_{60}$  plus a histogram of the ends of the knots ( $k_-, k_+$ ) and their extension ( $\Delta k$ , measured as the number of residues contained inside the knot) in the structures formed by sets  $Q_{60}$  and  $V_{60}$ . An example of a  $5_2$  one from  $Q_{60}$  and a  $3_1$  from  $V_{60}$  are shown in S7 Fig. Significantly, not only does  $V_{60}$  form fewer and less stable knots than  $Q_{60}$ , but also the extension of the  $V_{60}$  knots is typically larger than the  $Q_{60}$  ones. Furthermore, the average extension of the knotted  $Q_{60}$  conformers is 36 (with a 0.12% error), which corresponds to the median threshold value for most polyglutamine-expansion-related diseases such as HD. We note that knotted structures would have been found experimentally as putative events in [16], since the final length would be reduced and thus they



**Fig 4. Knots in the studied conformers.** The top left panel shows an example of a  $Q_{60}$  conformation containing a trefoil ( $3_1$ ) knot with the knot ends highlighted with yellow spheres. To its right, the same conformation has been partially stretched, and the region inside the knot is highlighted in red and zoomed in. The middle panels represent histograms of the knot end positions,  $k_{\pm}$ , for  $Q_{60}$  (left) and  $V_{60}$  (right). The bottom panel shows their corresponding extension,  $\Delta k$ . The percentage of knotted structures relative to the total number of independent conformers found for  $Q_{60}$  and  $V_{60}$  are  $(9.3 \pm 1.8)\%$  and  $(3.6 \pm 0.5)\%$ , respectively. Shallow knots have an extension closer to 60 (the system size). Protein representations have been done with VMD [48].

doi:10.1371/journal.pcbi.1004541.g004

would render molecules with lower total contour length increase. Furthermore, knotted proteins have previously been found in nature especially in enzymes such as methyltransferases and carbonic anhydrases [45–47]. Nonetheless, the only hypothesized function of the knot itself—as opposed to the whole protein—is to prevent the unfolding of the protein in a case where the proteasome were to try to degrade it [46].

The presence of knots on its own is not indicative of their relevance: they need to last long enough to be able to have any effect. To that end, we performed 200 ns all-atom simulations with explicit TIP3P water of three randomly chosen knotted conformers to see the behaviour of these knots with time. The knot ends fluctuate along the protein as does the knot size, and in some cases the knot unties just to be formed again some time later—the time of the protein being in the untied conformation lasting for as long as 200 ps. Also, in two of the three cases, preferred places for the left and right ends can be seen, the right end being the same for both of them. The results of these simulations can be seen in [S8 Fig](#).

### Other lengths in polyQ chains

Given that the average extension of the knots corresponds with the median of the threshold of the polyQ diseases, we applied the same methodology to other  $Q_n$  tracts, with  $n = 16, 25, 33, 38, 40$  and  $80$ . As expected, no knots were found for  $n < 35$ ; but there were no knots in sets  $Q_{38}, Q_{40}$  or  $Q_{80}$  either. This may be attributed to a low probability of knot formation combined with small statistics, which would imply that BEMD took  $Q_{60}$  through a knot-forming path while taking the rest of  $Q_n$  studied through non-forming ones. This is reinforced by the fact that the greater statistics of  $V_{60}$  do find knotted conformers. Therefore, an increase in the sampling may catch these knotted structures in  $Q_{80}$  and  $Q_{40}$ , while their formation is fairly improbable for  $n$  below 35 since the typical knot size is about this length.

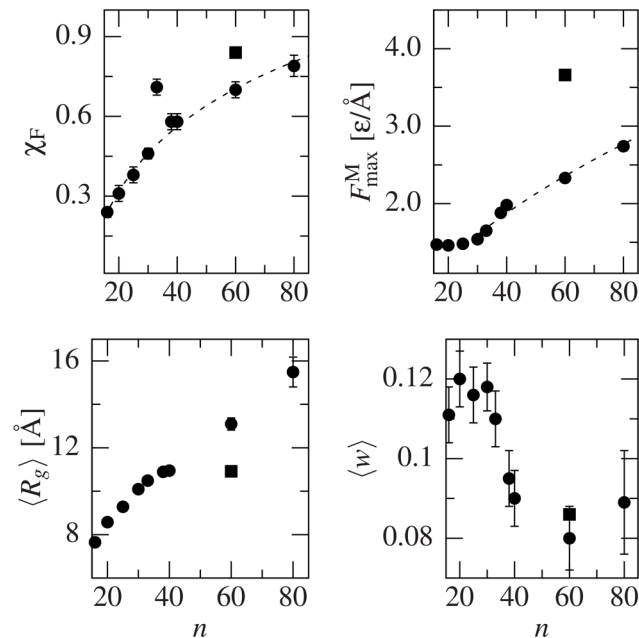
[Fig 5](#) shows the evolution of the mechanical stability and shape with the chain length. In particular, the fraction of conformers with  $F_{\max} > 0$ , which we name  $\chi_F$ , follow a logarithmic law, while the maximum  $F_{\max}$  for each set, denoted as  $F_{\max}^M$  behaves like an avalanche system: it has a constant value until  $n = 33$ , and then starts growing as a power law with exponent 0.562. The average  $R_g$  appears to be saturating as  $n$  approaches 40, but it suddenly jumps for  $n = 60$  and  $80$ . Judging by  $w$ , the shapes of the conformers change around  $n = 35$  from elongated to more globular. Interestingly,  $V_{60}$  behaves differently than  $Q_{60}$  except that the average  $w$  (lower right panel) is similar, suggesting the similarity of shapes. We also conclude that the fraction of mechanically stable conformers increases uniformly with  $n$ , while the maximum  $F_{\max}$  presents an avalanche behaviour for  $n > 30$ , once again close to HD's threshold of 35.

### Discussion

In this study, we have generated an ensemble of structurally independent conformers for glutamine expansions with  $n$  residues. We have focused on  $n = 60$  which, if present in huntingtin protein, would result in Huntington disease, and on  $n = 20$ , which would not.

We have then expanded the study to  $n = 16, 25, 30, 33, 38, 40$  and  $80$  in order to further explore the structural nature of the  $n \approx 35$  threshold in most polyQ-related diseases.

We find that proteins related to the disease exhibit less conformational polymorphism than the ones unrelated to it in terms of independent structures and transition kinetics, even though the former show much more mechanical variability (in terms of  $F_{\max}$  and  $n_p$ ) as well as structural (measured by  $\text{SS}$  and  $\langle z \rangle$ ). We also conclude that, contrary to intuition,  $\langle z \rangle$ ,  $CO$ ,  $\text{SS}$  and  $\beta$ -content are not good predictors of either temporal or mechanical stability. This conclusion extends not only for polyQ but also generally for all proteins in CATH.



**Fig 5. Variability of the specified parameters with the length,  $n$ , of the polyQ chain (circles).** The values for  $V_{60}$  are indicated by a square.  $\chi_F$  represents the fraction of conformers with at least one force peak for that particular length. The dotted fits correspond to a logarithmic function (top left,  $352 \ln(x/8.115)$ ) and a polynomial behavior (top right,  $y = 0.236x^{0.562}$ ), which is typical for avalanches. The bottom panels show average over the structures of  $R_g$  and  $w$ .  $\langle R_g \rangle$  has a saturating behavior up to  $n = 40$ , but jumps for higher values.  $\langle w \rangle$  presents a transition around  $n = 35$  from slightly elongated to more globular proteins.

doi:10.1371/journal.pcbi.1004541.g005

Finally, we prove the presence of knots of length 35 at least in  $Q_{60}$ . The sequential size of these knots suggests a relationship to HD. One of the possible mechanisms for the relevance of the knots in pathology is impairing the process of proteasomal degradation, as suggested in [46] and [16]. Moreover, although there is evidence for the toxicity of the monomeric polyQ species [49], even if the toxicity was due mainly to the oligomers (see *e.g.* Ref. [50]), the blockade of the degradation machinery by a knotted monomer would induce an increase of the concentration of aggregating protein, and thus toxicity may be caused by the monomers even if they are not toxic themselves.

## Supporting Information

**S1 Text. The details of structure generation and selection are explained here, together with the stability associated to the coordination number.** Furthermore, a statistical analysis of the lack of relation between  $F_{\max}$  and the rest of the descriptors used in this work is also presented. (PDF)

**S1 Fig. Five conformers with highest mechanical stability in set  $Q_{60}$ .** The structure with the biggest  $F_{\max}$  is at the top. The left column shows snapshots of the structures. The red ribbons represent  $\beta$  strands and the red lines correspond to  $\beta$  bridges. The black lines indicate hydrogen-bonded turns. The orange spheres mark the termini, from which the molecule is pulled. The center column displays the unfolding  $F - d$  curve (left axis) together with the unfolding scenario diagram (right axis), *i.e.* the time a contact is broken vs. the distance between the residues that are in contact. The column on the right shows the values of the relevant descriptors.

All molecule cartoons were generated using VMD [48].  
(TIF)

**S2 Fig. Five conformers with highest mechanical stability in set Q<sub>20</sub>.** The structure with the biggest  $F_{\max}$  is at the top. The left column shows snapshots of the structures. The red ribbons represent  $\beta$  strands and the red lines correspond to  $\beta$  bridges, while blue helices are  $\alpha$  helices. The black lines indicate hydrogen-bonded turns. The center column displays the unfolding  $F-d$  curve (left axis) together with the unfolding scenario diagram (right axis). The column on the right shows the values of the relevant descriptors.  
(TIF)

**S3 Fig. Five conformers with highest mechanical stability in set V<sub>60</sub>.** The structure with the biggest  $F_{\max}$  is at the top. The left column shows snapshots of the structures. The red ribbons represent  $\beta$  strands and the red lines correspond to  $\beta$  bridges. The black lines indicate hydrogen-bonded turns and  $\alpha$ -helices are depicted in blue. The center column displays the unfolding  $F-d$  curve (left axis) together with the unfolding scenario diagram (right axis). The column on the right shows the values of the relevant descriptors.  
(TIF)

**S4 Fig. Scatter plot of  $F_{\max}$  vs.  $\alpha$ ,  $\beta$  and hydrogen-bonded turns ( $\tau$ ) content for polyQ chains.** The horizontal dashed lines mark off the top five values of  $F_{\max}$ .  
(TIF)

**S5 Fig. Scatter plot of  $F_{\max}$  vs. CO for the specified sets.** The horizontal dashed lines mark off the top five values of  $F_{\max}$ .  
(TIF)

**S6 Fig. Kinetics of independent structure formation for Q<sub>60</sub> (circles), V<sub>60</sub> (triangles) and Q<sub>20</sub> (diamonds).** Although more complete plots should be fit with a double exponential function [4], short trajectories correspond to a linear behavior. The fitted slopes are .28, .62 and .98 respectively. Data for V<sub>60</sub> were taken from [4].  
(TIF)

**S7 Fig. Examples of knotted structures.** The top structure corresponds to a three-twist ( $5_2$ ) knot in Q<sub>60</sub>, while the lower panels are for a trefoil knot from V<sub>60</sub>, where no other knots were found. Left column shows a representation of the molecule before stretching, with the knot ends highlighted with yellow spheres. Right panels show the molecules partially stretched, and the region inside the knot is highlighted in red and zoomed in.  
(TIF)

**S8 Fig. Time evolution of the knots.** Three randomly chosen knotted conformers were simulated with all-atom and explicit solvent. One of them is shown in Fig 4. The top panel shows the evolution of the knot size with time for one of the simulations. The middle panel shows a histogram of the knot sizes along this time for the three simulations, each with a different color. The bottom panel shows a histogram of the respective knot ends, the left end ( $k_-$ , inverted) and the right ones ( $k_+$ ).  
(TIF)

**S9 Fig. An example of the SS sieve and time clustering stages.** The gray line in the top panel shows evolution of SS with time for one of the replicas. Structures with SS > 30% (the thin horizontal line) are taken for clustering. A cluster ends whenever the gap between successive structured conformers becomes greater than 50 ps. The black dots correspond to structures that represent clusters: these are the structures with the highest SS in the cluster. The red

box in the top panel is shown zoomed in the middle panel, where clusters are represented by red lines. The bottom panel shows the RMSD of each cluster representative relative to the previous one. All of these RMSD's are greater than 2 Å so the clusters can be considered to be uncorrelated in time.

(TIF)

**S10 Fig. Color-map plots of the difference between the joint CDF and the product of the independent CDFs of  $F_{\max}$  and the specified descriptor.** Differences are always below 0.1, and below 0.05 in three of the descriptors (SS, CO and  $\tau$ ). Therefore,  $F_{\max}$  is statistically independent of the descriptors studied.

(TIF)

**S1 Table. Parameters of a linear regression for the dependence of  $F_{\max}$  on various structural descriptors.** The top panel lists the values of the Pearson  $R^2$  coefficients with a 95% confidence interval. The lower panel lists the slopes of the linear fits together with the error bars. The number in the parenthesis is the corresponding  $p$ -value. Even though the slope for each correlation is significantly different from zero;  $R^2$  is never close to one, so no correlation can be established between the descriptors and  $F_{\max}$ . This is also assessed in [S10 Fig](#).

(PDF)

## Acknowledgments

We thank many critical discussions with P. Cossio, R. Hervás, A. Rodríguez and J.M. Ortiz. We also acknowledge the contribution of TeideHPC and SVG High-Performance Computing facilities. TeideHPC facilities are provided by the Instituto Tecnológico y de Energías Renovables (ITER), S.A., [www.iter.es](http://www.iter.es). SVG HPC facilities are provided by the Galician Supercomputing Center (CESGA), [www.cesga.es](http://www.cesga.es).

## Author Contributions

Conceived and designed the experiments: ÀGS MC MCV. Performed the experiments: ÀGS MS. Analyzed the data: ÀGS MC. Contributed reagents/materials/analysis tools: MS MC. Wrote the paper: ÀGS MC MCV MS.

## References

1. Chothia C, Finkelstein AV. The classification and origins of protein folding patterns. *Annual Review of Biochemistry*. 1990; 59(1):1007–1035. PMID: [2197975](#)
2. Chothia C. One thousand families for the molecular biologist. *Nature*. 1992 Jun; 357(6379):543–544. doi: [10.1038/357543a0](#) PMID: [1608464](#)
3. Sillitoe I, Cuff AL, Dessailly BH, Dawson NL, Furnham N, Lee D, et al. New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. *Nucleic Acids Research*. 2013; 41(D1):D490–D498. Available from: <http://nar.oxfordjournals.org/content/41/D1/D490.abstract>. doi: [10.1093/nar/gks1211](#) PMID: [23203873](#)
4. Cossio P, Trovato A, Pietrucci F, Seno F, Maritan A, Laio A. Exploring the universe of protein structures beyond the Protein Data Bank. *PLoS Comput Biol*. 2010; 6(11):e1000957. doi: [10.1371/journal.pcbi.1000957](#) PMID: [21079678](#)
5. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*. 2004; 57(4):702–710. doi: [10.1002/prot.20264](#)
6. Magrane M, Consortium U. UniProt Knowledgebase: a hub of integrated protein data. *Database*. 2011; 2011. doi: [10.1093/database/bar009](#) PMID: [21447597](#)
7. Nasir J, Floresco SB, O'Kusky JR, Diewert VM, Richman JM, Zeisler J, et al. Targeted disruption of the Huntington's disease gene results in embryonic lethality and behavioral and morphological changes in heterozygotes. *Cell*. 1995; 81(5):811–823. doi: [10.1016/0092-8674\(95\)90542-1](#) PMID: [7774020](#)

8. Zuccato C, Ciammola A, Rigamonti D, Leavitt BR, Goffredo D, Conti L, et al. Loss of Huntingtin-Mediated BDNF Gene Transcription in Huntington's Disease. *Science*. 2001; 293(5529):493–498. Available from: <http://www.sciencemag.org/content/293/5529/493.abstract>. doi: [10.1126/science.1059581](https://doi.org/10.1126/science.1059581) PMID: [11408619](https://pubmed.ncbi.nlm.nih.gov/11408619/)
9. Velier J, Kim M, Schwarz C, Kim TW, Sapp E, Chase K, et al. Wild-Type and Mutant Huntingtins Function in Vesicle Trafficking in the Secretory and Endocytic Pathways. *Experimental Neurology*. 1998; 152(1):34–40. Available from: <http://www.sciencedirect.com/science/article/pii/S0014488698968327>. doi: [10.1006/exnr.1998.6832](https://doi.org/10.1006/exnr.1998.6832) PMID: [9682010](https://pubmed.ncbi.nlm.nih.gov/9682010/)
10. Petruska J, Hartenstine MJ, Goodman MF. Analysis of Strand Slippage in DNA Polymerase Expansions of CAG/CTG Triplet Repeats Associated with Neurodegenerative Disease. *Journal of Biological Chemistry*. 1998; 273(9):5204–5210. doi: [10.1074/jbc.273.9.5204](https://doi.org/10.1074/jbc.273.9.5204) PMID: [9478975](https://pubmed.ncbi.nlm.nih.gov/9478975/)
11. Ross CA. Polyglutamine Pathogenesis: Emergence of Unifying Mechanisms for Huntington's Disease and Related Disorders. *Neuron*. 2002; 35(5):819–822. doi: [10.1016/S0896-6273\(02\)00872-3](https://doi.org/10.1016/S0896-6273(02)00872-3) PMID: [12372277](https://pubmed.ncbi.nlm.nih.gov/12372277/)
12. Pla P, Orvoen S, Saudou F, DAVID DJ, Humbert S. Mood disorders in Huntington's disease: from behavior to cellular and molecular mechanisms. *Frontiers in Behavioral Neuroscience*. 2014; 8(135). doi: [10.3389/fnbeh.2014.00135](https://doi.org/10.3389/fnbeh.2014.00135) PMID: [24795586](https://pubmed.ncbi.nlm.nih.gov/24795586/)
13. Fan HC, Ho LI, Chi CS, Chen SJ, Peng GS, Chan TM, et al. Polyglutamine (PolyQ) Diseases: Genetics to Treatments. *Cell Transplantation*. 2014-04-09T00:00:00; 23(4–5):441–458. Available from: <http://www.ingentaconnect.com/content/cog/ct/2014/00000023/F0020004/art00006>. PMID: [24816443](https://pubmed.ncbi.nlm.nih.gov/24816443/)
14. Albrecht A, Mundlos S. The other trinucleotide repeat: polyalanine expansion disorders. *Current Opinion in Genetics & Development*. 2005; 15(3):285–293. Genetics of disease. Available from: <http://www.sciencedirect.com/science/article/pii/S0959437X05000559>. doi: [10.1016/j.gde.2005.04.003](https://doi.org/10.1016/j.gde.2005.04.003)
15. Amiel J, Trochet D, Clément-Ziza M, Munnich A, Lyonnet S. Polyalanine expansions in human. *Human Molecular Genetics*. 2004; 13(suppl 2):R235–R243. Available from: [http://hmg.oxfordjournals.org/content/13/suppl\\_2/R235.abstract](http://hmg.oxfordjournals.org/content/13/suppl_2/R235.abstract). doi: [10.1093/hmg/ddh251](https://doi.org/10.1093/hmg/ddh251) PMID: [15358730](https://pubmed.ncbi.nlm.nih.gov/15358730/)
16. Hervás R, Oroz J, Galera-Prat A, Goñi O, Valbuena A, Vera AM, et al. Common features at the start of the neurodegeneration cascade. *PLoS Biol*. 2012; 10(5):e1001335. doi: [10.1371/journal.pbio.1001335](https://doi.org/10.1371/journal.pbio.1001335) PMID: [22666178](https://pubmed.ncbi.nlm.nih.gov/22666178/)
17. Piana S, Laio A. A bias-exchange approach to protein folding. *The Journal of Physical Chemistry B*. 2007 May; 111(17):4553–4559. doi: [10.1021/jp067873i](https://doi.org/10.1021/jp067873i) PMID: [17419610](https://pubmed.ncbi.nlm.nih.gov/17419610/)
18. Sułkowska JI, Cieplak M. Mechanical stretching of proteins—a theoretical survey of the Protein Data Bank. *Journal of Physics: Condensed Matter*. 2007; 19(28):283201.
19. Sikora M, Sułkowska JI, Cieplak M. Mechanical strength of 17,134 model proteins and cysteine slip-knots. *PLoS Comput Biol*. 2009 Oct; 5(10):e1000547. doi: [10.1371/journal.pcbi.1000547](https://doi.org/10.1371/journal.pcbi.1000547) PMID: [19876372](https://pubmed.ncbi.nlm.nih.gov/19876372/)
20. Ferreon ACM, Moran CR, Gambin Y, Deniz AA. Single-molecule fluorescence studies of intrinsically disordered proteins. *Methods in Enzymology*. 2010; 472:179–204. doi: [10.1016/S0076-6879\(10\)72010-3](https://doi.org/10.1016/S0076-6879(10)72010-3) PMID: [20580965](https://pubmed.ncbi.nlm.nih.gov/20580965/)
21. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*. 2008; 4(3):435–447. doi: [10.1021/ct700301q](https://doi.org/10.1021/ct700301q)
22. Bonomi M, Branduardi D, Bussi G, Camilloni C, Provasi D, Raiteri P, et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications*. 2009; 180(10):1961–1972. Available from: <http://www.sciencedirect.com/science/article/pii/S001046550900157X>. doi: [10.1016/j.cpc.2009.05.011](https://doi.org/10.1016/j.cpc.2009.05.011)
23. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*. 1995; 117(19):5179–5197. doi: [10.1021/ja00124a002](https://doi.org/10.1021/ja00124a002)
24. Qiu D, Shenkin PS, Hollinger FP, Still WC. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *The Journal of Physical Chemistry A*. 1997; 101(16):3005–3014. doi: [10.1021/jp961992r](https://doi.org/10.1021/jp961992r)
25. Pietrucci F, Laio A. A Collective Variable for the Efficient Exploration of Protein Beta-Sheet Structures: Application to SH3 and GB1. *Journal of Chemical Theory and Computation*. 2009; 5(9):2197–2201. doi: [10.1021/ct900202f](https://doi.org/10.1021/ct900202f)
26. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen My, et al. In: *Comparative Protein Structure Modeling Using Modeller*. John Wiley & Sons, Inc.; 2002. Available from: <http://dx.doi.org/10.1002/0471250953.bi0506s15>.

27. Hestenes MR, Stiefel E. Methods of conjugate gradients for solving linear systems. vol. 49. National Bureau of Standards Washington, DC; 1952.
28. Miettinen MS, Knecht V, Monticelli L, Ignatova Z. Assessing polyglutamine conformation in the nucleating event by molecular dynamics simulations. *The Journal of Physical Chemistry B*. 2012; 116(34):10259–10265. doi: [10.1021/jp305065c](https://doi.org/10.1021/jp305065c) PMID: [22770401](https://pubmed.ncbi.nlm.nih.gov/22770401/)
29. Khare SD, Ding F, Gwanmesia KN, Dokholyan NV. Molecular origin of polyglutamine aggregation in neurodegenerative diseases. *PLoS Comput Biol*. 2005; 1(3):230–235. doi: [10.1371/journal.pcbi.0010030](https://doi.org/10.1371/journal.pcbi.0010030) PMID: [16158094](https://pubmed.ncbi.nlm.nih.gov/16158094/)
30. Laghaei R, Mousseau N. Spontaneous formation of polyglutamine nanotubes with molecular dynamics simulations. *The Journal of Chemical Physics*. 2010; 132(16):165102. doi: [10.1063/1.3383244](https://doi.org/10.1063/1.3383244) PMID: [20441310](https://pubmed.ncbi.nlm.nih.gov/20441310/)
31. Cieplak M, Allan DB, Leheny RL, Reich DH. Proteins at Air–Water Interfaces: A Coarse-Grained Model. *Langmuir*. 2014; 30:1288–96. Available from: <http://dx.doi.org/10.1021/la502465m>.
32. Sikora M, Szymczak P, Thompson D, Cieplak M. Linker-mediated assembly of gold nanoparticles into multimeric motifs. *Nanotechnology*. 2011; 22(44):445601. doi: [10.1088/0957-4484/22/44/445601](https://doi.org/10.1088/0957-4484/22/44/445601) PMID: [21979426](https://pubmed.ncbi.nlm.nih.gov/21979426/)
33. Joosten RP, te Beek TAH, Krieger E, Hekkelman ML, Hoof RWW, Schneider R, et al. A series of PDB related databases for everyday needs. *Nucleic Acids Res*. 2011 Jan; 39(Database issue):D411–D419. doi: [10.1093/nar/gkq1105](https://doi.org/10.1093/nar/gkq1105) PMID: [21071423](https://pubmed.ncbi.nlm.nih.gov/21071423/)
34. Tsai J, Taylor R, Chothia C, Gerstein M. The packing density in proteins: standard radii and volumes. *Journal of Molecular Biology*. 1999; 290(1):253–266. Available from: <http://www.sciencedirect.com/science/article/pii/S002228369928292>. doi: [10.1006/jmbi.1999.2829](https://doi.org/10.1006/jmbi.1999.2829) PMID: [10388571](https://pubmed.ncbi.nlm.nih.gov/10388571/)
35. Maxwell JC. L. on the calculation of the equilibrium and stiffness of frames. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Sciences*. 1864; 27(182):294–299.
36. Cieplak M, Robbins MO. Nanoindentation of 35 virus capsids in a molecular model: relating mechanical properties to structure. *PloS One*. 2013; 8(6):e63640. doi: [10.1371/journal.pone.0063640](https://doi.org/10.1371/journal.pone.0063640) PMID: [23785395](https://pubmed.ncbi.nlm.nih.gov/23785395/)
37. Sułkowska JI, Cieplak M. Selection of optimal variants of G-like models of proteins through studies of stretching. *Biophysical Journal*. 2008; 95(7):3174–3191. doi: [10.1529/biophysj.107.127233](https://doi.org/10.1529/biophysj.107.127233) PMID: [18567634](https://pubmed.ncbi.nlm.nih.gov/18567634/)
38. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *Journal of Molecular Biology*. 1998; 277(4):985–994. doi: [10.1006/jmbi.1998.1645](https://doi.org/10.1006/jmbi.1998.1645) PMID: [9545386](https://pubmed.ncbi.nlm.nih.gov/9545386/)
39. Kesner BA, Ding F, Temple BR, Dokholyan NV. N-terminal strands of filamin Ig domains act as a conformational switch under biological forces. *Proteins: Structure, Function, and Bioinformatics*. 2010; 78(1):12–24. doi: [10.1002/prot.22479](https://doi.org/10.1002/prot.22479)
40. Cieplak M, Hoang TX. Universality classes in folding times of proteins. *Biophysical Journal*. 2003; 84(1):475–488. doi: [10.1016/S0006-3495\(03\)74867-X](https://doi.org/10.1016/S0006-3495(03)74867-X) PMID: [12524300](https://pubmed.ncbi.nlm.nih.gov/12524300/)
41. Berkovich R, Garcia-Manyes S, Klafter J, Urbakh M, Fernández JM. Hopping around an entropic barrier created by force. *Biochemical and Biophysical Research Communications*. 2010; 403(1):133–137. doi: [10.1016/j.bbrc.2010.10.133](https://doi.org/10.1016/j.bbrc.2010.10.133) PMID: [21050839](https://pubmed.ncbi.nlm.nih.gov/21050839/)
42. Rief M, Gautel M, Oesterhelt F, Fernandez JM, Gaub HE. Reversible Unfolding of Individual Titin Immunoglobulin Domains by AFM. *Science*. 1997; 276(5315):1109–1112. doi: [10.1126/science.276.5315.1109](https://doi.org/10.1126/science.276.5315.1109) PMID: [9148804](https://pubmed.ncbi.nlm.nih.gov/9148804/)
43. Carrion-Vazquez M, Oberhauser AF, Fowler SB, Marszalek PE, Broedel SE, Clarke J, et al. Mechanical and chemical unfolding of a single protein: A comparison. *Proceedings of the National Academy of Sciences USA*. 1999; 96(7):3694–3699. doi: [10.1073/pnas.96.7.3694](https://doi.org/10.1073/pnas.96.7.3694)
44. Sułkowska JI, Sułkowski P, Szymczak P, Cieplak M. Tightening of knots in proteins. *Physical Review Letters*. 2008; 100(5):058106. doi: [10.1103/PhysRevLett.100.058106](https://doi.org/10.1103/PhysRevLett.100.058106) PMID: [18352439](https://pubmed.ncbi.nlm.nih.gov/18352439/)
45. Taylor WR. A deeply knotted protein structure and how it might fold. *Nature*. 2000; 406(6798):916–919. doi: [10.1038/35022623](https://doi.org/10.1038/35022623) PMID: [10972297](https://pubmed.ncbi.nlm.nih.gov/10972297/)
46. Virnau P, Mirny LA, Kardar M. Intricate knots in proteins: Function and evolution. *PLoS Computational Biology*. 2006; 2(9):e122. doi: [10.1371/journal.pcbi.0020122](https://doi.org/10.1371/journal.pcbi.0020122) PMID: [16978047](https://pubmed.ncbi.nlm.nih.gov/16978047/)
47. Sułkowska JI, Rawdon EJ, Millett KC, Onuchic JN, Stasiak A. Conservation of complex knotting and slipknotting patterns in proteins. *Proceedings of the National Academy of Sciences USA*. 2012; 109(26):E1715–E1723. doi: [10.1073/pnas.1205918109](https://doi.org/10.1073/pnas.1205918109)
48. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *Journal of Molecular Graphics*. 1996; 14(1):33–38. doi: [10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) PMID: [8744570](https://pubmed.ncbi.nlm.nih.gov/8744570/)



49. Nagai Y, Inui T, Popiel HA, Fujikake N, Hasegawa K, Urade Y, et al. A toxic monomeric conformer of the polyglutamine protein. *Nature Structural & Molecular Biology*. 2007; 14(4):332–340. doi: [10.1038/nsmb1215](https://doi.org/10.1038/nsmb1215)
50. Ripaud L, Chumakova V, Antonin M, Hastie AR, Pinkert S, Körner R, et al. Overexpression of Q-rich prion-like proteins suppresses polyQ cytotoxicity and alters the polyQ interactome. *Proceedings of the National Academy of Sciences*. 2014; 111(51):18219–18224. doi: [10.1073/pnas.1421313111](https://doi.org/10.1073/pnas.1421313111)