Supporting Information

Dougan et al. 10.1073/pnas.0900678106

SI Text

Polyprotein Engineering and Purification. To allow for efficient single molecule experiments, we first constructed polyproteins using protein engineering. All constructs made are shown in Table 1 including; the number of amino acids in the linkers and handles N_{Link}, the number of amino acids in the polypeptide segment N_{PP}, the average initial extension $L_{Initial}$, and the expected extension of the folded I27 proteins, the linkers, handles, and the polypeptide segment calculate using the worm-like chain (WLC) model of polymer elasticity (1), L_{WLC}. The details of the I27 (2) and PEVK (3) polyprotein engineering and purification have been reported previously. Briefly, we constructed a 3-domain N-C linked polyprotein of I27, the 27th Ig-like domain of cardiac titin, through successive cloning in modified pT7Blue vectors and then expressed the gene using vector pQE30 in *Escherichia coli* strain BLR(DE3). The (I27-PEVK) construct is based on the human titin I27 and PEVK (exon 161) sequences. The core sequence for the chimera consists of the I27 module followed by the PEVK exon. This core sequence is repeated once with the addition of 1 I27 module, yielding (I27-PEVK-I27-PEVK-I27). For the I27 and polypeptide segments chimera constructs we used successive cloning in pT7Blue vectors and then expressed the gene using vector pQE80L in *Escherichia coli* strain BLR(DE3). Coomassie blue staining of the purified construct separated using SDS/PAGE and MALDI-TOF mass spectroscopy measurements were used to confirm the correct molecular weight for each engineered polyprotein construct.

Force-clamp Spectroscopy of Single Polyproteins. ***Single molecule-force spectroscopy.*** We used a custom-built atomic force microscope equipped with a PicoCube P363.3-CD piezoelectric translator (Physik Instrumente) controlled by an analog PID feedback system. Silicon-nitride cantilevers (Veeco) were calibrated for their spring constant using the equipartition theorem as reported. The average spring constant was approximately 15 pN/nm. All data were obtained and analyzed using custom software written for use in Igor 5.0 (Wavemetrics). There was approximately 0.5 nm of peak-to-peak noise and a feedback response time of approximately 5 ms in all experiments.

A significant advantage of force-clamp experiments is that using polyprotein engineering we can ascertain that a single construct is probed. There are 2 stringent tests that provide this certainty. The first test is the identification of a mechanical fingerprint, and the second is the unfolding rate of the construct of interest. Both are described in detail below.

Identifying the mechanical fingerprint. The number of unfolding modules in each polyprotein construct is well-defined, for example 3 modules in the polyprotein $(I27)_3$. Obtaining the full fingerprint will depend on where the protein is picked up, that is, where the cantilever tip and substrate are in contact with the protein. If the attachment is not at the 2 ends of the monomer unit then only 1 or 2 I27-unfolding steps will be measured. Conversely, if more than 1 construct or an aggregate is picked up, more than 3 I27-unfolding steps will be measured. Analysis of all of the trajectories obtained in the experiments determines the likelihood of picking up the full construct as opposed to a partial construct. In *Supporting Information* (*SI*) Figure S1 the number of experimental traces containing 1–8 steps are shown for all constructs. In the case of the $(I27)_3$ construct (Fig S1*A*) traces with 1–3 steps are frequently observed, with 3 steps being the most common. However, traces with more than 3 I27-unfolding steps are rarely observed (≈5% of traces). In the case of the polyglutamine (polyQ) construct (I27-Q25) (Fig. S1*C*) traces with 1–3 steps are observed, with 3 steps being the most common. Again, traces with more than 3 I27-unfolding steps are rarely observed (≈4%). As a further control, we examined the (I27-Q15) construct. For this construct, 5 repeats of the I27 protein are necessary since this allows for 4 repeats of the shorter polyQ chain of only 15 repeats (Table 1). Traces with 1–5 steps are observed (Fig. S1*B*), with 5 steps being most common. Traces with more than 5 I27-unfolding steps are rarely observed (≈4%). Therefore, analysis of all experimental data shows no evidence of an increased number of I27-unfolding steps, implying that these experiments are indeed probing the behavior of single constructs.

Calculating the unfolding rate. The unfolding rate (α) of a polyprotein is sensitive to the way in which the force is distributed throughout the protein. If 2 polyproteins, or more, are connected in parallel or had formed part of a complicated aggregate array then the force (F) would be distributed throughout the I27 protein in a very different way and the measured unfolding rate would be different. Indeed, this has previously been observed for a construct were we identified the mechanical fingerprint for protein dimerization (4). Only a single polyprotein will give the correct number of unfolding modules and the correct unfolding rate constant at a particular force (α(F)). The α(F) can be obtained by collecting many traces ($n > 30$) which contain 3 I27-unfolding steps and summing and normalizing them. This normalized trace P(t) is then fit with a single exponential were the time constant is the α(F) (Fig. S2). To estimate the error on our experimentally obtained rate constant, we carried out the nonparametric bootstrap method (5). The unfolding rate of the I27 proteins in our HPP constructs is determined by collecting many traces which contain 3 I27-unfolding steps. The unfolding rate for $(I27)_3$ is 14.62 ± 0.05 s^{-1}, for (I27-Q25) the unfolding rate is 10.53 ± 0.05 $s^{-1}$and for (I27-Q15) the unfolding rate is 11.68 ± 0.05 s^{-1}. For all constructs the unfolding rate is unchanged from that measured in the $(I27)_3$ construct. Therefore, in force-clamp experiments probing the properties of HPP chains there are 2 stringent mechanical fingerprints for the identification of a single construct; 3 unfolding I27 proteins and the correct unfolding rate for I27 unfolding at a particular force. These experiments show that individual polyQ chains, ranging in length from 15 to 75 repeats, form compact and highly stable conformations which are insensitive to an applied force of 180 pN. By contrast, if an $(I27)_3$ polyprotein was pulled along with other $(I27)_3$ polyproteins the pulling force would be distributed across all of the proteins and the unfolding rate would significantly decrease. As an example we can consider the unfolding rate if an $(I27)_3$ polyprotein is pulled in tandem with another $(I27)_3$ polyprotein. In this scenario the force will be evenly distributed across both proteins. If the pulling force is 180 pN, then each protein will experience a force of only 90 pN. We can calculate the unfolding rate α(F) at a particular force from an Arrhenius term where $\alpha(F) = \alpha_0 \exp(F\Delta x/kT)$ where α_0 andΔ x are known quantities for the I27 protein (6). In Figure S2, the gray dashed line represents the expected P(t) for this arrangement. It is clear that the unfolding rate is substantially different from that of a single polyprotein construct. Indeed, the measured unfolding rate would reduce from 14.62 ± 0.05 s^{-1} to 0.05 ± 0.001 s^{-1}. In the case of a 3 polyproteins pulled in tandem, the force would be equally distributed across 3 modules and each protein would experience a third of the pulling force (black

dotted line in Fig. S2). The measured unfolding rate would subsequently reduce from 14.62 ± 0.05 s^{-1} to 7.28 × 10^{-3} ± 1 × 10^{-4} s^{-1}.

Probing the Mechanical Properties of Polyalanine Chains. To test whether the mechanical resilience observed in polyQ chains is a unique property of HPP chains we designed and engineered a construct containing a different HPP sequence. We made the chimera construct (I27-A25), containing both I27 and 25 alanine residues (see *Materials and Methods*). Specifically, the construct was (I27-A25-I27-A25-I27) (Table S1). It is noteworthy that alanine expansions have also been linked to HPP expansion diseases (7). We measured the molecular properties of (I27-A25) by applying a constant force of 180 pN, resulting in a series of step increases in length of 24 nm. A histogram of $L_{Initial}$ is shown in Fig. S5 which has an average $L_{Initial}$ of 20.19 ± 9.85 nm. The measured length of $L_{Initial}$ is significantly lower than that expected L_{WLC} (black shaded regions Fig. S5). Instead, the average $L_{Initial}$ is close to that expected for extension of only the folded I27 and amino acid handles and linkers (gray shaded regions Fig. S5). These measurements demonstrate that a different HPP chain, polyalanine, also forms a highly compact structure which is insensitive to an applied force of 180 pN. Strikingly, these measurements suggest that mechanical resilience may be a common feature of HPP chains, in contrast with heteropolypeptide chains such as PEVK.

Calculating the normalized length extension of a polypeptide segment. To directly compare the molecular properties of HPP and heteropolypeptide segments we calculated the normalized length extension of only the polypeptide segment in each construct, L_{PP}.

$$L_{PP} = \frac{L_{Initial} - (L_{127}^{WLC} + L_{Link}^{WLC})}{N_{PP} \times L_{AA}}$$

where $L_{Initial}$ is the experimentally measured initial extension for each construct, $L_{I27}{}^{WLC}$ is the calculated extension of the folded I27 proteins, and $L_{Link}{}^{WLC}$ is the calculated extension of the handles and linkers using the WLC model of polymer elasticity at a force of 180 pN, N_{PP} is the number of amino acid residues in the polypeptide segment and L_{AA} is the average length of an amino acid. We use this approach to calculate L_{PP} for HPP and heteropolypeptides chains of similar length. In particular, we calculate the L_{PP} for the (I27-Q50) construct (black bars) and the [I27-$(Q11P)_4$] construct (blue bars), which both have N_{PP} approximately 100 (Table 1). Gaussian fits to the curves yield average values of 0.05 ± 0.21 for (I27-Q50) and 1.15 ± 0.35 for [I27-$(Q11P)_4$] (Fig. S6). Strikingly, the calculated L_{PP} demonstrate that while the heteropolypeptide segment Q11P readily extends under an applied force, the HPP segment Q50 does not extend.

Molecular Dynamics Simulations. With the aim of gaining insight into the molecular architecture of polyQ chains we completed molecular dynamics simulations to probe the possible configurations of polyQ and the corresponding mechanical stability. Simulations were carried out at constant pressure (1 bar) and temperature (300 K, except heat-annealing simulations) with the Gromacs 4.0 simulation suite (http://www.gromacs.org) (8). The GROMOS force field was applied to proteins (9) and a time step of 2 fs was used. A typical 1-nm cutoff distance for the calculation of Lennard−Jones and electrostatic interactions was chosen. The particle-mesh Ewald method was used to treat the long-range electrostatic interactions (10).

To efficiently search through the conformational space a heat annealing procedure was used and the solvation environment was treated with an implicit solvent model, that is, Generalized Born model (11). The heat-annealing simulations were started from the equilibrated configuration of a polyQ chain of length Q = 25 (Q25) after a 1-ns equilibration in an implicit solvation environment. Since in our AFM experiments an extended polyQ state was never observed (Fig. 2), the distance between the 2 terminal residues of the Q25 chains was constrained to be within 6 A in the heat-annealing simulations. By comparison, we also performed the heat-annealing simulations without constraint. The Q25 chains were first heated to a high temperature (T = 700, 750, 800 K, separately) for up to 100 ns where the configurations were chosen every 200 ps. The molten configurations were then annealed to 300 K using 6 different schemes. For scheme 1, the Q25s were first annealed quickly to 500 K in 50 ps and then slowly to 300 K in 200 ps, whereas in scheme 2 the cooling to 500 K was done for 100 ps and the subsequent cooling to 300 K was done in 400 ps; for scheme 3 and 4, Q25s were first annealed slowly to 500 K in 200 ps and then quickly to 300 K in 50 ps, whereas in scheme 4 the cooling to 500 K was done for 400 ps and the subsequent cooling to 300 K was done in 100 ps; for scheme 5, Q25s were constantly annealed to 300 K in 300 ps, whereas in scheme 6 the cooling to 300 K was done for 600 ps. By comparing the heat-annealing simulation with and without constraints we find that polyQ chains exhibit a strong tendency toward a collapsed structure with small end-to-end distance. It is interesting to note that the distance between the 2 terminal residues fluctuates around approximately 6 Å at 700 K even in the absence of constraint probably because the collapsed structure is entropically driven and on subsequent heat annealing the end-to-end distance remains small. The end-to-end distance constraint is thus really unnecessary for generating the collapsed configuration of polyQ. After the heat-annealing simulations, the collapsed configurations were equilibrated at 300 K for 1 ns. In summary, 10,000 different configurations have been completed to obtain a representative ensemble of collapsed structures. Each collapsed Q25 configuration was then pulled at a constant rate to trigger unraveling, with the pulling rate of 0.01 nm/ps. Upon pulling a wide range of unraveling forces were measured for the ensemble of collapsed configurations (Fig. 5). Significantly, 125 out of 9,000 collapsed configurations required very large forces to unravel (more than 1,000 pN).

The procedure described above permitted the identification of configurations exhibiting high mechanical stability with an implicit solvent model. A number of representative configurations which display high mechanical stability are shown in Fig. 5*B* in the main text. In these structures, the residues are bound together by more than 40 hydrogen bonds between them. These 125 configurations were further equilibrated for 10 ns in an implicit solvation environment, and then solvated in a 4 × 4 × 8 nm^3 box of SPC/E water molecules (12). The entire system was subsequently equilibrated for 2 ns. Due to the competition of explicit water, the number of hydrogen bonds of all these Q25 configurations quickly decreased to various extents. In some configurations, the hydrogen bond number decreased to 10, and the Q25s essentially lost their structures. However, there are 23 configurations that possess a very compact geometry and which preserve their structure, maintaining around 20 hydrogen bonds. After further 5-ns equilibration in explicit solvation environment, those 23 configurations were then pulled at a constant rate of 0.01 nm/ps to trigger unraveling. A number of properties of these representative structures were measured. Namely, the unfolding force required to unravel the collapsed structure F_{UN}, the initial radius of gyration R_g and number of hydrogen bonds in the collapsed structure N_{HB}. Interestingly, 5 configurations exhibit very high mechanical stability with rupture forces of more than 900 pN (Fig. S3). Remarkably, the configuration with highest mechanical stability requires up to 1,500 pN to unravel.

Similar to the AFM experiment, SMD simulations of I27 have also been performed using the same pulling rate. The rupture force of I27 is around 850 pN, which is comparable with value of 900 pN for the most stable configurations of polyQ. To further

compare the mechanical stability of Q25 with I27, both I27 and Q25 were submitted to a lower pulling rate: 0.002 nm/ps. At this lower pulling rate, for I27 the rupture force significantly decreases to 510 pN. The behavior of rupture force with pulling rate for the I27 protein is consistent with the previous work (13, 14). On the other hand, it is interesting to note that the forces of unraveling Q25s decrease much less. For example, the unraveling force of the most mechanical stable configuration slightly decreases from 1,500 pN to 1420 pN. This leads us to speculate that at the much lower pulling rates of actual AFM experiments, the difference between rupture forces for Q25 and I27 could be larger than found in the simulations. Compared to the AFM experiments, the ratio of mechanically resilient configurations of Q25 is quite rare in the heat-annealing protocol. One of the possible reasons is the choice of implicit solvent model, and the relatively short equilibration time. In this way, we further performed the 20-ns equilibration of 5 highly stable configurations. The mechanical stability increased to various extents, e.g., the highest mechanical stability slightly increases to 1,550 pN with pulling rate of 0.01 nm/ps.

There are several common characteristics shared by these 5 highly stable, collapsed configurations: both N- and C-terminal glutamine residues are bound together and wrapped by the surrounding residues and many hydrogen bonds are coupled together to resist the unraveling force. Furthermore, upon close examination of the force induced unraveling we find that up to 17 hydrogen bonds must be rupture simultaneously before unraveling occurs (Fig. S4). Due to the relatively long side-chain of glutamine, the surrounding glutamines can wrap around both terminal residues, and the hydrogen bonds between the neighboring strands containing the termini and distal glutamines form a network that resists the rupture force. This is in direct contrast with a protein such as I27 where the main barrier to unfolding is although to be the rupture of only 6 hydrogen bonds. Given that I27 easily unfolds at a pulling force of 180 pN and that 6 hydrogen bonds are broken we can speculate that the rupture of 17 hydrogen bonds would require a higher pulling force, as indeed seems to be the case. Significantly, this force could be of the order of a covalent bond. Given that in single molecule experiments the construct is likely covalently attached to the cantilever tip and substrate, we can speculate that construct detachment occurs before or concomitant with polyQ unraveling. More importantly, while it is likely that a broad ensemble of collapsed structures exist, the representative structures obtained from MD simulations of polyQ in water provide support for the existence of highly collapsed and mechanical resilient structures.

1. Bustamante C, Marko JF, Siggia ED, Smith S (1994) Entropic elasticity of lambda-phage DNA. *Science* 265:1599–1600.
2. Carrion-Vazquez M, et al. (1999) Atomic force microscopy reveals the mechanical design of a modular protein. *Proc Natl Acad Sci USA* 96:3694–3699.
3. Li HB, et al. (2001) Multiple conformations of PEVK proteins detected with single molecule techniques. *Proc Natl Acad Sci USA* 98:10682–10686.
4. Sarkar A, Caamano S, Fernandez JM (2007) The mechanical fingerprint of a parallel polyprotein dimer. *Biophys J* 92:L36–L38.
5. Efron B (1982) in *The Jackknife, the Bootstrap, and Other Resampling Plans* (Society for Industrial and Applied Mathematics, Philadelphia, PA).
6. Dougan L, Fang G, Lu H, Fernandez JM (2008) Solvent molecules bridge the mechanical unfolding transition state of a protein. *Proc Natl Acad Sci USA* 105:3185–3190.
7. Brown LY, Brown SA (2004) Alanine tracts: The expanding story of human illness and trinucleotide repeats. *Trends Genet* 20:51–58.
8. Lindahl E, Hess B, van der Spoel D (2001) Determining the shear viscosity of model liquids from molecular dynamics simulations. *J Mol Model* 7:306–317.
9. Lukas D, Schuler LD, Daura X, van Gunsteren WF (2001) An improved GROMOS96 force field for alphatic hydrocarbons in the condensed phase. *J Comp Chem* 22:1205–1218.
10. Darden T, York D, Pedersen L (1993) Particle mesh Ewald (PME): A N log(N) method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092.
11. Roux B, Simonson T (1999) Implicit solvent models. *Biophys Chem* 78:1–20.
12. Berendsen HJC, Grigera JR, Straatsma TP (1987) The missing term in effective pair potentials. *J Phys Chem* 91:6269–6271.
13. Best RB, Li B, Steward A, Daggett V, Clarke J (2001) Can non-mechanical proteins withstand force? Stretching barnase by AFM and MD simulation. *Biophys J* 81:2344–2356.
14. Lu H, Schulten K (1999) The key event in force-induced unfolding of titin's immunoglobulin domains. *Chem Phys* 247:141–153.

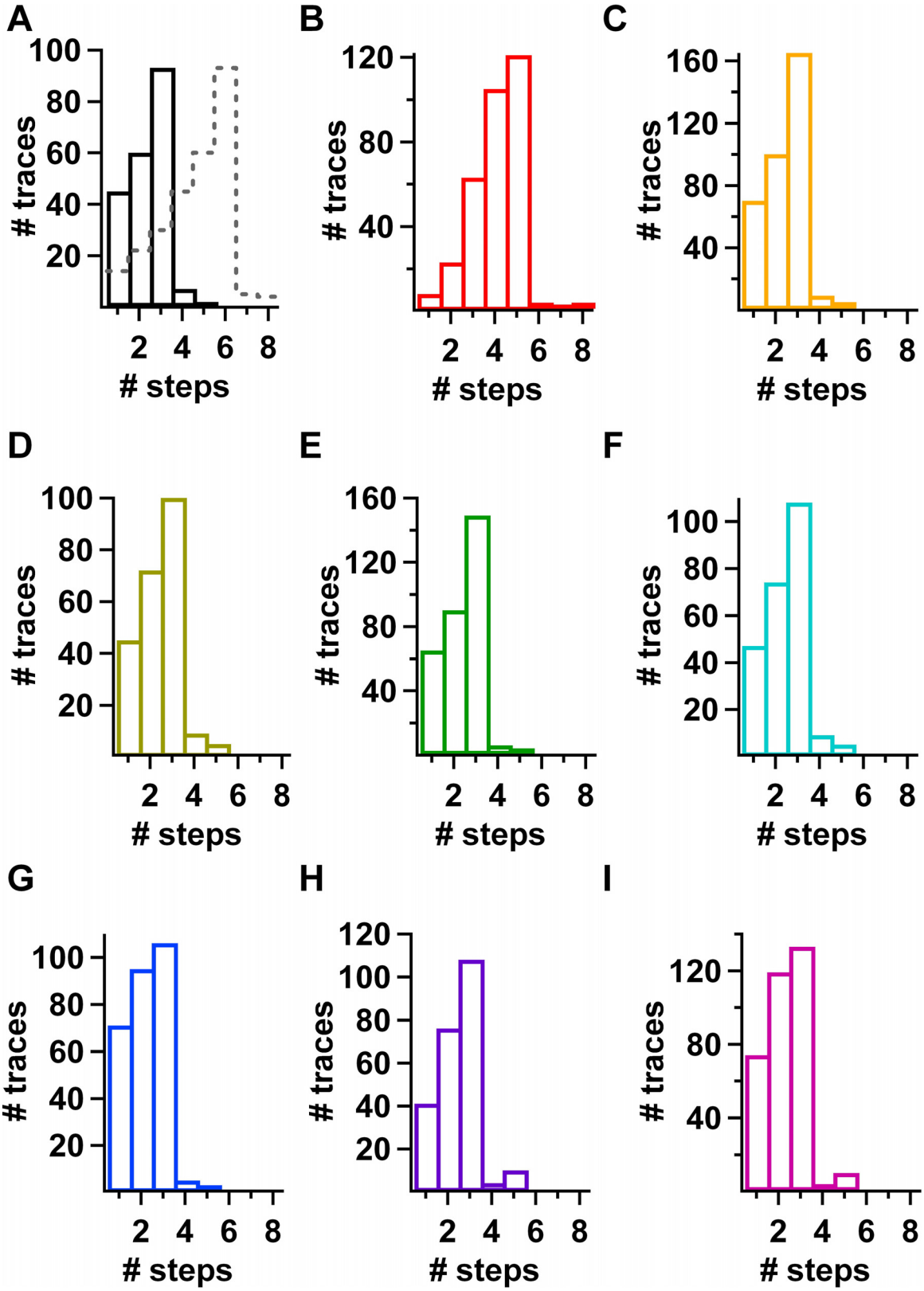

Fig. S1. Number of traces containing 1–8 I27-unfolding steps for; (*A*) $(I27)_3$, (*B*) (I27-Q15), (*C*) (I27-Q25), (*D*) (I27-Q50), (*E*) (I27-Q75), (*F*) (I27-A25), (*G*) (I27-PEVK), (*H*) [I27-$(QP)_{24}$], and (*B*) [I27-$(QP)_{11}$]. Also, in (*A*), a hypothetical distribution of steps (gray dashed line) for a protein dimmer of the $(I27)_3$ construct.

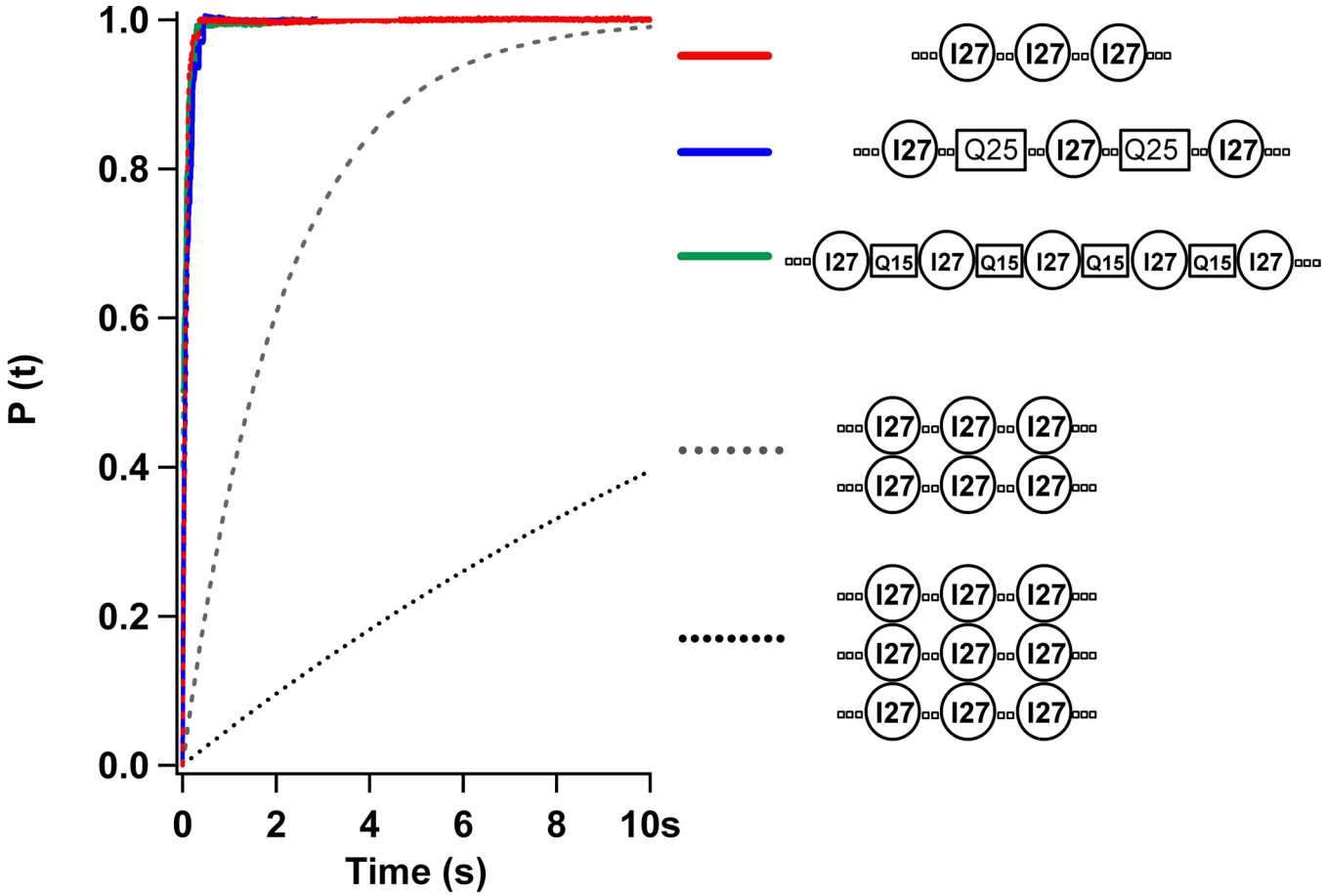

Fig. S2. The unfolding rate of 3 different constructs, $(I27)_3$ (red), (I27-Q25) (blue), and (I27-Q15) (green), is found by summation and normalization of $n > 30$ unfolding traces containing 3 I27-unfolding steps at a force of 200 pN. The unfolding rate for $(I27)_3$ is 14.62 $\pm$ 0.05 s^{-1}, for (I27-Q25) the unfolding rate is 10.53 $\pm$ 0.05 $s^{-}1$, and for (I27-Q15) the unfolding rate is 11.68 $\pm$ 0.05 $s^{-1.}$ The calculated unfolding rate for 2 $(I27)_3$ constructs pulled in tandem is 0.05 $\pm$ 0.001 s^{-1} (gray dashed line). The calculated unfolding rate for 3 $(I27)_3$ constructs pulled in tandem is $7.28 \times 10^{-3} \pm 1 \times 10^{-4}$ s^{-1} (black dotted line).

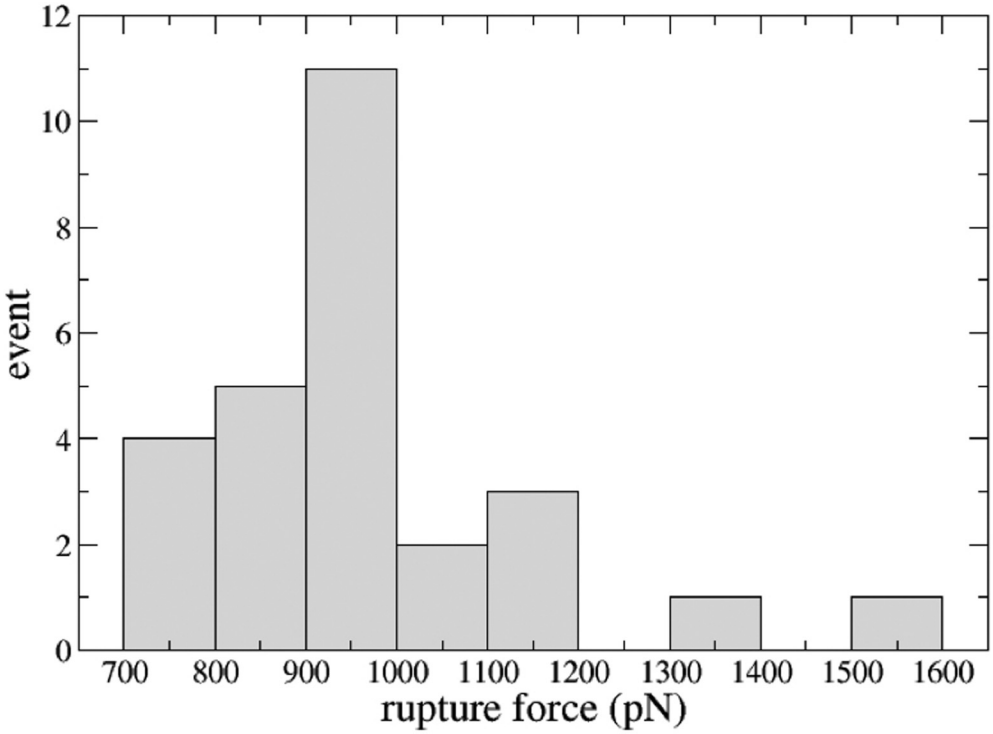

Fig. S3. Extreme mechanical stability in polyglutamine chains The distribution of extreme rupture forces for a polyQ chain of length $n = 25$ obtained from the steered molecular dynamics simulations.

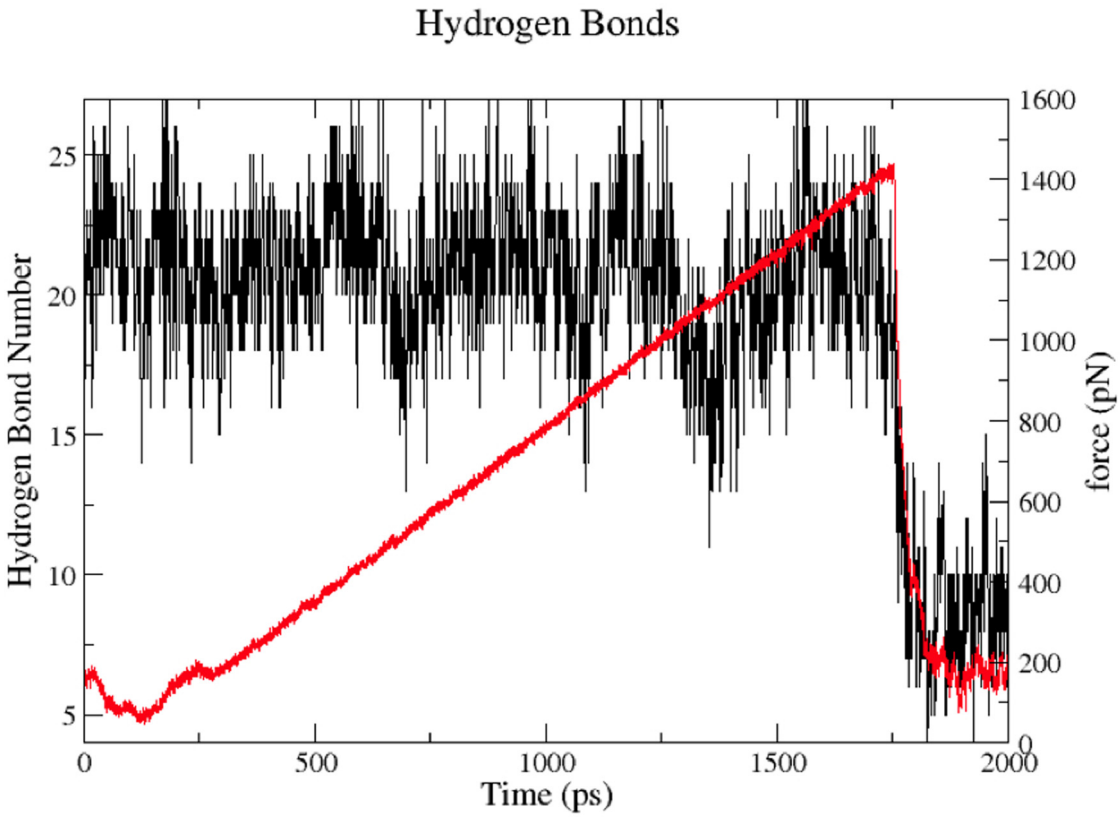

Fig. S4. Hydrogen bond rupture in polyglutamine chains in the steered molecular dynamics simulation. We measured the number of hydrogen bonds in the polyQ chain of length Q = 25 chain during the simulation. In the example shown, we find that the number of hydrogen bonds decreases from 23 to 6 at the point when the polyQ chain increases in length. Therefore, within 40 ps (between 1,750 ps and 1,790 ps) 17 hydrogen bonds rupture before unraveling occurs.

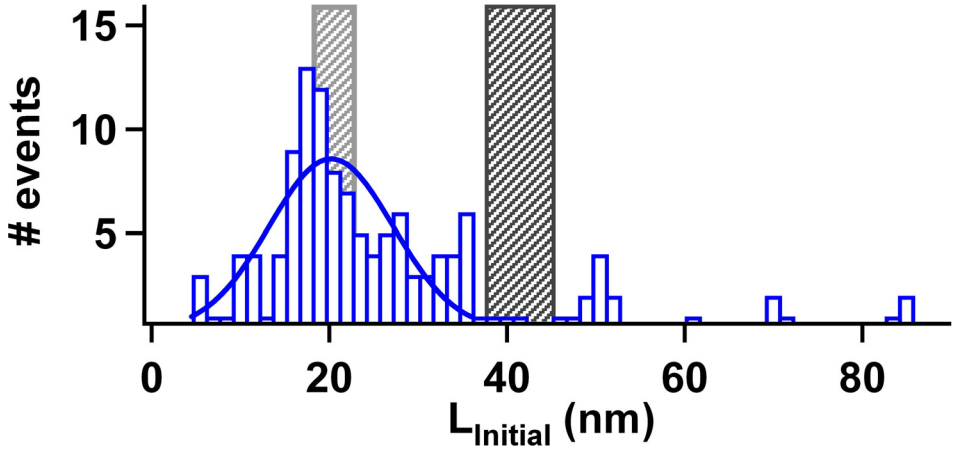

Fig. S5. Probing the molecular properties of polyalanine chains. We construct chimeras containing the I27 protein and polypeptide chains of the amino acid alanine, A25. Using force-clamp spectroscopy we apply a constant force of 180 pN along the end to end length of the constructs. We identify extension of a complete construct by the presence of 3 I27-unfolding steps, where each I27-unfolding step is 24 nm in length. We measure the initial extension $L_{Initial}$ for each trajectory which satisfies these stringent criteria. A histogram of $L_{Initial}$ is shown for $n = 108$ traces. A Gaussian fit to the histograms (solid line) gives an average $L_{Initial} = 20.19 \pm 9.85$ nm. The measured $L_{Initial}$ is significantly shorter than that expected for full extension of the construct (black shaded area in histogram). Instead $L_{Initial}$ is in close agreement with the expected length extension of only the folded I27 proteins and linkers (gray shaded area). The data suggests that the polyalanine chains have not extended.

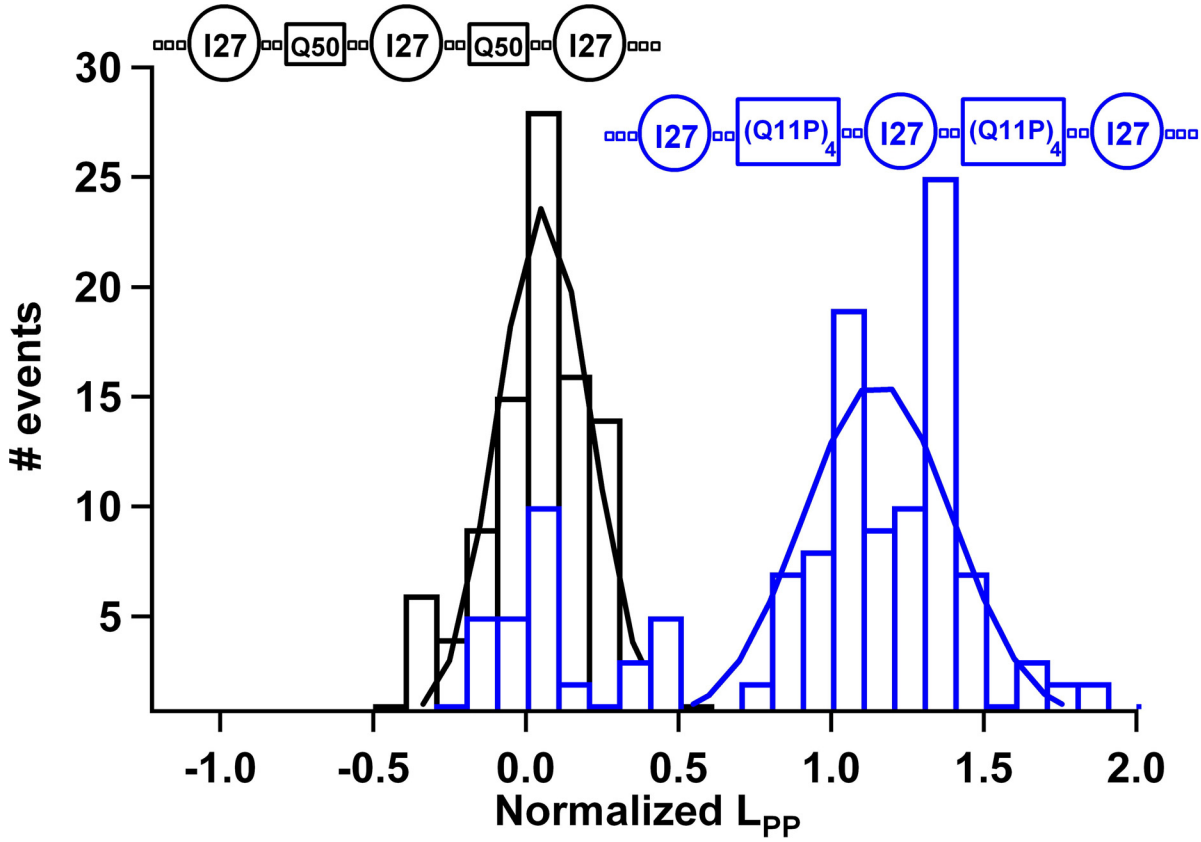

Fig. S6. Homopolypeptide chains do not readily extend under an applied force while heteropolypeptide chains behave as random coils. (*A*) The normalized extension of the HPP and heteropolypeptide chain L_{PP} is calculated for each of the constructs using the experimentally measured initial extension $L_{Initial}$. We calculate L_{PP} for the (I27-Q50) construct (black bars) and the [I27-$(Q11P)_4$] construct (blue bars). Gaussian fits to the curves yield average values of 0.05 $\pm$ 0.21 for (I27-Q50) and 1.15 $\pm$ 0.35 for [I27-$(Q11P)_4$]. The calculated L_{PP} demonstrate that while the heteropolypeptide chains PEVK and Q11P readily extend under an applied force, the HPP chains Q25 and Q50 do not extend.

Table S1. Homopolypeptide and heteropolypeptide polyprotein constructs

Construct	N_{PP}	N_{Link}	Av. $L_{Initial}$ (nm)	L_{WLC} (nm) $P = 0.2$ nm $P = 2.3$ nm
$(I27)_3$	0	20	13.30 ± 7.10	18.06 21.07
$(I27\text{-}Q15)_4\text{-}I27$	60	32	31.16 ± 8.21	49.81 58.11
$(I27\text{-}Q25)_2\text{-}I27$	50	24	13.79 ± 9.63	36.20 42.24
$(I27\text{-}Q50)_2\text{-}I27$	100	28	26.09 ± 7.93	54.35 63.41
$(I27\text{-}Q75)_2\text{-}I27$	150	32	20.65 ± 6.98	72.49 84.57
$(I27\text{-}A25)_2\text{-}I27$	50	24	20.19 ± 9.85	36.20 42.24
$(I27\text{-}PEVK)_2\text{-}I27$	56	24	39.65 ± 17.26	38.22 44.59
$(I27\text{-}QP)_{24})_2\text{-}I27$	96	32	40.06 ± 17.26	61.91 72.23
$(I27\text{-}Q11P)_4)_2\text{-}I27$	96	28	70.00 ± 12.39	53.00 61.84

Details of the constructs made including the number of amino acids in the linkers and handles N_{Link}, the number of amino acids in the polypeptide segment N_{PP}, the average initial extension $L_{Initial}$ and the expected extension of the folded I27 proteins, the linkers and handles and the polypeptide segment calculated using the Worm Like Chain model of polymer elasticity [Bustamante C, Marko JF, Siggia ED, Smith S (1994) *Science* 265:1599–1600], L_{WLC}, at a force of 180 pN and a persistence length of 0.2 nm and 2.3 nm.