

How Sequence Determines Elasticity of Disordered Proteins

Shanmei Cheng,[†] Murat Cetinkaya,^{†‡} and Frauke Gräter^{†‡*}

[†]German Max Planck Society-Chinese Academy of Sciences Partner Institute and Key Laboratory for Computational Biology, Shanghai Institute for Biological Sciences, Chinese Academy of Sciences, Shanghai, China; and [‡]Bioquant, Heidelberg University, Heidelberg, Germany

ABSTRACT How nature tunes sequences of disordered protein to yield the desired coiling properties is not yet well understood. To shed light on the relationship between protein sequence and elasticity, we here investigate four different natural disordered proteins with elastomeric function, namely: FG repeats in the nucleoporins; resilin in the wing tendon of dragonfly; PPAK in the muscle protein titin; and spider silk. We obtain force-extension curves for these proteins from extensive explicit solvent molecular dynamics simulations, which we compare to purely entropic coiling by modeling the four proteins as entropic chains. Although proline and glycine content are in general indicators for the entropic elasticity as expected, divergence from simple additivity is observed. Namely, coiling propensities correlate with polyproline II content more strongly than with proline content, and given a preponderance of glycines for sufficient backbone flexibility, nonlocal interactions such as electrostatic forces can result in strongly enhanced coiling, which results for the case of resilin in a distinct hump in the force-extension curve. Our results, which are directly testable by force spectroscopy experiments, shed light on how evolution has designed unfolded elastomeric proteins for different functions.

INTRODUCTION

Elastomeric proteins are present in a wide range of living organisms, and are utilized for their toughness and flexibility. Structural disorder and associated hydration are critical features of these elastomeric proteins (1). Although experimental techniques such as NMR and x-ray crystallography can measure well-defined tertiary structures that have specific functions in the cell, a majority of natural proteins are intrinsically unstructured, featuring an elasticity specifically tailored for their distinct force-bearing and sensing functions (2). Their secondary structures are typically transient and confined to the exchanges among Polyproline II (PPII), β -strand, β -turn, and irregular structures (3).

Disordered proteins show low sequence complexity and significant amino-acid compositional bias. Their unstructured conformations are mainly due to their sequence biases. Glycine and proline are two particular residues that often appear in the sequence of disordered proteins. Both of them contribute to disorder but for opposite reasons. Glycine, lacking any side chain, is so flexible that order is entropically unfavorable. Proline's cyclic side chain, by contrast, is stiff and therefore highly restricted in, and mostly disturbing, regular secondary structure formation. A recent study has proven that the glycine and proline compositions of peptides affect their amyloid formation tendency and elasticity (1). More recently, the elasticity of silk was shown to correlate highly with glycine content by measurements from circular dichroism spectroscopy (4).

How amino-acid sequence can affect the entropic and enthalpic contributions to the elasticity of a disordered protein is still unknown. The elasticity of disordered or unfolded proteins can be studied experimentally with single-molecule AFM (5,6). The persistence length commonly used to analyze AFM force-extension data reflects an effective protein elasticity incorporating entropic as well as attractive interactions, most likely of hydrophobic nature, which lower the persistence length (6,7). Various models have been proposed to study the elasticity of DNA and protein. The wormlike chain (WLC) model (8) is the most commonly used by

$$F(z) = \frac{k_B T}{4L_p} \left[\frac{1}{(1 - z/L_0)^2} - 1 + \frac{4z}{L_0} \right], \quad (1)$$

where F is the force, z is the end-to-end extension, L_p is the persistence length, k_B is the Boltzmann constant, T is the temperature, and L_0 is the contour length of the polymer. At low extensions, force grows linearly as a Hookean spring, and at high extensions the force diverges as

$$(k_B T/L_p) [4(1 - z/L_0)^2]^{-1}.$$

This model is able to describe the mechanical response of dsDNA and protein chains fairly well. A high persistence length reflects low flexibility and vice versa.

Four different intrinsically disordered proteins have been chosen for this study of molecular determinants of elasticity. First, phenylalanine-glycine (FG) repeats are disordered proteins which form a major component of nuclear pore complexes. They consist of up to 50 repeat units and appear intrinsically unfolded, containing short clusters of hydrophobic amino acids. They facilitate the passage of protein

Submitted June 25, 2010, and accepted for publication October 6, 2010.

*Correspondence: frauke.graeter@h-its.org

Frauke Gräter's present address is Heidelberg Institute for Theoretical Studies gGmbH, Schloss-Wolfsbrunnengweg 35, 69118 Heidelberg, Germany.

Editor: Peter Hinterdorfer.

© 2010 by the Biophysical Society
0006-3495/10/12/3863/7 \$2.00

doi: 10.1016/j.bpj.2010.10.011

and RNA through the nuclear pore by forming a three-dimensional meshwork with hydrogel-like properties. Mutation of phenylalanine to serine has a smaller impact on aggregation (9), but will destroy the formation of the hydrogel (10). Hybrid approaches have been applied to explore the behavior of FG-domains at the nanometer scale (11). A coarse-grained model combined with all-atom simulations of nuclear pore complex arrays showed a brushlike structure with a height of 77.1–88.1 Å, larger than the radius of gyration of a single chain of 13.7–20.0 Å (12). Single-molecule force spectroscopy has been used to study the flexibility of individual cNup153 molecules and mechanically verify that they are unfolded and exhibit entropic elasticity with a mean persistence length of 0.39 ± 0.14 nm (5). We here chose the repeated motif

SDTSKPAFSFGAKPDEKKDS.

Resilin from the wing tendons of the adult dragonfly is a member of a family of elastomeric proteins characterized by high resilience, low stiffness, high strain, and efficient energy storage (13). In a recent study, resilin-based near-perfect rubber was produced, achieving 97% of the resilience of natural resilin, using ultraviolet illumination (13,14). We chose the elastic repeat motif

GGRPSDSYGAPGGNGGRPS

from the *Drosophila* CG15920: 42–51 gene (15).

The Pro-Glu-Val-Lys (PEVK) domain, an unstructured domain of titin, has been shown to be an important factor for the passive elasticity of muscle (16). Two main motifs were identified in the PEVK sequence: one is the PPAK motif with a 28-residue-long sequence, and the other is the polyE motif containing a preponderance of glutamate. The PEVK segment of cardiac titin is assumed to take up multiple mechanical conformations, as indicated by a broad range of persistence lengths ($p = 0.3$ – 2.3 nm) obtained in experiment (16). The comparison between individual PEVK exons that varied greatly in their proline content and total PEVK content suggested that this wide range of persistence lengths is independent from the specific amino-acid composition (17). This indicates that, within a certain range, elastic properties are robust with regard to the detailed PEVK content of the polypeptides. In this article, we chose the PPAK motif comprising the first 20 conserved amino acids

PPAKVPEVPKPKVPPEEKVPV

from motif alignment (18).

Silk fibers constitute an interesting class of materials due to their exceptional mechanical characteristics involving high tensile strength and toughness with a low mass density (19). The ability of high impact absorbency and thus, high elasticity of silk fibers, arise from the amorphous, glycine-rich repeating units that are supposed to act like a cross-link-

ing matrix between strong crystalline blocks (20). We chose the amorphous repeating unit of the spider silk from major ampullate gland of the spider *Araneus diadematus* (19) having the sequence

GPGGYGPGSQGPSGPGGYGPGGGPG.

The above four representatives of elastomeric proteins have the advantage to widely vary in the amino-acid content. Resilin from dragonfly wings is rich in glycines (40%), PPAK is rich in prolines (35%), and the FG repeat from nuclear pore complex is in-between, with 10% proline and 5% glycine, while being relatively rich in hydrophobic residues. Silk is high in both glycines (54%) and prolines (25%). Both glycine, lacking a side chain, and proline, with a rigidifying ring, can be expected to alter a chain's entropic elasticity due to the increased, or decreased, respectively, dihedral degrees of freedom. In previous work we have dissected the driving forces for coiling of the unfolded state of typical globular proteins in water by molecular dynamics (MD) simulations into contributions from entropic coiling and hydrophobic forces (6,7). The entropic coiling propensity of an unfolded protein, with a persistence length of ~ 1.2 nm, is effectively increased—resulting in the frequently measured apparent persistence length of 0.3–0.6 nm under the poor solvent conditions of a largely hydrophobic protein chain in water (7).

Here, we first access the backbone elasticity of the above four disordered proteins by entropic chain simulations, and compare their entropic elasticity in terms of persistence lengths from WLC fits. Second, we investigate influences besides chain entropy, such as charged interactions and the hydrophobic effect on the protein elasticity via all-atom MD simulations.

MATERIALS AND METHODS

All simulations were carried out using the MD software package GROMACS 3.3.1 (21). The intrapeptide interactions were defined using the OPLS/AA force field (22). Water was represented by the transferable intermolecular potential (TIP4P) model (23). Simulations were run in the NpT ensemble. The temperature was kept constant at $T = 300$ K by coupling to a Nosé-Hoover thermostat with a coupling time of $\tau_T = 0.1$ ps. The pressure was kept constant at $p = 1$ bar using an isotropic coupling to a Berendsen barostat with a coupling time of $\tau_T = 0.1$ ps and an isotropic compressibility of 4×10^{-5} bar⁻¹. All bonds were constrained using the LINCS algorithm. An integration time step of 2 fs was used. Lennard-Jones interactions were calculated using a cutoff of 1.4 nm. At distances < 1.0 nm, the electrostatic interactions were calculated explicitly, whereas long-range electrostatic interactions were calculated by particle-mesh Ewald summation in umbrella sampling runs.

Calculated coiling propensities might heavily depend on the chosen force field. Before studying the elasticity of disordered proteins, four force fields were tested for two turns of ubiquitin (PDB code 1UBQ), i.e., turn I (residues 4–16) and turn II (residues 21–33), and results were compared to experiments. The force fields were GROMOS43A2 (24), GROMOS53A6 (25), and OPLS (22) with single point charge (SPC) water (26) as well as OPLS with TIP4P water (23). Force-clamp spectroscopy experiments showed that after a force quench from 100 pN to 10 pN on a single

polyubiquitin protein, proteins that folded collapsed to 0.42 ± 0.22 of the length at 100 pN, whereas those that failed to fold only showed a reduction in length to 0.72 ± 0.12 (6). We consider the latter value for nonfolding proteins as the one to be expected for the ubiquitin fragments. In our simulations, at a high force of 100 pN, all force fields gave an end-to-end distance of ~ 3.8 nm. At a low force of 10 pN, simulations were repeated three times for each force field. The distributions of normalized length of 10 pN vs. 100 pN show that the GROMOS43A2 force field with SPC water has the strongest coiling propensity with a normalized length of 0.12 ± 0.04 nm and 0.27 ± 0.11 nm, respectively, for the two turns. GROMOS53A6 with SPC water gave 0.73 ± 0.15 nm for both turn I and II and OPLS with TIP4P water gave 0.65 ± 0.29 nm and 0.68 ± 0.25 nm for turn I and II, respectively. Thus, both GROMOS53A6 with SPC and OPLS with TIP4P result in coiling tendencies of the two peptides, which are in quantitative agreement with the experiments.

In fact, the free energy of solvating the side chains of all amino acids except glycine and proline are overestimated by three of the four tested force fields as compared to the experimental results (27–29). Whereas GROMOS43A2 with SPC overestimates solvation free energies most strongly on average, GROMOS53A6 with SPC and OPLS with TIP4P have been found to perform best in this respect, in agreement with our simulation results. We chose OPLS with TIP4P for subsequent simulations.

Disordered structures were constructed using PyMOL and pulled under a constant stretching force of 500 pN applied to the N- and C-termini, acting outwards along the z axis in solution. The resulting extremely extended structures were placed in a box large enough to accommodate the protein, and water was added, along with ions, to obtain a salt concentration of 0.1 M. After a steepest-descent energy minimization, all nonhydrogen atoms were restrained for a 100-ps equilibrium simulation. Forty-nanosecond MD simulations with force quenched to 2 pN were then performed. In these collapse simulations, equilibrium configurations were clustered to get representative collapsed structures. These representative structures (11.7 ns, 34.5 ns, 27.9 ns, and 20 ns for FG, resilin, PPAK, and silk, respectively) were then pulled with a constant velocity of 4×10^{-5} nm/ps and a force constant of $500 \text{ kJ/mol} \cdot \text{nm}^2$ to generate starting structures span-

ning a wide range of extensions for subsequent umbrella simulations. Trajectories were used with the z component of the end-to-end distance ranging from 0.1 nm per bond to 0.35 nm per bond and at least nine windows were chosen with a width of ~ 0.025 nm per bond each. The simulation time in each umbrella window varied from 15 ns for large extensions to 100 ns for smaller extensions, depending on the convergence. A force constant of $100 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2}$ was used for all umbrella sampling simulations. Simulations were repeated with different starting structures and in total 1.44 μs , 1.69 μs , 0.56 μs , and 1.08 μs of MD simulations for FG, resilin, PPAK, and silk, respectively, were carried out. The potential of mean force was calculated using the weighted histogram analysis method (WHAM) to avoid biased sampling (30). The resulting potentials of mean force are shown later in Fig. 3.

The umbrella sampling simulations were repeated for entropic versions of the protein. The force field was modified by switching off all electrostatic interactions and the attractive part of the Lennard-Jones interactions (C6), leaving only bonded interactions and the repulsive part of the Lennard-Jones interactions (C12). A detailed description of this procedure has been given elsewhere (7). The resulting model represents a purely entropic chain, in which the conformational ensemble is only restricted by the bonded interactions, mainly the dihedral potentials along the protein backbone, and by the local steric repulsion of the atoms. This entropic chain model incorporates the correct volume exclusion, backbone geometry, and conformational freedom of the protein and sugar molecules, and ignores nonlocal interactions and solvent-induced effects. However, removal of the C6 term of the Lennard-Jones interaction will increase the effective size of each atom. To compensate for this effect, we divided the C12 coefficients by a factor of 4, to better mimic a Weeks-Chandler-Anderson potential (31). Entropic chain simulations time varied from 100 ns for large extensions to 500 ns for smaller extensions depending on the convergence. The total simulation time was 7.4 μs , 7.4 μs , 7 μs , and 9.5 μs for FG, resilin, PPAK, and silk, respectively. The results were also analyzed by WHAM and are shown in Fig. 1.

The first 1-ns of every simulation were omitted from calculations of means and standard errors, to allow for equilibration.

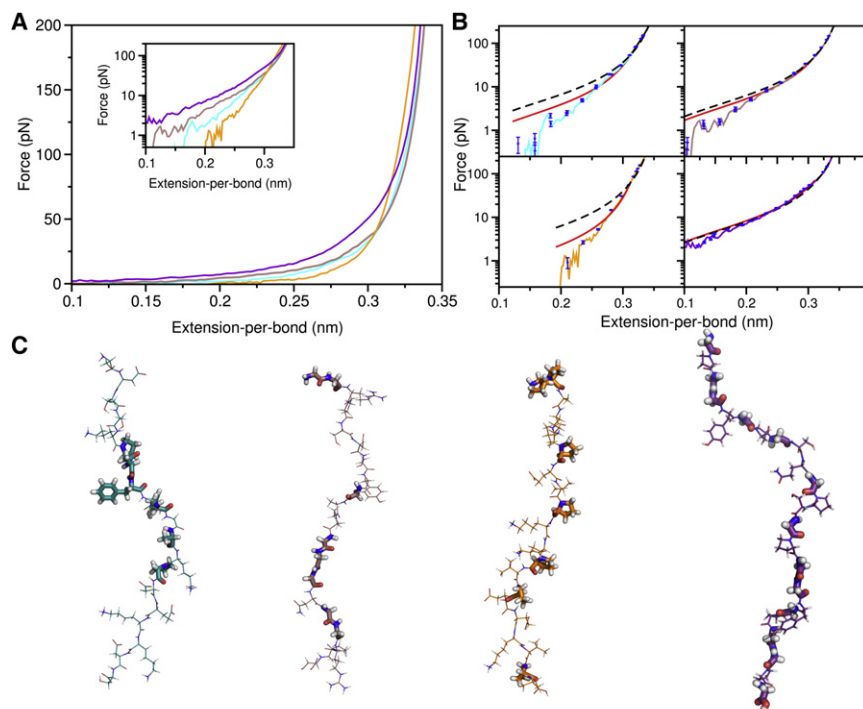


FIGURE 1 Chain entropy of the four disordered proteins. (A) Force extension curves of FG (cyan), resilin (brown), PPAK (orange), and silk (violet) obtained from entropic chain simulations and weighted histogram analysis method (WHAM). (Inset) Semilogarithmic plot of panel A. (B) Force extension curves for individual disordered protein from entropic chain simulations (colored lines as in panel A), wormlike chain (WLC) fit to all data (black dashed line), and WLC fit to only the data with forces < 60 pN (red solid lines). Average forces with standard error of the mean in umbrella sampling windows (blue circles). (C) Snapshots of FG, resilin, PPAK, and silk from left to right with an extension-per-bond of 0.25 nm. Highlighted are particular sequence features in each structure, namely Phe and Gly (FG), Gly (resilin), Pro (PPAK), and Gly (silk).

RESULTS AND DISCUSSION

Elasticity of entropic chains

The wormlike chain (WLC) model is a frequently-used model to measure protein stiffness. It regards the protein as a slender cylindrical elastic rod with a fixed contour length L_0 . It is an entropic model that does not consider the effects of hydration. We first treated the protein chains as entropic chains by switching off the electrostatic and attractive van der Waals interactions, leaving only excluded volume interactions (7). Force-extension curves for these entropic chains were obtained by umbrella sampling combined with the weighted histogram method (Fig. 1), in which the peptide chain was held at a range of extensions by a pair of harmonic potentials acting on the terminal atoms. To allow comparison between the four chains with different absolute lengths, we henceforth give extensions and contour lengths as distances per interresidue bond. Among the polypeptides, silk is the only one that can be well fitted by the WLC model, with a persistence length of 0.776 ± 0.003 nm and a contour length per bond of 0.365 nm. The others depart significantly from the best-fit WLC curves, especially at low extensions with forces <50 pN. By fitting only the data with an extension-per-bond below 0.31 nm, we obtained $L_p = 1.79 \pm 0.77$ nm, $L_0 = 0.349 \pm 0.001$ nm per bond for FG, $L_p = 1.26 \pm 0.035$ nm, $L_0 = 0.357 \pm 0.001$ nm per bond for resilin, $L_p = 3.36 \pm 0.17$ nm, $L_0 = 0.335 \pm 0.001$ nm per bond for PPAK, and $L_p = 0.74 \pm 0.02$ nm, and $L_0 = 0.364 \pm 0.001$ nm per bond for silk.

Silk and resilin show a higher contour length per bond and a comparably low persistence length due to the glycines for which steric clashes of the side chains are absent (Fig. 2 A, inset). In agreement, Dicko et al. (4) concluded that the emergence of elasticity in silks correlates highly with the glycine content, and that a stiffening effect due to prolines in the sequence is absent. A possible dichotomization of silk processing has been suggested, with glycine controlling the soluble precursor assembly and proline governing the solid fiber behavior (32,33).

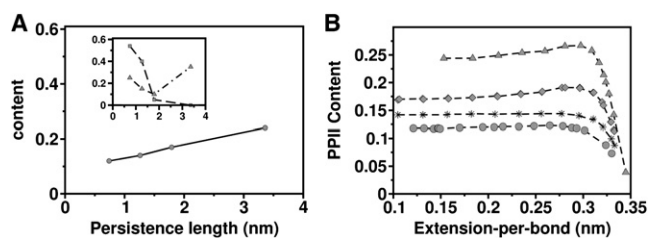


FIGURE 2 Influence of glycine and proline on entropic chain elasticity. (A) PPII content as a function of persistence length. (Inset) Glycine content (square) and proline content (triangle) versus persistence length. Persistence lengths are obtained from WLC fits (red curves in Fig. 1 B). A high correlation is found for the PPII content and glycine content. (B) PPII content of FG (diamond), resilin (star), PPAK (triangle), and silk (circle) as a function of extension-per-bond.

The distinctive cyclic structure of proline's side chain locks its ϕ -backbone dihedral angle at $\sim -75^\circ$, giving proline an exceptional conformational rigidity compared to other amino acids. Proline loses less conformational entropy upon unfolding. Hence, proline can render a protein more rigid. However, apparently, proline content is not the only determinant of protein rigidity because it is largely uncorrelated with the persistence length (Fig. 2 A, inset). For example, PPAK has the highest persistence length among the four disordered proteins not only due to the high proline content but also due to their locations in the sequence. Similarly, silk has a proline content of 25%, close to that of PPAK, but it is still the most elastic of the studied polypeptides because all the prolines are located close to glycines.

Instead of the proline content, we find the content of PPII conformation, a prevalent structure in disordered proteins, to highly correlate with entropic chain elasticity (Fig. 2 B). It is formed when sequential residues all adopt ϕ -, ψ -, and ω -backbone dihedral angles of roughly -75° , 150° , and 180° (34,35). It has no internal hydrogen bonds due to steric constraints. At low extensions, the PPII contents of chains converge to 0.25 for PPAK, 0.175 for FG, 0.14 for resilin, and 0.12 for silk (Fig. 2 B). The higher the PPII content, the more rigid the protein. The short contour length per bond of PPAK can thus be explained by its high PPII content. We note that not only proline polymers but also lysine polymers prefer a PPII conformation, as observed by experiment (36). Indeed, we here find FG, featuring a lysine content of 20%, to show a higher PPII content than resilin and silk.

In experiments (16), PPAK is found to show a wide range of persistence lengths. Possible contributions could be a heterogeneity in the *cis-trans* conformation of the X-Pro bonds, or variations in salt bridges formed between charged residues in PPAK. Our entropic chain simulations suggest that another factor could be the strong variation in PPII content, as exhibited by PPAK with a factor of 10 in the region of 0.3–0.35 nm extension-per-bond (Fig. 2 B).

Elasticity from full MD simulations

The entropic elasticity of a disordered protein correlates with its amino-acid sequence, in particular with the glycine content and the PPII structure content measured from the entropic chain simulations. However, other effects besides the entropy of the chain must also be taken into consideration. Umbrella sampling was performed for FG, resilin, PPAK, and silk polypeptides, this time including all interatomic interactions. In standard all-atom MD simulations, we again obtained the elasticities from force-extension curves (Fig. 3). Interestingly, results show the same overall tendency of the four disordered proteins as those obtained by only considering entropic chains, emphasizing the role of entropic chain elasticity in the coiling of disordered

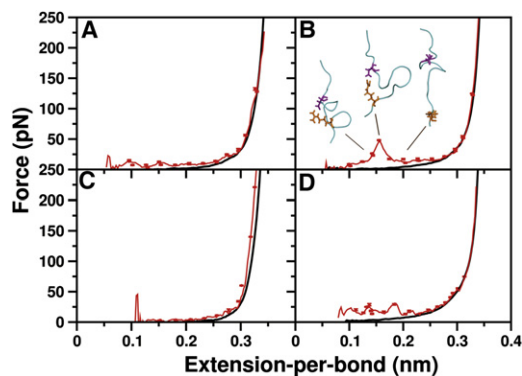


FIGURE 3 Force extension curves for FG (A), resilin (B), PPAK (C), and silk (D) from umbrella sampling and WHAM using standard MD simulations including electrostatic and solvent effects (red). For comparison, the force-extension curves of the respective entropic chains are shown (black). Sample conformations of resilin at extensions of ~ 3.0 nm, ~ 3.5 nm, and ~ 4.0 nm are shown in panel B. Charged residues of D6 (magentas) and R18 (orange) are shown in sticks.

proteins. However, there are significant alterations to the force extension curves, in particular at shorter extensions. As shown in the respective free energy profiles (see Fig. S1 in the Supporting Material), this results in an additional gain in free energy upon collapse of the peptides of up to $10 k_B T$, in which specific energy barriers are introduced. Thus, in contrast to the folding of a protein into its native state, unspecific collapse is a barrier-free diffusion limited process, at least for the four cases considered here.

For FG, hydrophobic residues are distributed over the first two-thirds of the peptide chain, starting from the N-terminus, whereas the C-terminus is richer in polar hydrophilic residues (Fig. 1 C, left). The initial burial of the central hydrophobic residues F8 and F10 leads to a reduction of the total solvent-accessible surface area (SASA) by 1.0 nm^2 , in contrast to the entropic chain model (see Fig. 4 A). A further decrease of SASA is observed upon the burial of the hydrophobic or polar residues K5, P6, S9, and P14. These events lead to the formation of a hydrophobic core by the N-terminal two-thirds of the peptide at an extension-per-bond of 0.15 nm , with the remaining one-third is still extended. This packing of the hydrophobic core is not very stable due to the competition between further burial of hydrophobic residues and repulsive interactions of the charged residues, as reflected by a zigzag shape of the SASA and electrostatic energy curves at low extensions (Fig. 4 A). Lim et al. (5) considered the flexible FG nucleoporins as entropic barriers to nucleocytoplasmic transport. In agreement, for extensions beyond $\sim 0.25 \text{ nm}$ per residue, supposedly those found in FG brushes, the force-extension curve closely follows the entropic chain behavior (Fig. 3 A). However, enhanced coiling due to the hydrophobic effect is dominant at smaller extensions.

For resilin, the force-extension curve features a pronounced hump with a peak force of $\sim 50 \text{ pN}$ at an exten-

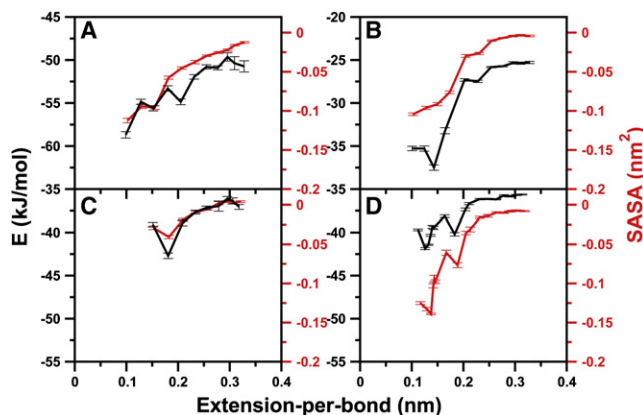


FIGURE 4 Electrostatic energy per residue (E , black) and solvent-accessible surface area per residue (SASA, red) for FG (A), resilin (B), PPAK (C), and silk (D) as a function of extension-per-bond.

sion-per-bond between 0.1 and 0.2 nm , a strong divergence from purely entropic behavior (Fig. 3 B). We analyzed the change in molecular interactions upon collapse to identify the sequence properties leading to this hump. An only minor initial collapse due to the burial of the central hydrophobic residue Y8 along with a decrease of the total SASA of 0.5 nm^2 (see Fig. 4 B) does not lead to a significant divergence of the force-extension curve from the purely entropic chain. The subsequent decrease of hydrophobic SASA by $\sim 0.93 \text{ nm}^2$ is connected to the further burial of hydrophobic residues Y8, A10, and P19, accounting for 35%, 10.7%, and 18.7% of the total decrease, respectively. The burial of these hydrophobic residues also causes the favorable burial of the neighboring charged residue R18, which leads to a reduction in electrostatic interactions by 100 kJ/mol . The attractive electrostatic interactions become stronger with a further loss of 80 kJ/mol of electrostatic energy due to the approach of polar residue R18 to D6.

We note that, here, we do not take into account any water-protein enthalpic contributions or entropic effects which are likely to partly compensate the change in internal protein electrostatic energy. Nevertheless, considering this contribution as a qualitative measure for the driving forces of collapse, we suggest that the hump at an extension-per-bond of $\sim 0.15 \text{ nm}$ in the force-extension curve (Fig. 3 B) might be caused by this electrostatic interaction. The interaction of these two charged residues leads to a distorted hairpin structure at the central glycine-rich part of the chain from G12 to G17. Due to the lack of side chains, the glycines cause very little steric hindrance when forming such a tight bend. In accordance, resilin was also observed to form a knot structure under ultraviolet illumination, mainly due to the long glycine chain (13).

In general, glycine and proline residues are frequently found in loop-and-turn structures of proteins and are also believed to play an important role during chain compaction early in folding (37). In the respective free energy profile

(Fig. S1 B), resilin features a steep drop in free energy upon hairpin formation at an extension of 0.15. We conclude that an interplay of high entropic chain flexibility and attractive electrostatic interactions along the chain can, in general, give rise to such sharp humplike elastic properties.

Due to the decreased tendency of PPAK to coil, we only were able to sample chain extension down to ~ 0.1 nm per bond (Fig. 3 A). Overall, the deviation from the entropic chain behavior is only minor, again emphasizing the high rigidity of this proline-rich disordered protein. A consequence is an only slight burial of hydrophobic residues by an area as small as 0.1 nm^2 at extensions-per-bond < 0.15 nm, except for K11, which shows a pronounced decrease of 0.25 nm^2 at an extension-per-bond of 0.185 nm. The marginally stronger coiling propensity of PPAK with respect to its entropic chain model at high extensions is caused by a higher PPII content (37% in full MD simulations versus 25% in entropic chain simulations at an extension-per-bond of 0.30 nm), stressing the role of PPII structure formation for the coiling of proline-rich peptides.

Silk exhibits a force-extension curve that significantly diverges from its entropic chain behavior at low extensions more strongly than its counterparts PPAK and FG, but without featuring the eye-striking hump of the related glycine-rich resilin peptide. With a burial of surface area of 0.15 nm^2 at an extension-per-bond of 0.15 nm (Fig. 4 D), it is the least hydrated among the four disordered proteins, due to its strong coiling propensity. Interestingly, this proline-rich *Araneus* MA silk has previously been shown to be even more hydrated than the proline-deficient *Nephila* MA silk (38). Apparently, although prolines add to the hydrophobicity and thus to the surface burial of silk, they here do not play a significant role in enhancing the PPII content (Fig. 2 A).

We also asked whether these four archetypal disordered peptides distinguish themselves from peptides which are part of well-folded globular proteins. To this end, we determined the force-extension curves of two fragments of ubiquitin, an α -helix and a β -turn, which are shown in Fig. S2. Although their entropic chain elasticity is highly similar to the FG repeat, as expected given their similarly low Gly/Pro content, they overall collapse more strongly starting already at extensions as high as 0.3 nm, supposedly due to the higher hydrophobic forces typical for the core of folded proteins like ubiquitin.

CONCLUSIONS

In general, the main driving force for the initial collapse of these disordered proteins is of entropic nature, which is largely defined by the glycine and PPII content, and less by the proline content. Nonlocal electrostatic and hydrophobic forces are setting in at lower end-to-end distances. Their impact on the coiling propensity again depends on glycines and prolines in the chain, with glycine allowing

and proline impeding formation of a compact core of tight side chain-side chain packings. As a consequence, proline renders PPAK rigid, whereas glycine renders silk and resilin elastic. FG forms a compact hydrophobic core comprising the first two-thirds of the chain. However, in comparison, resilin and silk can more strongly coil due to the extended glycine chain.

Recently, an empirical relation to predict the compaction of disordered proteins from its sequence has been put forward (39). The authors propose a correlation of the hydrodynamic radius with the protein's net charge and its proline content, and no significant correlation with the glycine content, in contrast to our findings. However, the dependencies showed strong scattering so that divergences from the correlations can be expected to occur frequently for individual disordered proteins. We tested whether the resulting relation for the hydrodynamic radius (Eq. 6 of Marsh and Forman-Kay (39)) is able to predict the relative coiling propensity observed for the four natural disordered proteins we have investigated in the MD simulations.

The sequence-based relation predicts a coiling propensity in the order

$$\text{FG} > \text{resilin} > \text{silk} > \text{PPAK},$$

which directly reflects the correlation with their proline content (net charges for our four proteins vary only between 0 and 1 and do not influence the prediction). However, we find an order

$$\text{silk} > \text{resilin} > \text{FG} > \text{PPAK},$$

as measured from the integral of the force-extension curves up to intermediate extensions (Fig. 3). In our case, despite the considerable proline content of resilin and silk of 15% and 25%, respectively, both disordered proteins show a strong tendency to coil. Apparently, the nonadditive impact of prolines and glycines in their respective sequence distribution and environment, resulting in varying degrees of PPII and hairpin formation, gives rise to the generally observed divergence from simple additivity.

SUPPORTING MATERIAL

Two figures are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(10\)01259-2](http://www.biophysj.org/biophysj/supplemental/S0006-3495(10)01259-2)

We thank Scott A. Edwards for helpful discussions and for carefully reading the manuscript.

REFERENCES

1. Rauscher, S., S. Baud, ..., R. Pomès. 2006. Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. *Structure*. 14:1667–1676.
2. Dunker, A. K., C. J. Oldfield, ..., V. Uversk. 2008. The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics*. 9:S1.

3. Boichicchio, B., A. Pepe, and A. M. Tamburro. 2008. Investigating by CD the molecular mechanism of elasticity of elastomeric proteins. *Chirality*. 20:985–994.
4. Dicko, C., D. Porter, ..., F. Vollrath. 2008. Structural disorder in silk proteins reveals the emergence of elastomericity. *Biomacromolecules*. 9:216–221.
5. Lim, R. Y. H., N. P. Huang, ..., U. Aebi. 2006. Flexible phenylalanine-glycine nucleoporins as entropic barriers to nucleocytoplasmic transport. *Proc. Natl. Acad. Sci. USA*. 103:9512–9517.
6. Walther, K. A., F. Gräter, ..., J. M. Fernandez. 2007. Signatures of hydrophobic collapse in extended proteins captured with force spectroscopy. *Proc. Natl. Acad. Sci. USA*. 104:7916–7921.
7. Gräter, F., P. Heider, ..., B. J. Berne. 2008. Dissecting entropic coiling and poor solvent effects in protein collapse. *J. Am. Chem. Soc.* 130:11578–11579.
8. Bustamante, C., J. F. Marko, ..., S. Smith. 1994. Entropic elasticity of λ -phage DNA. *Science*. 265:1599–1600.
9. Dölker, N., U. Zachariae, and H. Grubmüller. 2010. Hydrophilic linkers and polar contacts affect aggregation of FG repeat peptides. *Biophys. J.* 98:2653–2661.
10. Frey, S., R. P. Richter, and D. Görlich. 2006. FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science*. 314:815–817.
11. Elad, N., T. Maimon, ..., O. Medalia. 2009. Structural analysis of the nuclear pore complex by integrated approaches. *Curr. Opin. Struct. Biol.* 19:226–232.
12. Miao, L., and K. Schulten. 2009. Transport-related structures and processes of the nuclear pore complex studied through molecular dynamics. *Structure*. 17:449–459.
13. Elvin, C. M., A. G. Carr, ..., N. E. Dixon. 2005. Synthesis and properties of crosslinked recombinant pro-resilin. *Nature*. 437:999–1002.
14. Nairn, K. M., R. E. Lyons, ..., C. M. Elvin. 2008. A synthetic resilin is largely unstructured. *Biophys. J.* 95:3358–3365.
15. Ardell, D. H., and S. O. Andersen. 2001. Tentative identification of a resilin gene in *Drosophila melanogaster*. *Insect Biochem. Mol. Biol.* 31:965–970.
16. Li, H., A. F. Oberhauser, ..., J. M. Fernandez. 2001. Multiple conformations of PEVK proteins detected by single-molecule techniques. *Proc. Natl. Acad. Sci. USA*. 98:10682–10686.
17. Sarkar, A., S. Caamano, and J. M. Fernandez. 2005. The elasticity of individual titin PEVK exons measured by single molecule atomic force microscopy. *J. Biol. Chem.* 280:6261–6264.
18. Greaser, M. 2001. Identification of new repeating motifs in titin. *Proteins*. 43:145–149.
19. Gosline, J. M., P. A. Guerette, ..., K. N. Savage. 1999. The mechanical design of spider silks: from fibroin sequence to mechanical function. *J. Exp. Biol.* 202:3295–3303.
20. Oroudjev, E., J. Soares, ..., H. G. Hansma. 2002. Segmented nanofibers of spider dragline silk: atomic force microscopy and single-molecule force spectroscopy. *Proc. Natl. Acad. Sci. USA*. 99 (Suppl 2):6460–6465.
21. van der Spoel, D., E. Lindahl, ..., H. J. Berendsen. 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.
22. Jorgensen, W. L., and J. Tirado-Rives. 1988. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 110:1657–1666.
23. Jorgensen, W. L., J. Chandrasekhar, and J. D. Madura. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
24. van Gunsteren, W. F., S. R. Billeter, ..., I. G. Tironi. 1996. Biomolecular Simulation: The GROMOS96 Manual and User Guide. Hochschulverlag AG an der ETH, Zürich, Switzerland.
25. Oostenbrink, C., T. A. Soares, ..., W. F. van Gunsteren. 2005. Validation of the 53A6 GROMOS force field. *Eur. Biophys. J.* 34:273–284.
26. Bernard, P., editor. 1981. Intermolecular Forces. Reidel, Dordrecht, The Netherlands.
27. Geerke, D. P., and W. F. van Gunsteren. 2006. Force field evaluation for biomolecular simulation: free enthalpies of solvation of polar and apolar compounds in various solvents. *ChemPhysChem*. 7:671–678.
28. Oostenbrink, C., A. Villa, ..., W. F. van Gunsteren. 2004. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* 25:1656–1676.
29. MacCallum, J. L., and D. P. Tieleman. 2003. Calculation of the water-cyclohexane transfer free energies of neutral amino acid side-chain analogs using the OPLS all-atom force field. *J. Comput. Chem.* 24:1930–1935.
30. Kumar, S., J. M. Rosenberg, ..., P. A. Kollman. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13:1011–1021.
31. Weeks, J. D., D. Chandler, and H. C. Andersen. 1971. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.* 54:5237–5247.
32. Liu, Y., A. Sponner, ..., F. Vollrath. 2008. Proline and processing of spider silks. *Biomacromolecules*. 9:116–121.
33. Savage, K. N., and J. M. Gosline. 2008. The role of proline in the elastic mechanism of hydrated spider silks. *J. Exp. Biol.* 211:1948–1957.
34. Boichicchio, B., and A. M. Tamburro. 2002. Polyproline II structure in proteins: identification by chiroptical spectroscopies, stability, and functions. *Chirality*. 14:782–792.
35. Zagrovic, B., J. Lipfert, ..., V. S. Pande. 2005. Unusual compactness of a polyproline type II structure. *Proc. Natl. Acad. Sci. USA*. 102:11698–11703.
36. Rucker, A. L., and T. P. Creamer. 2002. Polyproline II helical structure in protein unfolded states: lysine peptides revisited. *Protein Sci.* 11:980–985.
37. Krieger, F., A. Möglich, and T. Kiefhaber. 2005. Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. *J. Am. Chem. Soc.* 127:3346–3352.
38. Savage, K. N., and J. M. Gosline. 2008. The effect of proline on the network structure of major ampullate silks as inferred from their mechanical and optical properties. *J. Exp. Biol.* 211:1937–1947.
39. Marsh, J. A., and J. D. Forman-Kay. 2010. Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* 98:2383–2390.