Statistical Analysis of Single-Molecule AFM Force Spectroscopy Curves

Alexei Valiaev[1], Stefan Zauscher[1], and Scott C. Schmidler[2*]

[1]Department of Mechanical Engineering and Materials Science, and
Center for Biologically Inspired Materials and Materials Systems
144 Hudson Hall, Box 90300
Duke University, Durham, NC, USA

[2]Department of Statistical Science, and
Program in Structural Biology and Biophysics
223 Old Chemistry Bldg, Box 90251
Duke University, Durham, NC, USA

Abstract

We develop a statistical framework for the quantitative characterization and analysis of force-extension curves obtained from single-molecule force spectroscopy (SMFS) measurements. We apply this methodology to force-extension data obtained for elastin-like polypeptides (ELPs) with precisely engineered molecular architectures, where we demonstrate that our approach enables SMFS to be used to study hydrophobic hydration in intrinsically unstructured biomacromolecules. Our results obtained for ELPs suggest that hydrophobic hydration, rather than local backbone conformational entropy, is the key contributor to modulating the molecular elasticity of ELPs under changes in amino acid sequence.

As with previous analysis, we parametrize SMFS curves using models from polymer statistical mechanics; however, we introduce several statistical innovations that dramatically improve the precision of the estimated parameters. Our approach (i) accounts for increased thermal noise in the data at low forces, (ii) provides confidence intervals for fitted polymer-theory parameters obtained from nonparametric bootstrapping, (iii) treats multiple curves simultaneously to reduce variability in the fitted parameters, and (iv) treats noise in both force and extension measurements simultaneously. A key advantage of our approach is the ability to quantify uncertainty in the fitted parameters, allowing comparison of parameters obtained under different experimental

* *Corresponding author:* Scott C. Schmidler, Department of Statistical Science, Duke University, Durham, NC 27708-0251. Tel: (919) 684-8064; Fax: (919) 684-8594; Email: schmidler@stat.duke.edu

conditions or for structural variants. Our approach is able to distinguish previously unresolvable small differences in the molecular architecture of ELPs from SMFS experiments, resolving differences in Kuhn segment lengths as small as 0.01nm, significantly smaller than previously possible. This approach is universally applicable to SMFS data and of general interest for the analysis of structure-property relations in polymers and biomacromolecules.

Running title: Statistical Analysis of AFM Curves
Keywords: atomic force microscopy, freely-jointed chain, worm-like chain, bootstrapping, Bayesian analysis, elastin-like polypeptides

1 Introduction

Over the last decade, atomic force microscope (AFM) force spectroscopy and optical tweezers have been increasingly used to study structure and conformation, inter- and intra-molecular interactions, and mechanical properties of various biological and synthetic macromolecules (1–12). Advances in both the experimental (3, 11, 13, 14) and theoretical treatments (3, 15, 16) have been applied to study DNA (14, 17–19), proteins (6, 7, 9, 20) and other macromolecules (5, 13, 21) on molecular length scales.

When individual globular proteins or DNA are stretched with an AFM, characteristic force-extension "fingerprints" are often observed that arise from force-induced changes in the secondary and tertiary conformation of the molecule (7–9, 14, 22). However, many macromolecules do not exhibit such characteristic fingerprints, and behave instead like random polymer chains. The force-extension curves of such molecules are often well described by statistical mechanical polymer elasticity models (3, 10, 11). The most commonly used models for this purpose are the freely jointed chain (FJC) model (23), the worm-like chain (WLC) model (15, 17), and their modifications (18). These models have been applied successfully to fit and interpret the force-extension behavior of various biological and synthetic polymers (2–7, 10, 13, 24, 25). Although polymer elasticity models do not provide atomic-level information, they can generate valuable insight onto mechanical properties and structural variants of single molecules (2, 3, 10, 17). Alternatives to polymer elasticity models include the use of molecular dynamics (MD) simulations under applied mechanical force, which have been used to study detailed molecular rearrangements in proteins, polysaccharides, and DNA (9, 26–28), and the use of simplified kinetic models and Monte Carlo simulations (6, 29, 30).

Despite significant recent work on theoretical aspects of polymer elasticity models (16, 31–33), there has been little discussion about the statistical methodology used for obtaining and interpreting model parameters from AFM force spectroscopy measurements. In particular, little has been published regarding the accuracy, reproducibility, and uncertainty of polymer elasticity model parameters obtained from fitting to experimental data.

In this paper we present a set of statistical procedures and tools for the analysis of AFM force-extension curves and their fitting with polymer elasticity models. As with previous studies, we parameterize SMFS curves using models from polymer statistical mechanics; however, we introduce several statistical innovations that dramatically narrow the resulting distribution of fitted polymer-theory parameters such as the Kuhn segment length. Our approach (i) accounts for increased thermal noise in the data at low forces, (ii) provides

confidence intervals for the fitted polymer-theory parameters obtained from nonparametric bootstrapping, (iii) treats multiple curves simultaneously to reduce variability in the fitted parameters, and (iv) treats noise in both force and extension measurements simultaneously. A key advantage of our approach is the ability to quantify uncertainty in the fitted parameters, allowing comparison of parameters obtained under different experimental conditions or for structural variants.

We first demonstrate the applicability of our approach by analyzing the force extension behavior of poly(ethylene glycol) (PEG) in hexadecane. As has been shown previously, the conformational behavior of PEG in hexadecane is well described by a FJC model (11). We then demonstrate the power of our approach for biophysical studies by showing that SMFS, when combined with our approach for data analysis, can be used to study hydrophobic hydration in intrinsically unstructured biomacromolecules. We perform SMFS to study the conformational mechanics of stimulus-responsive elastin-like polypeptides (ELPs) (34, 35), using our data analysis approach to quantify the effects of solvent condition and guest residue substitutions by comparing Kuhn segment lengths obtained by fitting a FJC model. ELPs are well suited to our approach, because (1) the primary structure of ELPs can be precisely controlled and easily modified by genetic engineering methods (35, 36), (2) the force-extension behavior follows closely that of an intrinsically disordered protein and is thus amenable to fits with the FJC model, and (3) ELPs are stimulus-responsive biomacromolecules that undergo subtle changes in their conformational mechanics when solvent conditions are changed. Our indicate that hydrophobic hydration, rather than local backbone conformational entropy, modulates the molecular elasticity of ELPs under changes in amino acid sequence.

Our approach is able to resolve differences in Kuhn segment lengths as small as 0.01nm. Such precision allows the study of subtle conformational and structural differences in biological or synthetic macromolecules that cannot be resolved by visual inspection of the force-extension behavior (2, 8, 11) or by conventional data analysis techniques (1, 2, 10, 13, 21, 24).

These results demonstrate that SMFS, when combined with our approach for data analysis, can be used to study the subtleties of polypeptide-water interactions and thus provides a basis for the study of hydrophobic hydration in intrinsically unstructured biomacromolecules. Although we focus primarily on fitting the FJC model, our approach is directly applicable to the WLC model and other polymer elasticity models.

2 Materials and Methods

ELPs were synthesized in the laboratory of Dr. Ashutosh Chilkoti (Department of Biomedical Engineering, Duke University) using methods described previously (34, 36). Three ELP libraries (Figure 1) were used in this study, all contain Val-Pro-Gly-Xaa-Gly (VPGXG) (X is a guest residue) pentapeptide repeats flanked by a leader (Ser-Lys-Gly-Pro-Gly) and a trailer (Trp-Pro). ELP1-180 contains Val, Ala, and Gly at the guest residue positions in a 5:2:3 ratio and consist of 180 pentapeptides with a total molecular weight of 71.9 kDa. ELP4-120 contains Val at all guest positions and has a molecular weight of 50 kDa. ELP2-4 is a block copolymer of 50 kDa (ELP2-60/ELP4-64) which includes an ELP2 containing Val, Gly, Ala guest residues in a 1:7:8 ratio, and an ELP4 block containing only Val guest residues. Polyethylene glycol (PEG) with a molecular weight of 35 kDa (Sigma Chemical

Co., cat # 94646), and hexadecane (Sigma Chemical Co., cat # H0255) were purchased from SigmaAldrich.

Sample Preparation ELPs were covalently attached to functionalized gold surfaces that were prepared by vapor deposition of a 10nm chromium adhesion layer on to a glass substrate followed by vapor deposition of 100nm gold. Before deposition, the glass surfaces were cleaned for 20 min in a 1:3 (v/v) solution of H_2O_2/H_2SO_4 (Piranha) at 80°C (Caution: Piranha solution reacts violently with organic matter!). To minimize unspecific interactions between ELPs and the gold surface we used a mixed SAM of oligoethylene-glycol terminated alkanethiols (Prochimia, cat.#: TH 011-01). In this mixture CH_3 terminated EG_3 thiols provide a nonfouling background for the ELPs whereas COOH terminated EG_6 provide chemical functionality to graft ELPs via amine coupling. We chose a ratio of 5% EG_6 and 95% EG_3 which provided sufficiently low grafting densities for single-molecule measurements. The COOH groups of the EG_6 thiols were reacted for 30 minutes with 1-ethyl-3-(dimethylamino) propyl carbodiimide (EDAC) (0.4 M, Aldrich) and N-hydroxysuccinimide (NHS) (0.1 M, Aldrich) in Milli-Q^{TM}grade water. Prior to incubation with ELPs, the substrates were rinsed with ethanol and water and dried in a stream of N_2 gas. Next, a drop of the desired ELP solution (5μM in PBS buffer) was placed on the functionalized gold surface for 2-3 hours in a sealed Petri dish. After the incubation step, the samples were thoroughly washed with Milli-Q grade water. Polyethylene glycol (PEG) was dissolved in Milli-Q water to a concentration of 10μM and incubated on a bare gold surface for 2 hours.

Force Spectroscopy AFM force spectroscopy experiments were performed with a MultiMode AFM with Nanoscope IIIa controller (Veeco, Digital Instruments) using a fluid cell attachment. The sensitivity of the photodetector was determined from the constant compliance regime upon approach at large applied normal forces. Furthermore, force-distance curves were converted into force-extension curves by accounting for the effect of cantilever bending. A constant pulling rate of 1 μm/s was maintained throughout all experiments. However, when a tethered molecule is stretched at a constant pulling rate, the microcantilever bends and thus the actual tip velocity, relative to the substrate surface, will change. This results in hydrodynamic drag forces that act on the cantilever during force-extension experiments. We accounted for this additional force component by calculating the hydrodynamic drag force on the cantilever (see Hydrodynamic Drag Subtraction in Section 4).Rectangular Si_3N_4 cantilevers (TM Microscopes) were used and their spring constants (typically 20-25pN/nm) were estimated before the experiments from the power spectral density of the thermal noise fluctuations (37).

3 Polymer Elasticity Models

To describe force-extension curves we use parametric elasticity models derived from polymer statistical mechanics. Such models have been extensively used in the literature (2, 3, 17, 18). The most commonly used models for this purpose are the *freely jointed chain* (FJC), the *worm-like chain* (WLC), and their extensions.

Freely-jointed chain (FJC) and extended FJC When stretched, many macromolecules exhibit 'random-coil' behavior described by the random-walk statistics of the freely-jointed chain (FJC) model (18, 23). The FJC model represents a polymer chain by n rigid segments of length l_K connected by freely-rotating joints with no long range interactions (2, 5), and yields elasticity law:

$$x(F) = L\left(\coth\left(\frac{Fl_K}{k_BT}\right) - \frac{k_BT}{Fl_K}\right) = L\mathcal{L}(\beta) \tag{1}$$

where $\beta = \frac{Fl_K}{k_BT}$ and $\mathcal{L}(\beta) = \coth(\beta) - \frac{1}{\beta}$ is the Langevin function. Here $L = nl_K$ is the contour length, l_K is the Kuhn segment length, k_B is Boltzmann's constant and T is absolute temperature. The elastic behavior of many macromolecules has been shown to be well approximated by the FJC model; examples include polydimethylsiloxane in heptane (3), poly(methacrylic acid) (10), and polyethylene glycol (PEG) in hexadecane (11). Deviations from the predictions of the FJC model can often be explained by higher-order structure arising from interactions among chain segments or with solvent molecules (3, 11).

With increasingly larger applied forces and extensions, the molecular response becomes increasingly enthalpic as the polymer backbone is stretched and bond-angles are deformed. The FJC model assumes that the dependence of force upon extension is purely entropic, up to a maximum extension given by contour length L. Discrepancies between the model and empirical observations at high forces led to the introduction of the *extended freely jointed chain (EFJC)* to account for enthalpic contributions such as backbone deformation (18). The EFJC model incorporates an additional "segmental stiffness" parameter λ:

$$x_{Ext}(F) = x(F)\left(1 - \frac{F}{\lambda l_K}\right) \tag{2}$$

where λ describes the elasticity of an individual segment when stretched. Differences between the FJC and EFJC models for describing experimental data are demonstrated in Section 4.

Worm-like chain and extended WLC An alternative polymer representation is given by the linear bending elasticity of a thin homogeneous rod described by the Kratky-Porod or wormlike chain (WLC) model (15, 17, 38). The WLC can be obtained as a limit of a freely-rotating chain as $l_K \to 0$ and $n \to \infty$ simultaneously such that $nl_K = L$. The WLC is parameterized by its contour length L and the *persistence length* l_p, which is the exponential decay rate of the autocorrelation function. The energy of a stretched WLC is given (38, 39) by the line integral

$$E = k_BT\int_0^L \frac{l_p}{2}\kappa^2 \mathrm{d}s - fd$$

where $\kappa = \left|\frac{\partial^2 t(s)}{\partial s^2}\right|$ is the curvature, $t(s)$ is a unit tangent vector, and s the arc length. The elasticity law cannot be obtained analytically but is given by the solution of a variational problem, and an approximation which is valid for both small- and large-force limits was given by (17):

$$F(x) = \frac{k_BT}{l_p}\left(\frac{1}{4(1-\frac{x}{L})^2} - \frac{1}{4} + \frac{x}{L}\right) \tag{3}$$

This analytical approximation is often used to fit the WLC model to experimental force-extension curves; an additional series expansion term is given by (15). The WLC model has been applied to modeling force extension behavior for proteins (6, 7) and especially DNA (17, 19), where in some cases the WLC model works well up to forces of several hundred pN (3). To include stiffness of the chain, a modified WLC chain with "elastic modulus" parameter Φ can be used (40):

$$F(x) = \frac{k_B T}{l_p}\left(\frac{1}{4(1-\frac{x}{L}+\frac{F(x)}{\Phi})^2} - \frac{F(x)}{\Phi} - \frac{1}{4} + \frac{x}{L}\right) \tag{4}$$

Although the polymer elasticity models described above do not explicitly account for the detailed molecular composition of a polymer (*e.g.,* polypeptide sequence), the molecular elastic response and therefore the fitted parameters depend on molecular composition.

4 Data selection and preprocessing

In force spectroscopy experiments we collected several thousand force-extension curves for each experimental condition. We denote the set of curves by $(\mathbf{d}, \mathbf{f}) = ((\mathbf{d}_1, \mathbf{f}_1), \ldots, (\mathbf{d}_m, \mathbf{f}_m))$ where $\mathbf{d}_j = (d_{1j}, \ldots, d_{n_j j})$ and $\mathbf{f}_j = (f_{1j}, \ldots, f_{n_j j})$ denote the finitely-sampled measurements of the j^{th} curve, with d_{ij} the separation distance measured at applied force f_{ij}.

Only a fraction of the measured curves, however, represent the actual single-molecule force-extension event of interest. Many exhibit artifacts such as multiple force-extension events, simultaneous extension of more than one molecule, or large nonspecific adhesion. A custom built Matlab program was developed to automatically filter the data based on several criteria (see below) established to reject curves that exhibit these effects.

Smoothing Due to significant noise (in the order of $\pm 20pN$) arising from cantilever thermal fluctuations (see Section 5.1) we smoothed individual force-extension curves using a local linear regression smoother (loess) (41):

$$\hat{\mathbf{f}}_j(d) = \mathbf{d}_j(\mathbf{d}_j^T \mathbf{W}_d \mathbf{d}_j)^{-1} \mathbf{d}_j^T \mathbf{W}_d \mathbf{f}_j$$

where $\mathbf{W}_d = \mathrm{diag}(w(\mathbf{d_j} - d\mathbf{1}))$ and $w(x) \propto (1-|x|^3)^3$ for x lying in the k-nearest neighborhood of d and $|x| \leq 1$, and 0 otherwise, with $k = 30$ chosen to give good performance.

Rupture peak identification First, all rupture events in a curve were identified using the criteria:

$$(\Delta \mathbf{f}_j)_i > F_r$$

where $(\Delta \mathbf{f}_j)_i = (f_{ij} - f_{ij-1})/(d_{ij} - d_{ij-1})$ and F_r is a constant (25pN) chosen to identify sudden force decreases corresponding to rupture events. Let f_j^* denote the maximum force f_i obtained at this rupture event for curve j, and d_j^* the associated separation distance. Curves with more than one such event indicate multiple probe attachments and were discarded.

Because of unspecific and adhesion interactions some of the force curves exhibited a significant force offset at small and intermediate extensions. These offsets represent noticeable

deviations from the FJC model and significantly affect the fitted parameters. To remedy this, curves with an offset greater than 60pN were eliminated from further analysis, as were curves with too low of a force threshold. These filtering criteria impose the force threshold requirements:

$$\min_{i:d_{ij}<d_j^*} \hat{f}_{ij} < 60\text{pN} \qquad \text{and} \qquad f_j^* \geq 200\text{pN}$$

Normalization test Because an AFM cantilever tip attaches to a surface-tethered molecule at a random location along the backbone, the contour (maximum extension) length varies across pulls. The Kuhn segment length, however, is an intrinsic property which should be independent of the attachment length. This suggests that the force extension behavior of a molecule can be described by a canonical force extension curve obtained by normalization of the curve for an individual pull with respect to attachment length. After normalizing individual curves by the extension at a fixed force (200pN), normalized curves which superimpose closely were assumed to represent a valid single-attachment pull (3). Outlier curves (which cannot be normalized with other curves obtained under the same experimental conditions) were identified and removed as follows:

i) Smoothed curves $(\mathbf{d}_j, \hat{\mathbf{f}}_j)$ were normalized to unit distance by division by contour length at 200pN ($\frac{\mathbf{d}_j}{d^{200}}$) (Figure 2a).

ii) An average normalized curve was estimated robustly using a pointwise median curve:

$$\bar{f}(\alpha) = \operatorname*{median}_j \hat{f}_j(\alpha d_j^{200}) \qquad \alpha \in (0,1)$$

Note: $\hat{f}_j$'s and thus $\bar{f}$ are defined everywhere on $[0,1]$, not just at observed points d_{ij}.

iii) A robust estimate of curve variance was obtained as the median absolute deviation:

$$\hat{\sigma}_f(d) = c \operatorname*{median}_j \left| f_j(d) - \bar{f}(d) \right|$$

and curves were rejected which deviated more than $\pm 2\sigma_f(d)$ from the median curve (Figure 2b):

$$\max_i \left| f_{ij} - \bar{f}(d_{ij}) \right| \geq 2\hat{\sigma}_f(d)$$

The remaining set of curves that satisfy all of the above criteria were then used for further analysis.

Hydrodynamic Drag Subtraction The pulling rates used in force spectroscopy experiments (11, 22, 42) are often sufficient to produce hydrodynamic drag on the cantilever. To obtain more accurate force estimates this drag can be estimated and subtracted from force measurements. Since the flow velocities are generally small (viscous flow regime), the drag force F_H can be described by

$$F_H = C\mu V \tag{5}$$

where C is a constant dependent on cantilever geometry, μ is the viscosity of the liquid, and V is the velocity of the cantilever relative to the surrounding fluid phase.

To estimate the hydrodynamic drag contribution associated with a particular cantilever type we performed force measurements at different sample displacement rates. As expected, the force values obtained from cantilever deflection measurements depended linearly on pulling speed (Figure 3). The coefficient $C\mu$ can then be determined by linear regression.

In a typical single molecule pulling experiment performed at constant displacement rate, the relative velocity between cantilever and surrounding fluid changes as the molecule is stretched. To account for this, we estimated the relative cantilever velocity by differentiating the measured, smoothed force-separation data and then subtracted the drag force (Eq. 5) at every point along the measured force-extension curve

$$f_i = F_i - \hat{C}\mu\hat{v}_i \tag{6}$$

where F_i is the measured force and $\hat{v}_i = (d_{i+1} - d_{i-1})/2\Delta t$ is the estimated cantilever velocity. Figure 4 shows force-extension curves obtained for ELP1-180 in PBS buffer solution before and after the drag force correction.

Force Window Selection Values of the fit parameters such as Kuhn length can vary significantly depending on the force range chosen for fitting (2, 3, 13). To remove this ambiguity we choose a force window between 60 pN and 200 pN for all curves when fitting the FJC model to a force extension profile. These force values were chosen to maximize the range of the force region in which all of the selected curves were well described by the FJC (Eq. 1). Extending the force range to a larger force results in the deviation from the FJC fit at low and large force ranges (Figure 5).

Figure 6 compares the fit of the FJC and EFJC to the same experimental force-extension curve, where it can be seen that the two models yield very similar fits to the data in the [60-200] pN force window. We therefore chose to work with the FJC model to remove the unnecessary additional parameter of the EFJC.

5 Data Reduction and Modeling

5.1 Statistical model fitting

We now turn to the estimation of polymer elasticity model parameters. Given an experimentally measured force-extension curve, we wish to determine the parameters of the polymer elasticity model which best fit the observed data. Since our approach is applicable for any of the various polymer models, we use the notation $g(f;\theta)$ for the elasticity law of a generic polymer model with parameter vector θ, except where it is necessary to specify the particular model in use. (For the WLC, $g^{-1}(d;\theta)$ is used instead.) Thus the model fitting methods described in this section are applicable to *all* the polymer models discussed in Section 3, with appropriate choice of θ given in Table 1. For example, in the FJC model used in most of our examples, $\theta = (l_K, L)$ will be a vector containing Kuhn and contour lengths.To simplify notation in this section we suppress the curve index and simply denote an arbitrary curve by $(\mathbf{d}, \mathbf{f}) = (d_i, f_i)_{i=1}^n$.

The commonly used approach to fitting such models is *least squares* (LS) error minimization:

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} (d_i - g(f_i;\theta))^2$$

Since $g()$ is nonlinear in θ, this yields a nonlinear least-squares (NLS) problem which must be solved numerically (see Section 7).

While $\hat{\theta}$ provides a point estimate of the parameters, for purposes of comparing parameters obtained from different curves or molecules it is critically important to account for uncertainty $var(\hat{\theta})$ in these fitted parameters. This can be done by adopting a statistical model for the measurement errors, in order to obtain confidence intervals or perform significance tests. Under certain assumptions ($d_i = g(f_i;\theta) + \epsilon_i$ with ϵ_i independent normally distributed errors with constant variance), LS fitting gives *maximum likelihood* estimates of the parameters θ (see e.g. (43)), and under such assumptions the uncertainty in the estimated parameters $\hat{\theta}$ may be quantified by estimating the noise variance σ^2 from the residuals $r_i = d_i - g(f_i;\hat{\theta})$:

$$\hat{\sigma}^2 = \frac{1}{(n-p^*)} \sum_{i=1}^{n} (d_i - g(f_i;\hat{\theta}))^2$$

where p^* is the number of parameters being fit, approximately $(m+1)$ for the FJC model and $(2m+1)$ for the EFJC, and using $\hat{\sigma}^2$ to obtain confidence intervals for the parameters.

Unequal Variances and Weighted Least Squares However, a plot of the residuals (Fig. 7) from fitting typical force-extension curves shows a clear violation of the constant variance assumption. This is reasonable as observations at low forces have significantly larger contributions from thermal fluctuations of the end-to-end distances of the polymer chain than those at higher forces, so that σ^2 decreases as a function $\sigma^2(f)$ of f. When σ^2 is not constant, maximum likelihood estimates are obtained by a *weighted* least-squares (WLS) minimization:

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} w_i \left(d_i - g(f_i;\theta)\right)^2 - \log w_i \tag{7}$$

where $w_i = \sigma^2(f_i)^{-1}$. To calculate the weights we need the variance function $\sigma^2(f)$ providing the dependence on f.

The FJC elasticity law (Eq. 1) provides the mean (ensemble average) separation distance d under applied force f. The full distribution of d may be obtained by noting that under the FJC model

$$d = \sum_{i=1}^{n_K} \mathbf{x}_i^T \mathbf{f}$$

is the projection onto the unit force vector $\mathbf{f}$ of a sum of independent random vectors with *Langevin* or *von Mises-Fisher* distribution:

$$f(\mathbf{x};\mathbf{f},\beta) = \left(\frac{\beta}{2}\right)^{\frac{1}{2}} \frac{\mathrm{e}^{\frac{\beta}{l_K}\mathbf{x}^T\mathbf{f}}}{l_K \Gamma(\frac{3}{2}) I_{\frac{1}{2}}(\beta)}$$

where $\|\mathbf{x}\| = l_K$, $\|\mathbf{f}\| = 1$, and β is defined by (Eq. 1). Here I_ν is the modified Bessel function of the first kind and order ν. The marginal distribution of $d_i = \mathbf{x}_i^T \mathbf{f}$ is given by (44):

$$f(d_i;\beta) = \left(\frac{\beta}{2}\right)^{\frac{1}{2}} \frac{\mathrm{e}^{\beta \frac{d_i}{l_K}}}{l_K \Gamma(\frac{1}{2}) I_1(\beta)} \qquad d_i \in [-l_K, l_K]$$

so that $d = \sum_i d_i$ has mean (Eq. 1) and for large n_K has asymptotic variance:

$$V_d(f) \;=\; x'(f) \;=\; n_K l_K^2 (1 - \mathcal{L}(\beta)^2 - \frac{2}{\beta}\mathcal{L}(\beta)) \tag{8}$$

Thus we take $\sigma^2(f) = \hat{\sigma}^2 V_d(f)$ where $\hat{\sigma}^2$ is a scale parameter to be estimated:

$$\hat{\sigma}^2 = \frac{1}{(n-p)} \sum_{i=1}^{n} V_d(f_i)^{-1} \left(d_i - g(f_i; \hat{l}_K, \hat{L})\right)^2$$

Estimated values for $\hat{\sigma}$ were around 3, on the order of experimental noise. Figure 7 shows that this theoretical prediction of the force-dependence of fluctuations describes the observed deviations from the FJC model well.

Nonlinear Least Squares Calculations Because the weights $w_i(\theta)$ are functions of other parameters θ, the resulting model is fit by an iteratively-reweighted least-squares procedure (Appendix 7).

Confidence intervals via bootstrapping A critical aspect of comparing parameters obtained under different experimental conditions or from structural variants is to quantify the uncertainty in the estimated parameters. We use a nonparametric bootstrap procedure (45) to obtain confidence intervals for the fitted parameters by Monte Carlo resampling; details are given in Appendix 7.

5.2 Multiple curve fitting

Several force-extension curves are usually recorded during measurements under the same experimental conditions. Because each curve represents the same molecular sequence, each provides additional information for estimating the elasticity parameters of the molecule. Here we wish to pool information from all of the curves to improve our estimate of l_K and reduce the uncertainty $var(\hat{l}_K)$ of the estimated parameters. However, the random location of AFM cantilever tip attachment to the molecule means that each curve has a distinct value of L. We thus need to simultaneously estimate the parameters $\theta = (l_K, L_1, \ldots, L_m)$, where l_K is the *common* Kuhn length parameter.

The maximum likelihood estimates $\hat{\theta}$ are again obtained by minimizing a weighted LS criteria:

$$\hat{\theta} = \arg \min_{l_K, L_1, \ldots, L_m} \sum_{j=1}^{m} \sum_{i=1}^{n_j} w_{ij} \left(d_{ij} - g(f_{ij}; l_K, L_j)\right)^2 - \log \sigma_j^2 w_{ij} \tag{9}$$

with $w_{ij} = V_d(f_i)^{-1}$, using a single shared l_K. Minimization of (Eq. 9) involves $m(p-1)+1$ parameters and again requires iteratively-reweighted least-squares calculations; an efficient algorithm is given in Appendix 7. The noise variance estimate becomes

$$\hat{\sigma}^2 = \frac{1}{(n-p^*)} \sum_{j=1}^{m} \sum_{i=1}^{n_j} V_d(f_{ij})^{-1} \left(d_{ij} - g(f_{ij}; \hat{l}_K, \hat{L}_j) \right)^2$$

where $n = \sum_{j=1}^{m} n_j$ and $p^* = m(p-1)+1$. Confidence intervals for multiple curve analysis are again obtained by a nonparameteric bootstrap procedure (Appendix 7).

5.3 Accounting for Measurement Error in Force and Distance

The statistical fitting procedures described here, like other fitting procedures used in the literature, rely on minimizing deviations between theoretical and measured distances at prescribed forces (Section 5.1), or in the case of the WLC, fitting force at prescribed distances. As described in Section 5.1, this approach implicitly assumes a statistical model of the form $d = g(f;\theta) + \epsilon$ having measurement error in d, with f known. However, in AFM force spectroscopy, measures for both f and d contain random noise due to thermal fluctuations and other uncertainties. Fitting procedures that account for noise in one axis (force or distance) but not both can provide biased parameter estimates for regression models where measurement error is of comparable scale in both axes (46).

A more efficient parameter estimation approach in such cases is to model the random error in both measurements, yielding a model of the form:

$$d_i = g(z_i;\theta) + \epsilon_i \qquad\qquad f_i = z_i + \delta_i \tag{10}$$

which has the interpretation that z_i and $g(z_i)$ are the "true" force and distance respectively, while f_i and d_i are the observed force and distance measured with respective independent random errors ϵ and δ. Models of the form (Eq. 10) can be more realistic and provide better parameter estimates; however, they are significantly more difficult to fit, leading to "total" least squares minimization problems of the form:

$$\hat{\theta} = \arg \min_{z_1,\ldots,z_n,\theta} \sum_{i=1}^{n} w_i(\theta) \frac{(d_i - g(f_i;\theta))^2}{\sigma_\epsilon^2} + \frac{(f_i - z_i)^2}{\sigma_\delta^2} - \log w_i(\theta) \tag{11}$$

which provide maximum likelihood estimates under similar assumptions on ϵ and δ.

Minimization criteria of the form (Eq. 11) are called "total least-squares" problems, and in the case of nonlinear g can be difficult to solve numerically (43, 46). Instead of maximum likelihood, we use a Bayesian approach to parameter estimation (47, 48), by imposing a Gaussian process prior on the $\mathbf{z} = (z_1, \ldots, z_n)$:

$$z_i \mid \mathbf{z}_{j<i} \sim N(z_{i-1}, \sigma_z^2)$$

and standard non- or weakly-informative priors on the elasticity model parameters θ: for the FJC, we use broad uniform priors for the polymer parameters ($l_K \sim U(0,100)$, $L \sim U(0,1000)$) and diffuse gamma $Ga(.1,.1)$ priors for inverse scale parameters σ_ϵ^2,σ_δ^2, and σ_z^2.

Combining these priors with the likelihood obtained from (Eq. 10) yields the posterior distribution over unknowns:

$$
\begin{aligned}
&P(\theta, \mathbf{z}, \sigma_\epsilon^2, \sigma_\delta^2, \sigma_z^2 \mid \mathbf{d}, \mathbf{f}) \\
&\qquad \propto P(\theta, \sigma_\epsilon^2, \sigma_\delta^2, \sigma_z^2) \prod_{i=1}^{n} P(d_i \mid z_i, \theta, \sigma_\epsilon^2) P(f_i \mid z_i, \sigma_\delta^2) P(z_i \mid z_{i-1}, \sigma_z^2) \\
&\qquad \propto w_i^{-n} (\sigma_\epsilon \sigma_\delta \sigma_z)^{-(n+2\alpha-2)} \mathrm{e}^{-\left[\beta(\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_\delta}+\frac{1}{\sigma_z^2})+\frac{1}{2}\left(\sum_{i=1}^{n} w_i \frac{(d_i - g(z_i;\theta))^2}{\sigma_\epsilon^2} + \frac{(f_i - z_i)^2}{\sigma_\delta^2} + \frac{(z_i - z_{i-1})^2}{\sigma_z^2}\right)\right]}
\end{aligned}
$$

and the marginal posterior distribution over parameters θ such as the Kuhn length l_k is given by:

$$
P(l_K \mid \mathbf{d}, \mathbf{f}) = \int \dots \int P(\theta, \mathbf{z}, \sigma_\epsilon^2, \sigma_\delta^2, \sigma_z^2 \mid \mathbf{d}, \mathbf{f}) \, \mathrm{d}\sigma_\epsilon^2 \, \mathrm{d}\sigma_\delta^2 \, \mathrm{d}\sigma_z^2 \, \mathrm{d}\mathbf{z} \, \mathrm{d}L \tag{12}
$$

The $(n+4)$-dimensional integral (12) can be approximated by Monte Carlo integration using Markov chain simulation (49); here we used the Gibbs sampling package WinBugs (50, 51) for ease of implementation, although other MCMC strategies will be more efficient for this problem. The resulting force residuals shown in Figure 8 indicate that the constant variance assumption is adequate in the force dimension.

6 RESULTS

Before turning to ELPs, we first illustrate the approach on a well-studied FJC system, polyethylene glycol (PEG) in hexadecane. PEG in hexadecane is generally accepted to behave as a FJC as shown in previous studies(11). This can be explained by the fact that hexadecane is apolar, hence solvent-mediated supra-molecular structures in PEG molecules are unlikely to form and PEG adopts a random coil conformation in this solvent. Thus PEG serves as an ideal illustrative example system. We applied our data analysis approach to fit a FJC model to force-extension data generated by SMFS studies on PEG as described in Section 2. The resulting representative fit is shown in Figure 9, where the FJC model is seen to fit the data well.

Elucidating the Molecular Elasticity of ELPs

It is of significant scientific and engineering interest to understand the mechanochemical properties of elastin-like polypeptides. ELPs are stimulus-responsive polypeptides(35) consisting of pentapeptide repeats Val-Pro-Gly-X-Gly (VPGXG), where X is any amino acid except Pro (52). ELPs are attractive for a variety of applications requiring molecular level control of polymer structure, as they are genetically encodable and can be synthesized easily by heterologous overexpression from a synthetic gene, with precise control over their composition and chain length (35, 36). Typically SMFS experiments with globular proteins, DNA, and synthetic macromolecules have focused on the analysis of molecular "fingerprints" arising from force-induced changes in secondary and tertiary structures (7, 22, 53) and changes in intra-chain hydrogen bonding (2, 21). However, systematic experiments on the single

molecule level that link differences in the elastic behavior of polypeptides to changes in their hydration behavior have been missing.

Force-induced molecular stretching of polypeptides in SMFS experiments results in a change in the equilibrium conformation of the macromolecule, likely increasing its solvent accessible surface area by exposing previously buried sidechains and contact surfaces, especially for hydrophobic polypeptides like ELPs. The force required to stretch a single ELP molecule in aqueous solvent thus reflects two main components: (1) the restoring force arising from the stretch-induced entropic elasticity of the polypeptide backbone, and (2) the force arising from changes in the solvent-polypeptide interactions.

Here we explore the effect of solvent ionic strength and guest residue substitution on the force-extension behavior (elasticity) of ELPs at large molecular extensions (60-80% of the contour length) in terms of an effective Kuhn segment length. Our results suggest that hydrophobic hydration, rather than local backbone conformational entropy, is the key contributor to modulating the molecular elasticity of ELPs under changes in amino acid sequence. These results also demonstrate that, using our improved-precision data analysis methodology, SMFS can be used to study subtleties of polypeptide-water interactions. Although we focus on ELPs, the approach is applicable to a range of intrinsically unstructured biomacromolecules.

Effects of ionic strength on ELP elasticity Previous studies have shown that hydrophobic collapse of ELPs on surfaces and in solution can be induced isothermally by changing the ionic strength of the medium (35, 54). An increase in ionic strength leads to reduced interaction of solvation shell waters with the ELP side chains exposed upon stretching. Such change in hydrophobic hydration is likely to affect the mechanical properties of ELP molecules in solution.

To test this hypothesis, we performed SMFS on ELP1-180 in PBS and PBS+1.5M NaCl to measure the dependence of the restorative elastic force, and thus Kuhn segment length, on solution ionic strength. However the change in the conformational mechanics of ELPs as a function of solvent condition is subtle, and cannot be resolved by comparison of individual force-extension curves. Figure 10 shows the estimated Kuhn length parameter values and 95% confidence intervals obtained using the single- and multiple-curve fitting procedures of Sections 5.1 and 5.2. We observe an increase in effective Kuhn segment length, indicating that less energy is required to stretch the polypeptide at high ionic strength.

Our results indicate that changes in the solvent conditions affect the apparent stiffness of ELPs and are in good agreement with the two-phase model for elastin elasticity proposed by Weis-Fogh and Anderson (55). However, changes in the elasticity of the ELP molecule due to changes in the hydration of its hydrophobic groups are subtle, as they are dominated by the entropic elasticity of the backbone. Encouraged by our results, we proceeded to explore the effect of hydrophobic hydration on ELP elasticity as a function of the type of guest residue (see below).

Single curve vs. multiple curve fitting: The average width of confidence intervals obtained for individual pulls of ELP in PBS ($\approx$0.04nm) and in PBS + 1.5M NaCl ($\approx$0.045nm) is more than 2.5 times the width of confidence intervals obtained using the multiple-curve

shared parameter model (0.016nm and 0.018nm, respectively). The single curve fitting approach leaves sufficient uncertainty to allow the 95% intervals for curves under the two conditions to overlap, making it impossible to distinguish a statistically significant difference between the two. However the use of multiple curves to estimate a shared parameter yields significantly narrower confidence intervals which do not overlap, allowing straightforward discrimination between average ELP behavior in the different solvents. Thus this experiment also highlights the additional power of biophysical studies using SMFS with our multiple curve-fitting approach compared with the standard single-curve approach.

Hydrophobic hydration of ELPs The interaction of water with hydrophobic groups is important in protein folding/unfolding and profoundly affects their conformational properties. Protein conformation and flexibility are intimately linked to the hydration water structure, in which the hydrogen-bonded hydration water network transmits information around the protein and controls its dynamics. The interaction of water with hydrophobic protein surfaces (hydrophobic hydration) produces a reduction in water density and an increase in heat capacity, both of which are consequences of more ordering in the solvent which also causes a loss of entropy. The effect of hydrophobic hydration on ELP elasticity can be explored by varying the molecular composition of the ELP constructs. Genetic engineering allows the synthesis of ELP constructs with a specific sequence of aliphatic amino acids at the guest residue position. This provides us with molecular constructs of similar molecular weight but significantly different side-group hydrophobicity.

To demonstrate the utility of our approach in providing insight into the subtleties of hydrophobic hydration of elastin-like polypeptides that exhibit different guest residue substitutions, we use the effective Kuhn length as a reporter to study differences in their hydrophobic hydration. We collected SMFS force-extension curves from ELPs with slightly different molecular sequences obtained by the guest residue substitutions shown in Figure 1, and applied our multiple curves fitting approach to their analysis. Figure 11 shows the resulting Kuhn length parameter estimate comparisons for the three ELP constructs with different guest residues.

As described previously, changes in elasticity may be ascribed either to changes in backbone conformational entropy or to changes in hydrophobic hydration. Our choice of ELP constructs allows us to infer which effect is dominant. Primary amino acid sequence largely determines random-coil backbone entropy of unfolded proteins (56). In particular, backbone mobility typically decreases with increasing side chain volume, which would entail a corresponding increase in the Kuhn length. For the ELP constructs studied here, the guest residues of ELP4 (all Val) contain bulkier side groups than the guest residues in ELP1 (50% Val, 20% Ala) or ELP2-4 (52% Val, 26% Ala). This would suggest that flexibility is reduced for the ELP4 backbone and should result in a larger Kuhn length for ELP4 than for ELP2-4 or ELP1.

However, applying our data analysis approach to resolve the associated small differences in Kuhn length, we find instead that the effective Kuhn length scales with the averaged *hydrophobicity* (57, 58). ELP4-120, the most hydrophobic ELP containing only aliphatic Val guest residues, has the shortest Kuhn length of 0.29nm; ELP1-180, the least hydrophobic, has the longest Kuhn length of 0.38nm; and ELP2-4 with intermediate hydrophobicity has

an intermediate Kuhn length of 0.36nm. (Valine has a larger sidechain volume and thus an increased tendency to form ordered (clathrate) water if exposed to solvent, incurring larger entropic penalties when compared to the hydration of alanine or glycine.) Based on hydrophobic hydration, we would expect ELP4 to show the largest free energy loss due to hydrophobic hydration upon stretching, therefore requiring the highest energy required for stretching and having the lowest Kuhn length; while ELP1 has the highest Kuhn length and therefore requires the lowest energy for stretching. These results also agree with the ordering obtained by estimating the hydrophobicity of the distinct pentapeptide constructs, either using standard hydrophobicity scales (59, 60), or by calculating solvent-accessible surface area (ASA) of the nonpolar groups in an extended state from that of the extended tripeptide Gly-X-Gly (58), which yields ASA of 645, 621, and 618Å^2/pentapeptide for ELP4, ELP2-4, and ELP1, respectively. The ordering of ELP elasticities we observe in our SMFS experiments, as quantified by effective Kuhn length, is therefore consistent with the hypothesis that hydrophobic hydration rather than local backbone conformational entropy forms the dominant contribution to modulating the molecular elasticity of ELPs under primary sequence variation.

It should be noted that the various ELP pentapeptides differ in only a single guest position, and therefore the observed differences in effective Kuhn length due to residue substitutions are small. Such effects could not be resolved above the noise level prior to application of the data analysis methodology outlined in this paper. Further studies using our approach to analyze additional effects on ELP elasticity due to solvent quality, temperature, and related quantities will be reported elsewhere (61).

Comparison with Previous Data Analysis Approach Our data analysis approach was motivated by the comparison of fitted Kuhn segment lengths across molecules and experimental conditions as demonstrated above. Initially, we fit an extended FJC model (Kuhn length, contour length, and segmental stiffness) to a set (n = 20 - 40) of force-extension curves which showed a single force pull and collapsed onto one universal curve after normalization by extension at a common, constant force (200 pN). The fit was not constrained by a force window, *i.e.*, all data were considered from approximately zero force to rupture, and fits were performed using standard least-squares according to previously published procedures (1, 2, 10, 21, 24).

The individual pull Kuhn segment lengths obtained with this approach ($l_K = 0.6 \pm 0.3$nm) were consistent with typical Kuhn lengths obtained in force spectroscopy experiments on other polymers (2, 3). However, the large range of fitted parameters obtained across pulls prohibits meaningful comparison of fit results obtained for one type of ELP under a range of experimental conditions (*i.e.*, above and below the transition temperature), and comparison of different structural variants of ELPs (*i.e.*, different guest residues), see below. To reduce this parameter variability we developed and applied our approach as detailed in Section 5. The distribution of Kuhn segment lengths obtained using traditional (old) approach are compared with the bootstrap sampling distribution of the shared Kuhn length obtained from our approach in Figure 12. It should be noted that our approach estimates the Kuhn length for an "average" pull for the given molecule and experimental conditions; through the statistical estimation procedures described in Section 5 this average Kuhn length may

be estimated with significantly reduced variability relative to individual pull Kuhn lengths, as is evident from the figure.

Sensitivity to Spring Constant An important source of error in SMFS experiments arises from uncertainties in the spring constant calibration. This uncertainty in spring constant (typically ±5-10% (37, 62–64)) leads to uncertainties in the measurements of force and our estimations for the Kuhn segment lengths. We performed sensitivity analysis to ascertain the effect of this spring constant uncertainty on the mean and variance of the fitted Kuhn length parameters.

Table 2 shows the effect of the spring constant on the distribution of Kuhn segment lengths for ELP1-180 in PBS+1.5 M NaCl. Varying the spring constant has a nonlinear effect on the mean and variance of the Kuhn segment length distribution. As can be seen however, the difference between means remains fairly stable. Thus an uncertainty in the cantilever spring constant of ±10% still permits differentiation between estimated Kuhn lengths under different conditions.

Measurement Error in Both Force and Distance In a typical SMFS experiment both force and distance are measured with some amount of noise or uncertainty. Noise in force measurements arises due to thermal vibrations of the cantilever, and is typically on the order of tens of piconewtons. Uncertainty in distance measurements may arise from noise in the capacitance sensor or strain gauge (65). Even when a position sensor is not used (as in some commercially available AFMs), distance measurements are influenced by the discretization of the digital to analog converters and effects such as creep in the piezoelectric material. When fitting a polymer elasticity model using least-squares minimization, only noise in the measured distance (FJC) is accounted for. Conversely, when fitting the WLC or inverse-FJC, only noise in measured force is accounted for. As described in Section 5.3, this can lead to bias and increased variance in the estimated parameters.

We repeated the fitting of several ELP curves to the FJC model using the Bayesian model described in Section 5.3 which accounts for noise in both force and distance measurements simultaneously. MCMC simulation was performed using the freely available WinBugs package. Figure 13 shows resulting Kuhn length distributions compared with those obtained in the previous section. The resulting 95% posterior intervals show a 20-50% reduction from those obtained from the model which accounts only for noise in the distance. Shifts in the interval locations suggests that the bias is also significantly reduced.

The histograms in Figure 14 compare the results from the two models for a typical force-extension curves. In this case, the effect of including noise in force is a moderate but noticeable improvement. Whether this improvement justifies the additional computational effort required to generate the Bayesian intervals probably depends on the magnitude of differences one is attempting to discern. In many cases the original bootstrap intervals may be sufficient.

7 Discussion and Conclusions

Force spectroscopy by AFM and magnetic and optical tweezers has attracted significant interest over the past decade for the study of mechanical and conformational behavior of single molecules on surfaces (2, 3, 14). Many publications in this area have used polymer elasticity models to analyze the force extension behavior (2, 3, 5, 11, 13).

In this paper we have developed a statistical approach for the quantitative characterization and comparison of force-extension curves obtained from force spectroscopy measurements based on estimating parameters of polymer elasticity models. We demonstrated this approach by extracting and analyzing the Kuhn segment length parameters of a freely-jointed chain model from ELP force extension curves, but our approach applies directly to other polymer elasticity models such as the extended freely-jointed chain and the worm-like chain. As part of our approach, we identified a force window in which the experimental data for ELP is well-described by the FJC model, with the model breaking down at higher forces due to enthalpic contributions and a lower forces due to noise from thermal vibrations of the cantilever. Our parameter estimation approach involves fitting multiple force-extension curves for each molecular construct and quantification of uncertainty in the parameter values by bootstrap resampling. We also demonstrate a further refinement that handles noise in force as well as distance via a Bayesian model estimated by Markov chain Monte Carlo (MCMC) simulation.

While the single curve model is widely used for the analysis of force extension curves (2, 10, 13, 24), the use of multiple curves to fit a shared parameter for each experimental data significantly reduced the uncertainty associated with the resulting (Kuhn segment length) parameters, allowing more precise comparison between Kuhn segment lengths obtained from different molecular constructs or under varying experimental conditions (Figure 10). In addition, modeling noise simultaneously in force and distance further increases the precision of the parameter estimation (Figure 14). The reported approach produces narrow Kuhn segment length distributions with a 95% confidence interval of $\approx$0.01-0.04nm, and enables us to quantify subtle changes in elastic response.

We have applied our data analysis approach to enable us to perform previously impossible studies to quantify and elucidate the molecular mechanics of stimulus-responsive ELPs under varying solvent conditions and molecular composition. By reducing uncertainty in fitted parameters, we were able to resolve small differences in Kuhn length. Our results for ELPs indicate that hydrophobic hydration, rather than local backbone conformational entropy, is the key contributor to modulating the molecular elasticity of ELPs under primary sequence variation.

We have provided a detailed description of the procedure that can be used by other researchers working with single molecule force spectroscopy. Our statistical methodology provides a new approach to obtaining polymer elasticity model parameters from force-extension curves. The approach improves significantly on methods in common use and yields improved precision in parameter estimates. These improvements have allowed us to demonstrate that subtle conformational and structural variations in single macromolecules can be resolved, which is of significant interest since conformation and chemical composition control many important characteristics of biological and synthetic macromolecules (61). We anticipate that even smaller variations in Kuhn segment length may be discernible by including a larger

number of curves in the analysis, and by using using smaller cantilevers (66), minimizing thermal noise contributions. These results suggest that the data analysis methodology reported here enables SMFS to be used to study detailed molecular properties in intrinsically unstructured biomacromolecules.

Acknowledgments

SCS was partially supported by NSF grant DMS-0605141.

APPENDIX

A: Nonlinear Least Squares Calculations

Single curves Fitting the above models requires minimization of (weighted) least-squares criteria (Eq. 7). Since the w_i's are a function of θ, this is done iteratively by solving a sequence of WLS problems

$$\hat{\theta}^{(t+1)} = \arg\min_{\theta} \sum_{i=1}^{n} w_i(\hat{\theta}^{(t)})(d_i - g(f_i;\theta))^2 - \log w_i(\hat{\theta}^{(t)}) \tag{13}$$

Each successive minimization of the form (Eq. 13) is itself a NLS problem requiring iterative solution. Here we use a simplex method to obtain $\hat{\theta} = (\hat{l}_k, \hat{L}, \hat{\lambda})$ using the default Matlab nonlinear optimization routine.

Multiple curves Whereas simplex methods work well for solving (Eq. 13) in $p = 2$ (l_K, n_K) or $p = 3$ (l_K, n_K, λ) parameters, minimization of (Eq. 9) involves $m(p-1)+1$ parameters. In high dimensions nonlinear solvers such as simplex often perform poorly. However, we may rewrite the minimization criteria (Eq. 9) as:

$$\min_{l_K} \sum_{j=1}^{m} \min_{L_j} \sum_{i=1}^{n_j} w_{ij} \left(d_{ij} - g(f_{ij}; l_K, L_j)\right)^2 - \log \sigma_j^2 w_{ij}$$

which immediately suggests a stage-wise or iterative decomposition. We performed stage-wise minimization by discretizing l_K in a range of plausible values $\left[l_K^{\min}, l_K^{\max}\right]$, so $l_K \in [l_K^{(1)} \ldots l_K^{(s)}]$ with $l_K^{(s)} - l_K^{(s-1)} = \Delta l_K$ chosen approximately two orders of magnitude smaller than a mean value of l_K at the same experimental condition. For each value of Kuhn length $l_K = l_K^{(s)}$, we found optimal contour lengths $(\hat{L}_1^{(s)}, \ldots, \hat{L}_m^{(s)})$ by solving

$$\hat{L}_j^{(s)} = \arg\min_{L_j} \sum_{i=1}^{n_j} w_{ij}(l_K^{(s)}, L_j) \left(d_{ij} - g(f_{ij}; l_K^{(s)}, L_j)\right)^2 - \log \sigma_j^2 w_{ij}$$

for $j = 1, \ldots, m$ independently, and then computing:

$$\gamma(s) = \sum_{j=1}^{m} \sum_{i=1}^{n_j} w_{ij} \left(d_{ij} - g(f_{ij}; l_K^{(s)}, \hat{L}_j^{(s)})\right)^2 - \log \sigma_j^2 w_{ij}$$

Then $\hat{\theta}$ is given by $l_K^{(s^*)}, \hat{L}_1^{(s^*)}, \ldots, \hat{L}_m^{(s^*)}$ where $s^* = \arg\min_{s \in \{1,\ldots,S\}} \gamma(s)$.

Using such large-scale discretization is computationally intensive and the precision of the resulting estimates is limited by Δl_K. A faster alternative that does not rely on discretization is to perform an iterative minimization by iteratively computing:

$$\hat{l}_K^{(t+1)} = \arg\min_{l_K} \sum_{j=1}^{m} \sum_{i=1}^{n_j} w_{ij} \left(d_{ij} - g(f_{ij}; l_K, \hat{L}_j^{(t)}) \right)^2 - \log \sigma_j^2 w_{ij}$$

$$\hat{L}_j^{(t+1)} = \arg\min_{L_j} \sum_{i=1}^{n_j} w_{ij} \left(d_{ij} - g(f_{ij}; \hat{l}_K^{(t+1)}, L_j) \right)^2 - \log \sigma_j^2 w_{ij}$$

until convergence. This iterative minimization is significantly faster in practice, and by comparing the results of both methods, we have found it to be robust to initial starting values (data not shown), which may be chosen randomly. To speed convergence we initialize by setting $L_j^{(0)} = d_j^*$ which tends to be near the minima.

B: Bootstrap Confidence Intervals

A critical aspect of comparing parameters obtained under different experimental conditions or from structural variants is to quantify the uncertainty in the estimated parameters. We use a nonparametric bootstrap procedure (45) to obtain confidence intervals for the fitted parameters by Monte Carlo resampling. A brief sequence of steps is outlined below.

1) Generate B random samples, each drawn as follows:

   (a) *Single curve fitting:* Draw n pairs with replacement from the n observed values (d_i, f_i). Denote the b^{th} sample by pairs $(d_i^b, f_i^b)_{i=1}^n$.

   (b) *Multiple curve fitting:* Draw m full curves from the m observed curves $(\mathbf{d}_j, \mathbf{f}_j)$. Denote the b^{th} sample by $(\mathbf{d}_j^b, \mathbf{f}_j^b)_{j=1}^m$.

2) For each bootstrap sample $b = 1, \ldots, B$, obtain parameter estimates $\hat{\theta}^b$ by fitting the polymer model $g(f; \theta)$

   (a) to single curve data $(d_i^b, f_i^b)_{i=1}^n$ as described in Section 5.1, or

   (b) to multiple curve data $(\mathbf{d}_j^b, \mathbf{f}_j^b)_{j=1}^m$ as described in Section 5.2

3) Construct the 95% confidence interval $\hat{\theta} \pm 2\hat{se}_{bstp}$ where

$$\hat{se}_{bstp} = \sqrt{\frac{1}{B-1} \sum_{b=1}^{B} (\hat{\theta}^b - \bar{\theta})^2} \qquad \bar{\theta} = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^b$$

The number of bootstrap samples B is chosen to balance precision of standard error estimates against computational limitations. All intervals reported in this paper use $B = 100$ for the single curve model and $B = 500$ for the multiple curve model.

More accurate confidence intervals may be obtained directly from the percentiles of the bootstrap samples, by $[\hat{\theta}^{b(.025)}, \hat{\theta}^{b(.975)}]$, where $\hat{\theta}^{b(\alpha)}$ is the $(100 \times \alpha)$th empirical percentile, obtained by sorting the bootstrap samples and taking the $(B \times \alpha)$th value. This procedure requires significantly more sampling ($B = 500$ or 1000) (45), and thus this was done here only for the multiple curve model examples.

References

1. Wang, C., W. Q. Shi, W. K. Zhang, X. Zhang, Y. Katsumoto, and Y. Ozaki, 2002. Force spectroscopy study on poly(acrylamide) derivatives: Effects of substitutes and buffers on single-chain elasticity. *Nano Lett.* 2:1169–1172.

2. Zhang, W., and X. Zhang, 2003. Single molecule mechanochemistry of macromolecules. *Prog. Polym. Sci.* 28:1271–1295.

3. Janshoff, A., M. Neitzert, Y. Oberdorfer, and H. Fuchs, 2000. Force spectroscopy of molecular systems - Single molecule spectroscopy of polymers and biomolecules. *Angew. Chem., Int. Ed.* 39:3213–3237.

4. Zou, S., W. K. Zhang, X. Zhang, and B. Z. Jiang, 2001. Study on polymer micelles of hydrophobically modified ethyl hydroxyethyl cellulose using single-molecule force spectroscopy. *Langmuir* 17:4799–4808.

5. Li, H. B., M. Rief, F. Oesterhelt, H. E. Gaub, X. Zhang, and J. C. Shen, 1999. Single-molecule force spectroscopy on polysaccharides by AFM - nanomechanical fingerprint of alpha-(1,4)-linked polysaccharides. *Chem. Phys. Lett.* 305:197–201.

6. Rief, M., J. M. Fernandez, and H. E. Gaub, 1998. Elastically coupled two-level systems as a model for biopolymer extensibility. *Phys. Rev. Lett.* 81:4764–4767.

7. Rief, M., M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, 1997. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* 276:1109–1112.

8. Marszalek, P. E., H. B. Li, and J. M. Fernandez, 2001. Fingerprinting polysaccharides with single molecule atomic force microscopy. *Biophys. J.* 80:156a–156a.

9. Marszalek, P. E., H. Lu, H. B. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten, and J. M. Fernandez, 1999. Mechanical unfolding intermediates in titin modules. *Nature* 402:100–103.

10. Ortiz, C., and G. Hadziioannou, 1999. Entropic elasticity of single polymer chains of poly(methacrylic acid) measured by atomic force microscopy. *Macromolecules* 32:780–787.

11. Oesterhelt, F., M. Rief, and H. Gaub, 1999. Single-molecule force spectroscopy by AFM indicates helical structure of poly(ethelene-glycol) in water. *New J. Phys.* 6:1–11.

12. Zhang, W. K., S. Zou, C. Wang, and X. Zhang, 2000. Single polymer chain elongation of poly(N-isopropylacrylamide) and poly(acrylamide) by atomic force microscopy. *J. Phys. Chem. B* 104:10258–10264.

13. Hugel, T., M. Grosholz, H. Clausen-Schaumann, A. Pfau, H. Gaub, and M. Seitz, 2001. Elasticity of single polyelectrolyte chains and their desorption from solid supports studied by AFM based single molecule force spectroscopy. *Macromolecules* 34:1039–1047.

14. Bustamante, C., S. B. Smith, J. Liphardt, and D. Smith, 2000. Single-molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol.* 10:279–285.

15. Bouchiat, C., M. D. Wang, J. F. Allemand, T. Strick, S. M. Block, and V. Croquette, 1999. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys. J.* 76:409–413.

16. Mazars, M., 1999. Freely jointed chains in external potentials: analytical computations. *J. Phys. A* 32:1841–1861.

17. Marko, J. F., and E. D. Siggia, 1995. Stretching DNA. *Macromolecules* 28:8759–8770.

18. Smith, S. B., Y. J. Cui, and C. Bustamante, 1996. Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules. *Science* 271:795–799.

19. Onoa, B., S. Dumont, J. Liphardt, S. B. Smith, I. Tinoco, and C. Bustamante, 2003. Identifying kinetic barriers to mechanical unfolding of the T-thermophila ribozyme. *Science* 299:1892–1895.

20. Yang, G., C. Cecconi, J. Haack, W. Bryer, W. Baase, F. W. Dahlquist, M. S. Kellermayer, and C. Bustamante, 1998. Force induced unfolding and refolding of individual protein molecules by SFM. *Biophys. J.* 74:A227–A227.

21. Li, H. B., W. K. Zhang, X. Zhang, J. C. Shen, B. B. Liu, C. X. Gao, and G. T. Zou, 1998. Single molecule force spectroscopy on poly(vinyl alcohol) by atomic force microscopy. *Macromol. Rapid Commun.* 19:609–611.

22. Eckel, R., R. Ros, A. Ros, S. D. Wilking, N. Sewald, and D. Anselmetti, 2003. Identification of binding mechanisms in single molecule-DNA complexes. *Biophys. J.* 85:1968–1973.

23. Flory, P. J., 1969. Statistical Mechanics of Chain Molecules. John Wiley & Sons.

24. Zhang, W. K., S. Zou, C. Wang, and X. Zhang, 2000. Single polymer chain elongation of poly(N-isopropylacrylamide) and poly(acrylamide) by atomic force microscopy. *J. Phys. Chem. B* 104:10258–10264.

25. Bemis, J. E., B. B. Akhremitchev, and G. C. Walker, 1999. Single polymer chain elongation by atomic force microscopy. *Langmuir* 15:2799–2805.

26. Lu, H., B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten, 1998. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.* 75:662–671.

27. Lee, G., W. Nowak, J. Jaroniec, Q. M. Zhang, and P. E. Marszalek, 2004. Molecular dynamics simulations of forced conformational transitions in 1,6-linked polysaccharides. *Biophys. J.* 87:1456–1465.

28. Orzechowski, M., and P. Cieplak, 2005. Application of steered molecular dynamics (SMD) to study DNA-drug complexes and probing helical propensity of amino acids. *J. Phys: Condens. Matter* 17:S1627–S1640.

29. Morris, S., S. Hanna, and M. J. Miles, 2004. The self-assembly of plant cell wall components by single-molecule force spectroscopy and Monte Carlo modelling. *Nanotechnology* 15:1296–1301.

30. Friedsam, C., A. K. Wehle, F. Kuhner, and H. E. Gaub, 2003. Dynamic single-molecule force spectroscopy: bond rupture analysis with variable spacer length. *J. Phys.: Condens. Matter* 15:S1709–S1723.

31. Titantah, J. T., C. Pierleoni, and J. P. Ryckaert, 1999. Different statistical mechanical ensembles for a stretched polymer. *Phys. Rev. E* 60:7010–7021.

32. Mazars, M., 1998. Canonical partition functions of freely jointed chains. *J. Phys. A* 31:1949–1964.

33. Mazars, M., 1996. Statistical physics of the freely jointed chain. *Phys. Rev. E* 53:6297–6319.

34. Meyer, D. E., K. Trabbic-Carlson, and A. Chilkoti, 2001. Protein purification by fusion with an environmentally responsive elastin-like polypeptide: Effect of polypeptide length on the purification of thioredoxin. *Biotechnol. Progr.* 17:720–728.

35. Meyer, D. E., and A. Chilkoti, 1999. Purification of recombinant proteins by fusion with thermally-responsive polypeptides. *Nat. Biotech.* 17:1112–1115.

36. Meyer, D. E., and A. Chilkoti, 2002. Genetically encoded synthesis of protein-based polymers with precisely specified molecular weight and sequence by recursive directional ligation: Examples from the elastin-like polypeptide system. *Biomacromolecules* 3:357–367.

37. Hutter, J. L., and J. Bechhoefer, 1993. Calibration of Atomic-Force Microscope Tips. *Rev. of Sci. Instrum.* 64:1868–1873.

38. Marko, J. F., and E. D. Siggia, 1994. Fluctuations and Supercoiling of DNA. *Science* 265:506–508.

39. Fixman, M., and J. Kovac, 1973. Polymer conformational statistics. III. Modified Gaussian models with stiff chains. *J. Chem. Phys.* 58:1564–1568.

40. Wang, M. D., H. Yin, R. Landick, J. Gelles, and S. M. Block, 1997. Stretching DNA with Optical Tweezers. *Biophysical Journal* 72:1335–1346.

41. Cleveland, W. S., 1979. Robust locally-weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* 74:829–836.

42. Haupt, B. J., T. J. Senden, and E. M. Sevick, 2002. AFM evidence of Rayleigh instability in single polymer chains. *Langmuir* 18:2174–2182.

43. Seber, G. A. F., and C. J. Wild, 1989. Nonlinear Regression. Wiley.

44. Mardia, K. V., and P. E. Jupp, 2000. Directional Statistics. Wiley.

45. Efron, B., and R. J. Tibshirani, 1993. An Introduction to the Bootstrap. Chapman & Hall.

46. Carroll, R. J., D. Ruppert, and L. Stefanski, 1995. Measurement Error in Nonlinear Models. Chapman & Hall.

47. Dellaportas, P., and D. A. Stephens, 1995. Bayesian analysis of errors-in-variables regression models. *Biometrics* 51:1085–1095.

48. Richardson, S., 1996. Measurement Error, Chapman & Hall, 401–417. *In* Gilks et al. (49).

49. Gilks, W. R., S. Richardson, and D. J. Spiegelhalter, editors, 1996. Markov Chain Monte Carlo in Practice. Chapman & Hall.

50. Spiegelhalter, D., and A. Thomas, 1998. Graphical modeling for complex stochastic systems: the BUGS project. *IEEE Intel. Sys. & Appl.* 13:14–15.

51. Best, N. G., D. J. Spiegelhalter, A. Thomas, and C. E. G. Brayne, 1996. Bayesian analysis of realistically complex models. *J. Roy. Stat. Soc. A* 159:323–342.

52. Urry, D. W., 1997. Physical chemistry of biological free energy transduction as demonstrated by elastic protein-based polymers. *J. Phys. Chem. B* 101:11007–11028.

53. Kellermayer, M. S., S. B. Smith, H. L. Granzier, and C. Bustamante, 1997. Folding-Unfolding Transitions in Single Titin Molecules Characterized with Laser Tweezers. *Science* 276:1112–1116.

54. Hyun, J., S. J. Ahn, W. K. Lee, A. Chilkoti, and S. Zauscher, 2002. Molecular recognition-mediated fabrication of protein nanostructures by dip-pen lithography. *Nano Lett.* 2:1203–1207.

55. Weis-Fogh, T., and S. Anderson, 1970. New Molecular Model for the Long-range Elasticity of Elastin. *Nature* 227:718–721.

56. Schwarzinger, S., P. E. Wright, and H. J. Dyson, 2002. Molecular Hinges in Protein Folding: the Urea-Denatured State of Apomyoglobin. *Biochemistry* 41:12681–6.

57. Karplus, P. A., 1997. Hydrophobicity regained. *Prot. Sci.* 6:1302–1307.

58. Rose, G. D., A. R. Geselowitz, G. J. Lesser, R. H. Lee, and M. H. Zehfus, 1985. Hydrophobicity of Amino-Acid Residues in Globular-Proteins. *Science* 229:834–838.

59. Hopp, T. P., and K. R. Woods, 1981. Prediction of Protein Antigenic Determinants From Amino Acid Sequences. *Proc. Natl. Acad. Sci. USA* 78:3824–3828.

60. Kyte, J., and R. F. Doolittle, 1982. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* 157:105–132.

61. Valiaev, A., D. W. Lim, S. C. Schmidler, R. L. Clark, A. Chilkoti, and S. Zauscher, 2008. Hydration and Conformational Mechanics of Single End-Tethered Elastin-Like Polypeptides. *J. Amer. Chem. Soc. (to appear)*.

62. Maeda, N., and T. J. Senden, 2000. A method for the calibration of force microscopy cantilevers via hydrodynamic drag. *Langmuir* 16:9282–9286.

63. Cleveland, J. P., S. Manne, D. Bocek, and P. K. Hansma, 1993. A Nondestructive Method for Determining the Spring Constant of Cantilevers for Scanning Force Microscopy. *Rev. Sci. Instrum.* 64:403–405.

64. Sader, J. E., and L. White, 1993. Theoretical-Analysis of the Static Deflection of Plates for Atomic-Force Microscope Applications. *J. Appl. Phys.* 74:1–9.

65. Kuijpers, A. A., G. J. M. Krijnen, R. J. Wiegerink, T. S. J. Lammerink, and M. Elwenspoek, 2003. 2D-finite-element simulations for long-range capacitive position sensor. *J. Micromech. Microeng.* 13:S183–S189.

66. Viani, M. B., L. I. Pietrasanta, J. B. Thompson, A. Chand, I. C. Gebeshuber, J. H. Kindt, M. Richter, H. G. Hansma, and P. K. Hansma, 2000. Probing protein-protein interactions in real time. *Nat. Struct. Biol.* 7:644–647.

Polymer Model	Parameter vector vector θ	Parameter names	Elasticity law $g(f;\theta)$
FJC	(l_K, L)	Kuhn length, contour length	(Eq. 1)
EFJC	(l_K, L, λ)	Kuhn length, contour length, segmental stiffness	(Eq. 2)
WLC	(l_p, L)	persistence length, contour length	(Eq. 3)
EWLC	(l_p, L, Φ)	persistence length, contour length, elastic modulus	(Eq. 4)

Table 1: Specific choices of θ and $g(f;\theta)$ corresponding to the various polymer elasticity models described in Section 3. Parameter fitting equations in text are given in terms of generic θ, g to enable application of the described techniques to various polymer models.

	ELP in PBS			ELP in PBS+1.5M NaCl			Difference	
$K_c{}^*$	Mean	95% int	se_{bstp}	Mean	95% int	se_{bstp}	Mean	95% int
[pN/nm]	[nm]	[nm]	[nm]	[nm]	[nm]	[nm]	[nm]	[nm]
$0.8 \cdot K_c$	-	-	-	.426	[.410 .443]	.009	-	-
$0.9 \cdot K_c$	.380	[.366 .396]	.008	.421	[.400 .445]	.012	.041	.004
K_c	.382	[.357 .407]	.013	.435	[.411 .458]	.012	.053	.004
$1.1 \cdot K_c$	.374	[.357 .392]	.009	.424	[.406 .442]	.009	.050	.014
$1.2 \cdot K_c$	.360	[.339 .381]	.011	.422	[.402 .440]	.009	.062	.021

Table 2: Sensitivity analysis of Kuhn length estimates to changes in cantilever spring constant, obtained for ELP in PBS and ELP in PBS+1.5M NaCl. Although changes in mean are observed, differences between means remain relatively stable. Some values (-) were unavailable when the spring constant perturbation lowered the rupture force below the 200pN window.

Figure Legends

1) Schematic structure of the ELP constructs used in the experiments.

2) Normalization and filtering of (a random subset of) force-separation curves. Show are (a) original unnormalized curves (inset), and curves after normalization by contour length at 200pN; (b) normalized curves, along with the median curve (black bold line), +/- confidence bands, and two rejected curves (indicated by the arrows) which fall outside the confidence bands.

3) The drag force plotted as a function of pulling velocity. Solid lines are linear fits to the data and their slopes yield the coefficient $C \cdot \mu_{\text{Hex}} = 24 \frac{\text{pN}}{\mu\text{m/sec}}$ for hexadecane and $C \cdot \mu_{\text{H}_2\text{O}} = 7 \frac{\text{pN}}{\mu\text{m/sec}}$ for water. The resulting ratio of the slopes for hexadecane and water is 3.43, comparable with the ratio of 3.76 obtained from the viscosities of hexadecane and water at room temperature ($3.35cP/0.89cP = 3.76$). (I) and (II) denote different cantilevers with the same geometry.

4) Estimated hydrodynamic drag force correction (Eq. 6) applied to force-extension profile obtained for ELP1-180 in PBS buffer solution. Open circles indicate the force extension data before the hydrodynamic drag subtraction, while the solid line corresponds to data after drag force subtraction. Grey and black solid lines indicate FJC fits to the data before and after subtraction of hydrodynamic drag contribution.

5) (a) Fitting the FJC model to a wide force range results in deviation between actual and fit data at low and high forces. (b) Solid circles show separation residuals obtained by subtracting measured data from the fitted FJC model. Dashed lines indicate the 60-200pN force window chosen for fitting.

6) Fitted polymer elasticity models in the 60-200 pN force window for an example force extension curve. (a) Extended FJC fit (b) FJC fit. Very little difference between the models is observed in the chosen force window. Inset shows separation residuals.

7) Separation residual plots from least-squares fitting of two force-extension curves using the FJC model. (a) ELP, (b) PEG. Also shown are the variance functions (Eq. 8) obtained from weighted least-squares shown as $\pm 2\hat{\sigma}\sqrt{V_d(f)}$.

8) Force residuals obtained from fitting the model of Section 5.3 which accounts for noise in both force and distance. No evidence of heteroscedasticity is seen.

9) Force extension curve for PEG in hexadecane. The solid line shows the FJC model fit to the data in 30-470pN force window.

10) Box plots obtained by the multiple curves model for ELP1-180 in PBS+1.5M NaCl and ELP1-180 in PBS (center plot) and single curve mode (right and left plots). The multiple (or shared) model utilizes the curves shown on the right and left sides of the figure (i.e., ten curves were used for ELP in PBS and nine curves for ELP in PBS+1.5 M

NaCl.). The use of the multiple curves approach produces narrower confidence intervals and enables differentiation between smaller differences in Kuhn segment length.

11) Box plots showing the Kuhn length distributions obtained from the multiple curves fitting approach for three different types of ELPs shown in Figure 1. The elastic response behaviors of different ELP molecules are clearly distinguishable, despite differing at only a single guest residue.

12) Comparison between Kuhn segment length distributions obtained for ELP1-180 in PBS+1.5M NaCl using previously published data reduction procedures and that obtained from the shared parameter estimate described in Section 5.

13) Reduction in bias and uncertainty of Kuhn length obtained when noise in both force and distance are accounted for in the statistical model. Boxplots show the 95% intervals for nine force extension curves. The reduction in interval width is given for each curve.

14) Histograms comparing results of Kuhn length estimation under models with and without accounting for noise in force measurements. The force-and-distance model yields a more narrow 95% (posterior) interval for l_K compared to the 95% confidence interval (shown as brackets) of the distance-only model.

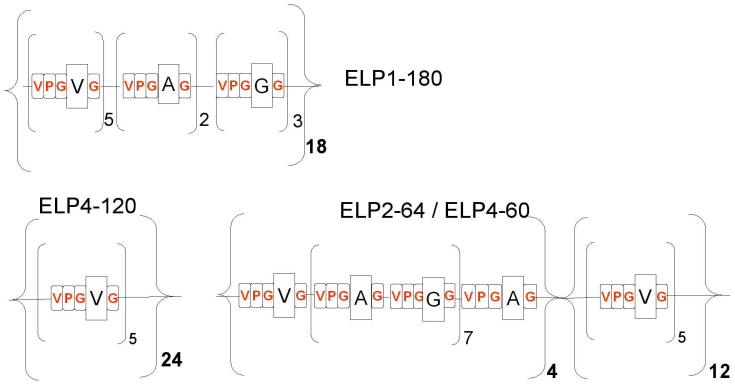
VPG V G
5
VPG A G
2
VPG G G
3
18
ELP1-180
ELP4-120
VPG V G
5
24
ELP2-64 / ELP4-60
VPG V G
VPG A G
VPG G G
VPG A G
7
4
VPG V G
5
12

Figure 1:

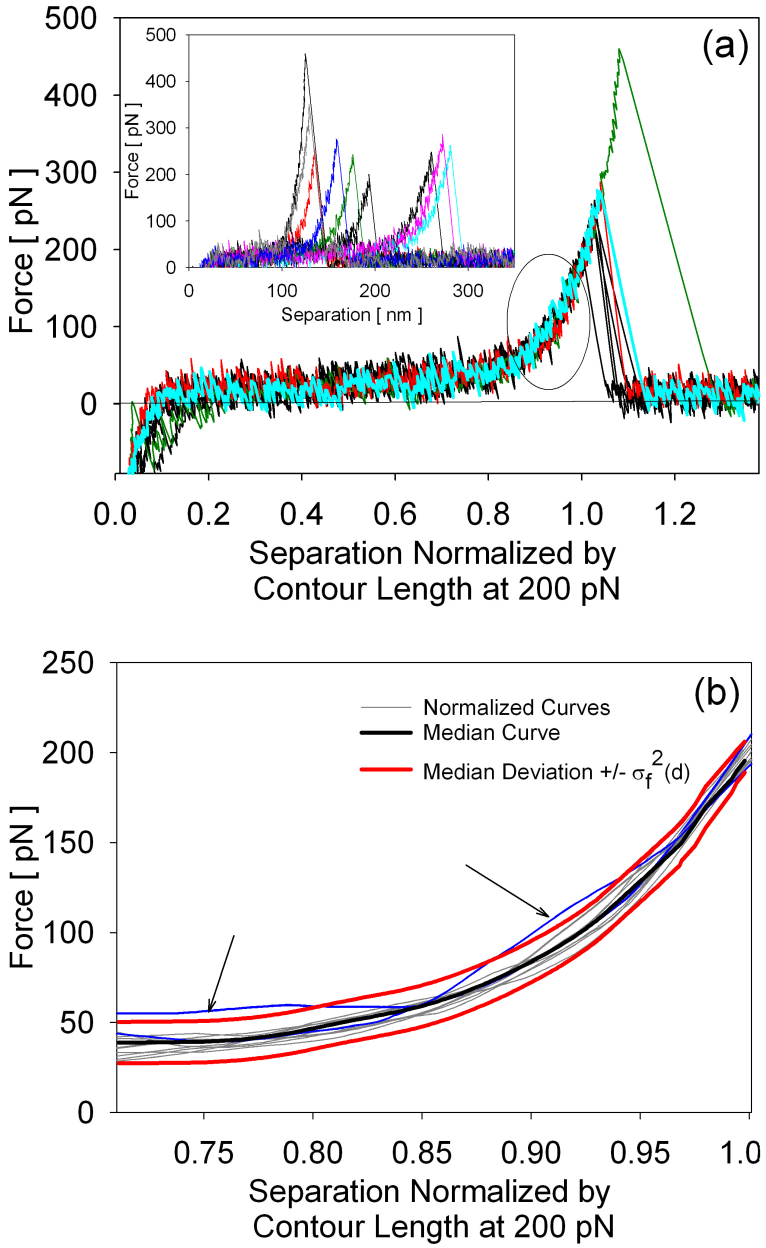
(a)
Force [pN]
Separation Normalized by
Contour Length at 200 pN
Separation [nm]
(b)
Normalized Curves
Median Curve
Median Deviation +/- σf2(d)
Force [pN]
Separation Normalized by
Contour Length at 200 pN

Figure 2:

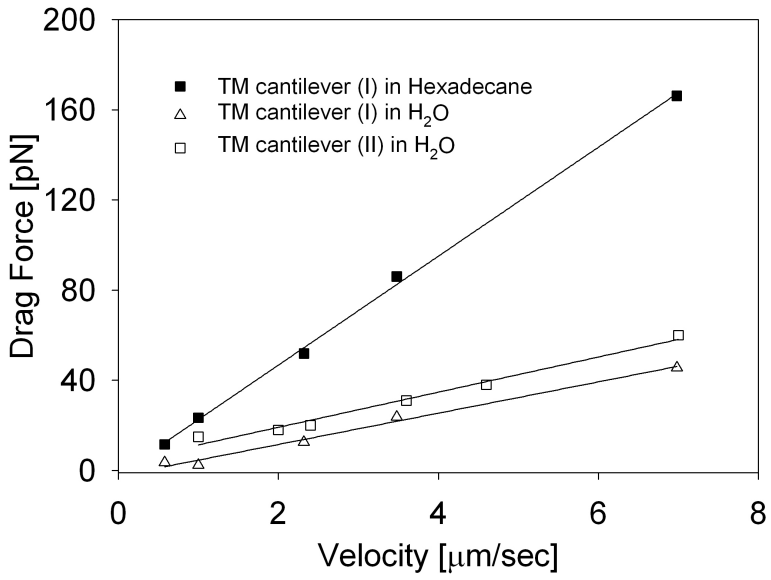

Figure 3:

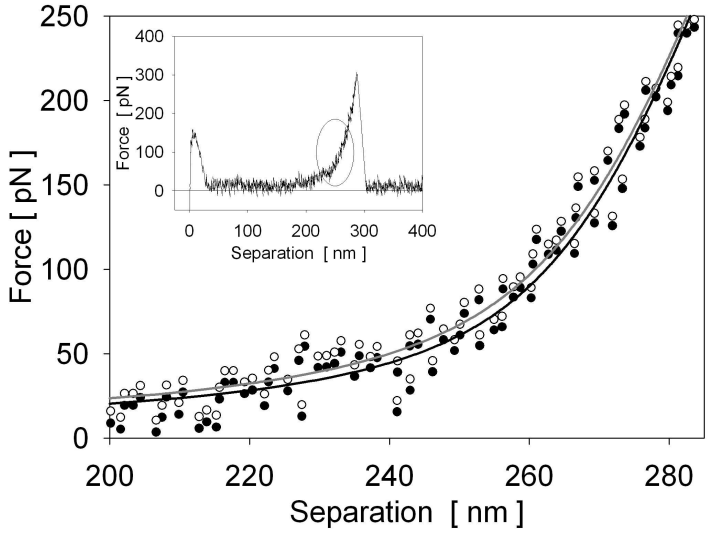

Figure 4:

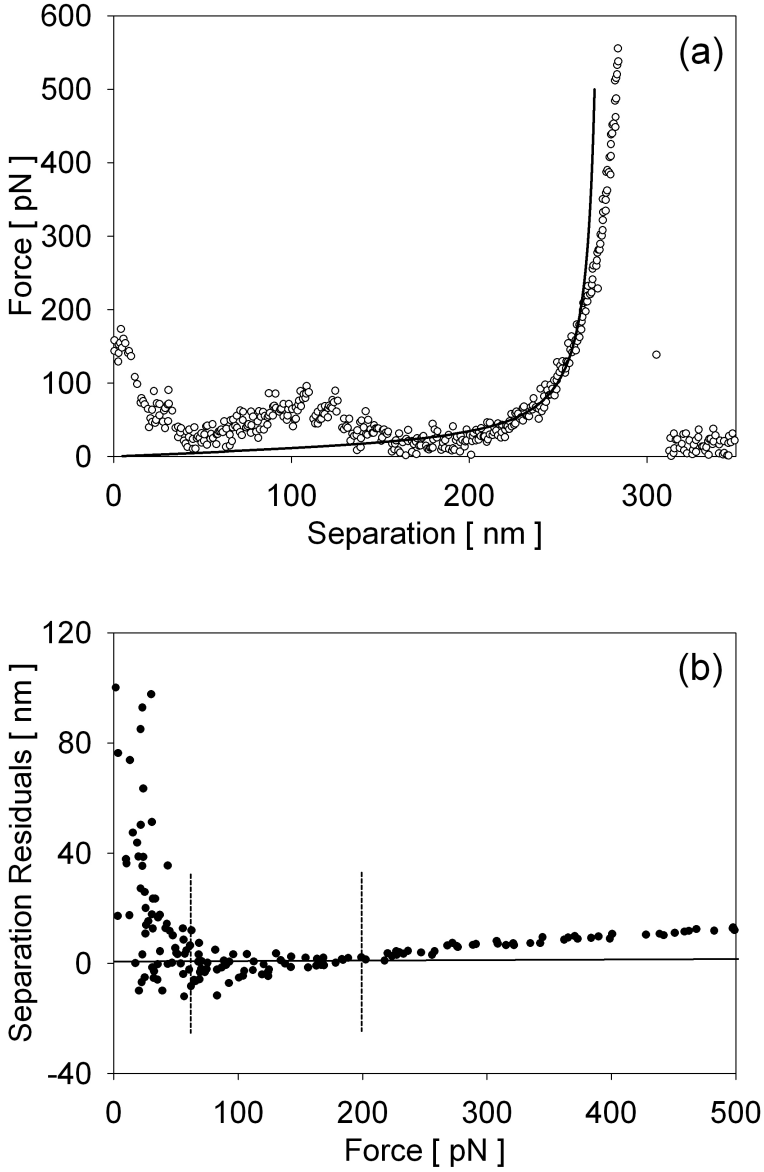
(a)
Force [pN]
Separation [nm]
0
100
200
300
400
500
600
(b)
Separation Residuals [nm]
Force [pN]
-40
0
40
80
120
100
200
300
400
500

Figure 5:

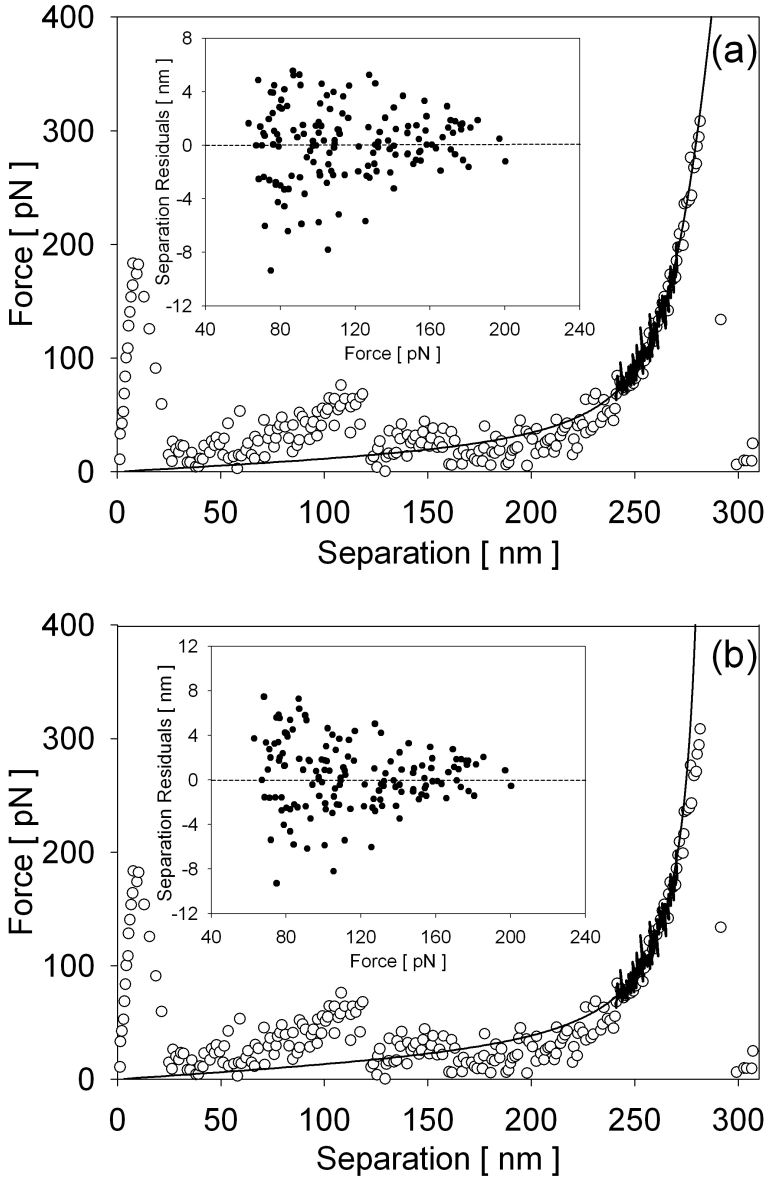
(a)
Force [pN]
Separation [nm]
Separation Residuals [nm]
Force [pN]
(b)
Force [pN]
Separation [nm]
Separation Residuals [nm]
Force [pN]

Figure 6:

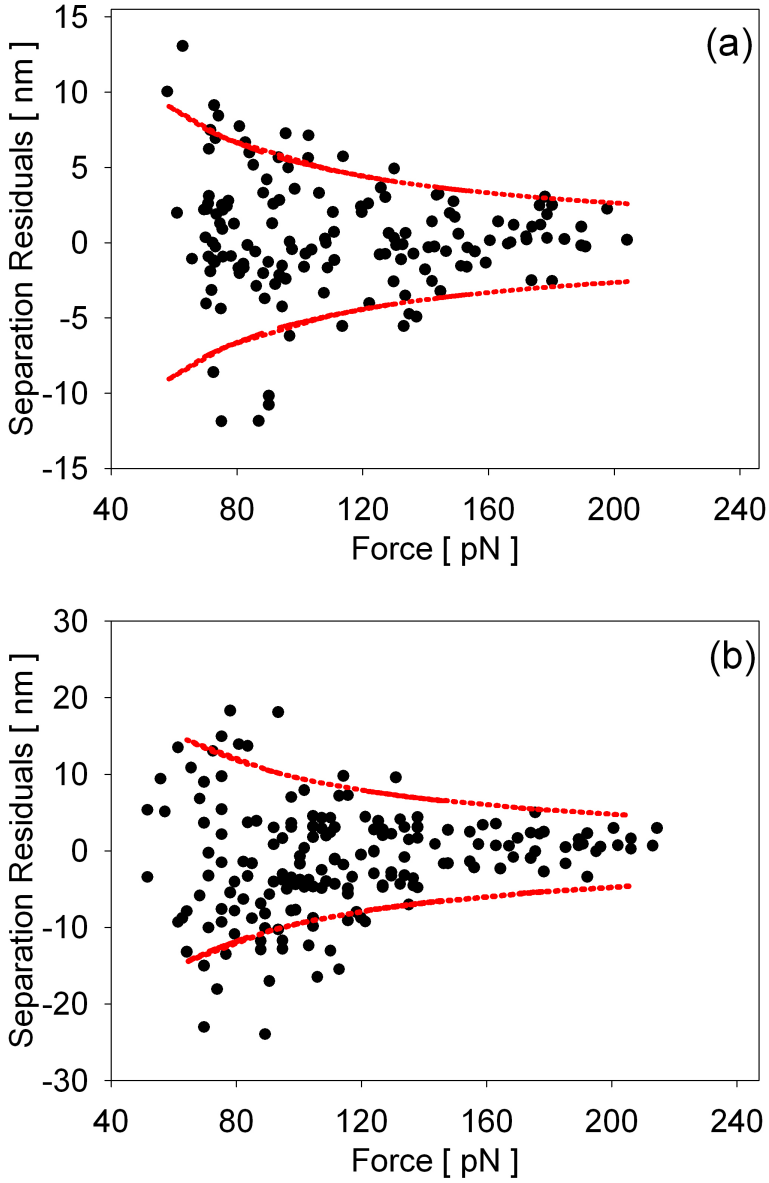
(a)
Separation Residuals [nm]
15
10
5
0
-5
-10
-15
40
80
120
160
200
240
Force [pN]
(b)
Separation Residuals [nm]
30
20
10
0
-10
-20
-30
40
80
120
160
200
240
Force [pN]

Figure 7:

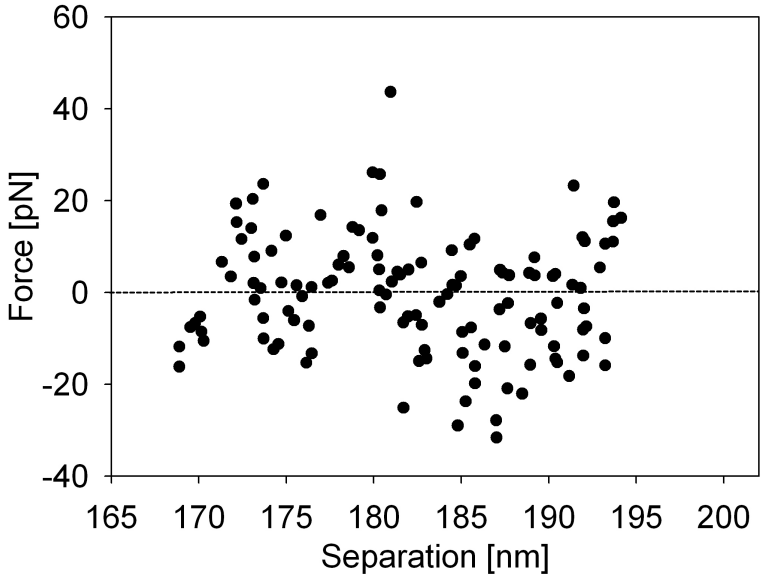

Figure 8:

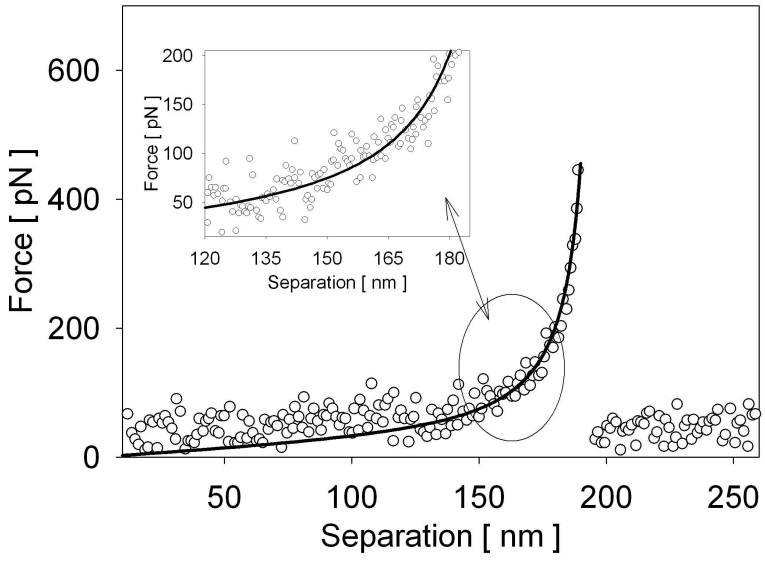

Figure 9:

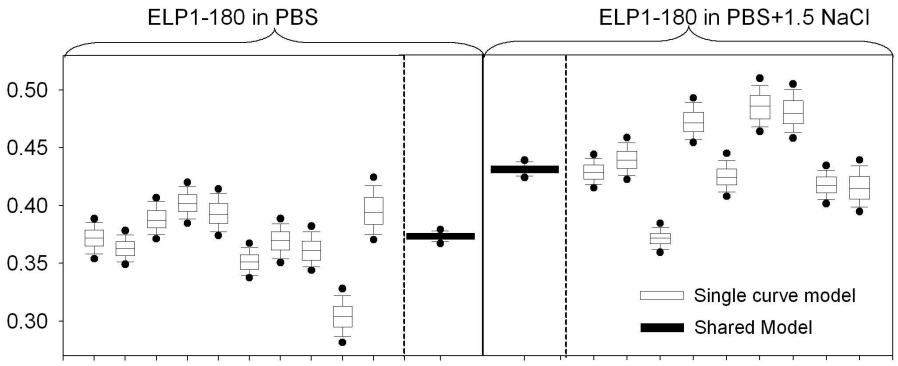

Figure 10:

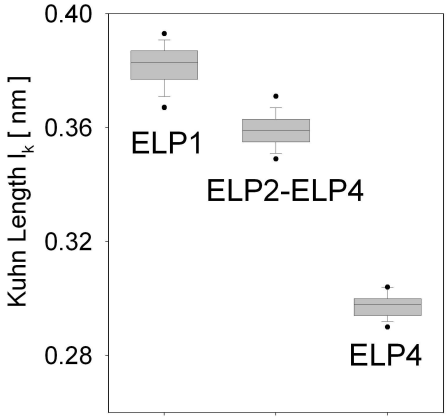

Figure 11:

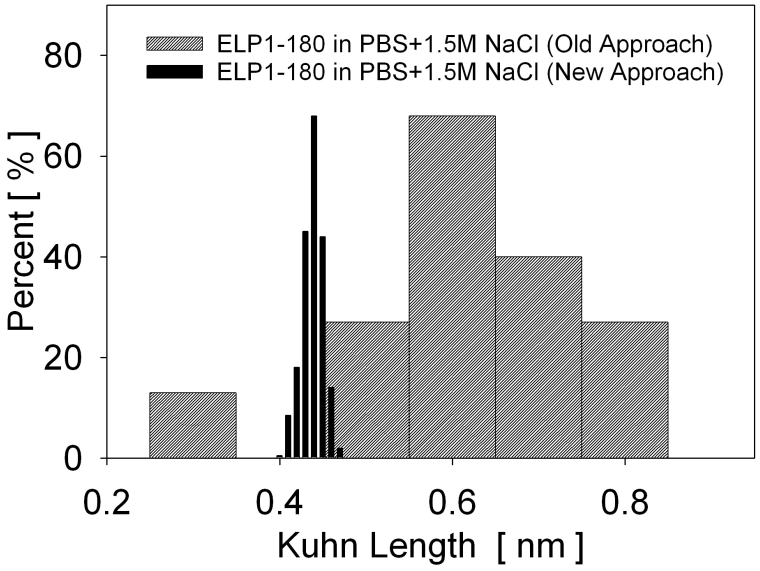

Figure 12:

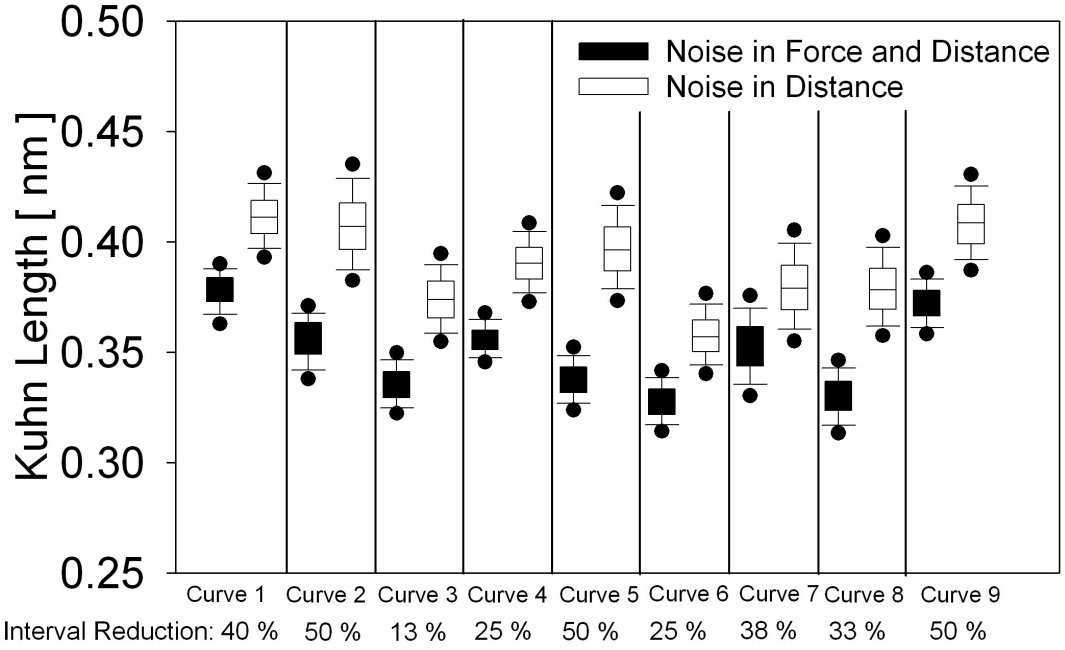

Figure 13:

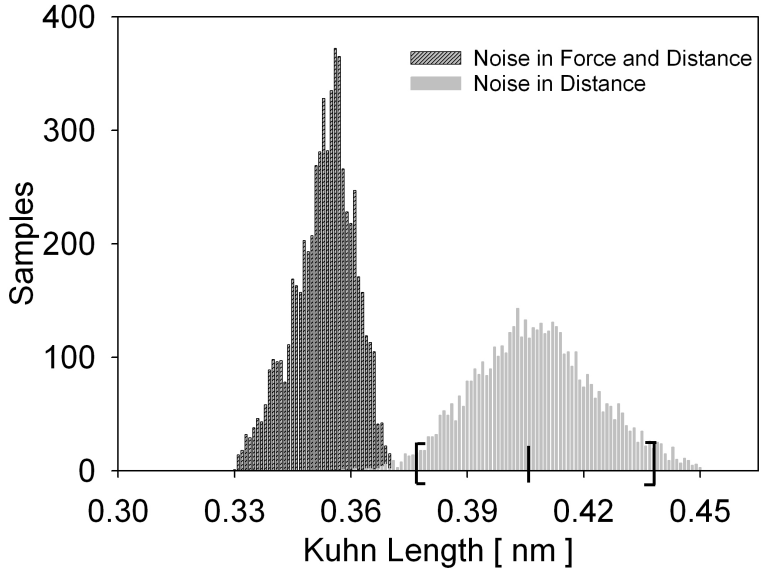

Figure 14: