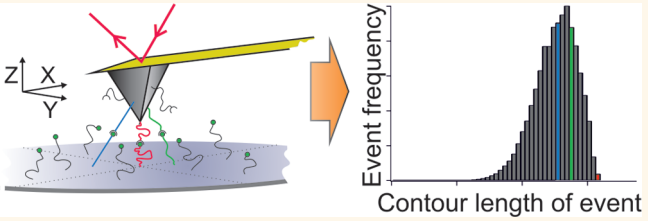
# Extraction of Accurate Biomolecular Parameters from Single-Molecule Force Spectroscopy Experiments

**Oliver E. Farrance, Emanuele Paci, Sheena E. Radford, and David J. Brockwell***

Astbury Centre for Structural and Molecular Biology and School of Molecular and Cellular Biology, University of Leeds, Leeds, West Yorkshire, LS2 9JT, U.K.

**ABSTRACT** The atomic force microscope (AFM) is able to manipulate biomolecules and their complexes with exquisite force sensitivity and distance resolution. This capability, complemented by theoretical models, has greatly improved our understanding of the determinants of mechanical strength in proteins and revealed the diverse effects of directional forces on the energy landscape of biomolecules. In unbinding experiments, the interacting partners are usually immobilized on their respective substrates *via* extensible linkers. These linkers affect both the force and contour length ($L_c$) of the complex at rupture. Surprisingly, while the former effect is well understood, the latter is largely neglected, leading to incorrect estimations of $L_c$, a parameter that is often used as evidence for the detection of specific interactions and remodeling events and for the inference of interaction regions. To address this problem, a model that predicts contour length measurements from single-molecule forced-dissociation experiments is presented that considers attachment position on the AFM tip, geometric effects, and polymer dynamics of the linkers. Modeled data are compared with measured contour length distributions from several different experimental systems, revealing that current methods underestimate contour lengths. The model enables nonspecific interactions to be identified unequivocally, allows accurate determination of $L_c$, and, by comparing experimental and modeled distributions, enables partial unfolding events before rupture to be identified unequivocally.




**KEYWORDS:** contour length · worm-like chain (WLC) · freely jointed chain (FJC) · single-molecule force spectroscopy (SMFS) · atomic force microscopy (AFM)

Single-molecule force spectroscopy (SMFS) techniques are increasingly being applied in biology to understand biomolecular interactions and protein dissociation under the application of force.[1−4] Usually employed in nonequilibrium conditions, SMFS offers unique insights into the kinetics of bond dissociation of inter- and intramolecular interactions under applied force, allowing exploration of regions of the underlying free energy landscape that are inaccessible to traditional ensemble methods.[5] In addition to slip bonds (where lifetime of the intermolecular bond decreases exponentially with applied force), first described by Bell,[6] SMFS experiments have revealed that biomolecular complexes have evolved a wide variety of responses to mechanical deformation (*e.g.*, catch, flex, and trip bonds).[7−9]

In SMFS experiments, the interactions of binding partners, identified by their dissociation (bond breakage) at a characteristic tip−substrate separation, are measured at many retraction velocities. Fitting force−extension data to appropriate models, such as the worm-like chain (WLC)[10] or freely jointed chain (FJC),[11] that describe the force−extension profile of a long, single molecule under stretching inclusive of entropic effects allows the total combined length of tethers and binding partners between the two objects (the contour length ($L_c$)) to be determined.

In addition to filtering real events from noise, the $L_c$ contains information about the extent of unfolding before unfolding or dissociation if the structure of the protein/complex is known.[12−14] For more unstructured polypeptides, the length of each polypeptide chain sequestered in the binding interface can be determined.[15,16] This is achieved by fitting the experimentally obtained contour length distribution with single or multiple normal (Gaussian) distributions[17−22] and comparing the modal value with an estimated value of $L_c$ obtained by

* Address correspondence to D.J.Brockwell@leeds.ac.uk.

Received for review September 11, 2014 and accepted February 3, 2015.

**Published online** 10.1021/nn505135d

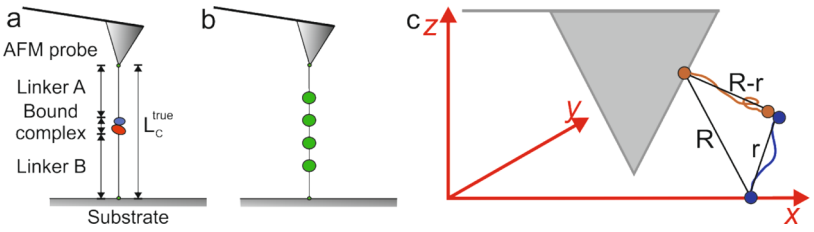



**Figure 1. (a, b) Single-molecule AFM pulling experiment. $L_c^{true}$ is the maximum observable contour length for a particular combination of linkers and binding partners corresponding to the case when those linkers are immobilized at the tip apex and a position directly beneath it on the substrate shown for typical (a) unbinding and (b) unfolding experimental setups. (c) Geometry of the AFM tip (gray triangle) and the linkers (blue and brown lines with filled circles at their termini) used in this model. *R* is the separation between immobilization points, and *r* and *R*−*r* are the separation between ends of polymer chains.**

summing the individual components that contribute to the tip−surface separation at complex rupture (Figure 1a). An identity between the experimental and theoretical $L_c$ values provides evidence that a specific interaction is being quantified and that the complex is native-like at rupture. By contrast, an experimental value that is greater than the theoretical value is used to indicate that unfolding or remodeling of the complex takes place prior to dissociation.[9,23] In the latter case the difference between the estimated and measured contour lengths can be used to calculate the length gain in terms of number of amino acids involved in the conformational change and mapped onto the known structure of the protein, revealing which specific interactions unfold prior to dissociation under the application of force.[12−16]

Interpretation of observed contour length distributions in this manner relies on presuming that each tethering end is immobilized at, or close to, the points at which the two objects come into closest proximity (Figure 1a). The validity of this simplistic approach is thus determined by the manner in which the proteins are immobilized onto the AFM tip and substrate (Figure 1). Generally biomolecules (usually proteins) are covalently immobilized onto the AFM tip and substrate by one of two methods, depending on the type of investigation undertaken. In mechanical unfolding experiments, where a single protein is extended between the tip and AFM substrate, the protein is usually immobilized directly to the mica,[24] glass,[25−27] or gold[28−30] substrate and attached to the tip *via* nonspecific adsorption. As the latter requires the application of pressure onto the polypeptide by the tip, immobilization presumably occurs close to the apex, and the contour length that is measured (found by fitting of an appropriate model to force−distance curves) is usually close to the end-to-end length of the immobilized protein (Figure 1b), irrespective of differences in pulling geometry.[31]

By contrast, in mechanical unbinding experiments, each of the binding partners are immobilized to different surfaces *via* flexible linkers. This minimizes artifacts, maximizes steric freedom to allow complex formation, and aids identification of rupture events (*via* their characteristic force−extension profiles) from nonspecific

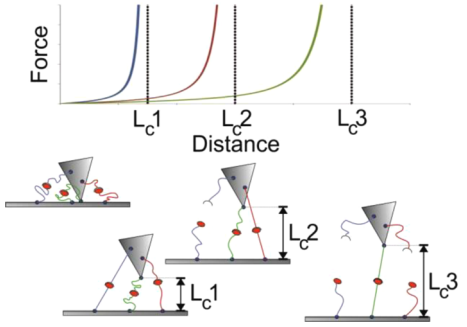



**Figure 2. Schematic demonstrating the principle that the observed contour length from an SMFS dissociation experiment measured using AFM can vary depending on linker immobilization location.**

unbinding events or noise in the experiment. In these experiments, interpretation of $L_c$ is complicated by the presence of the extensible linkers that are attached over the entire AFM tip surface and on the substrate, yielding apparently different $L_c$'s (Figure 2). In addition, as the surface area of the AFM tip increases with increasing distance from its tip, the number of binding sites for attachment of the linker also increases, while the probability of complex formation decreases (as the through-space end-to-end distance between linker ends is required to increase). This convolution of competing effects produces a distribution of apparent contour lengths that are shorter than the true contour length ($L_c^{true}$). In order to explore the relationship between the true contour length and the experimentally observed distribution, we have here developed and numerically solved a model that considers the single-molecule linkers and binding partners as a WLC. The effects of linker length, persistence length, and tip angle/curvature are included in the model, and the relationship between linker length, observed modal contour length value ($L_c^{mod}$), and $L_c^{true}$ is determined empirically. The model is then benchmarked against the observed contour length distributions from SMFS AFM unbinding experiments from three very different protein:ligand complexes: the interaction between two native proteins, a protein and a peptide, and an antibody and its antigen. The model provides a useful tool for the analysis of SMFS AFM data, allowing the extraction of accurate length

estimates of biomolecular complexes under force and the identification of remodeling/unfolding events prior to dissociation.

## RESULTS

**Description of Model.** As described above, the contour length that is measured in an AFM-based SMFS experiment corresponds to the separation between the two surfaces (tip−substrate separation) when the linkers, tethered by complex formation, are fully extended (Figure 1). In the ideal situation, where one immobilization point is located at the apex of the AFM tip and the other directly opposite on the substrate (as in most mechanical unfolding experiments), the observed contour length will be at its maximum and is simply a sum of the length of the components of the tethered complex ($L_c^{true}$). However, it is possible that binding may occur between partners whose linkers are immobilized at other locations. With the tip geometry considered here (square-based pyramid), the greater the distance from either the tip apex or the location directly beneath the apex to which the linker is immobilized, the greater surface area available for linker attachment and thus protein immobilization. Binding between partners immobilized at these locations results in extension of the complex occurring at a shorter tip−substrate separation. This will result in the observation of an apparent contour length that is smaller than $L_c^{true}$ (Figure 2).

To generate theoretical contour length probability distributions accounting for the factors described above, an iterative Monte Carlo procedure was developed. An idealized system was modeled that consists of two tethering surfaces: a flat square substrate surface of length $L_1 + L_2$ (where $L_1$ and $L_2$ are the end-to-end lengths of the linker and protein tethered to each surface), centered below a pyramidal AFM tip with a perfect point and flat sides (an alternative semispherical tip geometry is described in the Supporting Information). Flexible linkers that have complementary binding partners located at one end are immobilized onto these surfaces at their other end. The potential immobilization points of the linkers are modeled as being uniformly distributed over each surface. For each iteration, the immobilization point on the AFM tip was determined by randomly selecting one face and the location on that face. The location of the substrate-bound linker was again selected randomly. Using these coordinates, the separation between immobilization points for the binding partners is calculated for the case when the AFM tip is in contact with the surface. This separation ($R$) (Figure 1c) is then used to calculate the binding probability ($\pi(\vec{R})$).

A function that describes $\pi(\vec{R})$ has been derived by assuming that $P(r)$ is the probability that two ends of a polymer chain, separated by a vector $r$, is known (Figure 1c). Then the probability that two ends of two chains are in contact when the other two ends are separated by a vector $R$ is

$$\pi(\boldsymbol{R}) = \int \mathrm{d}\boldsymbol{r}\, P(\boldsymbol{r})\, P(\boldsymbol{R} - \boldsymbol{r}) \tag{1}$$

where the integral is extended over the whole Cartesian space.

We assume that the linkers behave like worm-like chain. We observe that for a WLC with contour length $L$ and persistence length $p$ the mean square end-to-end distance is

$$\langle R^2 \rangle = 2pL\left[1 - \frac{p}{L}(1 - \mathrm{e}^{-L/p})\right] \tag{2}$$

and that for a freely jointed chain with $N$ monomers linked by rigid rods of length $l$ (the Kuhn length) it is $\langle R^2 \rangle = Nl^2$, while the contour length is $L = Nl$. In the limit $L \gg p$ and $\langle R^2 \rangle = 2pL$, we can assume that WLC and FJC are equivalent if $l = 2p$.

In the following we use this equivalence and that a freely jointed chain containing $N$ monomers each of length $l$ has the probability of the chain ends being separated by a distance $R$ as given by

$$P(\boldsymbol{r}) = \left(\frac{3}{2\pi N l^2}\right)^{3/2} \mathrm{e}^{(-3r^2/2Nl^2)} \tag{3}$$

With this, eq 1 above becomes

$$\pi(\vec{\mathrm{R}}) = \frac{1}{2}\left(\frac{AB}{\pi(A+B)}\right)^{3/2} \mathrm{e}^{-R^2\left(\frac{AB}{A+B}\right)}\left(1 + \mathrm{erf}\left(\frac{BR}{\sqrt{A+B}}\right)\right) \tag{4}$$

where $A = 3/(2N_A l_A^2) = 3/(4L_A p_A^2)$, $B = 3/(2N_B l_B^2) = 3/(4L_B p_B^2)$, and $\mathrm{erf}(x) = (2/\pi)\int_0^x \mathrm{e}^{-t^2}\,\mathrm{d}t$ (see Supporting Information for full derivation).

Binding probabilities from eq 4 are then binned and summed for corresponding contour lengths where the contour length that would be observed during retraction of the AFM tip is calculated by finding the tip sample separation at full extension of the linkers. This procedure was repeated until the contour length distribution was converged ($10^7$ to $10^9$ times). The model thus contains only three parameters: the tip angle and two parameters that describe the linkers ($N$ and l for linkers A and B assuming, as is common practice, that the same linker type is used for both tip and substrate). Forced unbinding of biomolecular complexes using the AFM typically employs heterobifunctional polyethylene glycol (PEG) linkers comprising distinct reactive groups at their termini separated by a defined number of ethylene glycol repeats.[32,33] Previous studies, analyzing single-molecule force−extension profiles, have shown that both unstructured polypeptide chains and PEG possess similar persistence lengths,[34,35] which are usually fixed to 0.4 nm for fitting to data.[36] The resulting contour length−frequency distribution for this model using a tip angle of 60°, a persistence length

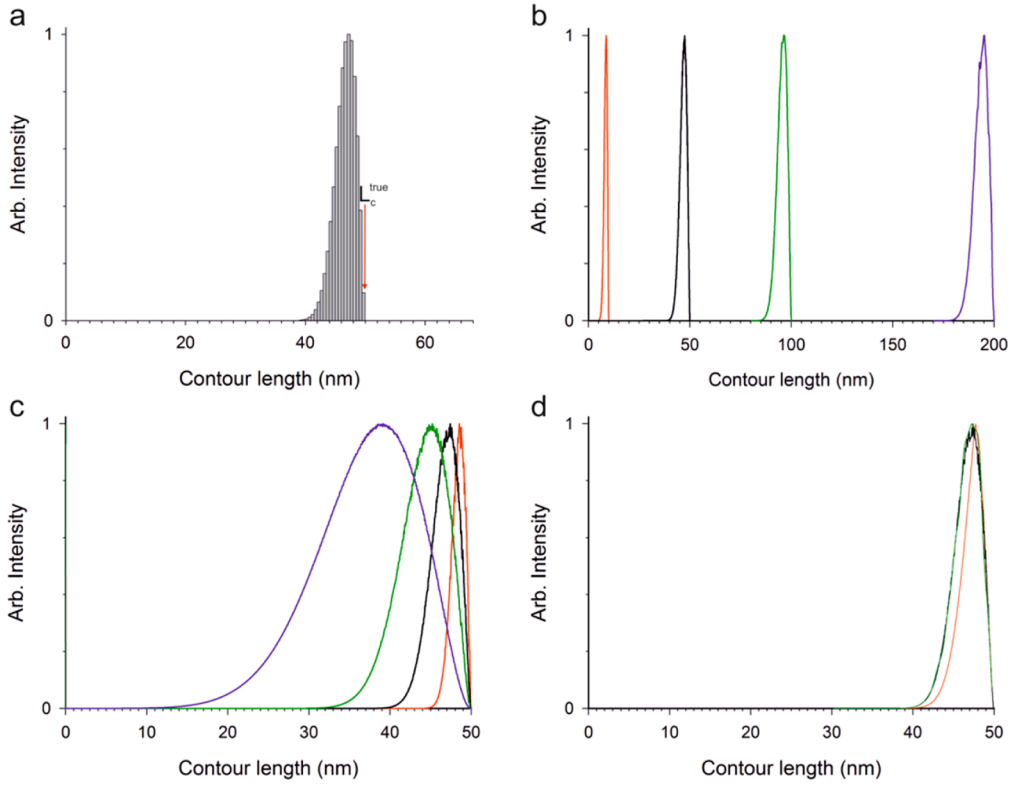



**Figure 3. Examples of contour length distributions generated using the described model. (a) Example of a modeled distribution where persistence length, $L_c^{true}$, and the AFM tip angle were set at 0.4 nm, 50 nm, and 60°, respectively. (b) Four examples of modeled distributions where persistence length and AFM tip angle were held constant at 0.4 nm and 60°, respectively, and the value for $L_c^{true}$ was varied as 10 nm (red), 50 nm (black), 100 nm (green), and 200 nm (blue). (c) Four examples of modeled distributions where $L_c^{true}$ and AFM tip angle were held constant at 50 nm and 60°, respectively, and the persistence length of the linkers was varied as 0.1 nm (red), 0.4 nm (black), 1.2 nm (green), and 5 nm (blue). (d) Three examples of modeled distributions where persistence length and $L_c^{true}$ were held constant at 0.4 and 50 nm, respectively, and the AFM tip angle was varied as 5° (red), 60° (black), and 160° (green).**

of 0.4 nm, and an $L_c^{true}$ of 50 nm is shown in Figure 3a. The generated distribution is asymmetric about the modal contour length ($L_c^{mod}$ = 47 nm) with a "tail" toward the lower contour length side of the distribution. Importantly, $L_c^{true}$ does not correspond to the modal value, as normally assumed, but is the highest observable contour length. In this example the difference between these values ($L_c^{diff}$) is 3 nm (equivalent to ~8 amino-acid residues), but is strongly dependent on total linker length (see below).

For the model to be a useful experimental tool it is necessary to understand how experimental variables modify the shape and position of the distribution of $L_c$ values. Figure 3b,c,d reveal the effect of linker length, persistence length, and AFM tip angle, respectively, on the contour length distributions obtained. In Figure 3b the numbers of monomers ($N_A$ and $N_B$) were selected to give total combined linker lengths of 10 nm (red), 50 nm (black), 100 nm (green), or 200 nm (blue) while keeping the persistence length (0.4 nm) and tip angle (60°) constant. Expectedly, longer-length linkers give rise to the observation of longer contour lengths. Also, as the total linker length is increased, there is an increase in the width of the generated distributions and an increase in $L_c^{diff}$ (1.3, 2.6, 3.8, and 4.8 nm for linker lengths of 10, 50, 100, and 200 nm). This observation is important, as the linker length (but not the chemical nature of the linker) is often varied experimentally to optimize the position of the unbinding events of interest in a force–extension profile. The persistence length was found to have the greatest effect on the position and width of the contour length distribution. Figure 3c shows the effect of increasing the persistence lengths for both linkers from 0.1 nm (red distribution) to 0.4 nm (black), 1.2 nm (green), and 5 nm (blue), while the AFM tip angle and total combined linker length were held constant (60° and 50 nm, respectively). Comparison of these generated distributions shows that, for linkers consisting of monomers with a longer persistence length, broader contour length distributions are predicted with a corresponding decrease in $L_c^{mod}$. This results from the fact that the longer the persistence length, the smaller the number of WLC segments constituting a given length of polymer chain. This is akin to a random walk where the size of the step has increased and the number of steps has decreased, giving rise to a larger average journey length or amount of space explored, which are directly proportional to step size (persistence length) and the square root of the number of steps (number of

monomers). While PEG linkers of differing length have identical persistence lengths, it is noteworthy that $L_c^{mod}$ will be significantly different for identical complexes immobilized by linkers of identical length but different persistence lengths. For example, if 100 nm PEG and ssDNA handles were used for immobilization of the same complex, $L_c^{mod}$ would be 96.2 and 94.7 nm, respectively (persistence length for PEG and ssDNA is 0.4 and 0.75 nm, respectively[35,37,38]). The effect of AFM tip angle was also investigated, as the aspect ratio of AFM tips varies depending on the type of cantilever and manufacturer. Figure 3d shows three example distributions where the tip angle was varied between 5° and 160° while maintaining the persistence length and total linker length at 0.4 and 50 nm, respectively. No clear difference in the generated distributions is observed when varying the AFM tip angle. This arises because there is a linear relationship between the distance along the AFM tip away from its apex and the surface area available for linker attachment. If the shape of the tip is changed to a geometry different from that of a perfect pyramid, this may no longer be the case (see Supporting Information for a description of the semispherical tip).

**Comparison of Experimental and Simulated Contour Length Distributions.** After understanding how experimental variables alter the predicted contour length distributions, the model was next benchmarked against SMFS dissociation data obtained for three different protein:protein or protein:peptide complexes (Figure 4): E9:Im9 (a typical protein:protein complex that covers a large surface area (1575 Å$^2$, see refs 39−41) (Figure 4a); TolB:Tol binding epitope (TBE) (an interaction between a protein and a 16-residue peptide) (Figure 4b);[42] and B1-8 $F_{ab}$:4-hydroxy-3-iodo-5-nitrophenylhexamine (NIP) (an interaction of an antibody fragment with its antigen) (Figure 4c).[43] For E9:Im9 and Tol B:TBE each partner was immobilized to the AFM tip and substrate by a standard surface derivitization protocol (Methods)[9] using NHS-$(PEG)_n$-maleimide linkers (succinimidyl-[*N*-maleimidopropionamido]-(ethylene glycol)$_n$ ester, $n$ = 12, 24, 70 ± 7, and 265 ± 27). To study the B1-8 $F_{ab}$:NIP interaction, B1-8 $F_{ab}$ was immobilized directly onto the derivatized surface after activation by EDC/sulfo-NHS (Methods). Nitrophenylhexamine was immobilized onto a mercapto-derivatized surface *via* the amino moiety using the same linkers described for E9:Im9 and TolB:TBE.

To obviate biased contour length selection, all bond rupture events (*i.e.*, those due to real single and multiple events and those due to nonspecific interactions) in all force−extension profiles were analyzed by extraction of an unbinding force and an $L_c$ obtained by fitting a WLC (persistence length fixed at 0.4 nm). Example force−extension profiles showing nonspecific and specific rupture events observed for TolB:TBE are shown in Figure 5a and b, respectively. These data were then used to generate $L_c$ frequency histograms (Figure 5c,d,e,f) and force−$L_c$ contour plots (inset in Figure 5c,d,e,f). These data and their comparison to controls measured in the presence of an excess of one binding partner in solution (Supplementary Figure S2) demonstrate that specific interactions can be detected readily above a background of nonspecific events.

Figure 5c shows that the predicted (black line) contour length distribution for E9:Im9 unbinding events closely matches the observed distribution (red bars) in terms of the modal value and width of the distribution using values of 60°, 17.8 nm, and 0.4 nm for tip angle, total combined linker and protein length (Figure 4a for protein length), and persistence length, respectively. In this case, use of a relatively short linker leads to an $L_c^{diff}$ of 1.8 nm. The similarity of $L_c^{mod}$ obtained by experiment and simulation (15 and 16 nm) suggests that unbinding occurs from a native-like state.

We next studied the TolB:TBE interaction (Figure 4b, Figure 5a,b,d) using longer linkers with a molecular mass of 12000 ± 1200 Da corresponding to 267 ± 27 PEG units. At first glance the contour length

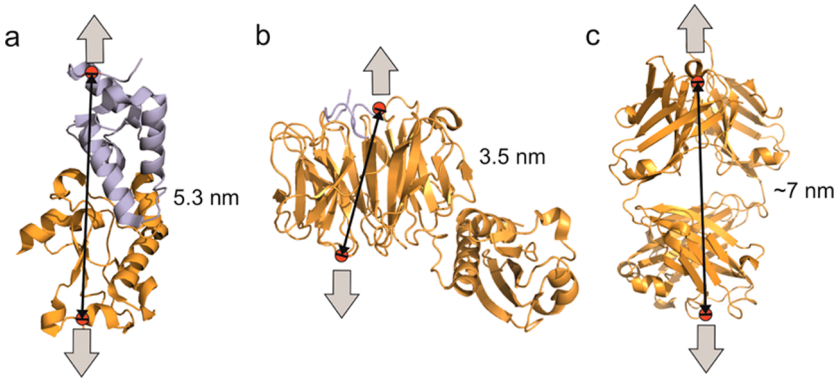



**Figure 4. Through-space distances for pulling locations used for presented SMFS data overlaid onto crystal structures. (a) E9:Im9 pulling geometry (PDB: 1EMV[41]). E9 and Im9 proteins are colored gold and lavender, respectively. (b) TolB:TBE pulling geometry (PDB: 2IVZ[52]). TolB and TBE are colored gold and lavender, respectively. (c) Pulling geometry applied for B1-8 $F_{ab}$:NIP where the pulling location for B1-8 $F_{ab}$ has been arbitrarily chosen to be the C-terminus of the heavy-chain fragment. (Note: the structure of B1-8 $F_{ab}$:NIP has not been solved. Structure of a typical $F_{ab}$ fragment is shown (PDB: 1YNL).[44])**

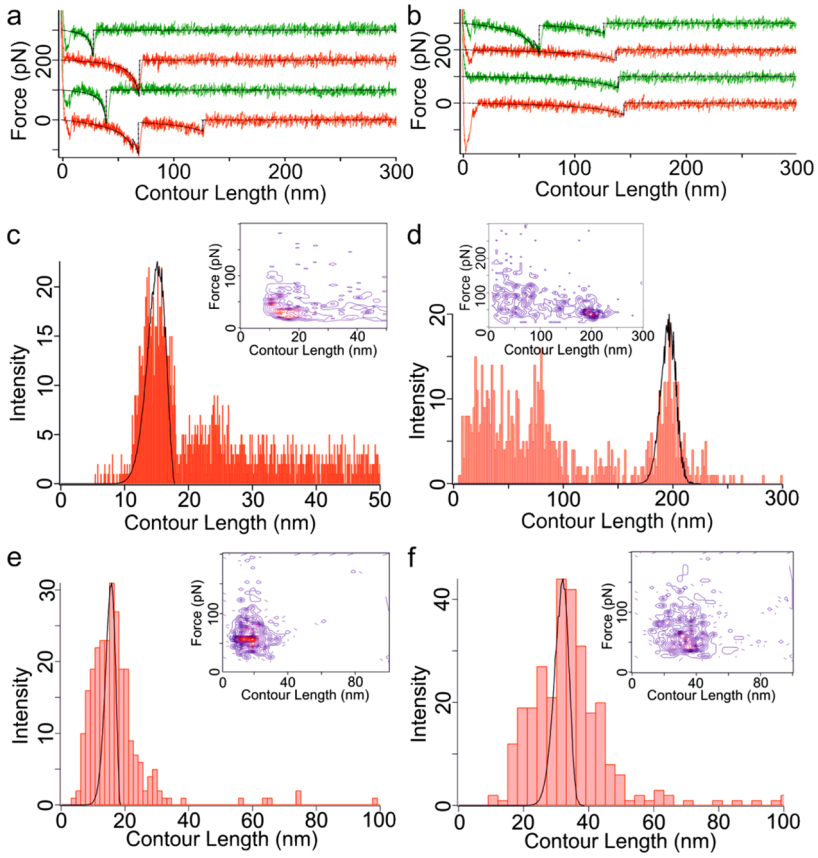



**Figure 5. (a) Typical force curves corresponding to nonspecific interactions from TolB:TBE data. (b) Typical force curves corresponding to specific interactions between TolB and TBE. (c) Contour length distribution for AFM SMFS data obtained for E9:Im9 (red distribution) compared with modeled distribution (black line). Modeled values for $L_c^{true}$, persistence length, and tip angle were input as 17.8 nm, 0.4 nm, and 60°, respectively. Inset: Contour plot showing relative frequency (color scale from blue (low) to red (high)) of events with specific force and contour length. (d) Contour length distribution for AFM SMFS data obtained on the TolB:TBE protein−peptide interaction (red distribution) compared with modeled distribution (black line). Modeled values for $L_c^{true}$, persistence length, and tip angle were input as 205.5 nm, 0.4 nm, and 60°, respectively. Inset: Contour plot showing relative frequency of events with specific force and contour length. (e and f) Contour length distributions for AFM SMFS data obtained on the B1-8 $F_{ab}$:NIP interaction (red distributions) compared with modeled distribution (black line) using two different length PEG linkers ((e) 18.5 nm and (f) 37.4 nm). Modeled values for persistence length and tip angle were input as 0.4 nm and 60°, respectively.**

distribution (Figure 5d) appears complex, but taken together with the force−contour length scattergram (insets Figure 5d), the data reveal a single specific unbinding event (at approximately 200 nm and 40 pN) and a series of nonspecific events at shorter extensions over a range of forces. Once more, excellent agreement is observed between experimental and simulated distributions for the specific unbinding events, but both are significantly wider than that observed for E9:Im9 (approximately 9 and 40 nm for the simulated E9:Im9 and TolB:TBE interactions, respectively). This broadening is due to increased linker length (see Figure 3) and the polydispersity of the linkers (accounted for in the model, see Supporting Information). Similar to E9:Im9 the $L_c^{mod}$ values obtained by simulation and experiment are 200 and 197 nm, respectively, suggesting the complex is native-like at dissociation.

As a final test of the model, we characterized the forced dissociation of an antigen (NIP) from an $F_{ab}$ fragment of an antibody (B1-8).[43] The antigen was immobilized *via* a short monodisperse PEG linker (24 units) or a longer PEG linker (70 units) with 10% dispersity (Figures 5e and f, respectively). No linkers were used for B1-8 $F_{ab}$, which was immobilized to an APTES-treated substrate *via* EDC activation of solvent-exposed carboxylate groups and converted to an amine-reactive species by inclusion of sulfo-NHS (Methods). In this case a poorer correlation between simulated and experimental data for longer lengths was observed (simulated *versus* experimental $L_c^{mod}$ are 17 and 16 nm and 35 and 30 nm for short and long linkers, respectively, in Figure 5e and f). As no structure of this complex is available, the distance between linker attachment positions was estimated to be 7 nm (from the C-terminal carboxyl group of the heavy chain to the center of the complementarity-determining region) using the high-resolution structure of an arbitrary $F_{ab}$:antigen complex.[44] In this case, as B1-8 $F_{ab}$:NIP can be assumed to be native-like upon rupture (see discussion in ref 9), the disparity

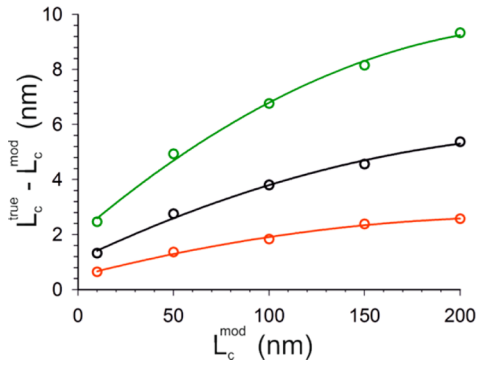



**Figure 6. Empirically determined relationship between distribution width ($L_c^{true} - L_c^{mod}$) and $L_c^{mod}$ for persistence lengths of 0.1, 0.4, and 1.2 nm shown in red, black, and green, respectively. Lines are included as guides for the eyes.**

between experimental and modeled $L_c^{mod}$ reveals either poor structural homology between the experimental and reference $F_{ab}$:antigen complex, the presence of other immobilization sites in addition to the C-terminus, or both. Additionally, the experimental and simulated distributions are significantly wider than that found for E9:Im9 and TolB:TBE. When the complex is immobilized by shorter linkers, broadening is due to the heterogeneity in $L_c$ introduced by the availability of many potential immobilization points on B1-8 $F_{ab}$ (acidic side-chains and the C-terminus). Nonspecific immobilization of B1-8 $F_{ab}$ is also probably the major factor for the simulated $L_c^{mod}$ value being greater than the observed value, as the C-terminus of the heavy chain is more distal to the antigen binding than all other potential immobilization sites. The use of longer linkers of less defined length (Figure 5f) further broadens the contour length distribution as described for the TolB:TBE complex above. Comparison of experimental and simulated $L_c$ distributions thus reveals greater $L_c$ heterogeneity and a less specific immobilization strategy used for B1-8 $F_{ab}$:NIP, demonstrating the power of the model to highlight subtle effects. The three biomolecular complexes that have been studied are diverse in terms of the type of interaction involved in complex stabilization, the molecular weight of the complexes, surface area of interaction, and the immobilization method. The ability of a simple model to accurately predict the experimental contour length distributions of these diverse systems suggests that it accurately captures the key determinants of contour length and shows that the presence of a dynamic ensemble of linker end-to-end lengths plays an important role in sculpting the distribution of contour lengths observed. While these data show that the actual $L_c$ of the complex and scaffold is always larger than the often cited $L_c^{mod}$, the latter value is more straightforward to measure for often nonideal experimentally derived distributions. To obviate this problem, the model was used to calculate $L_c^{diff}$ as a function of linker length using a tip angle of 60° and persistence lengths of 0.1, 0.4, and 1.2 nm. Figure 6 shows this relationship and demonstrates that for longer linkers, $L_c^{diff}$ is relatively large.

## CONCLUSION

Measured contour length values from SMFS AFM-based experiments are generally used to confirm that a specific interaction of interest is being measured and to infer low-resolution information on the interaction surface and/or unbinding pathway. For both of these applications it is necessary to extract accurate $L_c^{true}$ values. When filtering data, an event is typically accepted when the observed contour length falls close to an expected contour length. The range of contour lengths that are often reported as corresponding to the same specific interaction can, however, be as large as 100 nm in some cases.[17,45] In the work presented here a model has been derived that establishes a physical basis for predicting $L_c$ values and their distribution from AFM-based SMFS experiments. By considering the geometry of two idealized tethering surfaces and the probability of two idealized chains meeting, the model is able to predict contour length distributions from AFM-based SMFS pulling experiments. Often the distribution of measured contour lengths is not shown or discussed when SMFS data are presented in the literature. For those cases where they are presented and where a similar linker chemistry to that used in the experiments discussed above is employed, it is possible to simulate contour length distributions for comparison with the presented distributions. In a number of cases good agreement with the model discussed in this work is observed.[20,46,47] In the case of refs 20 and 47 a single peak in the observed contour length distribution is present in the same location and with the same approximate width as that generated using the model. This is also true in the case of ref 46, where, in addition, the shape of the presented distribution shows a similar asymmetry to modeled distributions.

As the model allows the precise shape and width of the contour length distribution to be predicted, analysis and interpretation of contour length distributions become feasible. For example, the derived model could be used to set limits on a range of contour lengths that correspond to a specific unbinding event and hence allow the identification of multiple specific unbinding events within a single contour length distribution. Currently, identification of multiple, distinct events from a single contour length distribution is not performed in accordance with any physical model. As such, the interpretation that individual events correspond to a specific interaction (such as those identified in refs 17, 21, and 48) could be disputed. The model presented here provides justification for such interpretations of contour length distributions. Further to this, deviations of experimental data from modeled

distributions (in terms of both shape and position of the $L_c$ distribution) should allow presumed values of particular variables to be tested directly. For example, where unfolding or remodeling of a portion of a protein assumed to be structured in the estimation of $L_c^{true}$ takes place, a shift in the observed contour length distribution to length values higher than those predicted from the model should be observed. As such, the model allows detection and quantification of any protein unfolding prior to unbinding to be measured and interpreted accurately in terms of the number of amino-acid residues involved. Additionally, the sensitivity of the contour length distributions to the persistence length of the tether in this model may provide a method to quantify this parameter.

Overall, therefore, the model derived in this work provides the experimentalist with a means of interpreting contour length distributions for unbinding events from AFM-based SMFS measurements within a physical framework. For a particular pair of single-molecule linkers (with associated length and persistence length) the model is able to generate an ideal contour length distribution. Comparison of this with experimentally determined contour length distributions allows individual and distinct events to be extracted from an ensemble and the nature of these events (*i.e.*, number of amino acids involved) to be elucidated.

## METHODS

**Proteins.** *E9:Im9.* Colicin E9 and its immunity protein (Im9) were provided purified and immobilized as described previously.[9] Briefly, both proteins are pseudo-wild-type variants containing a single solvent-exposed cysteine residue (S3C for E9 and S81C for Im9) to allow immobilization to an amine-derivatized surface *via* a maleimide/NHS succinimidyl heterobifunctional polyethylene glycol linker.

*TolB:Tol B Binding Epitope.* The pseudo-wild-type variant of TolB containing a hexahistidine tag and a single solvent-exposed cysteine (S299C) for immobilization as described for E9:Im9 above was provided purified as described previously.[49] TBE peptide with the addition of a C-terminal cysteine residue (GASDGSGWSSENNPWGGGSGSGC) was obtained from a commercial source (Peptide Synthetics).

*B1-8 Fab:Nitrophenol.* B1-8 Fab was provided purified as described previously.[43]

**AFM Measurements.** SMFS measurements were performed on an Asylum 3D AFM using functionalized MLCT cantilevers (Bruker) whose spring constants were determined using the thermal method.[50,51] All SMFS measurements were obtained at a retraction velocity of 1 $\mu$m s$^{-1}$, and data were collected at a rate of 20 kHz. Total retraction distance was set to be three times greater than the expected $L_c^{true}$ value for the particular interaction under investigation. For each complex, forced unbinding experiments were performed in the respective reaction buffer described below.

**Surface Functionalization.** Silicon substrates and silicon nitride AFM tips were first cleaned using piranha solution (3:1 sulfuric acid to 30% hydrogen peroxide) followed by irradiation under UV (254 nm) for 30 min. These surfaces were then held under vacuum in the presence of 20 $\mu$L of *N*,*N*-diisopropylethylamine (DIPEA) and 80 $\mu$L of (3-aminopropyl)triethoxysilane (APTES) (for all functionalizations except NIP) or 20 $\mu$L of DIPEA and 80 $\mu$L of (3-mercaptopropyl)trimethoxysilane (MPTES) (for NIP functionalization) for a period of 2 h. Following this, the chemicals were removed and the surfaces were left to cure for 24 h under a nitrogen atmosphere.

For E9:Im9 and TolB:TBE SMFS experiments, the APTES-treated surfaces were reacted with a heterobifunctional PEG linker (NHS-(PEG)-maleimide) (linkers containing 12 or 24 PEG domains were obtained from Pierce, and longer linkers with average $M_w$ = 3.4 or 12 kDa were obtained from Nanocs) by incubating the surfaces in 1 mL of chloroform containing 15 $\mu$L of 250 mM PEG linkers in DMSO for 1 h. The surfaces were then washed with chloroform and dried under nitrogen. Reaction of these linker-functionalized surfaces to the relevant protein/peptide was performed under aqueous reaction buffer (phosphate-buffered saline (PBS) for E9:Im9 and 50 mM HEPES, 50 mM NaCl, 5 mM $CaCl_2$ for TolB:TBE) by incubating the surface with the protein/peptide at a concentration of 1 mg mL$^{-1}$ for 30 min, followed by washing with reaction buffer.

For B1-8 $F_{ab}$:NIP SMFS experiments, NIP was immobilized by first reacting the MPTES-treated surfaces with NHS-(PEG)-maleimide in the same manner as APTES-treated surfaces described above. NIP-hexamine (Biosearch Technologies) in PBS at 1 mg mL$^{-1}$ was then incubated with the linker-functionalized surface for 30 min. The B1-8 $F_{ab}$ was attached directly to the surface by first converting surface-exposed carboxyl groups (C-termini and acidic side-chains) to an amine-reactive ester by reaction with EDC/Sulfo-NHS. Then 0.2 mg of EDC and 0.6 mg of sulfo-NHS were added to 0.5 mL of 1 mg mL$^{-1}$ B1-8 $F_{ab}$ in 0.1 M MES and 0.5 M NaCl at pH 6.0 for 15 min. Then 0.7 $\mu$L of 2-mercaptoethanol was added to quench the EDC. The amine-reactive B1-8 $F_{ab}$ was then exchanged into reaction buffer (PBS) and incubated with an APTES-treated surface for 30 min at room temperature.

These functionalized substrates and AFM probes were then used to perform the SMFS experiments described.

*Conflict of Interest:* The authors declare no competing financial interest.

*Acknowledgment.* This work was funded by the BBSRC (BB/G019452/1). The authors would like to thank J. Gowdy for assistance with Python code, C. Kleanthous and R. Kaminska at Oxford University for supplying the E9, Im9, and TolB proteins, and P. Tolar at the MRC National Institute for Medical Research for providing B1-8 $F_{ab}$.

*Supporting Information Available:* Derivation of eq 4. Alternative semispherical AFM tip approximation incorporated into the model and discussed. Polydispersity effect also incorporated into the model. This material is available free of charge *via* the Internet at http://pubs.acs.org.

## REFERENCES AND NOTES

1. Zlatanova, J.; Lindsay, S. M.; Leuba, S. H. Single Molecule Force Spectroscopy in Biology Using the Atomic Force Microscope. *Prog. Biophys. Mol. Biol.* **2000**, *74*, 37–61.
2. Hugel, T.; Seitz, M. The Study of Molecular Interactions by AFM Force Spectroscopy. *Macromol. Rapid Commun.* **2001**, *22*, 989–1016.
3. Neuman, K. C.; Nagy, A. Single-Molecule Force Spectroscopy: Optical Tweezers, Magnetic Tweezers and Atomic Force Microscopy. *Nat. Methods* **2008**, *5*, 491–505.
4. Crampton, N.; Brockwell, D. J. Unravelling the Design Principles for Single Protein Mechanical Strength. *Curr. Opin. Struct. Biol.* **2010**, *20*, 508–517.
5. Merkel, R.; Nassoy, P.; Leung, A.; Ritchie, K.; Evans, E. Energy Landscapes of Receptor-Ligand Bonds Explored with Dynamic Force Spectroscopy. *Nature* **1999**, *397*, 50–53.
6. Bell, G. I. Models for Specific Adhesion of Cells to Cells. *Science* **1978**, *200*, 618–627.

7. Thomas, W. E.; Vogel, V.; Sokurenko, E. Biophysics of Catch Bonds. *Annu. Rev. Biophys.* **2008**, *37*, 399–416.
8. Kim, J.; Zhang, C. Z.; Zhang, X. H.; Springer, T. A. A Mechanically Stabilized Receptor-Ligand Flex-Bond Important in the Vasculature. *Nature* **2010**, *466*, 992–995.
9. Farrance, O. E.; Hann, E.; Kaminska, R.; Housden, N. G.; Derrington, S. R.; Kleanthous, C.; Radford, S. E.; Brockwell, D. J. A Force-Activated Trip Switch Triggers Rapid Dissociation of a Colicin from its Immunity Protein. *PLoS Biol.* **2013**, *11*, e1001489.
10. Doi, M.; Edwards, S. F. *Theory of Polymer Dynamics*; Clerndon Press: Oxford, 1986.
11. Flory, P. J. *Statistical Mechanics of Chain Molecules*; Interscience Publishers: New York, 1969.
12. Bertz, M.; Rief, M. Ligand Binding Mechanics of Maltose Binding Protein. *J. Mol. Biol.* **2009**, *393*, 1097–1105.
13. Carrion-Vazquez, M.; Oberhauser, A. F.; Fowler, S. B.; Marszalek, P. E.; Broedel, S. E.; Clarke, J.; Fernandez, J. M. Mechanical and Chemical Unfolding of a Single Protein: A Comparison. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3694–3699.
14. Peng, Q.; Li, H. B. Atomic Force Microscopy Reveals Parallel Mechanical Unfolding Pathways of T4 Lysozyme: Evidence for a Kinetic Partitioning Mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1885–1890.
15. Dougan, L.; Li, J. Y.; Badilla, C. L.; Berne, B. J.; Fernandez, J. M. Single Homopolypeptide Chains Collapse into Mechanically Rigid Conformations. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 12605–12610.
16. Sandal, M.; Valle, F.; Tessari, I.; Mammi, S.; Bergantino, E.; Musiani, F.; Brucale, M.; Bubacco, L.; Samori, B. Conformational Equilibria in Monomeric Alpha-Synuclein at the Single-Molecule Level. *PLoS Biol.* **2008**, *6*, 99–108.
17. Lv, Z. J.; Roychaudhuri, R.; Condron, M. M.; Teplow, D. B.; Lyubchenko, Y. L. Mechanism of Amyloid Beta-Protein Dimerization Determined Using Single-Molecule AFM Force Spectroscopy. *Sci. Rep.* **2013**, *3*, 2880.
18. Krasnoslobodtsev, A. V.; Shlyakhtenko, L. S.; Lyubchenko, Y. L. Probing Interactions within the Synaptic DNA-Sfil Complex by AFM Force Spectroscopy. *J. Mol. Biol.* **2007**, *365*, 1407–1416.
19. Tong, Z.; Mikheikin, A.; Krasnoslobodtsev, A.; Lv, Z.; Lyubchenko, Y. L. Novel Polymer Linkers for Single Molecule AFM Force Spectroscopy. *Methods* **2013**, *60*, 161–168.
20. Yan, C.; Yersin, A.; Afrin, R.; Sekiguchi, H.; Ikai, A. Single Molecular Dynamic Interactions between Glycophorin A and Lectin as Probed by Atomic Force Microscopy. *Biophys. Chem.* **2009**, *144*, 72–77.
21. Krasnoslobodtsev, A. V.; Peng, J.; Asiago, J. M.; Hindupur, J.; Rochet, J. C.; Lyubchenko, Y. L. Effect of Spermidine on Misfolding and Interactions of Alpha-Synuclein. *PLoS One* **2012**, 7.
22. Jiao, F.; Fan, H.; Yang, G.; Zhang, F.; He, P. Directly Investigating the Interaction between Aptamers and Thrombin by Atomic Force Microscopy. *J. Mol. Recognit.* **2013**, *26*, 672–678.
23. Dong, J. J.; Castro, C. E.; Boyce, M. C.; Lang, M. J.; Lindquist, S. Optical Trapping with High Forces Reveals Unexpected Behaviors of Prion Fibrils. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1422–1430.
24. Gao, X.; Qin, M.; Yin, P. G.; Liang, J. Y.; Wang, J.; Cao, Y.; Wang, W. Single-Molecule Experiments Reveal the Flexibility of a Per-Arnt-Sim Domain and the Kinetic Partitioning in the Unfolding Pathway under Force. *Biophys. J.* **2012**, *102*, 2149–2157.
25. Karsai, A.; Kellermayer, M. S. Z.; Harris, S. P. Cross-Species Mechanical Fingerprinting of Cardiac Myosin Binding Protein-C. *Biophys. J.* **2013**, *104*, 2465–2475.
26. Li, Y. D.; Lamour, G.; Gsponer, J.; Zheng, P.; Li, H. B. The Molecular Mechanism Underlying Mechanical Anisotropy of the Protein GB1. *Biophys. J.* **2012**, *103*, 2361–2368.
27. Shen, T.; Cao, Y.; Zhuang, S. L.; Li, H. B. Engineered Bi-Histidine Metal Chelation Sites Map the Structure of the Mechanical Unfolding Transition State of an Elastomeric Protein Domain GB1. *Biophys. J.* **2012**, *103*, 807–816.
28. Sadler, D. P.; Petrik, E.; Taniguchi, Y.; Pullen, J. R.; Kawakami, M.; Radford, S. E.; Brockwell, D. J. Identification of a Mechanical Rheostat in the Hydrophobic Core of Protein L. *J. Mol. Biol.* **2009**, *393*, 237–248.
29. Kotamarthi, H. C.; Sharma, R.; Narayan, S.; Ray, S.; Ainavarapu, S. R. K. Multiple Unfolding Pathways of Leucine Binding Protein (LBP) Probed by Single-Molecule Force Spectroscopy (SMFS). *J. Am. Chem. Soc.* **2013**, *135*, 14768–14774.
30. Hoffmann, T.; Tych, K. M.; Brockwell, D. J.; Dougan, L. Single-Molecule Force Spectroscopy Identifies a Small Cold Shock Protein as Being Mechanically Robust. *J. Phys. Chem. B* **2013**, *117*, 1819–1826.
31. Carrion-Vazquez, M.; Marszalek, P. E.; Oberhauser, A. F.; Fernandez, J. M. Atomic Force Microscopy Captures Length Phenotypes in Single Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11288–11292.
32. Zimmermann, J. L.; Nicolaus, T.; Neuert, G.; Blank, K. Thiol-Based, Site-Specific and Covalent Immobilization of Biomolecules for Single-Molecule Experiments. *Nat. Protoc.* **2010**, *5*, 975–985.
33. Fuhrmann, A.; Ros, R. Single-Molecule Force Spectroscopy: A Method for Quantitative Analysis of Ligand-Receptor Interactions. *Nanomedicine* **2010**, *5*, 657–666.
34. Oesterhelt, F.; Rief, M.; Gaub, H. E. Single Molecule Force Spectroscopy by AFM Indicates Helical Structure of Poly-(Ethylene-Glycol) in Water. *New J. Phys.* **1999**, *1*, 6.1–6.11.
35. Kienberger, F.; Pastushenko, V. P.; Kada, G.; Gruber, H. J.; Riener, C.; Schindler, H.; Hinterdorfer, P. Static and Dynamical Properties of Single Poly(Ethylene Glycol) Molecules Investigated by Force Spectroscopy. *Single Mol.* **2000**, *1*, 123–128.
36. Wang, K.; Forbes, J. G.; Jin, A. J. Single Molecule Measurements of Titin Elasticity. *Prog. Biophys. Mol. Biol.* **2001**, *77*, 1–44.
37. Bosco, A.; Camunas-Soler, J.; Ritort, F. Elastic Properties and Secondary Structure Formation of Single-Stranded DNA at Monovalent and Divalent Salt Conditions. *Nucleic Acids Res.* **2014**, *42*, 2064–2074.
38. Ainavarapu, R. K.; Brujic, J.; Huang, H. H.; Wiita, A. P.; Lu, H.; Li, L. W.; Walther, K. A.; Carrion-Vazquez, M.; Li, H. B.; Fernandez, J. M. Contour Length and Refolding Rate of a Small Protein Controlled by Engineered Disulfide Bonds. *Biophys. J.* **2007**, *92*, 225–233.
39. Keeble, A. H.; Kleanthous, C. The Kinetic Basis for Dual Recognition in Colicin Endonuclease-Immunity Protein Complexes. *J. Mol. Biol.* **2005**, *352*, 656–671.
40. Li, W.; Hamill, S. J.; Hemmings, A. M.; Moore, G. R.; James, R.; Kleanthous, C. Dual Recognition and the Role of Specificity-Determining Residues in Colicin E9 Dnase-Immunity Protein Interactions. *Biochemistry* **1998**, *37*, 11771–11779.
41. Kuhlmann, U. C.; Pommer, A. J.; Moore, G. R.; James, R.; Kleanthous, C. Specificity in Protein-Protein Interactions: the Structural Basis for Dual Recognition in Endonuclease Colicin-Immunity Protein Complexes. *J. Mol. Biol.* **2000**, *301*, 1163–1178.
42. Papadakos, G.; Wojdyla, J. A.; Kleanthous, C. Nuclease Colicins and their Immunity Proteins. *Q. Rev. Biophys.* **2012**, *45*, 57–103.
43. Natkanski, E.; Lee, W. Y.; Mistry, B.; Casal, A.; Molloy, J. E.; Tolar, P. B Cells Use Mechanical Energy to Discriminate Antigen Affinities. *Science* **2013**, *340*, 1587–1590.
44. Gokulan, K.; Khare, S.; Ronning, D. R.; Linthicum, S. D.; Sacchettini, J. C.; Rupp, B. Co-Crystal Structures of NC6.8 Fab Identify Key Interactions for High Potency Sweetener Recognition: Implications for the Design of Synthetic Sweeteners. *Biochemistry* **2005**, *44*, 9889–9898.
45. Kudera, M.; Eschbaumer, C.; Gaub, H. E.; Schubert, U. S. Analysis of Metallo-Supramolecular Systems Using Single-Molecule Force Spectroscopy. *Adv. Funct. Mater.* **2003**, *13*, 615–620.
46. Morfill, J.; Blank, K.; Zahnd, C.; Luginbuhl, B.; Kuhner, F.; Gottschalk, K. E.; Pluckthun, A.; Gaub, H. E. Affinity-Matured Recombinant Antibody Fragments Analyzed by Single-Molecule Force Spectroscopy. *Biophys. J.* **2007**, *93*, 3583–3590.

47. Yersin, A.; Osada, T.; Ikai, A. Exploring Transferrin-Receptor Interactions at the Single-Molecule Level. *Biophys. J.* **2008**, *94*, 230–240.
48. Krasnoslobodtsev, A. V.; Volkov, I. L.; Asiago, J. M.; Hindupur, J.; Rochet, J. C.; Lyubchenko, Y. L. Alpha-Synuclein Misfolding Assessed with Single Molecule AFM Force Spectroscopy: Effect of Pathogenic Mutations. *Biochemistry* **2013**, *52*, 7377–7386.
49. Papadakos, G.; Housden, N. G.; Lilly, K. J.; Kaminska, R.; Kleanthous, C. Kinetic Basis for the Competitive Recruitment of TolB by the Intrinsically Disordered Translocation Domain of Colicin E9. *J. Mol. Biol.* **2012**, *418*, 269–280.
50. Hutter, J. L.; Bechhoefer, J. Calibration of Atomic-Force Microscope Tips. *Rev. Sci. Instrum.* **1993**, *64*, 1868–1873.
51. Butt, H. J.; Jaschke, M. Calculation of Thermal Noise in Atomic-Force Microscopy. *Nanotechnology* **1995**, *6*, 1–7.
52. Loftus, S. R.; Walker, D.; Mate, M. J.; Bonsor, D. A.; James, R.; Moore, G. R.; Kleanthous, C. Competitive Recruitment of the Periplasmic Translocation Portal TolB by a Natively Disordered Domain of Colicin E9. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12353–12358.