

Modular, Nondegenerate Polyprotein Scaffolds for Atomic Force Spectroscopy

By: Zackary N. Scholl, [Eric A. Josephs](#), and Piotr E. Marszalek

Scholl, Z. N.,[‡] Josephs, E. A.,[‡] Marszalek, P. E. “Modular, Nondegenerate Polyprotein Scaffolds for Atomic Force Spectroscopy.” *Biomacromolecules*, **2016**. *17* (7), 2502-2505. <https://doi.org/10.1021/acs.biomac.6b00548> [‡]Authors contributed equally to this work

This document is the Accepted Manuscript version of a Published Work that appeared in final form in *Biomacromolecules*, copyright © American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see <https://doi.org/10.1021/acs.biomac.6b00548>.

Abstract:

Atomic force spectroscopy (AFS) is a technique in which a cantilever probe of an atomic force microscope (AFM) is used to directly manipulate and pull on individual protein molecules. In an AFS experiment proteins of interest are deposited onto a surface and, after a protein comes into contact with the probe and is mechanically stretched between the surface and the probe, the forces used to unfold the proteins to an unstructured conformation can be directly determined (reviewed in refs 1–7). Such experiments can provide important insights into the energetics of that protein’s structure and protein folding or unfolding behavior.⁽⁸⁻¹⁴⁾

Keywords: polyproteins | atomic force spectroscopy | codon shuffling

Article:

Introduction

Atomic force spectroscopy (AFS) is a technique in which a cantilever probe of an atomic force microscope (AFM) is used to directly manipulate and pull on individual protein molecules. In an AFS experiment proteins of interest are deposited onto a surface and, after a protein comes into contact with the probe and is mechanically stretched between the surface and the probe, the forces used to unfold the proteins to an unstructured conformation can be directly determined (reviewed in refs 1–7). Such experiments can provide important insights into the energetics of that protein’s structure and protein folding or unfolding behavior.⁽⁸⁻¹⁴⁾

“Polyproteins” are polypeptide macromolecules that are composed of tandem identical repeats of protein domains which are each connected by a short linker.⁽¹⁾ Polyproteins are often used in AFS experiments because they provide several benefits. First, using a larger molecule increases the likelihood of contact between the cantilever probe and the protein. Second, since these incidences of contacts are rare, it is useful to have several proteins on the same molecule so that multiple measurements can be obtained from a single successful pulling event. Third, and most importantly, when these tandem identical repeats are used to flank another protein of interest, the polyprotein can serve (i) as pulling handles that protect the protein of interest from direct contact with the substrate and the AFM tip, and (ii) as a positive control such that the experimenter can

be certain that the probe is interacting with a single molecule of interest by the unique “force–extension” signature exhibited by the unfolding of the flanking repeats, which can then be differentiated by those produced by the protein of interest. Identical tandem repeats with well-characterized force–extension signatures that have been commonly used as polyproteins include titin I91 domain (formerly known as I27),^(12, 15) immunoglobulin G domain GB1,⁽⁸⁾ ubiquitin,⁽¹⁶⁾ fibronectin,⁽¹⁷⁾ protein G,⁽¹⁸⁾ and SNase,⁽¹⁹⁾ among others.

Though immensely useful, the experimental creation of polyproteins presents a major bottleneck in AFS experiments. To streamline this process, Steward et al. produced a plasmid containing repeating DNA cassettes that code for tandem repeat domains, each of which were separated by unique restriction sites, in a scheme that allows for the DNA coding for the protein of interest to replace individual cassettes with simple cloning techniques.⁽²⁰⁾ More recently, Hoffman et al. used a Gibson assembly cloning technique to combine DNA cassettes with specially designed linkers to assemble a DNA plasmid coding for a polyprotein,⁽²¹⁾ and Ott et al.⁽²²⁾ used a Golden Gate-based cloning technique to create tracts of repetitive DNA sequences with controllable lengths. However, in order to perform an AFS experiment successfully, the exact polypeptide sequences of the molecules of interest must be known precisely, while in practice it is very difficult to perform sequencing of the entire DNA sequence coding for the polyprotein to guarantee its fidelity because of the sequence degeneracy of the identical tandem repeat domains. These identical repeats are also subject to uncontrolled expansion or deletion by way of homologous recombination (e.g., as in Supporting Information of ref 13). Nonidentical protein repeats can be used to circumvent this problem,^(23, 24) although the force–extension signatures of these polyproteins can be more complex than those from polyproteins derived from a single identical protein, a complication that is nonideal for many force-spectroscopy experiments.

To eliminate this problem, we have improved on earlier polyprotein designs to generate a new plasmid backbone, pEMI91, that resolves these issues by using tandem repeats with “shuffled” codons^(25, 26) that can be easily sequenced (Figure 1). The plasmid for the polyprotein was designed to contain nine consecutive cassettes that would each translate to the I91 domain from *Homo sapiens* TTN gene titin. Each of these cassettes is separated by unique restriction sites for simple replacement of individual I91 cassettes with the DNA coding for a protein of interest, and by virtue of using “unique” I91-coding DNA sequences, a series of primers have been developed that can be used to sequence across the entire polyprotein (Table 1). The polyprotein itself is flanked by a His-tag and a Strep-tag to allow for simplified purification of the expressed protein and possesses two C-terminal cysteine residues to aid in protein display.

Table 1. Sequencing Primers for Nondegenerate I91 Polyprotein

sequencing primer name (in Figure 1)	DNA sequence
Seq1	5'-TAACACTAAGAGCGCCGCAA-3'
Seq2	5'-CAAAGAACTTCGCAGCGAGC-3'
Seq3	5'-GTAAAAGAACTGCGCTCGCT-3'
Seq4	5'-ATCTGAACCGGATGTCCACG-3'
Seq5	5'-GACTGGGGAGGTCAGCTTTC-3'
Seq6	5'-TGTCCTTTCAGCCGCTAAT-3'
Seq7	5'-ATTGCGTAGCGTCGACCTTA-3'

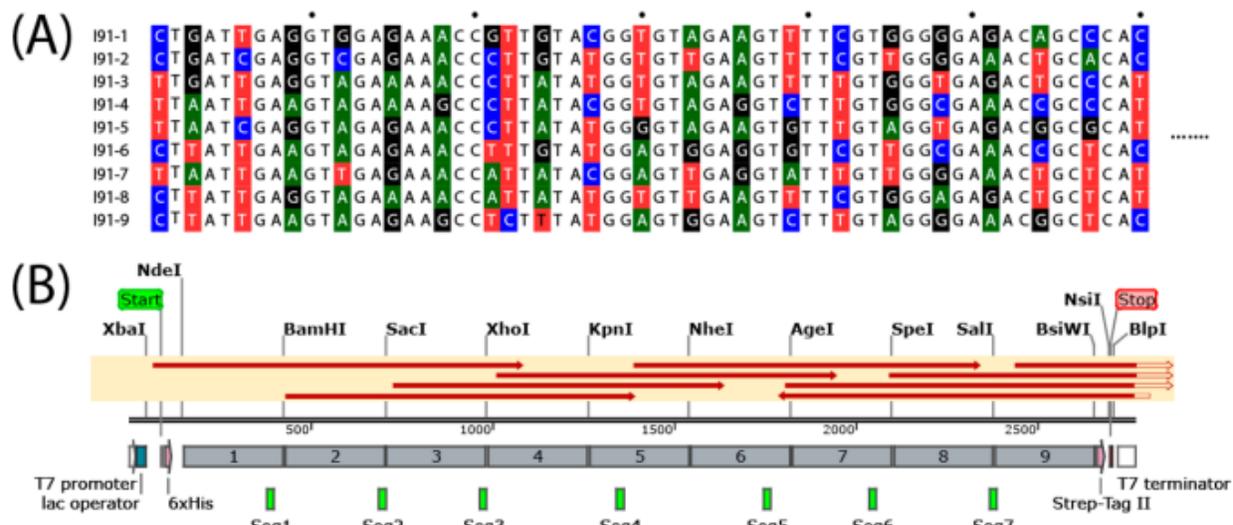


Figure 1. (A) The DNA sequences from the first 60 bp of each of the nine I91 cassettes colored by their differences to show the shuffling of the codons dispersed throughout the sequence. (B) Schematic of the I91 polyprotein design (gray blocks). The unique restriction sites between each cassette is shown, as well as the placement of a highly specific primer sequence (green blocks labeled Seq1, Seq2, ...). The sequencing coverage by the seven primers plus the T7 terminator and T7 promoter are shown as red arrows above (see also Supporting Information (SI)). The full sequence map is provided at AddGene.org.

Since recent studies on codon shuffling have revealed that modifying codons can affect protein expression,⁽²⁷⁾ protein structure,⁽²⁸⁾ and folding,⁽²⁹⁾ we have verified that our new plasmid has similar properties as earlier examples of titin I91 polyproteins by conducting standard force-spectroscopy experiments to determine that the loading-rate dependence, the distance to transition state, and the unfolding forces. The plasmid coding for this modular, nondegenerate polyprotein scaffold, known as pEMI91, has been deposited into the AddGene repository (#74888) to be made available for use by researchers.

Materials and Methods

Protein Purification

The engineered plasmid was transformed into *Escherichia coli* BL21(DE3)pLysS cells, and expression was induced using isopropyl β -d-thiogalactopyranoside. Cell lysate was run through a Strep-tag column (IBA, Goettingen, Germany) and the protein was dialyzed into phosphate-buffered saline (PBS) pH 7.4 and stored in 40% glycerol and 60% phosphate-buffered saline (PBS, pH 7.4) at -20 °C.

Sequencing

The full plasmid map was sequenced as shown in Figure 1 using the T7 promoter (5'-TAATACGACTCACTATAGGG-3') and the 5'-3' primers in Table 1, along with the 3'-5' T7

terminator sequencing primer (5'-GGCTTGGTTATGCCGGTACT-3'). Sanger sequencing was performed by Eton Biosciences, Inc.

Atomic Force Microscopy

Force-spectroscopy measurements were obtained using a custom-built AFM instrument.⁽³⁰⁾ Automation routines to control the AFM⁽³¹⁾ were implemented in LabView (National Instruments, Austin, Texas). Calibration of cantilever spring constants were done in the buffer solution using the energy equipartition theorem.⁽³²⁾ All measurements were performed in a PBS pH 7.4 solution at room temperature. Force spectroscopy experiments were performed using pulling rates of 50, 300, 1500, and 3000 nm/s using MLCT cantilevers (Bruker, Camarillo, CA) with the spring constant that varied between 16 and 150 pN/nm. In all experiments the purified protein was diluted to ~100 µg/mL in PBS and applied to recently evaporated gold and then incubated for an hour. A worm-like chain (WLC) model⁽³³⁾ with persistence length of 0.4 nm was fit to each peak in order to measure contour length increments in the force–extension (FE) data.

Data Analysis

Between 2000 and 70000 pulls were attempted for each loading rate, until ~20 high quality curves were obtained with more than three I91 domain unfolding events. The statistics of these attempts and the frequency of the number of domain unfolding events is shown in Supporting Information (Tables S1 and S2). Recordings that had nonspecific events (high force in the beginning of force curve) and recordings that contained events from multiple molecules (contour-length increments much less than prototypical 28 nm) were not used for analysis. Typically only 0.03–1.3% of total pulls resulted in a high quality force–extension curve, while ~63–75% were empty, and the rest contained nonspecific events or signatures of multiple molecules.

Unfolding forces from usable recordings were binned according to the Freedman-Diaconis rule⁽³⁴⁾ and loading rates dF/dt were calculated using a highly accurate approximation,⁽³⁵⁾

$$\frac{dF}{dt} = v \left[\frac{1}{k_c} + \frac{2\beta Lp(1 + \beta Fp)}{3 + 5\beta Fp + 8(\beta Fp)^{5/2}} \right]^{-1}$$

where F is the force, v is the ramp speed, k_c is the cantilever spring constant, β is the reciprocal of thermal energy $\left(\frac{1}{k_b T}\right)$, and p is the persistence length. A persistence length of 0.365 nm was used for the fitting.^(4, 36)

The histogram of unfolding forces was then combined with the loading rate information to generate the force-dependent unfolding rate data.^(37, 38) These data were fit using a model of force-induced unfolding over a barrier,^(38, 39) which yielded the unfolding rate, distance to transition state, and the associated errors of estimating these properties.

Results and Discussion

A new plasmid sequence containing tandem protein repeats was generated using the pET-15b plasmid sequence as a backbone (Figures 1A and S1, see also Supporting Information for full sequence). To generate nine unique DNA sequences (Supporting Information (SI)), we used amino acid sequence from I91 (UniProtKB Q8WZ42) and reverse translated and then codon shuffled using the Optimizer web service^(40, 41) using a random selection of codons weighted by their frequency of use in *E. coli*. The nine tandem I91 domain sequences were checked for sequence disparity using phylogenetic clustering⁽⁴²⁻⁴⁴⁾ which showed that the minimum proportion of substitutions between any pair of sequences was 8% (Figures 1B and SI). This high sequence disparity allowed individual primers for each cassette to be designed with high target specificity against the backbone pET-15b plasmid sequence (Table 1).⁽⁴⁵⁾ This polyprotein-coding DNA insert was synthesized by Genscript (Piscataway, NJ) and inserted into the pET-15b backbone.

We tested whether this sequence degeneracy was sufficient to allow us to unambiguously sequence the entire polyprotein-coding DNA using eight designed primers (Materials and Methods) in addition to primers derived from the sequences of the T7 promoter and T7 terminator sites that flank the I91 cassettes. Each primer was able to provide high fidelity sequencing results (Figure S2) up to 800 base pairs and the corresponding sequences matched the designed sequence as intended (red arrows in Figure 1B, see also SI). We verified the protein induction by purifying full-length proteins from the plasmid (details in Materials and Methods) which showed the expected size of ~100 kDa on a SDS-Gel (Figure 2A).

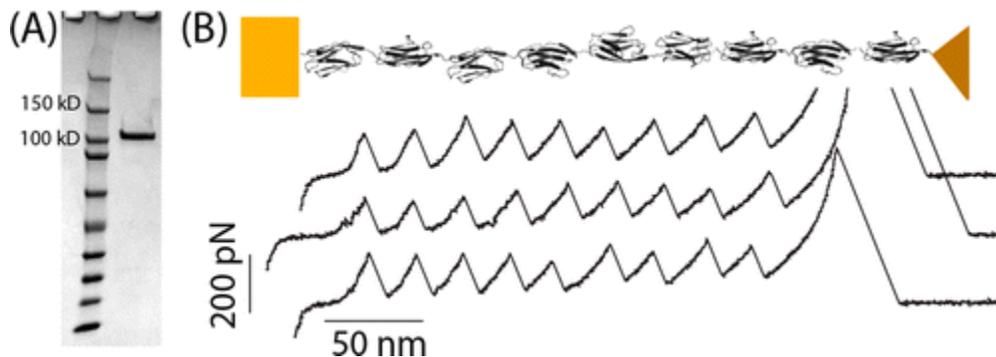


Figure 2. (A) SDS-gel of the denatured nine I91 domain polyprotein which runs at the expected size ~ 100 kDa. (B) Schematic of the pulling experiment where the I91 polyprotein is tethered to a gold substrate and pulled from the end by a cantilever, with representative examples of the force–extension signatures that result from unfolding of the nine tandem I91 domains.

To further verify that our shuffled-sequence polyprotein displayed correct mechanical and folding behavior identical to that of the wild-type I91 polyprotein, we performed standard force-spectroscopy experiments on the expressed polyprotein. The polyprotein comprised of nine tandem I91 domains was pulled at 300 nm/s using a cantilever with a spring constant of 16 pN/nm (more details in Materials and Methods). The schematic of the experiment and representative force–extension curves are shown in Figure 2B. The measured contour-length increment was 28.2 ± 2.2 nm (mean and standard deviation) and the unfolding force 202 ± 22 pN (Figure 3). These values match well the previously reported values of 28.4 ± 0.3 nm and unfolding force of ≈ 200 pN.⁽¹⁵⁾

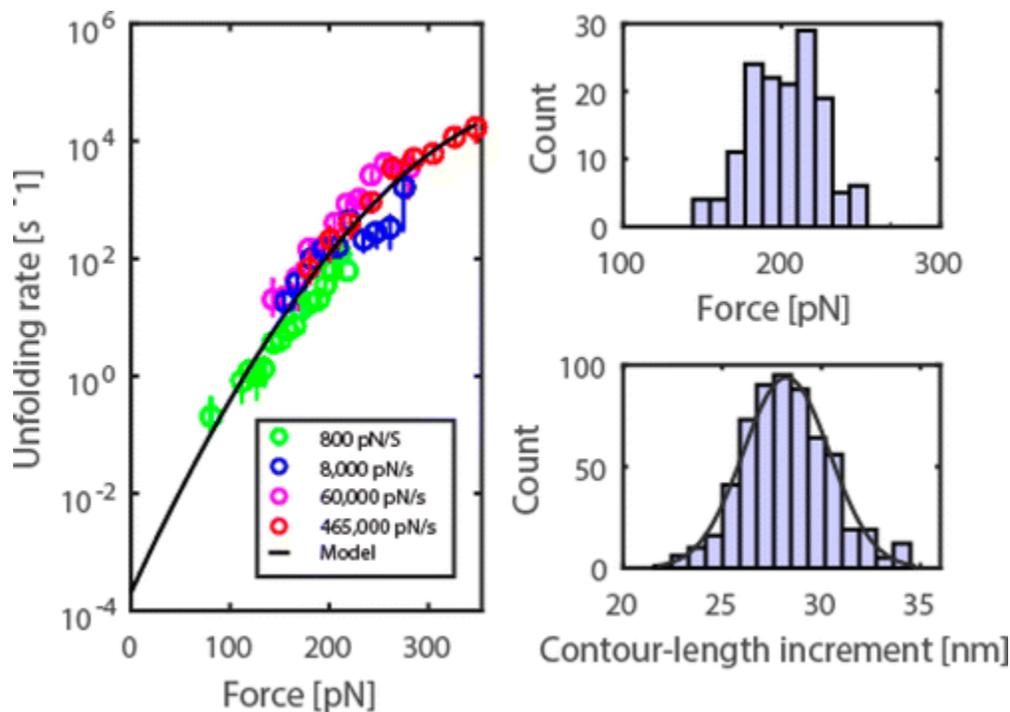


Figure 3. Left: The unfolding rate versus force data with error bars (dots) fit by a model of the force-induced unfolding (black line) with an intrinsic unfolding rate of $2.1 \times 10^{-4} s^{-1}$ and a distance to transition of 0.35 nm (see Materials and Methods) Right top: The unfolding force for experiments performed at 300 nm/s and spring constant of 16 pN/nm. Right bottom: The contour-length increment of the I91 domains for all experiments.

We also verified the loading-rate dependence of the I91 domains unfolding by performing experiments for an additional three different loading rates (distributions of unfolding forces in Figure S3). We followed established procedures for extracting the intrinsic unfolding rate and distance to transition state parameters (see Methods and Materials). The intrinsic unfolding rate determined is $2.1 \pm 1.2 \times 10^{-4} s^{-1}$ (mean and standard deviation) and the distance to transition state is 0.35 ± 0.03 nm. These values compare very well to previous measurements of an unfolding rate of $3.3 \times 10^{-4} s^{-1}$ and distance to transition state of 0.25–0.3 nm.^(15, 46)

Conclusions

A major bottleneck in AFS experiments is the cloning and creation of plasmid sequences for engineered polyproteins that flank a protein of interest, which must contain identical tandem domains to serve as positive control and tethering points. Here we present an improved plasmid backbone with a polyprotein insert containing tandem titin I91 domains with shuffled DNA codons. These domains translate to the prototypical I91 sequence, but allow facile sequencing through their sequence disparity and simple restriction digestion through the incorporation of unique restriction sites. We have verified that this sequence carries the expected properties of previous designs of polyproteins, with the added benefit of full facile sequencing. We expect similar strategies of codon shuffling can be used to generate different polyproteins of interest.

The nine I91 polyprotein insert was introduced into a plasmid which is available as a resource to other researchers (AddGene plasmid #74888).

Supporting Information

The Supporting Information is available at <https://doi.org/10.1021/acs.biomac.6b00548>.

Acknowledgment

This work is supported by the National Science Foundation (NSF) GRFP 1106401 and the Katherine Goodman Stern Fellowship to Z.N.S. and by the NSF MCB-1517245 and MCB-1244297 to P.E.M. P.E.M. acknowledges kind support from the Kimberly-Clark Corporation. E.A.J. is supported by the National Institute of General Medical Sciences of the National Institutes of Health (NIH; F32GM112502).

References

1. Crépin, T.; Swale, C.; Monod, A.; Garzoni, F.; Chaillet, M.; Berger, I. *Curr. Opin. Struct. Biol.* **2015**, *32*, 139 DOI: 10.1016/j.sbi.2015.04.007 [Google Scholar](#)
2. Ott, W.; Jobst, M. A.; Schoeler, C.; Gaub, H. E.; Nash, M. A. J. *Struct. Biol.* **2016**, DOI: 10.1016/j.jsb.2016.02.011 [Google Scholar](#)
3. Popa, I.; Kosuri, P.; Alegre-Cebollada, J.; Garcia-Manyes, S.; Fernandez, J. M. *Nat. Protoc.* **2013**, *8*, 1261 DOI: 10.1038/nprot.2013.056 [Google Scholar](#)
4. Scholl, Z. N.; Li, Q.; Marszalek, P. E. *Wiley Interdiscip. Rev.: Nanomed. and Nanobiotechnol* **2014**, *6*, 211 DOI: 10.1002/wnan.1253 [Google Scholar](#)
5. Tych, K. M.; Hoffmann, T.; Batchelor, M.; Hughes, M. L.; Kendrick, K. E.; Walsh, D. L.; Wilson, M.; Brockwell, D. J.; Dougan, L. *Biochem. Soc. Trans.* **2015**, *43*, 179 DOI: 10.1042/BST20140274 [Google Scholar](#)
6. Žoldák, G.; Rief, M. *Curr. Opin. Struct. Biol.* **2013**, *23*, 48 DOI: 10.1016/j.sbi.2012.11.007 [Google Scholar](#)
7. Hoffmann, T.; Dougan, L. *Chem. Soc. Rev.* **2012**, *41*, 4781 DOI: 10.1039/c2cs35033e [Google Scholar](#)
8. Cao, Y.; Li, H. *Nat. Mater.* **2007**, *6*, 109 DOI: 10.1038/nmat1825 [Google Scholar](#)
9. Garcia-Manyes, S.; Giganti, D.; Badilla, C. L.; Lezamiz, A.; Perales-Calvo, J.; Beedle, A. E.; Fernández, J. M. *J. Biol. Chem.* **2016**, *291*, 4226 DOI: 10.1074/jbc.M115.673871 [Google Scholar](#)
10. He, C.; Hu, C.; Hu, X.; Hu, X.; Xiao, A.; Perkins, T. T.; Li, H. *Angew. Chem.* **2015**, *127*, 10059 DOI: 10.1002/ange.201502938 [Google Scholar](#)
11. Kotamarthi, H. C.; Sharma, R.; Ainaravapu, S. R. K. *Biophys. J.* **2013**, *104*, 167a DOI: 10.1016/j.bpj.2012.11.939 [Google Scholar](#)

12. Rico, F.; Gonzalez, L.; Casuso, I.; Puig-Vidal, M.; Scheuring, S. *Science* **2013**, 342, 741 DOI: 10.1126/science.1239764 [Google Scholar](#)
13. Scholl, Z. N.; Yang, W.; Marszalek, P. E. *ACS Nano* **2015**, 9, 1189 DOI: 10.1021/nn504686f [Google Scholar](#)
14. Valle-Orero, J.; Eckels, E. C.; Stirnemann, G.; Popa, I.; Berkovich, R.; Fernandez, J. M. *Biochem. Biophys. Res. Commun.* **2015**, 460, 434 DOI: 10.1016/j.bbrc.2015.03.051 [Google Scholar](#)
15. Carrion-Vazquez, M.; Oberhauser, A. F.; Fowler, S. B.; Marszalek, P. E.; Broedel, S. E.; Clarke, J.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, 96, 3694 DOI: 10.1073/pnas.96.7.3694 [Google Scholar](#)
16. Brujić, J.; Walther, K. A.; Fernandez, J. M. *Nat. Phys.* **2006**, 2, 282 DOI: 10.1038/nphys269 [Google Scholar](#)
17. Li, L.; Huang, H. H.-L.; Badilla, C. L.; Fernandez, J. M. *J. Mol. Biol.* **2005**, 345, 817 DOI: 10.1016/j.jmb.2004.11.021 [Google Scholar](#)
18. Cao, Y.; Balamurali, M.; Sharma, D.; Li, H. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, 104, 15677 DOI: 10.1073/pnas.0705367104 [Google Scholar](#)
19. Wang, C.-C.; Tsong, T.-Y.; Hsu, Y.-H.; Marszalek, P. E. *Biophys. J.* **2011**, 100, 1094 DOI: 10.1016/j.bpj.2011.01.011 [Google Scholar](#)
20. Steward, A.; Toca-Herrera, J. L.; Clarke, J. *Protein Sci.* **2002**, 11, 2179 DOI: 10.1110/ps.0212702 [Google Scholar](#)
21. Hoffmann, T.; Tych, K. M.; Crosskey, T.; Schiffrin, B.; Brockwell, D. J.; Dougan, L. *ACS Nano* **2015**, 9, 8811 DOI: 10.1021/acsnano.5b01962 [Google Scholar](#)
22. Ott, W.; Nicolaus, T.; Gaub, H. E.; Nash, M. A. *Biomacromolecules* **2016**, 17, 1330 DOI: 10.1021/acs.biomac.5b01726 [Google Scholar](#)
23. Schlierf, M.; Rief, M. *J. Mol. Biol.* **2005**, 354, 497 DOI: 10.1016/j.jmb.2005.09.070 [Google Scholar](#)
24. Li, H.; Oberhauser, A. F.; Fowler, S. B.; Clarke, J.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, 97, 6527 DOI: 10.1073/pnas.120048697 [Google Scholar](#)
25. Tang, N. C.; Chilkoti, A. *Nat. Mater.* **2016**, 15, 419 DOI: 10.1038/nmat4521 [Google Scholar](#)
26. Mi, L. *Biomacromolecules* **2006**, 7, 2099 DOI: 10.1021/bm050158h [Google Scholar](#)
27. Gustafsson, C.; Govindarajan, S.; Minshull, J. *Trends Biotechnol.* **2004**, 22, 346 DOI: 10.1016/j.tibtech.2004.04.006 [Google Scholar](#)
28. Zhou, M.; Guo, J.; Cha, J.; Chae, M.; Chen, S.; Barral, J. M.; Sachs, M. S.; Liu, Y. *Nature* **2013**, 495, 111 DOI: 10.1038/nature11833 [Google Scholar](#)
29. Angov, E. *Biotechnol. J.* **2011**, 6, 650 DOI: 10.1002/biot.201000332 [Google Scholar](#)

30. Oberhauser, A. F.; Marszalek, P. E.; Erickson, H. P.; Fernandez, J. M. *Nature* **1998**, 393, 181 DOI: 10.1038/30270 [Google Scholar](#)
31. Scholl, Z. N.; Marszalek, P. E. *Ultramicroscopy* **2014**, 136, 7 DOI: 10.1016/j.ultramic.2013.07.020 [Google Scholar](#)
32. Florin, E. L.; Rief, M.; Lehmann, H.; Ludwig, M.; Dornmair, C.; Moy, V. T.; Gaub, H. E. *Biosens. Bioelectron.* **1995**, 10, 895 DOI: 10.1016/0956-5663(95)99227-C [Google Scholar](#)
33. Marko, J. F.; Siggia, E. D. *Macromolecules* **1995**, 28, 8759 DOI: 10.1021/ma00130a008 [Google Scholar](#)
34. Freedman, D.; Diaconis, P. *Probab. Theory Related Fields* **1981**, 57, 453 DOI: 10.1007/BF01025868 [Google Scholar](#)
35. Dudko, O. K.; Hummer, G.; Szabo, A. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105, 15755 DOI: 10.1073/pnas.0806085105 [Google Scholar](#)
36. Dietz, H.; Rief, M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, 103, 1244 DOI: 10.1073/pnas.0509217103 [Google Scholar](#)
37. Evans, E.; Halvorsen, K.; Kinoshita, K.; Wong, W. P. *Handbook of Single-Molecule Biophysics*; Springer, 2009; p 571. [Google Scholar](#)
38. Zhang, Y.; Dudko, O. K. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, 110, 16432 DOI: 10.1073/pnas.1309101110 [Google Scholar](#)
39. Dudko, O. K.; Hummer, G.; Szabo, A. *Phys. Rev. Lett.* **2006**, 96, 108101 DOI: 10.1103/PhysRevLett.96.108101 [Google Scholar](#)
40. Puigbò, P.; Guzmán, E.; Romeu, A.; Garcia-Vallvé, S. *Nucleic Acids Res.* **2007**, 35, W126 DOI: 10.1093/nar/gkm219 [Google Scholar](#)
41. Puigbò, P.; Romeu, A.; Garcia-Vallvé, S. *Nucleic Acids Res.* **2007**, 36, D524 DOI: 10.1093/nar/gkm831 [Google Scholar](#)
42. Goujon, M.; McWilliam, H.; Li, W.; Valentin, F.; Squizzato, S.; Paern, J.; Lopez, R. *Nucleic Acids Res.* **2010**, 38, W695 DOI: 10.1093/nar/gkq313 [Google Scholar](#)
43. McWilliam, H.; Li, W.; Uludag, M.; Squizzato, S.; Park, Y. M.; Buso, N.; Cowley, A. P.; Lopez, R. *Nucleic Acids Res.* **2013**, 41, W597 DOI: 10.1093/nar/gkt376 [Google Scholar](#)
44. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J. *Mol. Syst. Biol.* **2011**, 7, 539 DOI: 10.1038/msb.2011.75 [Google Scholar](#)
45. Ye, J.; Coulouris, G.; Zaretskaya, I.; Cutcutache, I.; Rozen, S.; Madden, T. L. *BMC Bioinf.* **2012**, 13, 134 DOI: 10.1186/1471-2105-13-134 [Google Scholar](#)
46. Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. *Science* **1997**, 276, 1109 DOI: 10.1126/science.276.5315.1109 [Google Scholar](#)