2017

Advances in single molecule microscopy: Protein characterization, force analysis and fluorescence localization

Chi-Fu Yen
Iowa State University

Advances in single molecule microscopy: Protein characterization, force analysis and fluorescence localization

by

Chi-Fu Yen

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Electrical Engineering

Program of Study Committee:
Sanjeevi Sivasankar, Major Professor
Anumantha G. Kanthasamy
Jaeyoun Kim
John G. Lajoie
Meng Lu

The student author and the program of study committee are solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2017

TABLE OF CONTENTS

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my advisor, Dr. Sanjeevi Sivasankar, who has consistently inspired me in this study. I consider myself extremely blessed to be his student. Without his attentive guidance and endless encouragement, my doctoral journey would not have been such a fruitful and eventful one.

In addition, I would like to thank Dr. Kai-Ming Ho and my dissertation committees, Dr. Anumantha G. Kanthasamy, Dr. Jaeyoun Kim, Dr. Meng Lu and Dr. John G. Lajoie for their valuable time and feedback throughout the course of this study.

I have been fortunate to have a fantastic group of friends that have been there for me in the good times and bad. I would like to specifically thank my labmates: Dr. Hui Li, Dr. Sabyasachi Rakshit, Dr. Sunae Kim, Dr. Kristine Manibog, Shafraz Omer, Patrick Schmidt, Andrew Priest, Ramesh Koirala and Seth Heerschap for the memorable times we have shared.

Finally, I want to thank my wife, Yin-Sin Chang, for being a bright light at the end of the tunnel, and my family for their never ending love and support from the other side of the ocean.

ABSTRACT

Recent advances in single molecule techniques have allowed scientists to address biological questions which cannot be resolved by traditional ensemble measurements. In this dissertation, I integrate single molecule and bulk measurements to establish a direct link between copper exposure and neurotoxicity in prion disease. Furthermore, I develop a new analysis method to improve the accuracy of kinetic parameter estimation in single molecule Atomic Force Microscope (AFM) experiments. Finally, I develop a new fluorescence localization microscopy to identify the axial position of a single fluorescent object with sub-nanometer accuracy.

Prion diseases are characterized by the misfolding and oligomerization of prion protein (PrP). Copper exposure has been linked to prion pathogenesis; however, the molecular mechanism is still unknown. In the first part of this dissertation, I use single molecule fluorescence assay, dynamic force spectroscopy (DFS) with AFM, and real-time quaking-induced conversion (RT-QuIC) assay to resolve, for the first time, the mechanistic basis by which Cu^{2+} ions induce a structural change in PrP, further promote oligomerization, template amyloid formation and neurotoxicity.

In the second part of this dissertation, I established a more accurate analysis method for single molecule DFS experiments. DFS is a widely used technique to characterize the dissociation kinetic between individual biomolecules. In AFM-DFS, receptor-ligand complexes are ruptured at different force rates by varying the speed at which the AFM-tip and substrate are pulled away from each other. The rupture events are grouped according to their pulling speeds and the mean force and loading rate of

each group is calculated. This data is subsequently fit to the established models to extract the kinetic parameters such as the intrinsic off-rate (k_{off}) and the width of the potential energy barrier (x_β). However, due to the large uncertainty in determining the mean forces and loading rates, errors in the estimated k_{off} and x_β can be substantial. Here, I demonstrate that these errors can be dramatically reduced by sorting rupture events into groups using cluster analysis. Monte Carlo simulations show that the cluster analysis is very effective at improving the accuracy of parameter estimation, especially when the number of unbinding events are limited and not well separated into distinct groups.

Finally, I describe a new technique, standing wave axial nanometry (SWAN), to image the axial location of a single nanoscale fluorescent object with sub-nanometer accuracy and 3.7 nm precision. A standing wave, generated by positioning an AFM tip over a focused laser beam, is used to excite fluorescence; axial position is determined from the phase of the emission intensity. I use SWAN to measure the orientation of single DNA molecules of different lengths, grafted on surfaces with different functionalities.

CHAPTER 1. INTRODUCTION

1.1 General Introduction

This dissertation advances the field of bioengineering by developing new tools to observe single biological molecules, new methods to analyze single molecule data and by using single molecule techniques to address critical biological problems. This dissertation is divided into three parts. In the first part of the dissertation, I integrate both ensemble and single molecule techniques to resolve and understand the mechanisms of protein misfolding, aggregation and neurotoxicity in prion neurodegenerative disease. Next, I develop a new analysis method to improve the accuracy of kinetic parameter estimation in single molecule Atomic Force Microscope (AFM) experiments. Finally, I develop a new fluorescence localization microscope to identify the axial position of a single fluorescent molecule with nanometer accuracy and precision.

An important feature of many major neurodegenerative disorders is the misfolding and aggregation of protein which causes tissue damage and cell death [1]. Transmissible spongiform encephalopathies (TSEs) or prion diseases, such as scrapie in sheep and Creutzfeldt-Jakob disease (CJD) in humans, are characterized by the misfolding and accumulation of prion protein (PrP) [2]. In Alzheimer's diseases, the most common cause of dementia among the elderly, the amyloid-β peptide (Aβ) monomers, aggregate to form amyloid plaques [3, 4]. Similarly, in Parkinson's disease, the natively unstructured protein, α-synuclein (αSyn), accumulates into fibers inside nerve cells [5].

Propagation of protein misfolding in these disorders is believed to proceed through a common prion-like mechanism, which means that proteins misfold, impose their structures on natively folded proteins, and template their aggregation [6]. Multiple lines of evidence suggest that metal binding is strongly correlated to protein misfolding in neurodegenerative diseases [7-9]. For example, Aβ forms cytotoxic aggregates in high concentrations of Cu^{2+} and Zn^{2+} [9]. Similarly, upon binding to Cu^{2+} and Fe^{3+}, αSyn coverts to a different conformation which enhances its aggregation [9]. Furthermore, it has been reported that Cu^{2+} promotes PrP aggregation and produces neurotoxic aggregates [10]. However, the mechanisms by which metal ions induces neurodegenerative protein misfolding, aggregation, and neurotoxicity are not clear.

The misfolding and structural change of proteins are a complex process, which involve multiple pathways and cause different conformations to coexist in a test tube [11, 12]. In contrast to conventional ensemble experiments which measure the averaged behavior of thousands of proteins, single molecule approaches are well suited to map protein misfolding and structural changes since they can directly characterize the biophysical properties of individual molecules and resolve heterogeneous misfolding pathways [13-15]. In the first part of this dissertation, I use single molecule force and fluorescence measurements, along with bulk biophysical techniques, to understand the relationship between metal exposure, protein misfolding, aggregation and neurotoxicity in prion diseases.

One of the techniques that I used to quantify the dissociation kinetics of individual PrP molecules is dynamic force spectroscopy (DFS) measurements using an AFM [15]. However, due to limitations in current data analysis methods, the errors in estimated

kinetic parameters can be substantial. The second part of this dissertation aims to address this concern by incorporating clustering algorithms into DFS analysis to improve the accuracy of kinetic parameter estimation. I employ computer simulations to mimic real DFS experiments and evaluate the performance of this new analysis method.

Another technique that I used to measure the misfolding of individual PrP molecules is fluorescence microscopy. Currently, the resolution of conventional fluorescence microscopes is limited by the diffraction of light: with a high numerical aperture objective and visible excitation, resolution is about 200 nm in the lateral direction and 500 nm along the optical axis. Although a single fluorescent molecule can be localized with nanometer accuracy along the x- and y-axis [16, 17], improving resolution and single molecule localization accuracy along the optical axis is more challenging [18]. In the final part of this dissertation, I develop a fluorescence localization method with an axial accuracy and precision superior to previous techniques. This technique, called standing wave axial nanometry (SWAN), utilizes a commercial AFM mounted on a single molecule confocal microscope [19]. A standing wave, generated by positioning an AFM tip over a focused laser beam, is used to excite the fluorophore; axial position is determined from the phase of the emission intensity. As a proof of principle, I use SWAN to measure the orientation of single DNA molecules of different lengths, grafted on surfaces with different functionalities.

1.2 Thesis Organization

This dissertation is organized as follows:

Chapter 2 presents our paper published in *Science Advances*, resolving the mechanistic basis for copper induced PrP misfolding and oligomerization [20]. This work combines single molecule fluorescence and AFM force measurements, along with circular dichroism (CD) spectroscopy, dynamic light scattering (DLS) and real-time quaking induced conversion (RT-QuIC) protein conversion assay, to investigate the relationship between copper exposure and neurotoxicity in prion diseases at the single molecule level.

Chapter 3 presents a new analysis method for DFS experiments. This work is motivated by our *Science Advances* paper [20], where we realized the lack of accuracy of current methods in analyzing DFS data. Chapter 3 addresses this concern by using simple cluster analysis algorithms, including Gaussian mixture, logistic regression and K-means clustering to sort and analyze data. The results of this part of the dissertation have been submitted for publication.

Chapter 4 presents our paper published in *Nano Letters* [21] describing a new fluorescence localization technique, SWAN, to image the axial location of a single molecule with nanometer accuracy and precision.

Finally, chapter 5 summarizes the significance of this dissertation, its impact on understanding the molecular mechanism of protein aggregation in neurodegenerative diseases and improvement of single molecule techniques. I also briefly describe future applications of the single molecule methods described in this dissertation.

1.3 References

[1] C.M. Dobson, Protein folding and misfolding, Nature 426 (2003) 884-890.

[2] S.B. Prusiner, Prions, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 13363-13383.

[3] S.A. Funke, D. Willbold, Peptides for Therapy and Diagnosis of Alzheimer's Disease, Curr. Pharm. Design 18 (2012) 755-767.

[4] J. Hardy, D.J. Selkoe, Medicine - The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics, Science 297 (2002) 353-356.

[5] M.H. Polymeropoulos, C. Lavedan, E. Leroy, S.E. Ide, A. Dehejia, A. Dutra, B. Pike, H. Root, J. Rubenstein, R. Boyer, E.S. Stenroos, S. Chandrasekharappa, A. Athanassiadou, T. Papapetropoulos, W.G. Johnson, A.M. Lazzarini, R.C. Duvoisin, G. DiIorio, L.I. Golbe, R.L. Nussbaum, Mutation in the alpha-synuclein gene identified in families with Parkinson's disease, Science 276 (1997) 2045-2047.

[6] B. Frost, M.I. Diamond, Prion-like mechanisms in neurodegenerative diseases, Nat. Rev. Neurosci. 11 (2010) 155-159.

[7] A.I. Bush, Metals and neuroscience, Curr. Opin. Chem. Biol. 4 (2000) 184-191.

[8] J.H. Viles, Metal ions and amyloid fiber formation in neurodegenerative diseases. Copper, zinc and iron in Alzheimer's, Parkinson's and prion diseases, Coord. Chem. Rev. 256 (2012) 2271-2284.

[9] S.S. Leal, H.M. Botelho, C.M. Gomes, Metal ions as modulators of protein conformation and misfolding in neurodegeneration, Coord. Chem. Rev. 256 (2012) 2253-2270.

[10] M.F. Jobling, X.D. Huang, L.R. Stewart, K.J. Barnham, C. Curtain, I. Volitakis, M. Perugini, A.R. White, R.A. Cherny, C.L. Masters, C.J. Barrow, S.J. Collins, A.I. Bush, R. Cappai, Copper and zinc binding modulates the aggregation and neurotoxic properties of the prion peptide PrP106-126, Biochemistry 40 (2001) 8073-8084.

[11] F. Chiti, C.M. Dobson, Protein misfolding, functional amyloid, and human disease, Annu. Rev. Biochem. 75 (2006) 333-366.

[12] H. Yu, X. Liu, K. Neupane, A.N. Gupta, A.M. Brigley, A. Solanki, I. Sosova, M.T. Woodside, Direct observation of multiple misfolding pathways in a single prion protein molecule, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 5283-5288.

[13] A.C.M. Ferreon, A.A. Deniz, Protein folding at single-molecule resolution, Biochim. Biophys. Acta, Proteins Proteomics 1814 (2011) 1021-1029.

[14] A. Borgia, P.M. Williams, J. Clarke, Single-molecule studies of protein folding, Annu. Rev. Biochem. 77 (2008) 101-125.

[15] K.C. Neuman, A. Nagy, Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy, Nat. Methods 5 (2008) 491-505.

[16] R.E. Thompson, D.R. Larson, W.W. Webb, Precise nanometer localization analysis for individual fluorescent probes, Biophys. J. 82 (2002) 2775-2783.

[17] A. Yildiz, J.N. Forkey, S.A. McKinney, T. Ha, Y.E. Goldman, P.R. Selvin, Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization, Science 300 (2003) 2061-2065.

[18] B. Huang, An in-depth view, Nat. Methods 8 (2011) 304-305.

[19] S. Sivasankar, S. Chu, Nanoparticle-mediated nonfluorescent bonding of microspheres to atomic force microscope cantilevers and imaging fluorescence from bonded cantilevers with single molecule sensitivity, Nano Lett. 9 (2009) 2120-2124.

[20] C.F. Yen, D.S. Harischandra, A. Kanthasamy, S. Sivasankar, Copper-induced structural conversion templates prion protein oligomerization and neurotoxicity, Sci. Adv. 2 (2016) e1600014.

[21] H. Li, C.F. Yen, S. Sivasankar, Fluorescence Axial Localization with Nanometer Accuracy and Precision, Nano Lett. 12 (2012) 3731-3735.

CHAPTER 2. COPPER-INDUCED STRUCTURAL CONVERSION TEMPLATES PRION PROTEIN OLIGOMERIZATION AND NEUROTOXICITY

This chapter is published in *Science Advances*†

Chi-Fu Yen, Dilshan S. Harischandra, Anumantha Kanthasamy and Sanjeevi Sivasankar

2.1 Abstract

Prion protein (PrP) misfolding and oligomerization are key pathogenic events in prion disease. Copper exposure has been linked to prion pathogenesis, however, its mechanistic basis is unknown. Here we resolve, with single molecule precision, conditions for Cu^{2+} mediated PrP misfolding, oligomerization and neurotoxicity. Using a single molecule fluorescence assay, we demonstrate that Cu^{2+} induces PrP monomers to misfold before oligomer assembly; the disordered N-terminal region mediate this structural change. Single molecule force spectroscopy measurements show that the misfolded monomers have a 900 fold higher binding affinity which promotes their oligomerization. Real-time quaking-induced conversion demonstrates that misfolded PrP serve as seeds that template amyloid formation. Finally, organotypic slice cultures show that misfolded PrP mediate inflammation and degeneration of neuronal tissue. Our study provides direct proof that Cu^{2+} misfolds PrP, subsequently promoting oligomerization and cytotoxicity.

---

† C.F. Yen, D.S. Harischandra, A. Kanthasamy, S. Sivasankar, Copper-induced structural conversion templates prion protein oligomerization and neurotoxicity, Sci. Adv. 2 (2016) e1600014.

2.2 Introduction

Protein misfolding and oligomeric protein accumulation are key pathogenic mechanisms in many major neurodegenerative disorders including prion diseases, Parkinson's disease and Alzheimer's disease [1]. Propagation of protein misfolding in these disorders is believed to proceed through a common prion-like mechanism, meaning that proteins misfold, impose their structures upon natively folded proteins and template their aggregation [2]. Although the structure of natively-folded PrP has been resolved [3-6], the mechanistic basis for its misfolding, aggregation and neurotoxicity is unknown.

PrP is a metal-binding protein; the unstructured N-terminal tail of PrP contains four copies of an octarepeat sequence which bind a number of divalent ions including Cu^{2+}, Ni^{2+} and Mn^{2+} (Fig.1A) [7-9]. Metal ions also bind to an additional site in the flexible region between the octarepeats and the C-terminal domain [10]. It has been proposed that Cu^{2+} ions play a role in PrP structural conversion [11]; however, direct molecular proof for the role of Cu^{2+} in PrP misfolding and aggregation is lacking. Biochemical experiments show that PrPs incubated with Cu^{2+} ions misfold from a native, α-helix rich structure, to a predominantly β-sheet conformation that is resistant to proteolytic digestion [12, 13]. Along the same lines, aggregation of PrP is enhanced either by the insertion of additional octapeptide repeats or by exposure to high concentrations of Cu^{2+} [14, 15]. However, the mechanisms by which Cu^{2+} induces PrP misfolding, aggregation and neurotoxicity is unknown.

Here we resolve, with single molecule precision, Cu^{2+} induced misfolding of PrP under physiological conditions. We also demonstrate that the misfolded PrP serve as seeds for the templated formation of aggregates which mediate inflammation and

degeneration of neuronal tissue. We compare the biophysical properties of full-length, human PrP with globular protein lacking the N-terminal octapeptide repeats in the presence and absence of Cu^{2+}, Ni^{2+} and Mn^{2+} (Fig. 1). Using a fluorescence based single

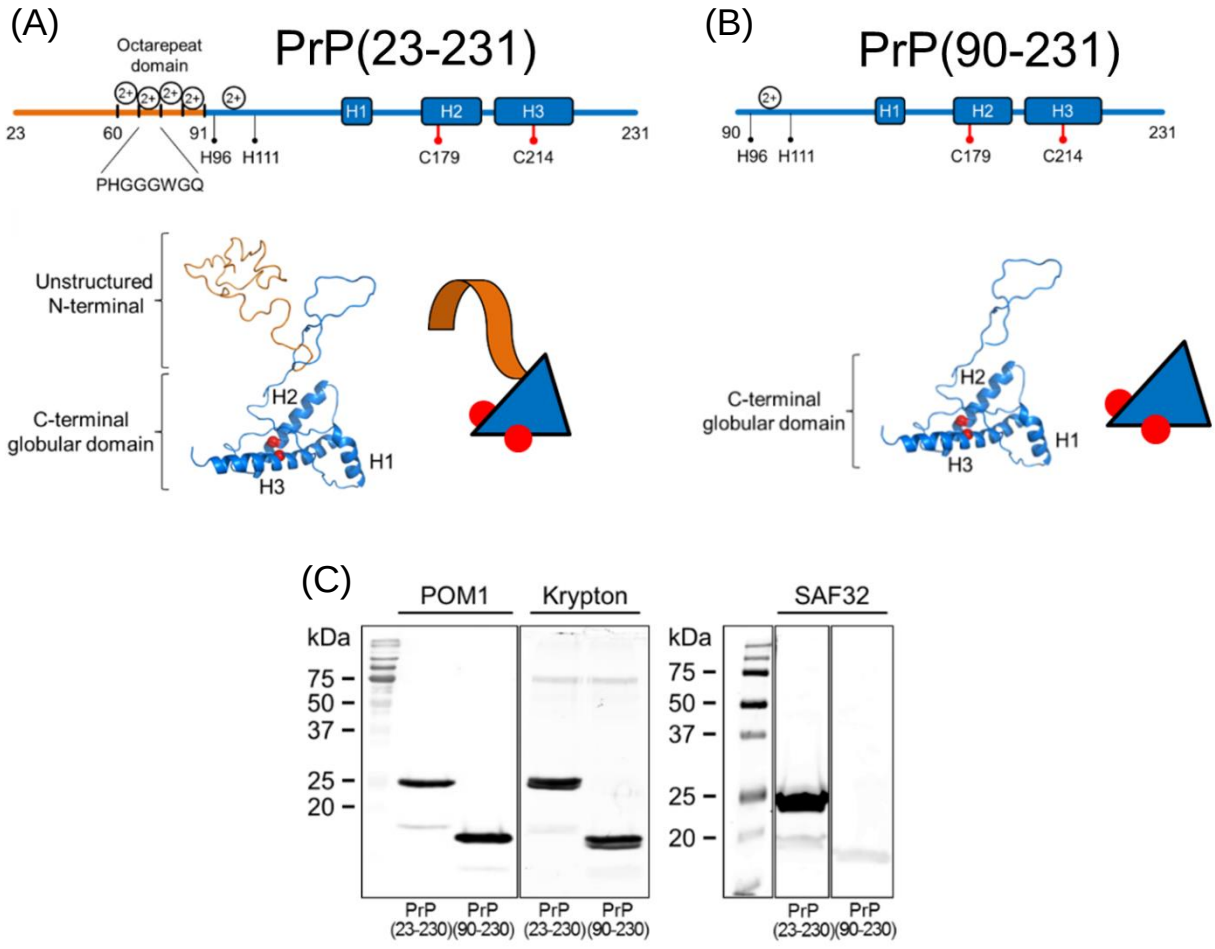

Figure 1. Prion protein (PrP) constructs used in the experiments. Two constructs of PrP were used in the experiments: **(A)** Full-length human PrP (PrP(23-231) or PrP(23-230)). The unstructured N-terminal domain (orange) has four metal binding octapeptide repeats. The C-terminal is globular and rich in α-helices (blue). Red dots indicate cysteine 179 and cysteine 214 which were used to tether PrP to the surface. **(B)** PrP without unstructured N-terminal region (PrP(90-231) or PrP(90-230)). **(C)** SDS-PAGE and Western blot of purified PrP(23-230) (Molecular Weight = 25.2 kD) and PrP(90-230) (Molecular Weight = 18.5 kD). SDS-PAGE was stained by Krypton fluorescent protein stain. Anti-prion antibodies POM1 and SAF32 were used to recognize the globular domain of PrP and the octapeptide repeats respectively.

molecule assay, we demonstrate that PrP monomers misfold to a protease resistant conformation before oligomer assembly; the N-terminal region and Cu^{2+} ions mediate misfolding. This is a striking result since the conventional model is that protease-resistance is acquired only by oligomers; in fact protease-resistance is often used as a read-out for PrP oligomerization [16]. Using single molecule force measurements with an Atomic Force Microscope (AFM), we show that misfolded PrP monomers have an almost 900 fold higher affinity (K_A) compared to the native isoform and rapidly form oligomers. Using real-time quaking-induced conversion (RT-QuIC) [17], a cell-free seeding assay designed for rapid clinical diagnosis of prion disease, we show that misfolded PrP form active seeds that template aggregation. Finally, to investigate the effect of Cu^{2+} induced PrP misfolding on neuronal viability, we quantify protein markers for inflammation, apoptosis and oxidative stress in brain explants exposed to native and misfolded PrP. Analysis of these organotypic slice cultures provide direct molecular proof that the misfolded PrPs mediate neuroinflammation and neurodegeneration. Taken together, our results identify conditions for Cu^{2+} induced PrP misfolding, oligomerization and neurotoxicity.

2.3 Results

2.3.1 N-terminal region and Cu^{2+} ions mediate protease resistance in PrP monomers

A common characteristic of PrP misfolding is the conversion of native, proteinase-K (PK) sensitive PrP (PrP^{sen}) into an isoform that is prone to aggregation and resists PK digestion (PrP^{res}) [18-20]. In order to identify misfolded PrP and determine the role of different metal ligands and of the N- and C-terminal domains in PrP misfolding, we

developed a fluorescence based single molecule PK-resistance assay. The single molecule resolution of our assay allowed us measure the misfolding of PrP monomers and avoid artifacts introduced by higher order oligomerization. In our assay, the PrPs were first biotinylated and then covalently immobilized on a glass substrate decorated with maleimide functionalized Poly Ethylene Glycol (PEG) tethers (Methods). Surface density measurements (Methods; figs. S1 and S2) showed that the PrP bound to the surface as well separated monomers (table S1). The immobilized PrP monomers were incubated overnight either in the presence or absence of 1 mM Ni^{2+}, Mn^{2+}, or Cu^{2+} ions; dissociation constants indicate that the PrP would be saturated with the divalent ions under these experimental conditions [11]. The samples were then treated with PK for different time periods (0-13 hours) and subsequently labeled with fluorescent streptavidin (Fig. 2, A to C; Methods). If the biotinylated PrP converted to a PrP^{res} conformation, it was not digested by PK and could therefore bind to the fluorescent streptavidin (Fig. 2C). On the other hand, biotinylated PrP^{sen} was readily digested by PK which resulted in a decreased binding of fluorescent streptavidin (Fig. 2, A and B). As a positive control, we covalently immobilized PK-resistant, fluorescently labeled, double-stranded DNA to identical PEG functionalized substrates. As a negative control, PK-sensitive, biotin-conjugated bovine serum albumin (BSA) was covalently immobilized on the PEG monolayers (Methods).

Our data demonstrated that only full-length PrP (PrP(23-231)) incubated in Cu^{2+} converted to a stable PrP^{res} conformation. Even after a ten hour exposure to PK, the fluorescence intensity of PrP(23-231) was unaltered (Fig. 2D). Similarly, the fluorescence intensity from the positive control remained unchanged after 10 hour PK incubation. On

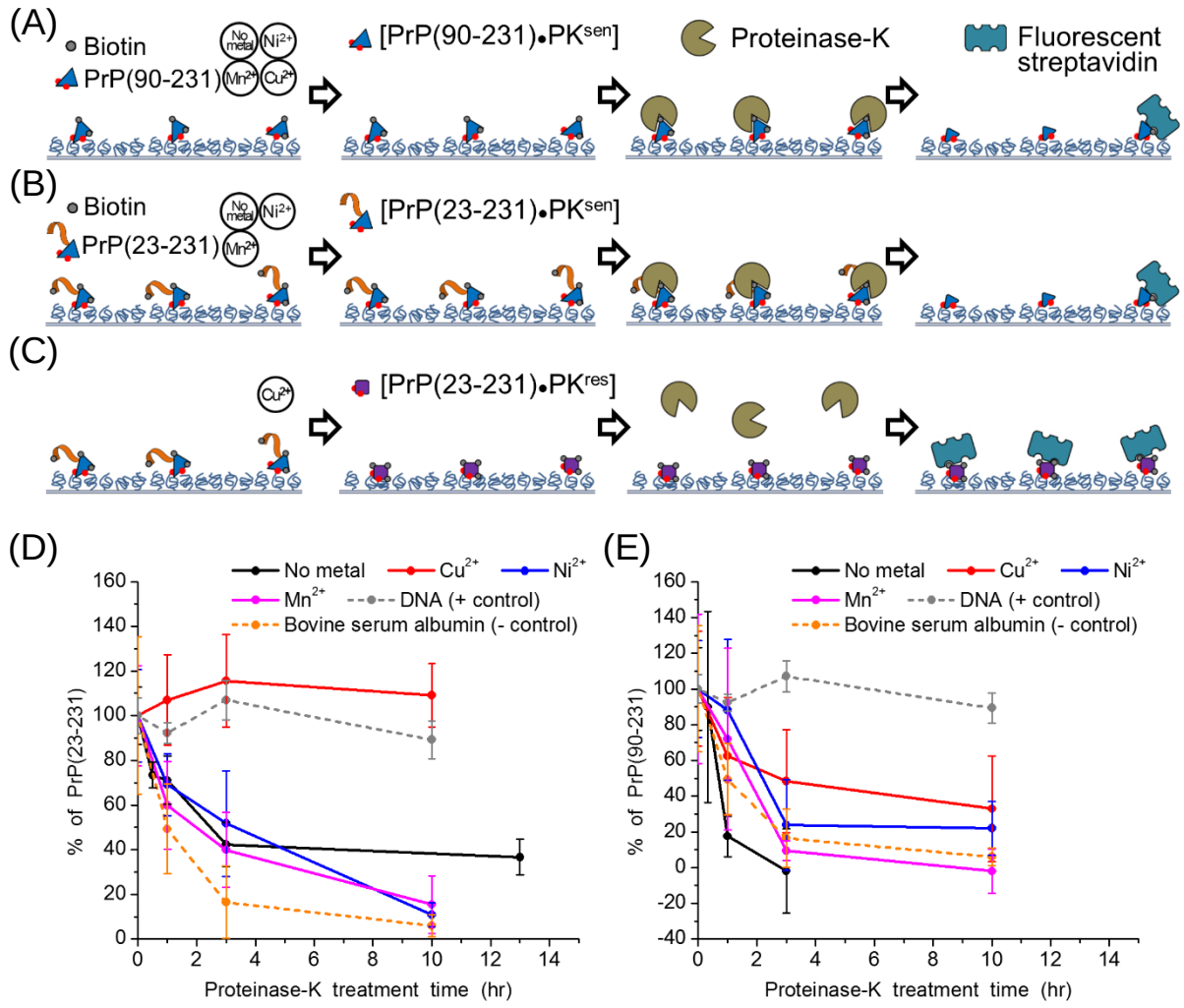

Figure 2. Monitoring protease resistance in single prion protein (PrP) monomers. PrP(90-231) or PrP(23-231) was biotinylated and covalently immobilized on a substrate functionalized with a Poly Ethylene Glycol (PEG) monolayer. The immobilized PrP was incubated **(A)** either without or with 1 mM Cu^{2+}/Mn^{2+}/Ni^{2+} (PrP(90-231)), **(B)** either without or with 1 mM Mn^{2+}/Ni^{2+} (PrP(23-231)) or **(C)** with 1mM Cu^{2+} (PrP(23-231)). Following incubation, the immobilized PrP was digested with Proteinase-K (PK); undigested PrP was detected by labeling with fluorescent streptavidin. Structural conversion to a PK-resistant conformation resulted in a fluorescence signal. Measured fluorescence from **(D)** PrP(23-231) and **(E)** PrP(90-231) after PK digestion for different time courses normalized by the fluorescence measured from undigested PrP in each condition. Only PrP(23-231) incubated in Cu^{2+} converted to a stable PK-resistant conformation. PEG substrates functionalized with either fluorescently labeled DNA or biotin-conjugated bovine serum albumin (BSA) were used as positive and negative controls respectively. The data in D was acquired from a total of 32652, 17638, 4776 and 29207 molecules in the absence and presence of Cu^{2+}, Mn^{2+} and Ni^{2+} respectively. The data in E was acquired from a total of 9634, 12438, 10680 and 8869 molecules in the absence and presence of Cu^{2+}, Mn^{2+} and Ni^{2+} respectively. Error bars correspond to the s.e. calculated using a bootstrap with replacement protocol.

the other hand, PK-resistance was significantly reduced for PrP(23-231) bound to Mn^{2+} or Ni^{2+} and for PrP(23-231) in the absence of divalent ions; the fluorescence intensity decreased between 60% and 90% after 10 hour PK incubation (Fig. 2D). The data in figure 2D was acquired from a total (summed across all time-points) of 32652, 17638, 4776 and 29207 PrP(23-231) molecules in the absence and presence of Cu^{2+}, Mn^{2+} and Ni^{2+} respectively.

Similarly, globular PrP that lacked the unstructured N-terminal region (PrP(90-231)) was sensitive to PK digestion both in the absence and presence of Cu^{2+}, Mn^{2+} and Ni^{2+}; after a 10 hour digestion with PK, the fluorescence intensity decreased between 60% and 100% (Fig. 2E). In contrast, fluorescence signal from the negative control was completely eliminated after 10 hours of PK digestion (Fig. 2, D and E). The data in figure 2E was acquired from a total (summed across all time-points) of 9634, 12438, 10680 and 8869 PrP(90-231) molecules in the absence and presence of Cu^{2+}, Mn^{2+} and Ni^{2+} respectively. These results show that monomeric PrP^{res} formation requires both the intrinsically disordered N-terminal region and Cu^{2+} ions; elimination of either results, predominantly, in PrP^{sen} conformation.

2.3.2 Full-length PrPs incubated in Cu^{2+} ions have a higher binding affinity

Next, we proceeded to use single molecule AFM force measurements [21, 22] to characterize the kinetics of the initial stage of PrP aggregation and to identify the role of different divalent metal ions and protein domains in this process. AFM force spectroscopy has previously been employed to study the misfolding and interactions of neurodegenerative proteins [23, 24]. Single molecule measurements with tethered PrP

monomers enabled us to directly determine association and dissociation rates while avoiding artifacts due to the formation of higher order oligomers. Relative dimerization rates (relative on-rates) were determined from the measured binding probabilities while off-rates were measured using single molecule dynamic force spectroscopy. In both sets of experiments, the PrP constructs were covalently immobilized on an AFM tip and substrate (Fig. 3, A and B; Methods) and incubated in either the presence or absence of 1 mM Cu^{2+}, Mn^{2+} or Ni^{2+}. The AFM tip and surface was first brought into contact enabling opposing PrPs to dimerize. The tip was then withdrawn from the substrate and the force required to rupture the PrP dimer was recorded. If the PrPs did not interact, no unbinding forces were measured (Fig. 3C). However, dimer formation resulted in unbinding events characterized by the non-linear stretching of the PEG tethers (Fig. 3D); the stretching of PEG served as a molecular fingerprint for PrP dimer formation since its extension under load has been extensively characterized [25]. As described in the methods section, the unbinding events were fit to an extended freely-jointed chain (FJC) model [25] using a total least squares fitting protocol, only unbinding events that occurred at a distance corresponding to the contour length of two PEGs in series were used in further analysis (Methods). Since each PrP can be tethered to the AFM tip /substrate via either one or two PEGs (covalently bound to either Cys 179 or/and Cys 214), we fit the unbinding events to alternate FJC models that assumed stretching of either a single tether or two parallel tethers [26]. Our analysis showed that in 78% of the selected events, PrP was immobilized on both the AFM tip and substrate via a single PEG tether. In contrast, 21% of unbinding corresponded to rupture events where one of the PrPs was tethered via two PEG linkers while the other PrP was immobilized via a single PEG. Only 1% of the events,

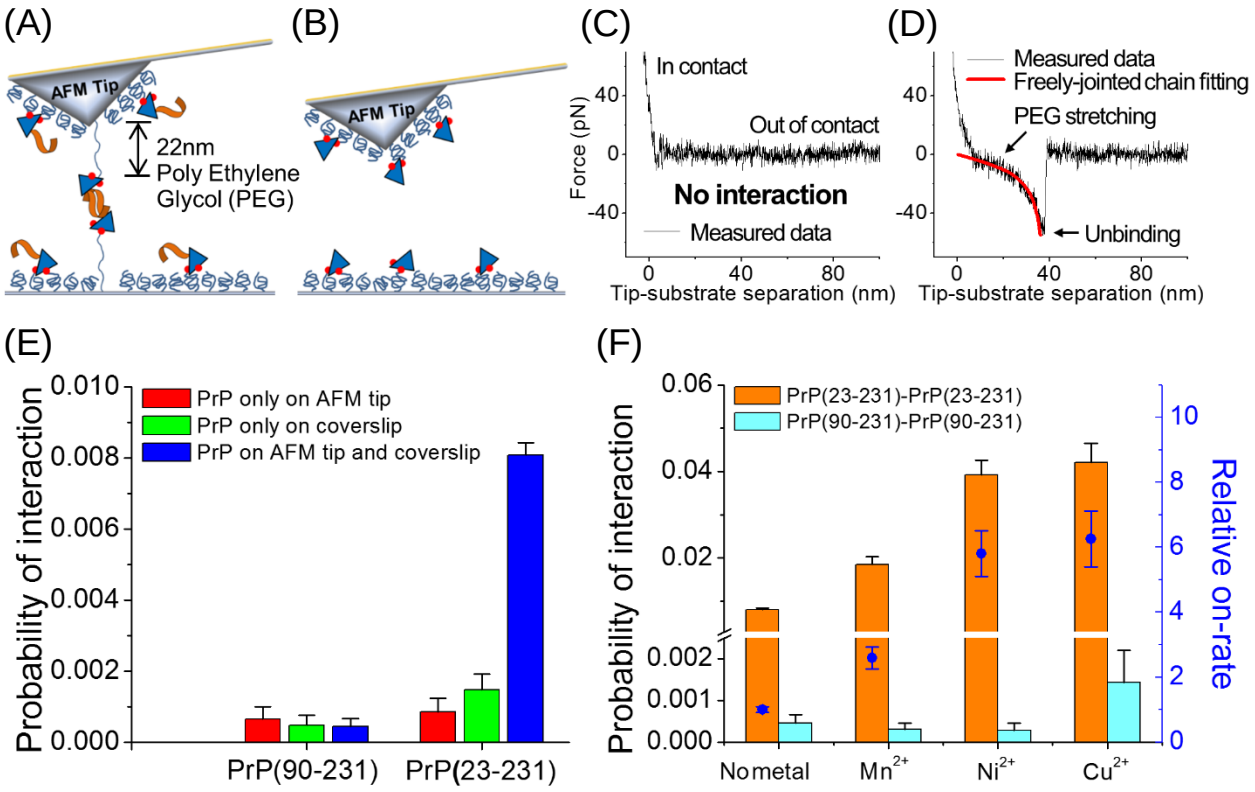

Figure 3. Binding probabilities of full-length PrP and truncated PrP measured using single molecule Atomic Force Microscopy (AFM). Probability of binding was measured for **(A)** PrP(23-231) bound to both the AFM tip and substrate and **(B)** PrP(90-231) bound to both AFM tip and substrate. Representative force *vs.* tip-substrate separation when opposing PrPs **(C)** do not interact, or **(D)** interact with each other. In the case of PrP interaction, the PEG tether is stretched as a freely-jointed chain (FJC) before PrP unbinding. **(E)** Interaction probabilities were measured for PrP only on AFM tip (red), PrP only on substrate (green) and PrP on both AFM tip and substrate (blue). Red and green bars indicate the probability of non-specific interactions. Only PrP(23-231) on both tip and substrate shows a higher binding rate than non-specific interactions. **(F)** Binding probability (orange) and relative on-rate (blue) between opposing PrP(23-231) and binding probability (cyan) between opposing PrP(90-231) in the absence and presence of 1mM Mn^{2+}, 1mM Ni^{2+}, and 1mM Cu^{2+}. The total number of PrP(23-231) measurements were 62523, 20440, 24315 and 19070 in the absence of divalent metal ions and in the presence of Mn^{2+}, Ni^{2+}, and Cu^{2+} respectively. The total number of PrP(90-231) measurements were 22048, 18866, 8293 and 8368 in the absence of divalent metal ions and in the presence of Mn^{2+}, Ni^{2+}, and Cu^{2+} respectively. Error bars of binding probability indicate the s.e., were estimated using a bootstrap with replacement protocol. Error bars of relative on-rate indicate the s.e., were propagated from the s.e. of binding probability.

corresponded to unbinding events where both PrPs were immobilized via two tethers. This heterogeneous tethering distribution however did not affect the measured binding probability or unbinding force; identical results were obtained when events corresponding to the stretching of only single tethers were compared to measurements that correspond to the stretching of both single and parallel PEG tethers. The measurements were repeated several thousand times at 6 or 11 different pulling velocities and at different positions of substrate. Surface density of PrP was estimated for every condition using single molecule fluorescence microscopy (Methods; figs. S1 and S2); measured probabilities were normalized for direct comparison.

To identify the role of the unstructured N-terminal region in PrP interactions, we first compared the binding probability of full-length PrP(23-231) and globular PrP(90-231) in the absence of divalent metal ions. We established that our selected unbinding events had a non-specific binding probability of 0.15±0.04% by eliminating PrP from either the tip or the substrate (Fig. 3E). A comparable binding probability of 0.05±0.02% was measured for opposing PrP(90-231) suggesting that globular PrP lacking the N-terminal domain do not form dimers (Fig. 3E). In contrast, PrP(23-231) dimerized with a five-fold higher probability (0.81±0.04%) showing that the N-terminal tail plays a role in dimerization (Fig. 3E).

Next, we quantified the effect of divalent ions on the homotypic binding probabilities between opposing PrP(23-231) (Fig. 3A) and opposing PrP(90-231) (Fig. 3B). Homotypic PrP(90-231) binding probabilities were comparable to non-specific adhesion; interaction frequency in the presence of Mn^{2+}, Ni^{2+}, Cu^{2+} and in the absence of divalent ions was 0.03±0.02%, 0.03±0.02%, 0.14±0.08%, and 0.05±0.02% respectively

(Fig. 3F). However, in striking contrast, homotypic PrP(23-231) binding probability was significantly enhanced by the presence of divalent ions. While a binding probability of 0.81±0.04% was measured in the absence of divalent ions, the interaction frequency increased to 1.85±0.18%, 3.92±0.34% and 4.21±0.45% when Mn^{2+}, Ni^{2+} and Cu^{2+} were added to solution (Fig. 3F).

Based on the probabilities of interaction, we calculated the relative on-rates for PrP(23-231) interactions (table S2) [27]. Since the association rates depend on factors besides microscopic binding rates, such as the protein surface density, the local geometry of the tip and substrate and the location of the proteins, all rates were calculated relative to the association rate of homotypic PrP(23-231) interactions in the absence of divalent ions (Methods) [27]. The on-rate of homotypic PrP(23-231) was increased to 6.3±0.9 times in 1 mM Cu^{2+}, 5.8±0.7 times in 1 mM Ni^{2+} and 2.6±0.3 in 1 mM Mn^{2+} (Fig. 3F; table S2).

Next, using PrP(23-231), we calculated the dissociation rate by measuring the most probable rupture force at different rates of application of force; the data was fitted to the Bell-Evans equation [28] to extract the intrinsic off-rate (k_{off}) and the width of the potential energy barrier (x_β) that inhibits dimer dissociation (Fig. 4; Methods; figs. S3 to S6). Our data shows that PrP dimer off-rates in the absence of divalent ions and in the presence Mn^{2+}, Ni^{2+} were comparable; the measured off-rates corresponded to 3.1×10^{-2} sec^{-1}, 1.3×10^{-2} sec^{-1} and 1.0×10^{-2} sec^{-1} respectively (Fig. 4, A to C; table S2). In contrast, the measured off-rate in Cu^{2+} was two orders of magnitude smaller with a value of 2.2×10^{-4} sec^{-1} (Fig. 4D; table S2). This indicates that dimers of PrP in Cu^{2+} have lifetimes that are around 100 times longer than PrP^{sen} dimers.

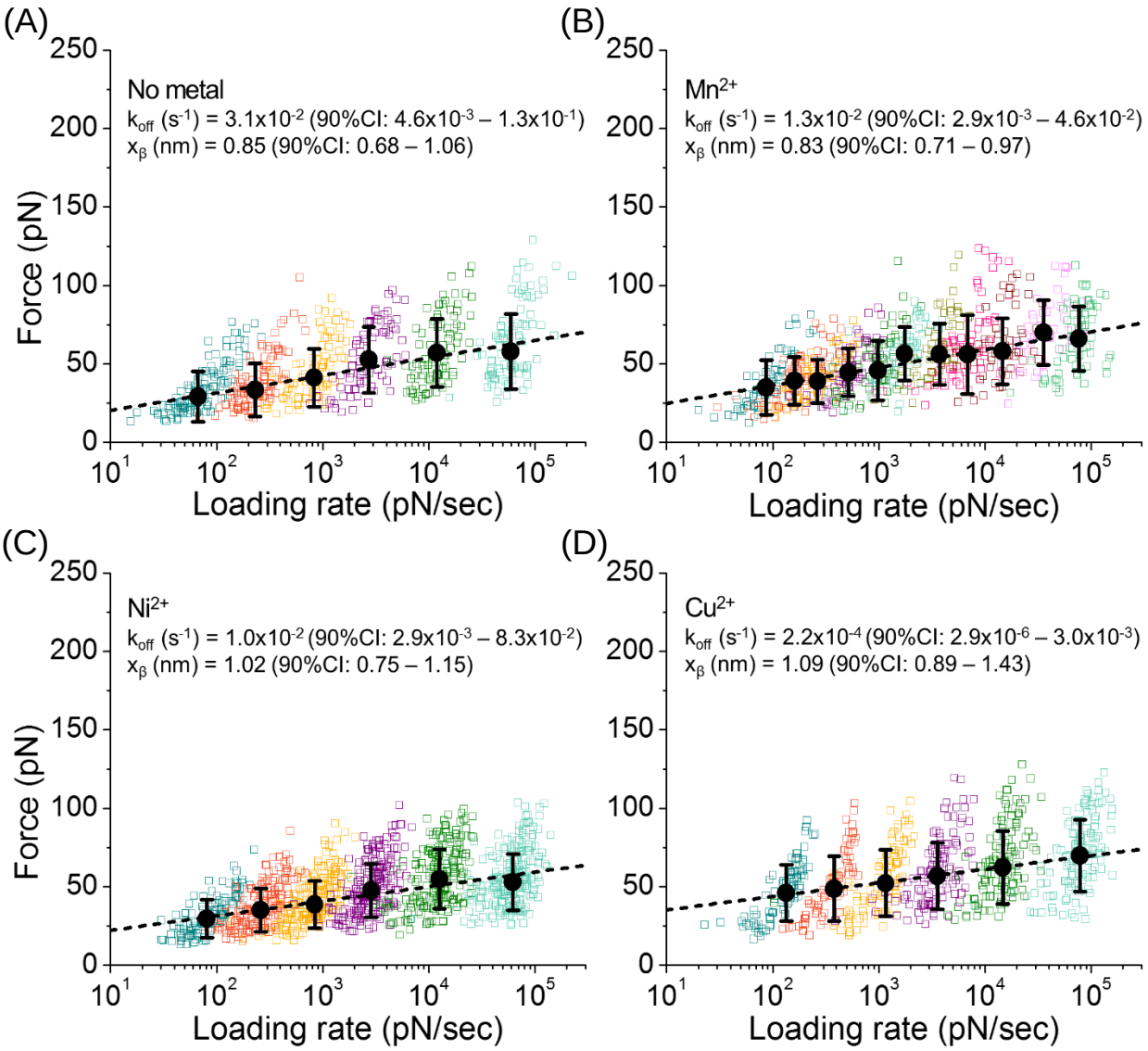

Figure 4. Dissociation rate of PrP(23-231) dimers measured using single molecule dynamic force spectroscopy. (A) Plots of the measured rupture forces and loading rates (LRs) in the absence of divalent ions. Colored open squares correspond to individual unbinding events at different LRs (556 events). Black circles representing the most probable rupture forces were fit to the Bell-Evans model (black dashed line) to determine the off-rates (k_{off}) and energy barrier width (x_β). Error bars indicate the standard deviation of forces. 90% confidence interval (CI) was calculated using a bootstrap with replacement protocol. Similar analysis for PrP(23-231) in **(B)** 1mM Mn^{2+} (709 events); **(C)** 1mM Ni^{2+} (1133 events); **(D)** 1mM Cu^{2+} (606 events).

Using the measured off-rates and the relative on-rates determined from the binding probabilities, we calculated the relative association constant (K_A) for PrP(23-231) in the presence of different metal ions (Methods). The K_A of PrP incubated in Cu^{2+}, Ni^{2+} and Mn^{2+} was 863, 18 and 6 times higher than PrP in the absence of divalent ions respectively (table S2).

2.3.3 PrPs incubated in Cu^{2+} serve as seeds for templated aggregation

Next, we used RT-QuIC [17], an *in vitro* seeding assay, to test if PrP exposed to divalent metal ions form seeds that template aggregation. In a conventional RT-QuIC assay, trace amounts of scrapie seeds are added to a recombinant PrP substrate and repeatedly agitated [17]; amyloid fiber formation is monitored from an increase in fluorescence intensity when Thioflavin T (ThT) binds to the aggregate. If the seeds are active, a more rapid onset of aggregation is measured with increasing seed concentration. In our measurements, we generated PrP seeds by incubating human recombinant PrP(23-230) or PrP(90-230) in either the presence or absence of 10 μM Mn^{2+}, Ni^{2+} or Cu^{2+} ions at 37°C. As the PrP began to misfold to form seeds, their aggregation was monitored using ThT fluorescence intensity; seed formation was accelerated by periodically agitating the solution (Methods). While PrP(23-230) incubated in Mn^{2+}, Ni^{2+} and Cu^{2+} ions, began forming seeds between 15 hrs and 25 hrs, PrP(23-230) in the absence of divalent ions began forming seeds only after 55 hours (fig. S7). When PrP(90-230) was incubated in the presence of Ni^{2+} and Cu^{2+} ions for 95 hours, no seed formation was measured (fig. S8). However, PrP(90-230) incubated in Mn^{2+} ions began forming seeds after 85 hours (fig. S8).

We used the PrP(23-230) seeds to template aggregation of recombinant human PrP(23-230) in the absence of metal ions (Methods). Since the presence of trace amounts of divalent ions may affect *in vitro* conversion [29], 1 mM of EDTA, a metal ion chelator, was included in assay (Methods). When seeds were added to the substrate of natively folded PrP, a lag phase (T_{th}) preceded the onset of aggregation (Fig. 5A). The duration of T_{th} indicated the efficiency of the seeding reaction [30, 31]; for the same concentration of seeds, a shorter T_{th} indicated an earlier onset of aggregation and more efficient seed propagation. Previous theoretical studies of templated protein aggregation predict that the lag phase depends on the amount of initial seed; a decrease in T_{th} is predicted as the seed concentration increases (Fig. 5B) [30]. We therefore measured T_{th} when 150 fg, 1.5 pg, 15 pg, 150 pg and 1.5 ng of seeds were added to 15 μg of recombinant human PrP(23-230) (Fig. 5, C to F; Methods; figs. S9 to S11). Only PrP(23-230) seeds prepared in 10 μM Cu^{2+} ions showed a decrease in T_{th} with increasing seed concentration (one-tailed t-test, $p = 0.05$), suggesting an inverse correlation as predicted by theory (Fig. 5D). Similarly, when PrP(23-230) seeds prepared in 1 μM Cu^{2+} ions were used in the experiments, a more rapid onset of aggregation with increasing seed concentration was measured (fig. S12). In contrast, a decrease in T_{th} with increasing seed concentration was not observed when seeds were prepared in Mn^{2+} and Ni^{2+} ions (one-tailed t-test, $p = 1$, 0.81 respectively) (Fig. 5, E and F). These results demonstrates that only PrP(23-230) exposed to Cu^{2+} ions serve as seeds for formation of prion protein aggregates.

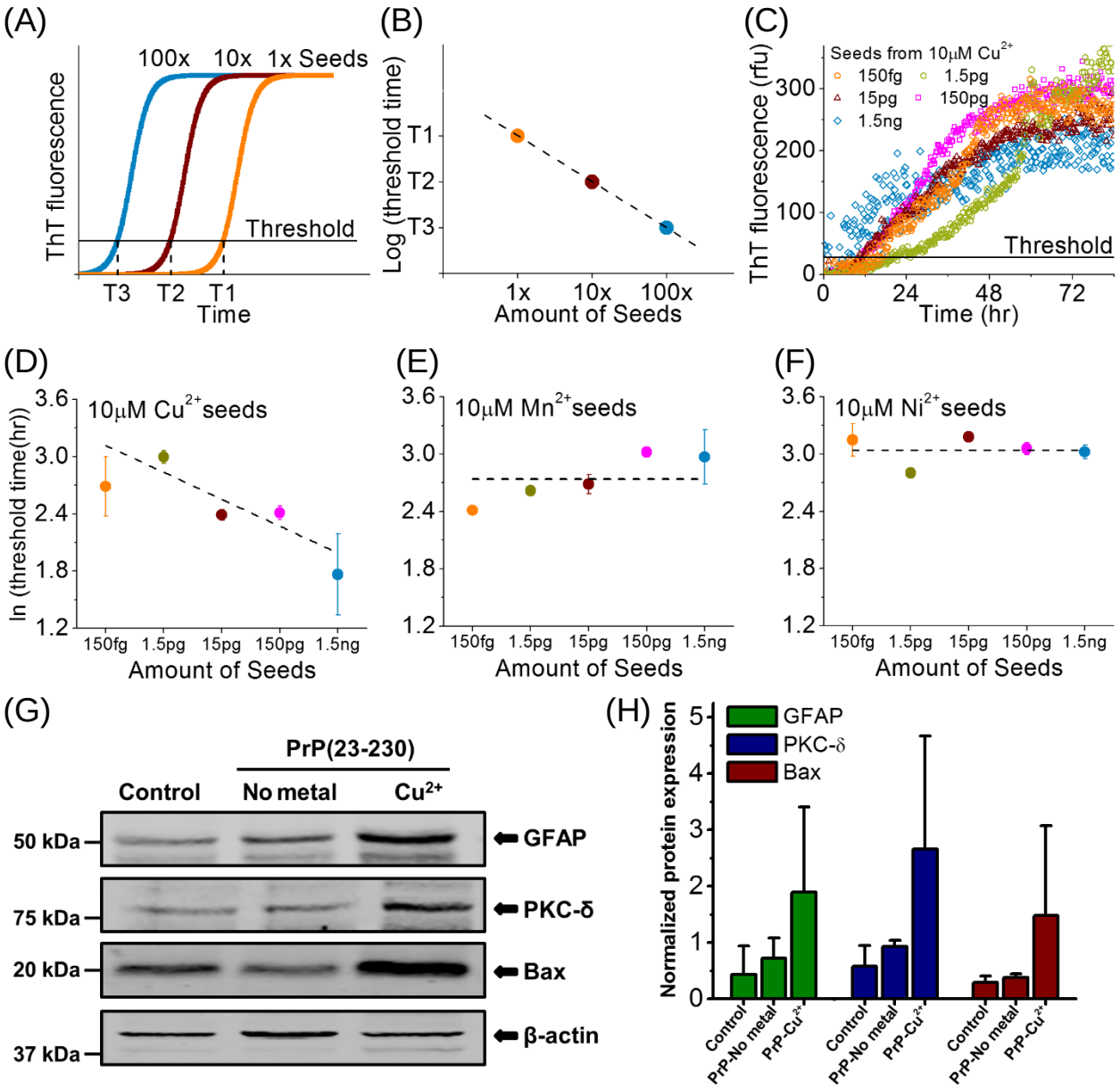

Figure 5. Seeding activity and neurotoxicity of PrP(23-230) measured using real-time quaking induced conversion (RT-QuIC) and organotypic slice cultures. (A) In the RT-QuIC assay, enhancement of thioflavin T (ThT) fluorescence is measured upon binding to amyloid fibrils. For nucleated polymerization, duration of the lag phase increases when the initial amount of seeds decreases. **(B)** For templated aggregation, the duration of the lag phase is predicted to be inversely correlated to the seed concentration. **(C)** RT-QuIC traces for 15 μg human recombinant PrP(23-230) substrate seeded with PrP(23-230) that had been pre-exposed to 10 μM Cu^{2+}. Fluorescence signals from PrP aggregates were averaged across five replicates and baseline corrected. Duration of the lag phase was determined to be the time point where ThT fluorescence intensity first increased beyond a predetermined threshold (5 times the standard deviation of a blank samples without PrP seeds and substrate). **(D)** Log-log plot of lag phase

duration for different amounts of PrP(23-230) seeds formed in 10 μM Cu^{2+}. The linear decrease in lag phase duration with increasing seed concentration is suggestive of an inverse correlation as predicted by theory. Similar analysis using seeds formed in **(E)** 10 μM Mn^{2+} and **(F)** 10 μM Ni^{2+} ions. In contrast to seeds formed in 10 μM Cu^{2+}, the threshold time does not decrease with increasing seed concentration indicating that aggregation in E and F are not templated. Error bars correspond to s. e. of mean. **(G)** Representative western blot of protein markers for neuroinflammation and neurodegeneration after organotypic slice cultures were exposed to PrP(23-230) seeds. **(H)** Quantification of Western blot band intensities show increased GFAP, PKC-δ and Bax protein expression upon exposure to copper-induced prion amyloids. Each group is represented by the mean ± SEM from at least four separate measurements.

2.3.4 PrPs incubated in Cu^{2+} induce neuroinflammation and neurodegeneration in brain tissue

Finally, we demonstrated the cytotoxic nature of Cu^{2+} induced PrP amyloids using a mouse organotypic slice culture assay. Slices of brain tissue from young mice were incubated with either misfolded PrP(23-230) prepared by a pre-exposure to Cu^{2+} ions (PrP-Cu^{2+}) or to PrP(23-230) monomers that were not pre-exposed to Cu^{2+} (PrP-No Metal). PrP-Cu^{2+} and PrP-No Metal, were prepared using an identical protocol as the RT-QuIC assay but without adding ThT (Methods). At the end of the treatment period, whole cell lysates were prepared from the brain slices and three protein markers for neuroinflammation and neurodegeneration were assayed using Western Blot analysis (Methods). We first characterized the levels of glial fibrillary acidic protein (GFAP), an important indicator of astrocytic activation during neuroinflammation. Western blot analysis indicated that PrP-Cu^{2+} results in higher GFAP immunoreactivity than PrP-No Metal or control slices which weren't exposed to PrP (Fig. 5, G and H). Next, since elevated gliosis enhances oxidative stress and downstream neuronal degeneration, we also analyzed the expression of Protein Kinase C δ (PKC-δ), a key oxidative stress-sensitive kinase of the novel PKC family. Activation of PKC-δ has been shown to induce

neuronal cell death and its relevance to oxidative stress mediated neurodegeneration is seen in many neurodegenerative diseases including Alzheimer's, Parkinson's and Prion disease [32-35]. We observed significantly increased levels of PKC-δ when brain slices were incubated with PrP-Cu^{2+}. Finally, since PKC-δ is a pro-apoptotic kinase, its activation results in a downstream apoptotic protein cascade and neuronal degeneration. We therefore investigated the activation of one such pro-apoptotic protein, Bax. Our western blot analysis showed an increase in Bax activity in mouse brain slices incubated with PrP-Cu^{2+}. We ruled out any role of the miniscule amounts of Cu^{2+} ions, introduced into the slice cultures upon addition of protein seeds, in enhancing these protein markers (fig. S13). Taken together, the notable increase in the levels of GFAP, PKC-δ, and Bax upon addition of PrP-Cu^{2+} suggest that Cu^{2+} induced misfolding contributes to neurotoxic signaling and mediates neuronal cell death (Fig. 5, G and H).

2.4 Discussion

Since PK degrades accessible regions of proteins that have flexible secondary structures, proteolytic digestion is routinely used to identify PrP domains that have misfolded into a β-sheet rich conformation. We therefore developed a fluorescence assay, with single-molecule sensitivity, to measure the PK-resistance, and consequently the misfolding, of PrP monomers. Our experiments demonstrate that the unstructured N-terminal region and Cu^{2+} ions mediate PrP monomer misfolding and PK-resistance. It has been reported that Cu^{2+} binding reduces the folding stability of PrP, making a conformation transition more thermodynamically favorable [36]. In fact, ensemble biophysical experiments with PrP(23-231) show that while exposure to Cu^{2+} ions results

in a PK-resistant conformation, PrP^{sen} isoforms are observed in Zn^{2+}, Mg^{2+}, Ca^{2+}, Mn^{2+} and Fe^{2+} [12]. In contrast, PrP(90-231) forms oligomeric aggregates rich in β-sheets only at acidic pH [37].

It has been suggested that the direct inhibition of PK by copper may be responsible for previous reports of Cu^{2+} induced PK-resistance in ensemble biochemical experiments [38]. However, in our experiments, the Cu^{2+} ions are washed out of solution prior to PK incubation. Thus our fluorescence experiments demonstrate that Cu^{2+} promotes PK-resistance through a direct interaction with PrP(23-231) and not by inhibiting PK itself. Furthermore, we show that PrP(90-231) exposed to Cu^{2+} remains more PK-sensitive confirming that the Cu^{2+} ions do not inhibit PK activity. Previous studies show that PK degrades the N-terminal region of scrapie PrP without compromising pathogenicity [39, 40]. Experiments with transgenic mice also show that globular PrP is sufficient for propagating scrapie *in vivo* [41]. In contrast, both our single molecule and RT-QuIC data shows that in an *in-vitro* system, the N-terminal region is essential for Cu^{2+} induced misfolding and aggregation. One possible explanation for this inconsistency is that Cu^{2+} ions trigger PrP(23-231) monomers to misfold into a distinct, non-infective, strain that resists PK digestion [42].

In solution, unmodified PrP^{sen} as well as PrP^{sen} tagged with biotin and PEG tethers are equally sensitive to digestion by PK. We demonstrated this by using PK to digest unlabeled-PrP(23-230), biotin/PEG-labeled PrP(23-230), unlabeled-PrP(90-230) and biotin/PEG-labeled PrP(90-230), in solution (fig. S14). However, since the PrPs are tethered to a surface in our single molecule protease resistance assay, biotinylated fragments of the digested PrP that remain immobilized following PK treatment likely

account for the 40% fluorescence background measured when PrP^{sen} is digested with PK for 10 hours. A candidate biotinylated fragment is the region of PrP containing three biotinylation sites (Lys 185, Lys 194 and Lys 204), that lies in between the two Cys tether points. It is however important to note that since the tethered fragments are expected to contribute a similar fluorescence background signal when PrP^{res} is digested with PK, this isoform is at least 60% more PK-resistant than PrP^{sen}. Our hypothesis on the origin of the background signal is supported by the complete elimination of fluorescence signal measured in the negative control (biotinylated BSA), which lacks tethered, biotinylated peptide fragments (Fig. 2, D and E).

By comparing the interaction probabilities of PrP(23-231) and PrP(90-231), we show that the unstructured N-terminal tail has dominant influence on the first stage of oligomerization: *i.e.* PrP dimerization. Our data suggests that in the absence of the N-terminal tail, PrP(90-231) does not dimerize within the 100 ms time window that the AFM tip and surface are held in contact. It has previously been suggested that one of the key functions of the unstructured N-terminal region is in enhancing the efficiency of aggregation [43]. This effect may be mediated by the polybasic region (residues 23-31) in the disordered N-terminal region which is believed to be involved in PrP propagation [44]. Our data also indicates that the relative association constant of full-length PrP incubated with Cu^{2+} ions was 863 times higher than the affinity in the absence of divalent ions. This increase in relative K_A could arise from a Cu^{2+} induced tertiary structural conversion of PrP(23-231) [45].

In our single molecule force and fluorescence assay, PrPs were immobilized by reducing the intramolecular disulfide bond between Cys 179 and Cys 214 and

functionalizing the Cys residues with PEG tethers. Since tagging the Cys residues with PEG tethers can potentially destabilize PrP and promote misfolding, we used NMR spectroscopy, dynamic light-scattering, CD spectroscopy and thermal denaturation experiments to confirm that PrP molecules are in a stable, native, monomeric conformation upon functionalization. We confirmed that the PrPs remained in their native conformation following disulfide bond reduction and PEG labeling using CD spectroscopy (Supplementary Methods; fig. S15) and ^{1}H-NMR (Supplementary Methods; Supplementary Results; figs. S16 to S23). Our CD data shows that native, reduced and PEG-functionalized PrP(23-230) and PrP(90-230) retain an α-helix content that is characteristic of natively folded PrP (fig. S15) [46]. While PrP(23-230) and PrP (90-230) with intact disulfide bonds have α-helix contents of 28% and 43% respectively, the α-helix content, upon reduction of this linkage, is 27% and 35%. Similarly, the α-helix content of PrP(23-230) and PrP(90-230) bound to PEG is 21% and 32%. In contrast, previous CD measurements show that misfolded PrP is typically characterized by a low α-helix content of ~ 7% [47]. We also used ^{1}H-NMR to confirm that the native structure of PrP(23-230) and PrP (90-230) was preserved when the disulfide bond between Cys179 and Cys214 was reduced (figs. S20 and S21) and when PrP(23-230) and PrP (90-230) were functionalized with PEG linkers (figs. S22 and S23); the chemical shift dispersions of all these structures were similar to natively folded PrP.

Furthermore, by monitoring temperature-induced unfolding of PrP(23-230) and PrP(90-230) before and after attaching PEG tethers, we confirmed that PEG functionalization does not alter the stability of the folded PrP (fig. S24) [48]. Identical CD thermal denaturation curves with melting temperatures of 61.4 ± 0.4 °C and 62.9 ± 0.7 °C

were measured for PrP(23-230) before and after PEG functionalization (fig. S24A). Similarly, the melting temperatures (65.1 ± 0.4 °C and 65.9 ± 0.8 °C) of PrP(90-230) before and after attaching PEG tethers were similar (fig. S24B). Finally, light scattering data showed that while 100% of PrP(23-230) and 96% of PrP(90-230) with intact disulfide bonds existed as monomers, reducing this linkage and functionalizing the Cys residues with PEG linkers does not significantly decrease the fraction of monomers (Supplementary Methods; fig. S25). In agreement with our results, previous studies show that reducing the disulfide bond between Cys179 and Cys214 does not alter the pathway for PrP misfolding [49] and that misfolding occurs without disulfide exchange [50].

Since previous studies have shown that aging of PrP promotes oligomerization and copper-induced PK resistance [12], we used PrP(23-230) and PrP(90-230) within one month after purification. Furthermore, we used CD and dynamic light-scattering to confirm that the proteins retained their native conformation and remained in a monomeric state, even after 2 months of storage (Supplementary Methods; figs. S26 and S27). After a 2 month storage period, 96% of PrP(23-230) and 95% of PrP(90-230) remained monomeric (fig. S26) with α-helix contents of 38% and 33% respectively (fig. S27). Since the presence of even trace amounts of misfolded PrP will result in spontaneous aggregation in an RT-QuIC experiment, the absence of large deviations between replicates (figs. S9 to S11) confirms that the PrP substrate used in our experiments is in a native conformation.

We also confirmed that the presence of an N-terminal His-tag does not promote PrP denaturation or aggregation by using CD and dynamic light-scattering (Supplementary Methods; figs. S26 and S27). After His-tag cleavage, 92% of PrP(23-

230) and 97% of PrP(90-230) remained in a monomeric state (fig. S26) and had α-helix contents of 43% (fig. S27). Finally, it is important to note that the Cu^{2+} ion induced structural conversion of full-length PrP is not an artifact arising from the presence of His-tag since the globular PrP constructs, which also contain an identical His-tag, do not show a Cu^{2+} induced misfolding and aggregation.

Since the prion hypothesis proposes that misfolded seeds can template the aggregation of normal PrP into amyloid fibrils, we tested if PrP exposed to different divalent ions can serve as seeds using RT-QuIC. Our data showed that a more rapid onset of aggregation was measured as the seed concentration increased. This correlation is predicted by theory and has previously been used to quantify the aggregation kinetics of two major forms of amyloid-beta peptides [30, 51].

Several studies have shown that different synthetic prion fragments promote neuroinflammatory and apoptotic responses in neuronal cells without evidence of being infectious [52-54]. Therefore, to investigate the neurotoxic effect of Cu^{2+} induced misfolded-PrP, we employed a mouse organotypic slice culture assay, which had been previously adopted as an excellent *ex vivo* model of progressive neurodegenerative disorders [35, 55]; a key advantage of this platform is that intact tissue explants preserve the basal cellular and molecular environment of local circuits similar to that of the whole brain. In the CNS, many neuropathological disorders are characteristically accompanied by activation and proliferation of glial cells that promote neuroinflammation, which is initiated relatively early in the disease process [56-59]. Consistent with this, we observed an increase in the levels of the neuroinflammatory marker GFAP, upon exposing brain slices to misfolded PrP. Furthermore, since oxidative stress is a critical initiator of

neurotoxic insult [32-35], we analyzed the expression of the oxidative stress-sensitive kinase PKC-δ; notably, an increase in PKC-δ expression was measured when we exposed brain slice cultures to PrP amyloids. Finally, consistent with the finding that activated PKC-δ triggers the redistribution and activation of Bax, a Bcl-2 family protein that can directly induce cytochrome c release and the downstream apoptotic cascade [60, 61], we observed markedly increased levels Bax, in brain slices exposed to misfolded PrP. Overall, our experiments demonstrate the neurotoxicity of amyloidogenic structures resulting from prion protein interaction with Cu^{2+} ions. The neuronal cell death we observed in the mouse organotypic slice culture assay directly confirms the effect of abnormal prion protein, ruling out any role for the minuscule amount of Cu^{2+} ions (100 nM) in the oligomeric protein solution (fig. S13). In fact, copper-induced neurotoxicity normally requires concentrations ranging from 0.1 – 1 mM in *ex vivo* and *in vivo* experimental models [62, 63].

While its physiological function is still uncertain, it has been proposed that cellular PrP plays a role in regulating Cu^{2+} homeostasis [64-66] and protecting neurons against apoptosis induced by Bax [67]. Previous studies show that PrP is associated with Cu^{2+} ions *in vivo* [9, 68]. While the average Cu^{2+} concentration in the human brain is about 80 μM [65], concentrations as high as 1.3 mM and 0.4 mM are found in the *locus ceruleus* and the *substantia nigra* respectively [69]. Consequently, our experimental conditions mimic these physiological Cu^{2+} concentrations. Our data suggests that an excess of Cu^{2+} ions contributes to the structural instability and oligomerization of cellular PrP and induces neuroinflammation and neurodegeneration.

2.5 Materials and Methods

2.5.1 PrP constructs used

We used recombinant, full-length, human PrP that was either purified in house (PrP(23-230)) or purchased commercially (PrP(23-231); Jena Bioscience). The biophysical properties of full-length PrP was compared with home grown, (PrP(90-230)) or commercially obtained (PrP(90-231); Jena Bioscience) globular PrP. ^{1}H-NMR spectra showed that the structure of PrP(23-231) and PrP(90-231) are similar to PrP(23-230) and PrP(90-230) respectively (Supplementary Results; figs. S16 to S19). Plasmids for recombinant human PrP(23-230) and PrP(90-230) with an N-terminal histidine tail and an engineered thrombin cleavage site were a generous gift from Prof. Kurt Wüthrich (University of Zürich, Switzerland) [70]. The PrP constructs were expressed and purified as described previously with minor modifications [71]. Expression of human PrP(23-230) and PrP(90-230) proteins were performed in *E.coli* BL21(DE3) in a pRSET-A vector expression system. At OD_{600} = 0.5, prion protein expression was induced with isopropyl β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM; after 8 hours of IPTG induction, bacterial cell pellets were harvested by centrifugation. Cell lysis was carried in 25 mL buffer A (6M GdmCl, 10 mM Tris-HCl, 100 mM Na_2HPO_4, 10 mM reduced glutathione, pH 8.0). The lysate was incubated with 5 mL Ni-NTA agarose beads (Qiagen) on a rocking shaker at room temperature for 1 hour; 5 mM imidazole for PrP(90-230) or 10 mM imidazole for PrP(23-230) was added during the incubation to prevent the non-specific binding of protein. The resin was transferred to a column and the beads were washed with a 200 mL gradient of buffer A to buffer B (10 mM Tris-HCl, 100 mM Na_2HPO_4, pH 8.0) to induce PrP refolding. The column was washed with 20 mL of buffer B plus 50

mM imidazole at 4 °C and PrP was eluted with buffer C (10 mM Tris-HCl, 100 mM Na_2HPO_4, 500 mM imidazole, pH 5.8) and dialyzed against 10 mM sodium phosphate buffer, pH 5.8. Protein concentrations were determined by Bradford protein assay using bovine serum albumin (BSA) as reference. Protein fractions collected after elution were assayed for presence and purity of prion protein using SDS-PAGE with Krypton fluorescent protein staining (Pierce) and with Western blot analysis using either POM1 (Prionatis AG) or SAF32 (Cayman Chemical) [72, 73] (Fig. 1C). The purified PrP was stored at 4 °C and used within one month.

2.5.2 Surface and AFM tip functionalization and PrP immobilization

The protocols for functionalization of AFM cantilevers and substrates for AFM-FS experiment have been described elsewhere [74, 75]. Silicon nitride AFM cantilevers (Olympus) and glass substrates (Fisher) were sequentially cleaned in a solution of piranha (H_2SO_4:H_2O_2 in 3:1 vol/vol ratio), deionized water and acetone. The cantilevers and substrates were then silanized using a 2% (vol/vol) 3-aminopropyltriethoxysilane (Sigma) solution in acetone for 30 minutes and functionalized with a monolayer of maleimide-PEG (maleimide-PEG-succinimidyl valerate; MW 3,400; Laysan Bio). For PK digestion experiments, glass-coverslips were functionalized with a mixture of 20% maleimide-PEG and 80% mPEG-succinimidyl valerate (MW 2,000; Laysan Bio). Prior to PrP immobilization, the disulfide bond was reduced with an equal volume of 10 mM tris(2-carboxyethyl)phosphine (TCEP) for 3 hours at 4 °C. The reduced PrP(23-231) or PrP(90-231), at a concentration of 1 μg/mL, was incubated on the surface in reaction buffer (50 mM sodium phosphate, 50 mM NaCl, 10 mM EDTA, pH7.2) for 1 or 4 hours respectively

at room temperature. Unlinked maleimide groups were quenched using 1 mM cysteine and 10 mM TCEP in reaction buffer overnight at 4 °C.

2.5.3 Proteinase-K digestion experiments

Biotinylation of PrP for fluorescence measurements was performed by mixing 10 mM Sulfo-NHS-LC-Biotin (Thermo Scientific) with 10 μg/mL of human recombinant PrP(23-231) or PrP(90-231) in reaction buffer for 3 hours at 4 °C. The reaction was quenched by adding 1 mg/10μL L-Lysine (Sigma) for 1 hour at room temperature. The biotinylated PrP was immobilized on PEG functionalized glass coverslips and incubated overnight either with or without 1 mM Mn^{2+}, 1 mM Ni^{2+} or 1 mM Cu^{2+}. Following incubation, the divalent ions were washed away and the immobilized PrP was treated with 100 μg/mL PK (Sigma) in buffer (25 mM Tris-HCl, 5 mM $CaCl_2$, 1 mM EDTA, pH 7.4) for different time courses at 37 °C. Following PK digestion, the PrP was labeled using 0.7 μg/mL Alexa-555 conjugated streptavidin (Invitrogen) for 10 minutes at room temperature and rinsed with buffer (25 mM Tris-HCl, 100 mM NaCl, pH 7.4). Prior to incubation with fluorescent streptavidin, the substrate was incubated with BSA (1 mg/mL, 2 hours) to minimize non-specific fluorescence background. For every measurement, a blank-control sample without PrP was prepared in parallel using an identical protocol. A 3' Tetramethylrhodamine (TMR) modified DNA (IDT) was used as positive-control because of its resistance to PK digestion. The 5' terminal of the DNA was modified by di-thiol for surface immobilization. A biotin-conjugated BSA (Sigma), which is sensitive to PK digestion, was used as negative-control. A 50 nM solution of biotin-BSA was immobilized on maleimide PEG as described above.

For each time point, eight to twenty fluorescence images were acquired for both the fluorescent samples (either experiment, positive-control or negative-control) and the blank-control samples. We randomly selected one image each from the fluorescent and blank-control datasets and recorded their difference in fluorescence intensity. This process was repeated 100,000 times. The mean and standard deviation of these values were normalized to the intensity of an identically prepared sample without PK treatment.

2.5.4 Single molecule AFM-FS

Force-ramp experiments were performed in a pH 7.4 measurement buffer (25 mM Tris-HCl, 100 mM NaCl) using an Agilent 5500 AFM system with closed-loop scanner. The AFM tip and substrate were brought into contact for 0.1 sec, then withdrawn at 6 or 11 different speeds ranging from 27 $nm \cdot sec^{-1}$ to 10,000 $nm \cdot sec^{-1}$. Cantilever spring constants were measured using the thermal fluctuation method. Force measurements were performed continuously for 24 hours at room temperature; a LabView script allowed us to operate the AFM in an automated mode to eliminate operator bias. For every experiment performed with divalent metal ions, we first acquired data for 24 hours in the measurement buffer and then acquired data for the next 24 hours in a pH 7.4 buffer with 1 mM divalent metal ions (25 mM Tris-HCl, 97 mM NaCl, 1 mM $NiCl_2$/$MnCl_2$ /$CuCl_2$) to enable a direct comparison. Prior to the experiment, the AFM cantilever and substrate functionalized with PrP were incubated in 1 mg/mL BSA for 3 hours to minimize nonspecific protein adhesion. Each unbinding force trace was fitted to an extended freely-jointed chain (FJC) model (Eq. 1) [25] which describes the stretching of a polymer, *L*, under stretching force, *F*.

$$L = L_C \times \left(coth\left(\frac{FL_K}{k_BT}\right) - \frac{k_BT}{FL_K}\right) + \frac{FL_C}{L_m K_S} \quad (1)$$

The model contains only one fitting parameter: the contour length of polyethylene glycol (PEG) tethers, L_C. The values of all other parameters were obtained from the literature [25]. We used a value of 0.7 nm for L_K (the Kuhn length of PEG), 0.2837 nm for L_m (the average length of a PEG monomer) and 150,000 pN/nm for K_S (the stiffness of a PEG monomer). Force curves were fit using a total least squares (TLS) fitting protocol; goodness of fit was estimated from the residuals. For each experimental condition, the distribution of L_C from selected events was fitted to a Gaussian distribution. Only events with fitted L_C within the range of Gaussian center +/- one standard deviation were used to calculate the probability of interaction and off-rate.

2.5.5 Calculation of relative on-rates

The detailed derivation of the formulas used for the calculation of relative on-rates can be found in [27]. Briefly, the specific binding probability was calculated as $P_{corrected}=(P_m-P_{control})/(1-P_{control})$, where P_m is the measured binding probability when PrP is immobilized on both tip and substrate and $P_{control}$ is the measured binding probability when PrP is immobilized only on the substrate. The average number of bonds formed between opposing PrPs during each tip-substrate contact is a function of contact area between AFM tip and substrate A_c, the protein density on opposing surfaces ρ_a and ρ_b, the contact time t and the on-rate k_{on} and off-rate k_{off}. Their mathematical relation is given by $-\ln(1-P_{corrected}) = A_c\rho_a\rho_b(1-e^{-k_{off}t})k_{on}/k_{off} \cong A_c\rho_a\rho_b t\, k_{on}$. Assuming A_c, ρ_a, ρ_b are fixed, the on-rate k_{on} is proportional to $-\ln(1-P_{corrected})$.

2.5.6 RT-QuIC assay

PrP seeds for the RT-QuIC assay were created by incubating 6 μM of PrP with 10 μM divalent metal ions ($NiCl_2$/$MnCl_2$/$CuCl_2$) in Thioflavin T (ThT) assay buffer (20 mM Tris-HCl, 150 mM NaCl, 10 μM ThT, pH 7.4). The sample was incubated in a Cytation3 multi-mode microplate reader (BioTek) at 37°C for 64-95 hours with continuous cycles of 1-min shaking alternating with 1-min resting periods. The formation of prion seeds was detected as an increase in ThT fluorescence intensity, which was measured every 15 min. Seeds were created with 3 replicates for each condition. The PrP(23-230) seeds were serially diluted to a final amount of 1.5 ng, 150 pg, 15 pg, 1.5 pg or 150 fg in 10 μL of buffer (20 mM Tris-HCl, 150 mM NaCl, pH 7.4). Reactions were prepared in a black 96-well, optical-bottomed plate (Nunc) with 90 μL of RT-QuIC master mix (final concentration of 20 mM Tris-HCl, 150 mM NaCl, 10 μM ThT, 1 mM EDTA and 6 μM of recombinant human PrP(23-230), pH 7.4) and then 10 μL of seed were loaded into wells. Sealed plates were inserted into a Cytation3 multi-mode microplate reader, incubated at 37ºC and shaken intermittently (1-min shake/rest cycle) at 807 cpm in a double orbital configuration for 71-95 h. ThT fluorescence was recorded every 15 min throughout the experiment. In each experiment, 3-5 replicates of blank samples (150 mM NaCl, 10 uM ThT, 1 mM EDTA and 20 mM Tris-HCl, pH 7.4) were used. Experimental samples using seeds generated in either $NiCl_2$/$MnCl_2$ or $CuCl_2$ were examined in RT-QuIC assay with 4 and 5 replicates, respectively. For the same seeding condition, fluorescence intensity was averaged across 3-5 replicates, followed by baseline subtraction and robust locally weighted scatterplot smoothing. The duration of lag phase (T_{th}) was determined from the point where the ThT fluorescence intensity first reached a threshold value, where the

presence of amyloid can be detected [31]. This threshold was defined as 5x standard deviation of the fluorescence intensity from blank sample. Standard error of mean of T_{th} was calculated using a bootstrap with replacement protocol.

2.5.7 Organotypic slice culture assay and analysis

All procedures involving animal handling were approved by the Institutional Animal Care and Use Committee (IACUC) at Iowa State University and conducted in compliance with Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC) accreditation. Organotypic slice cultures were prepared as previously described with some modifications [34, 55]. Briefly, 9–12 day-old C57BL/6 pups were anesthetized using isoflurane and after decapitation, brains were quickly removed and brain blocks were prepared in 2% (w/v) low-melting-point agarose (Invitrogen 15517–022) in HBSS. Once the agarose was cooled on ice, 350-µm thick cortical-striatal slices were prepared using a Compresstome™ Vf-300 microtome (Precisionary Instruments Inc.). Slices were transferred to Millicell-CM Biopore PTFE membrane inserts (Millipore PICM03050) and maintained in slice culture medium (50% MEM, 25% Basal Eagle medium, 25% horse serum, 0.65% glucose supplemented with penicillin/streptomycin and glutamax) in a humidified 37 °C incubator with 5% CO_2 and 95% air for 7 days. The PrP(23-230) seeds used in the organotypic assay were prepared similarly to seeds used in the RT-QuIC assay except that ThT was not used in the buffer. Seed materials were incubated and shaken using a thermomixer as same condition as the ThT assay either in the presence (PrP-Cu^2) or absence (PrP-No Metal) of Cu^{2+} ions. A separate ThT assay was carried out in parallel in order to monitor the seeds formation. The formation of PrP-No Metal and

PrP-Cu^{2+} seeds were terminated when we observed the increase of ThT fluorescence in the ThT assay. After one week, organotypic slice cultures were exposed to 10μL of the seed in 1ml of slice culture medium for 48hr and 90% of the media was exchanged every other day for another week. At the end of the treatment, slices were collected, washed with PBS and whole cell tissue lysates were prepared using modified RIPA buffer containing protease and phosphatase inhibitor cocktail (Thermo Scientific, Waltham, MA), as described previously [35, 76-78]. Western blot analysis on the lysates were performed using GFAP (1:2000; Millipore), Bax (1:1000; Cell signaling) or PKCδ (1:1000; Santa Cruz) antibodies. To confirm equal protein loading, blots were re-probed with β-actin (1:15000; Sigma) antibody.

2.6 Acknowledgments

A.K. acknowledges the W. Eugene and Linda Lloyd Endowed Chair. We thank Kai-Sin Wang for advice on statistical analysis of the data and Dr. Bruce Fulton for technical assistance with acquisition of NMR spectra.

2.7 Author Contributions

S.S. directed the research. C.-F.Y. and S.S. designed the biophysical experiments while D.S.H. and A.K. designed the neurotoxicological assays. C.-F.Y. and D.S.H. performed the experiments and analyzed the data. S.S and C.-F.Y wrote the manuscript with assistance from A.K. and D.S.H.

2.8 References

[1] M. Jucker, L.C. Walker, Self-propagation of pathogenic protein aggregates in neurodegenerative diseases, Nature 501 (2013) 45-51.

[2] B. Frost, M.I. Diamond, Prion-like mechanisms in neurodegenerative diseases, Nat. Rev. Neurosci. 11 (2010) 155-159.

[3] R. Riek, S. Hornemann, G. Wider, M. Billeter, R. Glockshuber, K. Wüthrich, NMR structure of the mouse prion protein domain PrP(121-231), Nature 382 (1996) 180-182.

[4] R. Zahn, A. Liu, T. Lührs, R. Riek, C. von Schroetter, F. López García, M. Billeter, L. Calzolai, G. Wider, K. Wüthrich, NMR solution structure of the human prion protein, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 145-150.

[5] S. Hornemann, C. Schorn, K. Wüthrich, NMR structure of the bovine prion protein isolated from healthy calf brains, EMBO Rep. 5 (2004) 1159-1164.

[6] D.G. Donne, J.H. Viles, D. Groth, I. Mehlhorn, T.L. James, F.E. Cohen, S.B. Prusiner, P.E. Wright, H.J. Dyson, Structure of the recombinant full-length hamster prion protein PrP(29-231): the N terminus is highly flexible, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 13452-13457.

[7] G.S. Jackson, I. Murray, L.L. Hosszu, N. Gibbs, J.P. Waltho, A.R. Clarke, J. Collinge, Location and properties of metal-binding sites on the human prion protein, Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 8531-8535.

[8] E.D. Walter, D.J. Stevens, M.P. Visconte, G.L. Millhauser, The prion protein is a combined zinc and copper binding protein: Zn2+ alters the distribution of Cu2+ coordination modes, J. Am. Chem. Soc. 129 (2007) 15440-15441.

[9] D.R. Brown, K. Qin, J.W. Herms, A. Madlung, J. Manson, R. Strome, P.E. Fraser, T. Kruck, A. von Bohlen, W. Schulz-Schaeffer, A. Giese, D. Westaway, H. Kretzschmar, The cellular prion protein binds copper in vivo, Nature 390 (1997) 684-687.

[10] C.S. Burns, E. Aronoff-Spencer, G. Legname, S.B. Prusiner, W.E. Antholine, G.J. Gerfen, J. Peisach, G.L. Millhauser, Copper coordination in the full-length, recombinant prion protein, Biochemistry 42 (2003) 6794-6803.

[11] S.S. Leal, H.M. Botelho, C.M. Gomes, Metal ions as modulators of protein conformation and misfolding in neurodegeneration, Coord. Chem. Rev. 256 (2012) 2253-2270.

[12] K.F. Qin, D.S. Yang, Y. Yang, M.A. Chishti, L.J. Meng, H.A. Kretzschmar, C.M. Yip, P.E. Fraser, D. Westaway, Copper(II)-induced conformational changes and

protease resistance in recombinant and cellular PrP. Effect of protein age and deamidation, J. Biol. Chem. 275 (2000) 19121-19131.

[13] B.S. Wong, C. Vénien-Bryan, R.A. Williamson, D.R. Burton, P. Gambetti, M.S. Sy, D.R. Brown, I.M. Jones, Copper refolding of prion protein, Biochem. Biophys. Res. Commun. 276 (2000) 1217-1224.

[14] R.A. Moore, C. Herzog, J. Errett, D.A. Kocisko, K.M. Arnold, S.F. Hayes, S.A. Priola, Octapeptide repeat insertions increase the rate of protease-resistant prion protein formation, Protein Sci. 15 (2006) 609-619.

[15] S.L. Yu, S.M. Yin, N. Pham, P. Wong, S.C. Kang, R.B. Petersen, C.Y. Li, M.S. Sy, Ligand binding promotes prion protein aggregation--role of the octapeptide repeats, FEBS J. 275 (2008) 5564-5575.

[16] B. Caughey, G.S. Baron, B. Chesebro, M. Jeffrey, Getting a grip on prions: oligomers, amyloids, and pathological membrane interactions, Annu. Rev. Biochem. 78 (2009) 177-204.

[17] R. Atarashi, R.A. Moore, V.L. Sim, A.G. Hughson, D.W. Dorward, H.A. Onwubiko, S.A. Priola, B. Caughey, Ultrasensitive detection of scrapie prion protein using seeded conversion of recombinant prion protein, Nat. Methods 4 (2007) 645-650.

[18] J. Collinge, A.R. Clarke, A general model of prion strains and their pathogenicity, Science 318 (2007) 930-936.

[19] B. Oesch, D. Westaway, M. Wälchli, M.P. McKinley, S.B. Kent, R. Aebersold, R.A. Barry, P. Tempst, D.B. Teplow, L.E. Hood, A cellular gene encodes scrapie PrP 27-30 protein, Cell 40 (1985) 735-746.

[20] J.R. Silveira, G.J. Raymond, A.G. Hughson, R.E. Race, V.L. Sim, S.F. Hayes, B. Caughey, The most infectious prion protein particles, Nature 437 (2005) 257-261.

[21] E.L. Florin, V.T. Moy, H.E. Gaub, Adhesion forces between individual ligand-receptor pairs, Science 264 (1994) 415-417.

[22] P. Hinterdorfer, W. Baumgartner, H.J. Gruber, K. Schilcher, H. Schindler, Detection and localization of individual antibody-antigen recognition events by atomic force microscopy, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 3477-3481.

[23] B.H. Kim, N.Y. Palermo, S. Lovas, T. Zaikova, J.F.W. Keana, Y.L. Lyubchenko, Single-molecule atomic force microscopy force spectroscopy study of Aβ-40 interactions, Biochemistry 50 (2011) 5154-5162.

[24] R. Hervás, J. Oroz, A. Galera-Prat, O. Goñi, A. Valbuena, A.M. Vera, A. Gómez-Sicilia, F. Losada-Urzáiz, V.N. Uversky, M. Menéndez, D.V. Laurents, M. Bruix, M. Carrión-Vázquez, Common features at the start of the neurodegeneration cascade, PLoS Biol. 10 (2012) e1001335.

[25] F. Oesterhelt, M. Rief, H.E. Gaub, Single molecule force spectroscopy by AFM indicates helical structure of poly(ethylene-glycol) in water, New J. Phys. 1 (1999) 6.1-6.11.

[26] T.A. Sulchek, R.W. Friddle, K. Langry, E.Y. Lau, H. Albrecht, T.V. Ratto, S.J. DeNardo, M.E. Colvin, A. Noy, Dynamic force spectroscopy of parallel individual Mucin1-antibody bonds, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 16638-16643.

[27] S.E. Chesla, P. Selvaraj, C. Zhu, Measuring two-dimensional receptor-ligand binding kinetics by micropipette, Biophys. J. 75 (1998) 1553-1572.

[28] E. Evans, K. Ritchie, Dynamic strength of molecular adhesion bonds, Biophys. J. 72 (1997) 1541-1555.

[29] O.V. Bocharova, L. Breydo, V.V. Salnikov, I.V. Baskakov, Copper(II) inhibits in vitro conversion of prion protein into amyloid fibrils, Biochemistry 44 (2005) 6776-6787.

[30] S.I.A. Cohen, M. Vendruscolo, M.E. Welland, C.M. Dobson, E.M. Terentjev, T.P.J. Knowles, Nucleated polymerization with secondary pathways. I. Time evolution of the principal moments, J. Chem. Phys. 135 (2011) 065105.

[31] D.M. Henderson, K.A. Davenport, N.J. Haley, N.D. Denkers, C.K. Mathiason, E.A. Hoover, Quantitative assessment of prion infectivity in tissues and body fluids by real-time quaking-induced conversion, J. Gen. Virol. 96 (2015) 210-219.

[32] M. Alfa Cissé, K. Louis, U. Braun, B. Mari, M. Leitges, B.E. Slack, A. Fisher, P. Auberger, F. Checler, B. Vincent, Isoform-specific contribution of protein kinase C to prion processing, Mol. Cell. Neurosci. 39 (2008) 400-410.

[33] F. Ciccocioppo, P. Lanuti, M. Marchisio, F. Gambi, E. Santavenere, L. Pierdomenico, A. Bascelli, L. Velluto, D. Gambi, S. Miscia, Expression and phosphorylation of protein kinase C isoforms in Abeta(1-42) activated T lymphocytes from Alzheimers disease, Int. J. Immunopathol. Pharmacol. 21 (2008) 23-33.

[34] D.S. Harischandra, N. Kondru, D.P. Martin, A. Kanthasamy, H. Jin, V. Anantharam, A.G. Kanthasamy, Role of proteolytic activation of protein kinase Cδ in the pathogenesis of prion disease, Prion 8 (2014) 143-153.

[35] H. Jin, A. Kanthasamy, D.S. Harischandra, N. Kondru, A. Ghosh, N. Panicker, V. Anantharam, A. Rana, A.G. Kanthasamy, Histone hyperacetylation up-regulates protein kinase Cδ in dopaminergic neurons to induce cell death: relevance to epigenetic mechanisms of neurodegeneration in Parkinson disease, J. Biol. Chem. 289 (2014) 34743-34767.

[36] N.D. Younan, M. Klewpatinond, P. Davies, A.V. Ruban, D.R. Brown, J.H. Viles, Copper(II)-induced secondary structure changes and reduced folding stability of the prion protein, J. Mol. Biol. 410 (2011) 369-382.

[37] W. Swietnicki, M. Morillas, S.G. Chen, P. Gambetti, W.K. Surewicz, Aggregation and fibrillization of the recombinant human prion protein huPrP90-231, Biochemistry 39 (2000) 424-431.

[38] L.A. Stone, G.S. Jackson, J. Collinge, J.D.F. Wadsworth, A.R. Clarke, Inhibition of proteinase K activity by copper(II) ions, Biochemistry 46 (2007) 245-252.

[39] S.B. Prusiner, D.C. Bolton, D.F. Groth, K.A. Bowman, S.P. Cochran, M.P. McKinley, Further purification and characterization of scrapie prions, Biochemistry 21 (1982) 6942-6950.

[40] S.B. Prusiner, D.F. Groth, D.C. Bolton, S.B. Kent, L.E. Hood, Purification and structural studies of a major scrapie prion protein, Cell 38 (1984) 127-134.

[41] S. Supattapone, P. Bosque, T. Muramoto, H. Wille, C. Aagaard, D. Peretz, H.O. Nguyen, C. Heinrich, M. Torchia, J. Safar, F.E. Cohen, S.J. DeArmond, S.B. Prusiner, M. Scott, Prion protein of 106 residues creates an artifical transmission barrier for prion replication in transgenic mice, Cell 96 (1999) 869-878.

[42] E. Quaglio, R. Chiesa, D.A. Harris, Copper converts the cellular prion protein into a protease-resistant species that is distinct from the scrapie isoform, J. Biol. Chem. 276 (2001) 11432-11438.

[43] S. Supattapone, T. Muramoto, G. Legname, I. Mehlhorn, F.E. Cohen, S.J. DeArmond, S.B. Prusiner, M.R. Scott, Identification of two prion protein regions that modify scrapie incubation time, J. Virol. 75 (2001) 1408-1413.

[44] J.A. Turnbaugh, U. Unterberger, P. Saá, T. Massignan, B.R. Fluharty, F.P. Bowman, M.B. Miller, S. Supattapone, E. Biasini, D.A. Harris, The N-terminal, polybasic region of PrP(C) dictates the efficiency of prion propagation by binding to PrP(Sc), J. Neurosci. 32 (2012) 8817-8830.

[45] A.K. Thakur, A.K. Srivastava, V. Srinivas, K.V.R. Chary, C.M. Rao, Copper alters aggregation behavior of prion protein and induces novel interactions between its N- and C-terminal regions, J. Biol. Chem. 286 (2011) 38533-38545.

[46] L. Whitmore, B.A. Wallace, DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data, Nucleic Acids Res. 32 (2004) W668-673.

[47] J.C. Sang, C.-Y. Lee, F.Y. Luh, Y.-W. Huang, Y.-W. Chiang, R.P.Y. Chen, Slow spontaneous α-to-β structural conversion in a non-denaturing neutral condition reveals the intrinsically disordered property of the disulfide-reduced recombinant mouse prion protein, Prion 6 (2012) 489-497.

[48] N.J. Greenfield, Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions, Nat. Protoc. 1 (2006) 2527-2535.

[49] H. Yu, X. Liu, K. Neupane, A.N. Gupta, A.M. Brigley, A. Solanki, I. Sosova, M.T. Woodside, Direct observation of multiple misfolding pathways in a single prion protein molecule, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 5283-5288.

[50] E. Welker, L.D. Raymond, H.A. Scheraga, B. Caughey, Intramolecular versus intermolecular disulfide bonds in prion proteins, J. Biol. Chem. 277 (2002) 33477-33481.

[51] G. Meisl, X. Yang, E. Hellstrand, B. Frohm, J.B. Kirkegaard, S.I.A. Cohen, C.M. Dobson, S. Linse, T.P.J. Knowles, Differences in nucleation behavior underlie the contrasting aggregation kinetics of the Aβ40 and Aβ42 peptides, Proc. Natl. Acad. Sci. U.S.A. 111 (2014) 9384-9389.

[52] G. Forloni, N. Angeretti, R. Chiesa, E. Monzani, M. Salmona, O. Bugiani, F. Tagliavini, Neurotoxicity of a prion protein fragment, Nature 362 (1993) 543-546.

[53] J. Gong, A. Jellali, V. Forster, J. Mutterer, E. Dubus, W.D. Altrock, J.A. Sahel, A. Rendon, S. Picaud, The toxicity of the PrP106-126 prion peptide on cultured photoreceptors correlates with the prion protein distribution in the mammalian and human retina, Am. J. Pathol. 170 (2007) 1314-1324.

[54] S. Vilches, C. Vergara, O. Nicolás, G. Sanclimens, S. Merino, S. Varón, G.A. Acosta, F. Albericio, M. Royo, J.A. Del Río, R. Gavín, Neurotoxicity of prion peptides mimicking the central domain of the cellular prion protein, Plos One 8 (2013) e70881.

[55] J. Falsig, A. Aguzzi, The prion organotypic slice culture assay--POSCA, Nat. Protoc. 3 (2008) 555-562.

[56] B.C. Gray, P. Skipp, V.M. O'Connor, V.H. Perry, Increased expression of glial fibrillary acidic protein fragments and mu-calpain activation within the hippocampus of prion-infected mice, Biochem. Soc. Trans. 34 (2006) 51-54.

[57] L.F. Eng, R.S. Ghirnikar, GFAP and Astrogliosis, Brain Pathol. 4 (1994) 229-237.

[58] W. Kamphuis, C. Mamber, M. Moeton, L. Kooijman, J.A. Sluijs, A.H.P. Jansen, M. Verveer, L.R. de Groot, V.D. Smith, S. Rangarajan, J.J. Rodríguez, M. Orre, E.M. Hol, GFAP isoforms in adult mouse brain with a focus on neurogenic astrocytes and reactive astrogliosis in mouse models of Alzheimer disease, PLoS One 7 (2012) e42823.

[59] R. Kordek, P.P. Liberski, R. Yanagihara, S. Isaacson, D.C. Gajdusek, Molecular analysis of prion protein (PrP) and glial fibrillary acidic protein (GFAP) transcripts in experimental Creutzfeldt-Jakob disease in mice, Acta Neurobiol. Exp. 57 (1997) 85-90.

[60] S.R. Cerda, R. Mustafi, H. Little, G. Cohen, S. Khare, C. Moore, P. Majumder, M. Bissonnette, Protein kinase C delta inhibits Caco-2 cell proliferation by selective changes in cell cycle and cell death regulators, Oncogene 25 (2006) 3123-3138.

[61] L.A. Sitailo, S.S. Tibudan, M.F. Denning, Bax activation and induction of apoptosis in human keratinocytes by the protein kinase C delta catalytic domain, J. Invest. Dermatol. 123 (2004) 434-443.

[62] I. Paris, A. Dagnino-Subiabre, K. Marcelain, L.B. Bennett, P. Caviedes, R. Caviedes, C.O. Azar, J. Segura-Aguilar, Copper neurotoxicity is dependent on dopamine-mediated copper uptake and one-electron reduction of aminochrome in a rat substantia nigra neuronal cell line, J. Neurochem. 77 (2001) 519-529.

[63] A.R. White, G. Multhaup, F. Maher, S. Bellingham, J. Camakaris, H. Zheng, A.I. Bush, K. Beyreuther, C.L. Masters, R. Cappai, The Alzheimer's disease amyloid precursor protein modulates copper-induced toxicity and oxidative stress in primary neuronal cultures, J. Neurosci. 19 (1999) 9170-9179.

[64] E. Gaggelli, H. Kozlowski, D. Valensin, G. Valensin, Copper homeostasis and neurodegenerative disorders (Alzheimer's, prion, and Parkinson's diseases and amyotrophic lateral sclerosis), Chem. Rev. 106 (2006) 1995-2044.

[65] J. Stöckel, J. Safar, A.C. Wallace, F.E. Cohen, S.B. Prusiner, Prion protein selectively binds copper(II) ions, Biochemistry 37 (1998) 7185-7193.

[66] C.J. Choi, A. Kanthasamy, V. Anantharam, A.G. Kanthasamy, Interaction of metals with prion protein: possible role of divalent cations in the pathogenesis of prion diseases, Neurotoxicology 27 (2006) 777-787.

[67] L. Westergard, H.M. Christensen, D.A. Harris, The cellular prion protein (PrP(C)): its physiological function and role in disease, Biochim. Biophys. Acta 1772 (2007) 629-644.

[68] M.L. Kramer, H.D. Kratzin, B. Schmidt, A. Römer, O. Windl, S. Liemann, S. Hornemann, H. Kretzschmar, Prion protein binds copper within the physiological concentration range, J. Biol. Chem. 276 (2001) 16711-16719.

[69] I.E. Dreosti, R.M. Smith, Neurobiology of the Trace Elements: Volume 2: Neurotoxicology and Neuropharmacology, Humana Press, New York, 1983.

[70] S. Hornemann, B. Christen, C. von Schroetter, D.R. Pérez, K. Wüthrich, Prion protein library of recombinant constructs for structural biology, FEBS J. 276 (2009) 2359-2367.

[71] R. Zahn, C. von Schroetter, K. Wüthrich, Human prion proteins expressed in Escherichia coli and purified by high-affinity column refolding, FEBS Lett. 417 (1997) 400-404.

[72] M. Polymenidou, R. Moos, M. Scott, C. Sigurdson, Y.-Z. Shi, B. Yajima, I. Hafner-Bratkovic, R. Jerala, S. Hornemann, K. Wuthrich, A. Bellon, M. Vey, G. Garen, M.N.G. James, N. Kav, A. Aguzzi, The POM monoclonals: a comprehensive set of antibodies to non-overlapping prion protein epitopes, PLoS One 3 (2008) e3872.

[73] C. Féraudet, N. Morel, S. Simon, H. Volland, Y. Frobert, C. Créminon, D. Vilette, S. Lehmann, J. Grassi, Screening of 145 anti-PrP monoclonal antibodies for their capacity to inhibit PrPSc replication in infected cells, J. Biol. Chem. 280 (2005) 11247-11258.

[74] S. Rakshit, Y. Zhang, K. Manibog, O. Shafraz, S. Sivasankar, Ideal, catch, and slip bonds in cadherin adhesion, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 18815-18820.

[75] K. Manibog, H. Li, S. Rakshit, S. Sivasankar, Resolving the molecular mechanism of cadherin catch bond formation, Nat. Commun. 5 (2014) 3941.

[76] D.S. Harischandra, H. Jin, V. Anantharam, A. Kanthasamy, A.G. Kanthasamy, α-Synuclein protects against manganese neurotoxic insult during the early stages of exposure in a dopaminergic cell model of Parkinson's disease, Toxicol. Sci. 143 (2015) 454-468.

[77] A.G. Kanthasamy, C. Choi, H. Jin, D.S. Harischandra, V. Anantharam, A. Kanthasamy, Effect of divalent metals on the neuronal proteasomal system, prion protein ubiquitination and aggregation, Toxicol. Lett. 214 (2012) 288-295.

[78] A. Ghosh, H. Saminathan, A. Kanthasamy, V. Anantharam, H. Jin, G. Sondarva, D.S. Harischandra, Z. Qian, A. Rana, A.G. Kanthasamy, The peptidyl-prolyl isomerase Pin1 up-regulation and proapoptotic function in dopaminergic neurons: relevance to the pathogenesis of Parkinson disease, J. Biol. Chem. 288 (2013) 21955-21971.

2.9 Supplementary Materials

Supplementary Methods

PrP surface density estimation: PrP surface density was determined using a calibration plot of fluorescence intensity *vs.* number of fluorescent-streptavidin (figs. S1 and S2). Cleaned glass coverslips were functionalized with 1 mg/mL biotin-conjugated BSA (Sigma) and incubated with Alexa-555 conjugated streptavidin (Invitrogen) at different concentrations (0.1-1 μg/mL). After rinsing with buffer (25 mM Tris-HCl, 100 mM NaCl,

pH 7.4), the sample was imaged on a home-built, single molecule sample scanning confocal fluorescence microscope. The number of immobilized fluorescent streptavidin was varied by changing incubation time (1-10 minutes); molecules in each 20x20 μm image were counted (fig. S1, B and C) and the total fluorescence intensity was measured. Background fluorescence intensity was estimated from the averaged integrated intensities of nine control samples without any fluorescent streptavidin (fig. S1A); this background intensity was subtracted from the intensity of all measured images. A calibration plot was generated using 24 images and fit to a first order polynomial (fig. S2).

When biotinylated PrP was immobilized on the surface and decorated with fluorescent streptavidin, the integrated fluorescence intensity of the image was recorded and the PrP surface density was determined using the fitted coefficients of the calibration plot. Since each PrP has 8-11 primary amine groups that can be labeled with biotins, every PrP was functionalized with at least 1 biotin. Furthermore, because streptavidin is two times larger in size than a PrP, no more than one streptavidin could bind to a single immobilized PrP. Consequently, when the immobilized, biotinylated PrP was incubated with fluorescent streptavidin, the fluorescence intensity saturated after 10 minutes and did not increase further even when the incubation period was lengthened by an order of magnitude.

NMR spectroscopy: One-dimensional ^{1}H-NMR spectra of purified PrP(23-230) and PrP(90-230) were recorded on a Bruker Avance II 700 spectrometer at a ^{1}H-frequency of 700 MHz at 20 °C with an acquisition time of 1.47 sec and the sweep width of 11,161 Hz (figs. S18 and S19). Samples were concentrated by using Centriprep® centrifugal filter

with 10 kDa membrane (EMD Millipore) at 1,000x g to a final concentration of 80 μM for PrP(23-230) and 90 μM for PrP(90-230). Spectra were measured before and after adding 5 mM TCEP as reducing agent in a mixed solvent of 95% H_2O/ 5% D_2O and 10 mM sodium phosphate buffer, pH 5.8 (figs. S18 to S21). The spectra were analyzed by using the program Topspin 2.1 (Bruker). For PrP(23-231) and PrP(90-231), ^{1}H-NMR NMR spectra were acquired on an 800 MHz Varian spectrometer (Agilent) at 20 °C by Jena Bioscience (figs. S16 and S17). The protein concentration was 50 μM in 10 mM sodium acetate buffer, pH 4.5.

To confirm that PrP remained natively folded after PEG functionalization, ^{1}H-NMR spectra was measured for 21 μM PrP(23-230) in 10 mM sodium phosphate buffer at pH 7.2. The PrP(23-230) was first incubated with 630 μM TCEP at 4 °C for 3 hours and subsequently incubated with 210 μM maleimide-PEG at 20 °C for 1 hour (figs. S22 and S23). The measured spectra was compared to ^{1}H-NMR spectra acquired from a sample containing an identical concentration of PrP(23-230) and maleimide-PEG, but lacking TCEP. Similarly, ^{1}H-NMR spectra was measured for 141 μM PrP(90-230), incubated with 4.2 mM TCEP and 1.4 mM maleimide-PEG. The PEG-functionalized PrP(90-230) spectra was compared to an identically prepared control sample lacking TCEP. D_2O (final concentration 5%) was added right before the measurement.

Dynamic light scattering: Dynamic light scattering measurements (DLS) were performed at 20 °C on Zetasizer Nano ZS (Malvern) using disposable cuvette. The translational diffusion coefficients of protein molecules were derived from the autocorrelation function of scattered light intensity. Particle sizes and size distributions

were analyzed by using software Zetasizer (Malvern). DLS experiments testing the effects of disulfide bond reduction and maleimide-PEG functionalization was carried out with protein that was 3 weeks old (fig. S25). Protein concentration was 15.6 μM for PrP(23-230) and 105 μM for PrP(90-230) in 10 mM sodium phosphate buffer, pH 7.2. Measurements were performed in three different conditions: before and after incubating with 1 mM TCEP at 4 °C for 3 hours and after further incubating with 300 μM maleimide-PEG at 20 °C for 3 hours. As a reference, a control sample containing only 300 μM maleimide-PEG incubated with 1 mM TCEP at 20 °C for 3 hours was measured.

Experiments on the effects of protein aging and the role of the N-terminal his-tag were performed with 5.2 μM of 2-month-old PrP(23-230) and 9.1 μM of 2-month-old PrP(90-230) (fig. S26). Measurement were performed before and after incubating with 0.1 unit/mL thrombin enzyme at 20 °C for 1 hour [71]. The cleavage of His-tag was confirmed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (fig. S26A).

Circular dichroism (CD) spectroscopy: CD spectra of 5 μM PrP(23-230) or 5 μM PrP(90-230) in 10 mM sodium phosphate buffer, pH 7.2 was measured before and after reduction with 150 μM TCEP at 4 °C for 3 hours. Samples were further incubated with 50 μM maleimide-PEG at 20 °C for 1 hour and CD spectra was re-recorded (fig. S15). The spectra were acquired at 4 °C on a JASCO J-715 spectrometer (JASCO), using a water-cooled quartz cuvette. Three spectra were recorded in 0.2 nm intervals from 180 to 260 nm for each sample and then averaged over three scans for each sample. Secondary structure was calculated using K2D algorithm with the online server DICHROWEB [46].

To study the effect of the N-terminal His-tag on secondary structure, we cleaved the His-tag by incubating 2-month-old PrP with 0.1 unit/mL thrombin enzyme at 20 °C for 1 hour (figs. S26A and S27). The effect of age on secondary structure was tested under identical data acquisition conditions using 2 month old PrP (fig. S27).

To study the thermal stability of PrP(23-230) and PrP(90-230) before and after pegylation (fig. S24), CD signal at 222 nm was recorded every 2 °C from 44 to 88 °C at a rate of 0.4 °C per minute. The melting temperature (T_m) was calculated by fitting data to a two-state transition model of a PrP converting between folded and unfolded forms. The data was corrected for pre- and post-transition linear changes in ellipticity as a function of temperature [48].

Supplementary Results

NMR spectroscopy: The one-dimensional ^{1}H-NMR spectra of PrP(23-231) and PrP(90-231) (figs. S16 and S17) and PrP(23-230) and PrP(90-230) (figs. S18 and S19) share several common features. Their chemical shift dispersion is typical for natively folded prion protein (PrP). The resonance lines near 0 ppm indicate that the globular C-terminal domain is conserved and PrP are folded natively [71]. Higher indole N-H resonance lines near 10 ppm are observed in the spectra of full-length PrP due to the tryptophan residues located in the unstructured N-terminal tail [71]. The differences between PrP(23-231) and PrP(23-230) or between PrP(90-231) and PrP(90-230) are largely due to the different signal-to-noise ratios. The difference in intensity in the range of 7.5 to 9 ppm is caused

by differences in pH. Overall, ^{1}H-NMR spectra indicate that the structure of PrP(23-231) and PrP(90-231) are similar to PrP(23-230) and PrP(90-230) respectively.

In AFM and PK digestion experiments, the disulfide bond between Cys179 and Cys214 was reduced in order to tether PrP molecules to the surface. To examine if reduced PrP remains in its native conformation, ^{1}H-NMR spectra of PrP(23-230) and PrP(90-230) were measured in the presence of reducing agent (figs. S20 and S21) and in the presence of maleimide-PEG (figs. S22 and S23). The spectra were nearly identical to the spectra without reducing agent, indicating that the native structure of PrP was not altered when the disulfide bond was reduced or when the Cys residues were linked to a PEG tether.

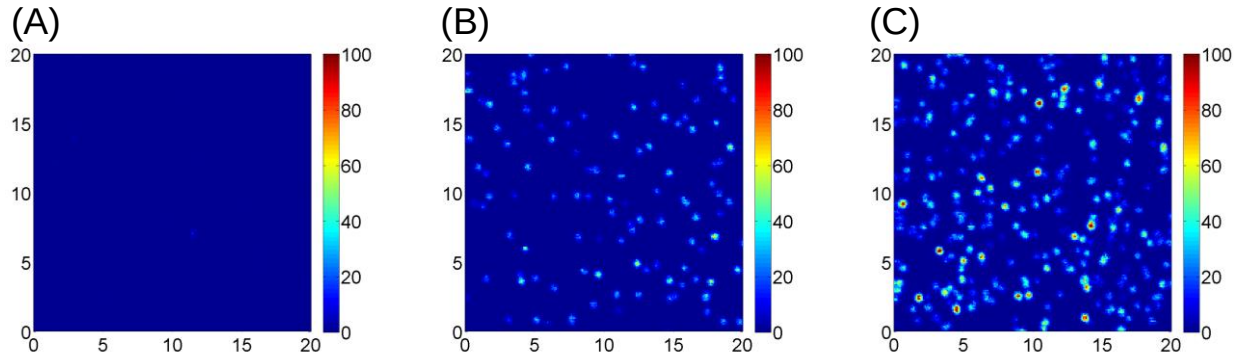

Figure S1. Fluorescence images with different amount of fluorescent streptavidin on the substrate. (A) Control sample with biotin-conjugated bovine serum albumin on the substrate without streptavidin; **(B)** with 96 and **(C)** 195 streptavidin molecules on the substrate. Scan size: 20x20 μm.

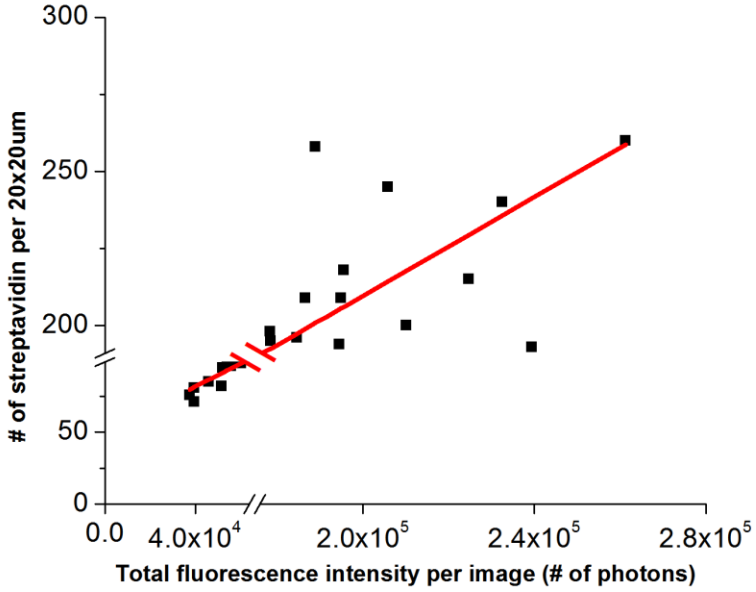

Figure S2. Calibration plot of number of streptavidin vs. fluorescence intensity. A total of 24 images with different streptavidin densities were used to generate this calibration plot. The slope and intercept of the least-squares regression line (red line) allowed us to estimate the number of prion proteins in each experiment, from their corresponding fluorescence intensity.

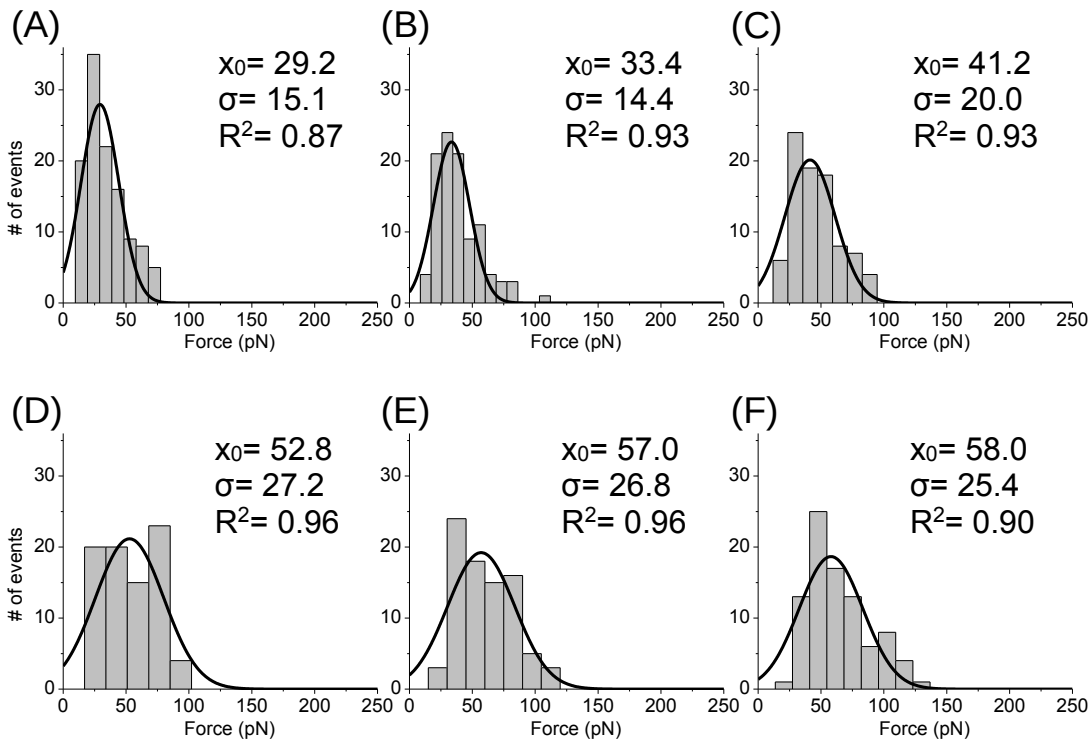

Figure S3. Histogram of unbinding force for PrP(23-231) measured in the absence of divalent ions. The histogram of rupture forces measured at pulling speed **(A)** 27 nm/s **(B)** 74 nm/s **(C)** 200 nm/s **(D)** 556 nm/s **(E)** 2000 nm/s and **(F)** 10000 nm/s. The bin width of histogram is determined by Freedman-Diaconis rule. The kernel density estimation of each force distribution is fitted to a Gaussian equation (black line): $f(x)=A*exp(-(x-x_0)^2/2\sigma^2)$.

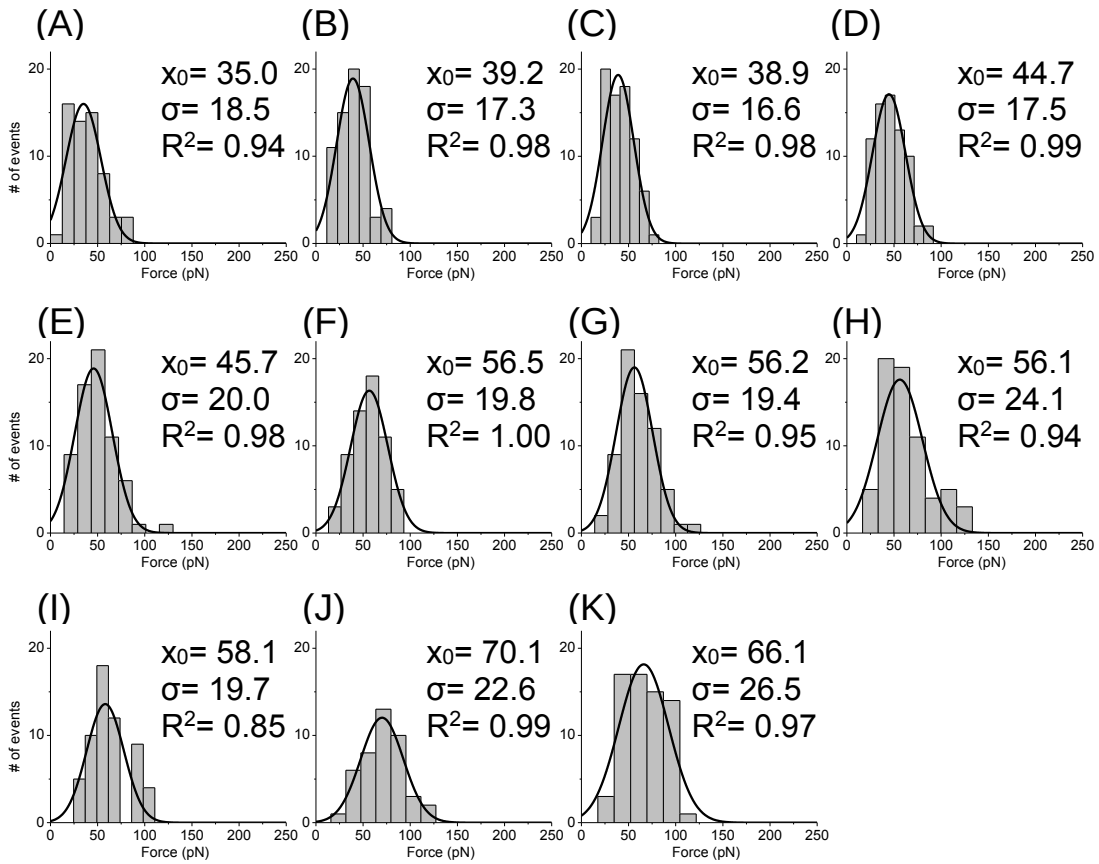

Figure S4. Histogram of unbinding force for PrP(23-231) measured in 1 mM Mn^{2+}. The histogram of rupture forces measured at pulling speed **(A)** 27 nm/s **(B)** 45 nm/s **(C)** 74 nm/s **(D)** 122 nm/s **(E)** 200 nm/s **(F)** 333 nm/s **(G)** 556 nm/s **(H)** 1000 nm/s **(I)** 2000 nm/s **(J)** 5000 nm/s and **(K)** 10000 nm/s. The bin width of histogram is determined by Freedman-Diaconis rule. The kernel density estimation of each force distribution is fitted to a Gaussian equation (black line): $f(x)=A*exp(-(x-x_0)^2/2\sigma^2)$.

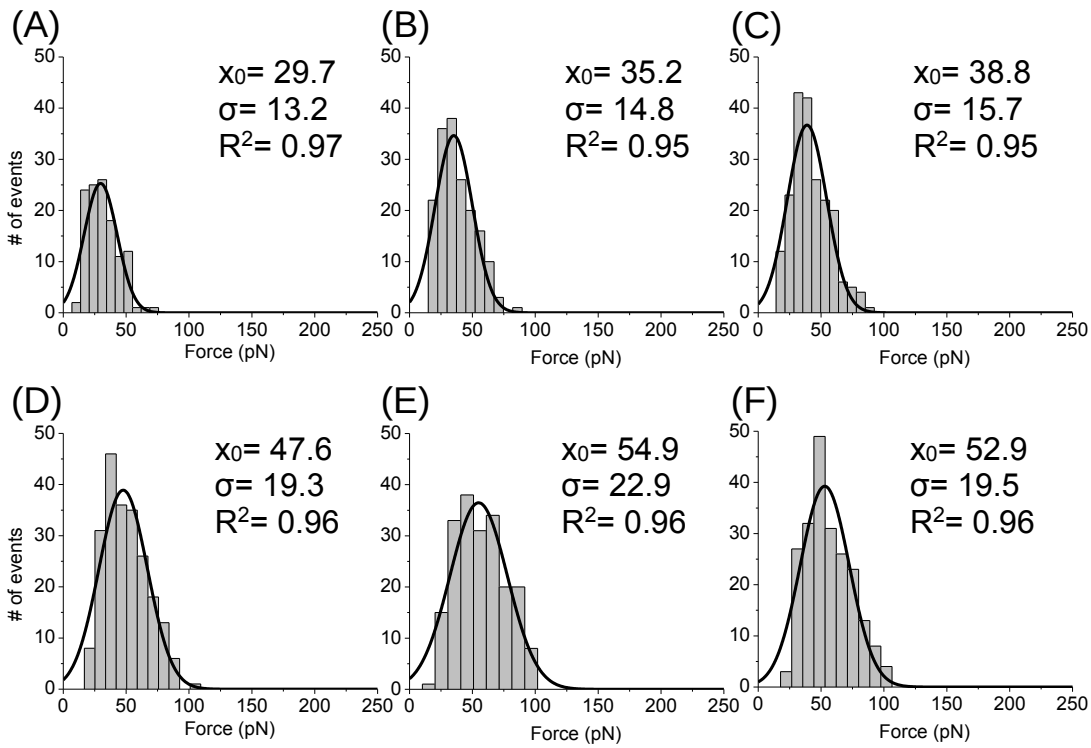

Figure S5. Histogram of unbinding force for PrP(23-231) measured in 1 mM Ni^{2+}. The histogram of rupture forces measured at pulling speed **(A)** 27 nm/s **(B)** 74 nm/s **(C)** 200 nm/s **(D)** 556 nm/s **(E)** 2000 nm/s and **(F)** 10000 nm/s. The bin width of histogram is determined by Freedman-Diaconis rule. The kernel density estimation of each force distribution is fitted to a Gaussian equation (black line): $f(x)=A*exp(-(x-x_0)^2/2\sigma^2)$.

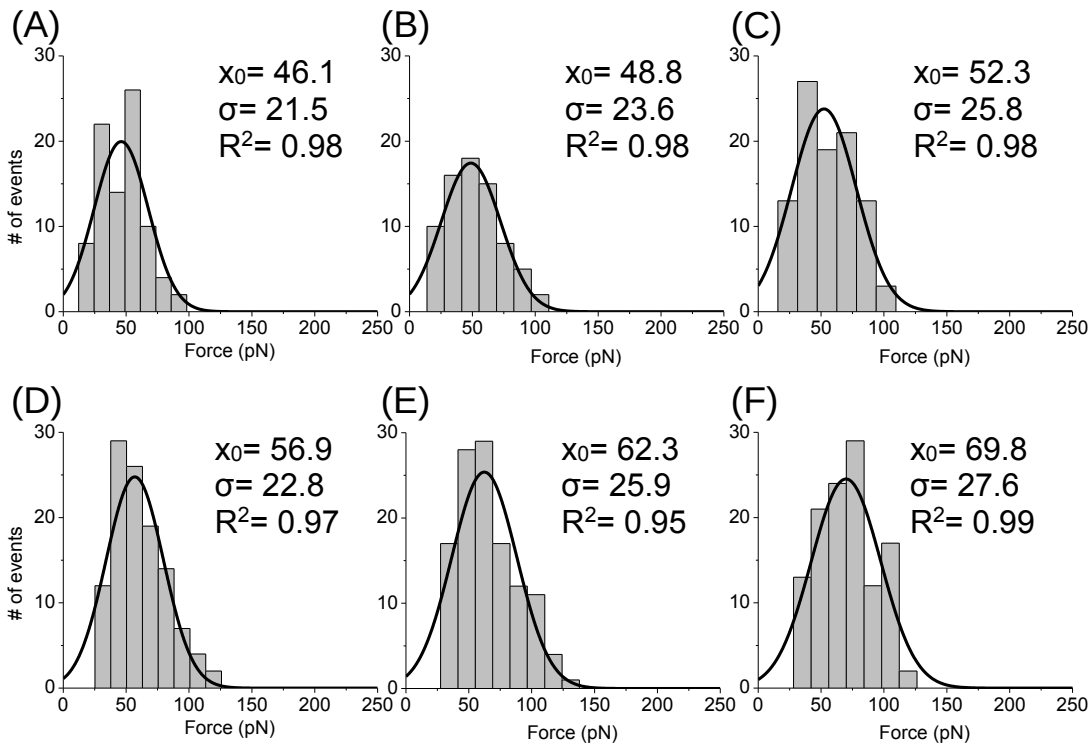

Figure S6. Histogram of unbinding force for PrP(23-231) measured in 1 mM Cu^{2+}. The histogram of rupture forces measured at pulling speed **(A)** 27 nm/s **(B)** 74 nm/s **(C)** 200 nm/s **(D)** 556 nm/s **(E)** 2000 nm/s and **(F)** 10000 nm/s. The bin width of histogram is determined by Freedman-Diaconis rule. The kernel density estimation of each force distribution is fitted to a Gaussian equation (black line): $f(x)=A*exp(-(x-x_0)^2/2\sigma^2)$.

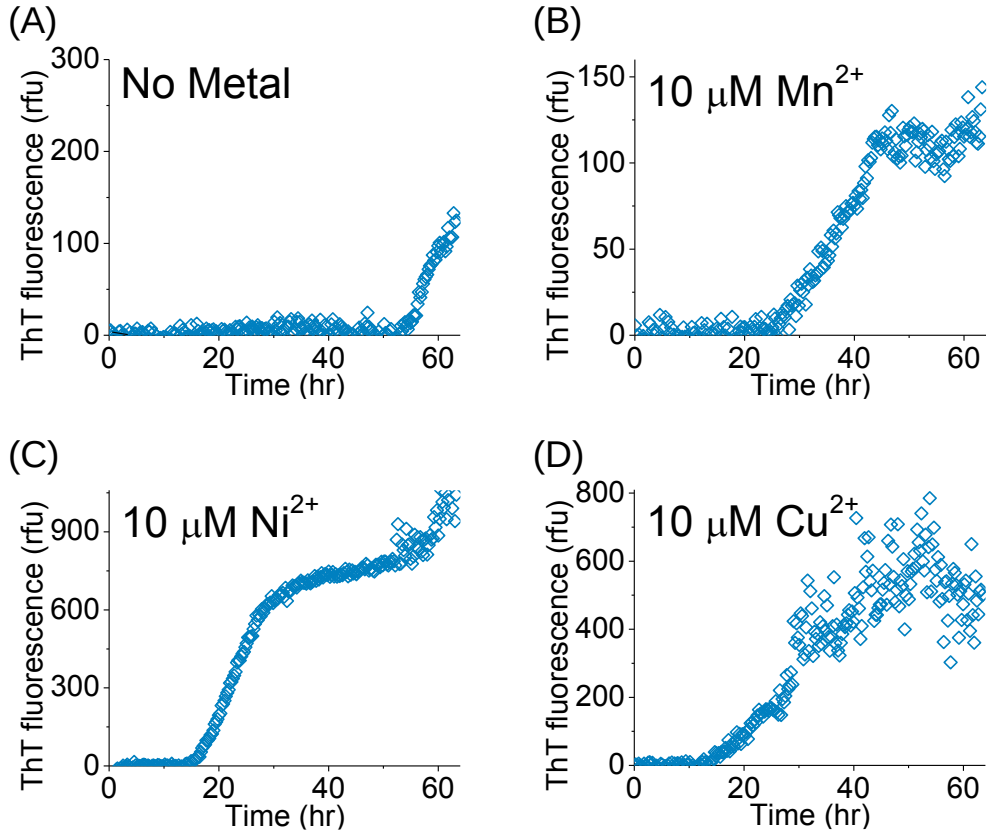

Figure S7. Formation of PrP(23-230) seeds monitored in real time using Thioflavin T (ThT) fluorescence intensity. Seeds were generated by incubating 6 μM human recombinant PrP(23-230) with 10 μM ThT **(A)** in the absence of divalent ions or in the presence of either **(B)** 10 μM Mn^{2+} **(C)** 10 μM Ni^{2+} or **(D)** 10 μM Cu^{2+} in a pH 7.4 buffer (20 mM Tris-HCl, 150 mM NaCl). Formation of PrP seeds was monitored in real time from an increase in ThT fluorescence intensity. In the presence of 10 μM divalent ions, PrP(23-230) seeds started forming within 15 to 25 hours. In contrast, without divalent ions, ThT fluorescence start increasing only after 55 hours. Traces represent baseline corrected, weighted averages across three replicates.

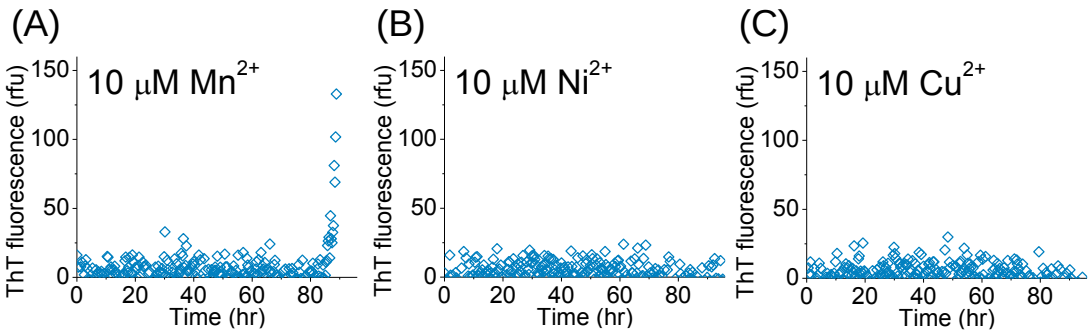

Figure S8. No seeds were formed when PrP(90-230) was incubated with divalent metal ions for 84 hours. Formation of PrP(90-230) seeds monitored in real time using Thioflavin T (ThT) fluorescence intensity. PrP(90-230) seeds were created by incubating 6 μM human recombinant PrP(90-230) with 10 μM ThT in the presence of either **(A)** 10 μM Mn^{2+} **(B)** 10 μM Ni^{2+} or **(C)** 10 μM Cu^{2+} in a pH 7.4 buffer (20 mM Tris-HCl, 150 mM NaCl). Traces represent baseline corrected, weighted averages across three replicates. ThT fluorescence start increasing after 84 hours only in the presence of 10 μM Mn^{2+}.

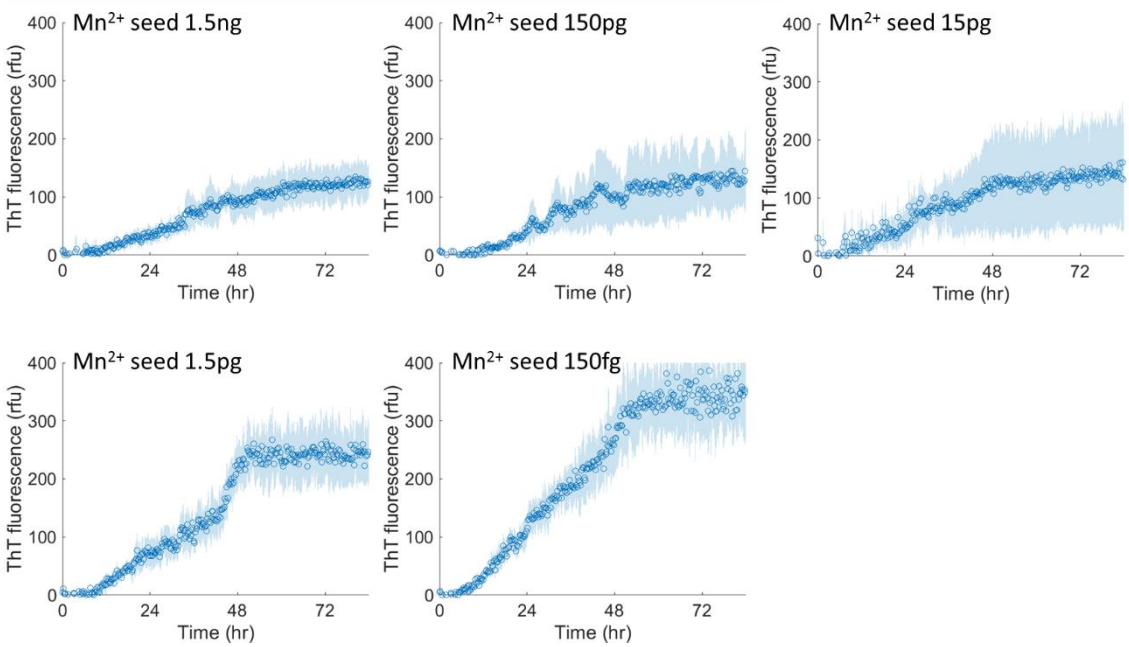

Figure S9. Seeding activity of PrP(23-230) seeds generated in 10 μM Mn^{2+} measured using real-time quaking induced conversion (RT-QuIC). RT-QuIC traces for 15 μg human recombinant PrP(23-230) substrate seeded with PrP(23-230) that had been pre-exposed to 10 μM Mn^{2+}. Fluorescence signals from PrP aggregates were averaged across four replicates and baseline corrected. The shaded area corresponds to the average ± one standard deviation. Duration of the lag phase (T_{th}) was determined to be the time point where ThT fluorescence intensity first increased beyond a predetermined threshold (5 times the standard deviation of a blank samples without PrP seeds and substrate). The determined T_{th} are 19.5 ± 12.7 hr (1.5 ng), 20.5 ± 0.6 hr (150 pg), 14.7 ± 1.7 hr (15 pg), 13.7 ± 0.5 hr (1.5 pg) and 11.2 ± 0.5 hr (150 fg). Error corresponds to standard error of mean was calculated using a bootstrap with replacement protocol.

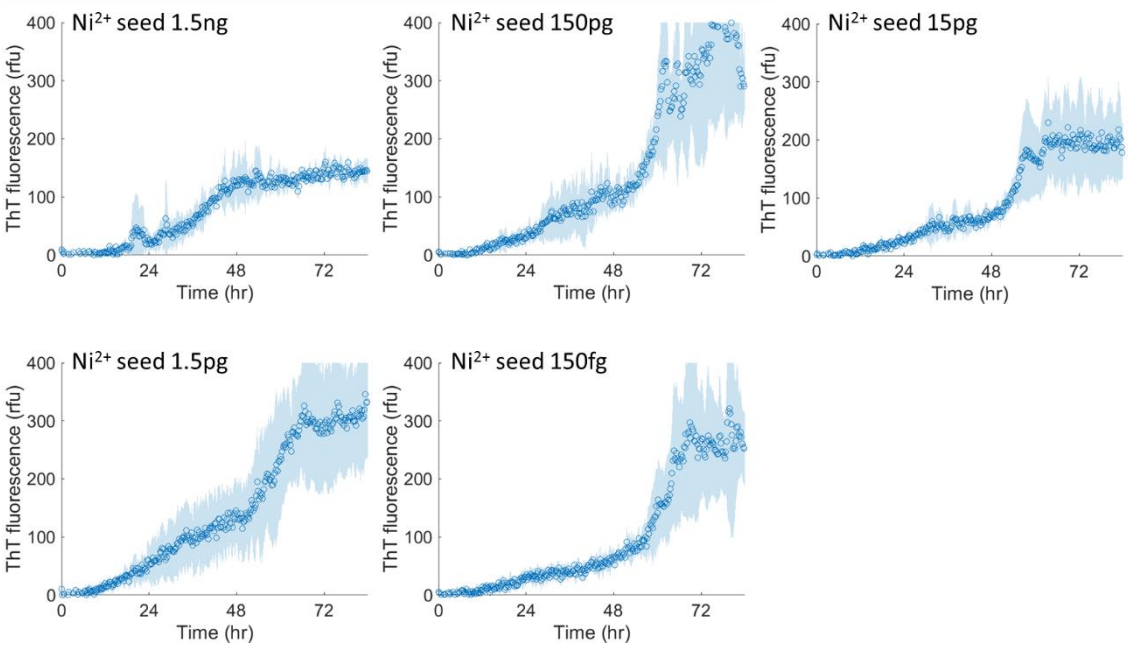

Figure S10. Seeding activity of PrP(23-230) seeds generated in 10 μM Ni^{2+} measured using real-time quaking induced conversion (RT-QuIC). RT-QuIC traces for 15 μg human recombinant PrP(23-230) substrate seeded with PrP(23-230) that had been pre-exposed to 10 μM Ni^{2+}. Fluorescence signals from PrP aggregates were averaged across four replicates and baseline corrected. The shaded area corresponds to the average ± one standard deviation. Duration of the lag phase (T_{th}) was determined to be the time point where ThT fluorescence intensity first increased beyond a predetermined threshold (5 times the standard deviation of a blank samples without PrP seeds and substrate). The determined T_{th} are 20.5 ± 1.6 hr (1.5 ng), 21.3 ± 1.3 hr (150 pg), 24.1 ± 0.9 hr (15 pg), 16.5 ± 0.7 hr (1.5 pg) and 23.3 ± 6.1 hr (150 fg). Error corresponds to standard error of mean was calculated using a bootstrap with replacement protocol.

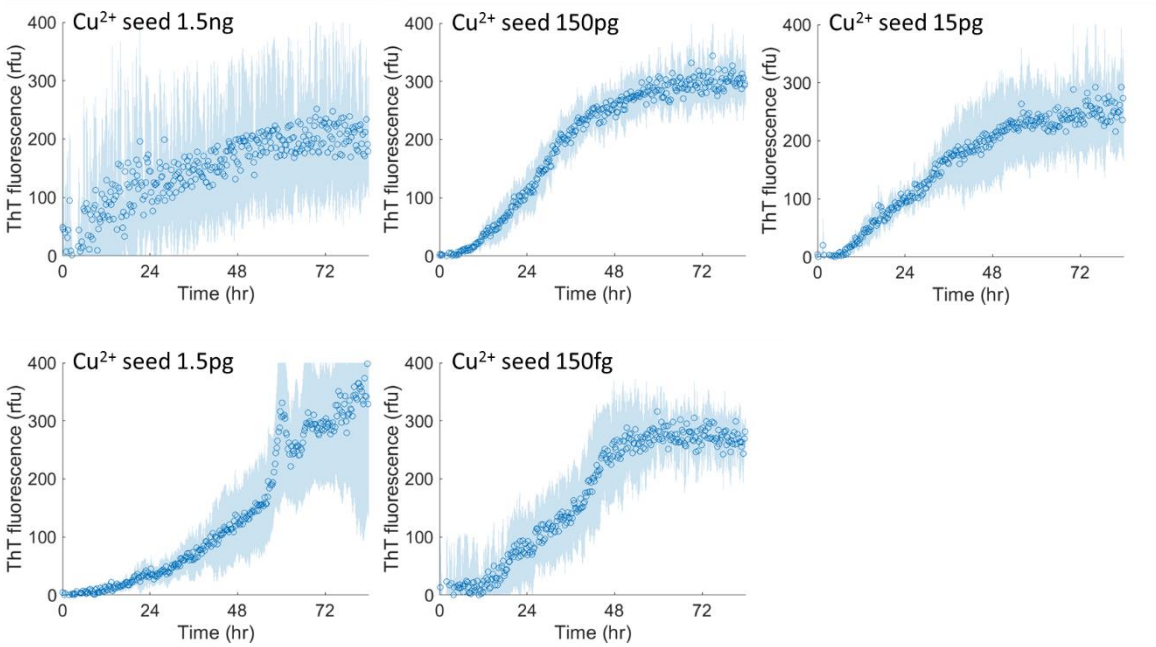

Figure S11. Seeding activity of PrP(23-230) seeds generated in 10 μM Cu^{2+} measured using real-time quaking induced conversion (RT-QuIC). RT-QuIC traces for 15 μg human recombinant PrP(23-230) substrate seeded with PrP(23-230) that had been pre-exposed to 10 μM Cu^{2+}. Fluorescence signals from PrP aggregates were averaged across five replicates and baseline corrected. The shaded area corresponds to the average ± one standard deviation. Duration of the lag phase (T_{th}) was determined to be the time point where ThT fluorescence intensity first increased beyond a predetermined threshold (5 times the standard deviation of a blank samples without PrP seeds and substrate). The determined T_{th} are 5.8 ± 14.1 hr (1.5 ng), 11.2 ± 1.1 hr (150 pg), 10.9 ± 0.7 hr (15 pg), 20.0 ± 2.1 hr (1.5 pg) and 14.7 ± 14.3 hr (150 fg). Error corresponds to standard error of mean was calculated using a bootstrap with replacement protocol.

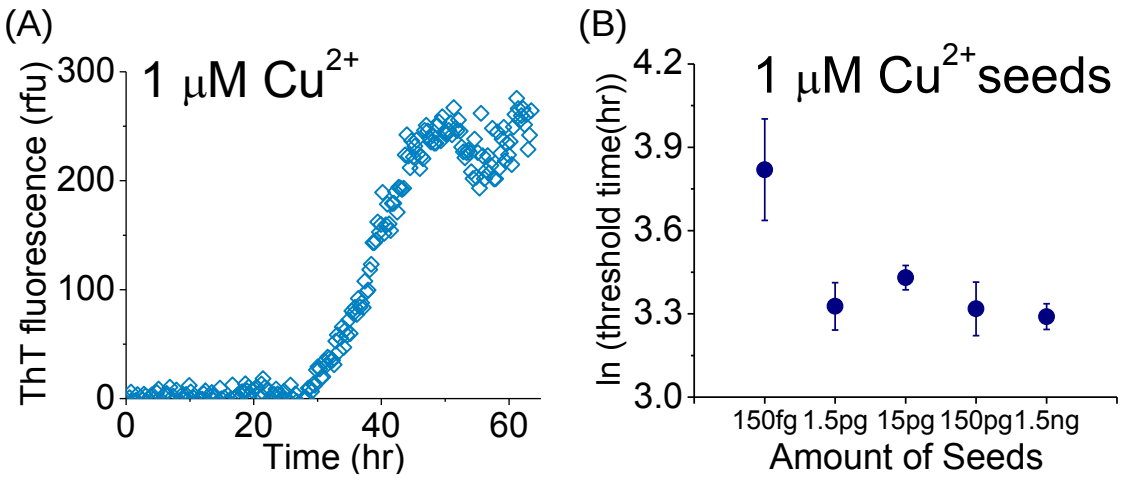

Figure S12. Formation of PrP(23-230) seeds using 1 μM Cu^{2+} and its corresponding seeding activity. (A) PrP(23-230) seeds were generated by incubating 6 μM human recombinant PrP(23-230) with 10 μM ThT in the presence of 1 μM Cu^{2+} in a pH 7.4 buffer (20 mM Tris-HCl, 150 mM NaCl). Formation of PrP seeds was monitored in real time from an increase in ThT fluorescence intensity. Traces represent baseline corrected, weighted averages across three replicates. **(B)** Plot of lag phase duration for different amounts of PrP(23-230) seeds formed in 1 μM Cu^{2+}. The decrease in lag phase duration with increasing seed concentration is suggested by theory.

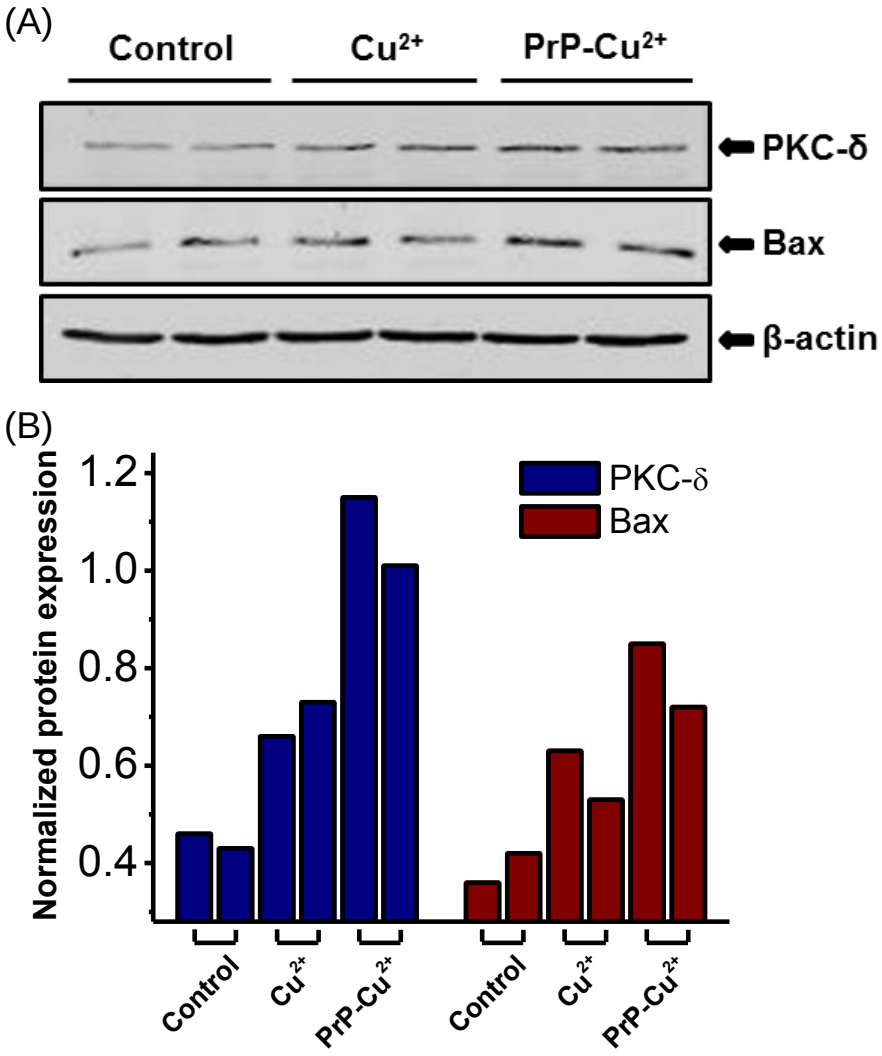

Figure S13. Residual copper upon addition of protein seed to brain slice culture, does not increase levels of PKC-δ and Bax. (A) Representative western blot of PKC-δ and Bax. Organotypic slice cultures were untreated (control), exposed to either 100 nM Cu^{2+} or copper-induced PrP(23-230) seeds (PrP-Cu^{2+}). **(B)** Quantification of Western blot band intensities show higher PKC-δ and Bax protein expression upon exposure to copper-induced prion oligomers. Each condition was repeated two times.

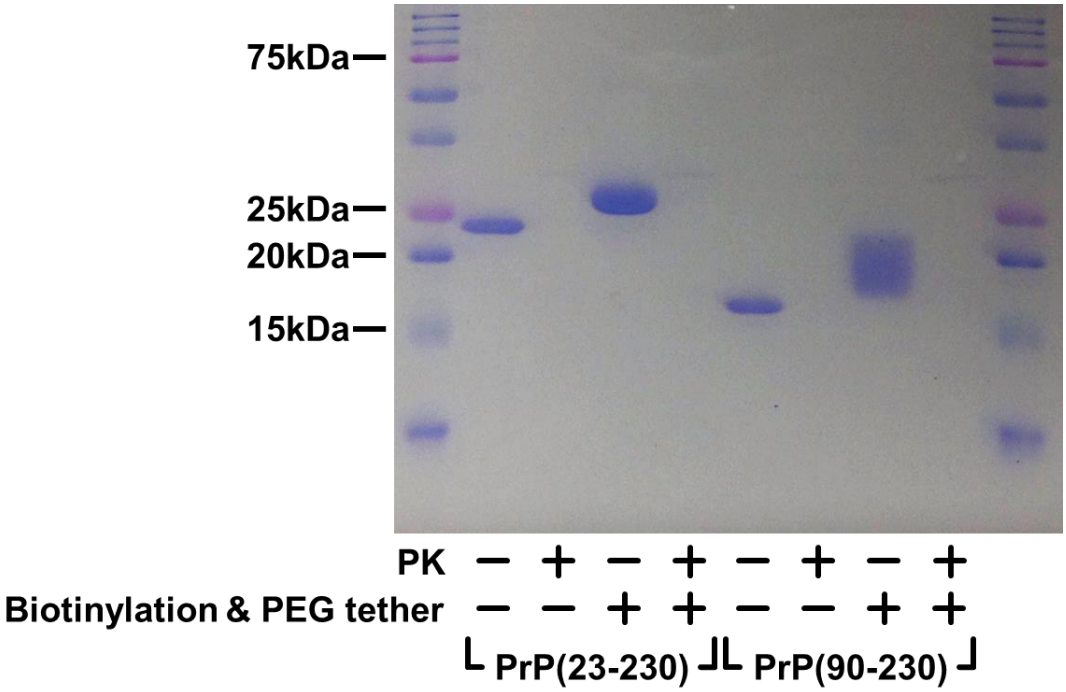

Figure S14. Biotinylation of PrP and functionalization with PEG tethers does not alter sensitivity to proteinase K (PK) digestion. The PK digests were monitored using SDS-PAGE. Under digestion conditions that are identical to those used in our single molecule PK resistance assay, both PrP(23-230) and PrP(90-230) were completely digested by PK before and after being biotinylated and tagged with PEG tethers. Biotinylation was performed by mixing either 10 μM of PrP(23-230) or 15 μM of PrP(90-230) with 500 μM Sulfo-NHS-LC-Biotin in reaction buffer (50 mM sodium phosphate, 50 mM NaCl, 10 mM EDTA, pH7.2) at 4 °C for 3 hours. The reaction was quenched by adding 5mM L-Lysine at 20 °C for 1 hour. Biotinylated PrP was reduced using 300 μM TCEP at 4 °C for 3 hours and then PEGylated using 100 μM maleimide-PEG at 20 °C for 1 hour. Due to the 3.4 kDa molecular weight of the PEG tether, the pegylated PrP has a larger molecular weight. PK digestion was performed using 100 μg/mL PK at 37 °C for 3 hours.

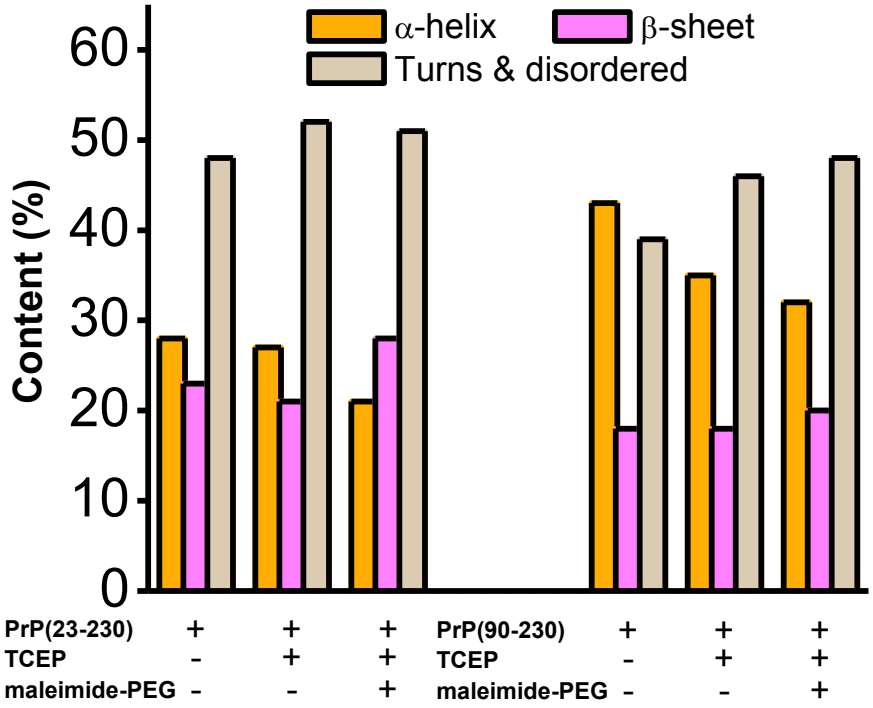

Figure S15. PrP(23-230) and PrP(90-230) remain in a native conformation after reduction of disulfide bond and functionalization with PEG tether. Secondary structure of PrP(23-230) and PrP(90-230) before and after TCEP reduction and binding to PEG measured using Circular Dichroism. 5 µM PrP(23-230) and 5 µM PrP(90-230) in 10 mM sodium phosphate buffer, pH 7.2 were reduced in 150 µM TCEP at 4 °C for 3 hours and then incubated with 50 µM maleimide-PEG at 20 °C for 1 hour. Spectra was measured at 4 °C for each condition. Spectral deconvolution was performed using online server DICHROWEB. Secondary structure calculated from K2D is plotted. The measured α-helix content indicates that PrP retains its native structure even after disulfide bond reduction and binding to PEG.

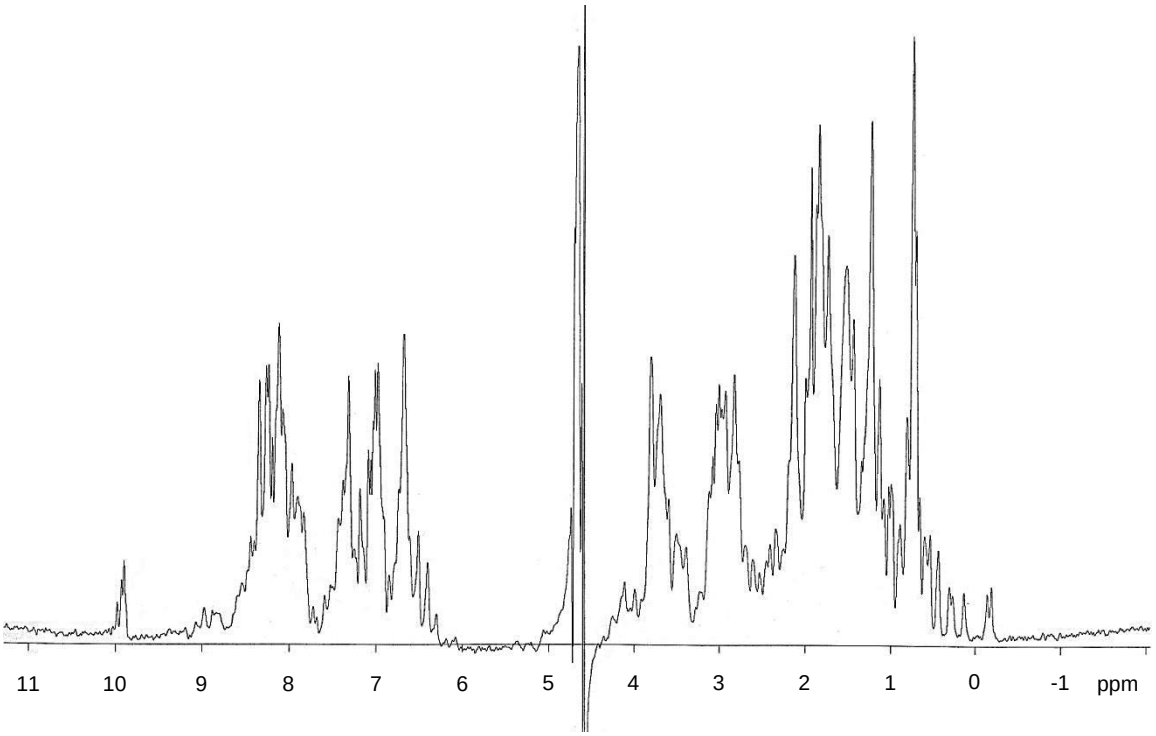

Figure S16. One-dimensional ^{1}H-NMR spectra of PrP(23-231) shows that protein is in a natively folded conformation. 1D ^{1}H-NMR spectrum of PrP(23-231) was acquired in 10 mM sodium acetate buffer at pH 4.5 using 800 MHz Varian spectrometer (Agilent). The protein concentration was 50 μM. Spectrum was measured at 20 °C by Jena Bioscience.

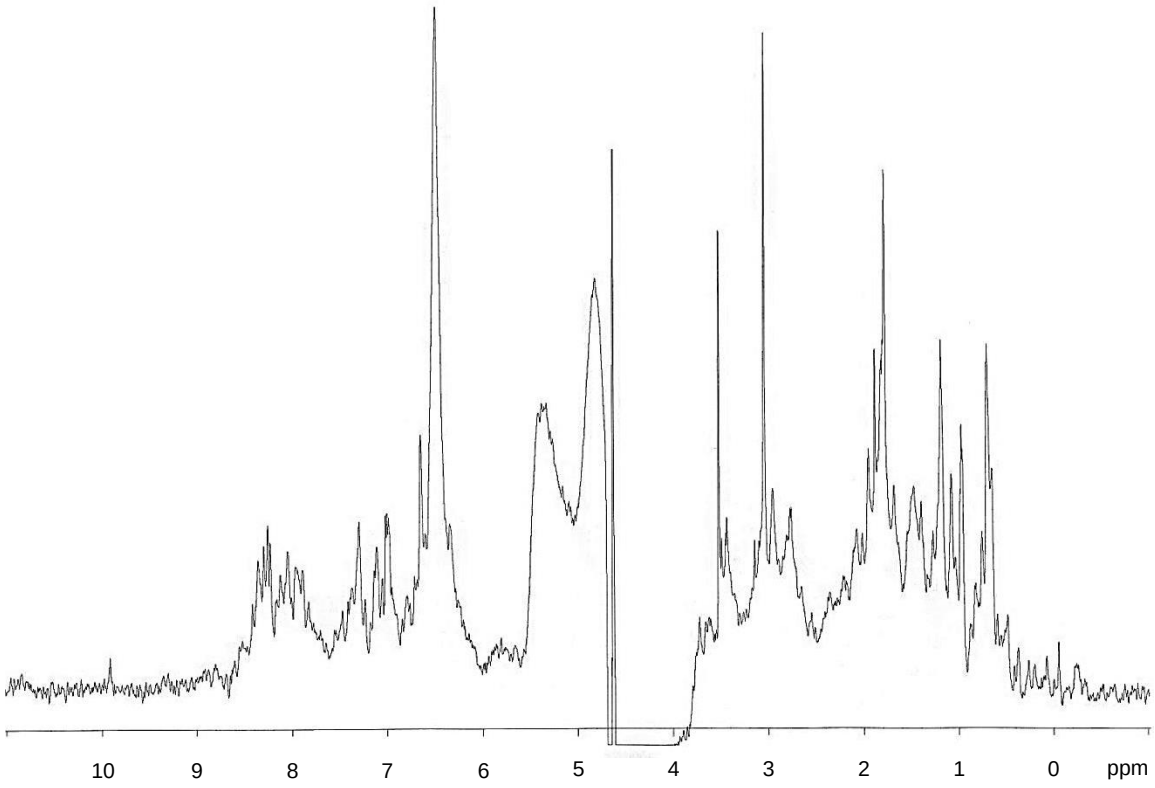

Figure S17. One-dimensional ^{1}H-NMR spectra of PrP(90-231) shows that protein is in a natively folded conformation. 1D ^{1}H-NMR spectrum of PrP(90-231) was acquired in 10 mM sodium acetate buffer at pH 4.5 using 800 MHz Varian spectrometer (Agilent). The protein concentration was 50 μM. Spectrum was measured at 20 °C by Jena Bioscience.

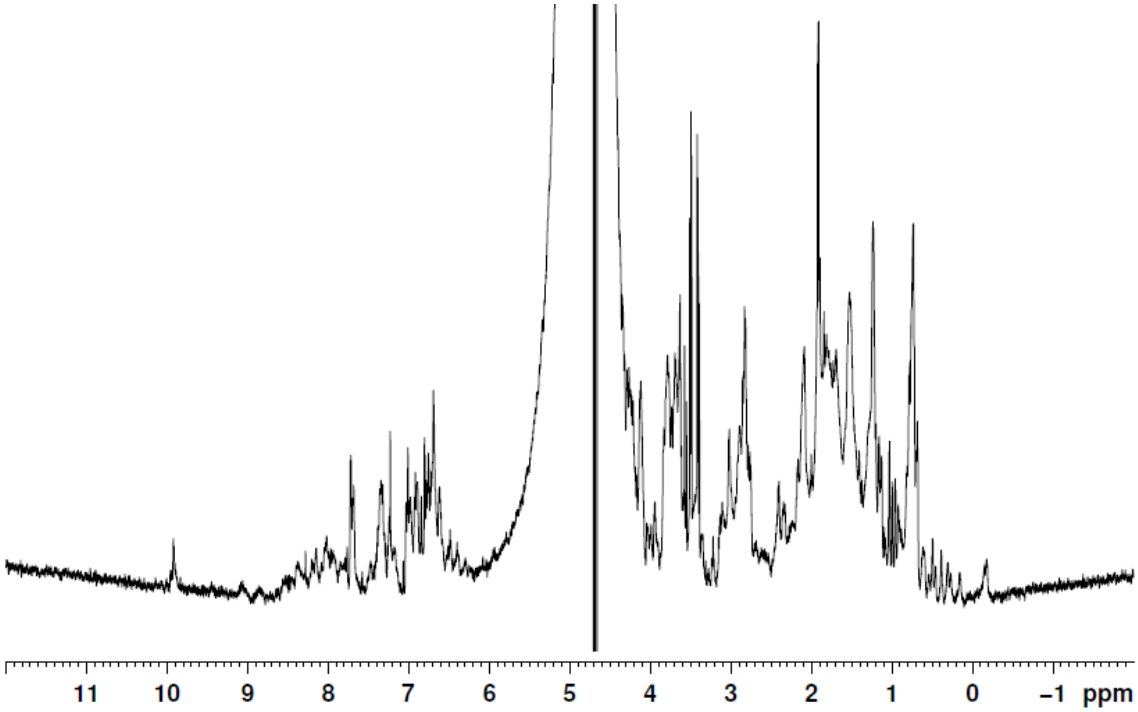

Figure S18. One-dimensional ^{1}H-NMR spectra of PrP(23-230) shows that protein is in a natively folded conformation. 1D ^{1}H-NMR spectrum of PrP(23-230) was acquired using Bruker Avance II 700 spectrometer at a ^{1}H-frequency of 700 MHz. Spectrum was measured at 20 °C. Protein concentration was 80 μM in 10 mM sodium phosphate buffer, pH 5.8.

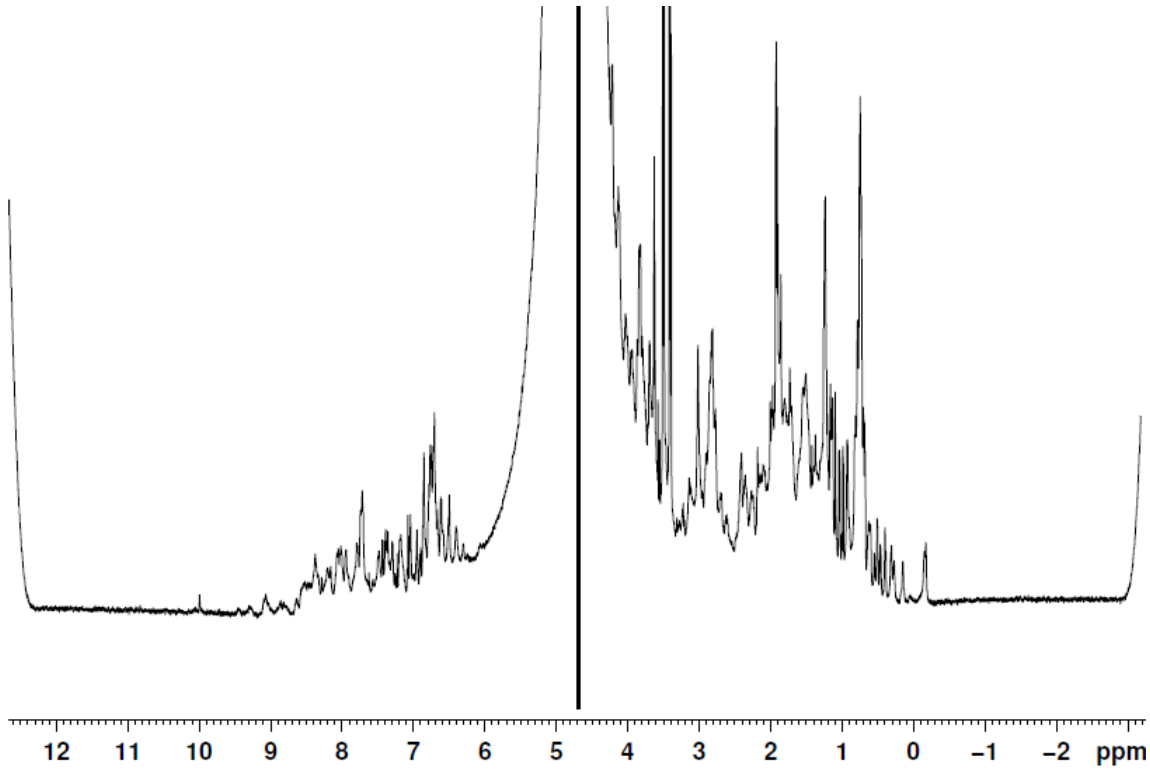

Figure S19. One-dimensional ^{1}H-NMR spectra of PrP(90-230) shows that protein is in a natively folded conformation. 1D ^{1}H-NMR spectrum of PrP(90-230) was acquired using Bruker Avance II 700 spectrometer at a ^{1}H-frequency of 700 MHz. Spectrum was measured at 20 °C. Protein concentration was 90 μM in 10 mM sodium phosphate buffer, pH 5.8.

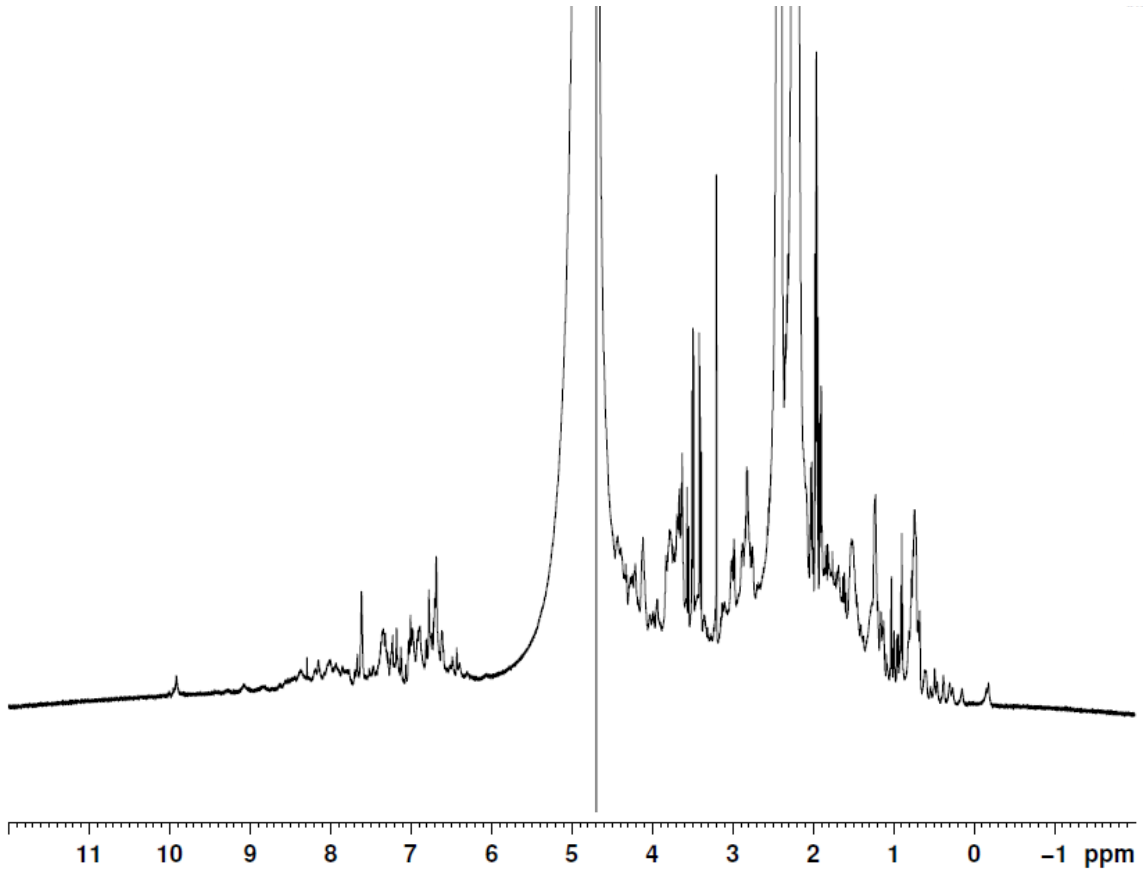

Figure S20. One-dimensional ^{1}H-NMR spectra of PrP(23-230) after disulfide bond reduction shows that protein remains in a natively folded conformation. 1D ^{1}H-NMR spectrum of PrP(23-230) was acquired using Bruker Avance II 700 spectrometer at a ^{1}H-frequency of 700 MHz. Spectrum was measured at 20 °C. Protein concentration was 80 μM in 5 mM TCEP and 10 mM sodium phosphate buffer, pH 5.8.

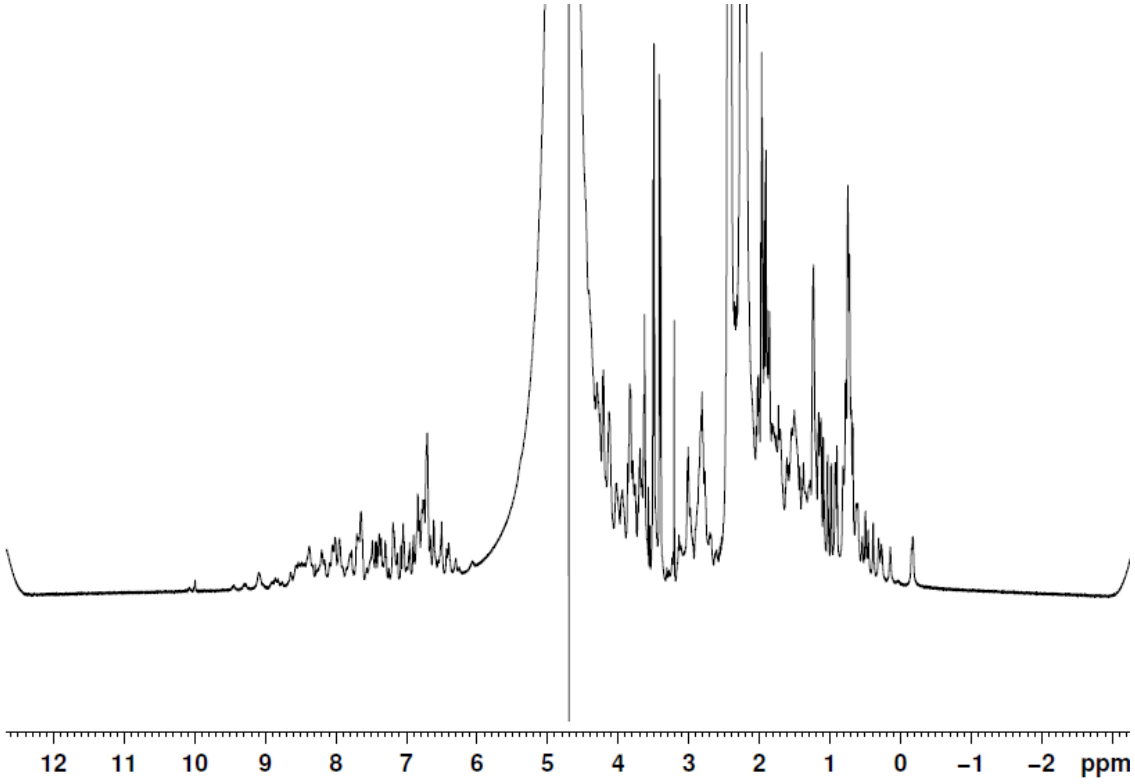

Figure S21. One-dimensional ^{1}H-NMR spectra of PrP(90-230) after disulfide bond reduction shows that protein remains in a natively folded conformation. 1D ^{1}H-NMR spectrum of PrP(90-230) was acquired using Bruker Avance II 700 spectrometer at a ^{1}H-frequency of 700 MHz. Spectrum was measured at 20 °C. Protein concentration was 90 μM in 5 mM TCEP and 10 mM sodium phosphate buffer, pH 5.8.

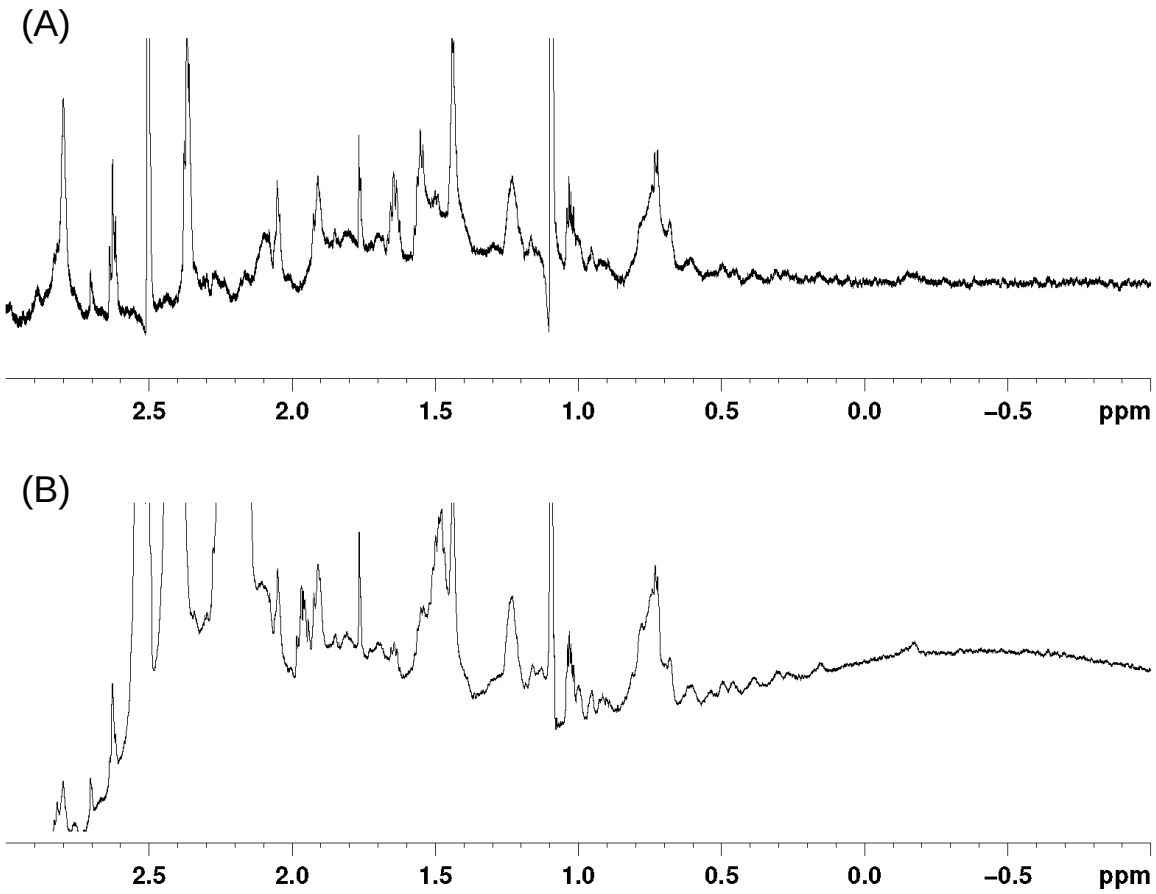

Figure S22. One-dimensional ^{1}H-NMR spectra of PrP(23-230) linked to PEG tethers shows that the protein remains in a natively folded conformation. Spectra was acquired using Bruker Avance II 700 spectrometer at a ^{1}H-frequency of 700 MHz. Spectrum was measured at 20 °C. Protein concentration was 21 μM in 10 mM sodium phosphate buffer at pH 7.2. **(A)** Spectra for PrP(23-230) that was not covalently tagged with PEG. PrP(23-230) with intact disulfide bonds was incubated with 210 μM maleimide-PEG at 20 °C for 1 hour. **(B)** Spectra for PrP(23-230) covalently functionalized with PEG tethers. The PrP(23-230) was first incubated with 630 μM TCEP at 4 °C for 3 hours to reduce disulfide bonds and subsequently incubated with 210 μM maleimide-PEG at 20 °C for 1 hour. Spectra in (A) and (B) have similar chemical shift dispersions typical for natively folded PrP.

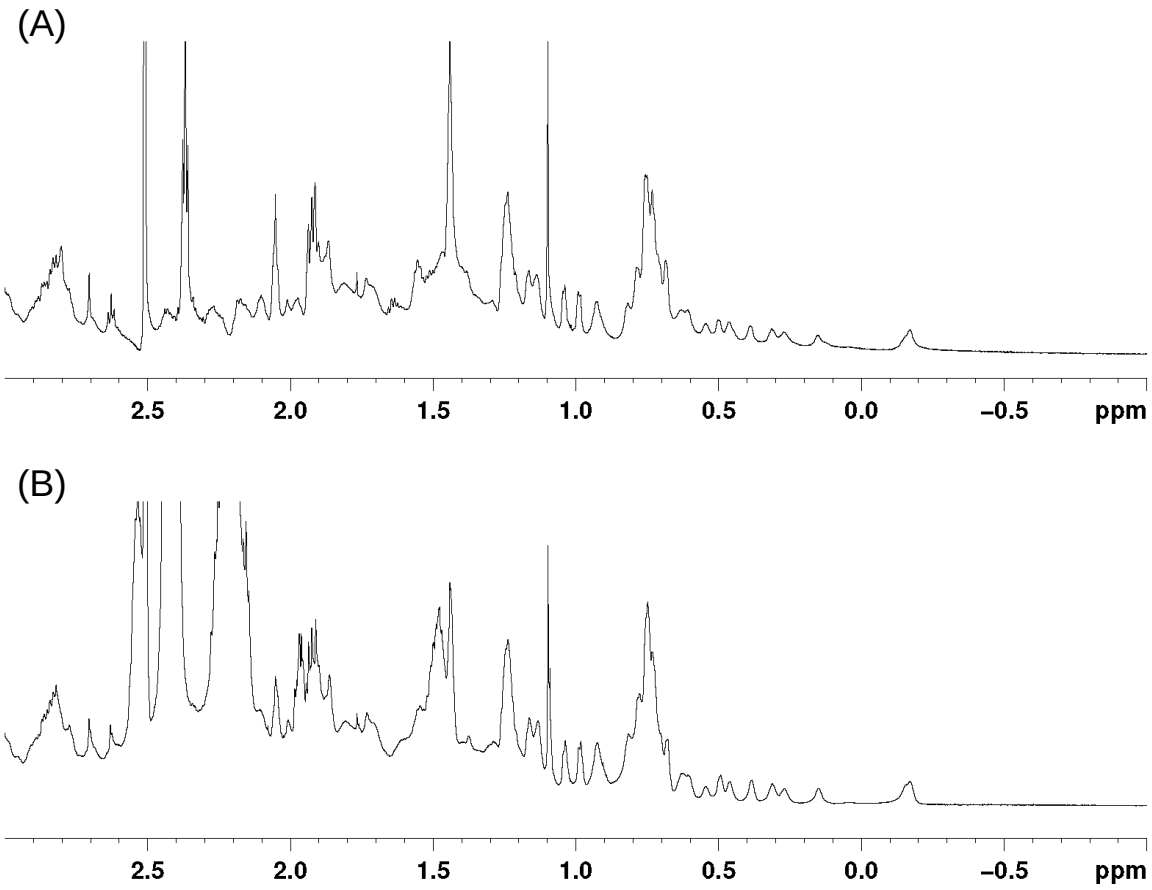

Figure S23. One-dimensional ^{1}H-NMR spectra of PrP(90-230) linked to PEG tethers shows that the protein remains in a natively folded conformation. Spectra was acquired using Bruker Avance II 700 spectrometer at a ^{1}H-frequency of 700 MHz. Spectrum was measured at 20 °C. Protein concentration was 141 μM in 10 mM sodium phosphate buffer at pH 7.2. **(A)** Spectra for PrP(90-230) that was not covalently tagged with PEG. PrP(90-230) with intact disulfide bonds was incubated with 1.4 mM maleimide-PEG at 20 °C for 1 hour. **(B)** Spectra for PrP(90-230) covalently functionalized with PEG tethers. The PrP(90-230) was first incubated with 4.2 mM TCEP at 4 °C for 3 hours to reduce disulfide bonds and subsequently incubated with 1.4 mM maleimide-PEG at 20 °C for 1 hour. Spectra in (A) and (B) have similar chemical shift dispersions typical for natively folded PrP.

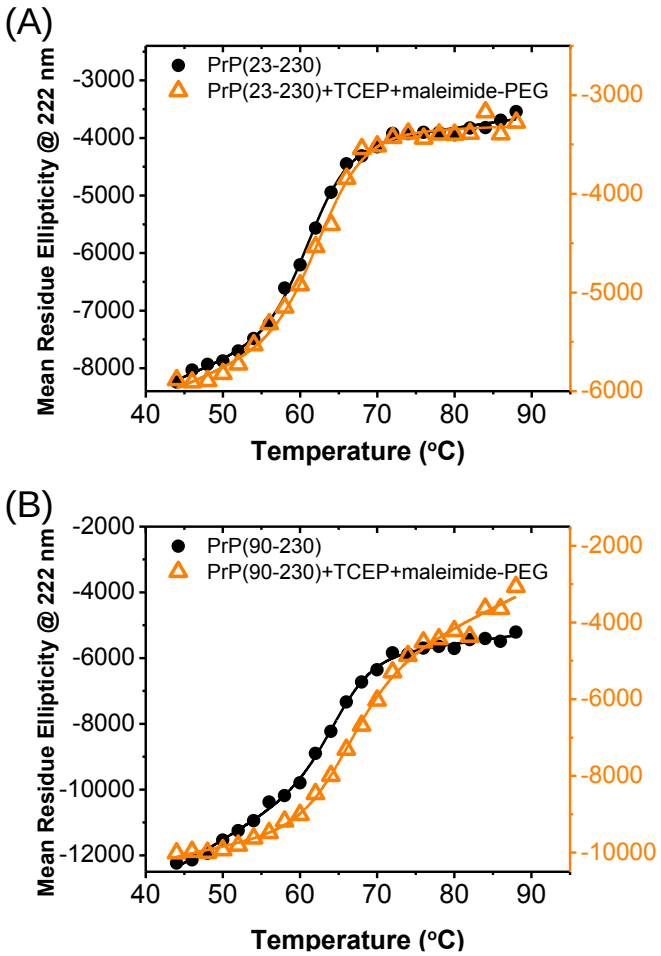

Figure S24. Functionalization with PEG tethers do not alter the thermal stabilities of PrP(23-230) and PrP(90-230). Thermal unfolding transition curves of **(A)** PrP(23-230) and **(B)** PrP(90-230) before and after functionalization with PEG tethers were measured using Circular Dichroism (CD) at 222 nm. 5 µM PrP in 10 mM sodium phosphate buffer, pH 7.2 was reduced using 150 µM TCEP at 4 °C for 3 hours and subsequently incubated with 50 µM maleimide-PEG at 20 °C for 1 hour. CD signals were recorded every 2 °C from 44 to 88 °C at a rate of 0.4 °C per minute. The melting temperature (T_m) was calculated by fitting the CD data to a two-state transition model which assumes that PrP monomers convert between folded and unfolded forms. The fitted curves are plotted as solid lines. T_m of PrP(23-230) before and after functionalization with PEG tether are 61.4 ± 0.4 °C and 62.9 ± 0.7 °C. T_m of PrP(90-230) before and after functionalization with PEG tether are 65.1 ± 0.4 °C and 65.9 ± 0.8 °C. These results indicate that the thermal stability of PrP does not change after functionalization with PEG tethers. Error corresponds to standard error of fitting.

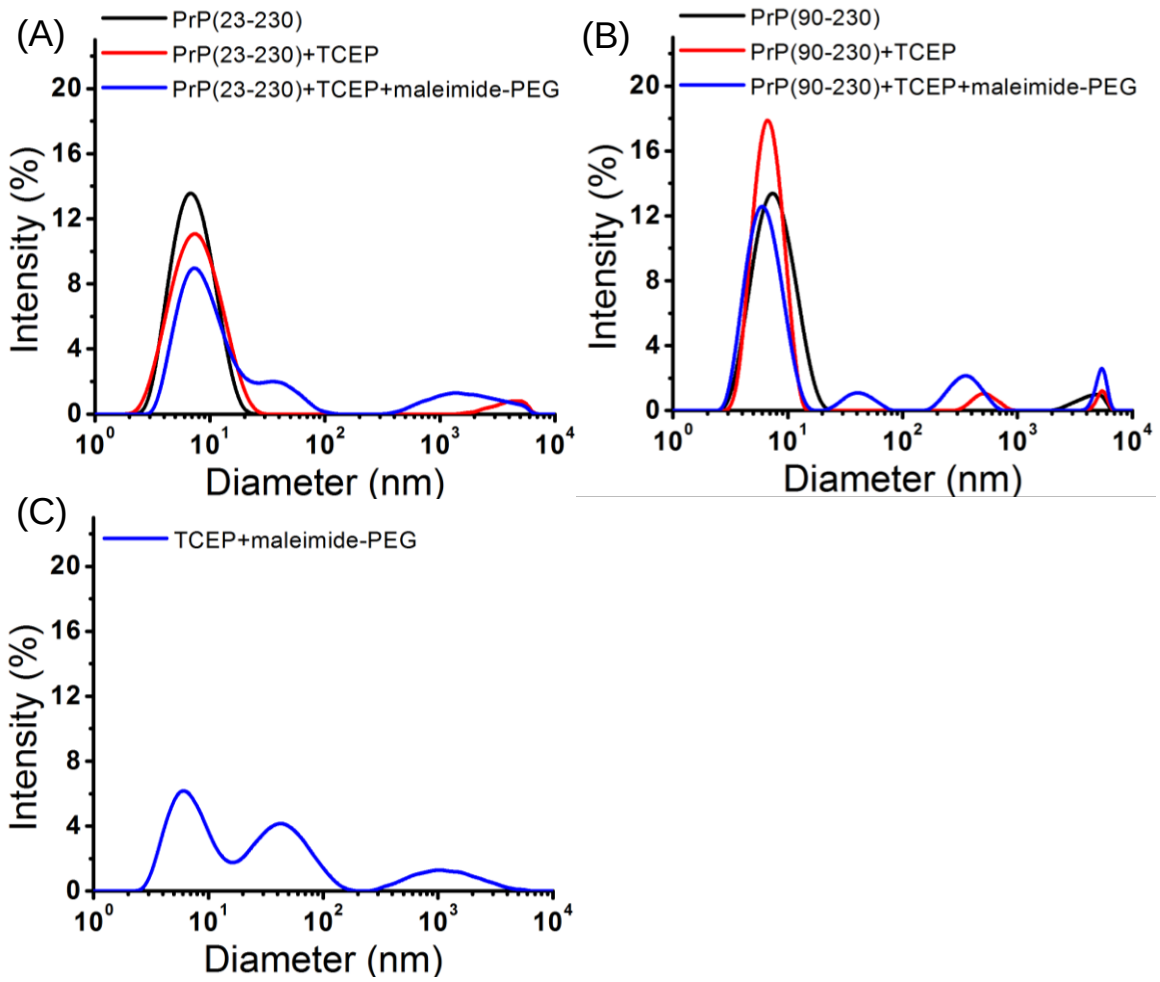

Figure S25. Reduction of disulfide bond and functionalization with PEG tether does not cause aggregation of PrP(23-230) and PrP(90-230). Size distribution of PrP(23-230) and PrP(90-230) before and after reduction of disulfide bond and after functionalization with PEG was measured using dynamic light scattering. **(A)** Size distribution of PrP(23-230) and **(B)** PrP(90-230) in 10 mM sodium phosphate buffer at pH 7.2 (black); after incubating with 1 mM TCEP at 4 °C for 3 hours (red), followed by incubation with 300 µM maleimide-PEG at 20 °C for another 3 hours (blue). Numbers inside parentheses indicate the percentage of PrP monomer. As shown in **(C)**, the appearance of higher order peaks upon addition of maleimide-PEG are due to formation of maleimide-PEG aggregates and not due to PrP aggregation. (C) Size distribution of maleimide-PEG in the presence of 1 mM TCEP without any PrP. The additional peaks measured with reduced maleimide-PEG are responsible for the three peaks measured in (A) and (B).

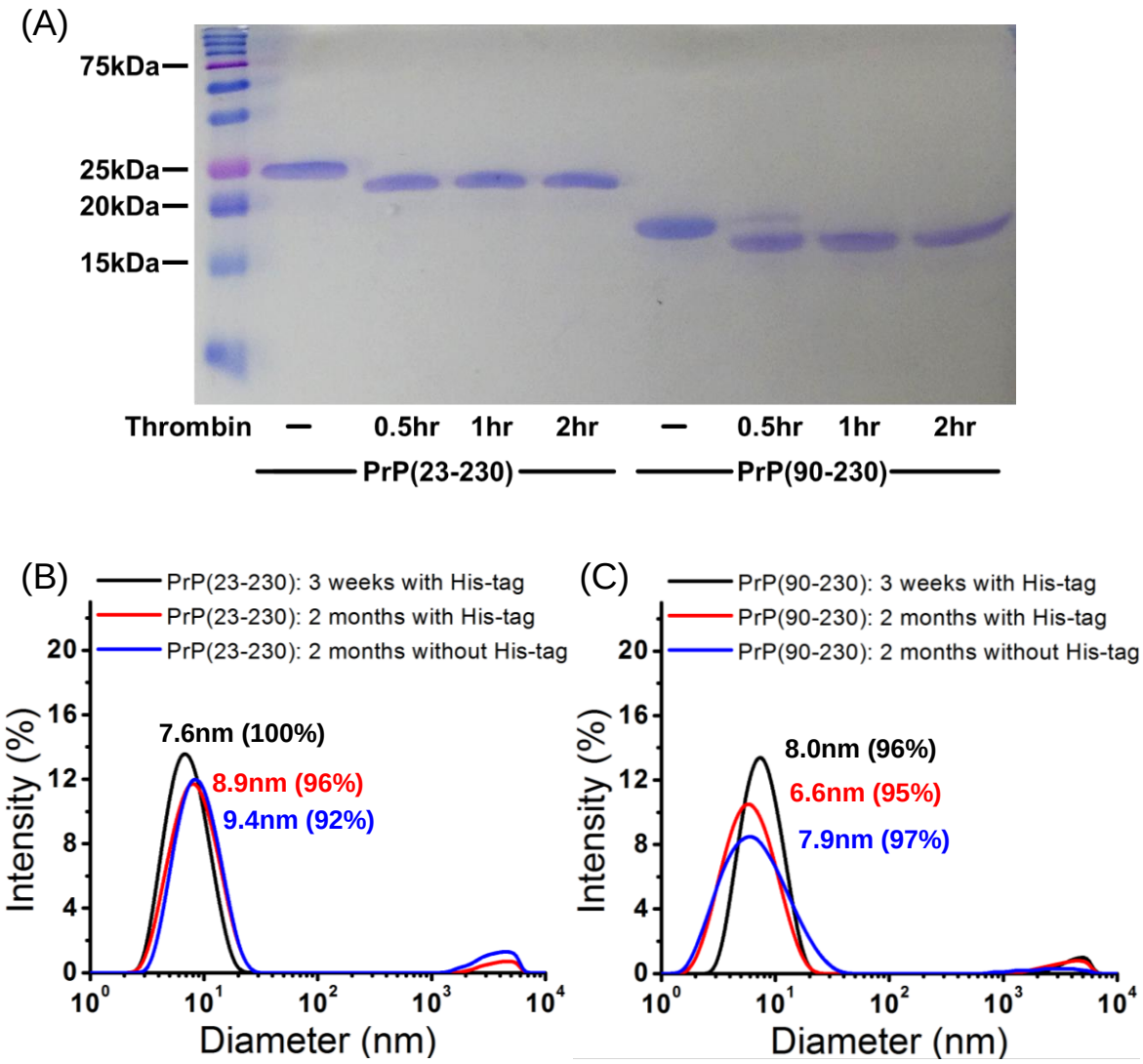

Figure S26. Protein age and presence of N-terminal His-tag does not cause aggregation of PrP(23-230) and PrP(90-230). (A) Cleavage of His-tag was confirmed by SDS-PAGE. His-tag cleavage results in a 1.4 kDa reduction in the molecular weight. After incubation with 0.1 unit/mL thrombin enzyme at 20 °C for 0.5-2 hours, PrP was heated to 95 °C in 0.1 mM DTT for 5 mins to terminate the reaction before loading the gel. Result shows that His-tag is totally cleaved from PrP after incubation with thrombin for 1 hour. Size distribution of **(B)** PrP(23-230) and **(C)** PrP(90-230) in 10 mM sodium phosphate buffer at pH 7.2 were measured at 4 °C using dynamic light scattering. After purification, His-tagged PrP was stored at 4 °C for 3 weeks (black) and 2 months (red). The His-tag on the 2 month PrP was cleaved by incubation with 0.1 unit/mL thrombin enzyme at 20 °C for 1 hour (blue). Numbers in parentheses indicate the percentage of PrP monomer.

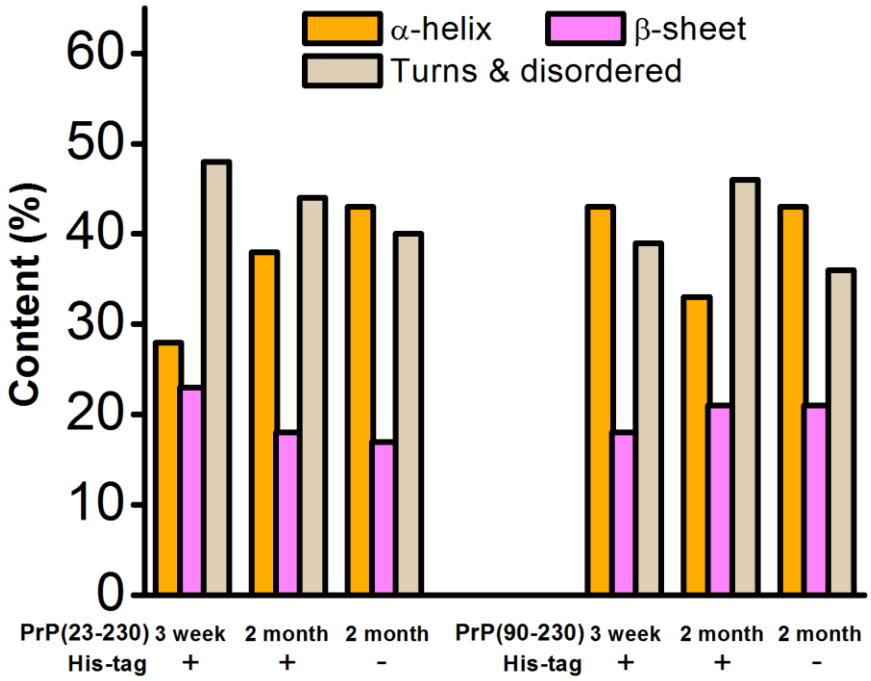

Figure S27. Protein age and presence of N-terminal His-tag does not alter the secondary structure of PrP(23-230) and PrP(90-230). Secondary structure of PrP(23-230) and PrP(90-230) in 10 mM sodium phosphate buffer, pH 7.2 was measured using Circular Dichroism. The His-tagged PrP(23-230) and PrP(90-230) were stored at 4 °C for 3 weeks and 2 months and then incubated with 0.1 unit/mL thrombin enzyme at 20 °C for 1 hour to cleave the His-tag. Spectra was measured at 4 °C for each condition. Spectral deconvolution was performed using online server DICHROWEB. Secondary structure calculated from K2D is plotted. The measured α-helix content indicates that PrP retains its native structure under all measurement conditions.

Table S1. Surface density of PrP in PK digestion experiments

Errors correspond to s.e. propagated from the fitting error in the calibration curve and the s.e. of fluorescence intensity estimation.

PrP surface density (# of molecule/μm^2)	PrP(23-231)	PrP(90-231)
No metal	3.11 ± 0.26	1.65 ± 0.25
Mn^{2+}	0.53 ± 0.15	1.25 ± 0.26
Ni^{2+}	1.86 ± 0.21	1.19 ± 0.23
Cu^{2+}	1.00 ± 0.18	0.85 ± 0.18

Table S2. Off-rate, relative on-rate and relative association constant (K_A) for PrP(23-231) binding

On-rate and K_A were normalized relative to the homotypic PrP(23-231) interactions in the absence of divalent ions. Errors in on-rate correspond to s.e., which were propagated from the s.e. of binding probability. 90% confidence intervals (CI) of off-rate were calculated using a bootstrap with replacement protocol. 90% CI of K_A were calculated using a bootstrap with replacement protocol and the s.e. of on-rate were propagated.

	Relative on-rate	Off-rate (90% CI)	Relative K_A (90% CI)
No metal	1.0 ± 0.1	$3.1x10^{-2}$ ($4.6x10^{-3}$ – $1.3 x 10^{-1}$) s^{-1}	1.0 (0.1 – 10.6)
Mn^{2+}	2.6 ± 0.3	$1.3x10^{-2}$ ($2.9x10^{-3}$ – $4.6 x 10^{-2}$) s^{-1}	6.0 (0.6 – 46.6)
Ni^{2+}	5.8 ± 0.7	$1.0x10^{-2}$ ($2.9x10^{-3}$ – $8.3 x 10^{-2}$) s^{-1}	17.7 (1.0 – 115.1)
Cu^{2+}	6.3 ± 0.9	$2.2x10^{-4}$ ($2.9x10^{-6}$ – $3.0 x 10^{-3}$) s^{-1}	862.6 (33.1 – 78926.0)

CHAPTER 3. IMPROVING ESTIMATION OF KINETIC PARAMETERS IN DYNAMIC FORCE SPECTROSCOPY USING CLUSTER ANALYSIS

A manuscript submitted to *Biophysical Journal*

Chi-Fu Yen and Sanjeevi Sivasankar

3.1 Abstract

Dynamic Force Spectroscopy (DFS) is a widely used technique to characterize the dissociation kinetics and interaction energy landscape of receptor-ligand complexes with single-molecule resolution. In an Atomic Force Microscope (AFM) based DFS experiment, receptor-ligand complexes, sandwiched between an AFM tip and substrate, are ruptured at different stress rates by varying the speed at which the AFM-tip and substrate are pulled away from each other. The rupture events are grouped according to their pulling speeds and the mean force and loading rate of each group is calculated. This data is subsequently fit to established models and energy landscape parameters such as the intrinsic off-rate (k_{off}) and the width of the potential energy barrier (x_β) are extracted. However, due to large uncertainties in determining mean forces and loading rates of the groups, errors in the estimated k_{off} and x_β can be substantial. Here, we demonstrate that the accuracy of fitted parameters in a DFS experiment, can be dramatically improved by sorting rupture events into groups using cluster analysis instead of sorting them according to their pulling speeds. We test different clustering algorithms including Gaussian mixture, logistic regression and K-means clustering, under conditions that closely mimic DFS experiments. Using Monte Carlo simulations, we benchmark the performance of these clustering algorithms over a wide range of k_{off} and x_β, under different levels of thermal noise, and as a function of both the number of unbinding events and the number of pulling

speeds. Our results demonstrate that cluster analysis, particularly K-means clustering, is very effective at improving the accuracy of parameter estimation, particularly when the number of unbinding events are limited and not well separated into distinct groups. Cluster analysis is easy to implement and our performance benchmarks serve as a guide in choosing an appropriate method for DFS data analysis.

3.2 Introduction

Dynamic force spectroscopy (DFS) experiments are widely used to characterize the dissociation kinetics and interaction energy landscape of protein-protein interactions [1-3], DNA-protein binding [4, 5], and the aggregation of misfolded proteins [6, 7]. While these measurement can be performed using different micromanipulation tools including atomic force microscopy (AFM), micro-needle manipulation, optical tweezers and magnetic tweezers [8], AFM based DFS experiments are widely used because of their sub-nanometer spatial resolution [8].

In a typical DFS experiment, an AFM cantilever and substrate functionalized with flexible polymer linkers, are decorated with the biomolecules of interest (Fig. 1) [9]. The functionalized AFM tip and substrate are brought into contact, enabling opposing molecules to interact, and then pulled apart at a range of pulling speeds. The force applied to the protein complex is sensed by the deflection of the cantilever while the rate of applied force (the loading rate) is controlled by varying the separation-speed of the AFM tip and substrate. Histograms of rupture forces for each pulling speed are plotted to determine the most probable unbinding force; from the dependence of the rupture forces on loading rates, the energy landscape parameters of the system can be predicted [10, 11]. In the

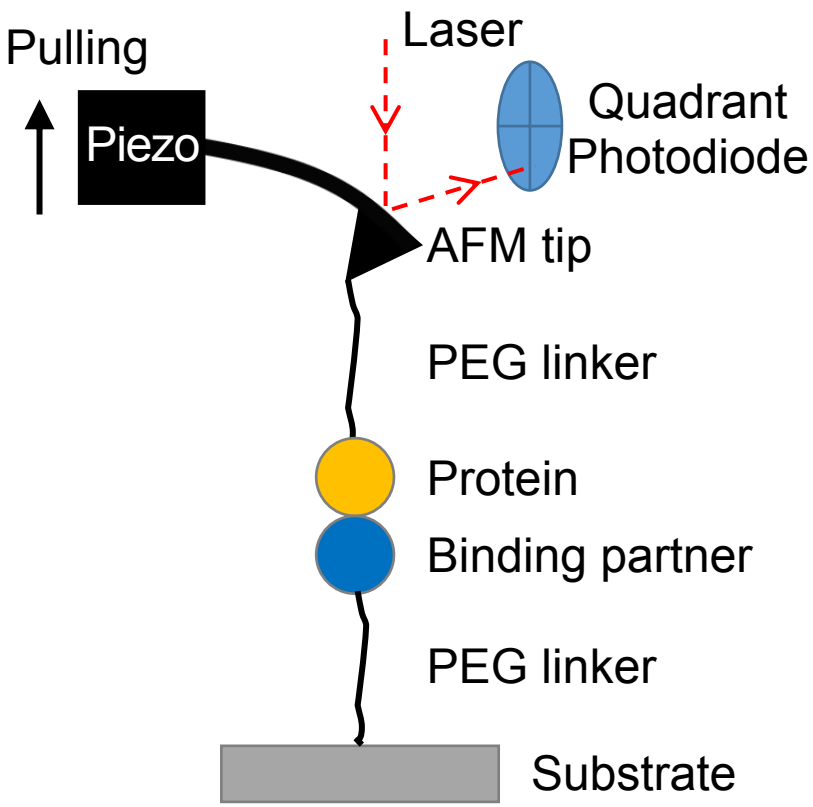

Figure 1. Schematic of the experimental setup for an AFM based DFS measurement. (*not to scale*). A receptor and its binding partner are tethered to the AFM tip and substrate via poly(ethylene glycol) (PEG) linkers. The receptor ligand complex is ruptured by translating the tip away from the substrate with a piezoelectric actuator. During this process, force and tip-surface distance are recorded.

widely used single barrier model, the intrinsic off-rate under zero force, k_{off} and the width of energy barrier that inhibit protein dissociation, x_β are determined by fitting the most probable force at different loading rates to the Bell-Evans model:

$$F^*(r) = F_\beta \ln(r/(k_{off}F_\beta)) \qquad (1)$$

where $F^*(r)$ is the most probable unbinding force, r is the loading rate, $F_\beta = k_B T/x_\beta$, k_B is the Boltzmann's constant and T is the absolute temperature [10, 12]. To increase the quality of the fit, several pulling speeds are used so that the loading rates cover a large dynamic range [13, 14].

In order to measure single molecule binding, DSF experiments are typically designed such that the chance of observing a specific unbinding event is less than 10% [15]. Under these conditions, collecting enough events to recover the unbinding force distribution and accurately estimate the most probable unbinding forces before the sample degrades is often impractical. Consequently, the mean or median rupture force is commonly used for data analysis instead of the most probable force [16-18]. Alternatively, the most probable force is determined by fitting the force histograms to a Gaussian distribution [19]. However, due to thermal fluctuations, heterogeneity of chemical bonds, and 'contaminating' multiple unbinding and nonspecific adhesion events, the measured force distribution often varies from the theoretical model which decreases the accuracy of parameter estimation using the simplified mean or median force methods [20, 21]. This distortion is most pronounced in the prediction of k_{off}, where errors are exponentiated with uncertainties of the same order of magnitude as the estimated value [22]. Different methods have been proposed to improve the accuracy of the estimated k_{off} and x_{β}, such as fitting the force and loading rate distribution with a probability density function or introducing correction algorithms [23, 24]. However, a simple high-accuracy method to improve parameter estimation in DFS experiments, which retains the simplicity of using mean forces, is still lacking.

To overcome this bottleneck, we use cluster analysis to group unbinding events and improve the accuracy of fitted k_{off} and x_{β} in a typical DFS experiment. Cluster analysis is a widely used technique in identifying specific patterns from a large database, such as determining biologically relevant genes in microarray experiments [25], identifying similar behaviors in marketing research [26] and pattern recognition in computer vision [27].

Here, we use three clustering models: Gaussian mixture, logistic regression and K-means to group single molecule unbinding events in DFS and to identify the most representative forces and loading rates for subsequent fitting. We simulated experimental data within a realistic range of k_{off}, x_β, thermal noise, number of pulling speeds and number of events by performing Monte Carlo simulations. The simulated data was analyzed using both conventional analysis and cluster analysis and the performances of different methods were compared. We show that clustering algorithms greatly improve the estimation of k_{off} and x_β, even with the amount of data is limited and where the unbinding events measured at multiple pulling speeds are not well separated from each other.

3.3 Methods

3.3.1 Force-Distance curve simulation

When a receptor-ligand complex is ruptured by withdrawing the cantilever away from the substrate using a piezoelectric actuator (Fig. 1), force-distance (FD) curves are the primary output of the measurement. We therefore simulated unbinding events as FD curves at a range of loading rates (Fig. 2). Our model parameters were chosen to relate the model in Figure 1 to a realistic DFS experiment. We assumed that the receptor and ligand were immobilized on an AFM tip and substrate through poly ethylene glycol (PEG; MW: 3,400 Dalton) linkers. Spring constant of the cantilever was set to 40 pN/nm since soft probes with 10 – 100 pN/nm stiffness are usually used to measure weak biological interactions [13]. The measurements were simulated to occur at 25 °C with k_BT equal to 4.1 pN·nm throughout the study. In order to mimic a realistic DFS experiment where

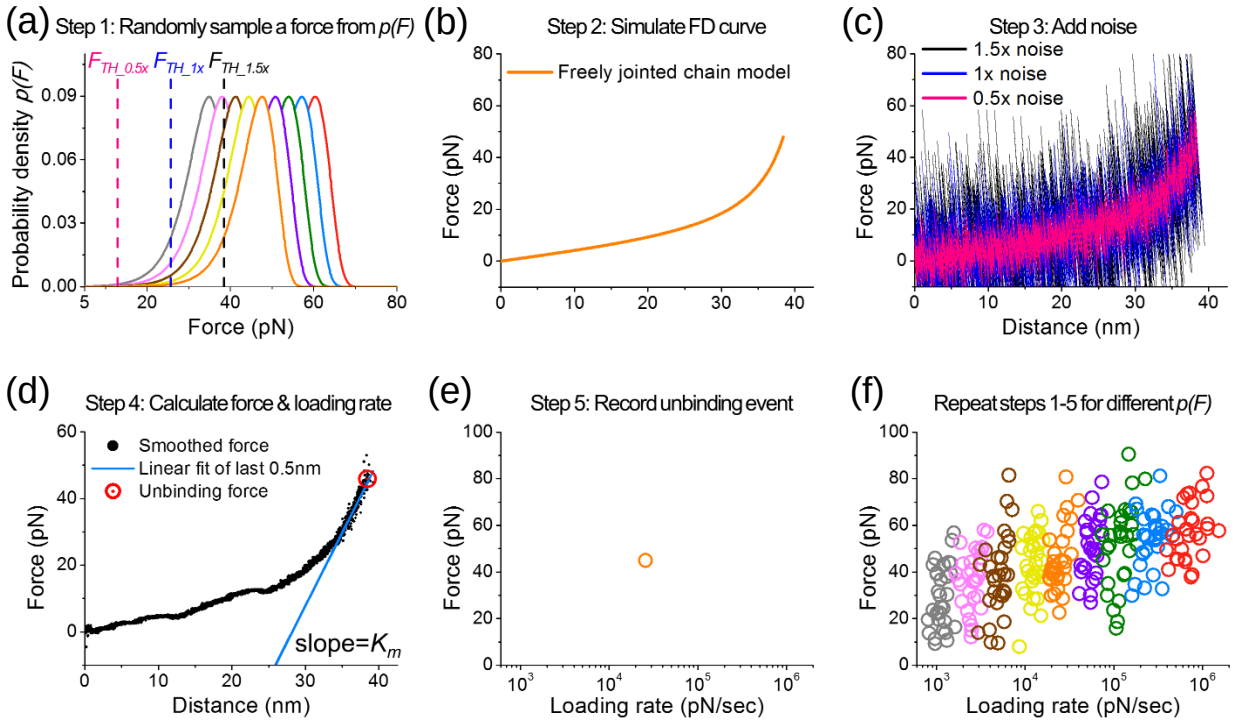

Figure 2. Workflow for simulation of single molecule unbinding events. **(a)** An unbinding force greater than the force threshold (F_{TH}) was randomly sampled from the probability distribution of force. F_{TH} was set to 12.8 pN ($F_{TH\_0.5\times}$), 25.6 pN ($F_{TH\_1\times}$) and 38.4 pN ($F_{TH\_1.5\times}$) for conditions with 0.5×, 1× and 1.5× thermal noise, respectively. **(b)** For the sampled unbinding force, a Force-Distance (FD) curve was simulated at 500 kHz using the freely jointed chain model to account for the stretching of PEG linkers. **(c)** The noise in force and distance due to thermal vibrations of the AFM cantilever was added to the FD curve at each time point. The calculated noise was normally distributed with standard deviations of 6.4 pN, 12.8 pN or 19.2 pN in force and 0.16 nm, 0.32 nm or 0.48 nm in distance, for conditions with 0.5× noise, 1× noise, and 1.5× noise respectively. **(d)** To determine the unbinding force and loading rate, we first smoothed the noisy FD curve using a 4 nm moving average window and estimated the spring constant of the molecule, K_m, by fitting the last 0.5 nm data to a straight line. The loading rate was calculated by substituting K_m into Eq. 4. The last force reading was used as unbinding force. **(e)** The calculated force and loading rate for each FD curve was recorded. **(f)** By repeating the process described in panels *(b)*-*(e),* the unbinding events for a DFS measurement were simulated. Colors represent different pulling speeds.

loading rates usually span only two to three orders of magnitude [13], we fixed the lower and upper bounds on loading rate to be 2,000 and 10^6 pN/sec.

We first calculated the probability distribution of rupture forces, *p(F)*, at a given loading rate, *r*, using the Bell-Evans model [10]:

$$p(F) = \frac{k_{off}}{r} exp\left[\frac{F}{F_\beta} - \frac{k_{off}F_\beta}{r}(e^{F/F_\beta} - 1)\right] \quad (2)$$

Since both the receptor and ligand were immobilized on the tip and substrate using flexible PEG linkers, a non-linear stretching of PEG tethers should be measured in each FD curve. We simulated the PEG stretching using the extended freely jointed chain model [28]:

$$D(F) = L_C \times \left(\coth\left(\frac{FL_K}{k_BT}\right) - \frac{k_BT}{FL_K}\right) + \frac{FL_C}{L_mK_s} \quad (3)$$

where L_C and L_K are the contour length and Kuhn length of the PEG tethers and L_m and K_S are the average length and stiffness of a PEG monomer. Based on previous studies [28], we used a value of 43.6898 nm for L_C, 0.7 nm for L_K, 0.2837 nm for L_m and 150,000 pN/nm for K_S. We simulated FD traces using Eq. 3 at 500 kHz data acquisition rate, with an unbinding force randomly sampled from the probability distribution of force (Eq. 2) (Fig. 2(a)-(b)).

Since the thermal fluctuations of the cantilever, which are detected by the AFM's Quadrant Photodiode (QPD) voltage, couple in as noise in both measured force and the calculated tip-surface distance, we calculated the QPD voltage at each time point of the FD curve and calculated the noise in both force and tip-surface distance, using an optical lever sensitivity of 30 nm/V to correlate changes in the QPD voltage to cantilever fluctuations [29]. Since noise varies with factors such as AFM design, quality factor of

cantilever and environmental noise, we considered three levels of thermal noise in our simulations: 0.5×, 1× or 1.5×. We accounted for these three thermal noise levels by adding normally distributed noise with standard deviations of 6.4 pN, 12.8 pN or 19.2 pN in force and 0.16 nm, 0.32 nm or 0.48 nm in distance to the FD curve (Fig. 2(c)) [30]. In an actual DFS experiment, unbinding forces lower than a force threshold (F_{TH}), which depends on the level of noise, cannot be detected. We accounted for this in our simulations, by setting an F_{TH} value of 12.8 pN, 25.6 pN or 38.4 pN for conditions with 0.5×, 1× or 1.5× thermal noise and only sampled forces greater than F_{TH} (Fig. 2(a)).

Next, we estimated the loading rate for each FD curve, by modeling the cantilever and PEG linker as two springs that were pulled in series. While the spring constant of cantilever, K_C, was fixed, the spring constant of the PEG linker, K_m, was calculated as the slope of the tangent line to the FD curve at the unbinding force. Consequently, the loading rate, r, was calculated as:

$$r = V_{pulling}(K_C K_m/(K_C + K_m)) \qquad (4)$$

where $V_{pulling}$ is the pulling speed. We smoothed the noisy FD trace using a moving 4 nm window and estimating the spring constant of molecule, K_m, by fitting the last 0.5 nm data to a straight line (Fig. 2(d)). Loading rate was determined by substituting K_m, K_C and $V_{pulling}$ into Eq. 4. To simulate the dataset for a DFS experiment, we generated FD curves for different pulling speeds and recorded their rupture forces (last force reading in FD curve) and loading rates (Fig. 2(e)-(f)).

3.3.2 Calculation of k_{off} and x_β

The simulated rupture events were sorted into groups using four methods (described below); the number of groups were limited to be the number of pulling speeds. To extract k_{off} and x_β, we determined the mean force and loading rate of each group and then fitted the mean force *vs.* loading rate to the Bell-Evans model (Eq. 1) using a nonlinear least-squares fitting with bisquare weights. Simulations were repeated 100 times for each condition and k_{off} and x_β were calculated for each simulation. Relative error in x_β was calculated as $\left[median\left(calculated\ x_\beta\right) - (preset\ x_\beta)\right]/\left[preset\ x_\beta\right]$. Relative error in k_{off} was calculated as $\left[e^{median\left(ln\left(calculated\ k_{off}\right)\right)} - (preset\ k_{off})\right]/\left[preset\ k_{off}\right]$. The algorithms for clustering have been derived in ref. [27] and Matlab code was directly adopted from ref. [31]. The methods we used to group data include

Method 1: Pulling speed. This is the standard method in DFS data analysis where unbinding events with the same pulling speed are grouped together.

Method 2: Clustering using 2D Gaussian mixture model (GMM). Forces and loading rates were normalized for GMM since the ranges they span can be dramatically different. Each unbinding force F_i was normalized using $(F_i\text{-}F_{min})/(F_{max}\text{-}F_{min})$, where F_{max} and F_{min} are the maximum and minimum values for force. Loading rate r_i was normalized using $(\ln(r_i)\text{-}\ln(r_{min}))/(\ln(r_{max})\text{-}\ln(r_{min}))$, where r_{max} and r_{min} are the maximum and minimum values for loading rate. As an initial guess for classification, events were grouped according to their loading rates; groups were assigned with equal number of events.

Method 3: Clustering using logistic regression model. Data was normalized as in *Method 2* and the initial guess was used as the training dataset for 2D logistic regression clustering.

Method 4: Clustering using 1D K-means. Events were separated into groups based on the normalized loading rates. The initial guess was the same as in GMM.

3.4 Results and Discussion

3.4.1 Overview of cluster analysis

Cluster analysis is the process of sorting data into different groups such that events within the same category share similar characteristics. The main idea in applying this approach to a DFS experiment is that when a specific interaction is probed repeatedly using the same tip-sample pulling speed, the measured unbinding forces and loading rates are expected be similar within a certain noise level. Therefore, unbinding events are expected to form clusters on a force versus loading rate plot. The mean forces and loading rates calculated from the clustered events share common characteristics such that the influence of outliers are reduced. To test this idea, we simulated unbinding events to closely mimic a realistic DFS experiment and grouped the events either using cluster analysis algorithms or according to pulling speed (the standard DFS analysis method where the unbinding events are grouped according to the tip-surface retraction speeds). We used three clustering analysis approaches in this work: Gaussian mixture model (GMM), logistic regression and K-means (*Methods*) [27, 32].

Our rationale in using GMM [27] is that since the rupture force distribution at a constant pulling speed resembles a skewed Gaussian with a long tail at low forces [10],

the distribution of forces collected with many pulling speeds can be approximated as a mixture of Gaussian distributions. GMM was used to assign each unbinding event to a group by maximizing the posterior probability that the data point belongs to its assigned cluster such that the force distribution within each group is most likely to be a Gaussian [27]. In contrast to GMM, logistic regression identifies the boundaries between groups based on a training dataset, while K-means partitions data into clusters by minimizing the distance from the data point to the mean of its assigned cluster [27, 32]. The theory and mathematical derivation of these methods are beyond the scope of this study; we merely adopt these clustering algorithms as data-analysis tools.

3.4.2 Overlap of data increases with increasing k_{off}, x_β, noise, number of data and pulling speeds

Since the goal of cluster analysis is to partition data into groups by relocating ambiguous events at the boundaries into their proper categories, this approach is beneficial when unbinding events across multiple pulling speeds overlap. To generate overlapping datasets, we first examined how each parameter in our simulation affects data overlap (Fig. 3).

With a fixed range of loading rates, one would intuitively expect data overlap to increase with the number of pulling speeds, the number of unbinding events, and the level of thermal noise. To confirm this we first simulated a DFS experiment consisting of 600 unbinding events across 6 pulling speeds, using a k_{off} = 0.1 sec^{-1}, x_β = 1 nm, and with 1× thermal noise (Fig. 3(a); *Methods*). Fixing all the other parameters in the simulation, we increased the number of pulling speeds (9 pulling speeds, Fig. 3(b)), decreased the number of events per pulling speed (30 events per pulling speed, Fig. 3(c)), decrease the

noise level (0.5× thermal noise, Fig. 3(d)), reduced the k_{off} (k_{off} = 0.001 sec^{-1}, Fig. 3(e)) and reduced x_β (x_β = 0.5 nm, Fig. 3(f)). As expected, the degree of data overlap increased with the number of pulling speeds, number of unbinding events and with a higher noise level (Fig. 3(a)-(d)). The simulated data also showed an increasing overlap as k_{off} and x_β increased (Fig. 3(a), (e)-(f)).

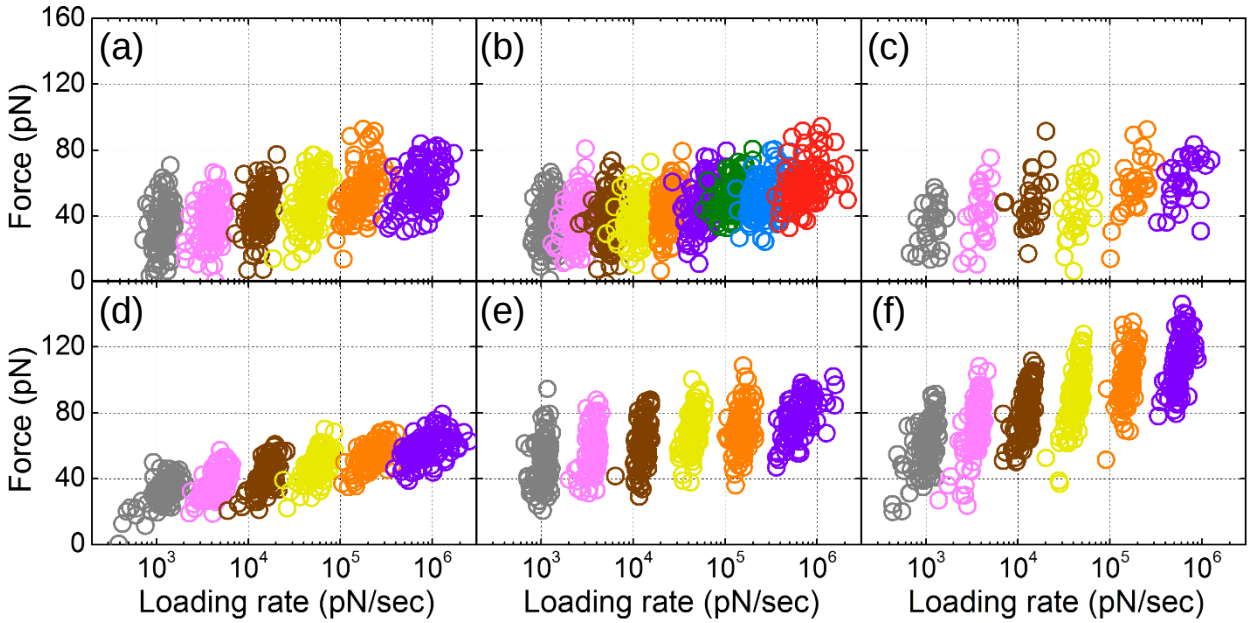

Figure 3. Unbinding events across multiple pulling speeds overlap when the number of pulling speeds, number of data points and thermal noise increases. (a) Data was simulated using k_{off} = 0.1 sec^{-1}, x_β = 1 nm, 6 pulling speeds, 100 events per pulling speed with 1× thermal noise. Colors represent different speeds. **(b)-(f)** Similar to panel *(a)*, unbinding events were simulated while varying only one parameter at a time. Simulations were carried out with **(b)** 9 pulling speeds; **(c)** 30 events per pulling speed; **(d)** 0.5× thermal noise; **(e)** k_{off} = 0.001 sec^{-1}; **(f)** x_β = 0.5 nm. Each circle represents one unbinding event.

3.4.3 Cluster analysis improves the estimation of k_{off} and x_β

Next, we compared the accuracy of different cluster analysis methods and the standard pulling speed method on the estimation k_{off} and x_β. An example dataset containing 270 simulated unbinding events evenly distributed over 9 tip-sample separation speeds with 1× thermal noise using k_{off} = 0.1 sec^{-1} and x_β = 1 nm is shown in Figure 4 (*Methods*). The events were separated into 9 groups using either pulling speeds, GMM, logistic regression or K-means (Fig. 4; *Methods*). The most probable unbinding forces and loading rates were calculated by averaging data within each group (Fig. 4, red squares); k_{off} and x_β were then extracted by fitting those mean values to the Bell-Evans equation (Fig. 4, red lines). We performed simulations where only one parameter (either the thermal noise, number of unbinding events, number of pulling speeds, k_{off} or x_β) was varied at a time, and the accuracy of different grouping methods on k_{off} and x_β estimation was compared. When thermal noise, number of data points or pulling speeds were varied, the k_{off} and x_β were fixed at 0.1 sec^{-1} and 1 nm, respectively. When the k_{off} or x_β were varied, 9 pulling speeds and 30 unbinding events per speed with 1× thermal noise were used. Simulations for each condition were repeated 100 times and the statistical distribution of estimated k_{off} and x_β were plotted (Fig. 5, 6). In the following discussion, we focus on estimated errors in k_{off} since the estimation of x_β is accurate within 10% error in all analysis.

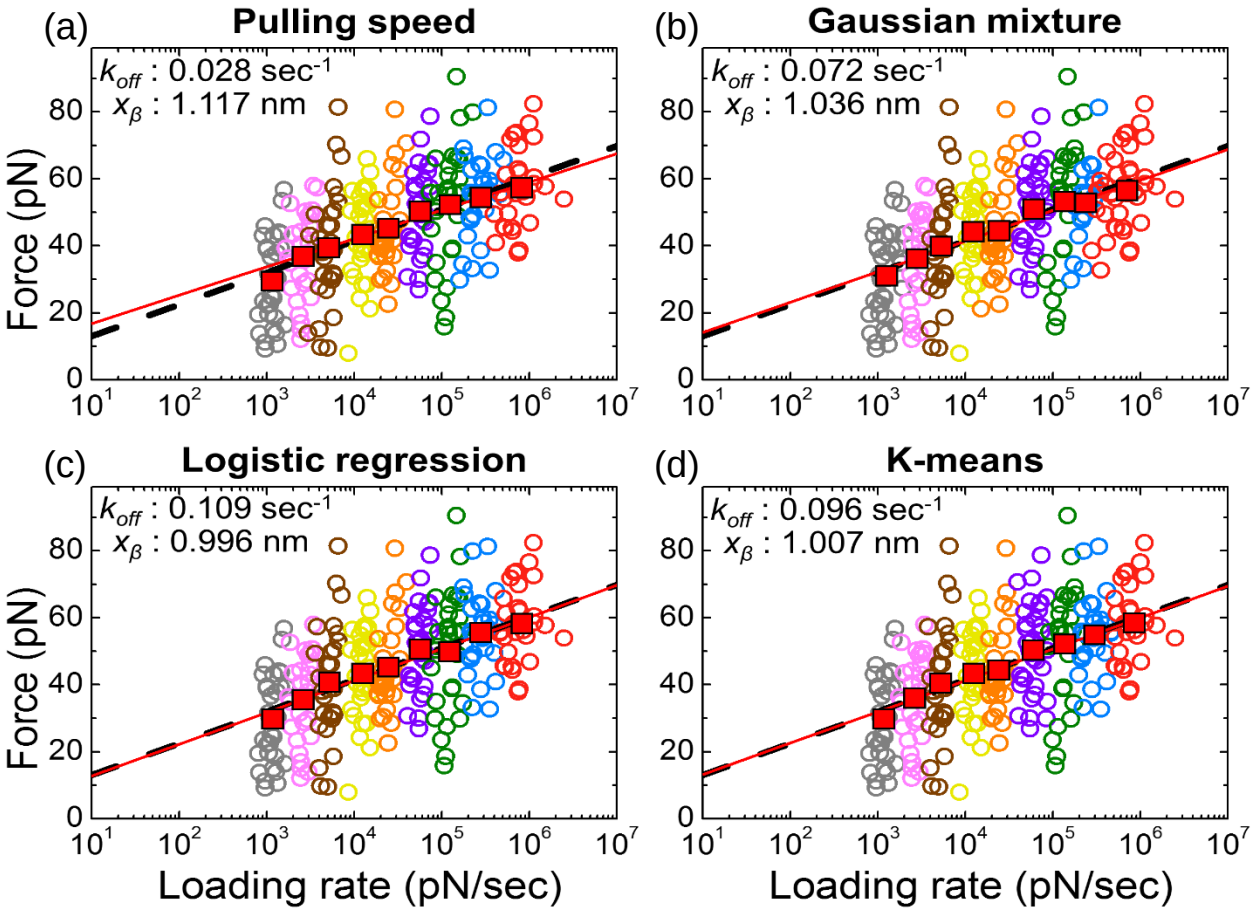

Figure 4. Visual comparison of different clustering algorithms used. Unbinding events were simulated using k_{off} = 0.1 sec^{-1}, x_β = 1 nm, 9 pulling speeds, 30 events per pulling speed with 1× thermal noise. Events were classified into different groups based on **(a)** pulling speeds; **(b)** Gaussian mixture cluster model; **(c)** logistic regression clustering; **(d)** K-means clustering. The results of grouping are indicated by colors. The average forces and loading rates (red squares) within each group were fit to the Bell-Evans model (red line) to estimate k_{off} and x_β. The estimated values are indicated. Black line represents the plot of Bell-Evans equation using k_{off} = 0.1 sec^{-1}, x_β = 1 nm. Differences between the different methods were observed at the boundaries between groups.

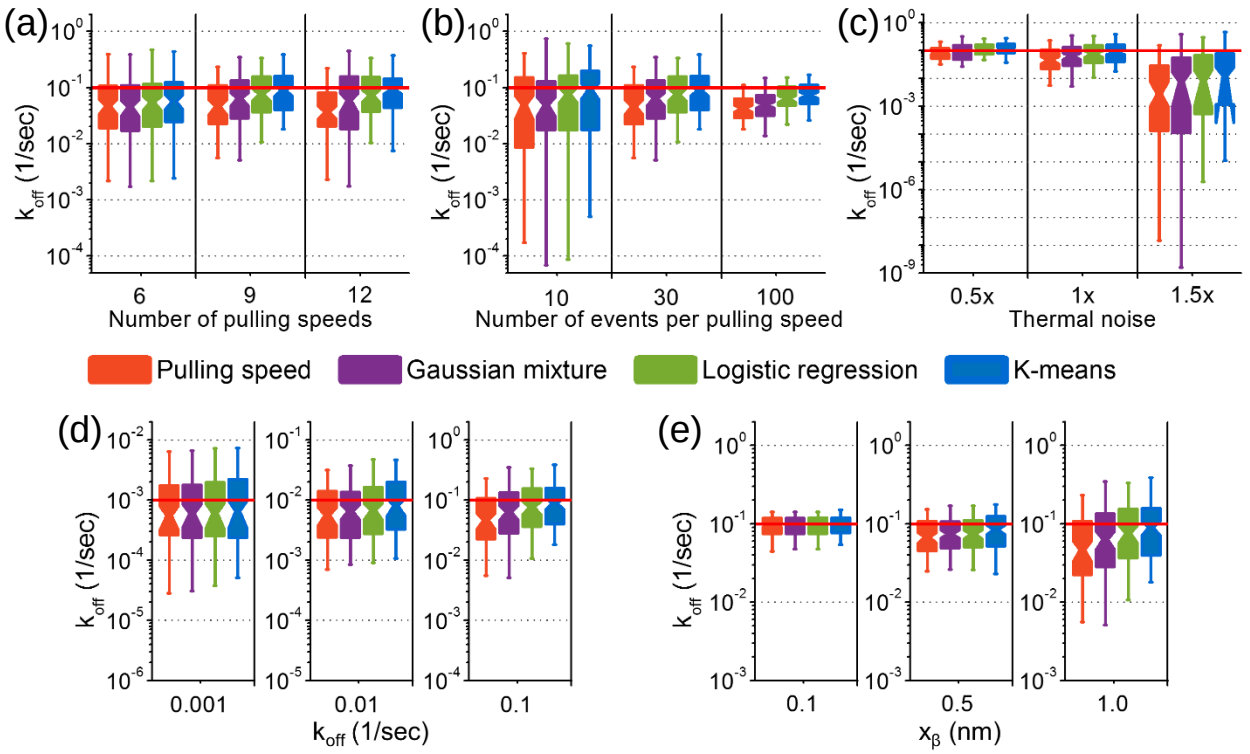

Figure 5. Evaluation of cluster analysis for estimating k_{off} in DFS experiments. DFS data was generated using Monte Carlo simulations with k_{off} = 0.1 sec^{-1}, x_β = 1 nm, 9 pulling speeds, 30 events per pulling speed and with 1× thermal noise. Four classification methods (pulling speeds, Gaussian mixture cluster model, logistic regression clustering, and K-means clustering) were used and kinetic parameters were estimated using Bell-Evans model. Simulations were repeated 100 times; the distribution of estimated k_{off} are plotted. Red line represents the preset values of k_{off}. The performance of methods was compared by varying **(a)** number of pulling speeds, **(b)** amount of data, **(c)** thermal noise, **(d)** k_{off} and **(e)** x_β. Cluster analysis significantly improved k_{off} estimation when the number of pulling speeds, number of unbinding events and the thermal noise increased. Cluster analysis was especially accurate at high dissociation rates and wide energy barriers.

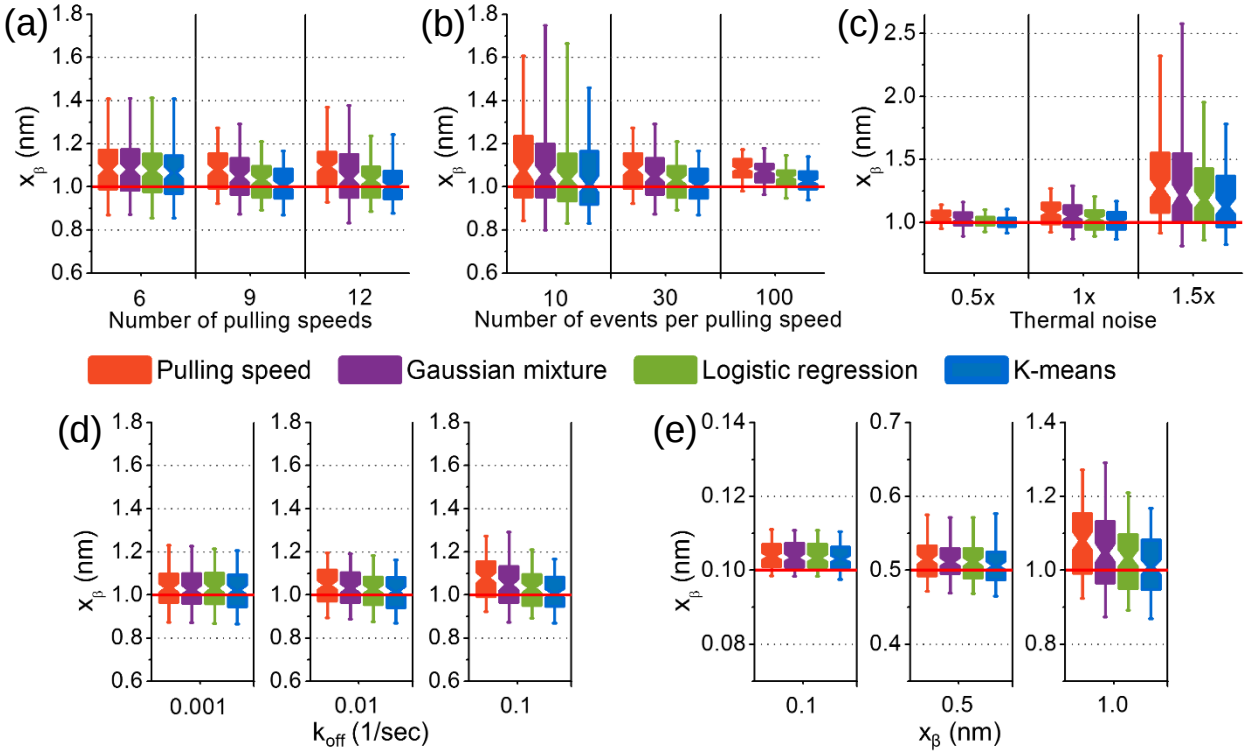

Figure 6. Evaluation of cluster analysis for estimation of x_β. Unbinding events were generated using Monte Carlo simulations with k_{off} = 0.1 sec^{-1}, x_β = 1 nm, 9 pulling speeds, 30 events per pulling speed and with 1× thermal noise. Pulling speeds, Gaussian mixture cluster model, logistic regression clustering, and K-means clustering were used to group data. Kinetic parameters were estimated using Bell-Evans model. Simulations were repeated 100 times. Red line represents the preset values of x_β. Clustering methods were compared by varying **(a)** number of pulling speeds, **(b)** amount of data, **(c)** thermal noise, **(d)** k_{off} and **(e)** x_β. As in the case of k_{off} estimates shown in figure 5, cluster analysis significantly improved x_β estimation when the number of pulling speeds, number of unbinding events, thermal noise, k_{off} and x_β increased.

First, we tested k_{off} estimation using different grouping methods when the number of pulling speeds were varied (6, 9, and 12 pulling speeds). Our data showed that as the number of pulling speeds was increased, cluster analysis improved k_{off} estimation (Fig. 5(a)). While relative errors in k_{off} using the standard pulling speed method were between -53% to -63% across all conditions, K-means analysis reduced the relative error from -43% at 6 pulling speeds to -11% and -8% at 9 and 12 pulling speeds respectively. Logistic regression also decreased the error from -46% to -24% and -27% as the number of tip-surface retraction speeds were increased. In contrast, GMM showed a less significant improvement, with errors of -55%, -36% and -42% at 6, 9, and 12 pulling speeds respectively.

Next, we measured k_{off} estimation when the number of data points in each group equaled 10, 30, and 100 unbinding events. Our results showed that k_{off} estimation using K-means method was superior, even when the amount of data was limited (Fig. 5(b)). When individual groups had 10, 30, and 100 rupture events, the relative errors using K-means clustering was -9%, -11% and -27% which was significantly lower than errors of -50%, -54% and -58% measured with the standard pulling speed analysis. The logistic regression model also showed better accuracy, with errors of -24%, -24% and -37% while GMM showed a less significant improvement, with errors of -51%, -36% and -51% at 10 events, 30 events and 100 events per group respectively. Interestingly, while increasing the amount of data mainly increased precision, k_{off} accuracy did not increase.

The accuracy of k_{off} estimation decreased with an increase in thermal noise using all grouping methods (Fig. 5(c)). When the force due to thermal fluctuations of the cantilever was 6.4 pN, 12.8 pN and 19.2 pN (0.5×, 1×, and 1.5× thermal noise), the

relative errors in estimated k_{off} using the standard pulling speed method (-15%, -54% and -97%), k-means analysis (18%, -11%, and -88%), logistic regression (14%, -24% and -92%), and GMM (-1%, -36% and -92%) were comparable.

Importantly, our data showed that cluster analysis was especially useful for studying molecular interactions with high dissociation rates and wide energy barriers (Fig. 5(d), (e)). At dissociation rates of 10^{-3} sec^{-1}, 0.01 sec^{-1}, and 0.1 sec^{-1} the relative error in k_{off} estimated by K-means was -34%, -19%, and -11% respectively. The relative error using logistic regression was comparable to K-means with values of -39% for 10^{-3} sec^{-1}, -32% for 0.01 sec^{-1}, and -24% for 0.1 sec^{-1} respectively. In contrast the k_{off} calculated using both GMM and the standard pulling speed method showed larger relative error (-40%, -35%, -36% for GMM and -44%, -44%, -54% for pulling speed) for off rates of 10^{-3} sec^{-1}, 0.01 sec^{-1}, and 0.1 sec^{-1} respectively.

Similarly, when the conventional pulling speed method was used, increasing the width of energy barrier increased the error in k_{off} estimation (-2% error at 0.1 nm, -33% error at 0.5 nm, and -54% error at 1 nm). However, this increase in relative error was not observed when cluster analysis was used (Fig. 5(e)). With barrier widths of 0.1 nm, 0.5 nm and 1 nm, the errors in estimated k_{off} were 0%, -16% and -11% for K-means; -1%, -26% and -24% for logistic regression; -1%, -26% and -36% for GMM.

Most importantly, the improvement in x_β estimation by cluster analysis followed a very similar trend as the k_{off} estimation (Fig. 6), indicating that the accuracy of both k_{off} and x_β estimates increased at the same time. This is particularly encouraging because it demonstrates that the increased accuracy of k_{off} by cluster analysis was not offset by a reduced accuracy in x_β estimation.

Finally, we validated the effect of clustering by simultaneously varying parameters across a range of values that have previously been measured in DFS experiments with biological systems including cell adhesion proteins and antigen-antibody complexes [13]. A total of 324 DFS experiments were simulated. We used four values of k_{off} (10^{-3}, 0.01, 0.1 and 1 sec^{-1}) in combination with three values of x_β (0.1, 0.5 and 1 nm). In order to account for different experimental conditions, we also varied the number of pulling speeds (6, 9 and 12), number of events per pulling speed (10, 30 and 100 events), and different levels of thermal noise (0.5×, 1× and 1.5×). We applied cluster analysis to each condition and compared the estimated k_{off} and x_β to the results obtained using the pulling speed method. The complete results of our simulations are presented in the *Supporting Information* (Figs. S1-S6; Tables S1-S3). In Figure 7 of the main manuscript, we just present results using 9 pulling speeds and 1× thermal noise analyzed using K-means, to illustrate its power in estimating k_{off} and x_β (Fig. 7). We use stars to indicate conditions where K-means is more accurate and triangles to indicate conditions where the standard pulling speed analysis is more accurate; increase in the number of stars/triangle indicates a proportionally higher accuracy using K-means/standard-analysis.

As seen in Fig. 7, K-means clustering improved the accuracy of k_{off} and x_β even when the number of unbinding events were low. The improved accuracy of parameter estimation was more pronounced for wide energy barriers, high off-rates and when the unbinding events across different loading rates overlap (Fig. 7). For instance, at a dissociation rate of 0.1 sec^{-1}, x_β of 1 nm and 30 events per pulling speed, K-means reduced the relative error in k_{off} to -11% as compared to a -54% relative error using pulling speed analysis. In contrast, when the energy barrier was narrow or the off-rate was small,

the unbinding events were already well-separated and cluster analysis did not significantly improve estimation of k_{off} and x_β.

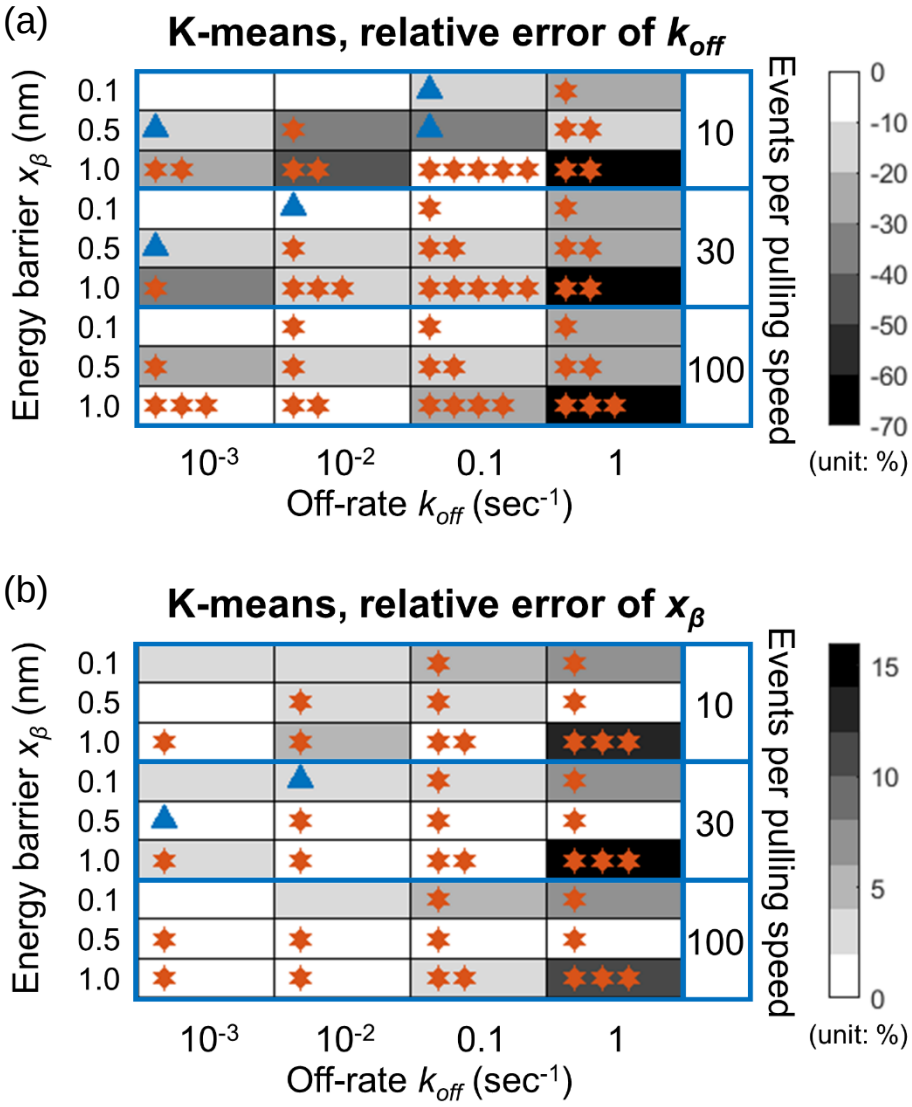

Figure 7. Relative error in estimated (a) k_{off} and (b) x_β using K-means clustering. DFS data was simulated using 9 pulling speeds and 1× thermal noise. Other parameters include k_{off} = 10^{-3}, 0.01, 0.1, 1 sec^{-1}; x_β = 0.1, 0.5, 1 nm; the number of events per pulling speed = 10, 30, 100. For each condition, the relative errors were compared with pulling speed method. Stars indicate conditions where K-means is more accurate (0 stars: no improvement; 1 star: relative error is reduced ≤ 10% for k_{off} and ≤ 5% for x_β; 2 stars: relative error is reduced by 10% – 20% for k_{off} and 5% – 10% for x_β; etc.). Triangles indicate conditions where the K-means analysis is less accurate; relative error is increased ≤ 10% for k_{off} and ≤ 5% for x_β.

3.5 Conclusion

This manuscript presents a high accuracy method using clustering algorithms, to improve kinetic parameter estimation while retaining the simplicity of data collection and analysis of a conventional DFS experiment. We benchmarked the performance of different clustering algorithms, by testing them across an extensive range of conditions that mimic real-world experiments. The parameters we varied included the number of unbinding events, pulling speeds, and noise levels, across a range of k_{off} and x_{β} typical of receptor-ligand pairs. Under these conditions, the K-means method had the highest accuracy in estimating k_{off} and x_{β}. Although logistic regression and GMM were more accurate than the conventional pulling speed method, the improvement was not as significant as K-means.

The cluster analysis used in this study could be further improved, by grouping unbinding events using Bell-Evans force distributions [10] or more sophisticated distributions described in the literature [33, 34]. The analysis method presented in our work can also be used to identify and eliminate artifacts due to the formation of multiple receptor ligand bonds and nonspecific binding events which are not tightly clustered on a force-loading rate plot [35]. Finally, although we used the classic Bell-Evans model in this manuscript, similar clustering analysis can be easily applied to other DFS models described in the literature [34, 36, 37].

3.6 Acknowledgments

This work was supported by Grant PHY-1607550 from the National Science Foundation. We thank Andrew Priest for help with the simulations. These simulations were supported in part by the Research IT team at Iowa State University, utilizing HPC@ISU equipment, some of which has been purchased through funding provided by NSF under MRI grant number CNS 1229081 and CRI grant number 1205413.

3.7 References

[1] V. Montana, W. Liu, U. Mohideen, V. Parpura, Single molecule measurements of mechanical interactions within ternary SNARE complexes and dynamics of their disassembly: SNAP25 vs. SNAP23, J. Physiol. 587 (2009) 1943-1960.

[2] A. Yersin, T. Osada, A. Ikai, Exploring transferrin-receptor interactions at the single-molecule level, Biophys. J. 94 (2008) 230-240.

[3] W. Baumgartner, P. Hinterdorfer, W. Ness, A. Raab, D. Vestweber, H. Schindler, D. Drenckhahn, Cadherin interaction probed by atomic force microscopy, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 4005-4010.

[4] F.W. Bartels, B. Baumgarth, D. Anselmetti, R. Ros, A. Becker, Specific binding of the regulatory protein ExpG to promoter regions of the galactoglucan biosynthesis gene cluster of Sinorhizobium meliloti--a combined molecular biology and force spectroscopy investigation, J. Struct. Biol. 143 (2003) 145-152.

[5] J. Yu, Y. Jiang, X. Ma, Y. Lin, X. Fang, Energy landscape of aptamer/protein complexes studied by single-molecule force spectroscopy, Chem. - Asian J. 2 (2007) 284-289.

[6] B.H. Kim, N.Y. Palermo, S. Lovas, T. Zaikova, J.F. Keana, Y.L. Lyubchenko, Single-molecule atomic force microscopy force spectroscopy study of Abeta-40 interactions, Biochemistry 50 (2011) 5154-5162.

[7] C.F. Yen, D.S. Harischandra, A. Kanthasamy, S. Sivasankar, Copper-induced structural conversion templates prion protein oligomerization and neurotoxicity, Sci. Adv. 2 (2016) e1600014.

[8] K.C. Neuman, A. Nagy, Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy, Nat. Methods 5 (2008) 491-505.

[9] P. Hinterdorfer, Y.F. Dufrene, Detection and localization of single molecular recognition events using atomic force microscopy, Nat. Methods 3 (2006) 347-355.

[10] E. Evans, K. Ritchie, Dynamic strength of molecular adhesion bonds, Biophys. J. 72 (1997) 1541-1555.

[11] J.T. Bullerjahn, S. Sturm, K. Kroy, Theory of rapid force spectroscopy, Nat. Commun. 5 (2014) 4463.

[12] G.I. Bell, Models for the specific adhesion of cells to cells, Science 200 (1978) 618-627.

[13] A.R. Bizzarri, S. Cannistraro, Dynamic force spectroscopy and biomolecular recognition, CRC Press, Boca Raton, FL, 2012.

[14] A. Noy, R.W. Friddle, Practical single molecule force spectroscopy: how to determine fundamental thermodynamic parameters of intermolecular bonds with an atomic force microscope, Methods 60 (2013) 142-150.

[15] D.F. Tees, R.E. Waugh, D.A. Hammer, A microcantilever device to assess the effect of force on the lifetime of selectin-carbohydrate bonds, Biophys. J. 80 (2001) 668-682.

[16] D.A. Simson, M. Strigl, M. Hohenadl, R. Merkel, Statistical breakage of single protein A-IgG bonds reveals crossover from spontaneous to force-induced bond dissociation, Phys. Rev. Lett. 83 (1999) 652-655.

[17] S. Loi, G. Sun, V. Franz, H.J. Butt, Rupture of molecular thin films observed in atomic force microscopy. II. Experiment, Phys. Rev. E 66 (2002) 031602.

[18] C. Gergely, J. Voegel, P. Schaaf, B. Senger, M. Maaloum, J.K. Horber, J. Hemmerle, Unbinding process of adsorbed proteins under external stress studied by atomic force microscopy spectroscopy, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 10802-10807.

[19] C. Ray, J.R. Brown, B.B. Akhremitchev, Single-molecule force spectroscopy measurements of "hydrophobic bond" between tethered hexadecane molecules, J. Phys. Chem. B 110 (2006) 17578-17583.

[20] M. Raible, M. Evstigneev, F.W. Bartels, R. Eckel, M. Nguyen-Duong, R. Merkel, R. Ros, D. Anselmetti, P. Reimann, Theoretical analysis of single-molecule force spectroscopy experiments: heterogeneity of chemical bonds, Biophys. J. 90 (2006) 3851-3864.

[21] S. Getfert, P. Reimann, Hidden multiple bond effects in dynamic force spectroscopy, Biophys. J. 102 (2012) 1184-1193.

[22] C. Ray, J.R. Brown, B.B. Akhremitchev, Rupture force analysis and the associated systematic errors in force spectroscopy by AFM, Langmuir 23 (2007) 6076-6083.

[23] C. Friedsam, A.K. Wehle, F. Kuhner, H.E. Gaub, Dynamic single-molecule force spectroscopy: bond rupture analysis with variable spacer length, J. Phys.: Condens. Matter 15 (2003) S1709-S1723.

[24] C. Ray, J.R. Brown, B.B. Akhremitchev, Correction of systematic errors in single-molecule force spectroscopy with polymeric tethers by atomic force microscopy, J. Phys. Chem. B 111 (2007) 1963-1974.

[25] P. D'haeseleer, How does gene expression clustering work?, Nat. Biotechnol. 23 (2005) 1499-1501.

[26] G. Punj, D.W. Stewart, Cluster-Analysis in Marketing-Research - Review and Suggestions for Application, J. Marketing Res. 20 (1983) 134-148.

[27] C.M. Bishop, Pattern recognition and machine learning, Springer, New York, 2006.

[28] F. Oesterhelt, M. Rief, H.E. Gaub, Single molecule force spectroscopy by AFM indicates helical structure of poly(ethylene-glycol) in water, New J. Phys. 1 (1999).

[29] R. Levy, M. Maaloum, Measuring the spring constant of atomic force microscope cantilevers: thermal fluctuations and other methods, Nanotechnology 13 (2002) 33-37.

[30] H.J. Butt, M. Jaschke, Calculation of Thermal Noise in Atomic-Force Microscopy, Nanotechnology 6 (1995) 1-7.

[31] M. Chen, Pattern Recognition and Machine Learning Toolbox http://prml.github.io/, 2016.

[32] J.C. Stoltzfus, Logistic Regression: A Brief Primer, Acad. Emerg. Med. 18 (2011) 1099-1104.

[33] O.K. Dudko, G. Hummer, A. Szabo, Theory, analysis, and interpretation of single-molecule force spectroscopy experiments, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 15755-15760.

[34] O.K. Dudko, G. Hummer, A. Szabo, Intrinsic rates and activation free energies from single-molecule pulling experiments, Phys. Rev. Lett. 96 (2006) 108101.

[35] T. Sulchek, R.W. Friddle, A. Noy, Strength of multiple parallel biological bonds, Biophys. J. 90 (2006) 4686-4691.

[36] G. Hummer, A. Szabo, Kinetics from nonequilibrium single-molecule pulling experiments, Biophys. J. 85 (2003) 5-15.

[37] R.W. Friddle, A. Noy, J.J. De Yoreo, Interpreting the widespread nonlinear force spectra of intermolecular bonds, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 13573-13578.

3.8 Supporting Information

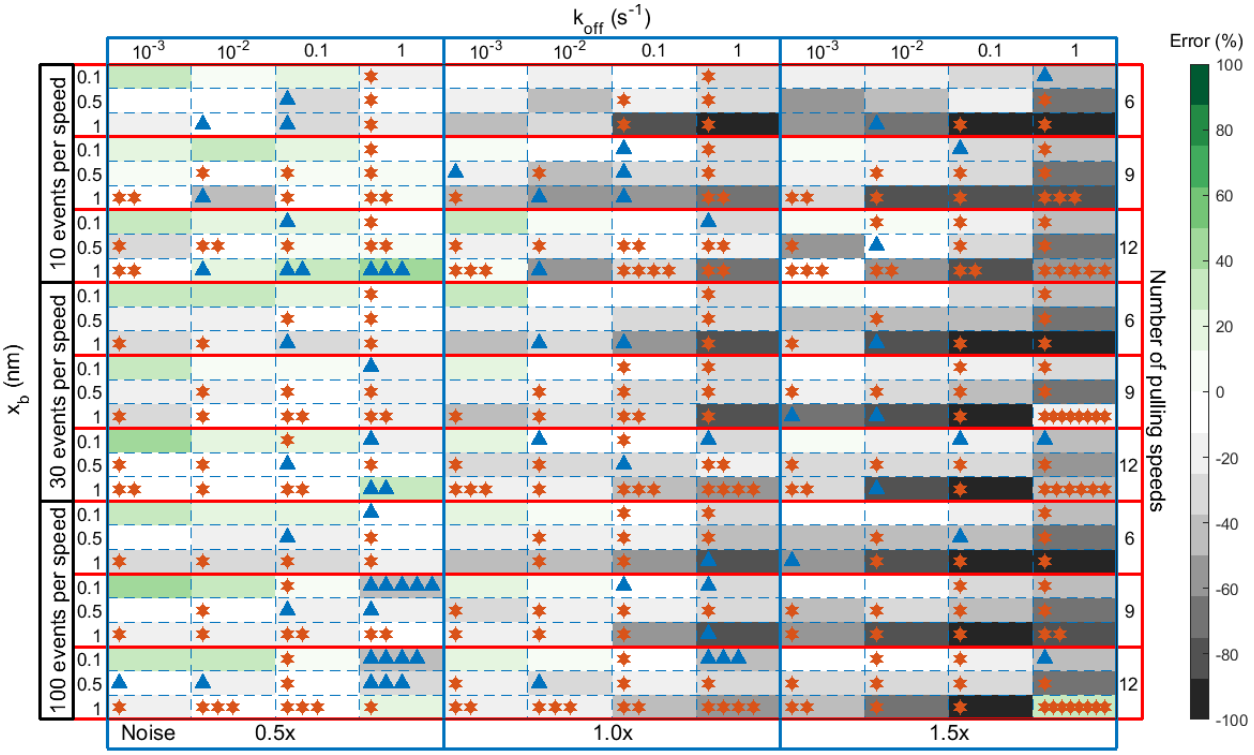

Figure S1. **Relative error in estimated k_{off} using Gaussian mixture model (GMM) cluster analysis.** The effect of GMM was validated by varying parameters across a realistic range of values. DFS data was simulated using four values of k_{off} (10^{-3}, 10^{-2}, 0.1 and 1 sec^{-1}) in combination with three values of x_β (0.1, 0.5 and 1.0 nm), three numbers of pulling speeds (6, 9 and 12), and three amounts of data (10, 30 and 100 events per pulling speed) under three levels of thermal noise (0.5×, 1× and 1.5×). For each condition, the relative error of k_{off} was compared with the standard pulling speed analysis. Stars indicate conditions where GMM is more accurate (0 stars: no improvement; 1 star: relative error is reduced ≤ 10%; 2 stars: relative error is reduced by 10% – 20%; etc.). Triangles indicate conditions where the GMM is less accurate (1 triangle: relative error is increased ≤ 10%; 2 triangles: relative error is increased by 10% – 20%; etc.).

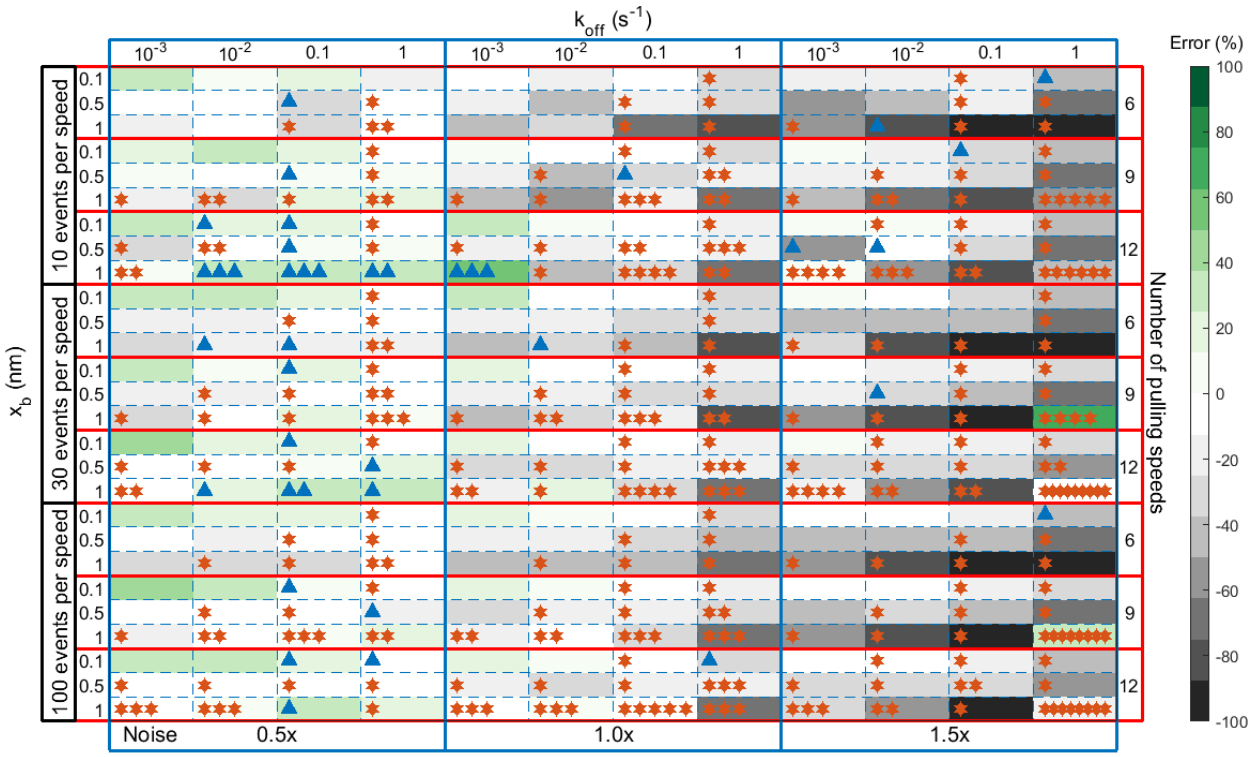

Figure S2. Relative error in estimated k_{off} using logistic regression cluster analysis. The effect of logistic regression clustering was validated by varying parameters across a realistic range of values. DFS data was simulated using four values of k_{off} (10^{-3}, 10^{-2}, 0.1 and 1 sec^{-1}) in combination with three values of x_β (0.1, 0.5 and 1.0 nm), three numbers of pulling speeds (6, 9 and 12), and three amounts of data (10, 30 and 100 events per pulling speed) under three levels of thermal noise (0.5×, 1× and 1.5×). For each condition, the relative error of k_{off} was compared with the standard pulling speed analysis. Stars indicate conditions where logistic regression is more accurate (0 stars: no improvement; 1 star: relative error is reduced ≤ 10%; 2 stars: relative error is reduced by 10% – 20%; etc.). Triangles indicate conditions where the logistic regression is less accurate (1 triangle: relative error is increased ≤ 10%; 2 triangles: relative error is increased by 10% – 20%; etc.).

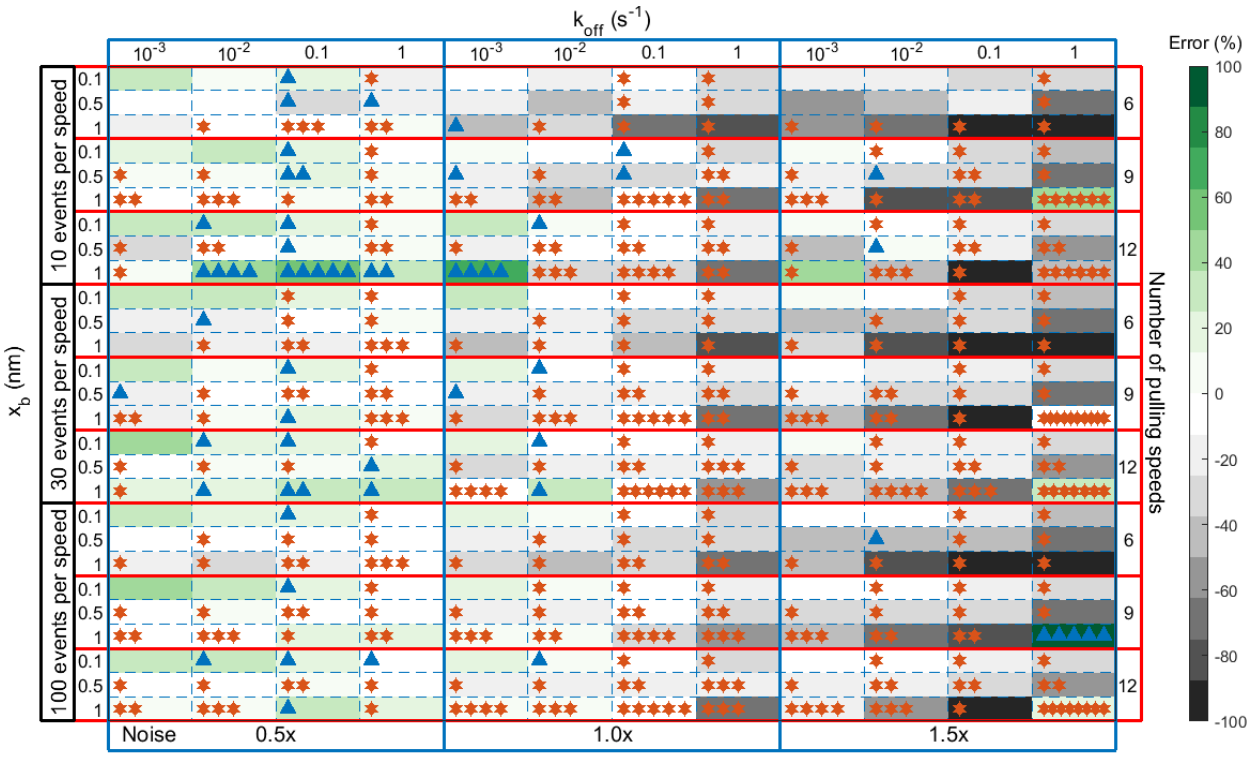

Figure S3. **Relative error in estimated k_{off} using K-means cluster analysis.** The effect of K-means clustering was validated by varying parameters across a realistic range of values. DFS data was simulated using four values of k_{off} (10^{-3}, 10^{-2}, 0.1 and 1 sec^{-1}) in combination with three values of x_β (0.1, 0.5 and 1.0 nm), three numbers of pulling speeds (6, 9 and 12), and three amounts of data (10, 30 and 100 events per pulling speed) under three levels of thermal noise (0.5×, 1× and 1.5×). For each condition, the relative error of k_{off} was compared with the standard pulling speed analysis. Stars indicate conditions where K-means is more accurate (0 stars: no improvement; 1 star: relative error is reduced ≤ 10%; 2 stars: relative error is reduced by 10% – 20%; etc.). Triangles indicate conditions where the K-means is less accurate (1 triangle: relative error is increased ≤ 10%; 2 triangles: relative error is increased by 10% – 20%; etc.).

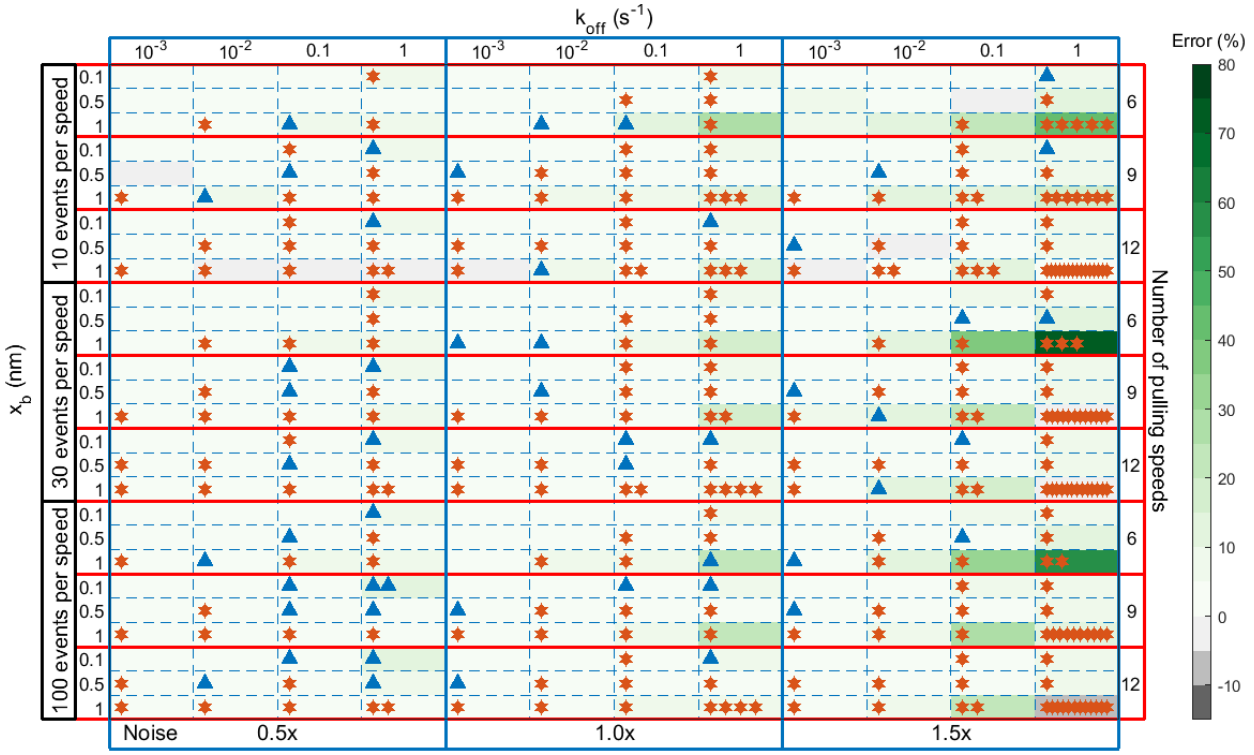

Figure S4. Relative error in estimated x_β using Gaussian mixture model (GMM) cluster analysis. The effect of GMM was validated by varying parameters across a realistic range of values. DFS data was simulated using four values of k_{off} (10^{-3}, 10^{-2}, 0.1 and 1 sec^{-1}) in combination with three values of x_β (0.1, 0.5 and 1.0 nm), three numbers of pulling speeds (6, 9 and 12), and three amounts of data (10, 30 and 100 events per pulling speed) under three levels of thermal noise (0.5×, 1× and 1.5×). For each condition, the relative error of x_β was compared with the standard pulling speed analysis. Stars indicate conditions where GMM is more accurate (0 stars: no improvement; 1 star: relative error is reduced ≤ 5%; 2 stars: relative error is reduced by 5% – 10%; etc.). Triangles indicate conditions where the GMM is less accurate (1 triangle: relative error is increased ≤ 5%; 2 triangles: relative error is increased by 5% – 10%).

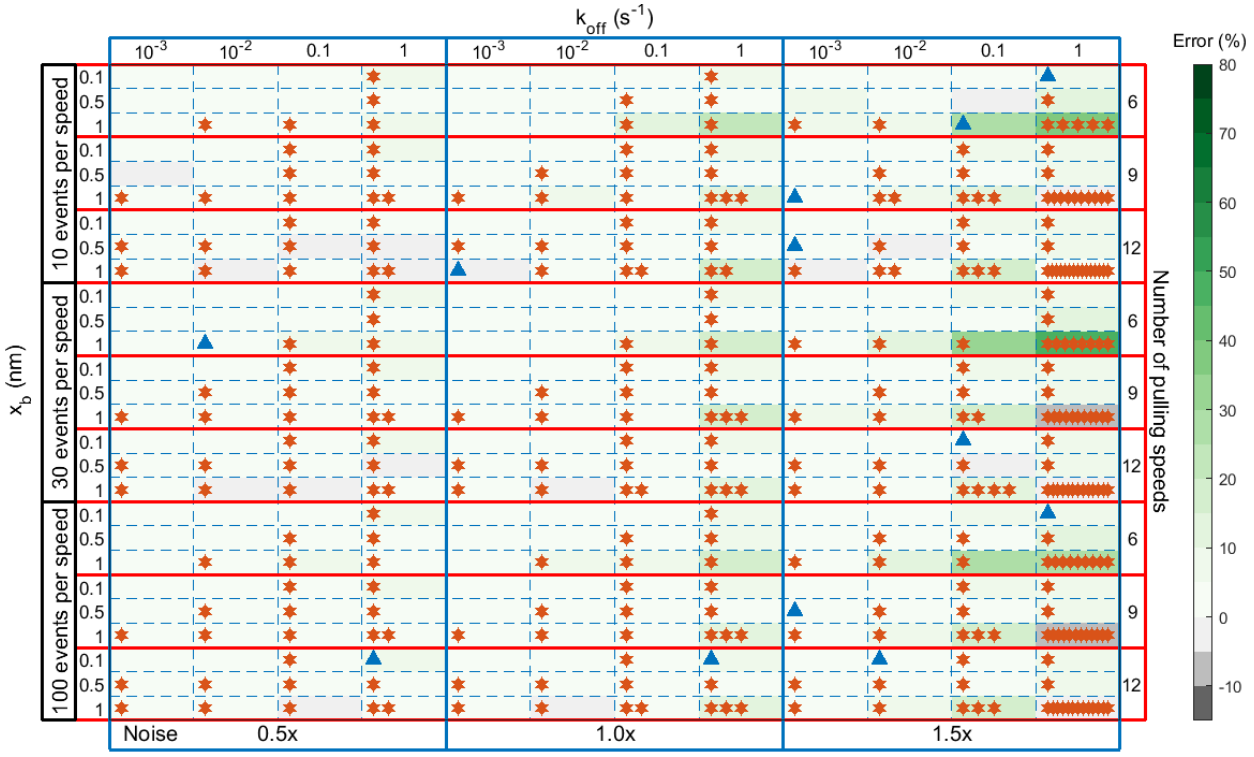

Figure S5. Relative error in estimated x_β using logistic regression cluster analysis. The effect of logistic regression clustering was validated by varying parameters across a realistic range of values. DFS data was simulated using four values of k_{off} (10^{-3}, 10^{-2}, 0.1 and 1 sec^{-1}) in combination with three values of x_β (0.1, 0.5 and 1.0 nm), three numbers of pulling speeds (6, 9 and 12), and three amounts of data (10, 30 and 100 events per pulling speed) under three levels of thermal noise (0.5×, 1× and 1.5×). For each condition, the relative error of x_β was compared with the standard pulling speed analysis. Stars indicate conditions where logistic regression is more accurate (0 stars: no improvement; 1 star: relative error is reduced ≤ 5%; 2 stars: relative error is reduced by 5% – 10%; etc.). Triangles indicate conditions where the logistic regression is less accurate; relative error is increased ≤ 5%.

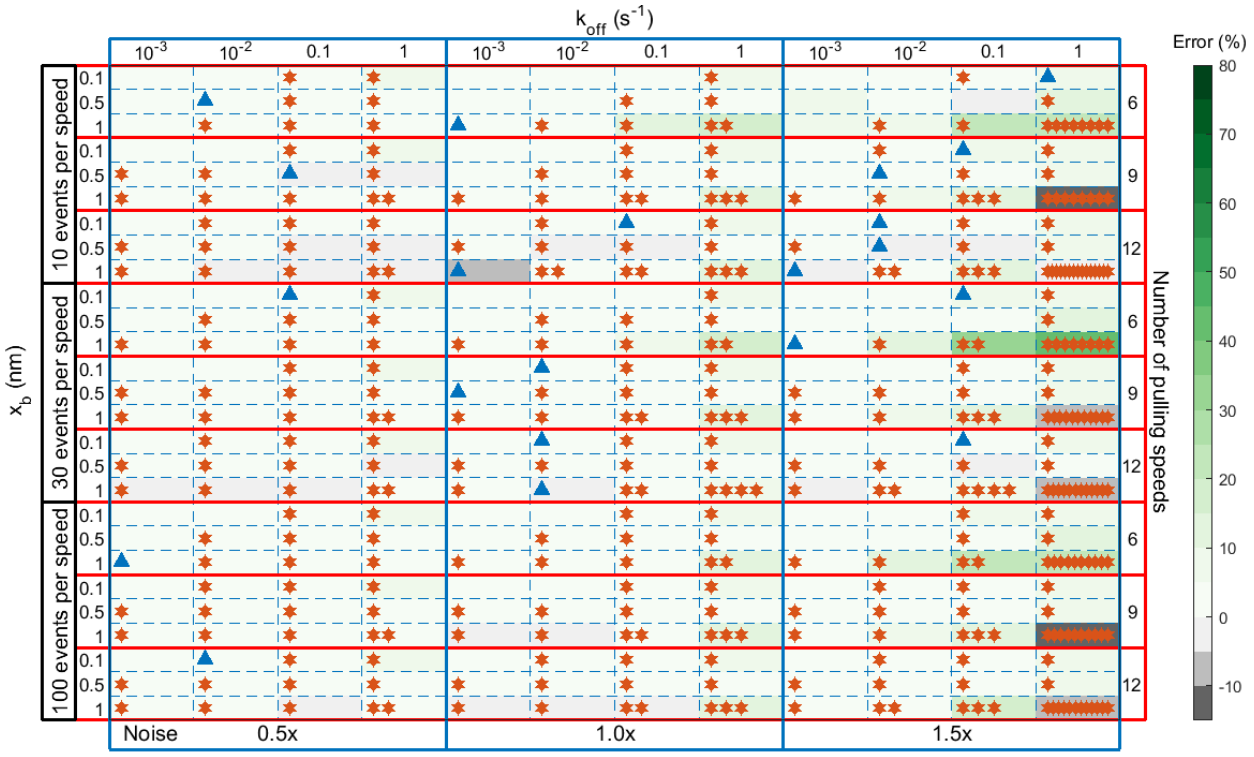

Figure S6. Relative error in estimated x_β using K-means cluster analysis. The effect of K-means clustering was validated by varying parameters across a realistic range of values. DFS data was simulated using four values of k_{off} (10^{-3}, 10^{-2}, 0.1 and 1 sec^{-1}) in combination with three values of x_β (0.1, 0.5 and 1.0 nm), three numbers of pulling speeds (6, 9 and 12), and three amounts of data (10, 30 and 100 events per pulling speed) under three levels of thermal noise (0.5×, 1× and 1.5×). For each condition, the relative error of x_β was compared with the standard pulling speed analysis. Stars indicate conditions where K-means is more accurate (0 stars: no improvement; 1 star: relative error is reduced ≤ 5%; 2 stars: relative error is reduced by 5% – 10%; etc.). Triangles indicate conditions where the K-means is less accurate; relative error is increased ≤ 5%.

Table S1. Relative errors of k_{off}, x_β with 0.5× thermal noise* Unit: %

Simulation conditions		Kinetic parameters		Methods of analysis							
Events per velocity	Number of velocities	True Value		Pulling speed		GMM		Logistic regression		K-means	
		x_β (nm)	k_{off} (s^{-1})	x_β	k_{off}	x_β	k_{off}	x_β	k_{off}	x_β	k_{off}
10	6	0.1	10^{-3}	2	30	2	30	2	30	2	30
			10^{-2}	3	12	3	12	3	12	3	12
			10^{-1}	2	18	2	18	2	18	2	21
			1	7	-15	7	-13	7	-15	7	-13
		0.5	10^{-3}	2	-4	2	-4	2	-4	2	-4
			10^{-2}	2	0	2	0	2	0	2	0
			10^{-1}	4	-25	4	-27	4	-27	3	-27
			1	3	-13	3	-11	2	-8	3	-16
		1.0	10^{-3}	2	-13	2	-13	2	-13	2	-13
			10^{-2}	2	-5	2	-6	2	-5	2	0
			10^{-1}	5	-31	5	-31	5	-30	4	-8
			1	7	-21	6	-17	6	-7	3	2
	9	0.1	10^{-3}	2	24	2	24	2	24	2	24
			10^{-2}	2	31	2	31	2	31	2	31
			10^{-1}	3	17	3	17	3	17	3	18
			1	7	-8	7	-8	7	-6	6	-3
		0.5	10^{-3}	0	7	0	7	0	7	0	6
			10^{-2}	2	-6	2	-5	2	-6	1	0
			10^{-1}	0	9	0	4	0	10	-1	19
			1	2	-12	1	4	1	4	0	9
		1.0	10^{-3}	3	-20	1	-6	2	-17	1	-9
			10^{-2}	4	-39	5	-47	3	-28	1	-12
			10^{-1}	4	-17	3	-11	1	16	2	9
			1	8	-27	4	10	2	14	2	9
	12	0.1	10^{-3}	2	27	2	27	2	27	2	27
			10^{-2}	2	21	2	21	2	22	2	31
			10^{-1}	2	18	2	21	2	21	2	22
			1	6	-8	6	-7	5	-4	5	1
		0.5	10^{-3}	4	-35	4	-32	4	-32	4	-29
			10^{-2}	2	-16	2	-5	1	-1	1	-3
			10^{-1}	2	-8	1	0	0	10	0	10
			1	4	-16	1	1	-1	10	0	6
		1.0	10^{-3}	2	-18	0	-1	1	1	0	12
			10^{-2}	2	-6	0	14	-1	34	-1	38
			10^{-1}	4	-12	0	28	0	37	-3	53
			1	7	-21	-1	42	0	35	-1	31

Events per velocity	Number of velocities	True Value		Pulling speed		GMM		Logistic regression		K-means	
		x_β (nm)	k_{off} (s^{-1})	x_β	k_{off}	x_β	k_{off}	x_β	k_{off}	x_β	k_{off}
30	6	0.1	10^{-3}	1	34	1	34	1	34	1	34
			10^{-2}	1	30	1	30	1	30	1	30
			10^{-1}	3	16	3	16	3	16	3	15
			1	7	-9	6	-7	6	-7	6	-6
		0.5	10^{-3}	2	-16	2	-16	2	-16	2	-16
			10^{-2}	3	-14	3	-14	3	-14	3	-16
			10^{-1}	2	-5	2	-4	2	-4	2	-2
			1	2	-7	2	-6	2	-4	1	3
		1.0	10^{-3}	4	-37	4	-37	4	-37	4	-37
			10^{-2}	3	-16	3	-16	3	-18	3	-13
			10^{-1}	5	-24	5	-28	5	-24	4	-13
			1	8	-32	7	-22	6	-15	5	-12
	9	0.1	10^{-3}	1	36	1	36	1	36	1	36
			10^{-2}	3	9	3	9	3	9	3	9
			10^{-1}	3	12	3	12	3	13	3	15
			1	7	-10	7	-14	6	-8	6	-8
		0.5	10^{-3}	3	-16	3	-16	3	-16	2	-17
			10^{-2}	2	-13	2	-13	2	-13	1	-5
			10^{-1}	2	-11	2	-11	2	-7	1	-1
			1	4	-18	3	-15	2	-5	1	-2
		1.0	10^{-3}	3	-30	3	-30	3	-25	2	-13
			10^{-2}	3	-18	2	-12	2	-9	1	9
			10^{-1}	4	-15	3	-1	1	14	0	18
			1	8	-32	6	-17	2	9	1	10
	12	0.1	10^{-3}	1	40	1	40	1	40	1	40
			10^{-2}	3	16	3	16	3	16	3	17
			10^{-1}	3	17	3	16	3	17	2	23
			1	6	-7	7	-16	6	-6	5	-7
		0.5	10^{-3}	2	-13	2	-12	2	-12	2	-7
			10^{-2}	2	-9	1	-7	1	-1	0	6
			10^{-1}	2	-9	2	-11	1	1	0	2
			1	2	-4	0	-3	-1	13	-1	13
		1.0	10^{-3}	3	-21	2	-10	1	-4	0	14
			10^{-2}	2	-13	1	-5	0	15	-1	21
			10^{-1}	4	-19	2	-7	-1	30	-1	36
			1	7	-25	0	36	0	28	1	28

Events per velocity	Number of velocities	True Value		Pulling speed		GMM		Logistic regression		K-means	
		x_β (nm)	k_{off} (s^{-1})	x_β	k_{off}	x_β	k_{off}	x_β	k_{off}	x_β	k_{off}
100	6	0.1	10^{-3}	2	27	2	27	2	27	2	27
			10^{-2}	2	21	2	21	2	21	2	21
			10^{-1}	3	15	3	15	3	15	3	15
			1	6	-9	7	-11	6	-9	6	-7
		0.5	10^{-3}	2	-11	2	-11	2	-11	2	-11
			10^{-2}	2	-13	2	-13	2	-13	2	-11
			10^{-1}	3	-12	3	-13	3	-12	2	-8
			1	4	-14	3	-12	3	-12	2	-7
		1.0	10^{-3}	3	-27	3	-25	3	-27	3	-24
			10^{-2}	4	-29	4	-28	4	-28	4	-25
			10^{-1}	6	-34	6	-33	6	-28	5	-20
			1	7	-25	7	-21	4	-7	3	-1
	9	0.1	10^{-3}	1	40	1	40	1	40	1	40
			10^{-2}	2	30	2	30	2	30	2	30
			10^{-1}	3	11	3	10	3	12	3	14
			1	6	-6	12	-49	5	-3	5	-2
		0.5	10^{-3}	2	-7	2	-7	2	-7	2	-6
			10^{-2}	2	-9	2	-8	2	-7	1	0
			10^{-1}	3	-14	3	-14	2	-10	1	-2
			1	4	-15	4	-18	3	-16	2	-9
		1.0	10^{-3}	3	-20	3	-18	2	-14	1	-1
			10^{-2}	3	-21	3	-18	2	-8	1	0
			10^{-1}	5	-25	4	-15	2	3	0	16
			1	8	-28	4	-9	2	14	2	14
	12	0.1	10^{-3}	1	32	1	32	1	32	1	32
			10^{-2}	2	27	2	27	2	27	2	28
			10^{-1}	3	16	4	9	3	17	2	19
			1	6	-7	10	-42	6	-11	6	-9
		0.5	10^{-3}	2	-12	2	-12	2	-11	1	-6
			10^{-2}	2	-13	2	-14	2	-10	1	-4
			10^{-1}	2	-13	2	-12	2	-9	1	1
			1	3	-13	6	-36	2	-8	2	-12
		1.0	10^{-3}	3	-26	3	-19	1	-2	0	7
			10^{-2}	4	-31	2	-10	1	2	0	10
			10^{-1}	5	-24	2	0	0	25	-1	28
			1	8	-28	1	24	1	22	1	24

* Thermal noise distributed normally about zero with standard deviations of 6.4 pN in force and 0.16 nm in distance.

Table S2. Relative errors of k_{off}, x_β with 1× thermal noise* Unit: %

Simulation conditions		Kinetic parameters		Methods of analysis							
Events per velocity	Number of velocities	True Value		Pulling speed		GMM		Logistic regression		K-means	
		x_β (nm)	k_{off} (s^{-1})	x_β	k_{off}	x_β	k_{off}	x_β	k_{off}	x_β	k_{off}
10	6	0.1	10^{-3}	2	-11	2	-11	2	-11	2	-11
			10^{-2}	4	-15	4	-15	4	-15	4	-15
			10^{-1}	4	-7	4	-7	4	-7	4	-5
			1	9	-26	8	-26	8	-26	8	-26
		0.5	10^{-3}	1	-25	1	-25	1	-25	1	-25
			10^{-2}	3	-41	3	-41	3	-41	3	-41
			10^{-1}	2	-25	1	-24	1	-24	1	-23
			1	4	-36	4	-34	4	-34	4	-32
		1.0	10^{-3}	4	-44	4	-44	4	-44	4	-46
			10^{-2}	3	-35	3	-35	3	-35	3	-28
			10^{-1}	15	-77	15	-75	13	-70	12	-71
			1	27	-91	26	-90	22	-87	19	-83
	9	0.1	10^{-3}	3	1	3	1	3	1	3	1
			10^{-2}	3	-6	3	-6	3	-6	3	-6
			10^{-1}	5	-9	4	-9	4	-8	4	-11
			1	8	-29	8	-29	7	-27	7	-26
		0.5	10^{-3}	1	-16	1	-18	1	-16	1	-16
			10^{-2}	4	-41	3	-38	2	-38	3	-36
			10^{-1}	4	-31	4	-35	3	-33	2	-32
			1	5	-33	3	-28	1	-16	2	-17
		1.0	10^{-3}	4	-43	3	-39	3	-39	1	-24
			10^{-2}	8	-57	8	-61	7	-51	4	-41
			10^{-1}	7	-50	6	-51	4	-24	1	-9
			1	26	-89	15	-70	13	-71	13	-73
	12	0.1	10^{-3}	0	33	0	33	0	33	0	33
			10^{-2}	2	8	2	8	2	8	2	9
			10^{-1}	3	-3	3	-3	3	-3	3	-2
			1	7	-28	8	-30	7	-25	6	-21
		0.5	10^{-3}	1	-24	1	-21	1	-21	1	-17
			10^{-2}	2	-29	1	-20	2	-24	0	-9
			10^{-1}	2	-27	0	-10	0	-12	0	-12
			1	5	-38	1	-20	0	-17	0	-18
		1.0	10^{-3}	-2	37	-2	12	-4	57	-5	70
			10^{-2}	6	-53	7	-61	3	-46	1	-25
			10^{-1}	10	-64	3	-29	1	-29	1	-26
			1	24	-88	11	-69	15	-73	12	-71

Events per velocity	Number of velocities	True Value		Pulling speed		GMM		Logistic regression		K-means	
		x_β (nm)	k_{off} (s^{-1})	x_β	k_{off}	x_β	k_{off}	x_β	k_{off}	x_β	k_{off}
30	6	0.1	10^{-3}	1	25	1	25	1	25	1	25
			10^{-2}	4	-9	4	-9	4	-9	4	-9
			10^{-1}	4	-7	4	-7	4	-7	4	-7
			1	8	-28	7	-25	7	-26	7	-23
		0.5	10^{-3}	2	-24	2	-24	2	-24	2	-24
			10^{-2}	2	-20	2	-20	2	-20	2	-19
			10^{-1}	4	-34	4	-34	4	-34	3	-29
			1	5	-37	4	-34	4	-35	2	-27
		1.0	10^{-3}	4	-38	4	-38	4	-38	4	-38
			10^{-2}	3	-30	4	-32	3	-31	2	-25
			10^{-1}	8	-53	8	-55	8	-46	6	-43
			1	22	-86	20	-84	17	-80	16	-76
	9	0.1	10^{-3}	2	14	2	14	2	14	2	14
			10^{-2}	3	-1	3	-1	3	-1	3	-1
			10^{-1}	4	-2	3	-1	3	-1	3	0
			1	8	-27	7	-26	7	-25	7	-22
		0.5	10^{-3}	1	-13	1	-13	1	-13	1	-18
			10^{-2}	1	-23	2	-23	1	-23	1	-18
			10^{-1}	3	-33	2	-26	2	-26	1	-16
			1	5	-38	4	-36	3	-30	2	-26
		1.0	10^{-3}	3	-44	2	-40	3	-39	2	-34
			10^{-2}	5	-44	3	-35	3	-32	2	-19
			10^{-1}	8	-54	5	-36	3	-24	1	-11
			1	26	-89	19	-81	15	-76	14	-72
	12	0.1	10^{-3}	1	18	1	18	1	18	1	18
			10^{-2}	4	-3	4	-3	4	-3	4	-3
			10^{-1}	4	-4	4	-3	4	-2	4	-4
			1	8	-28	8	-33	6	-25	6	-23
		0.5	10^{-3}	2	-31	2	-27	2	-27	2	-25
			10^{-2}	2	-29	2	-27	2	-28	1	-21
			10^{-1}	2	-26	3	-33	1	-20	1	-16
			1	5	-42	2	-23	0	-15	0	-15
		1.0	10^{-3}	4	-42	1	-19	1	-24	0	-3
			10^{-2}	2	-24	1	-22	-2	16	-3	33
			10^{-1}	9	-63	4	-42	3	-27	0	-8
			1	26	-90	9	-59	12	-68	9	-60

Events per velocity	Number of velocities	True Value		Pulling speed		GMM		Logistic regression		K-means	
		x_β (nm)	k_{off} (s^{-1})	x_β	k_{off}	x_β	k_{off}	x_β	k_{off}	x_β	k_{off}
100	6	0.1	10^{-3}	2	15	2	15	2	15	2	15
			10^{-2}	3	3	3	3	3	3	3	3
			10^{-1}	4	-7	4	-7	4	-7	4	-5
			1	8	-29	8	-28	8	-28	7	-26
		0.5	10^{-3}	2	-21	2	-21	2	-21	2	-21
			10^{-2}	2	-21	2	-21	2	-21	1	-18
			10^{-1}	2	-26	2	-24	2	-26	2	-25
			1	6	-42	5	-38	5	-39	4	-34
		1.0	10^{-3}	4	-41	4	-41	4	-41	3	-37
			10^{-2}	5	-50	5	-49	5	-48	4	-45
			10^{-1}	7	-54	7	-52	6	-48	4	-36
			1	21	-84	21	-85	16	-74	14	-71
	9	0.1	10^{-3}	2	17	2	17	2	17	2	17
			10^{-2}	2	7	2	7	2	7	2	6
			10^{-1}	5	-10	5	-11	4	-10	4	-9
			1	8	-27	8	-28	7	-24	6	-23
		0.5	10^{-3}	2	-26	2	-25	2	-26	2	-23
			10^{-2}	2	-26	2	-23	2	-23	1	-19
			10^{-1}	2	-25	2	-20	1	-21	1	-12
			1	5	-39	4	-34	2	-27	2	-25
		1.0	10^{-3}	2	-29	1	-21	1	-16	0	0
			10^{-2}	2	-25	1	-19	1	-11	-1	10
			10^{-1}	8	-58	7	-51	4	-37	2	-27
			1	23	-86	22	-87	12	-64	10	-61
	12	0.1	10^{-3}	2	13	2	13	2	13	2	13
			10^{-2}	3	6	3	6	3	6	2	6
			10^{-1}	4	-10	4	-9	4	-7	4	-5
			1	8	-28	10	-48	8	-30	7	-27
		0.5	10^{-3}	2	-26	2	-25	2	-22	1	-18
			10^{-2}	2	-31	2	-31	2	-30	1	-24
			10^{-1}	2	-26	2	-21	1	-20	0	-14
			1	5	-37	2	-29	0	-15	0	-15
		1.0	10^{-3}	2	-36	2	-23	1	-9	-1	0
			10^{-2}	4	-35	1	-15	-1	7	-1	9
			10^{-1}	8	-54	6	-43	1	-12	0	-8
			1	23	-86	8	-52	11	-64	11	-64

* Thermal noise distributed normally about zero with standard deviations of 12.8 pN in force and 0.32 nm in distance.

Table S3. Relative errors of k_{off}, x_β with 1.5× thermal noise* Unit: %

Simulation conditions		Kinetic parameters		Methods of analysis							
Events per velocity	Number of velocities	True Value		Pulling speed		GMM		Logistic regression		K-means	
		x_β (nm)	k_{off} (s^{-1})	x_β	k_{off}	x_β	k_{off}	x_β	k_{off}	x_β	k_{off}
10	6	0.1	10^{-3}	3	-15	3	-15	3	-15	3	-15
			10^{-2}	3	-13	3	-13	3	-13	3	-13
			10^{-1}	4	-27	4	-27	4	-25	4	-27
			1	8	-38	8	-38	8	-40	8	-36
		0.5	10^{-3}	7	-58	7	-58	7	-58	7	-58
			10^{-2}	0	-38	0	-38	0	-38	0	-38
			10^{-1}	-1	-21	-1	-21	-1	-17	-1	-21
			1	14	-72	13	-71	12	-71	12	-70
		1.0	10^{-3}	6	-56	6	-56	5	-52	6	-52
			10^{-2}	10	-72	10	-75	9	-78	10	-72
			10^{-1}	25	-96	24	-95	26	-96	23	-94
			1	61	-100	41	-98	38	-99	22	-91
	9	0.1	10^{-3}	2	8	2	8	2	8	2	8
			10^{-2}	3	-13	3	-13	3	-13	3	-9
			10^{-1}	6	-29	6	-30	6	-30	6	-28
			1	10	-45	10	-44	9	-42	9	-43
		0.5	10^{-3}	1	-20	1	-20	1	-20	1	-17
			10^{-2}	1	-25	1	-23	0	-21	1	-28
			10^{-1}	3	-40	2	-36	3	-36	2	-29
			1	12	-70	10	-63	9	-64	8	-64
		1.0	10^{-3}	4	-47	2	-31	5	-46	1	-22
			10^{-2}	12	-85	10	-82	7	-72	9	-83
			10^{-1}	23	-96	15	-87	12	-87	10	-81
			1	49	-100	17	-77	-1	-51	-11	49
	12	0.1	10^{-3}	2	3	2	3	2	3	2	3
			10^{-2}	2	5	2	1	2	1	2	-3
			10^{-1}	5	-23	5	-23	5	-23	5	-23
			1	8	-43	8	-42	8	-38	8	-37
		0.5	10^{-3}	3	-57	4	-55	4	-58	3	-49
			10^{-2}	-3	0	-2	-3	-2	-5	-3	8
			10^{-1}	3	-39	3	-36	1	-29	0	-22
			1	10	-71	7	-63	6	-64	6	-59
		1.0	10^{-3}	1	-40	0	-12	-1	9	-3	40
			10^{-2}	9	-75	3	-59	3	-54	2	-46
			10^{-1}	28	-98	14	-84	15	-87	15	-89
			1	75	-100	1	-50	5	-44	-3	-46

Events per velocity	Number of velocities	True Value		Pulling speed		GMM		Logistic regression		K-means	
		x_β (nm)	k_{off} (s^{-1})	x_β	k_{off}	x_β	k_{off}	x_β	k_{off}	x_β	k_{off}
30	6	0.1	10^{-3}	2	1	2	1	2	1	2	1
			10^{-2}	3	-8	3	-8	3	-8	3	-8
			10^{-1}	6	-29	6	-29	6	-29	6	-29
			1	9	-40	9	-38	9	-38	9	-38
		0.5	10^{-3}	3	-42	3	-42	3	-42	3	-42
			10^{-2}	3	-42	3	-41	3	-42	3	-41
			10^{-1}	3	-43	3	-43	3	-43	3	-37
			1	12	-74	12	-73	12	-73	11	-71
		1.0	10^{-3}	1	-37	1	-36	1	-34	1	-36
			10^{-2}	12	-80	11	-81	12	-77	10	-76
			10^{-1}	38	-99	36	-99	33	-99	31	-98
			1	81	-100	70	-100	46	-99	43	-99
	9	0.1	10^{-3}	3	-5	3	-5	3	-5	3	-5
			10^{-2}	4	-17	4	-17	4	-17	4	-17
			10^{-1}	5	-23	5	-20	5	-20	5	-20
			1	8	-38	8	-37	8	-35	7	-35
		0.5	10^{-3}	0	-21	0	-21	0	-21	0	-19
			10^{-2}	2	-37	1	-27	2	-37	1	-26
			10^{-1}	4	-46	3	-40	3	-42	2	-36
			1	11	-73	9	-69	9	-68	8	-64
		1.0	10^{-3}	6	-67	6	-67	5	-59	3	-46
			10^{-2}	10	-81	10	-83	9	-77	7	-71
			10^{-1}	27	-97	22	-92	18	-92	13	-88
			1	60	-100	0	-21	-7	68	-7	-5
	12	0.1	10^{-3}	2	7	2	7	2	7	2	7
			10^{-2}	4	-18	4	-18	4	-17	4	-17
			10^{-1}	4	-20	5	-21	4	-19	4	-17
			1	9	-41	9	-42	8	-37	7	-35
		0.5	10^{-3}	1	-34	1	-31	1	-30	1	-29
			10^{-2}	1	-29	1	-27	1	-26	0	-21
			10^{-1}	2	-36	1	-32	0	-28	-1	-21
			1	10	-70	6	-60	5	-58	5	-53
		1.0	10^{-3}	4	-53	1	-36	0	-17	0	-25
			10^{-2}	9	-75	10	-80	4	-58	2	-43
			10^{-1}	27	-97	18	-91	12	-84	10	-74
			1	65	-100	-2	-42	-3	-12	-7	30

Events per velocity	Number of velocities	True Value		Pulling speed		GMM		Logistic regression		K-means	
		x_β (nm)	k_{off} (s^{-1})	x_β	k_{off}	x_β	k_{off}	x_β	k_{off}	x_β	k_{off}
100	6	0.1	10^{-3}	3	-3	3	-3	3	-3	3	-3
			10^{-2}	3	-8	3	-8	3	-8	3	-8
			10^{-1}	5	-25	5	-25	5	-25	5	-24
			1	9	-40	9	-40	9	-41	8	-39
		0.5	10^{-3}	3	-44	3	-44	3	-44	3	-44
			10^{-2}	3	-40	3	-39	3	-40	3	-41
			10^{-1}	4	-46	4	-46	4	-45	4	-44
			1	12	-71	11	-70	11	-70	11	-70
		1.0	10^{-3}	4	-51	4	-51	3	-50	3	-46
			10^{-2}	13	-86	12	-85	12	-83	11	-82
			10^{-1}	31	-98	30	-98	27	-97	25	-96
			1	69	-100	60	-100	26	-95	21	-93
	9	0.1	10^{-3}	3	-1	3	-1	3	-1	3	-1
			10^{-2}	3	-12	3	-12	3	-12	3	-12
			10^{-1}	5	-25	5	-25	5	-25	5	-23
			1	9	-39	8	-39	8	-37	8	-36
		0.5	10^{-3}	2	-38	2	-38	2	-38	2	-36
			10^{-2}	2	-32	1	-27	2	-30	1	-25
			10^{-1}	3	-45	2	-39	2	-41	2	-37
			1	11	-70	10	-67	9	-66	8	-63
		1.0	10^{-3}	6	-64	5	-58	4	-55	2	-41
			10^{-2}	11	-81	10	-80	9	-76	7	-68
			10^{-1}	27	-97	27	-97	15	-89	14	-87
			1	61	-100	12	-83	-6	28	-14	143
	12	0.1	10^{-3}	3	-8	3	-8	3	-8	3	-8
			10^{-2}	3	-11	3	-11	3	-11	3	-9
			10^{-1}	5	-23	5	-23	5	-22	4	-19
			1	9	-41	9	-41	8	-39	8	-36
		0.5	10^{-3}	1	-28	1	-28	1	-26	0	-21
			10^{-2}	2	-35	1	-30	1	-29	0	-21
			10^{-1}	3	-39	2	-32	1	-27	1	-28
			1	10	-69	8	-63	7	-60	6	-56
		1.0	10^{-3}	4	-56	3	-41	1	-27	0	-20
			10^{-2}	10	-78	8	-72	5	-61	4	-56
			10^{-1}	30	-98	24	-96	19	-92	19	-92
			1	65	-100	-9	29	-2	-22	-8	21

* Thermal noise distributed normally about zero with standard deviations of 19.2 pN in force and 0.48 nm in distance.

CHAPTER 4. FLUORESCENCE AXIAL LOCALIZATION WITH NANOMETER ACCURACY AND PRECISION

This chapter is published in *Nano Letters*†

Hui Li‡, Chi-Fu Yen‡ and Sanjeevi Sivasankar

4.1 Abstract

We describe a new technique, standing wave axial nanometry (SWAN), to image the axial location of a single nanoscale fluorescent object with sub-nanometer accuracy and 3.7 nm precision. A standing wave, generated by positioning an atomic force microscope tip over a focused laser beam, is used to excite fluorescence; axial position is determined from the phase of the emission intensity. We use SWAN to measure the orientation of single DNA molecules of different lengths, grafted on surfaces with different functionalities.

4.2 Introduction

Fluorescence imaging of nanoscale biological assemblies rely on localizing molecules with high accuracy and measuring distances between them with high resolution. However, the resolution of conventional fluorescence microscopes is limited by the diffraction of light: with a high numerical aperture objective and visible excitation, resolution is about 200 nm in the lateral direction and 500 nm along the optical axis.

A single fluorescent molecule can be localized with nanometer accuracy along the *x*- and *y*-axis by determining the centroid of its point spread functions (PSF) [1], a

---

† H. Li, C.F. Yen, S. Sivasankar, Fluorescence Axial Localization with Nanometer Accuracy and Precision, Nano Lett. 12 (2012) 3731-3735.

‡ These authors contributed equally

known as fluorescence imaging with one nanometer accuracy (FIONA) [2]. This approach has also been used to resolve the lateral separation between two dyes of the same or of different colors within a diffraction-limited spot [3-5]. FIONA can been combined with the stochastic switching of single molecule fluorescence to obtain high-resolution images of microscopic biological objects such as cells, an approach alternatively known as stochastic optical reconstruction microscopy (STORM) [6], photoactivated localization microscopy (PALM) [7], and fluorescence photoactivated localization microscopy (FPALM) [8]. The lateral resolution of fluorescence imaging can also be improved by using stimulated emission depletion (STED) to narrow the effective width of the PSF [9, 10].

Unlike imaging in the *x*- and *y*-direction, improving resolution and single molecule localization accuracy along the optical axis is more challenging [11]. In STORM experiments, the *z*-position of a single fluorophore can be determined with 50 nm resolution using a cylindrical lens to distort the shape of the PSF [12]; resolution can be further improved to 20 nm by sandwiching the sample between two opposing objectives [13]. Better resolution, down to 10 nm, can be achieved using interferometry as demonstrated in interferometric photoactivated localization microscopy (iPALM) and 4Pi-single marker switching microscopy (4Pi-SMS) [14, 15]; this however requires the use of custom optics in a complicated layout for interference detection. Alternatively, fluorescence interference-contrast microscopy (FLIC) has been used to determine the height of dye monolayers with nanometer accuracy [16]; these experiments however require multiple replicas of the sample deposited on patterned silicon oxide surfaces which limits its applicability in single molecule biological imaging. Fluorescence

interference has also been used to monitor the movement of single motor proteins on microtubules [17].

Here, we describe a new localization method with an axial resolution superior to previous techniques to determine the *z*-position of a single nanoscale fluorescent object with sub-nanometer accuracy. Our technique, called standing wave axial nanometry (SWAN), utilizes a commercial atomic force microscope (AFM) mounted on a single molecule confocal microscope [18]. A standing wave excitation pattern is generated by positioning the AFM tip over a focused laser beam. A fluorescent object is positioned within the standing wave and its fluorescence phase difference is used to measure the molecule's axial location.

As a proof of principle, we use SWAN to measure the orientation of single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) of different lengths, grafted on surfaces with different functionalities. The conformation of immobilized DNA affects surface hybridization efficiency [19] and critically impacts techniques such as DNA microarrays and gene sequencing. Previously, ensemble methods such as ellipsometry [20], surface plasmon resonance [21], and fluorescence self-interference [22] have been used to characterize the orientation of DNA bound to surfaces at high densities, where steric effects from neighboring molecules influence conformation. However, the orientation of single, tethered DNA has not been measured. Using SWAN, we show that dsDNA of different lengths, grafted using polymer tethers, are oriented at an average tilt of 30° with respect to the surface. On the other hand, dsDNAs adsorbed nonspecifically to a positively charged surface lie flat. Finally, a single-stranded G-quadruplex sequence, folds into a structure where its 5′ and 3′ ends are adjacent to each other.

4.3 Principle of SWAN

In SWAN, an AFM tip, which serves as a mirror, is positioned over the focused laser beam of a confocal microscope (Supporting Information, Figure S1). Interference between the incident beam and the beam reflected off the AFM tip, results in a standing wave between the tip and surface (Figure 1a). As the tip is translated in the z-direction, the fluorescence emitted by a fluorophore positioned within the standing wave oscillates with a phase that corresponds to its distance from the surface. Axial location of the fluorophore can be measured from the phase difference relative to a fiduciary marker (Figure 1a).

As the AFM tip moves by a distance *Z* along the optical axis, the phase difference between the incident and reflected beams at a height Z_{mol} above the surface is given by

$$\emptyset = \frac{2\pi \times 2 \times (Z + Z_0 - Z_{mol})}{\frac{\lambda_{laser}}{n}} \qquad (1)$$

where Z_0 is the initial position of the AFM tip, λ_{laser} is the wavelength of the excitation laser and *n* is the refractive index of the medium. The excitation intensity of the incident beam along the optical axis corresponds to the shape of the PSF and can be described by a power law [23]. Ignoring higher order terms, the excitation intensity at Z_{mol} due to interference of the incident and reflected beams is given by

$$I = A^2 + \frac{B^2}{1 + CZ^2} + 2A\frac{B}{\sqrt{1 + CZ^2}} cos(\emptyset) \qquad (2)$$

where *A* and *B* are the amplitudes of the incident and reflected beams and *C* is the fitted amplitude decay that is related to the focal depth of the microscope [23]. Since the emitted fluorescence intensity is proportional to the excitation intensity, the height of a fluorescent object located at Z_{mol}, can be determined by fitting its fluorescence emission to eq 2.

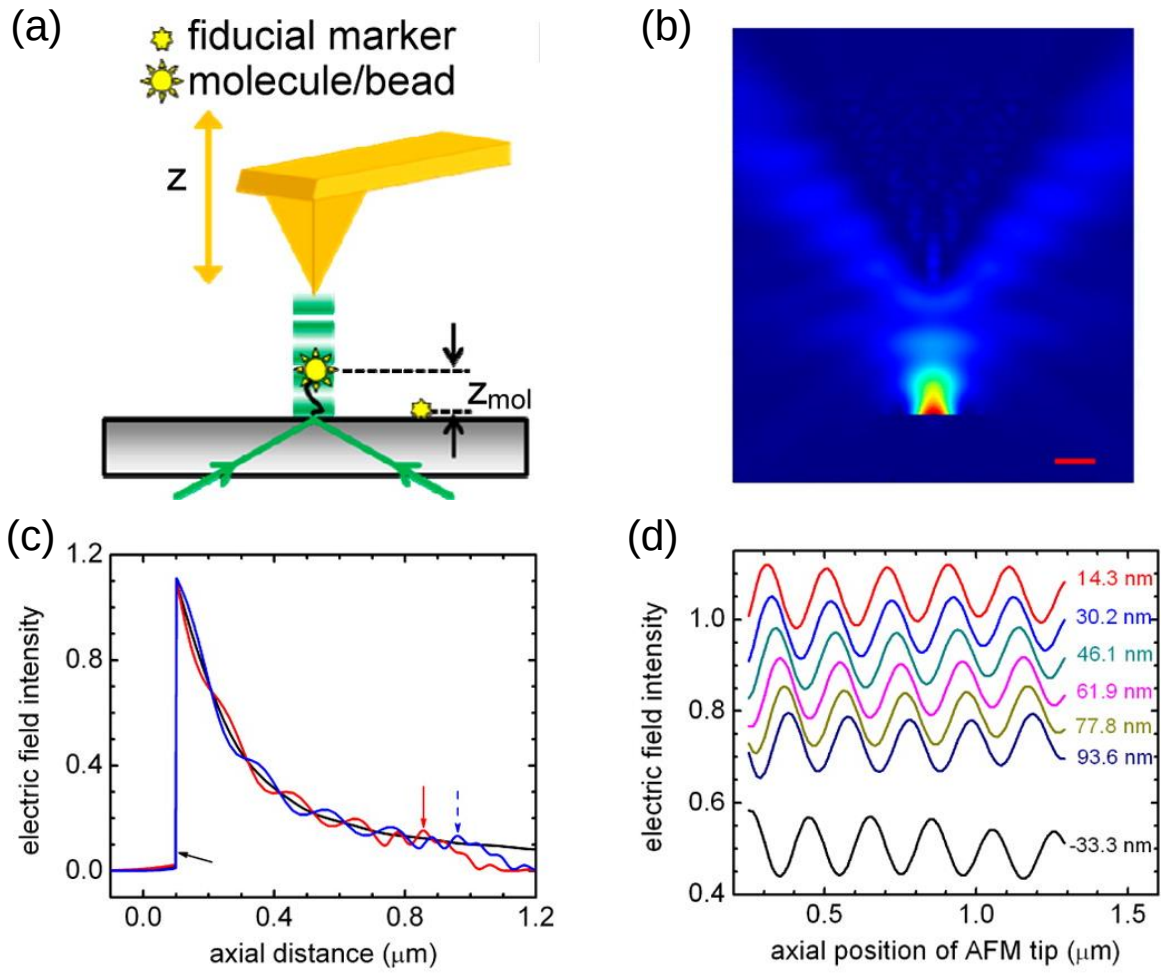

Figure 1. Schematic and validation of SWAN experiment. (a) Schematic of experiment. A standing wave, generated by positioning an atomic force microscope tip over a focused laser beam, is used to excite a nanoscale fluorescent object. Its axial position is determined from the phase difference of the fluorescence emission relative to a fiducial marker. **(b)** FDTD simulations show periodic oscillations in the electrical field intensity in the 0.75 μm gap between a silicon tip and coverslip. Scale bar = 200 nm. **(c)** Electric field intensity along the axial direction under three conditions: black, without tip; red, with tip 0.75 μm away from the surface (indicated by the red arrow); and blue, with tip 0.85 μm away from the surface (indicated by the blue arrow). Without a tip, the intensity along the optical axis decays rapidly due to the shape of the PSF. When an AFM tip is present, periodic oscillations in the electric field intensity are measured. Black arrow corresponds to location of coverslip surface which is also the position of laser focus. **(d)** Electric field intensities, monitored at six positions along the optical axis, as the AFM tip moves from 250 nm to 1.3 μm. The intensities oscillate with different phases. The black curve shows the intensity below the coverslip (magnified 10 times for clarity). Oscillations in the black curve, which result from the interference of light reflected from the coverslip and AFM tip, are used to determine the coverslip position.

We used finite difference time domain (FDTD) simulations to calculate the electric field intensity distribution when the AFM tip was located over a focused excitation laser (Supporting Information). The simulations show periodic oscillations in the electric field in the gap between the AFM tip and coverslip, corresponding to a standing wave generated due to the interference between the incident excitation and the beam reflected off the tip (Figure 1b). The electric field intensity along the optical axis is plotted in Figure 1c. Without a tip, the intensity decays rapidly in the axial direction corresponding to the shape of the PSF. When the AFM tip is present, periodic oscillations of the electric field are superimposed on this decay (Figure 1c). By changing the distance between the tip and surface, the phase of the oscillation can be modulated: for instance, the phase is reversed when a tip 0.75 μm from the coverslip is relocated 0.85 μm away from the surface (Figure 1c). We also monitored the electric field intensity at six positions along the optical axis while the tip was translated from 250 nm to 1.3 μm in 13 nm steps (Figure 1d). As the distance from the surface increased, the average intensity decreases. However, the field at the six positions oscillate with different phases (Figure 1d).

In order to excite a fluorophore using the standing wave, the fluorophore and AFM tip were aligned along the microscope's optical axis (Supporting Information, Figure S2). To ensure that the tip does not contact the molecule during the experiment, the AFM tip was held at a minimum Z_0 of 200 nm above the surface. The tip was translated in the axial direction from 200 to 1200 nm, while monitoring fluorescence emitted by the molecule. We determined the absolute tip-coverslip distance by measuring the interference of the excitation laser reflected off the coverslip and AFM tip and also by

comparing the fluorescence phase difference of the dye to fiduciary markers bound to the coverslip (Supporting Information, Figure S2).

4.4 Results

4.4.1 SWAN calibration

We determined the accuracy of SWAN by measuring the radius of polystyrene nanospheres of different sizes, uniformly loaded with fluorescent dyes (Supporting Information). Because each nanosphere contains a large number of dyes, SWAN measures the average height of all the fluorophores within a single nanosphere, that is, the nanosphere radius. We measured fluorescent nanospheres with manufacturer specified diameters of 50 nm (Bead A) and 100 nm (Bead B) relative to fiducial beads (Bead F) with a diameter of 25 nm (Figure 2a). We also verified the size of the nanospheres using tapping-mode AFM imaging; measured diameters were consistent with the manufacturer specified values. Average bead diameters (mean ± standard error of mean) of 25.2 ± 0.4 nm, 49.0 ± 0.6 nm, and 99.2 ± 1.3 nm were measured for Beads F (97 beads), Beads A (90 beads), and Beads B (85 beads), respectively (Supporting Information, Figure S3). Histograms of the measured diameters were fit to a Gaussian function to estimate bead polydispersity; fits to Beads F, Beads A, and Beads B had a full width at half maxima (FWHM) of 7.5, 10.3, and 18.0 nm, respectively (Supporting Information, Figure S3).

Typical SWAN data traces along with their fits to eq 2 are shown in Figure 2b. Relative to the fiducial beads, we expected to measure a height of 12.5 nm for Bead A

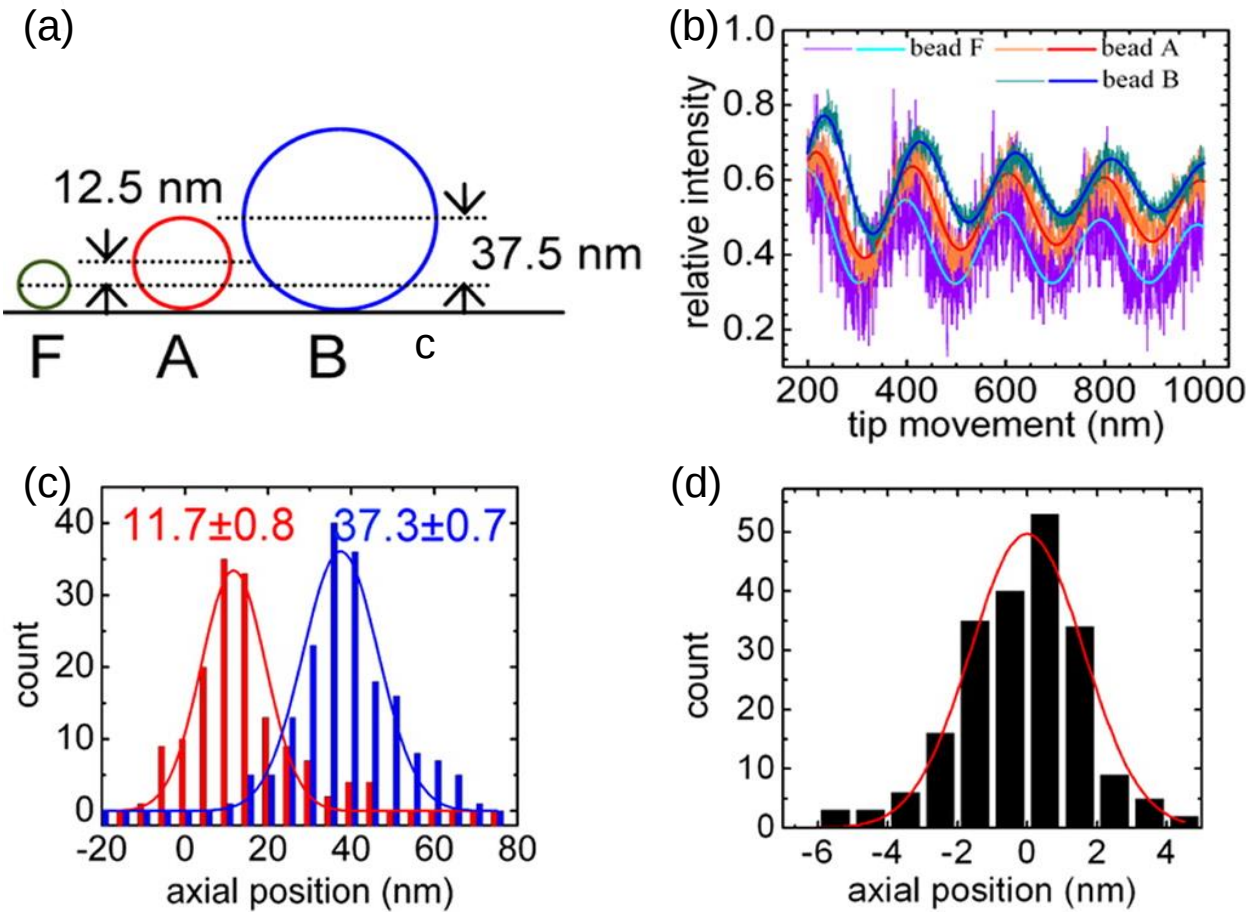

Figure 2. Calibration of SWAN using fluorescent nanospheres. (a) Schematic of experiment. The relative radii of 25 nm (Bead A) and 50 nm (Bead B) fluorescent nanospheres were measured relative to 12.5 nm (Bead F) fiducial beads. We expect to measure a relative radii of 12.5 and 37.5 nm for Beads A and B, respectively. **(b)** Typical experimental data traces (thin lines) along with their fits (thick lines) to eq 2. **(c)** Histograms of the measured radii were normally distributed; the peak of a Gaussian fit corresponded to the mean radii while the standard error of mean corresponded to the error. The measured relative radii for Beads A and Beads B were 11.7 ± 0.8 and 37.3 ± 0.7 nm, respectively. **(d)** The axial position of a single Bead F was measured 100 times; the measured values had a standard deviation of 1.6 nm, which corresponds to a localization precision (FWHM) of 3.7 nm.

and 37.5 nm for Bead B (Figure 2a). In close agreement, we measured relative radii (mean ± standard error of mean) of 11.7 ± 0.8 nm for Beads A (181 beads) and 37.3 ± 0.7 nm for Beads B (178 beads) relative to Beads F (181 beads) (Figure 2c). This demonstrates that the accuracy of SWAN is less than 1 nm. It should be noted that increasing the sample size (i.e., number of nanospheres) will not increase the accuracy of the measurement. The standard deviations of the measured radii were 8.1 and 7.2 nm, respectively; we ascribe this in part to the polydispersity in the measured bead sizes.

To quantify the resolution of SWAN, we measured the height of a single Bead F 100 times. The fluorescence intensity of the bead varied between 50 and 100 kHz and photons were collected for 1 s. To increase the precision of our measurement, we reduced instrumental drift by simultaneously recording the fluorescence intensity and excitation laser interference rather than measuring them sequentially. The measured bead radius had a standard deviation of 1.6 nm (Figure 2d), which corresponds to an axial resolution of 3.7 nm (FWHM). This resolution is almost three times better than interferometric techniques [14, 15] and more than five times better than optical astigmatism [13].

4.4.2 Measurement of single DNA orientation

Next, we used SWAN to measure the orientation of ssDNA and dsDNA molecules, labeled with a Cy3 fluorophore at one end and a biotin anchor at the other end. The DNA was tethered on a flexible polyethylene glycol (PEG) cushion decorated with nonfluorescent streptavidin (Figure 3a, Supporting Information). Average DNA orientation was measured relative to fluorescent Alexa Fluor 555 streptavidin that served as a fiduciary marker (Figure 3a).

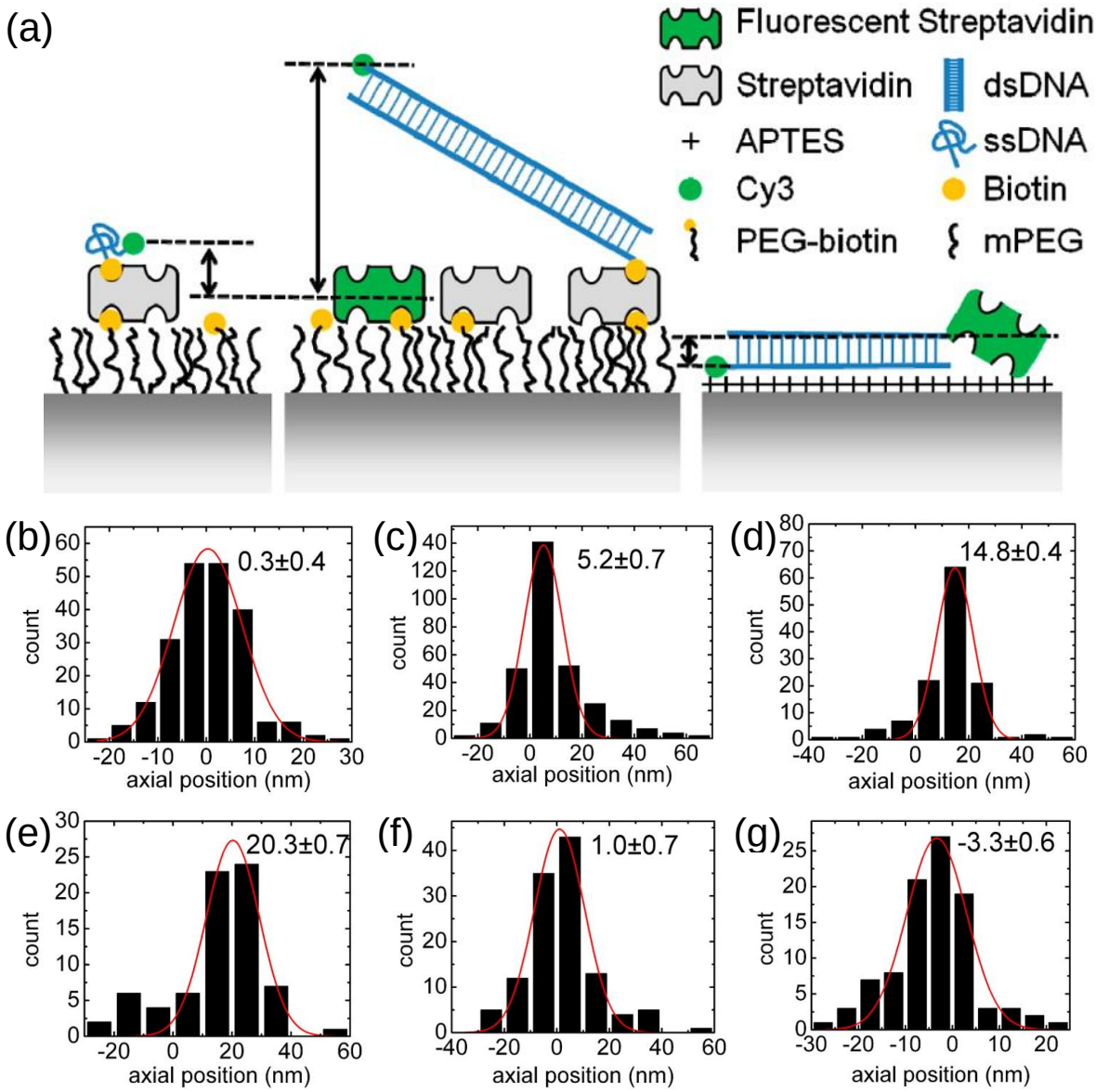

Figure 3. Orientation of immobilized DNA molecules measured using SWAN. (a) Schematic of the different DNA immobilization strategies used. ssDNA G-quadruplex (left panel) and dsDNA (middle panel) was tethered on a PEG cushion decorated with nonfluorescent streptavidin, via a biotin group. dsDNA (right panel) was also nonspecifically adsorbed to a self-assembled monolayer of positively charged silanes. Fluorescently labeled streptavidin served as a fiducial marker in all the experiments. **(b)** The measured height of 35 base ssDNA G-quadruplex sequence corresponds to a structure where the 5′ and 3′ ends are adjacent to each other. Histogram of measured heights for **(c)** 35 bp dsDNA, **(d)** 90 bp dsDNA and **(e)** 122 bp dsDNA anchored to surface via a flexible PEG linker. The measured heights correspond to tilt angles of 26°, 29°, and 34°, respectively. Histogram of measured heights for **(f)** 35 bp dsDNA and **(g)** 122 bp dsDNA nonspecifically adsorbed to a silane monolayer. The measured heights indicate that the DNA lies flat on the surface. Errors correspond to standard error of mean.

The ssDNA was a G-quadruplex sequence that folds into a four-stranded structure [24] in the presence of Na^+; when folded, the Cy3 is located at the same height as fluorescent streptavidin (Figure 3a). In the presence of Na^+, we therefore anticipate measuring a height of 0 nm for the Cy3 relative to the fiducial. Indeed, in 50 mM Na^+, we measured a height of 0.3 ± 0.4 nm (FWHM, 16.6 nm) for the ssDNA (Figure 3b).

We also measured the orientation of 3 dsDNA sequences with lengths of 35 base pairs (bp), 90 bp, and 122 bp. Since the contour lengths of the 35 bp and 90 bp fragments are much smaller than the 50 nm persistence length of dsDNA, these sequences are expected to behave as stiff rods. On the other hand, the 122 bp DNA fragment has a contour length of 41.5 nm that is comparable to the dsDNA persistence length; the end-to-end distance of this sequence was estimated to be 36.5 nm, 12% shorter than the contour length according to wormlike chain model [25]. It is anticipated that the tethered dsDNA freely rotates in the space above streptavidin [22]. Calculations show that under these conditions dsDNA is oriented at an average tilt of 30° with respect to the surface (Supporting Information, Figure S4). Using SWAN we measured a height of 5.2 ± 0.7 nm (FWHM, 17.2 nm), 14.8 ± 0.4 nm (FWHM, 15.5 nm), and 20.3 ± 0.7 nm (FWHM, 18.1 nm) for the three sequences, corresponding to tilt angles of 26°, 29°, and 34°, respectively (Figure 3c–e).

A widely used approach to lay DNA flat on a surface is to adsorb it nonspecifically to a self-assembled monolayer of positively charged aminopropyltriethoxysilane (APTES) (Figure 3a) [26]. We adsorbed 35 bp dsDNA, 122 bp dsDNA, and fluorescently labeled streptavidin on a freshly prepared APTES surface and used SWAN to measure the difference in height between the dsDNA and fluorescently labeled streptavidin which

served as a fiduciary marker. We measured a height of 1.0 ± 0.7 nm (FWHM, 22.6 nm) for 35 bp dsDNA (Figure 3f) and −3.3 ± 0.6 nm (FWHM, 15.0 nm) for 122 bp dsDNA (Figure 3g), indicating that both dsDNAs are lying flat on the surface.

4.5 Discussion and Conclusions

In summary, SWAN has an axial resolution that exceeds all previously described single molecule techniques and can be used to localize the *z*-position of single fluorescent objects with nanometer accuracy and precision. While a total of 5×10^4 to 10×10^4 photons were used in the nanosphere experiments to quantify the precision of SWAN, it is important to note that this count corresponds to the total number of photons measured over the 1 s duration of the experiment. This photon budget is merely a factor of 5 to 10 greater than the 8×10^3 to 2×10^4 photons collected in each experiment with Cy3 labeled DNA. To estimate the resolution of SWAN using single Cy3 fluorophores, we reduced the total photon budget in Figure 2d to either 4×10^4 photons, 2×10^4photons, or 8×10^3 photons by analyzing every 2nd, 4th, or 10th data point. The bead radii determined using this method had a FWHM of 4.3, 5.8, and 6.0 nm, respectively. A future study will experimentally determine the dependency of SWAN precision on photon budget.

We used SWAN to show that dsDNA, grafted at low densities using polymer tethers, are oriented at an average tilt of 30° with respect to the surface, which agrees with calculated tilt angles for freely rotating stiff rods. It is important to note that because of the long measurement time (approximately 1 s per molecule) relative to the time scale for molecular rearrangements, we measure only the average tilt angle adopted by the dsDNA and not its instantaneous orientation. Previous ensemble measurements using

densely grafted DNA, measured average tilt angles of 50° and 40° for 21 bp and 50 bp dsDNA [22]. It is likely that the tilt value measured using SWAN is smaller because the low grafting density of dsDNA reduces steric constraints and permits the DNA tether to sample a larger range of conformations. Using SWAN, we also show that dsDNA adsorbed to a positively charged surface lays flat. Finally, a single-stranded G-quadruplex sequence, folds into a structure where its 5′ and 3′ ends are adjacent to each other.

Errors in SWAN primarily arise from imprecise localization of the AFM tip over the fluorophore, instrumental drift between successive measurements as well as dye photophysics. To quantify these errors, we compared two different measurement strategies. In the first approach, we aligned the AFM tip over a single nanosphere (Bead F) and then measured its radius 100 times by simultaneously recording fluorescence oscillations and excitation laser interference. This measurement yielded a standard deviation in bead radius of 1.6 nm (Figure 2d). In the second approach, we performed 100 measurements where we sequentially aligned the AFM tip over the nanosphere, measured fluorescence oscillations, and then measured excitation laser interference. The measurement precision using this strategy was lower than the first approach;, the standard deviation in bead radius was now 5.4 nm (Supporting Information, Figure S5). We believe that the first strategy reduces instrumental drift by minimizing the time between successive measurements.

It is important to note that other interferometric methods assume that fluorescence interference occurs only at the peak wavelength of the fluorescence spectra [14, 15, 17]. Due to the broad emission spectra of fluorescent dyes, this approximation introduces systematic errors in axial localization. In contrast to these methods, periodic fluorescence

oscillations measured in SWAN occur due to standing wave excitation of the fluorophore and not due to fluorescence interference. This improves the localization accuracy of SWAN. Although a small fraction of the fluorescence emitted by a dye will reflect off the AFM tip and interfere with the directly emitted fluorescence, this effect is minimal due to the small radius of curvature of the AFM tip (~10 nm) and the large separation between the tip and surface (≥200 nm). To confirm this, we measured the oscillation period for the fluorescence and excitation laser (Supporting Information, Figure S6). Both oscillations have the same period ($\lambda_{laser}/2n$) confirming that the measured fluorescence oscillations are only due to standing wave excitation.

Unlike other interference-based approaches, SWAN does not require custom optics or specially engineered substrates. This makes the technique easy to use and well suited for biological samples such as lipid membranes, microfibers, protein complexes, and live cells. Moreover, unlike most interference-based techniques, where the interference pattern repeats itself and limits the working range to ~250 nm, the standing wave in SWAN decays with tip–surface distance, which allows successive periods to be distinguished and extends the working distance. SWAN can be easily integrated with other super-resolution and super-accuracy techniques to obtain nanometer resolution along both the lateral and axial directions. Finally, a unique advantage of SWAN is that it can be used to determine the axial position of molecules in single molecule AFM force measurements and in single molecule cut and paste applications for the bottom-up assembly of biological nanostructures [27].

4.6 Acknowledgments

The work was supported in part by an award from the Grow Iowa Values Fund. We thank Agilent Technologies for the loan of a 5500 AFM and for technical support. H.L., C.-F.Y. and S.S. designed the experiments. S.S. directed the research. H.L. and C.-F.Y. built the instrument and collected the data. H.L. did the FDTD simulations. H.L. and C.-F.Y. analyzed the data. S.S., H.L., and C.-F.Y. wrote the manuscript.

4.7 Author Contributions

H.L., C.-F.Y. and S.S. designed the experiments. S.S. directed the research. H.L. and C.-F.Y. built the instrument and collected the data. H.L. did the FDTD simulations. H.L. and C.-F.Y. analyzed the data. S.S., H.L. and C.-F.Y. wrote the manuscript.

4.8 References

[1] R.E. Thompson, D.R. Larson, W.W. Webb, Precise nanometer localization analysis for individual fluorescent probes, Biophys. J. 82 (2002) 2775-2783.

[2] A. Yildiz, J.N. Forkey, S.A. McKinney, T. Ha, Y.E. Goldman, P.R. Selvin, Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization, Science 300 (2003) 2061-2065.

[3] L.S. Churchman, Z. Okten, R.S. Rock, J.F. Dawson, J.A. Spudich, Single molecule high-resolution colocalization of Cy3 and Cy5 attached to macromolecules measures intramolecular distances through time, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 1419-1423.

[4] M.P. Gordon, T. Ha, P.R. Selvin, Single-molecule high-resolution imaging with photobleaching, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 6462-6465.

[5] X. Qu, D. Wu, L. Mets, N.F. Scherer, Nanometer-localized multiple single-molecule fluorescence microscopy, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 11298-11303.

[6] M.J. Rust, M. Bates, X. Zhuang, Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM), Nat. Methods 3 (2006) 793-795.

[7] E. Betzig, G.H. Patterson, R. Sougrat, O.W. Lindwasser, S. Olenych, J.S. Bonifacino, M.W. Davidson, J. Lippincott-Schwartz, H.F. Hess, Imaging intracellular fluorescent proteins at nanometer resolution, Science 313 (2006) 1642-1645.

[8] S.T. Hess, T.P.K. Girirajan, M.D. Mason, Ultra-high resolution imaging by fluorescence photoactivation localization microscopy, Biophys. J. 91 (2006) 4258-4272.

[9] T.A. Klar, S.W. Hell, Subdiffraction resolution in far-field fluorescence microscopy, Optics Letters 24 (1999) 954-956.

[10] S.W. Hell, J. Wichmann, Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy, Optics Letters 19 (1994) 780-782.

[11] B. Huang, An in-depth view, Nat. Methods 8 (2011) 304-305.

[12] B. Huang, W. Wang, M. Bates, X. Zhuang, Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy, Science 319 (2008) 810-813.

[13] K. Xu, H.P. Babcock, X. Zhuang, Dual-objective STORM reveals three-dimensional filament organization in the actin cytoskeleton, Nat. Methods 9 (2012) 185-188.

[14] G. Shtengel, J.A. Galbraith, C.G. Galbraith, J. Lippincott-Schwartz, J.M. Gillette, S. Manley, R. Sougrat, C.M. Waterman, P. Kanchanawong, M.W. Davidson, R.D. Fetter, H.F. Hess, Interferometric fluorescent super-resolution microscopy resolves 3D cellular ultrastructure, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 3125-3130.

[15] D. Aquino, A. Schönle, C. Geisler, C.V. Middendorff, C.A. Wurm, Y. Okamura, T. Lang, S.W. Hell, A. Egner, Two-color nanoscopy of three-dimensional volumes by 4Pi detection of stochastically switched fluorophores, Nat. Methods 8 (2011) 353-359.

[16] D. Braun, P. Fromherz, Fluorescence interference-contrast microscopy of cell adhesion on oxidized silicon, Applied Physics A: Materials Science & Processing 65 (1997) 341-348.

[17] J. Kerssemakers, J. Howard, H. Hess, S. Diez, The distance that kinesin-1 holds its cargo from the microtubule surface measured by fluorescence interference contrast microscopy, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 15812-15817.

[18] S. Sivasankar, S. Chu, Nanoparticle-mediated nonfluorescent bonding of microspheres to atomic force microscope cantilevers and imaging fluorescence

from bonded cantilevers with single molecule sensitivity, Nano Lett. 9 (2009) 2120-2124.

[19] E. Ozkumur, J.W. Needham, D.A. Bergstein, R. Gonzalez, M. Cabodi, J.M. Gershoni, B.B. Goldberg, M.S. Unlü, Label-free and dynamic detection of biomolecular interactions for high-throughput microarray applications, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 7988-7992.

[20] D.E. Gray, S.C. Case-Green, T.S. Fell, P.J. Dobson, E.M. Southern, Ellipsometric and interferometric characterization of DNA probes immobilized on a combinatorial array, Langmuir 13 (1997) 2833-2842.

[21] L.K. Wolf, Y. Gao, R.M. Georgiadis, Sequence-dependent DNA immobilization: specific versus nonspecific contributions, Langmuir 20 (2004) 3357-3361.

[22] L. Moiseev, M.S. Unlü, A.K. Swan, B.B. Goldberg, C.R. Cantor, DNA conformation on surfaces measured by fluorescence self-interference, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 2623-2628.

[23] M.C. DeSantis, S.K. Zareh, X. Li, R.E. Blankenship, Y.M. Wang, Single-image axial localization precision analysis for individual fluorophores, Optics Express 20 (2012) 3057-3065.

[24] F.W. Smith, J. Feigon, Quadruplex structure of Oxytricha telomeric DNA oligonucleotides, Nature 356 (1992) 164-168.

[25] S.B. Smith, L. Finzi, C. Bustamante, Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads, Science 258 (1992) 1122-1126.

[26] A. Bensimon, A. Simon, A. Chiffaudel, V. Croquette, F. Heslot, D. Bensimon, Alignment and sensitive detection of DNA by a moving interface, Science 265 (1994) 2096-2098.

[27] S.K. Kufer, E.M. Puchner, H. Gumpp, T. Liedl, H.E. Gaub, Single-molecule cut-and-paste surface assembly, Science 319 (2008) 594-596.

4.9 Supporting Information

SWAN setup (Figure S1): In SWAN, a closed-loop, tip-scanning, AFM (Agilent 5500) is mounted on a standard sample scanning, confocal fluorescence microscope. An excitation laser beam (532 nm) was focused to a diffraction-limited spot on a glass cover slip using a 60x, 1.42 N.A., oil-immersion objective (Olympus). The cover slip was

functionalized with fluorescent molecules, enclosed in a fluid-cell and mounted on a manual, 2-axes translation stage and a closed loop piezoelectric stage (100 µm x 100 µm, Physik Instrumente) for coarse and fine positioning. A custom-made, low vibration adaptor was used to mate the AFM to the confocal microscope. The AFM was modified with an 880 nm infrared Super Luminescent Diode (Q-Photonics) to spectrally separate the AFM light source from Cy3 fluorescence. Fluorescence from the sample was collected by the same objective and focused onto the detection face of an Avalanche Photodiode (APD, Micro Photon Devices). A motorized flipper equipped with a band-pass filter and a ND filter was positioned in front of the objective to transmit either Cy3 fluorescence or excitation light to the APD. Alternatively, to reduce data acquisition time, a dichroic beam splitter was used to separate the fluorescence and excitation light and a second APD was used to measure interference of the laser. Custom software written in LabView was used to control the microscope and synchronize AFM tip movement with fluorescence data collection.

Data collection (Figure S2): Non-fluorescent, silicon AFM cantilevers (Nanoworld, Arrow Cantilevers, Spring Constant 2.8 N/m) were used in the SWAN experiments. First, an image of the AFM tip was obtained by scanning the tip over the focused laser beam and recording light reflected off the tip; the tip position was determined from the centroid of its image. Then, a fluorescence image of the sample was collected by moving the piezoelectric stage. The position of the beads/molecules was determined by fitting their fluorescence to a 2D Gaussian function; each bead/molecule could then be positioned at the center of the focused laser beam. Using this strategy, the center of the focused

confocal laser beam, fluorescent molecule and AFM tip was aligned along the microscope's optical axis.

After that, the AFM tip was translated in the axial direction from 200 nm to 1200 nm, while monitoring fluorescence emitted by the beads/molecule positioned under the tip. Since the AFM tip does not make contact with the surface, the absolute tip-surface distance was not known. We corrected for this using two strategies. First, we measured the z-position of the coverslip by monitoring the interference of the excitation laser reflected off the coverslip and AFM tip and used it to correct the fluorescence phase difference of the dye. Ideally, the interference data should be described by equation (2) with $Z_{mol} = 0$; however, we measured a constant offset in Z_{mol}. This is caused by the large size of the diffraction limited, focused, laser beam relative to the sharp AFM tip, variability in the shape of tip and variations in the angle at which the AFM tip is mounted; for the same tip with the same mount, the offset was always constant. Second, we compared the fluorescence phase difference of the dye to a fiduciary marker on the same sample.

Finite Difference Time Domain (FDTD) simulations: FDTD simulations were performed using a commercial software package, FDTD Solutions (Lumercial Solutions Inc.). A 3.2 μm x 3.5 μm simulation region was assigned a refractive index of 1.33 to simulate the buffer environment; perfectly matched layers were used as boundary conditions to absorb all waves incident on them with no reflections. A rectangle of 2 μm x 0.6 μm with a refractive index of 1.46 was used to mimic the coverslip. An equilateral triangle (1.1 μm x 1.1 μm x 1.1 μm) with refractive index of 4.2 was used to mimic the

silicon AFM tip. A 0.2 μm thick layer around the surface of coverslip and AFM tip was discretized with mesh size of 1.6 nm (Mesh override region) to account for near field effects, other space was discretized using an auto non-uniform mesh with accuracy of 4 (corresponding to mesh size of about 22nm). A Gaussian laser source with wavelength of 532 nm, propagating towards the tip, was positioned at the surface of the coverslip. The beam had a spot size of 0.5 μm and diverged according to a N.A. 1.4 thin lens. Time domain simulations were run, mostly more than 30,000 iterations, until an auto-shutoff criterion was reached. Electric field intensity in the frequency domain was obtained by Fourier transform of the simulated time domain fields normalized by Fourier transform of the source pulse. The tip position was changed from 250 nm to 1.3 μm in 13 nm steps (81 simulations).

Measurement of nanosphere diameter with AFM (Figure S3): Diameter of *Beads F*, *Beads A* and *Beads B* were measured in air using a tapping-mode AFM. Nanospheres were diluted, adsorbed on freshly cleaved mica and dried at 70°C for 2 hours. The samples were then imaged with a Multi-mode AFM (Veeco, CA) using Si cantilevers (Model: TESPA, Bruker). Image scan speed and feedback parameters were tuned to obtain sharp images. A first-order flattening was performed on every image. Height of beads was determined by subtracting the maximum height of every bead from the average of the surrounding background.

Estimation of the average tilt angle of dsDNA (Figure S4): Since the length of the dsDNA used in our experiments is smaller than the persistence length, we modeled

dsDNA of length *R* as a stiff rod. We assumed that the dsDNA freely rotates in the half space above streptavidin due to its flexible linker. The average height, $\bar{h}$, of the dsDNA can be estimated as:

$$\bar{h} = \frac{\int_0^{\pi/2} h * 2\pi r * Rd\theta}{\int_0^{\pi/2} 2\pi r * Rd\theta} = \frac{\int_0^{\pi/2} R\sin(\theta) * 2\pi R\cos(\theta) * Rd\theta}{\int_0^{\pi/2} 2\pi R\cos(\theta) * Rd\theta}$$

$$= \frac{2\pi R^3 \int_0^{\pi/2} \sin(\theta)\cos(\theta)d\theta}{2\pi R^2 \int_0^{\pi/2} \cos(\theta)d\theta} = R * \frac{\frac{1}{2}\sin^2(\theta)\Big|_0^{\pi/2}}{\sin(\theta)\Big|_0^{\pi/2}} = R/2$$

So the average tilt angle is $\bar{\theta} = \arcsin(\bar{h}/R) = 30^\circ$

Preparation of fluorescent nanosphere sample: Fluorescent nanospheres (Thermo Scientific) with radii of 12.5 nm (*Beads F*), 25 nm (*Beads A*) and 50 nm (*Beads B*) were adsorbed onto freshly cleaned glass coverslips using 100mM $CaCl_2$. The nanospheres were pre-mixed and diluted to yield a surface density of approximately 1 bead/10 μm^2.

Preparation of DNA sample: Single strand DNA (35 bases) and three double strand DNA sequences (35 bp, 90 bp, 122 bp) were used to demonstrate single molecule axial localization by SWAN. The DNA molecules (Integrated DNA Technologies) were labeled with a single Cy3 fluorophore at one end and a biotin anchor at the other end. The DNA sequences used are listed below

35 bases ssDNA:

Cy3-5'-TGG GGT TTT GGG GTT TTG GGG TTT TGG GGA GAT GG-3'-biotin

35 bp dsDNA:

Strand 1: Cy3-5'-TGG GGT TTT GGG GTT TTG GGG TTT TGG GGA GAT GG-3'-biotin

Strand 2: 5'-CCA TCT CCC CAA AAC CCC AAA ACC CCA AAA CCC CA-3'

90 bp dsDNA:

Strand 1: 5'-CCC AGT TGA GCT GTG AGA ACC CCC TGT GCT TCA GGT TAT AAG ATT CCT CTA GGT AAA GTT GCG CCA CGG ACA ACA TCC GAT AGA ACG GCC-3'-biotin

Strand 2: 5'-GGC CGT TCT ATC GGA TGT TGT CCG TGG CGC AAC TTT ACC TAG AGG AAT CTT ATA ACC TGA AGC ACA GGG GGT TCT-3' ***and*** 5'-CAC AGC TCA ACT GGG-3'-Cy3

122 bp dsDNA:

Strand 1: 5'-CCC AGT TGA GCT GTG AGA ACC CCC TGT GCT TCA GGT TAT AAG ATT CCT CTA GGT AAA GTT GCG CCA CGG ACA ACA TCC GAT AGA ACG GCC GTC CAA CTG GCG TCA GGT ACA CCT CGC CAC CC-3'

Strand 2: biotin-5'-GGG TGG CGA GGT GTA CCT GAC GCC AGT TGG ACG GCC GTT

CTA TCG GAT GTT GTC CGT GGC GCA ACT TTA CCT AGA GGA ATC TTA TAA CCT

GAA GCA CAG GGG GTT CT-3' ***and*** 5'-CAC AGC TCA ACT GGG-3'-Cy3

The DNA was either tethered to a Polyethylene Glycol (PEG) polymer cushion (Laysan Bio Inc) decorated with non-fluorescent streptavidin (Sigma-Aldrich) or was non-specifically adsorbed to a self-assembled monolayer of positively charged APTES (Sigma-Aldrich). To immobilize the DNA, glass coverslips were first cleaned by heating in a piranha solution (25%H_2O_2 : 75% H_2SO_4) at 60 ^{0}C for 30 min and then rinsed with deionized water. The coverslips were subsequently sonicated in 1M KOH for 15 min, rinsed with deionized water, and dried using filtered, compressed air. The coverslips were finally incubated in 2% solution of APTES dissolved in Acetone for 30 min, rinsed with Acetone and deionized water, and dried using filtered air. The dsDNA was non-specifically adsorbed to the APTES monolayer. The APTES surface was also incubated with 20 pM fluorescent streptavidin (Invitrogen), for 30 min; the fluorescent streptavidin non-specifically bound to the surface and served as fiduciary markers. Since fluorescent streptavidin has 2 to 3 Alexa Fluor® 555 dyes per protein, it is bright and can be easily distinguished from the DNA molecules.

To tether DNA on PEG linkers, the silanized coverslips were first functionalized with PEG (MW 5000) containing an amine-reactive *N*-hydroxysuccinimide ester at one end. The silanized coverslips were incubated with 100 mg/ml of a 1:9 stoichiometric mixture of Biotin-PEG and m-PEG, dissolved in $NaHCO_3$ buffer (pH 8.0) for 3 hours. Then coverslips were then rinsed with deionized water, dried with filtered compressed air and stored in vacuum. Before an experiment, the coverslip was incubated with 20 pM of fluorescent streptavidin fiduciary markers for 10 min; the surface density of fluorescent streptavidin was ~ 1 molecule/10 μm^2. The biotinylated surface was then sequentially

incubated with 10 nM of non-fluorescent streptavidin and 20-100 pM DNA. All data was collected in a buffer containing 100mM Tris (pH 7.5), 2mM EDTA and 50 mM NaCl.

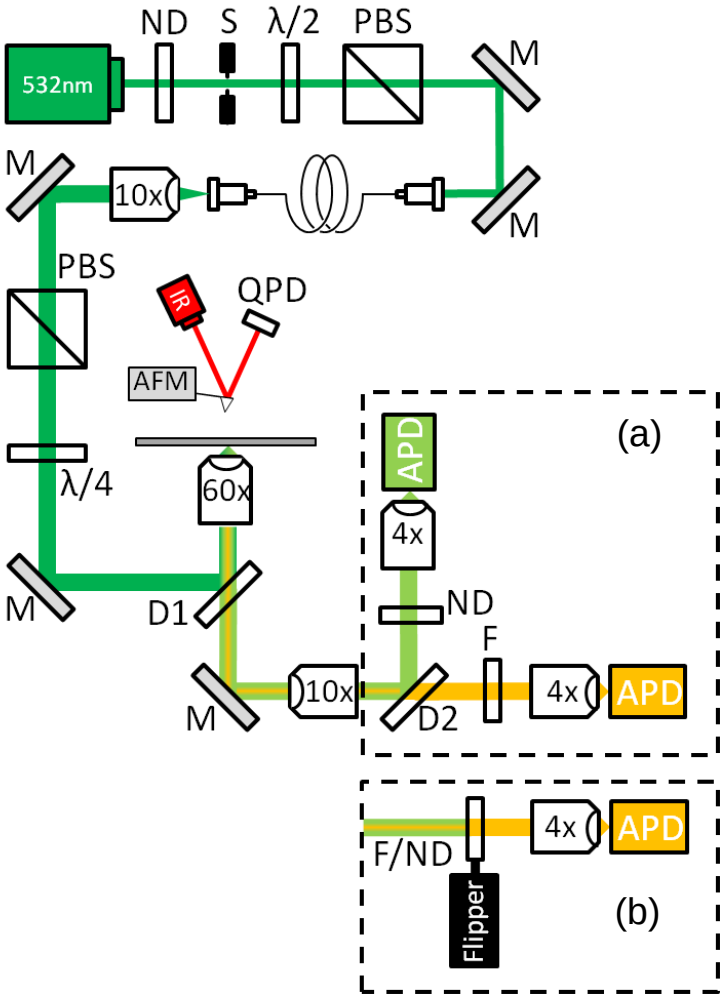

Figure S1. Schematic of SWAN microscope. A closed-loop, tip-scanning, AFM is mounted on a home-built, sample scanning, confocal fluorescence microscope. Abbreviations used in the figure: ND, neutral density filter; S, shutter; λ/2, ½ wave plate; PBS, polarizing cube beam splitter; M, mirror; D, dichroic beam splitter; λ/4, ¼ wave plate; F, band pass filter; QPD, Quadrant Photodiode; APD, Avalanche Photodiode. ***Inset:*** Two alternate detection schemes were used. **(a)** For simultaneous measurements of fluorescence intensity and excitation laser interference, two APDs were used. A dichroic beamsplitter separated the excitation laser from the fluorescence emission. **(b)** For sequential measurements of fluorescence intensity and excitation laser interference, a motorized flipper equipped with a band-pass filter and a ND filter was used to transmit either Cy3 fluorescence or excitation light to the APD.

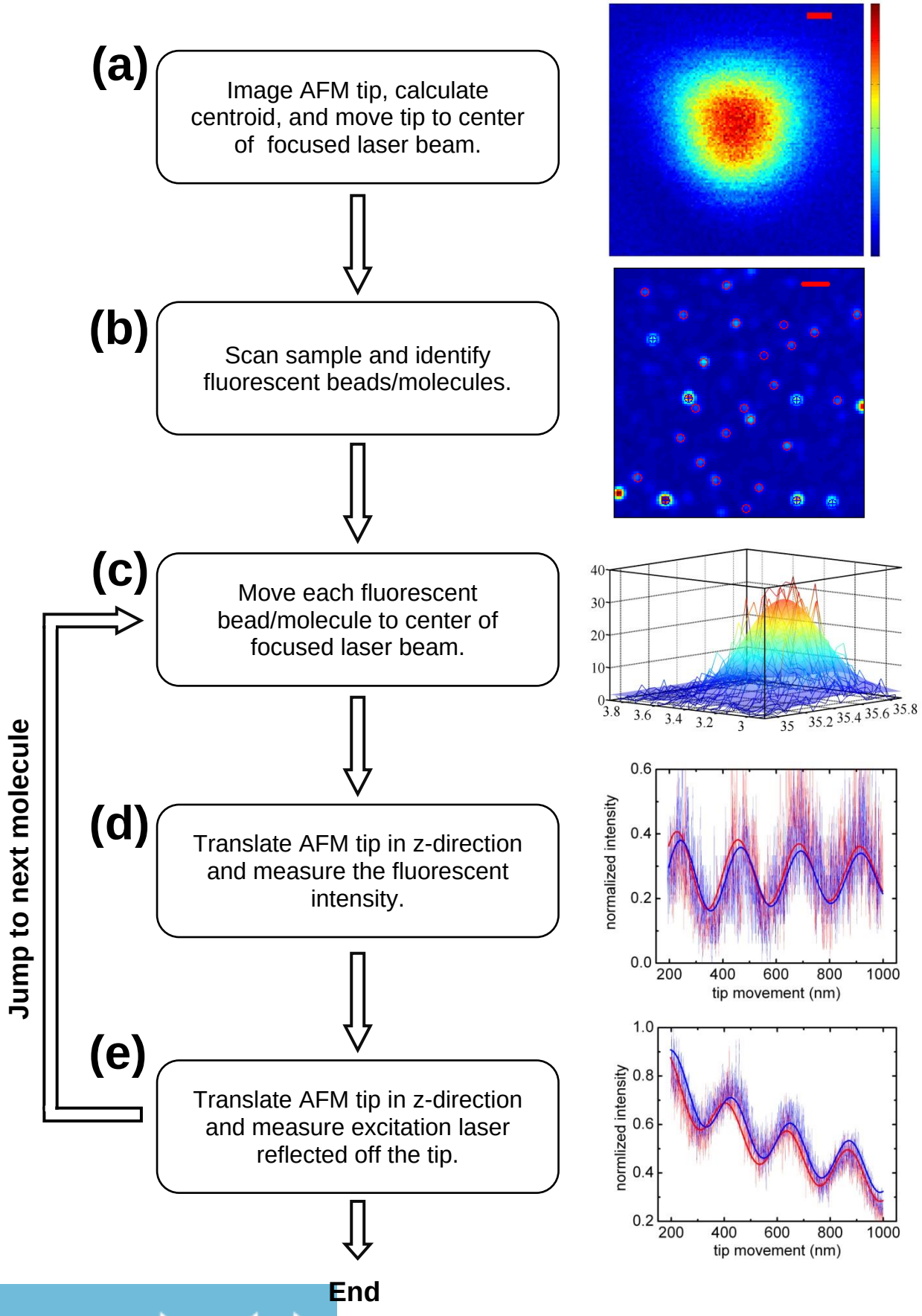

Figure S2. Steps in typical SWAN data collection. (a) An image of the AFM tip, held at a constant distance of 500nm above the surface, was obtained by using the APD to measure the light reflected off the AFM tip as it was scanned over the confocal laser beam. The tip position was determined from the centroid of its image. The tip was moved to the center of the focused laser beam and then withdrawn 7 μm away from the surface. *Right Panel:* Image of AFM tip. The image is not a perfect circle due to the pyramidal shape of the tip. Scale bar = 100 nm. **(b)** The sample containing immobilized fluorescent beads/molecules was scanned across a 10 μm x 10 μm region by moving the piezoelectric stage and the fluorescence at each pixel of the scan was collected using the APD. *Right Panel:* Typical image of DNA (red circle) and fluorescent streptavidin (blue cross), immobilized on surface. Scale bar = 1 μm. **(c)** The location of beads/molecules was calculated by fitting their fluorescence to a 2D Gaussian function; each bead/molecule was then sequentially positioned at the center of the focused laser beam. *Right Panel:* Image of a single fluorescent molecule fitted to a 2D Gaussian Function. **(d)** The AFM tip was now translated in the axial direction, while monitoring fluorescence emitted by the beads/molecule positioned under the tip. *Right Panel:* Fluorescence intensity during tip approach (red) or retraction (blue) from surface and their corresponding fits to equation (2) (thick lines). **(e)** To determine the absolute tip-surface distance, the z-position of the coverslip was measured by monitoring the interference of the excitation laser reflected off the coverslip and AFM tip. This was used to correct the fluorescence phase difference of the dye. *Right Panel:* Laser reflection intensity during tip approach (red) or retraction (blue) from surface and corresponding fits to equation (2) (thick lines).

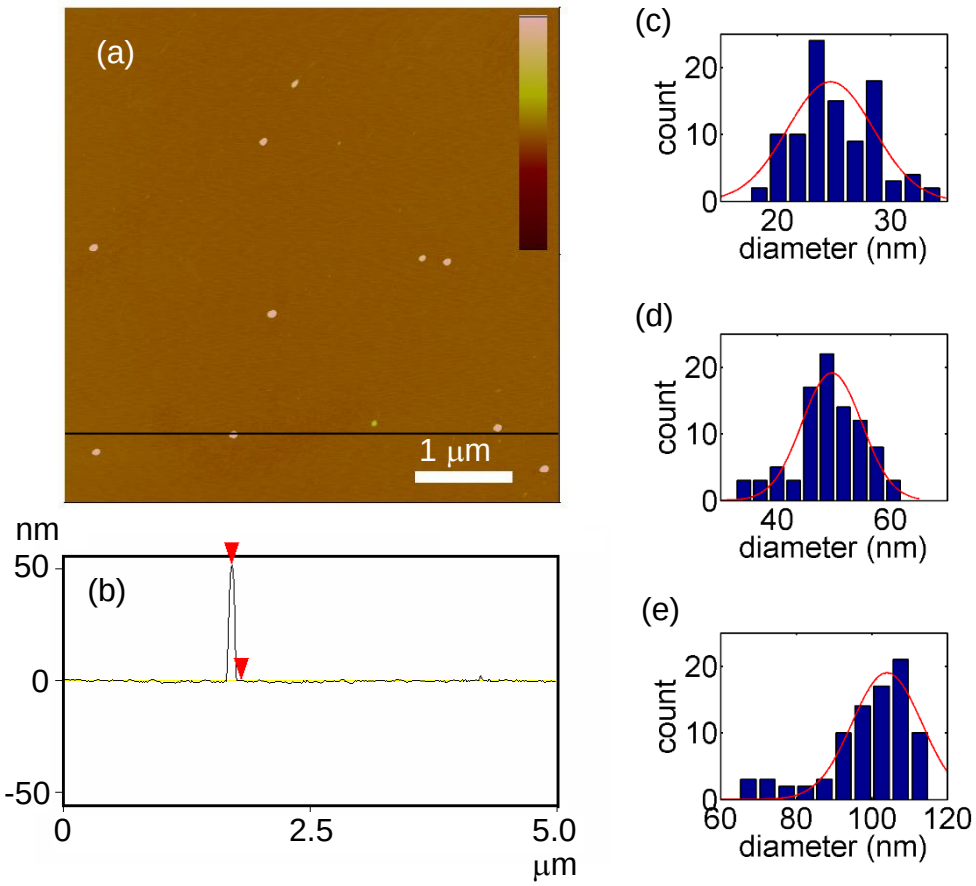

Figure S3. AFM measurement of Nanosphere diameter. (a) Typical AFM topography image of a *Bead A* sample. The color bar represents height from 0 to 50 nm. **(b)** Topography along the line in (a) shows height (diameter) of a bead. **(c)** Histogram of diameter of 97 *Beads F*. Gaussian fit has a peak value of 24.7 nm and FWHM of 7.5 nm. Arithmetic average of the nanosphere diameter (mean ± SEM) is 25.2 ± 0.4. **(d)** Histogram of diameter of 90 *Beads A.* Gaussian fit has a peak value of 49.6 nm and FWHM of 10.3 nm. Arithmetic average of the nanosphere diameter (mean ± SEM) is 49.0 ± 0.6. **(e)** Histogram of diameter of 85 *Beads B.* Gaussian fit has a peak value of 104.0 nm and FWHM of 18.0 nm. Arithmetic average of the nanosphere diameter (mean ± SEM) is 99.2 ± 1.3.

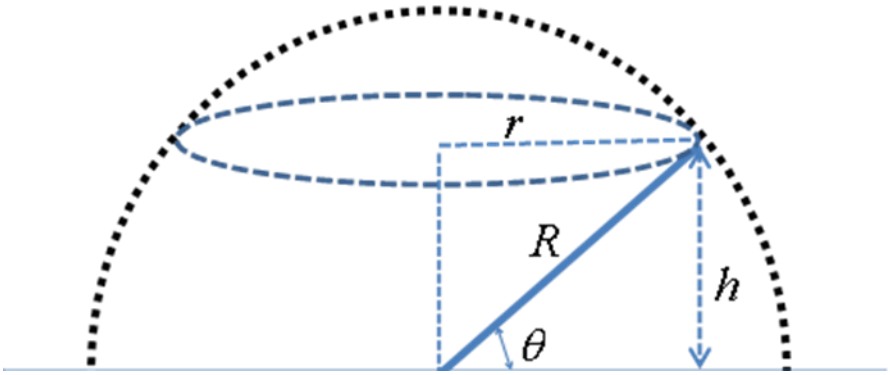

Figure S4. Schematic for calculation of the average tilt angle of a free rotating rod.

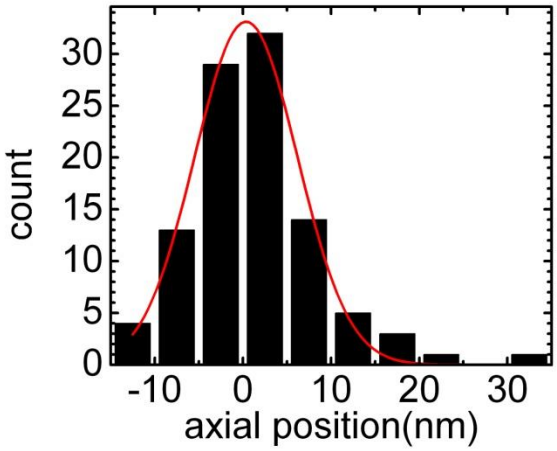

Figure S5. **Precision of sequential measurements.** The radius of a single nanosphere *(Bead F)* was measured 100 times. For each measurement, we sequentially aligned the AFM tip over the nanosphere, measured fluorescence oscillations and then measured excitation laser interference. The measurement precision using this strategy was lower than simultaneous measurements due to increased instrumental drift. The standard deviation in the measured bead radius was 5.4 nm.

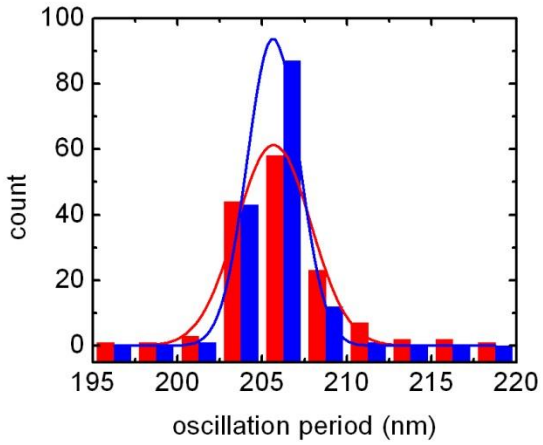

Figure S6. Oscillation period for fluorescence and excitation laser are similar. Distribution of fitted oscillation period (bar) and corresponding Gaussian distribution (solid line) for fluorescence (red) and laser reflection signal (blue). Both oscillations have the same period ($\lambda_{laser}/2n$) confirming that oscillation of fluorescence is only due to standing wave excitation and not due to fluorescence interference.

CHAPTER 5. CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Conclusions

I accomplished three goals in this dissertation. First, I resolved a critical biological problem in neuroscience, i.e. how does copper induce the misfolding, aggregation and neurotoxicity of PrP. Second, I established a more accurate analysis method for estimation of kinetic data from single molecule measurements. Third, I developed a new fluorescence localization microscopy with superior accuracy to previous methods. The significance and impact of each aspect of my work are summarized below.

In chapter 2, I integrated both ensemble and single molecule techniques to resolve the molecular mechanism of copper induced PrP misfolding, oligomerization and neurotoxicity [1]. Using a single molecule fluorescence assay, we demonstrated that Cu^{2+} ions misfold PrP monomers into a protease-resistant conformation before oligomer assembly. This is a noteworthy finding; before this work it was believed that protease resistance is acquired only by PrP oligomers, rather than PrP monomers. In fact, protease resistance was often used as a readout for PrP oligomerization [2]. In addition, using single molecule force measurements with an AFM, we showed that the binding affinity between PrP monomers increases 900 fold after misfolding, which promotes the rapid formation of PrP oligomers. Finally, using a cell-free seeding assay and organotypic slice cultures, we showed that misfolded PrPs form active seeds that template aggregation and subsequently lead to the inflammation and degeneration of neuronal tissues. These results demonstrated, for the first time that misfolded PrP serve

as seeds for templated formation of aggregates which mediate inflammation and degeneration of neuronal tissue.

While the DFS method we used to measure the binding affinity of PrPs in chapter 2, allowed us to quantify the dissociation kinetics of protein-protein interaction at the single molecule level, the conventional analysis method we used resulted in substantial errors in estimated kinetic parameters. In chapter 3, I addressed this concern by developing a new analysis method for single molecule DFS experiments. I presented a high accuracy method using clustering algorithms, to improve kinetic parameter estimation while retaining the simplicity of data collection and analysis of a conventional DFS experiment. I also benchmarked the performance of this new analysis method using computer simulations under an extensive range of conditions to serve as a guide for future users of DFS data analysis.

In my work with PrP, I used a single molecule fluorescence assay to indirectly determine the conformational change of individual PrP molecules [1]. With conventional fluorescence-based techniques, a single biological molecule can be localized with nanometer accuracy along the x- and y-axis [3]. However, along the z- axis, the best resolution that has been achieved, thus far, is ~10 nm using interferometry [4, 5]. In chapter 4, I described a novel fluorescence localization microscopy with a sub-nanometer accuracy and 3.7 nm precision along the optical axis, a technique called standing wave axial nanometry (SWAN). In SWAN, a standing wave generated between an AFM tip and microscope objective is used to excite a fluorophore; the axial position of the molecule is determined from the phase of the emission intensity. Compared to other interferometric techniques which require custom optics and a complicated layout

[4, 5], SWAN can be simply implemented by mounting a commercial AFM on a single molecule confocal microscope. Furthermore, other interferometric methods determine the axial position of molecules based on the assumption that fluorescence interference occurs only at the peak wavelength of the fluorescence spectra [4, 5]. Due to the broad emission spectra of fluorescent dyes, this approximation introduces systematic errors in axial localization. In contrast to these methods, periodic fluorescence oscillations measured in SWAN occur due to standing wave excitation of the fluorophore and not due to fluorescence interference which improves localization accuracy. Moreover, unlike most interference-based techniques, where the interference pattern repeats itself and limits the working range to ~250 nm, the standing wave in SWAN decays with tip-surface distance, which allows successive periods to be distinguished and extends the working distance.

5.2 Future Directions

Neurodegenerative disease research

In our study of PrP misfolding, a single molecule fluorescence assay was developed to distinguish the conformational change of PrP monomers based on their changes in protease resistance [1]. Since the acquisition of protease resistance upon protein misfolding and oligomerization is a common feature in many neurodegenerative disorders [6, 7], this method can be easily adopted to investigate the molecular mechanisms in similar diseases such as the structural conversion of α-synuclein protein in Parkinson's disease [7]. More importantly, our work has established a comprehensive workflow to quantify the biophysical properties of single PrP molecules.

In prion studies, one of the most challenging problems is that PrP molecules misfold and aggregate into various infectious conformations (strains), which result in different pathologies and clinical outcomes [8]. Different prion strains have different biophysical properties [9] which are averaged out in conventional bulk measurements. While multiple misfolding pathways in a single PrP molecule have been revealed using single molecule optical tweezers [10], a complete understanding of each prion strain is still lacking. A systematic investigation into aggregation kinetics, structural dynamics and biochemical properties of each prion strains at the single molecule level will be essential for future prion studies.

Single molecule force analysis

In AFM-DFS, protein complexes are ruptured at different force rates by varying the speed at which the AFM-tip and substrate are pulled away from each other. The rupture events are grouped according to their pulling speeds and the mean force and loading rate of each group is calculated to estimate the kinetic parameters. However, large uncertainties in determining the mean forces and loading rates can contribute significant errors in the estimated values. Using Monte Carlo simulations, I demonstrated that these errors can be dramatically reduced by sorting rupture events into groups using cluster analysis instead of sorting them according to their pulling speeds. Since this work is purely based on simulation results, a follow-up study validating our findings with real DFS data is required. By using DFS to measure protein complexes with known kinetics parameters, the performance of this new data analysis method can be confirmed. Furthermore, new clustering algorithms that sort rupture

events by describing the distribution of rupture forces using more sophisticated theoretical models [11, 12] would likely, further improve kinetic parameter estimation. Finally, since the principle of cluster analysis is to sort data into different groups such that events within the same category share similar characteristics, our method can also be used to identify and eliminate nonspecific artifacts which are not tightly clustered on a force-loading rate plot.

Single molecule fluorescence microscopy

Since SWAN uses an AFM tip as a micro mirror for light reflection, it is possible to adopt this technique to determine the axial position of molecules in single molecule AFM force measurements and in single molecule cut and paste applications for the bottom-up assembly of biological nanostructures [13]. Additionally, the high spatial accuracy of SWAN relies on the precise alignment of AFM tip over the focused laser beam between successive measurements to overcome the instrumental drift. This drift can be minimized by combining SWAN with ultra-stable AFM to avoid the time-consuming alignment before each measurement [14]. Finally, SWAN can be easily integrated with other lateral localization techniques for 3D tracking with nanometer accuracy [3].

5.3 References

[1] C.F. Yen, D.S. Harischandra, A. Kanthasamy, S. Sivasankar, Copper-induced structural conversion templates prion protein oligomerization and neurotoxicity, Sci. Adv. 2 (2016) e1600014.

[2] B. Caughey, G.S. Baron, B. Chesebro, M. Jeffrey, Getting a Grip on Prions: Oligomers, Amyloids, and Pathological Membrane Interactions, Annu. Rev. Biochem. 78 (2009) 177-204.

[3] A. Yildiz, P.R. Selvin, Fluorescence imaging with one manometer accuracy: Application to molecular motors, Acc. Chem. Res. 38 (2005) 574-582.

[4] G. Shtengel, J.A. Galbraith, C.G. Galbraith, J. Lippincott-Schwartz, J.M. Gillette, S. Manley, R. Sougrat, C.M. Waterman, P. Kanchanawong, M.W. Davidson, R.D. Fetter, H.F. Hess, Interferometric fluorescent super-resolution microscopy resolves 3D cellular ultrastructure, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 3125-3130.

[5] D. Aquino, A. Schönle, C. Geisler, C.V. Middendorff, C.A. Wurm, Y. Okamura, T. Lang, S.W. Hell, A. Egner, Two-color nanoscopy of three-dimensional volumes by 4Pi detection of stochastically switched fluorophores, Nat. Methods 8 (2011) 353-359.

[6] J. Stohr, J.C. Watts, Z.L. Mensinger, A. Oehler, S.K. Grillo, S.J. DeArmond, S.B. Prusiner, K. Giles, Purified and synthetic Alzheimer's amyloid beta (A beta) prions, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 11025-11030.

[7] M. Neumann, P.J. Kahle, B.I. Giasson, L. Ozmen, E. Borroni, W. Spooren, V. Muller, S. Odoy, H. Fujiwara, M. Hasegawa, T. Iwatsubo, J.Q. Trojanowski, H.A. Kretzschmar, C. Haass, Misfolded proteinase K-resistant hyperphosphorylated alpha-synuclein in aged transgenic mice with locomotor deterioration and in human alpha-synucleinopathies, J. Clin. Invest. 110 (2002) 1429-1439.

[8] R. Morales, K. Abid, C. Soto, The prion strain phenomenon: Molecular basis and unprecedented features, Biochim. Biophys. Acta, Mol. Basis Dis. 1772 (2007) 681-691.

[9] D. Peretz, M.R. Scott, D. Groth, R.A. Williamson, D.R. Burton, F.E. Cohen, S.B. Prusiner, Strain-specified relative conformational stability of the scrapie prion protein, Protein Sci. 10 (2001) 854-863.

[10] H. Yu, X. Liu, K. Neupane, A.N. Gupta, A.M. Brigley, A. Solanki, I. Sosova, M.T. Woodside, Direct observation of multiple misfolding pathways in a single prion protein molecule, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 5283-5288.

[11] O.K. Dudko, G. Hummer, A. Szabo, Theory, analysis, and interpretation of single-molecule force spectroscopy experiments, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 15755-15760.

[12] O.K. Dudko, G. Hummer, A. Szabo, Intrinsic rates and activation free energies from single-molecule pulling experiments, Phys. Rev. Lett. 96 (2006) 108101.

[13] S.K. Kufer, E.M. Puchner, H. Gumpp, T. Liedl, H.E. Gaub, Single-molecule cut-and-paste surface assembly, Science 319 (2008) 594-596.

[14] G.M. King, A.R. Carter, A.B. Churnside, L.S. Eberle, T.T. Perkins, Ultrastable Atomic Force Microscopy: Atomic-Scale Stability and Registration in Ambient Conditions, Nano Lett. 9 (2009) 1451-1456.