Refined procedure of evaluating experimental single-molecule force spectroscopy data

Alexander Fuhrmann,[*] Dario Anselmetti, and Robert Ros[*,†]
Experimental Biophysics, Physics Department, Bielefeld University, 33615 Bielefeld, Germany

Sebastian Getfert and Peter Reimann
Condensed Matter Theory, Physics Department, Bielefeld University, 33615 Bielefeld, Germany

Dynamic force spectroscopy is a well-established tool to study molecular recognition in a wide range of binding affinities on the single-molecule level. The theoretical interpretation of these data is still very challenging and the models describe the experimental data only partly. In this paper we reconsider the basic assumptions of the models on the basis of an experimental data set and propose an approach of analyzing and quantitatively evaluating dynamic force spectroscopy data on single ligand-receptor complexes. We present our procedure to process and analyze the force-distance curves, to detect the rupture events in an automated manner, and to calculate quantitative parameters for a biophysical characterization of the investigated interaction.

I. INTRODUCTION

Dynamic force spectroscopy is widely used to investigate molecular recognition on the single-molecule level (e.g., reviewed in [1]). From these experiments, quantitative data in terms of energy landscape parameters and kinetic constants of the interaction can be obtained [2]. The technique can be applied to a remarkable range of interactions; from the binding of complex biological molecules like antibodies [3–6], proteoglycans [7,8], cytochromes [9], chaperones [10], selectines [11], protein-DNA interactions [12–14] to small bioorganic or organic compounds like peptides [15] and supramolecular systems [16–18]. The binding affinities of the probed complexes can differ by several orders of magnitude. For example, the seminal early force spectroscopy works on streptavidin/avidin-biotin interactions [19,20] yield a dissociation constant K_D in the range of 10^{-15} M, whereas for weak calixarene-ion complexes one finds a K_D of 10^{-5} M [18].

Ligand-receptor interactions are mainly probed by dynamic force spectroscopy based on atomic force microscopy (AFM), but also alternative techniques like the biomembrane force probe [2,21] or optical tweezers [22] are available. A typical experimental setup is sketched in Fig. 1. One binding partner is connected to the force transducer (i.e., the tip of a soft AFM cantilever) and the other is bound to a surface. The ligand, the receptor-molecules, or both, are usually linked via polymeric tethers to the tip and to the surface, respectively [3,5,12,15,23]. In order to obtain force-distance curves, the AFM tip or the surface is cycled up and down while measuring the force acting on the cantilever. From these force curves rupture or unbinding forces are analyzed for various pulling velocities.

The interpretation and theoretical modeling of these experiments is challenging, as shown in various publications [24–27]. Starting with the original works of Bell [28] and Evans and Ritchie [29], almost all theoretical models rely on the assumption of a well-defined force-extension characteristic of all the involved elastic components of the experimental setup, which is independent of the velocity at which the pulling force increases and in particular does not change upon many repetitions of the same experiment. Since dispersive linker lengths, nonorthogonal pulling geometries, or partly mis- or unfolded binding partners can contribute to systemic spreads in force-extension characteristics, the main purpose of our present paper is a careful reconsideration of the above-mentioned basic assumption by analyzing the force-extension curves of specific protein-DNA interactions in the field of prokaryotic transcriptional regulation [30]. Ad-
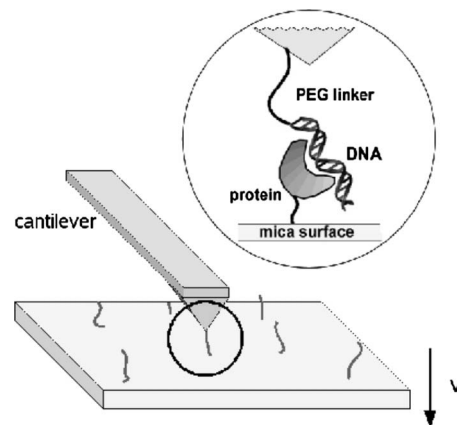

FIG. 1. Schematic illustration of a dynamic force spectroscopy experiment. The receptor (e.g., a protein) is immobilized on the surface (e.g., mica) and the ligand (e.g., a DNA-fragment) is connected via a linker [e.g., poly(ethylene glycol) (PEG)] to the tip of an AFM cantilever which serves as a force transducer. The distance between the mica surface and tip can be controlled with a piezo-electric element. When the surface is pulled down at a constant speed v, monotonically increasing forces act on the ligand-receptor complex.

*Present address: Department of Physics, Arizona State University, Tempe, AZ 85287, USA
†robert.ros@asu.edu

ditionally, we present our procedure to process and analyze the force-distance curves, to detect the rupture events in an automated manner, and to calculate quantitative parameters for a biophysical characterization of the investigated interaction. This results in a data analysis technique for dynamic force spectroscopy on single ligand-receptor complexes.

II. THEORETICAL BACKGROUND

In this section we briefly summarize the main theoretical framework relevant for our subsequent discussions. Nearly all theoretical models for bond rupture are based on the seminal works of Bell, Evans, and Ritchie [28,29]. Within this approach, a forced bond rupture is viewed as a thermally activated decay of a single slow reaction coordinate over a potential barrier.

Let $f(t)$ denote the force that acts on the bond at time t. Since it is increasing much slower than all relevant molecular relaxation processes, the reaction kinetics can be very well approximated by

$$\dot{n}(t) = -k(f(t))n(t), \qquad (1)$$

where $n(t)$ denotes the survival probability of the bond up to time t and $k(f)$ the dissociation rate at an arbitrary but fixed force f.

A further key assumption is that the instantaneous force $f(t)$ depends solely on the total instantaneous extension $s = s(t)$ of all elastic components of the complex (cantilever, linker, receptor, and ligand). In other words, there exists a common function $F(s)$ (later referred to as the master curve) such that

$$f(t) = F(s(t)) \qquad (2)$$

for all the pulling experiments under consideration, independently of any further details (pulling speed, linker properties, etc.) of the single repetitions of the experiment.

Finally, we restrict ourselves to the usual case that the surface is retracted at constant velocity v, i.e.,

$$s(t) = vt, \qquad (3)$$

and without loss of generality we choose the time offset such that $t=0$ when pulling starts. Hence it can be assumed that $f(t)$ is a continuous and monotonically increasing function of time. Then Eq. (1) can be expressed in terms of the force with formal solution

$$n(f) = n_v(f) = \exp\left(-\frac{1}{v}\int_{f_{\min}}^{f} df' \frac{k(f')}{F'(F^{-1}(f'))}\right) \qquad (4)$$

for $f > f_{\min}$, where $f_{\min}$ denotes the threshold value of the force below which the measurement is dominated by random fluctuations and artifacts so that rupture events cannot be detected in this regime (see Sec. III A). The subscript v refers to the velocity dependence of the survival probability.

In order to extract any useful and interesting quantitative information from the measurement, further assumptions about the force dependence of the dissociation rate $k(f)$ have to be made. A quite common approximation going back to Bell [28] is

$$k(f) = k_0 \exp\left(\frac{x_\beta f}{k_B T}\right), \qquad (5)$$

with k_0 the dissociation rate in the absence of an external force, x_β the distance between potential well and barrier, and $k_B T$ the thermal energy. In the following we refer to the model relying on Eqs. (1)–(5) as the standard or Bell model. Generalizations of this approximation have recently attracted considerable interest [31–33].

As can be seen from Eq. (4), specifying the force dependence of the rate $k(f)$ alone is not sufficient for analyzing the data: Also the full force-extension characteristic $F(s)$ including the threshold value $f_{\min}$ is required. It should be remarked that there exist a number of different models describing the mechanical response of idealized polymeric chains in terms of microscopic parameters like contour and Kuhn length (e.g., the freely jointed chain or the wormlike chain model, as reviewed, e.g., in [34]). Since the total elastic entity is composed of cantilever, linkers, and ligand-receptor molecules, no simple realistic model exists to describe the relevant force-extension curve $F(s)$. In the following we therefore just assume that an appropriate approximation for $F(s)$ is available from experimental data.

Given the survival probability $n_v(f)$, the probability density to observe a rupture event is

$$p(f|\vec{\mu},v,F,f_{\min}) = -\frac{d}{df} n_v(f), \qquad (6)$$

where the notation in terms of a conditional probability $p(f|\cdots)$ refers to the fact that this density is also conditioned on the model parameters $\vec{\mu}$ [e.g., in the Bell model $\vec{\mu} = (x_\beta, k_0)$], the pulling velocity v, the full force-extension curve, and the force offset $f_{\min}$. Under the above discussed assumption that $F(s)$ in Eq. (2) is the same function in all rupture measurements, a direct consequence of Eq. (4) is that $-v \ln(n_v(f))$ is independent of the pulling velocity [35]. For a large number of experimental data sets this has been demonstrated *not* to be the case [26]. In particular the distributions of measured rupture forces are much broader than those predicted by the standard model. A possible explanation for this discrepancy is given in [27], consisting of a heterogeneous bond model such that Eqs. (1)–(5) are still valid except that due to uncontrollable variations of the molecular complex or of the local environment of the bond, the parameter x_β is itself subjected to random variations. Sampling this parameter from a Gaussian distribution with mean $\bar{x}_\beta$ and variance σ_x^2 results in a good agreement with the analyzed data sets. The heterogeneous bond model thus involves the three parameters $\vec{\mu} = (\bar{x}_\beta, \sigma_x, k_0)$. Note that the standard model is a special case of the heterogeneous bond model corresponding to $\sigma_x = 0$, while the qualitative and quantitative differences between the two models become more and more pronounced with increasing variance σ_x.

III. DATA ANALYSIS

A dynamic force spectroscopy experiment consists typically of thousands of force-distance curves. This requires a

stable procedure to detect and process extended data sets. We developed MATLAB (Mathworks) based routines for the analysis of force-distance curves, which automatically detect, categorize, and quantify rupture events. In the following sections we motivate and develop step by step this procedure of processing and evaluating experimental single-molecule force spectroscopy data.

A. Raw data analysis

A raw data force-extension curve measured by the AFM for one rupture experiment is shown in Fig. 2, adopting the standard convention to plot on the abscissa the piezoposition (increasingly negative as time goes on) and on the ordinate the force acting on the cantilever, being positive if the cantilever in Fig. 1 is bent upward with respect to its rest position and negative for downward bending.

Our first step of processing the raw data is to fit the sensor response (dotted magenta straight line) and the reference baseline (solid green straight line: the cantilever is far away from the surface and the complex unbound, hence the only remaining forces are thermal fluctuations; the dotted green straight line represents the extrapolation of the observed baseline). Their intersection point is denoted by r_0 and can be identified with the moment at which the cantilever leaves the sample surface. We remark that in the measured force-distance curves the slope of the reference baseline can be slightly different from zero (e.g., due to thermal drift effects), and the zero values for both the raw force and the raw displacement data can be chosen arbitrarily. Therefore all forces will at some later stage be converted such that they represent the (positive) distance between the measured raw force and the reference baseline.

The next task is to identify discontinuities in the force-distance curves. An example is shown in Fig. 2 where the measured force value makes a large jump at point r_2. Clearly, such discontinuities are candidates for possible rupture events of the actual molecular bond of interest. If more than one discontinuity is found, only the last one is a potential candidate (inset of Fig. 2), while the preceding discontinuities must be disregarded as being caused by rupture of one out of several coexisting bonds.

As outlined in the previous section, an appropriate parametrization of the force-extension curve before rupture is required. We found that for our examined systems straight lines are not appropriate but that second degree polynomials are satisfactory for the parametrization within the whole relevant range of loading forces between r_2 and r_1 in Fig. 2. However, the point r_1, where the complex actually begins to be stretched, is not known *a priori* and has to be determined self-consistently together with the fitting procedure of the force-extension curve. To accomplish this task several points have to be considered. First, repulsive forces acting on the AFM tip can only be caused by noise or interactions between tip and surface. In both cases the corresponding force does not act on the bond (the complex of linkers, ligand molecules, and receptor molecules is not stretched). Consequently, these forces do not influence the lifetime of the bond and the concomitant part of the force-extension curve with
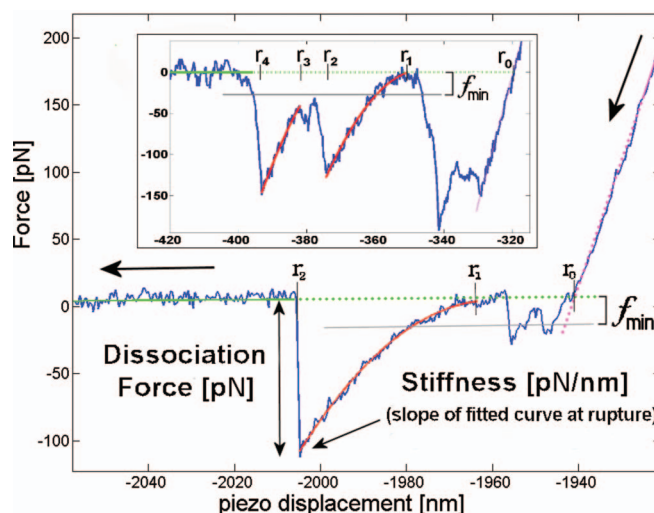

FIG. 2. (Color) Typical force-distance curve (only the so-called retracting part of the complete force-distance cycle is shown; direction of pulling from right to left, indicated by the two large arrows). Left of the point r_2, the connection between cantilever and surface has broken and the measured signal is caused by thermal fluctuations. Fitting this part of the curve with a straight line yields the "baseline" (solid green; extrapolated as dotted green line) which sets the force offset. Its intersection with the extrapolated dotted magenta line ("sensor response," fitted by a straight line) at r_0 defines the point where the cantilever leaves the surface. Between r_0 and r_1 the measured force is caused by adhesion, thermal fluctuations, and unspecific interactions. From (approximately) r_1 on, the complex composed of a cantilever, linker, and the ligand-receptor molecules is stretched until the bond breaks at r_2. The part of the rupture curve between r_1 and r_2 is called the force-extension characteristic or loading and can be fitted by a second degree polynomial (solid red curve). The slope of this polynomial at extension r_2 is called stiffness and the corresponding difference between the fitted red curve and the green baseline at r_2 is defined as the rupture force (or dissociation force; indicated as a vertical arrow). The gray solid line is a copy of the green line shifted downwards. The (positive) distance between the green and gray lines is denoted as force threshold f_{min}. For the purpose of identifying rupture events, only force values beyond this threshold are taken into account. Inset: Example of a force-distance curve for which no rupture event would be accepted for further analysis for the following reasons: The discontinuity at point r_2 is neither the last discontinuity in the force distance cycle nor does the cantilever immediately "jump back" to the green baseline. The observed event at point r_4 might correspond to the rupture of a single bond, but the corresponding red force-distance curve exceeds the threshold f_{min} from the very beginning.

positive forces is of no further interest. Second, the force has to be a monotonically increasing function of the extension of the complex which is expected to become stiffer upon being stretched (i.e., the second derivative of the curve in Fig. 2 has to be negative). Finally, discontinuities can only be explained by rupture events or adhesion effects. Hence there should be no discontinuity in the interval $[r_2, r_1]$.

The main idea of our procedure is to iteratively determine the point r_1. In every step the force-extension curve is fitted by a second degree polynomial and it is checked whether the

above-mentioned points are satisfied. In detail, our algorithm works as follows: (i) r_1 is chosen only slightly larger than r_2; (ii) a second degree polynomial is fitted to the raw data in the interval $[r_2, r_1]$; and (iii) it is checked whether the fitted polynomial meets one of the following conditions within the interval $[r_2, r_1]$: (a) it crosses the reference baseline; (b) it exhibits a maximum; (c) it has the "wrong" curvature as compared to the red line in Fig. 2, i.e., its second derivative is positive; and (d) another discontinuity has been detected in this interval. If none of the conditions are fulfilled, r_1 is increased by a small value Δr and steps (ii) and (iii) are repeated. Otherwise, r_1 is decreased by Δr and the resulting final r_1 value is defined as the position where the complex begins to be stretched. The corresponding final polynomial fit is chosen for the parametrization of the force-extension curve and the rupture force f_i is defined as the difference between this fit and the reference baseline at r_2 (see Fig. 2). In this way noise effects that especially dominate the measurement in the low-force regime can be reduced to a minimum. In practice, we found that an additional discontinuity [condition (d)] in fact always implies a maximum [condition (b)] of the fitted polynomial.

The force value at the beginning of the fit (at $r=r_1$) is the (force) offset for this single rupture curve. Most fits do not reach the zero force baseline for basically two reasons. First, due to a limited resolution of the apparatus and unspecific interactions between tip and surface, we cannot measure arbitrarily small forces. Second, as discussed above we only use the last rupture event in each force-distance cycle. However, in case that there are multiple rupture events in the cycle, it is likely that the cantilever does not "jump back" to the baseline after the second to last bond has ruptured (see inset of Fig. 2). We therefore introduce a threshold parallel to the baseline but at a distance $f_{\min}$ below it (see also Fig. 2). Note that by convention $f_{\min}$ is always positive like the rupture force. For a quantitative evaluation of the experiment it is required [see Eqs. (4) and (6)] that the bond has been formed at forces lower than this threshold value $f_{\min}$ and that all rupture events with rupture forces larger than this threshold value can be detected, i.e., $f_{\min}$ must in particular be larger than the thermal fluctuations. That means we only accept rupture events with the properties that the polynomial fit crosses the threshold between r_2 and r_1 before jumping back to the baseline at r_2. Consequently, when choosing $f_{\min}$ too small a lot of rupture events will not be accepted because the corresponding force value at the beginning of the pulling (at $r=r_1$) is already larger than the threshold force. On the other hand, when choosing $f_{\min}$ too large also a lot of rupture events will not be accepted because the rupture force is smaller than $f_{\min}$. Hence we usually choose this value such that as many rupture events as possible can be used for the evaluation of the experiment. In our experiment we found that noise effects increase with the pulling velocity. Therefore we generally admit that the offset $f_{\min}$ still depends on the pulling velocity, i.e., $f_{\min} = f_{\min}(v)$.

Other technical problems, when automatically processing the force-distance curves, are caused by adhesion effects. Near the sample surface, adhesion forces act on the cantilever which can cause a signal in the force-extension curve (cf. Fig. 2 left of point r_0). Most successful single-molecule force spectroscopy experiments use polymeric linkers of a defined length to connect the binding partners to the tip and the surface, respectively. This allows limiting the forced rupture events of interest to a distinct distance window [37]. Hence we use the distance from r_0 to r_1 as another selection criterion and exclude rupture events with small $|r_1 - r_0|$ from further analysis.

B. Construction of the force-extension master curve

One of the basic theoretical assumptions (see Sec. II) is a well-defined force-extension characteristic which, in particular, is independent of the pulling velocity and does not change upon repeating the experiment many times. In this section we show that this assumption does often not apply to the experimental reality and present a numerical scheme to remedy this discrepancy.

Experimental finding. In agreement with Ref. [38], we found that the second order polynomial fits resulting from the above described selection and fitting procedure do not give rise to a unique force-extension curve. This can easily be seen by considering the data pairs (f_i, f_i') with f_i a rupture force and f_i' the slope of the corresponding polynomial at the point of rupture. Binning these data (f_i, f_i') in a two-dimensional (2D) histogram, gives rise to a different kind of plot, featuring stiffness against force. A representative example of such a stiffness versus force plot is given in Fig. 3. If the assumption of a well-defined force-extension characteristic were correct, all points (f_i, f_i') should lie on a single curve (depending on the whole complex of linkers, ligand, and receptor molecules and the cantilever) apart from a slight broadening due to remnant noise effects. This assumption is thus incompatible with the experimental findings depicted in Fig. 3.

Main idea. Our main idea to remedy this discrepancy between theoretical prerequisite and experimental data is to introduce a further selection procedure with the very purpose that at the end only force-extension curves remain which all are very close to one single "master curve." In order to do this we have to focus on two points. First, the number of finally remaining force-extension curves should be as large as possible. Second, the master curve has to be characteristic for "typical" rupture events. We now describe our general procedure.

Construction of the master curve. In a first step we consider for *one pulling velocity* all rupture events (f_i, f_i') in a region

$$f_i \in [f_c - \Delta F/2; f_c + \Delta F/2], \tag{7}$$

$$f_i' \in [f_c' - \Delta S/2; f_c' + \Delta S/2] \tag{8}$$

with f_c a typical rupture force and f_c' a typical stiffness. In practice (f_c, f_c') is usually the position of the maximum in Fig. 3. Furthermore, the two interval widths ΔF, ΔS are chosen so that a "sufficient" number of points lie within the selected parameter windows. To exclude a possible contamination with double rupture events and obvious outliers, one can visibly inspect the corresponding force-extension curves and if necessary readjust the parameters.
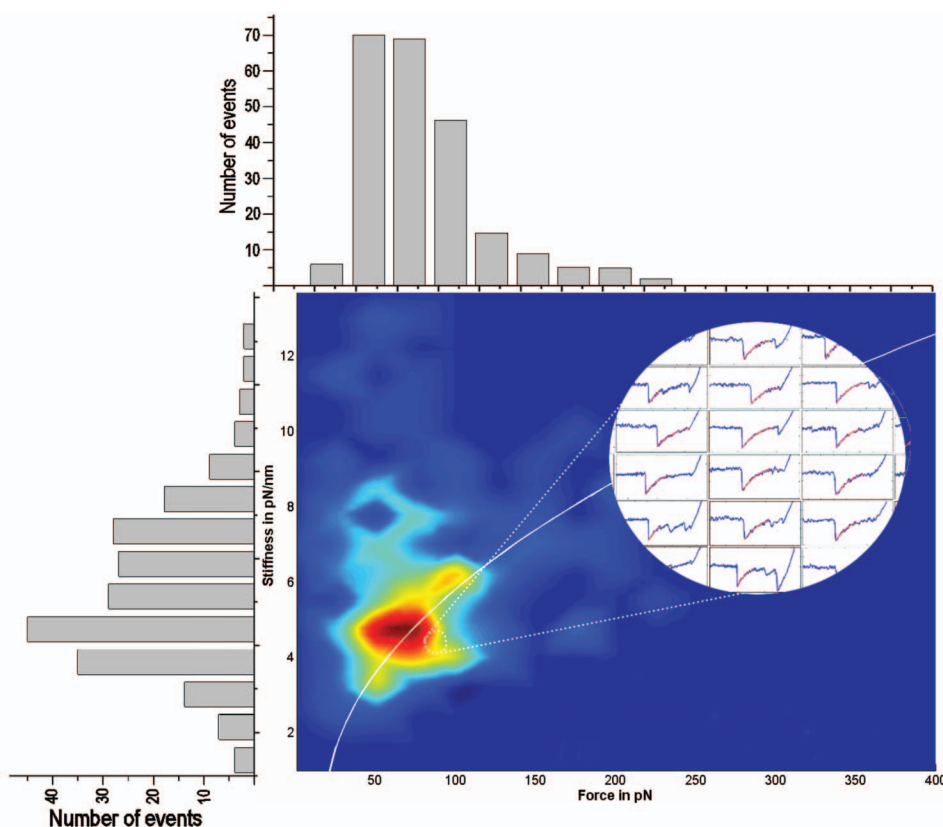
FIG. 3. (Color) Rupture force data for protein-DNA interaction previously studied in [30] are plotted against the slope of the force-extension curve at the point of rupture r_2 (stiffness) in a 2D histogram (red, high frequency, blue, low frequency). At a pulling velocity of $v=5000$ nm/s, 2317 rupture curves have been measured. After the automated analysis (see Sec. III A) with threshold force $f_{min}=42$ pN, 259 specific rupture events have been identified. The white solid line corresponds to the master curve which has been constructed out of the force-extension curves at the maximum of the stiffness against force plot. The cumulated distributions of the rupture force and of the stiffness are shown above and to the left of the 2D histogram. Inset: Illustration of raw data force-extension curves of the rupture events with (f_i, f_i') falling into the marked region of the 2D histogram.

After applying the above selection procedure, the remaining force-extension curves are automatically aligned by properly transforming abscissa and ordinate in Fig. 2 as follows: In a first step, the data are transformed such that the green reference baseline in Fig. 2 falls on top of the zero force line. Then the abscissa is shifted such that every adjusted force-extension curve crosses the gray threshold line at the same position $r=0$, see Fig. 4(b). Finally, this data set is used for a second degree polynomial fit, henceforth called the (force-extension) master curve and exemplified by the red line in Fig. 4(b).

Selection procedure. In a second step all force-extension curves for *all pulling velocities* are compared with this master curve. The decision of which curves are accepted for the further analysis is controlled by two parameters (δ_1, δ_2). The first of them sets the maximum relative deviation of the slope at the rupture point from the slope of the master curve at this point. The second defines an interval around the master curve, limiting the maximum relative deviation from the master curve as indicated by the green lines in Fig. 4(b). Only force-extension curves lying entirely in this interval will be accepted.

In the previous section we have described the necessity to introduce a threshold force f_{min} and to use only rupture events for the evaluation where the bond has formed below this force, but with rupture force $f_i > f_{min}$. However, it seems desirable that the master curve should also correctly describe the part of the force-extension curve for lower forces. Hence for the construction of the master curve, i.e., for choosing the subset of rupture forces out of which the master curve is constructed, a lower threshold force $\tilde{f}_{min} < f_{min}$ should be used sooner than later for the quantitative evaluation of the experiment.

It should be noted that the parameters (δ_1, δ_2) control how similar the force-extension curves of the accepted rupture events are. Ideally, one would wish to choose both as small as possible. However, the smaller they are, the lower is the number of accepted rupture events and, hence, the worse the statistic. Therefore one has to make a compromise between these two points. In contrast to the choice of (δ_1, δ_2) the choice of the interval widths ΔF and ΔS in Eqs. (7) and (8) is rather uncritical as they only determine the number of events used to construct the master curve. The detailed discussion of the explicit choice of the "filtering parameters" for a particular experiment will be given in Sec. IV.

A subtle point arising with the above selection of force-extension curves regards the question whether such a "filtering procedure" is not introducing certain hidden modifica-
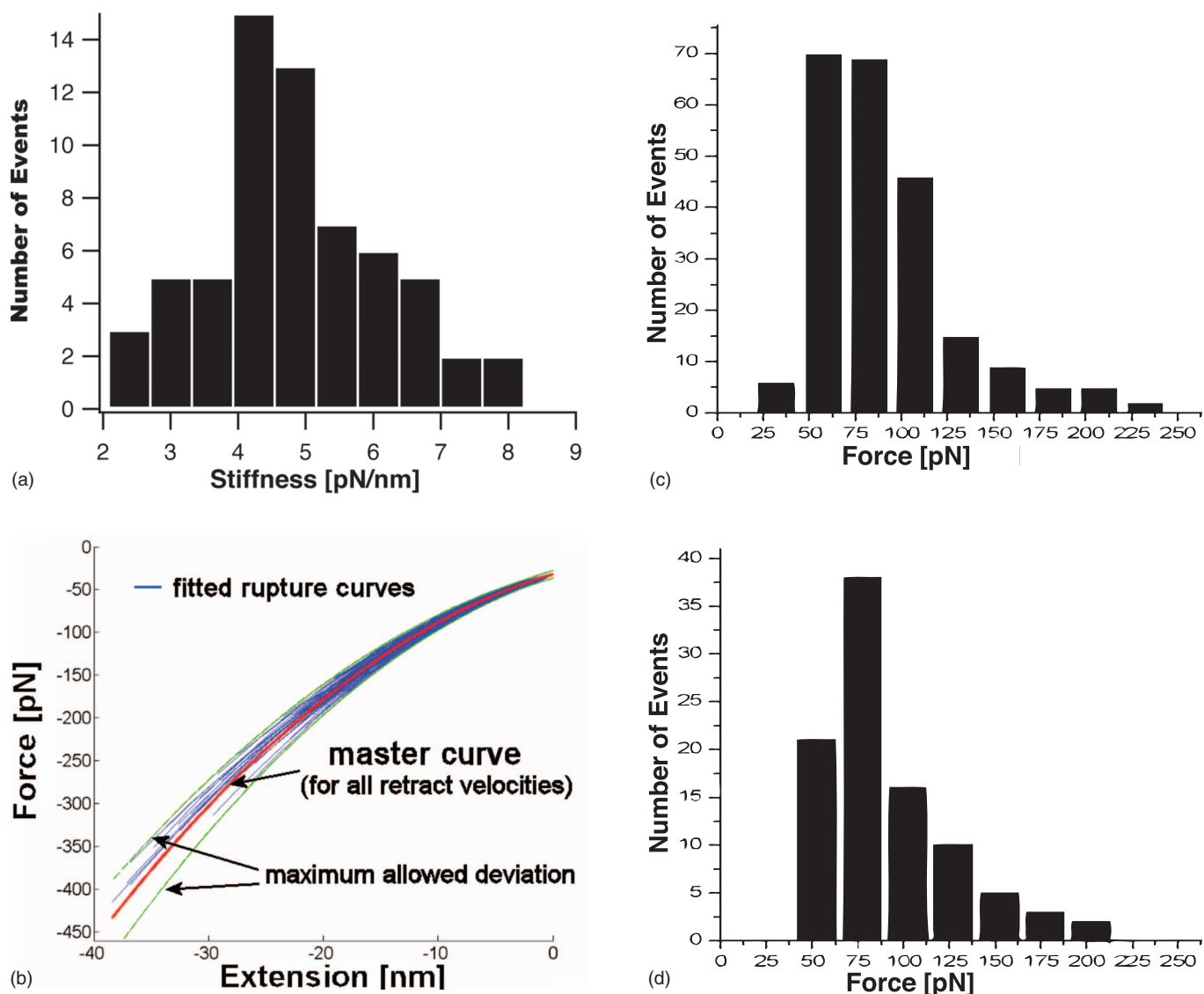
(a)

(b)

(c)

(d)

FIG. 4. (Color) Rupture force data from Fig. 3. (a) Histogram of the stiffness of all force-extension curves with rupture forces between 60 and 80 pN. (b) (Illustration) The master curve (red) is constructed as described in the text. The maximally allowed deviation of stiffness at the point of rupture is $\delta_1 = 0.2$ and the maximally allowed relative deviation from the master curve $\delta_2 = 0.12$. The blue thin lines are force-extension curves of single measurements (fitted by second degree polynomials) which have been accepted by the filtering procedure (see Sec. III B). All these events have force-extension curves which lie entirely between the two green lines that indicate the allowed deviation from the master curve. (c) and (d) Histograms of the rupture forces before and after this filtering procedure, respectively.

tions of the rupture event statistics. For example, the "long" force-extension segments, corresponding to large rupture forces, could have an over- or under-proportional probability to be eliminated, compared to the "short" ones. By comparing rupture force distributions for different filtering parameters, we will show in Sec. IV that such problems are apparently not very significant.

C. Further analysis of dynamic force spectroscopy data sets

The result of the so far data analysis is a set of rupture forces $\vec{f} = (f_1, \ldots, f_N)$ measured at pulling velocities $\vec{v} = (v_1, \ldots, v_N)$ with corresponding force offsets $\vec{f}_{\min} = [f_{\min}(v_1), \ldots, f_{\min}(v_N)]$. We recall that the threshold $f_{\min}$ de-

pends only on the pulling velocity and not on the actually measured rupture force, i.e., two rupture forces measured at the same pulling velocity have the same value for $f_{\min}$. For all these rupture measurements the force-extension curves follow a pathway close to a master curve as exemplified by Fig. 4(b).

To establish the connection with the theoretical description in Sec. II, one first of all has to invert the negative signs of both the extensions and the forces into their positive counterparts, resulting in the theoretical master curve $F(s)$. The next step is to estimate the model parameters $\vec{\mu} = (\bar{x}_\beta, \sigma_x, k_0)$ of the heterogeneous bond model. To this end we adopted a maximum likelihood procedure. Accordingly, the most probable parameters $\vec{\mu}^*$ are those that maximize the likelihood function

$$L(\vec{\mu}) = \prod_{i=1}^{N} p(f_i|\vec{\mu}, v_i, F, f_{\min}(v_i)). \qquad (9)$$

In doing so, it has been taken for granted that all force-extension curves follow exactly the master curve. This seems justified in view of the selection procedure described in the preceding section such that the actual deviations of the remaining force-extension curves from the master curve are very small. For a more detailed discussion of the adopted maximum likelihood procedure we refer to [36]. Here, we only point out that for estimating the parameters, no binning of the rupture forces to histograms is needed. The histograms are only used to compare the fitted distribution with the experimentally measured one. A further advantage is that estimates for the statistical error are easily obtained. The most probable parameters $\vec{\mu}^*$ are found by maximizing the likelihood function numerically. For practical purposes we have used $\lambda = \ln(k_0)$ as a fit parameter. Using k_0 itself would result in the same estimate k_0^* for the parameter but the corresponding error interval is not symmetrically around the most probable value. For the data presented in this work a commercial algorithm from the NAG (National Algorithms Group) library has been used for the maximization.

D. Software integration

The entire above described procedure for dynamic force spectroscopy data analysis is developed in MATLAB (Mathworks) and C programming language. Most functions of the program can be handled via a graphical user interface (GUI) enabling one to do a data analysis with a scan rate of about 200 force distance curves per minute with a standard PC. This force spectroscopy software can analyze both the data format of our homebuilt force spectroscopy instruments [12] and the data provided by the commercial MFP-3D (Asylum Research, USA).

IV. RESULTS AND DISCUSSION

In order to test and illustrate our data analysis method, the raw data originally published in [30] have been taken. The experiment conducted involved specific protein-DNA interactions relevant for prokaryotic transcriptional regulation. These interactions are mediated by small signaling molecules (N-acyl homoserine lactones, AHL), which are able to stimulate the protein-DNA binding by docking onto the proteins. In this particular experiment, the DNA target sequence is located on a 285 bp long DNA fragment, which is immobilized on the tip via an approximately 35 nm long poly(ethylene glycol)-linker, see also Fig. 1. The proteins are immobilized on mica surfaces via short linker molecules and stimulated by the AHL N-decanoyl-DL-homoserine lactone (C10-HL).

Each force-extension curve could be well-approximated by a second degree polynomial (for a particular example see Fig. 2), but no universal force-distance curve, neither for a specific pulling velocity, nor for the entire data set could be found, in disagreement with the theoretical model. Qualitatively this can be seen in Fig. 3, where the plot shows a broad single maximum with a significant number of outliers. To quantify this finding, we choose the subset of force-extension curves corresponding to rupture forces between 60 and 80 pN. For this subset, Fig. 4(a) shows the histogram of the slope (stiffness) of the fitted polynomials at the point of rupture. It has a pronounced maximum at 4.5 pN/nm but also a dispersion which cannot be neglected. Possible explanations are a dispersion of linker length [25] or that in some cases not single but multiple bonds are stretched. In the latter case, the corresponding complex is stiffer. Another possible reason is that the angle between the stretched complex and the surface normal is in most cases not zero [37]. As different angles yield different force-extension curves, variations of the pulling geometry might also explain our findings.

To account for the problem that no universal force-extension curve exists, a master curve can be constructed and only rupture events with a force-extension characteristic close to this curve are accepted for further analysis. For the construction of the master curve a subset of rupture events is chosen and a second degree polynomial is fitted to the corresponding force-extension curves. A fully objective criterion which subset should be chosen for this construction does not exist. However, as discussed in Sec. III B the master curve should be characteristic for typical rupture events. Therefore we have used the rupture curves corresponding to the peak of the 2D histogram in Fig. 3. The corresponding pulling velocity is $v = 5000$ nm/s being the highest pulling velocity used in [30]. The reason for considering just the largest v is simply that the rupture force increases with the pulling velocity. Hence the constructed master curve should be more appropriate for describing the force-extension characteristic in the high force regime than the corresponding curve constructed from rupture curves sampled at lower pulling velocities. Analogously, for an appropriate parametrization of the force-extension characteristic in the low-force regime it is desirable that the force-extension curves used to construct the master curve start well below the threshold force $f_{\min}$ which had to be introduced for a quantitative evaluation of the experiment (see Sec. III B). We have, thus, temporarily used a lower threshold $\tilde{f}_{\min}$ for this construction than later in the quantitative evaluation. In particular a threshold of 10 pN results in a very low number of force curves (10–20) which are then used for the construction. The red line in Fig. 4(b) shows the constructed master curve together with the force-extension curves of various single measurements which are used for a quantitative evaluation of the experiment. All these rupture curves lie within a preset level of deviation. Thus all rupture events that are used for further analysis have approximately the same force-extension pathway. Since it is very unlikely that a multiple and a single bond have the same force-extension curve, it follows that a contamination of the rupture force data by multiple-rupture events is reduced to a minimum. In order to examine whether the constructed master curve strongly depends on the selected subset of rupture forces, we have varied the selected area in the 2D histogram. As long as this area is near the peak, the constructed master curves were always similar. We have also compared the resulting master curve with constructed master curves from other pulling velocities, but no systematic deviations could be found.
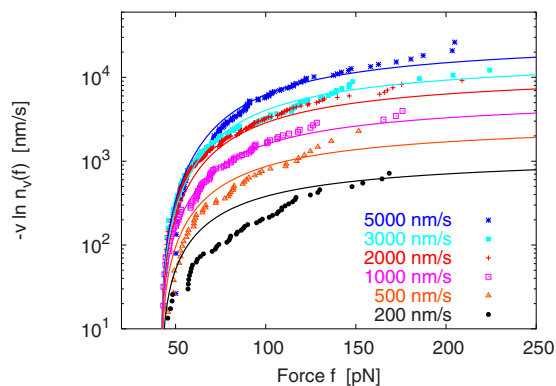
FIG. 5. (Color online) Symbols: The functions $-v \ln(\tilde{n}_v(f))$ for the experimental rupture force data from [30]. Each point corresponds to one observed rupture event. A unique threshold force $f_{\min}=42$ pN has been used for all pulling velocities and only events with a force-extension characteristic similar to the constructed master curve $F(s)$ have been taken into account (for more details see Sec. III B). The number N_v of experimental data points for the six different velocities are $N_{200\ \mathrm{nm/s}}=54$, $N_{500\ \mathrm{nm/s}}=48$, $N_{1000\ \mathrm{nm/s}}=80$, $N_{2000\ \mathrm{nm/s}}=151$, $N_{3000\ \mathrm{nm/s}}=87$, and $N_{5000\ \mathrm{nm/s}}=95$. A few large rupture forces are omitted for sake of better visibility. Solid lines: Theoretical functions for the heterogeneous bond model with force-extension characteristic $F(s)$.

In a next step, we have applied the filtering procedure for different choices of the tolerance levels (δ_1, δ_2). Our general observation is that with decreasing tolerance levels the width of the histogram of rupture forces initially decreases slightly. From a certain point on, however, the width is constant and the only effect of further lowering (δ_1, δ_2) is that the number of accepted rupture events gets smaller. For the particular choice $(\delta_1=0.2, \delta_2=0.12)$ the result of this filtering procedure is shown in Figs. 4(c) and 4(d) where the histograms of rupture forces before and after the filtering procedure are presented for one particular pulling velocity. Within their intrinsic statistical uncertainties, these histograms differ from each other only slightly. The same was found for all other pulling velocities and analyzed data sets. Even going to smaller tolerance levels (δ_1, δ_2) does not change this finding, implying that the distribution of rupture forces does not critically depend on details of the force-extension characteristic. Moreover, it follows that the filtering procedure itself does not introduce significant artifacts into the rupture force statistics.

Having ensured that the theoretical assumption of a unique force-extension characteristic is met at least for the considered subset of data, we can now check whether any model relying on Eqs. (1)–(3) is compatible to the experimental data. Given N_v rupture forces f_i, $i=1,\dots,N_v$, measured at a pulling velocity v, the survival probability of the bond can be approximated as $\tilde{n}_v(f)=N_v^{-1}\sum_{i=1}^{N_v}\Theta(f_i-f)$ with the Heaviside step function $\Theta(\cdot)$. As discussed in Sec. II, a direct consequence of Eqs. (1)–(4) is that $-v \ln(n_v(f))$ is independent of the pulling velocity (for a velocity independent threshold force $f_{\min}$). Figure 5 shows that this is definitively not the case for our experimental data set from [30]. Hence neither the standard Bell model [29] nor any of its extensions
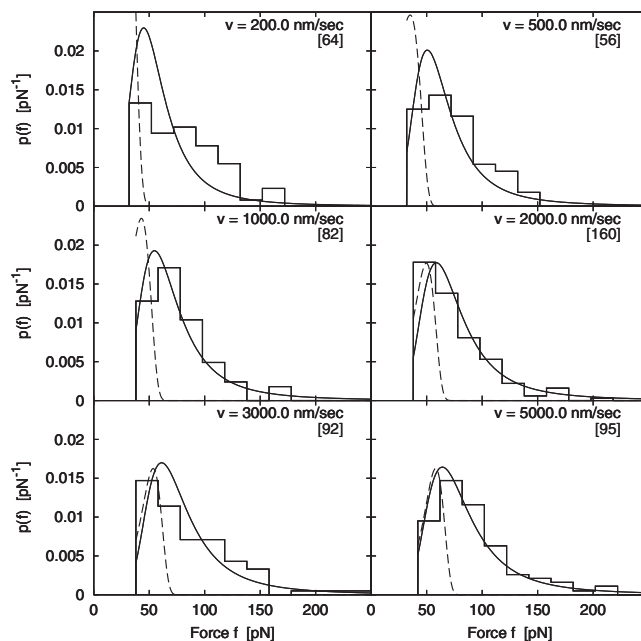

FIG. 6. Same rupture force data as in Fig. 5 but with velocity dependent threshold force: $f_{\min}=32$ pN for $v \leq 500$ nm/s, $f_{\min}=38$ pN for 1000 nm/s $\leq v \leq 3000$ nm/s, and $f_{\min}=42$ pN for $v=5000$ nm/s. The number of rupture events for each pulling velocity is indicated in brackets. (Solid lines) Maximum likelihood fit for the heterogeneous bond model. More details are given in the caption of Fig. 5. (Dashed lines) Theoretical distribution for the standard Bell model [Eqs. (1)–(6)] with parameters $k_0=1.43$ s^{-1} and $x_\beta=0.52$ nm rescaled by a factor 1/3. The force-extension master curve $F(s)$ has been used for the calculation of the rupture probability density according to Eqs. (4)–(6). The parameters have been estimated from the full (unfiltered) raw data set in [30] according to the standard Bell model under the assumption of a linear force extension characteristic.

[31–33] can explain the experimental findings.

On the other hand, the heterogeneous bond model has previously shown to be able to explain the splitting of the $-v \ln(n_v(f))$ curves. A maximum likelihood fit [36] for this model yields the most probable parameters $k_0^*=0.02$ s^{-1} [$\ln(k_0^* \times 1 \text{ s})=-3.86\pm1.55$], $\bar{x}_\beta^*=0.6\pm0.1$ nm, and $\sigma_x^*=0.23\pm0.04$ nm. For these parameters the heterogeneous bond model predicts approximately the same dependence of $-v \ln(n_v(f))$ on the pulling velocity as experimentally measured (Fig. 5). Finally in Fig. 6 the theoretical distributions are compared to the histograms of rupture forces measured at various pulling velocities. For all these velocities the theoretically predicted distribution is very similar to the experimentally measured one. Hence we can conclude that the heterogeneous bond model is able to explain the experimental findings.

The evaluation of the data without any filtering procedure and in the framework of the standard Bell model yielded the estimates for the model parameters: $k_0=1.43\pm0.45$ s^{-1} and $x_\beta=0.52\pm0.03$ nm [30]. It should be noted that for this evaluation only the dependence of the most probable rupture force on the pulling velocity has been used. Whereas the position of the potential barrier agrees very well with the

mean position of the barrier estimated above, the dissociation rate is two orders of magnitude larger. This clearly demonstrates the importance of consistent data analysis standards for extracting quantitative and comprehensive molecular data in a robust and independent manner. Assuming that the Bell model with these parameters is true, the distribution of rupture forces for the force-extension master curve and given force offsets can be calculated via Eqs. (4)–(6). Figure 6 shows clearly that these "expected" distributions explain the experimental findings only unsatisfactorily for all pulling velocities in agreement with our above conclusions.

V. SUMMARY AND CONCLUSIONS

In this paper we have described a data analysis method in the field of dynamic force spectroscopy on single ligand-receptor pairs. This method allows the careful processing and analysis of the force-distance characteristic in an automated manner. We have paid particular attention to guarantee that the assumptions made for the theoretical interpretation and evaluation of the experiment are really satisfied.

A common basic assumption in the context of theoretically interpreting and quantitatively evaluating single-molecule force spectroscopy experiments is that the force-extension characteristic of the considered molecular complex (including ligand, receptor, linkers, AFM, etc.) is always the same, i.e., it does not change from one pulling experiment to the next and does not depend on the pulling velocity. Here we have demonstrated by means of data from a representative experiment that this assumption is in fact not satisfied. In the absence of previously existing careful examinations of this assumption, we must conclude that similar discrepancies between theoretical models and experimental reality may be quite common and may require reconsideration of the experimental data and the hence deduced quantitative conclusions.

In a next step, we have put forward a selection procedure of rupture events such that this initial discrepancy between theory and experiment is remedied by construction. Yet, we found that this amendment is not sufficient to overcome yet another recently discovered inconsistency between experimental data and the common theoretical assumption that for any given instantaneous pulling force, the chemical bond between ligand and receptor always dissociates according to a unique instantaneous decay rate, independent of the pulling velocity or any other system property apart from the instantaneous pulling force.

A third often adopted assumption which is disproved by our present work is that the force-extension characteristic can be satisfactorily approximated by a linear behavior. Ideally, one would instead use a simple realistic model to describe the force-extension curve. Unfortunately, no such model exists which incorporates all components of the entire elastic entity composed of cantilever, linker, receptor, and ligand molecules. Realizing, however, that the parametrization of the curve is only an intermediate step in the evaluation, any function which satisfactorily describes the force-extension characteristic over the full range of pulling forces may be used. In our experiment second degree polynomials were sufficient. For other experiments more complicated functions may be necessary or for these systems physical models may exist (e.g., the freely jointed chain (FJC) or a worm-like chain (WLJ) model) where the model parameters provide interesting information about the system. To generalize our classification scheme to these cases only little modifications in the algorithm described in Sec. III A are necessary.

By using an extension of the usually applied model to evaluate single-molecule pulling experiments we were able to consistently explain the experimental findings from the representative experiment. In doing so the rate constant of dissociation, the distance between the potential minimum and the barrier, as well as the variations of this distance can be determined in a quantitative manner.

Finally, we like to point out that our method for processing and analyzing the force-distance characteristic has the potential to detect and analyze different subpopulations of the force-extension characteristic separately, which could give insights into different modes of binding between the receptor and the respective ligand.

ACKNOWLEDGMENT

[1] J. Zlatanova, S. M. Lindsay, and S. H. Leuba, Prog. Biophys. Mol. Biol. **74**, 37 (2000).

[2] R. Merkel *et al.*, Nature (London) **397**, 50 (1999).

[3] P. Hinterdorfer *et al.*, Proc. Natl. Acad. Sci. U.S.A. **93**, 3477 (1996).

[4] U. Dammer *et al.*, Biophys. J. **70**, 2437 (1996).

[5] R. Ros *et al.*, Proc. Natl. Acad. Sci. U.S.A. **95**, 7402 (1998).

[6] F. Schwesinger *et al.*, Proc. Natl. Acad. Sci. U.S.A. **97**, 9972 (2000).

[7] U. Dammer *et al.*, Science **267**, 1173 (1995).

[8] S. Garcia-Manyes *et al.*, J. Biol. Chem. **281**, 5992 (2006).

[9] B. Bonanni *et al.*, Biophys. J. **89**, 2783 (2005).

[10] A. Vinckier *et al.*, Biophys. J. **74**, 3256 (1998).

[11] J. Fritz, A. G. Katopodis, F. Kolbinger, and D. Anselmetti, Proc. Natl. Acad. Sci. U.S.A. **95**, 12283 (1998).

[12] F. W. Bartels *et al.*, J. Struct. Biol. **143**, 145 (2003).

[13] F. Kühner *et al.*, Biophys. J. **87**, 2683 (2004).

[14] B. Baumgarth *et al.*, Microbiology **151**, 259 (2005).

[15] R. Eckel *et al.*, Angew. Chem. Int. Ed. **44**, 3921 (2005).

[16] S. Zapotoczny *et al.*, Langmuir **18**, 6988 (2002).

[17] T. Auletta *et al.*, J. Am. Chem. Soc. **126**, 1577 (2004).

[18] R. Eckel *et al.*, Angew. Chem. Int. Ed. **44**, 484 (2005).

[19] E. L. Florin, V. T. Moy, and H. E. Gaub, Science **264**, 415 (1994).

[20] G. U. Lee, D. A. Kidwell, and R. J. Colton, Langmuir **10**, 354 (1994).

[21] E. Evans, K. Ritchie, and R. Merkel, Biophys. J. **68**, 2580 (1995).

[22] R. I. Litvinov, H. Shuman, J. S. Bennett, and J. W. Weisel, Proc. Natl. Acad. Sci. U.S.A. **99**, 7426 (2002).

[23] P. Hinterdorfer *et al.*, Single Mol. **1**, 99 (2000).

[24] W. Baumgartner, P. Hinterdorfer, and H. Schindler, Ultramicroscopy **82**, 85 (2000).

[25] C. Friedsam, A. K. Wehle, F. Kühner, and H. E. Gaub, J. Phys.: Condens. Matter **15**, S1709 (2003).

[26] M. Raible *et al.*, J. Biotechnol. **112**, 13 (2004).

[27] M. Raible *et al.*, Biophys. J. **90**, 3851 (2006).

[28] G. I. Bell, Science **200**, 618 (1978).

[29] E. Evans and K. Ritchie, Biophys. J. **72**, 1541 (1997).

[30] F. W. Bartels *et al.*, Biophys. J. **92**, 4391 (2007).

[31] O. K. Dudko, A. E. Filippov, J. Klafter, and M. Urbakh, Proc. Natl. Acad. Sci. U.S.A. **100**, 11378 (2003).

[32] G. Hummer and A. Szabo, Biophys. J. **85**, 5 (2003).

[33] O. K. Dudko, G. Hummer, and A. Szabo, Phys. Rev. Lett. **96**, 108101 (2006).

[34] R. Merkel, Phys. Rep. **346**, 343 (2001).

[35] M. Evstigneev and P. Reimann, Phys. Rev. E **68**, 045103 (2003).

[36] S. Getfert and P. Reimann, Phys. Rev. E **76**, 052901 (2007).

[37] T. V. Ratto *et al.*, Biophys. J. **86**, 2430 (2004).

[38] E. Thormann, P. L. Hansen, A. C. Simonsen, and O. G. Mouritsen, Colloids Surf., B **53**, 149 (2006).