PHYSICAL REVIEW E **68**, 045103(R) (2003)

Dynamic force spectroscopy: Optimized data analysis

Mykhaylo Evstigneev* and Peter Reimann
Universität Bielefeld, Fakultät für Physik, 33615 Bielefeld, Germany
(Received 16 July 2003; published 22 October 2003)

The forced rupture of single chemical bonds in biomolecular compounds (e.g., ligand-receptor systems) as observed in dynamic force spectroscopy experiments is addressed. An optimized method of data analysis is proposed. This method significantly outperforms the current standard one when applied to data from an idealized numerical computer simulation of an experiment with realistic parameter values. In particular, the force-free dissociation rate can be inferred with a considerably smaller statistical uncertainty and without the systematic overestimation of about 30%, which is shown to be inherent in the standard method.

DOI: 10.1103/PhysRevE.68.045103 PACS number(s): 02.50.−r

Dynamic force spectroscopy is a powerful and versatile method to characterize bond strengths of a variety of biomolecular compounds [1–5]. Using an atomic force microscope or some other experimental tool (see Table 1 of Ref. [6]), a single chemical bond (e.g., in a ligand-receptor complex) is exposed to a pulling force that increases linearly in time at a rate μ,

$$f(t)=\mu\, t, \tag{1}$$

until the bond breaks at a quite precisely measurable rupture force. In doing so, the biomolecule is permanently in contact with a thermal heat bath (usually a surrounding liquid) at a fixed temperature T. Upon repeating the experiment with the same compound and the same pulling rate μ one observes rupture forces that are statistically distributed over a rather wide range. Furthermore, for different pulling rates μ, one observes significantly different such statistical distributions. The theoretical challenge is to draw conclusions regarding the chemical bond under study from these experimental data. Milestones in this respect are due to Bell [7] and to Evans and Ritchie [2], forming the basis of the current standard method of data analysis and consisting of the following three main points.

(i) A rupture event is viewed as thermally activated decay of a metastable state across a potential barrier governed by a reaction kinetics of the form

$$\dot{p}(t)=-\nu(f(t))p(t), \tag{2}$$

where $p(t)$ is the probability of bond survival up to time t and $\nu(f)$ is the rate of decay in the presence of a pulling force f. The only assumption implicit in Eq. (2) is that intramolecular thermal relaxation processes are much faster than the instantaneous lifetime $1/\nu(t)$ of the bond and the time scale on which the applied force $f(t)$ notably changes. For the experimentally feasible pulling rates μ in Eq. (1), ranging from 10^{-1} to 10^{5} pN/sec [3], this condition, and hence Eq. (2) is very well satisfied in most cases.

(ii) Under the very same conditions, reaction rate theory [8] predicts an Arrhenius law for $\nu(f)$ of the form

$$\nu(f)=\omega(f)\exp\{-\Delta U(f)/k_BT\}, \tag{3}$$

where the potential barrier against decay $\Delta U(f)$ is large compared to the thermal energy k_BT and the prefactor $\omega(f)$ represents the intramolecular attempt frequency.

(iii) As long as the energy barrier separating the unbound and bound states is sufficiently sharp and deep and/or the applied force is sufficiently weak, one can approximately linearize $\Delta U(f)$ and neglect the f dependence of $\omega(f)$, with the result

$$\nu(f)=\nu_0\exp\{\alpha\, f\}, \tag{4}$$

where $\nu_0:=\nu(0)$ is the force-free dissociation rate of the chemical bond under study. Its determination is clearly of foremost interest in this context. The coefficient α times k_BT can be identified as the spatial distance between potential minimum and maximum (projected along the direction of the pulling force) and is therefore a second quantity of considerable interest.

The currently predominant method of estimating ν_0 and α from experimentally observed rupture forces is based on the above described theoretical insight in conjunction with the following line of reasoning. Taking into account Eqs. (1), (2), and (4), one readily finds [2] that the probability of rupture as a function of the acting force $f\geqslant 0$ is monotonically decreasing if the pulling rate μ is smaller than the critical value

$$\mu_0:=\nu_0/\alpha. \tag{5}$$

In other words, the most probable rupture force $f_*(\mu)$ is given by

$$f_*(\mu)=0 \quad \text{for} \quad \mu\leqslant\mu_0. \tag{6}$$

On the other hand, for $\mu>\mu_0$ the probability of rupture becomes a nonmonotonic function of the acting force f with a unique maximum at

$$f_*(\mu)=\frac{1}{\alpha}\ln\left(\frac{\alpha\mu}{\nu_0}\right) \quad \text{for} \quad \mu>\mu_0. \tag{7}$$

Hence, ν_0 and α can be determined by conducting experiments with several different pulling rates $\mu>\mu_0$. For each of them, the most probable rupture force $f_*(\mu)$ can be esti-

*Email address: mykhaylo@physik.uni-bielefeld.de

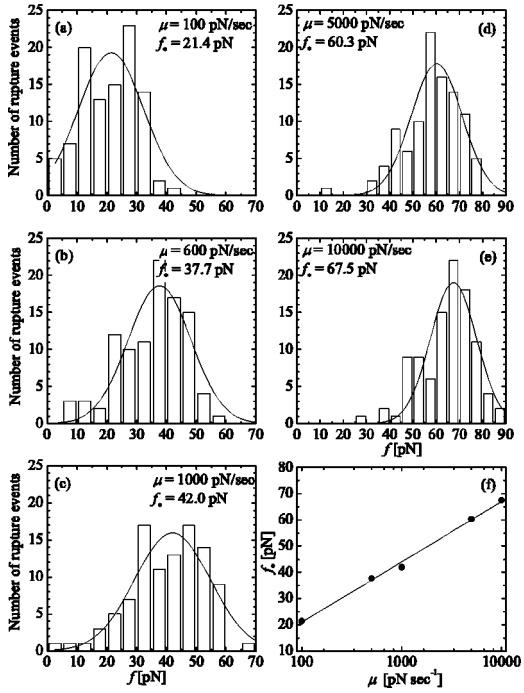

FIG. 1. (a)–(e) Number of rupture events versus rupture force f, represented as histograms with 5 pN bins, for five different pulling rates μ. For each μ, a total of $N_\mu = 100$ rupture events were randomly generated on a computer according to the probability density (19). In each plot, the solid line represents a Gaussian fit to the histogram, whose maximum is the indicated most probable rupture force $f_* = f_*(\mu)$. (f) Log-linear plot of the most probable rupture forces $f_*(\mu)$ versus pulling rates μ as obtained from (a)–(e). The linear best fit to these data points is indicated by the solid line, yielding for the fit parameters in Eq. (7) the estimates $\tilde{\nu}_0 \approx 1.24\ \mathrm{sec}^{-1}$ and $\tilde{\alpha} \approx 0.100\ \mathrm{pN}^{-1}$.

mated, usually by fitting a Gaussian to the observed distribution of rupture forces in the form of a histogram (cf. Fig. 1). Then, the resulting $f_*(\mu)$ values are plotted versus $\ln(\mu)$. Finally, a linear best fit yields an estimate for ν_0 and α. This procedure is henceforth referred to as *standard method*.

It has several obvious weak points. For instance, for typical values of the rupture parameters (see Table 3 in Ref. [5])

$$\nu_0 = 1\ \mathrm{sec}^{-1}, \quad \alpha = 0.1\ \mathrm{pN}^{-1} \tag{8}$$

the critical pulling rate in Eq. (5) is $\mu_0 = 10$ pN/sec. Hence, out of the experimentally feasible pulling rates μ between 10^{-1} and 10^5 pN/sec the first 2 decades are useless and the fitting regime is restricted to the remaining 3 decades. Moreover, μ_0 is not known *a priori*. Second, the uncontrolled theoretical approximation (4) is a key ingredient of the method. Third, one has the feeling that the method does not really make optimal use of the available data: a reduction of the statistical uncertainty by means of a more sophisticated approach seems possible. Finally, as we will show below, fitting the rupture force distribution by a Gaussian causes a systematic overestimation of ν_0 by about 30%.

In the following we propose an optimized method of rupture data analysis, which is free of the above shortcomings and is based on premises (1) and (2) only. In addition, in those exceptional cases where Eq. (2) is actually not satisfied, our method will allow us to recognize this fact.

Changing variables from t to f according to Eq. (1), kinetic equation (2) takes the form

$$d\,p_\mu(f)/df = -\nu(f)p_\mu(f)/\mu, \tag{9}$$

where $p_\mu(f) := \int_0^\infty dt p(t)\,\delta(t-(f/\mu))$ is the probability of bond survival up to a pulling force f at an arbitrary but fixed pulling speed μ. With the initial condition $p_\mu(0) = 1$ it follows that

$$p_\mu(f) = \exp\{-g(f)/\mu\}, \tag{10}$$

where we have introduced

$$g(f) := \int_0^f df'\ \nu(f'). \tag{11}$$

Our first observation is that once $g(f)$ is known, $\nu(f)$ follows immediately. Thus, we can focus on the determination of $g(f)$ from now on. Second, given a set of N_μ rupture forces f_n, $n = 1, \ldots, N_\mu$, at a fixed pulling rate μ, the best estimate for $p_\mu(f)$ that can be inferred from these data in the absence of any further *a priori* knowledge is

$$\tilde{p}_\mu(f) = \frac{1}{N_\mu} \sum_{n=1}^{N_\mu} \Theta(f_n - f), \tag{12}$$

where $\Theta(x)$ is the Heaviside step function [$\Theta(x<0) = 0$, $\Theta(x \geqslant 0) = 1$]. Furthermore, here and in the following, *a tilde indicates an estimate for the corresponding "true" quantity without tilde*, towards which it converges (with probability 1) for $N_\mu \to \infty$. However, note that this convergence is not uniform. Rather, for any fixed μ, the majority of rupture events f_n will sample a rather limited f interval (cf. Fig. 1). Only within this interval will an experimentally realistic finite number of pullings N_μ admit via Eq. (12) a reliable estimate

$$\tilde{g}_\mu(f) := -\mu \ln \tilde{p}_\mu(f) \tag{13}$$

for the "true" function $g(f)$ in Eq. (10).

We now come to the central point of our paper, namely, the simple observation that, according to Eq. (10), the function $-\mu \ln(p_\mu(f)) = g(f)$ is in fact independent of the pulling speed μ. By properly exploiting this universality, it should clearly be possible to reliably estimate $g(f)$ over a wide f range by combining data for several different pulling speeds μ. The technical details of how to do this in a way that makes optimal use of the information encapsulated in the available experimental data is the content of the following.

Consider an arbitrary but fixed $f > 0$. Then, for each pulling rate μ an estimate $\tilde{g}_\mu(f)$ for the true $g(f)$ follows from Eqs. (12) and (13). Its reliability is quantified by the variance $\tilde{\sigma}^2_{\tilde{g}_\mu(f)}$, whose explicit determination will be given shortly. With this amount of information at our disposition, according to the so-called method of weighted averages [9], the best

guess for the true $g(f)$ is represented by that argument x which minimizes the weighted sum of square deviations $\Sigma_\mu[x-\tilde{g}_\mu(f)]^2/\tilde{\sigma}^2_{\tilde{g}_\mu(f)}$, where Σ_μ indicates a summation over all pulling rates μ. In other words, this best guess for $g(f)$ is given by the weighted average

$$\tilde{g}(f)=\sum_\mu c_\mu(f)\tilde{g}_\mu(f), \tag{14}$$

$$c_\mu(f):=\frac{1}{\tilde{\sigma}^2_{\tilde{g}_\mu(f)}}\Big/\sum_\mu \frac{1}{\tilde{\sigma}^2_{\tilde{g}_\mu(f)}}. \tag{15}$$

In order to determine the variances $\tilde{\sigma}^2_{\tilde{g}_\mu(f)}$, we consider the number $N_\mu(f):=p_\mu(f)N_\mu$ of bonds surviving up to the pulling force f. It follows that $N_\mu(0)=N_\mu$ and that for any fixed f, μ, and N_μ, the number $N_\mu(f)$ is distributed binomially according to

$$W(N_\mu(f))=\frac{[p_\mu(f)]^{N_\mu(f)}[1-p_\mu(f)]^{N_\mu-N_\mu(f)}N_\mu!}{N_\mu(f)![N_\mu-N_\mu(f)]!},$$

implying for the associated variance $\sigma^2_{N_\mu(f)}$ the result

$$\sigma^2_{N_\mu(f)}=N_\mu p_\mu(f)[1-p_\mu(f)]. \tag{16}$$

An estimate $\tilde{\sigma}^2_{N_\mu(f)}$ for the true $\sigma^2_{N_\mu(f)}$ follows by replacing the true but unknown $p_\mu(f)$ in Eq. (16) by the estimate $\exp[-\tilde{g}(f)/\mu]$ [see Eq. (10)]. Then, by exploiting the error propagation law $\tilde{\sigma}^2_{\tilde{g}_\mu(f)}=[d\tilde{g}_\mu(f)/dN_\mu(f)]^2\,\tilde{\sigma}^2_{N_\mu(f)}$ and Eq. (13) one finds for the coefficients c_μ in Eq. (15) the result

$$c_\mu(f)=\frac{N_\mu\tilde{p}^2_\mu(f)e^{\tilde{g}(f)/\mu}}{\mu^2[1-e^{-\tilde{g}(f)/\mu}]}\sigma^2_{\tilde{g}(f)}, \tag{17}$$

$$\sigma^2_{\tilde{g}(f)}:=\left(\sum_\mu \frac{N_\mu\tilde{p}^2_\mu(f)e^{\tilde{g}(f)/\mu}}{\mu^2[1-e^{-\tilde{g}(f)/\mu}]}\right)^{-1}. \tag{18}$$

Finally, by taking into account Eq. (15) one readily verifies that $\sigma^2_{\tilde{g}(f)}$ from Eq. (18) indeed coincides with the variance $\Sigma_\mu c^2_\mu\,\tilde{\sigma}^2_{\tilde{g}_\mu(f)}$ describing the statistical uncertainty of $\tilde{g}(f)$ in Eq. (14).

In practice, the above method boils down to the following two steps. First, the functions $\tilde{p}_\mu(f)$ and $\tilde{g}_\mu(f)$ are determined from the experimentally observed rupture forces f_n, $n=1,\ldots,N_\mu$, for different pulling speeds μ according to Eqs. (12) and (13). Second, the weighted averages (14) and (17) are evaluated. Since the coefficients $c_\mu(f)$ in Eq. (14) depend themselves on the unknown quantity $\tilde{g}(f)$ according to Eq. (17), we are dealing with a transcendental equation for $\tilde{g}(f)$ for any fixed f value. Among many other well-known methods to solve such an equation, one particularly simple way is to iteratively update the value of $\tilde{g}(f)$ on the basis of Eq. (14) until stationarity is reached. The result is an estimate $\tilde{g}(f)$ for the true function $g(f)$ in Eq. (10) together with its statistical uncertainty (18). Once a specific f-dependence of the rate $\nu(f)$, as, e.g., in Eq. (4), is assumed, it is possible to decide using Eq. (11) whether this assumption is compatible with the measured data and to fit the parameters. In addition, one can check whether the functions $\tilde{g}_\mu(f)$ for different pulling rates μ indeed collapse within their statistical uncertainties onto a single master curve. If this is not the case, it follows that the basic kinetic law (2) is not satisfied for the bond rupture mechanism at hand.

In order to test the above procedure and compare it with the standard method, we generated the necessary input data by means of an idealized numerical computer simulation of an actual experiment. Exploiting Eqs. (4), (10), (11), one readily finds the explicit expression for the probability density $-d\,p_\mu(f)/df$ that a rupture event occurs at a given force f:

$$-\frac{d\,p_\mu(f)}{df}=\frac{\nu_0}{\mu}\exp\left\{\alpha f-\frac{\nu_0}{\alpha\,\mu}(e^{\alpha f}-1)\right\}. \tag{19}$$

According to this probability density with the parameters α and ν_0 given by Eq. (8), we have sampled rupture forces f_n, $n=1,\ldots,N_\mu$, for different pulling rates μ on the computer. As compared to real, experimentally measured rupture forces, these numerically generated data have the advantage of being completely clean and the underlying statistical probabilities are exactly known. Thus, our ''numerical experiment'' offers an ideal means for testing the efficiency of various approaches to data analysis. In particular, the generally uncontrolled theoretical approximation (4) is exactly satisfied by our test data, thus representing a considerable bias in favor of the standard method. The ultimate goal will then be to recover the preset, i.e., known, parameters ν_0 and α in Eq. (4) from the numerically generated rupture data using the standard method and the proposed one.

With the parameter values of Eq. (8), we have simulated $N_\mu=100$ rupture events for each of five pulling rates $\mu=100$, 500, 1000, 5000, and 10 000 pN/sec. Moreover, to arrange the data into histograms, as required by the standard method, the force axis was divided into 5 pN bins. All these numbers represent typical values in real experimental investigations [3,5].

The results obtained by the standard method are summarized in Fig. 1. In particular, the most probable rupture forces in Fig. 1(f) indeed satisfy to a very good approximation the linear dependence on $\ln\mu$ predicted by Eq. (7). The resulting estimates for the true parameter values in Eq. (8) are $\tilde{\nu}_0\approx 1.24\ \mathrm{sec}^{-1}$ and $\tilde{\alpha}\approx 0.100\ \mathrm{pN}^{-1}$.

The results obtained by our method are depicted in Fig. 2. The functions $\tilde{g}_\mu(f)$ for different pulling rates μ indeed collapse within their statistical uncertainties onto a single master curve. Further, the estimate $\tilde{g}(f)$ for the true function $g(f)$ is rather good over a wide range of pulling forces f. Finally, the resulting estimate $\tilde{\nu}_0\approx 0.95\ \mathrm{sec}^{-1}$ is considerably closer to the true parameter value in Eq. (8) than the one resulting from the standard method, while the estimate $\tilde{\alpha}\approx 0.098\ \mathrm{pN}^{-1}$ is slightly worse.

To gain further insight into the accuracy of both methods, we have repeated the above described numerical experiment 20 times. Taking averages over all 20 realizations,

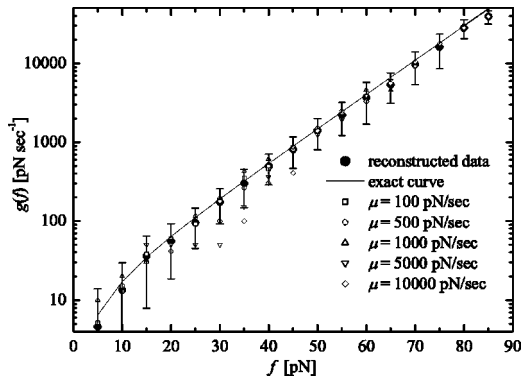

FIG. 2. The function $\tilde{g}_{\mu}(f)$ versus force f obtained from the same numerical simulation data as in Fig. 1 according to Eqs. (12) and (13). For the sake of clarity, only a small set of discrete f values are plotted with different symbols for the different pulling rates μ (some data points are outside the depicted parameter range). The resulting estimate for $\tilde{g}(f)$ according to our method (14), (17) is represented as filled circles. The true function $g(f)$ from Eqs. (4), (8), (11) is indicated as a solid line. The error bars were determined via Eq. (18) and then enlarged by a factor 5 for the sake of better visibility. Considering in the function $g(f)$ from Eqs. (4) and (11) the values of ν_0 and α not as fixed by Eq. (8) but rather as fitting parameters, a best fit to the solid dots (on a logarithmic scale) yields the estimates $\tilde{\nu}_0 \approx 0.95\ \mathrm{sec}^{-1}$ and $\tilde{\alpha} \approx 0.098\ \mathrm{pN}^{-1}$.

the standard method yielded $\tilde{\nu}_0 = 1.30 \pm 0.05\ \mathrm{sec}^{-1}$ and $\tilde{\alpha} = 0.101 \pm 0.001\ \mathrm{pN}^{-1}$. Our method yielded much more accurate estimates $\tilde{\nu}_0 = 0.97 \pm 0.01\ \mathrm{sec}^{-1}$, and for $\tilde{\alpha} = 0.1 \pm 0.0003\ \mathrm{pN}^{-1}$.

Our first observation is that the former estimate for ν_0 exceeds the true value by six standard deviations, indicating a systematic rather than a statistical error. We have found a similar 30% systematic overestimation of ν_0 also for several other realistic choices of the parameters ν_0, α, and N_{μ}. This systematic deviation is indeed universal, i.e., independent on the concrete values of ν_0 and α, because these parameters can always be set to unity by a proper choice of time and force scales. In essence it is rooted in the following property: for large μ the probability density (18) is an asymmetric function with negative skewness. Its fit by a symmetric Gaussian curve will always somewhat underestimate $f_*(\mu)$ and, according to Eq. (7), systematically overestimate ν_0.

Our second observation is that also the statistical uncertainties are significantly larger for the standard method in comparison with our method. This result verifies our above mentioned suspicion that the standard method does not really make optimal use of the available data.

In conclusion, in our present work we have restricted ourselves to a careful description and comparison of our method and the currently dominating standard method by using perfectly ''clean'' data from a computer simulation of the simplest possible real system. Obvious next steps include the following. (i) Comparing appropriate extensions of those methods with numerical simulation data for more complicated and realistic rate laws than in Eq. (4) as well as with real experimental data. (ii) In principle, the basic reaction kinetics (2) may be invalidated if the decay proceeds along different reaction pathways for different pulling speeds [10], due to contribution of recombination processes [11], or because of the influence of an intermediate energy barrier [12]. While our present method is able to recognize such a complication, its further quantitative analysis remains an open question.

Helpful discussions with R. Ros, F. Bartels, R. Eckel, D. Anselmetti, R. Eichhorn, and especially R. Merkel are gratefully acknowledged. This work was supported by Deutsche Forschungsgemeinschaft under Grant No. SFB 613 (Teilprojekt K7), the Alexander von Humboldt–Stiftung, and the ESF-program STOCHDYN.

[1] E.-L. Florin, V.T. Moy, and H.E. Gaub, Science **264**, 415 (1994).
[2] E. Evans and K. Ritchie, Biophys. J. **72**, 1541 (1997).
[3] E. Evans, Biophys. Chem. **82**, 83 (1999).
[4] F. Schwesinger, R. Ros, T. Strunz, D. Anselmetti, H.-J. Güntherodt, A. Honegger, L. Jermutus, L. Tiefenauer, and A. Plückthun, Proc. Natl. Acad. Sci. U.S.A. **97**, 9972 (2000).
[5] R. Merkel, Phys. Rep. **346**, 343 (2001).
[6] H. Clausen-Schaumann, M. Seitz, R. Krautbauer, and H.E. Gaub, Curr. Opin. Chem. Biol. **4**, 524 (2000).
[7] G.I. Bell, Science **200**, 618 (1978).
[8] P. Hänggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990).
[9] J.R. Taylor, *An Introduction to Error Analysis* (University Science Books, Mill Valley, CA, 1982).
[10] D. Bartolo, I. Derényi, and A. Ajdari, Phys. Rev. E **65**, 051910 (2002).
[11] U. Seifert, Europhys. Lett. **58**, 792 (2002).
[12] T. Strunz, K. Oroszlan, I. Schumakovitch, H.-J. Güntherodt, and M. Hegner, Biophys. J. **79**, 1206 (2000).