Maximum Likelihood Estimation of Multiple-Bond Kinetics from Single-Molecule Pulling Experiments

E. J. Hukkanen, J. A. Wieland, D. E. Leckband, and R. D. Braatz
University of Illinois at Urbana-Champaign

Abstract— **Kinetic parameters associated with bond dissociation have been identified from experiments in which mechanical forces exerted by atomic force microscopy or laser tweezers are used to break bonds in single molecules. Analysis of a single-bond microscopic model indicates that the breakage frequency distribution is most sensitive to the value of the distance to the free-energy minimum barrier and least sensitive to the molecular spring constant. This analysis plus limitations in the single-bond microscopic model motivate a double-bond microscopic model that requires only one additional kinetic parameter. In single-molecule experiments for the neural cellular adhesion molecule (NCAM), the double-bond microscopic model can more accurately describe the breakage frequency distribution for both high and low values of the applied force, which is consistent with the molecular structure of the NCAM determined by a surface force apparatus. This provides a systematic procedure for gaining information on the molecular structure of single molecules by analyzing breakage frequency distributions measured during single-molecule pulling experiments.**

I. Introduction

The extraction of kinetic information from single-molecule pulling experiments has been investigated by several researchers [1], [2], [3], [4], [5]. Most of the models are based on Bell's early cell adhesion model [6], with thorough explanations and derivations of the models reported in the aforementioned papers. Usually the kinetic parameters are estimated using linear least squares and the most probable force at which a rupture event occurs (i.e., the force that produces the most events). This may yield inaccurate or misleading parameter estimates, since the shape of the distribution is not considered. This motivates the use of the full distribution of events as a function of the applied force. Further, the occurrence of multiple bonds greatly affects the parameter estimates, especially if the estimation method is not tailored to handle this phenomenon. A multiple-bond model has been proposed for systems that exhibit such behavior [1], but as a function of the most probable force.

Here the kinetic parameters associated with the dissociation of adhesive bonds at both high and low loading rates are determined by maximum likelihood estimation applied to the breakage frequency distributions measured using atomic force microscopy (AFM) applied to the pulling of the neural cellular adhesion molecule (NCAM). Sensitivity analysis and multivariate statistical analysis are used to compute confidence intervals on the individual kinetic parameters for the microscopic model [4], which was the most consistent single-bond model for the observed distributions. These analyses motivate a double-bond microscopic model that more accurately describes the breakage frequency distributions for both high and low values of the applied force, using only one additional parameter, the distance to a free-energy barrier minimum, beyond the single-bond microscopic model. An F-test indicates that the double-bond microscopic model is statistically justified for the NCAM breakage frequency distributions, which is consistent with the molecular structure of NCAM determined from surface force apparatus (SFA) experiments [7]. This provides a systematic procedure for determining the number of bonds that can dissociate in a molecule during single-molecule pulling experiments, from breakage frequency distributions.

Corresponding author. Tel.: +1-217-333-5073; Fax: +1-217-333-5052; E-mail address: braatz@uiuc.edu (R. D. Braatz)

II. Theory

The cumulative distribution for the microscopic model [4] for rupture of adhesion bonds is

$$P_m(f) = 1 - \exp\left[-\frac{k_o e^{\frac{-\kappa_s x^2}{2k_B T}}}{r_f \frac{x}{k_B T}} \left(\frac{\kappa_m}{\kappa} \right)^{3/2} \left(e^{\left(\frac{fx}{k_B T} - \frac{f^2}{2\kappa k_B T} \right)} - 1 \right) \right] \tag{1}$$

where k_o is the intrinsic rate constant, κ_s is the spring constant, x is the distance from the free-energy minimum to the barrier, k_B is Boltzmann's constant, T is temperature (K), r_f is the loading rate, κ_m is the molecular spring constant, and κ is the effective force constant defined by

$$\kappa = \kappa_s + \kappa_m. \tag{2}$$

Analyzing the cumulative distribution results in more accurate parameter estimates than using a histogram for the distribution, which has binning errors. Analysis of the experimental distribution data indicated that it was acceptable to assume that the measurement errors are normally distributed and independent of each other (i.e., the measurement error covariance matrix $\mathbf{V}_i$ for the i^{th} experiment is defined by $\mathbf{V}_i = \sigma^2 \mathbf{I}$ where σ^2 is the variance). Under this assumption, parameter estimates computed from the following approach are both maximum likelihood and minimum variance estimates [8].

A. Microscopic Model

The microscopic model has three unknown parameters, k_o, x, and κ_m, that are stacked into a parameter vector:

$$\theta = [k_o, x, \kappa_m]^T \tag{3}$$

sensitivities	low loading rate r_f	high loading rate r_f
S_{k_o}	5.70×10^{-3}	2.86×10^{-4}
S_{κ_m}	8.25×10^{-5}	3.60×10^{-6}
S_x	2.22×10^{-2}	5.57×10^{-2}

TABLE I

PARAMETER SENSITIVITY ANALYSIS FOR THE MICROSCOPIC MODEL WITH OBJECTIVE FUNCTION (5).

The maximum likelihood parameters are computed by solving the optimization problem:

$$\min_{\theta} \Phi, \tag{4}$$

where

$$\Phi = \sum_{i=1}^{N} \sum_{j=1}^{M_i} \frac{1}{M_i} \left(P^{exp}(f_{i,j}) - P^{sim}(f_{i,j}) \right)^2, \tag{5}$$

M_i is the number of measurements collected in the ith experiment, $f_{i,j}$ is the jth measurement in the ith experiment, and N is the total number of experiments. This optimization was solved numerically using off-the-shelf sequential quadratic programming software [9], with global optimality verified by finely gridding each parameter.

B. Parameter Sensitivity Analysis

Parameter sensitivity analysis [8] quantifies the effect of perturbations in the parameters on the output variables (in this case, the cumulative distribution of the breakage frequencies). The parameter sensitivities determined analytically for the objective function (5) are reported in Table I for some representative parameter estimates. In the microscopic model, the molecular spring constant κ_m is the least sensitive parameter for both low and high loading rates. The distance to the free-energy barrier, x, is the most sensitive parameter for both low and high loading rates. Theoretically, estimates of most sensitive parameters from AFM data should be the most accurate, as quantified by smaller confidence intervals [8]. This agrees with intuition, that parameters with a low effect on the measured profiles should be difficult to determine accurately from the measurements. From Table I, it is clear that estimates of the molecular spring constant κ_m from AFM data will be much less accurate than estimates of the distance to the free-energy barrier x.

C. Double-Bond Microscopic Model

In the experimental section it is observed that the breakage frequency distributions are not accurately described by the single-bond microscopic model for large values of the applied force. Here a revision is made to the microscopic model that includes the effect of double-bond ruptures on the cumulative distribution (the extension to more than two bond ruptures is straightforward):

$$P_D(f) = \frac{P_d(f)}{P_d(f_{max})} \tag{6}$$

where the cumulative distribution $P_d(f)$ evaluated at the force, f, and the maximum force, f_{max}, is

$$P_d(f) = 2 - \exp\left[-\frac{k_o e^{\frac{-\kappa_s x_1^2}{2k_BT}}}{r_f \frac{x_1}{k_BT} \left(\frac{\kappa_m}{\kappa}\right)^{3/2}} \left(e^{\left(\frac{f x_1}{k_BT} - \frac{f^2}{2\kappa k_BT} \right)} - 1 \right) \right] - \exp\left[-\frac{k_o e^{\frac{-\kappa_s x_2^2}{2k_BT}}}{r_f \frac{x_2}{k_BT} \left(\frac{\kappa_m}{\kappa}\right)^{3/2}} \left(e^{\left(\frac{f x_2}{k_BT} - \frac{f^2}{2\kappa k_BT} \right)} - 1 \right) \right] \tag{7}$$

where k_o intrinsic rate constant, x_i is the i^{th} distance from the free-energy minimum to the barrier, κ_m is the molecular spring constant, and κ is the effective force constant defined by

$$\kappa = \kappa_s + \kappa_m. \tag{8}$$

Because the distance from the free-energy minimum to the barrier, x, was previously determined to be the most sensitive parameter for the microscopic model, only this parameter was duplicated to define this double-bond microscopic model.

The parameters, θ_D, for the double-bond microscopic model are

$$\theta_D = [k_o, \kappa_m, x_1, x_2]^T, \tag{9}$$

which are computed as the solution to the optimization problem

$$\min_{\theta_D,\ x_1 > x_2} \sum_{j=1}^{M_i} \frac{1}{M_i} \left(P_D^{exp}(f_{i,j}) - P_D^{sim}(f_{i,j}) \right)^2 \quad \text{for } i = 1, \ldots, N. \tag{10}$$

III. MATERIALS AND EXPERIMENTS

A. Preparation of Substrate and AFM Cantilevers for Experiments Involving NCAM

Substrates were glass microscope slides (Fischer Scientific) cut into 18 mm square pieces. Commercial Si_3N_4 V-shaped contact cantilevers with gold reflective coating were purchased from Digital Instruments. Gold coated tips and substrates are placed into an ethanolic thiol solution consisting of 1 mM 1, 8-octanedithiol and 10 mM 6-mercapto hexan-1-ol for approx. 18 hours. They are then washed with ethanol, dried with argon, and placed in a phosphate buffer solution (pH 7.4) containing 1 mg/mL NHS-PEG-MAL for 20 min. This allows any free thiol groups to react with the maleimide containing portion of the PEG spacer. The tip and surface is then washed with buffer solution and placed into the AFM cell. A 0.06 mg/ml protein solution is added to the cell. The cantilever is then mounted on the AFM apparatus, submerging it underneath the protein solution. Both the tip and substrate are allowed to incubate for 90 min. This allows the N-hydroxysuccinimide end of the PEG spacer to react with any available amine groups found on the NCAM. The cell is then flushed 10 times with buffer solution rinsing away any unreacted protein

loading rate r_f (pN/s)	number of rupture events
781	198
7423	123

TABLE II
CONDITIONS FOR NCAM EXPERIMENTS.

remaining in solution. Figures 1 and 2 show schematic representations of the tip and surface modification.

All NCAM experiments were performed using soluble NCAM with C-terminal oligohistidine tag and molecular weight of 120 kD expressed in Chinese hamster ovary cells. For details on the production and purification of NCAM refer to work performed by Johnson et al. [7].

B. AFM Set-Up

All force probe measurements were obtained using a commercial AFM apparatus (Pico AFM, Molecular Imaging) with commercial controller and data acquisition electronics (Digital Instruments). All experiments were performed at room temperature (21°C). Loading rates for the NCAM experiments ranged from approximately 400-7500 pN/s with 2000 force extension curves obtained per loading rate. The AFM cell was translated after every 500 measurements to allow the tip to sample a new area. The loading rate was calculated by multiplying the tip velocity (frequency times distance traveled per cycle) by the slope of the force-distance curve, k_s. See Figure 3 for an example of a typical force-distance curve obtained during a lipid pull out experiment.

IV. RESULTS AND DISCUSSION

Maximum likelihood parameter estimates and confidence intervals for the single-bond and double-bond microscopic models were determined for NCAM experiments at two different loading rates (see Table II). Table III reports the kinetic parameters, confidence intervals, and total residuals for the single-bond microscopic model. The total residual is defined as

$$R = \arg\min_{\theta} \Phi. \tag{11}$$

Large confidence intervals were observed for both loading rates, which is consistent with the inability of the single-bond microscopic model to describe the shape of the experimental breakage distributions for high and low values of the applied force (see Figures 4-5). The confidence interval for the molecular spring constant κ_m is the largest, which is consistent with the low sensitivity of the breakage frequency distribution to that parameter.

The kinetic parameters, confidence intervals, and total residuals for the double-bond microscopic model are reported in Table IV. For both loading rates, the distances to the free-energy barrier, x_1 and x_2, are different by a factor of 2-3. The confidence intervals and total residual are much smaller for the double-bond microscopic model

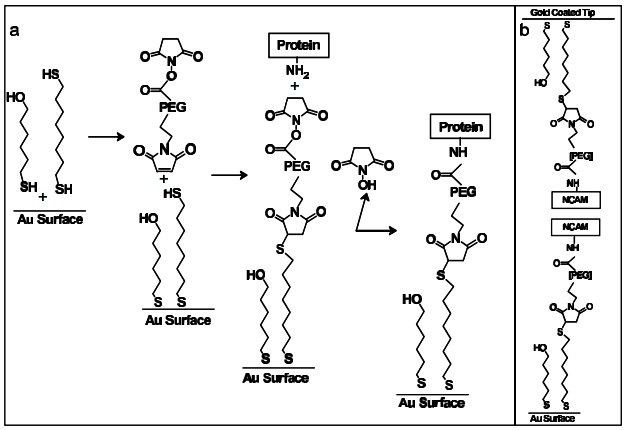

Fig. 1. a) Schematic showing step by step modification of gold coated tip and substrate. b) Full schematic of tip and surface orientation during the experiment.

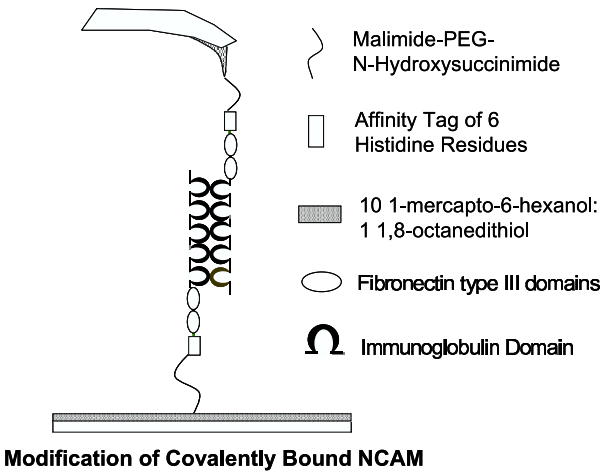

Fig. 2. Schematic showing the orientation of the modified tip and substrate during the experiment. Note that any free amine group on the protein can interact with the PEG spacer allowing many orientations of the protein. The most ideal case is pictured above.

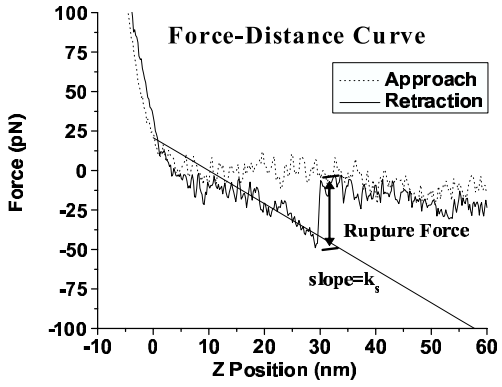

Fig. 3. Typical force-distance curve from the lipid pull out experiment showing rupture force and slope of the curve, k_s.

r_f (pN/s)	k_o (s^{-1})	κ_m (pN/nm)	x (nm)	total residual
781	1.868±0.869	172±651	0.3459±0.2531	0.9157
7423	18.08±5.24	1475±1465	0.0699±0.0301	0.4194

TABLE III

PARAMETER ESTIMATES AND CONFIDENCE INTERVALS, REPORTED AT 95%, USING THE MICROSCOPIC MODEL FOR EACH SET OF NCAM EXPERIMENTS.

θ	781 pN/s	7423 pN/s
k_o (s^{-1})	0.796 ± 0.188	5.402 ± 1.460
κ_m (pN/nm)	182 ± 320	1380 ± 548
x_1 (nm)	0.730 ± 0.142	0.188 ± 0.0225
x_2 (nm)	0.337 ± 0.108	0.0813 ± 0.0160
total residual	0.2906	0.0687

TABLE IV

PARAMETER ESTIMATES AND CONFIDENCE INTERVALS, REPORTED AT 95%, USING THE DOUBLE-BOND MICROSCOPIC MODEL (7) FOR INDIVIDUAL NCAM EXPERIMENTS.

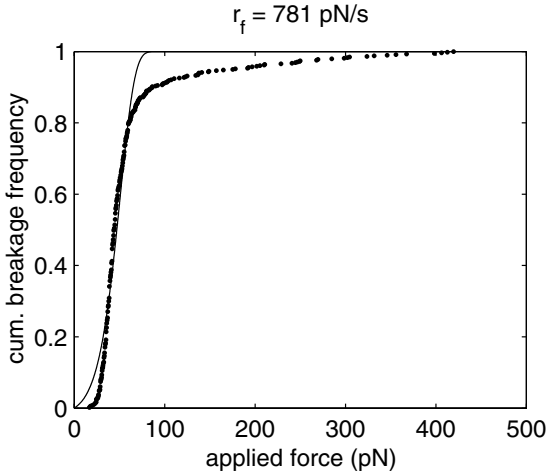

Fig. 4. Comparison of NCAM experimental and microscopic cumulative distributions for $r_f = 781$ pN/s.

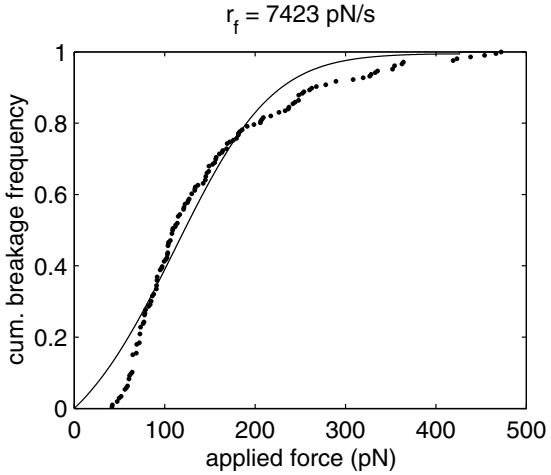

Fig. 5. Comparison of NCAM experimental and microscopic cumulative distributions for $r_f = 7423$ pN/s.

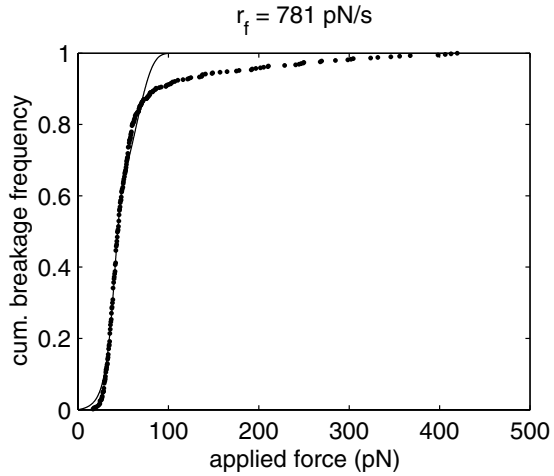

Fig. 6. Comparison of NCAM experimental and double-bond microscopic model distributions for a loading rate of $r_f = 781$ pN/s.

than for the single-bond microscopic model. The relative magnitudes of the confidence intervals are much smaller for the distances to the free-energy barrier, x_i, than for the other two parameters, which is consistent with the sensitivity analysis. The total residual is a factor of 3 and 10 smaller for the double-bond model than for the single-bond model for the loading rates of $r_f = 781$ pN/s and $r_f = 7423$ pN/s, respectively.

For the loading rate of $r_f = 781$ pN/s, the double-bond microscopic model provides a much more accurate fit to the experimental breakage frequency distribution for small values of the applied force (see Figure 6), resulting in smaller confidence intervals and total residual than the single-bond model (compare Tables III and IV). For this low loading rate, neither model is able to describe the long tail for high values of the applied force, suggesting that additional physical phenomena are relevant for this loading rate. The distances to the free-energy minimum barrier x_i are consistent with the molecular structure of NCAM. The total length of the five Ig domains of NCAM is approximately 18 nm, whereas x_i is approximately 0.08-0.73 nm. This indicates that the protein is relatively rigid, which was observed in SFA experiments [7].

For the higher loading rate of $r_f = 7423$ pN/s, the

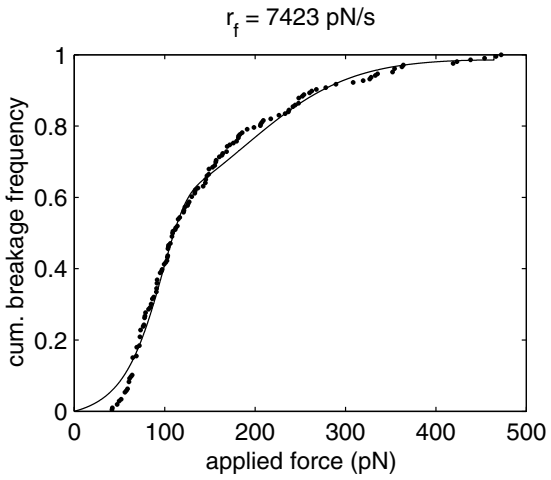

Fig. 7. Comparison of NCAM experimental and double-bond microscopic model distributions for a loading rate of $r_f = 7423$ pN/s.

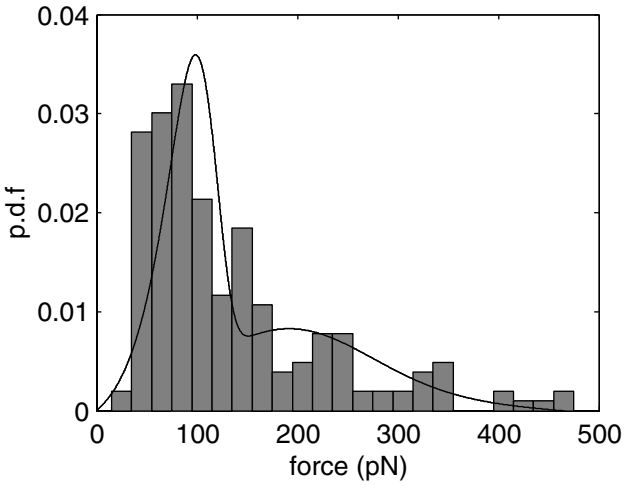

Fig. 8. Comparison of the differential distribution (12) and the experimental histogram for NCAM data for a loading rate of $r_f = 7423$ pN/s.

double-bond microscopic model provides a much more accurate fit to the experimental breakage frequency distribution for nearly all values of the applied force (see Figure 7). For this loading rate the confidence intervals of the double-bond microscopic model are approximately a factor of 3 smaller than for the single-bond model. For this loading rate the statistical analysis suggests that this double-bond microscopic model is able to capture the physical phenomena associated with the pulling of the NCAM molecules in these AFM experiments.

A comparison of the differential distribution determined analytically from (6) with a histogram constructed from experimental data (see Figure 8) illustrates the weakness of the common practice of fitting the differential distribution to histograms. The differential distribution is defined by

$$p_d(f) = \frac{dP_D(f)}{df}. \tag{12}$$

For the higher loading rate of $r_f = 7423$ pN/s, this differential distribution captures the long tail in the experimental data. It is much easier to visualize the two peaks in the model distribution, corresponding to the two bonds of different energies, than from the histogram of the data (see Figure 8). Fitting the kinetic parameters to the cumulative distribution instead of the differential distribution avoids the error associated with the binning of data that occurs when constructing a histogram.

The double-bond microscopic model only introduces one additional parameter over the single-bond microscopic model. The sensitivity analysis and previous statistical analyses indicated that considering two separate free-energy barrier distances x_i, instead of one, was the most promising extension to the single-bond model. Although additional kinetic parameters k_{io} and κ_{im} for each bond could have been considered, attempting to estimate a second molecular spring constant would lead to extremely large confidence

θ	7423 pN/s
k_{1o} (s^{-1})	5.20 ± 1.94
k_{2o} (s^{-1})	5.70 ± 4.06
κ_m (pN/nm)	1440 ± 1000
x_1 (nm)	0.189 ± 0.0239
x_2 (nm)	0.0791 ± 0.0300
total residual	0.0685

TABLE V

PARAMETER ESTIMATES AND CONFIDENCE INTERVALS, REPORTED AT 95%, USING THE DOUBLE-BOND MICROSCOPIC MODEL (14) FOR INDIVIDUAL NCAM EXPERIMENTS.

intervals since sensitivity analysis indicates that this parameter is much less sensitive than the other parameters. If the kinetic parameter vector,

$$\theta = [k_{1o}, k_{2o}, \kappa_m, x_1, x_2]^T, \tag{13}$$

is considered, the cumulative distribution, $P_d(f)$, of this double-bond model is

$$P_d(f) = 2 - \exp\left[-\frac{k_{1o}e^{\frac{-\kappa_s x_1^2}{2k_BT}}}{r_f \frac{x_1}{k_BT}\left(\frac{\kappa_m}{\kappa}\right)^{3/2}}\left(e^{\left(\frac{fx_1}{k_BT}-\frac{f^2}{2\kappa k_BT}\right)}-1\right)\right] - \exp\left[-\frac{k_{2o}e^{\frac{-\kappa_s x_2^2}{2k_BT}}}{r_f \frac{x_2}{k_BT}\left(\frac{\kappa_m}{\kappa}\right)^{3/2}}\left(e^{\left(\frac{fx_2}{k_BT}-\frac{f^2}{2\kappa k_BT}\right)}-1\right)\right]. \tag{14}$$

A simple quantification of the significance of the fifth parameter, k_{2o}, is achieved by computing the total residual (10) for varying k_{2o} (with the remaining parameters in Table IV fixed) for $r_f = 7423$ pN/s.

Kinetic parameters and confidence intervals were estimated for (14) (see Table V). The nominal parameter

estimates for this model are very similar to those for the original double-bond model that fixes $k_{1o} = k_{2o}$ (see Table IV). The confidence intervals for this extended double-bond model, however, are significantly larger for the intrinsic rate constants, k_{1o} and k_{2o}, and the molecular spring constant, κ_m. The original double-bond microscopic model (7) yields a residual of 0.0687 whereas the additional intrinsic rate constant, k_{2o}, does not significantly improve the total residual, where (14) yields a residual of 0.0685.

The F-test is used to determine whether a model that has additional parameters is statistically justified. The statistic F is defined by [8]

$$F = \frac{\Delta R/q}{R/(n-p)} \tag{15}$$

where R is the total residual of the proposed model, ΔR is the difference in total residual of the proposed model and prior model, n is the number of measurements, p is the number of parameters of the prior model, and q is the number of additional parameters required for the proposed model. The proposed model is justified with 100(1-α)% confidence if $F > F_\alpha(q, n-p)$, were some values of the F-distribution, $F_\alpha(q, n-p)$, are reported in Table VI with the results of applying the F-test to the single- and double-bond microscopic models. The double-bond model (7) that introduces one additional parameter, a distance to the free-energy minimum barrier, is statistically justified with >99.5% confidence. On the other hand, the double-bond model (14) that introduces two additional parameters does not meet the 95% confidence level for justifying the model with more parameters.

More formally, the null hypothesis was that the single-bond model (2) is valid. Using the F-test, it was concluded that the null hypothesis was invalid with a high level of confidence, so the double-bond model (7) was statistically justified. Applying the F-test with the null hypothesis that (7) is valid indicated that the second double-bond model (14) did not reduce the total residual enough to have a high level of confidence that the null hypothesis was invalid. Hence the original double-bond model (7) was the only model that was statistically justified.

The above statistical analysis does not imply that the intrinsic rate constants and molecular spring constants are the same for the two types of bonds; only that a distinction between these kinetic parameters cannot be discerned from the breakage frequency distributions. This statistical analysis is consistent with the sensitivity analysis, which indicated a much stronger effect of the distance to the free-energy minimum barrier on the breakage frequency distributions, compared to the other two kinetic parameters.

Model	p	R	ΔR	F	$F_{0.05}(q, n-p)$	α
(2)	3	0.4194	-	-	-	-
(7)	4	0.0687	0.3507	607.5	3.921	0.005
(14)	5	0.0685	0.0002	0.345	3.923	0.559

TABLE VI

F-TEST RESULTS FOR THE SINGLE- AND DOUBLE-BOND MICROSCOPIC MODELS AT THE 95% CONFIDENCE LEVEL ($q = 1, n = 123$).

V. CONCLUSIONS

This paper presents a systematic approach for analyzing distribution data and determining the kinetic parameters associated with bond dissociation from single-molecule pulling experiments of NCAM for low and high loading rates of applied force. Sensitivity analysis of a single-bond microscopic model indicates that the breakage frequency distribution is most sensitive to the value of the distance to the free-energy minimum barrier and least sensitive to the molecular spring constant. Experimental data and knowledge of the NCAM system [7] indicates that the single-bond microscopic model does not sufficiently describe the measured breakage frequency distributions. Therefore, a double-bond microscopic model was proposed, with only one additional parameter (i.e., an additional distance to the free-energy minimum barrier). The double-bond microscopic model was statistically justified (99.5% confidence) using an F-test. For high loading rate, the double-bond microscopic model provides a much more accurate fit to the experimental breakage frequency distribution for small values of the applied force, resulting in smaller confidence intervals and total residual than the single-bond model. The distances to the free-energy minimum barrier x_1, x_2 are consistent with the molecular structure of NCAM. The total length of the five Ig domains of NCAM is approximately 18 nm, whereas x_1, x_2 is approximately 0.08-0.73 nm. This indicates that the protein is relatively rigid, which was observed in SFA experiments [7].

REFERENCES

[1] E. Evans, "Probing the relation between force – lifetime – and chemistry in single molecular bonds," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 30, pp. 105–128, 2001.

[2] E. Evans, D. Berk, and A. Leung, "Detachment of agglutinin-bonded red blood cells. I. Forces to rupture molecular-point attachments," *Biophys. J.*, vol. 59, pp. 838–848, 1991.

[3] E. Evans and K. Ritchie, "Dynamic strength of molecular adhesion bonds," *Biophys. J.*, vol. 72, pp. 1541–1555, 1997.

[4] G. Hummer and A. Szabo, "Kinetics from nonequilibrium single-molecule pulling experiments," *Biophys. J.*, vol. 85, pp. 5–15, 2003.

[5] M. Rief, J. M. Fernandez, and H. E. Gaub, "Elastically coupled two-level systems as a model for biopolymer extensibility," *Phys. Rev. Lett.*, vol. 81, pp. 4764–4767, 1998.

[6] G. I. Bell, "Models for the specific adhesion of cells to cells," *Science*, vol. 200, pp. 618–627, 1978.

[7] C. P. Johnson, I. Fujimoto, C. Perrin-Tricaud, and D. Leckband, "Mechanism of homophilic adhesion by the neural cell adhesion molecule: Use of multiple domains and flexibility," *Proc. Natl. Acad. Sci. USA*, vol. 101, pp. 6963–6968, 2004.

[8] J. V. Beck and K. J. Arnold, *Parameter Estimation in Engineering and Science*. New York: Wiley, 1977.

[9] J. L. Zhou, A. L. Tits, and C. T. Lawrence, *User's Guide for FFSQP Version 3.7: A FORTRAN Code for Solving Constrained Nonlinear (Minmax) Optimization Problems, Generating Iterates Satisfying All Inequality and Linear Constraints*, Electrical Engineering Department and Institute for Systems Research, University of Maryland, College Park, MD 20742, April 1997.