Multiple-Bond Kinetics from Single-Molecule Pulling Experiments: Evidence for Multiple NCAM Bonds

E. J. Hukkanen,* J. A. Wieland,† A. Gewirth,† D. E. Leckband,*†‡ and R. D. Braatz*
*Department of Chemical and Biomolecular Engineering, †Department of Chemistry, and ‡Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois

ABSTRACT The kinetic parameters of single bonds between neural cell adhesion molecules were determined from atomic force microscope measurements of the forced dissociation of the homophilic protein-protein bonds. The analytical approach described provides a systematic procedure for obtaining rupture kinetics for single protein bonds from bond breakage frequency distributions obtained from single-molecule pulling experiments. For these studies, we used the neural cell adhesion molecule (NCAM), which was recently shown to form two independent protein bonds. The analysis of the bond rupture data at different loading rates, using the single-bond full microscopic model, indicates that the breakage frequency distribution is most sensitive to the distance to the transition state and least sensitive to the molecular spring constant. The analysis of bond failure data, however, motivates the use of a double-bond microscopic model that requires an additional kinetic parameter. This double-bond microscopic model assumes two independent NCAM-NCAM bonds, and more accurately describes the breakage frequency distribution, particularly at high loading rates. This finding agrees with recent surface-force measurements, which showed that NCAM forms two spatially distinct bonds between opposed proteins.

INTRODUCTION

The extraction of kinetic information from single-molecule pulling experiments has been investigated by several researchers (1–5). Most of the models derive from Bell's early model (6). Typically, the bond-rupture data are analyzed by constructing histograms of the rupture forces and then determining the most probable rupture force either by eye or from nonlinear least-squares fits to a probability distribution. Previous studies (3,4,7) indicate that the most probable rupture force f^* and the average rupture force $\bar{f}$, respectively, depend logarithmically on the loading rate r_f. A common method of obtaining kinetic parameters for individual bonds is to fit plots of f^* or $\bar{f}$ versus $\ln r_f$ to either of the two model equations (3,4).

This approach may yield inaccurate or misleading parameter estimates, since the shape of the initial force distribution is not considered when determining f^* or $\bar{f}$. This motivates the use of the full distribution of events as a function of the applied force. It has also been considered how multiple transition states along the unbinding trajectory alter the plots of f^* versus $\ln r_f$ (3). However, cases in which the molecules can form multiple, independent bonds will also greatly affect the parameter estimates, especially if the estimation method is not tailored to handle this phenomenon. This scenario is distinct from instances in which adhesion results from multiple simultaneous bonds in parallel between the tip and substrate (Fig. 1 *a*) (8,9). In the latter case, the force is shared simultaneously between the bonds. The model proposed in this study addresses the specific case where a receptor and ligand form more than one distinct intermolecular bond. In this instance, the bonds do not exist simultaneously, but sampling will detect two different bound states. We tested the ability of this double-bond model, together with rigorous statistical analyses, to describe rupture events in which the molecules can form multiple, independent bonds using the neural cell adhesion molecule (NCAM). NCAM is a particularly good test case for addressing these issues. It is one of the more abundant adhesion proteins in the brain (10). It mediates cell-cell adhesion and signaling by forming adhesive contacts between identical proteins on adjacent cells (10). The extracellular region of NCAM consists of two juxtamembrane fibronectin domains followed by five immunoglobulin (Ig) type domains. Based on the crystal structures of fragments of the NCAM ectodomain, different groups proposed different molecular models for homophilic binding (11,12). The majority of these models predict that NCAM proteins on adjacent cells adhere via a single interprotein bond (11,12). However, recent force measurements showed that NCAM forms either of two independent interprotein bonds that each involve different protein domains and relative protein alignments (Fig. 1 *a*) (13).

Here we used single-bond rupture investigations to quantify the parameters characteristic of NCAM bonds. The kinetic parameters associated with the dissociation of either of the two adhesive bonds, measured over a range of loading rates, are determined by maximum likelihood estimation applied to the breakage frequency distributions measured by atomic force microscopy (AFM).

This report further describes sensitivity analysis and multivariate statistical analysis, which were used to compute confidence intervals on the individual kinetic parameters for the full microscopic (FM) theory (4). The FM theory was

Submitted February 20, 2005, and accepted for publication July 8, 2005.

Address reprint requests to D. E. Leckband, Tel.: 217-244-0793; E-mail leckband@uiuc.edu; or to R. D. Braatz, Tel.: 217-333-5073; E-mail braatz@uiuc.edu.

0006-3495/05/11/3434/12 $2.00

doi: 10.1529/biophysj.105.061606

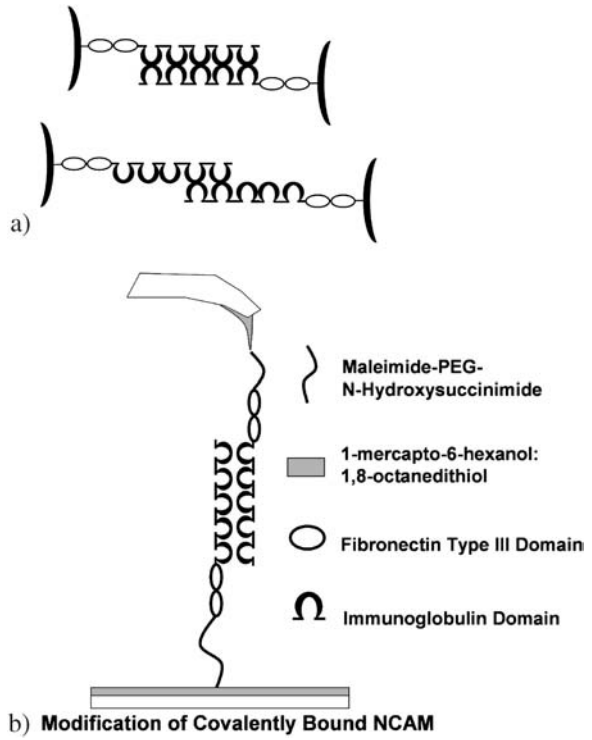

FIGURE 1 (*a*) Relative configuration of the opposing NCAM proteins in each of the two independent NCAM-NCAM bonds. The first bound state corresponds to the full overlap of opposing proteins, and the second bound state involves the overlap of the two outer two segments. (*b*) Schematic showing the orientation of the modified tip and substrate in the measurement. Although any free amine group on the protein can react with the NHS of the PEG spacer, the flexible PEG tether allows the proteins to align as shown.

found to be the most consistent single-bond model for the observed distributions. These analyses motivate the formulation of a double-bond microscopic model that more accurately describes the breakage frequency distributions between NCAM ectodomains over the range of values of the force ramp examined. This double-bond microscopic model, which assumes two independent bonds, uses one additional parameter, the distance from the energy minimum to the free energy barrier, beyond the single-bond FM model. An F-test confirms that the double-bond microscopic model is statistically justified for analyzing the NCAM breakage frequency distributions, and provides a better description of the rupture data than the single-bond FM theory, particularly at high loading rates. A comparison of the results of these analyses with histograms of the rupture data illustrates the disadvantages of relying on histograms to analyze these single-bond rupture data. This demonstration that the rupture data are described by two independent bound states is consistent with recently reported surface force measurements, which indicated that NCAM forms two independent homophilic bonds (13). This report provides a systematic procedure for analyzing bond rupture data, and for determining, from breakage frequency distributions, both the number and properties of bound states formed between two molecules.

THEORY

The cumulative distribution for the microscopic model (4) for the rupture of single intermolecular bonds is

$$P(f)=1-\exp\left[-\frac{k_o\exp\left(\frac{-\kappa_s x^2}{2k_BT}\right)}{r_f\sqrt{\frac{x}{k_BT}}(\kappa_m/\kappa)^{3/2}}\left(\exp\left(\frac{fx}{k_BT}-\frac{f^2}{2\kappa k_BT}\right)-1\right)\right], \tag{1}$$

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

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

The analysis of the cumulative distribution results in more accurate parameter estimates than using a histogram for the distribution, since the latter has binning errors. In this investigation, the 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 (14).

Full microscopic (FM) model

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

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

The maximum likelihood parameters are computed by solving the optimization problem (note that this optimization was solved numerically using off-the-shelf sequential quadratic programming software (15), with global optimality verified by finely gridding each parameter),

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

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 i^{th} experiment, $f_{i,j}$ corresponds to the j^{th} measurement in the i^{th} experiment, and N is the total number of experiments.

Confidence intervals

The parameter estimates are stochastic because of noise associated with measurements. An approximate confidence region for the parameters can be obtained by linearizing the model near the vicinity of the estimate (14),

$$\tilde{\mathbf{y}}_i(\theta) \approx \tilde{\mathbf{y}}_i(\theta^*) + \mathbf{F}_\iota(\theta^*)(\theta - \theta^*), \quad (6)$$

where $\tilde{\mathbf{y}}_i = \left[\tilde{\mathbf{y}}_{i,1}, \ldots, \tilde{\mathbf{y}}_{i,M_i}\right]^T$ is the vector of model predictions for the i^{th} experiment, θ^* is the maximum likelihood estimate for the parameter vector (i.e., the solution to the optimization problem Eq. 4), and $\mathbf{F}_i$ is the $M_i \times N_p$ matrix, where N_p is the number of parameters, given by

$$\mathbf{F}_i = \left.\frac{\partial \tilde{\mathbf{y}}_i}{\partial \theta}\right|_{\theta^*}, \quad (7)$$

which was calculated analytically. The parameter covariance matrix $\mathbf{V}_\theta$ for the linearized problem is given by

$$\mathbf{V}_\theta^{-1} = \sum_{i=1}^{N} \mathbf{F}_i^T \mathbf{V}_i^{-1} \mathbf{F}_i, \quad (8)$$

where $\mathbf{V}_i$ is defined to have compatible dimensions. The approximate $100(1 - \alpha)\%$ confidence region is the hyperellipsoid defined by

$$(\theta - \theta^*)^T \mathbf{V}_\theta^{-1} (\theta - \theta^*) \le \chi^2_{N_p}(\alpha), \quad (9)$$

where $\chi^2_{N_p}$ is the χ-squared distribution with N_p degrees of freedom, which is available in statistics textbooks (14). For this study, $\alpha = 0.05$ was used.

The eigenvectors of $\mathbf{V}_\theta{}^{-1}$ give the directions and the eigenvalues give the lengths of the axes of the hyperellipsoid. Since it is not possible to visualize the hyperellipsoid for higher than three dimensions, the confidence intervals are reported as

$$\theta_j^* - \sqrt{\chi^2_{N_p}(\alpha) V_{\theta,jj}} \le \theta_j \le \theta_j^* + \sqrt{\chi^2_{N_p}(\alpha) V_{\theta,jj}}, \quad (10)$$

where $V_{\theta,\,jj}$ is the (j, j) element of $\mathbf{V}_\theta$. (Note that the individual confidence intervals on each model parameter, although easier to interpret, do not provide as much information on the accuracy of the model parameters as the original confidence hyperellipsoid in Eq. 9.)

Parameter sensitivity analysis

Parameter sensitivity analysis quantifies the effect of perturbations in the parameters on the output variables (in this case, the cumulative distribution of the breakage frequencies). The sensitivities were determined from

$$S_i = \frac{\sqrt{V_{\theta,ii}}}{\sqrt{V_{\theta,jj}}} \left|\frac{\hat{\theta}_j}{\hat{\theta}_i}\right|, \quad (11)$$

where j was selected so that the maximum sensitivity was equal to 1.

The parameter sensitivities are reported in Table 1 for some representative conditions. In the FM 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 the most sensitive parameters from AFM data should be the most accurate, as quantified by smaller confidence intervals (14). 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 1, it is clear that estimates of the molecular spring constant κ_m from AFM data will be significantly less accurate than estimates of the distance to the transition state x.

TABLE 1 Parameter sensitivity analysis for the FM model

Sensitivities	Low loading rate r_f	High loading rate r_f
S_{k_o}	0.842	0.651
S_{κ_m}	0.344	0.554
S_x	1	1

Double-bond microscopic model

In Results and Discussion, below, it is observed that the breakage frequency distributions are not accurately described by the single-bond microscopic model. Influenced in part by recent surface-force measurements, a revision is made to the microscopic model that includes the effect of multiple (two) independent rupture events on the cumulative distribution (the extension to more than two bond ruptures is straightforward),

$$P_D(f) = \frac{P_d(f)}{P_d(f_{max})}, \quad (12)$$

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

$$P_d(f) = 2 - \exp\left[-\frac{k_o \exp\left(\frac{-\kappa_s x_1^2}{2 k_B T}\right)}{r_f \frac{x_1}{k_B T} (\kappa_m/\kappa)^{3/2}} \left(\exp\left(\frac{f x_1}{k_B T} - \frac{f^2}{2 \kappa k_B T}\right) - 1\right)\right] - \exp\left[-\frac{k_o \exp\left(\frac{-\kappa_s x_2^2}{2 k_B T}\right)}{r_f \frac{x_2}{k_B T} (\kappa_m/\kappa)^{3/2}} \left(\exp\left(\frac{f x_2}{k_B T} - \frac{f^2}{2 \kappa k_B T}\right) - 1\right)\right], \quad (13)$$

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

$$\kappa = \kappa_s + \kappa_m. \quad (14)$$

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 (see Results and Discussion, below). Duplicating the least-sensitive parameter, the molecular force constant, κ_m, would not result in more accurate parameter determinations.

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

$$\theta_D = [k_o, \kappa_m, x_1, x_2]^T, \quad (15)$$

which are computed as the solution to the optimization problem

$$\min_{\theta_D, x_1 > x_2} \sum_{j=1}^{M_i} \frac{1}{M_i} \left(P_D^{exp}(f_{i,j}) - P_D^{sim}(f_{i,j})\right)^2 \text{ for } i = 1, \ldots, N. \quad (16)$$

MATERIALS AND METHODS

Preparation of substrate and AFM cantilevers for NCAM measurements

The general sample configuration used in these measurements is shown in Fig. 1 *b*. Substrates were glass microscope slides (Fisher Scientific, Hampton, NH) cut into 18-mm-square pieces. Commercial Si_3N_4 V-shaped contact cantilevers with gold reflective coating were purchased from Digital Instruments (Sunnyvale, CA). Both the AFM tips and substrates were first cleaned in chloroform (Fisher Scientific) for 10 min. They were then dried with argon and soaked in a Piranha solution consisting of 70:30 v/v mixture of concentrated H_2SO_4 (Mallinckrodt Baker, Phillipsburg, NJ) and 30% H_2O_2 (Fisher Scientific) for 30 min. The tips and substrates were then washed with cold Milli-Q purified water (Millipore, Bedford, MA). The substrates were washed with boiling Milli-Q purified water, and then both were dried with argon, and immediately placed into a thermal evaporator (Cooke Vacuum Products, Norwalk, CT) for gold evaporation. Gold films were evaporated onto the glass substrates and tips. The evaporation involved two steps, each performed at a base pressure of 10^{-6} Torr without substrate heating. First a chromium adhesion layer with a thickness of ~30 Å was thermally deposited at a rate of ~0.2 Å/s. This was followed by the deposition of a gold layer of 800 Å at a rate of ~1.2 Å/s. The samples were rinsed thoroughly with ethanol, dried with argon, and then placed directly into an ethanolic thiol solution containing 1 mM 1,8-octanedithiol (Sigma-Aldrich, St. Louis, MO) and 10 mM 6-mercapto hexan-1-ol (Sigma-Aldrich). The incubation was carried out for ~18 h. Tips and substrates were then removed from the thiol solution, rinsed with ethanol, dried with argon, and placed into a phosphate-buffered solution containing 1 mg/mL of poly-(ethylene glycol)-α-maleimide, ω-N-hydroxysuccinimide ester (NHS-PEG-MAL; Shearwater, Huntsville, AL). The aqueous buffer used as the solvent for this and the following solutions contained 50 mM NaH_2PO_4 (Fisher Scientific), 100 mM NaCl (Mallinckrodt Baker), and 1 mM EDTA (Fisher Scientific) and was brought to pH 7.4 by adding 1 M NaOH (Fisher Scientific). After letting the coated materials set for 20 min, the tips and substrates were rinsed with buffer and immediately transferred to the AFM fluid cell. See Fig. 2 for schematic representations of the tip and substrate modification.

The coated glass slide was placed on the AFM stage and sandwiched under an O-ring and Teflon cell, to which ~1.5 mL of the NCAM solution, containing 0.06 mg/ml NCAM, was added. All NCAM experiments were performed using soluble NCAM extracellular domains (13) engineered with a C-terminal oligohistidine tag. In this investigation, the protein was covalently bound, and the histidine tag was only used for the protein purification. The protein, which has a molecular weight of ~120 kDa, was expressed in soluble form by stably transfected Chinese hamster ovary cells (13). The AFM tip was then mounted on the AFM head, thereby submerging it under the NCAM solution, and both the tip and sample were then allowed to incubate for 90 min. The cell was then flushed 10 times with buffer, keeping the tip and sample submerged, and rinsing off any nonspecifically adsorbed protein before conducting force measurements.

AFM setup

All force probe measurements were obtained with a commercial AFM apparatus (Pico AFM, Molecular Imaging, Tempe, AZ) interfaced with a commercial controller and data acquisition electronics (Digital Instruments, Buffalo, NY). Our version of the Digital Instruments software operating on the Nano III E platform provides triangle waveforms to control piezo movement during the acquisition of force-curve data. Another instrument, the Biomembrane Force Probe, is capable of generating more complicated force histories to identify possible substates contributing to rupture peaks (9). However, with this and many other commercial AFM instruments, such measurements require reprogramming the head-movement sequences. This is not possible with our instrument. All experiments were carried out at room temperature. Loading rates for the NCAM experiments ranged from ~400–7500 pN/s, with ~2000 force extension curves obtained per loading rate. The AFM cell was translated after every 500 measurements to minimize bias in the data due to aging at each substrate

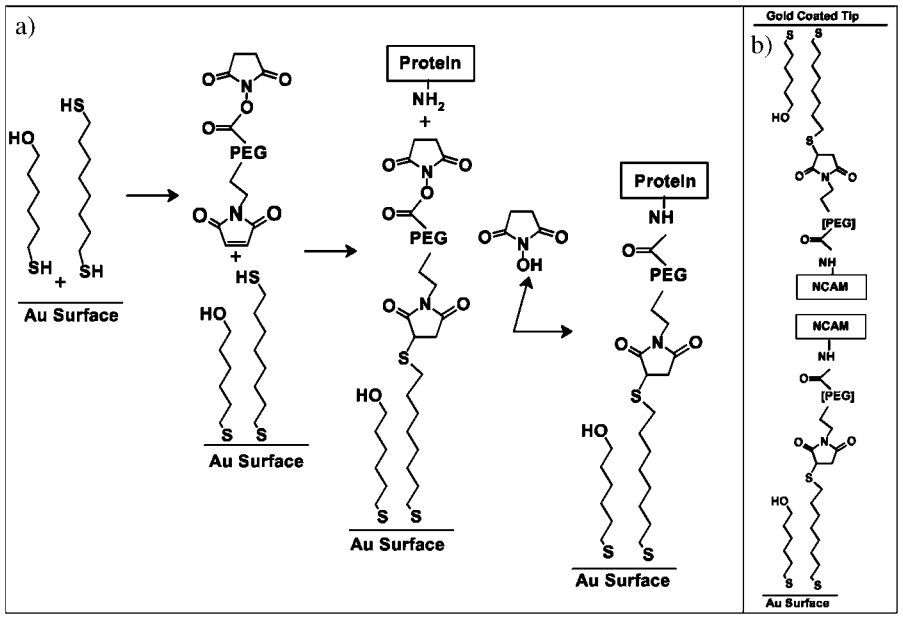

FIGURE 2 (*a*) Schematic showing the step-by-step modification of the gold-coated tip and substrate. The gold is first coated with a mixed monolayer of alkane thiols. The exposed thiols are activated with the MAL-PEG-NHS linker, which covalently binds proteins through their free amines. See text for details. (*b*) Full schematic showing the relative orientation of the immobilized proteins on the tip and modified substrate.

contact. Tip wear cannot, however, be ruled out. The loading rate was calculated by multiplying the tip velocity (frequency times distance traveled per cycle) by the slope of the force-distance curve just before bond rupture, k_s. Because the latter reflects the polymer extension, this approach accounts for the contribution of the polymer tether to the loading rate.

RESULTS AND DISCUSSION

Forced bond rupture

Histograms of the bond-rupture forces obtained at the different loading rates of 781, 1152, 1567, 4101, and 7423 pN/s are shown in Fig. 3. In these cases, the frequency of observed bond-rupture events was $<15\%$. That is, <15 binding events were detected for every 100 touches to the surface. This criterion increases the likelihood that the rupture events are due to single-bond events rather than to multiple-bond formation between the tip and surface. A histogram of the data obtained at 7423 pN/s suggests the existence of two peaks (compare to Fig. 7 *e*), which both broaden and shift to higher forces with increasing loading rate. This behavior is consistent with the existence of two separate bonds, although the distinction is less obvious if larger, statistically rigorous bin sizes are selected (Fig. 3 *e*). Furthermore, the apparent resolution of the peaks decreases at the lower loading rates as the two peaks appear to merge.

Analysis

Maximum likelihood parameter estimates and confidence intervals for the single-bond and double-bond microscopic models were determined for NCAM experiments at five

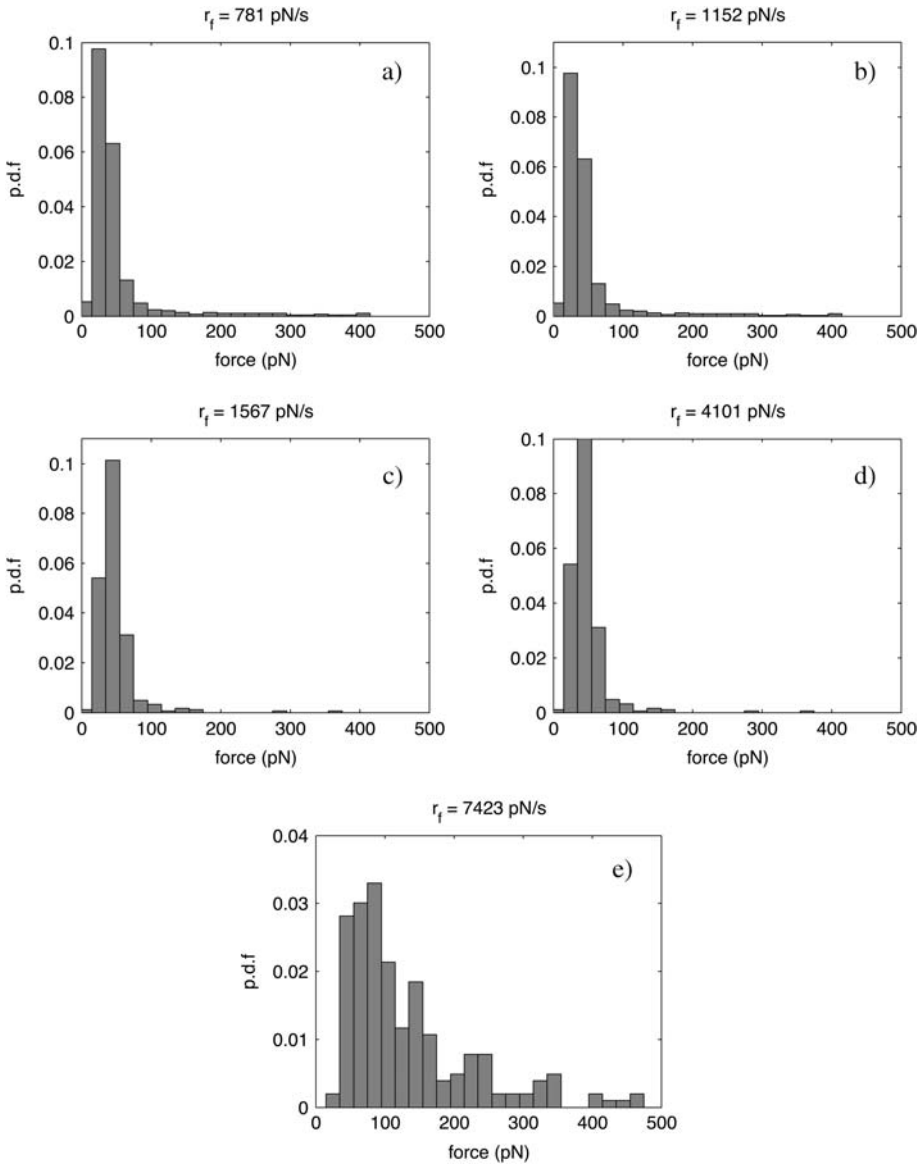

FIGURE 3 Histograms of forced bond ruptures at the indicated loading rates.

TABLE 2 Conditions for NCAM experiments

Loading rate r_f (pN/s)	Number of rupture events
781	198
1152	56
1567	53
4101	100
7423	123

different loading rates (see Table 2). Table 3 reports the kinetic parameters, confidence intervals, and total residuals for the FM model for a single bound state. The total residual is defined as

$$R = \arg\min_{\theta} \Phi. \quad (17)$$

Large confidence intervals were observed for all loading rates, which is consistent with the inability of the single-bond microscopic model to describe the shape of the experimental bond-rupture distributions for both high and low values of the applied force (see Fig. 4, *a–c*). The calculated 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 4. These results reveal three main features. First, for loading rates of 781 and 7423 pN/s, the distances to the free-energy barrier, x_1 and x_2, differ by a factor of 2–3. Although the distance to the transition state is assumed to be independent of the loading rate, these analyses show that this value decreases systematically with increasing loading rate. Similar parameter variations with the loading rate were observed in AFM measurements of the force to extract lipids from membranes (16). The location of the transition state reported in molecular dynamics simulations of lipid extraction similarly shifted to shorter distances with increasing pulling speeds (17). Recent theoretical analyses also showed that this parameter decreases with loading rate for smoothly varying potentials (D.E. Leckband, unpublished observations).

Second, at the high and low loading rates the confidence intervals and total residual are much smaller for the double-bond microscopic model than for the single-bond microscopic model. The relative magnitudes of the confidence intervals are smaller for the distances to the free-energy barrier, x_i, than for the molecular spring constant, κ_m, which is consistent with the sensitivity analysis. The total residual is a factor of 3 and 6 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. These analyses therefore indicate that the bond-rupture behavior is better described by two independent bonds with different values for x_i but with similar bond energies, E_b, which are related to the rate constants. Again, the findings are supported by direct force measurements, which identified two spatially distinct bonds with similar adhesion energies (13).

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 Fig. 4 *d*), resulting in smaller confidence intervals and total residual than the single-bond model (compare Tables 3 and 4). 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 such as infrequent, multiple, simultaneous rupture events are more prevalent at this loading rate.

For the higher loading rate of $r_f = 7423$ pN/s, the 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 Fig. 4 *f*). For this loading rate, the confidence intervals of the double-bond microscopic model are a factor of ~2–3 smaller than for the single-bond model. This double-bond microscopic model is, therefore, better able to capture the physical phenomena associated with the pulling of the adherent NCAM molecules in these AFM experiments.

Third, for the loading rates of $r_f = 1152$ and $r_f = 1567$ pN/s, the double-bond microscopic model is not statistically justified on the basis of these data (compare Tables 3 and 4). Therefore, Fig. 4, *b* and *e* (for $r_f = 1567$ pN/s), represents the model parameters for the single- and double-bond microscopic model (i.e., $x_1 = x_2$). The total residual does not change with the addition of a second distance parameter, although fitting distributions to histograms with particular bin sizes might lead to the opposite conclusion (see Fig. 7 *b*). The large size of the confidence intervals on x_1 for those loading rates (see Tables 3 and 4) and the smaller number of events in these experimental data sets (see Table 2) indicate that the lack of statistical significance of the double-bond model for these loading rates is likely due to inadequate sampling for those AFM data sets. This conclusion is supported by independent direct force measurements that identified two independent bonds (13). This physical

TABLE 3 Parameter estimates and confidence intervals, reported at 95%, using the FM model for each set of NCAM experiments

θ	781 pN/s	1152 pN/s	1567 pN/s	4101 pN/s	7423 pN/s
k_o (s^{-1})	1.87 ± 0.87	9.13 ± 2.55	1.76 ± 1.24	8.64 ± 10.79	18.08 ± 5.24
κ_m (pN/nm)	172 ± 651	3490 ± 6520	105 ± 364	636 ± 5380	1480 ± 1470
x (nm)	0.346 ± 0.253	0.0314 ± 0.0321	0.444 ± 0.373	0.164 ± 0.282	0.0699 ± 0.0301
Total residual	0.9157	0.1393	0.0347	0.3790	0.4194

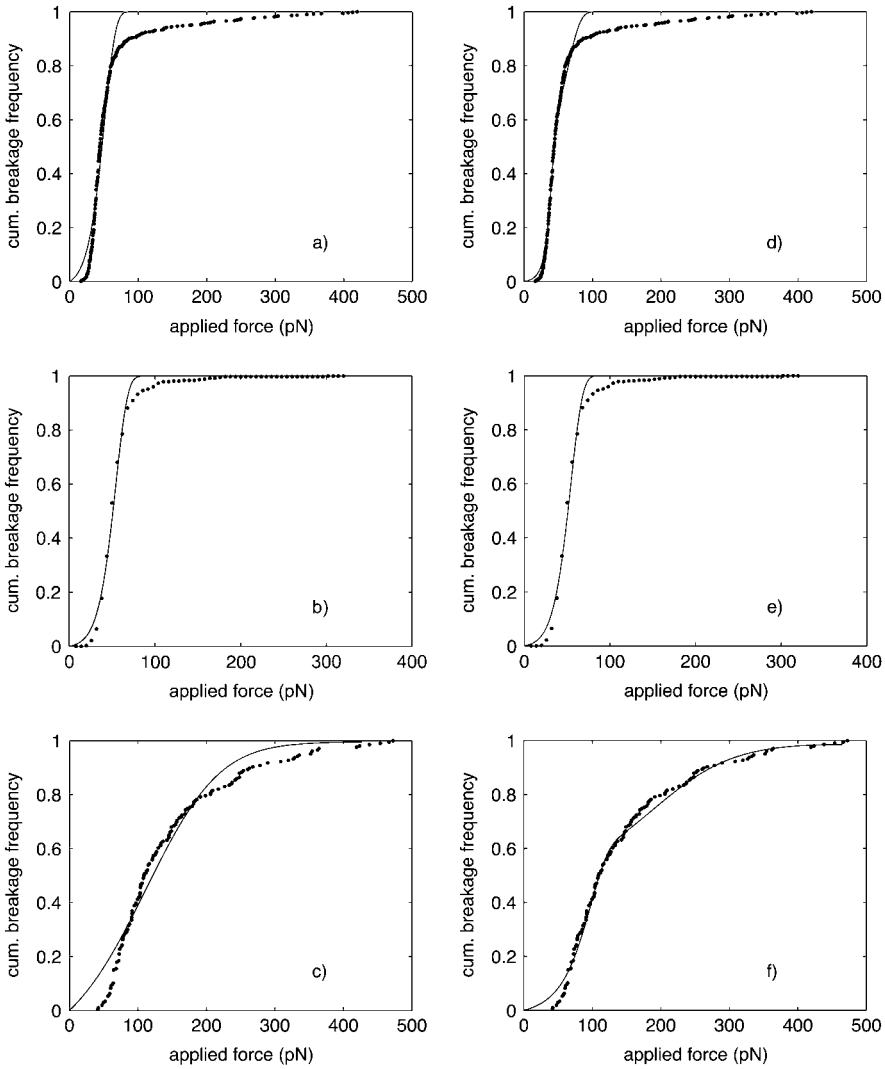

FIGURE 4 Comparison of NCAM experimental and single-bond FM cumulative distributions for (*a*) 781 pN/s, (*b*) 1567 pN/s, and (*c*) 7423 pN/s. Comparison of NCAM experimental and double-bond FM cumulative distributions for (*d*) 781 pN/s, (*e*) 1567 pN/s, and (*f*) 7423 pN/s.

situation would hold for all loading rates. Thus, the inability of the model to fit the data sets at intermediate loading reflects the limitations of the data. Therefore, further double-bond microscopic model analysis was not continued on this data set.

Lastly, for the loading rates of $r_f = 4101$ pN/s, the double-bond microscopic model improves the total residual, although not as much as for $r_f = 7423$ pN/s. The confidence interval for x_2 does not quite statistically justify a double-bond model at the 95% confidence level, but the double-bond model is statistically justified for a slightly lower confidence level.

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

TABLE 4 Parameter estimates and confidence intervals, reported at 95%, using the double-bond FM model Eq. 13 for individual NCAM experiments

θ	781 pN/s	1152 pN/s*	1567 pN/s*	4101 pN/s	7423 pN/s
k_o (s^{-1})	0.796 ± 0.188	9.13 ± 2.55	1.76 ± 1.24	10.53 ± 8.09	5.40 ± 1.46
κ_m pN/nm)	182 ± 320	3490 ± 6520	105 ± 364	1910 ± 6890	1380 ± 548
x_1 (nm)	0.730 ± 0.142	0.0314 ± 0.0321	0.444 ± 0.373	0.205 ± 0.120	0.188 ± 0.0225
x_2 (nm)	0.337 ± 0.108	—	—	0.0660 ± 0.0817	0.0813 ± 0.0160
Total residual	0.2906	0.1393	0.0347	0.2079	0.0687

*By applying the same statistical methods, the double-bond model is not appropriate for the experimental data.

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

$$\theta = [k_{1o}, k_{2o}, \kappa_m, x_1, x_2]^T, \tag{18}$$

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

$$P_d(f) = 2 - \exp\left[-\frac{k_{1o}\exp\left(\frac{-\kappa_s x_1^2}{2k_BT}\right)}{\frac{x_1}{r_f\frac{}{k_BT}(\kappa_m/\kappa)^{3/2}}}\left(\exp\left(\frac{fx_1}{k_BT} - \frac{f^2}{2\kappa k_BT}\right) - 1\right)\right] - \exp\left[-\frac{k_{2o}\exp\left(\frac{-\kappa_s x_2^2}{2k_BT}\right)}{\frac{x_2}{r_f\frac{}{k_BT}(\kappa_m/\kappa)^{3/2}}}\left(\exp\left(\frac{fx_2}{k_BT} - \frac{f^2}{2\kappa k_BT}\right) - 1\right)\right]. \tag{19}$$

A simple quantification of the significance of the fifth parameter, k_{2o}, is achieved by computing the total residual Eq. 16 for varying k_{2o} (with the remaining parameters in Table 4 fixed) for $r_f = 7423$ pN/s.

Kinetic parameters and confidence intervals were estimated for Eq. 19, which introduces the additional kinetic rate constant k_{2o} in addition to the second free-energy barrier distance x_2 (see Table 5). 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 4). 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 Eq. 13 yields a residual of 0.0687, whereas the additional intrinsic rate constant, k_{2o}, does not significantly improve the total residual, where Eq. 19 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 (14)

$$F = \frac{\Delta R/q}{R/(n-p)}, \tag{20}$$

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 $100(1-\alpha)\%$ confidence if $F > F_\alpha(q, n-p)$, where some values of the F-distribution, $F_\alpha(q, n-p)$, are reported in Table 6, with the results of applying the F-test to the single- and double-bond microscopic models. The double-bond model Eq. 13, which 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 Eq. 19, which introduces two additional parameters, does not meet the 95% confidence level for justifying the model with more parameters. In other words, the null hypothesis was that the single-bond model Eq. 1 is valid. Using the F-test, it is concluded that the null hypothesis is invalid with a high level of confidence, so the double-bond model Eq. 13 is statistically justified. Applying the F-test with the null hypothesis that Eq. 13 is valid indicates that the second double-bond model Eq. 19 does not reduce the total residual enough to have a high level of confidence that the null hypothesis is invalid. Hence the original double-bond model Eq. 13 is the only model that is statistically justified.

This 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. Note that 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.

A similar analysis for a three-bond model in which each bond has a different free-energy barrier distance indicated that such a three-bond model was not statistically justified.

TABLE 5 Parameter estimates and confidence intervals, reported at 95%, using the double-bond FM model Eq. 19 for individual NCAM experiments

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

Comparison of analytical methodologies

A comparison of the differential distribution determined analytically from the cumulative distribution Eq. 12 with a histogram constructed from experimental data (see Fig. 5)

TABLE 6 F-test results for the single- and double-bond FM models at the 95% confidence level

Model	n	p	q	R	ΔR	F	$F_{0.05}(q, n-p)$	α
Eq. 1	123	3	—	0.4194	—	—	—	—
Eq. 13	123	4	1	0.0687	0.3507	607.5	3.921	0.005
Eq. 19	123	5	1	0.0685	0.0002	0.345	3.923	0.559

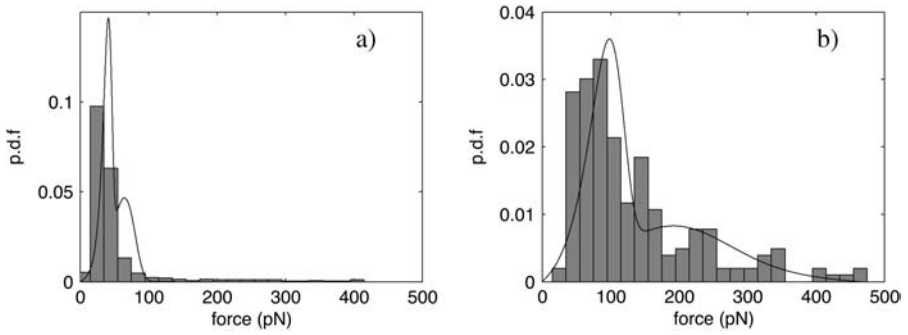

FIGURE 5 Comparison of the differential distribution Eq. 22 and the experimental histogram for NCAM data for a loading rate of (*a*) $r_f = 781$ pN/s and (*b*) $r_f = 7423$ pN/s.

illustrates the weakness of the common practice of fitting the differential distribution to histograms. The differential distribution is defined by

$$p_d(f) = \frac{dP_D(f)}{df}, \tag{21}$$

which is given by

$$p_d(f) = \frac{1}{P_D(f_{max})}\left\{\exp\left[-\frac{k_o \exp\left(\frac{-\kappa_s x_1^2}{2\,k_B T}\right)}{r_f \frac{x_1}{k_B T}(\kappa_m/\kappa)^{3/2}}\left(\exp\left(\frac{f x_1}{k_B T} - \frac{f^2}{2\,\kappa k_B T}\right) - 1\right)\right] \times \left(\frac{k_o \exp\left(\frac{-\kappa_s x_1^2}{2\,k_B T}\right)}{r_f \frac{x_1}{k_B T}(\kappa_m/\kappa)^{3/2}} \exp\left(\frac{f x_1}{k_B T} - \frac{f^2}{2\,\kappa k_B T}\right)\right)\right.$$
$$\times \left(\frac{x_1}{k_B T} - \frac{f}{\kappa k_B T}\right) + \exp\left[-\frac{k_o \exp\left(\frac{-\kappa_s x_2^2}{2\,k_B T}\right)}{r_f \frac{x_2}{k_B T}(\kappa_m/\kappa)^{3/2}}\left(\exp\left(\frac{f x_2}{k_B T} - \frac{f^2}{2\,\kappa k_B T}\right) - 1\right)\right]$$
$$\left. \times \left(\frac{k_o \exp\left(\frac{-\kappa_s x_2^2}{2\,k_B T}\right)}{r_f \frac{x_2}{k_B T}(\kappa_m/\kappa)^{3/2}} \exp\left(\frac{f x_2}{k_B T} - \frac{f^2}{2\,\kappa k_B T}\right)\right)\left(\frac{x_2}{k_B T} - \frac{f}{\kappa k_B T}\right)\right\}. \tag{22}$$

For the loading rate of $r_f = 7423$ pN/s, the cumulative distribution Eq. 12 fits the long tail in the experimental data (see Fig. 4 *f*). It is much easier to visualize the two peaks in its differential distribution, corresponding to the two bonds of different energies, than from the histogram of the data (see Fig. 5). Fitting the kinetic parameters to the cumulative distribution instead of fitting to a histogram avoids the error associated with the binning of data that occurs when constructing a histogram.

CONCLUSIONS

This article presents a systematic approach for analyzing the distribution of bond-rupture data and determining the kinetic parameters associated with bond dissociation. As a test case, we focused on single-molecule pulling experiments of NCAM for a range of 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. These experimental data, together with the knowledge of the NCAM system (13), indicate 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 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 and large values of the applied force, resulting in smaller confidence intervals and total residual than the single-bond model.

TABLE 7 Parameter estimates and confidence intervals, reported at 95%, using the DFS model for each set of NCAM experiments

θ	781 pN/s	7423 pN/s
t_{off} (s)	0.645 ± 0.162	0.0440 ± 0.0067
f_β (pN)	14.73 ± 1.46	107 ± 20
Total residual	0.9949	0.5331

These data and the analytical methods used in this study illustrate important aspects of single-bond rupture data and the methods commonly used for their analyses. First, although the use of different bin sizes may appear to reveal features such as multiple peaks, the cumulative distribution approach described here allows comparison of different models free of errors attributed to incorrect bin sizing. In this case, for example, prior surface force measurements and equilibrium binding studies both showed that NCAM forms two independent protein-protein bonds (13). Nevertheless, even though some histograms appeared to support this, rigorous analyses based only on histogram data did not support the conclusion. In this case, prior knowledge of the system indicated that the high standard deviations are attributed to limitations in the data sets rather than differences in the molecular mechanism of adhesion.

The tailing evident at high forces is a common feature in many single-bond rupture measurements. Fig. 7, *a–e*, exhibits some events at the largest rupture forces (the so-called ''tails''). The precise shape and number of events in the tails will fluctuate from one set of experiments to the next, due to the finite sample size and variations in sample preparation. These effects are the largest for Fig. 7, *b* and *c*, because those experiments have the fewest samples (see Table 2).

However, the shape of the tails also depends on the width of the bins used in the histograms and on the physical parameters of the system. In most instances, for example, tails are attributed to infrequent multipoint attachments in which n simultaneous bonds between the tip and surface share the force (8,9). In the latter case, two additional parameters x_i and k_{oi} are needed to describe each n^{th}-point attachment for $n > 1$. In these NCAM data, the higher forces could also be due to the existence of a second, stronger bond. We introduced the double-bond model, to account for this possibility. In contrast to multipoint attachments, the double-bond model introduced here requires only one additional parameter to describe the data. This simple extension of the FM theory, as opposed to models with additional parameters, accurately captures the tail at three of the r_f values used.

In conclusion, this study illustrates a rigorous approach to the analysis of single-bond rupture data. The findings highlight the potential pitfalls that can result from inappropriate data binning, and present an alternative method that avoids this. This methodology was, in turn, used to test whether a proposed double-bond model best describes the NCAM binding data. The results agree with prior force measurements and equilibrium binding studies that show that NCAM forms two different interprotein bonds (13).

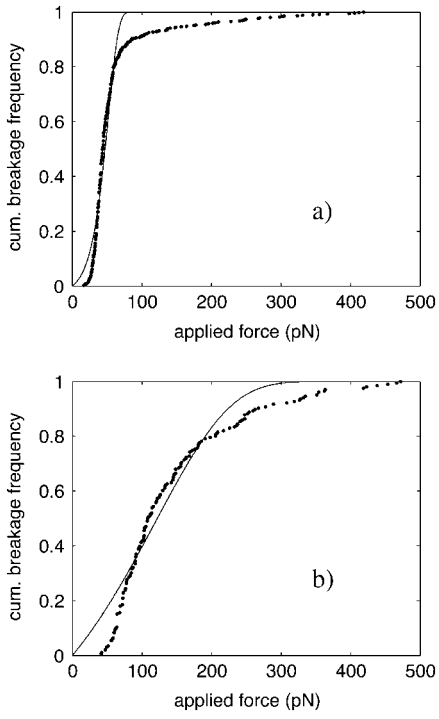

FIGURE 6 Comparison of NCAM experimental and DFS cumulative distributions for (*a*) $r_f = 781$ pN/s and (*b*) $r_f = 7423$ pN/s.

APPENDIX A: ALTERNATIVE MODEL

An alternative for modeling single-molecule pulling experiments is the dynamic force spectroscopy (DFS) model (1,2,3). The cumulative distribution of the DFS model for the rupture of adhesion bonds is given by

$$P_p(f) = 1 - \exp\left[-\frac{f_\beta}{r_f t_{off}}\left(\exp\left(\frac{f}{f_\beta}\right) - 1\right)\right], \qquad (23)$$

where f is the applied force, f_β is the thermal force barrier defined by

$$f_\beta = \frac{k_B T}{x}, \qquad (24)$$

where x is the distance from the free-energy minimum to the barrier, k_B is Boltzmann's constant, and T is the temperature (K). The spontaneous dissociation time t_{off} is the inverse of the intrinsic rate constant

$$t_{off} = \frac{1}{k_o}, \qquad (25)$$

and r_f is the applied loading rate. Parameter sensitivity analysis and confidence regions are determined in the same manner as for the microscopic model. A more thorough analysis of the DFS model is presented in a thesis (18).

Maximum likelihood parameter estimates and confidence intervals for the single-bond DFS model were determined for NCAM experiments at two different loading rates. The DFS model was not modified to account for two independent bonds. The DFS model fits to the specified loading rates are reported in Fig. 6, *a* and *b*. There do not appear to be noticeable differences in the DFS model fits in comparison to the FM model fits (see Fig. 4, *a* and *c*). The DFS parameter estimates and total residuals are reported in Table 7.

TABLE 8 Parameter estimates and confidence intervals, reported at 95%, using the DFS model for each set of NCAM experiments using Eqs. 24 and 25

θ	781 pN/s	7423 pN/s
k_o (s^{-1})	1.55 ± 0.39	22.7 ± 3.4
x (nm)	0.279 ± 0.028	0.0385 ± 0.0072
Total residual	0.9949	0.5331

Most notable is that the total residual of the DFS model is larger than the FM model for both loading rates. For a loading rate, $r_f = 781$ pN/s, an 8.7% increase in total residual is observed. For a loading rate, $r_f = 7423$ pN/s, a 27.1% increase in total residual is observed. A direct way of comparing the FM and DFS models is to compare similar parameters (i.e., by the relationships of Eqs. A2 and A3). Table 8 reports the DFS model parameters in terms of similar parameters associated with the FM model. Within the confidence intervals for both single-bond models, the parameter estimates are very similar in value. However, the smaller confidence intervals associated with the DFS model do not indicate that the model is a better fit (i.e., the total residuals are larger). Based on the FM model fits, it is not expected that a better fit with the double-bond DFS model is achieved than with the double-bond FM model. A similar trend is observed in further analysis of experiments (18).

APPENDIX B: ANALYSIS OF DATA WITH COMMERCIAL SOFTWARE

A commercial software package (Igor Pro 5.0, WaveMetrics, Lake Oswego, OR) was used to determine parameter estimates for the double-bond microscopic model. In this case, the parameter estimates are determined using

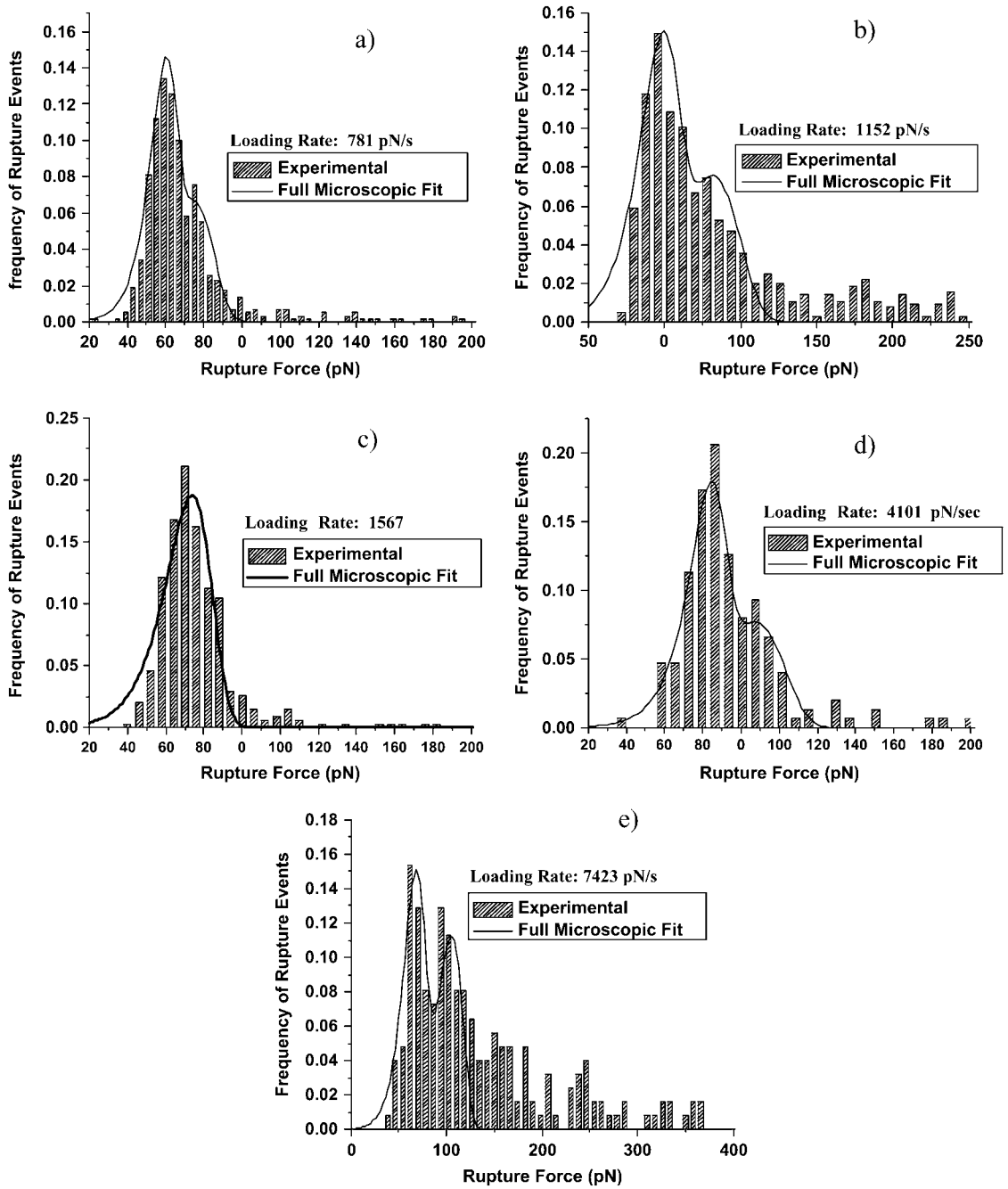

FIGURE 7 Comparison of the double-bond FM model to experimental data using Igor Pro 5.0 for (*a*) $r_f = 781$ pN/s, (*b*) $r_f = 1152$ pN/s, (*c*) $r_f = 1567$ pN/s, (*d*) $r_f = 4101$ pN/s, and (*e*) $r_f = 7423$ pN/s.

TABLE 9 Parameter estimates and confidence intervals, reported at 95%, using the double-bond FM model Eq. 13 for individual NCAM experiments and Igor Pro 5.0

θ	781 pN/s	1152 pN/s	1567 pN/s	4101 pN/s	7423 pN/s
k_{1o} (s^{-1})	0.152 ± 0.798	1.62 ± 1.68	2.11 ± 0.38	0.23 ± 1.19	0.352 ± 0.594
k_{2o} (s^{-1})	0.423 ± 0.300	0.93 ± 6.92	—	0.88 ± 0.64	0.053 ± 0.157
κ_{1m} (pN/nm)	170 ± 1480	260 ± 53	192 ± 8	377 ± 4580	138 ± 28
κ_{2m} (pN/nm)	112 ± 22	444 ± 7780	—	312 ± 57	267 ± 82
x_1 (nm)	0.701 ± 0.759	0.386 ± 0.079	0.432 ± 0.014	0.492 ± 0.621	0.522 ± 0.104
x_2 (nm)	0.519 ± 0.059	0.259 ± 0.895	—	0.420 ± 0.128	0.420 ± 0.128

nonlinear least-squares and the Levenberg-Marquardt method (provided within the software package). The parameter estimates and corresponding confidence intervals (95%) are reported in Table 9. A full six-parameter double-bond model was used, in which case previous parameter sensitivity analysis was not considered. A comparison of the results of this software package and the rigorous approach (i.e., the focus of this article) is made (see Table 4). For $r_f = 781$ pN/s, x_1 values are almost identical, although a much larger confidence interval is reported from the software package. The second distance to the free-energy minimum barrier x_2 differs by ~40%, but the confidence interval is significantly smaller. The intrinsic rate constants, determined by commercial software, are a factor of 2–5 smaller than the rigorous approach (i.e., use of cumulative distributions), and the confidence intervals are much larger. The least sensitive parameter, the molecular spring constant, is similar to that reported in Table 4. There is a significant difference in the parameter estimates for $r_f = 7423$ pN/s. The intrinsic rate constant and molecular spring constant are an order-of-magnitude smaller than the rigorous approach presented, whereas the distance to the free-energy minimum barrier is a factor of 3–5 larger. For $r_f = 1567$ pN/s, Igor predicts that the single-bond model is sufficient to model the experimental data, as seen in the presented approach. The distance to the free-energy minimum barrier is almost identical (see Tables 4 and 9). The other two parameters, k_o and κ_m, differ by ~50%. For $r_f = 1152$ and $r_f = 4101$ pN/s, Igor estimates parameter values that are grossly different from values reported in Table 4, with large confidence intervals. For the case of $r_f = 1152$ pN/s, Igor predicts that a double-bond model is plausible, although only a single-bond model is statistically justified (see Table 4). The difference in the commercial software and rigorous cumulative distribution approach is better understood by comparing the model prediction (software package) and experimental data (see Fig. 7, *a–e*). The fit (obtained through commercial software) only predicts rupture events for applied forces $f \leq 150$ pN, whereas the data clearly indicate that rupture events occur at $f \geq 150$ pN. The model prediction from the software package for $r_f = 7423$ pN/s ignores a significant portion of the data (see Fig. 5 *b* and 7 *e*), which are very well captured by the double-bond FM model with parameters fit according to the approach presented in this article (see Fig. 4 *f*). The rigorous approach presented in this article accounts for the entire data set, for both parameter estimation algorithms and comparisons to experiments.

This work was supported by National Institutes of Health grant No. RO1 GM63536.

REFERENCES

1. Evans, E. 2001. Probing the relation between force lifetime and chemistry in single molecular bonds. *Annu. Rev. Biophys. Biomol. Struct.* 30:105–128.
2. Evans, E., D. Berk, and A. Leung. 1991. Detachment of agglutinin-bonded red blood cells. I. Forces to rupture molecular-point attachments. *Biophys. J.* 59:838–848.
3. Evans, E., and K. Ritchie. 1997. Dynamic strength of molecular adhesion bonds. *Biophys. J.* 72:1541–1555.
4. Hummer, G., and A. Szabo. 2003. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophys. J.* 85:5–15.
5. Rief, M., J. M. Fernandez, and H. E. Gaub. 1998. Elastically coupled two-level systems as a model for biopolymer extensibility. *Phys. Rev. Lett.* 81:4764–4767.
6. Bell, G. I. 1978. Models for the specific adhesion of cells to cells. *Science.* 200:618–627.
7. Balsera, M., S. Stepaniants, S. Izrailev, Y. Oono, and K. Schulten. 1997. Reconstructing potential energy functions from simulated force-induced unbinding processes. *Biophys. J.* 73:1281–1287.
8. Marshall, B. T., M. Long, J. W. Piper, T. Yago, R. P. McEver, and C. Zhu. 2003. Direct observation of catch bonds involving cell-adhesion molecules. *Nature.* 423:190–193.
9. Perret, E., A. Leung, H. Feracci, and E. Evans. 2004. *Trans*-bonded pairs of E-cadherin exhibit a remarkable hierarchy of mechanical strengths. *Proc. Natl. Acad. Sci. USA.* 101:16472–16477.
10. Walsh, F. S., and P. Doherty. 1997. Neural cell adhesion molecules of the immunoglobulin superfamily: role in axon growth and guidance. *Annu. Rev. Cell Dev. Biol.* 13:425–456.
11. Jenson, P., V. Soroka, N. K. Thompson, I. Ralets, V. Berezin, E. Bock, and F. M. Poulsen. 1999. Structure and interactions of NCAM modules 1 and 2—basic elements in neural cell adhesion. *Nat. Struct. Biol.* 6: 6963–6968.
12. Ranheim, T. S., G. M. Edelman, and B. A. Cunningham. 1996. Homophilic adhesion mediated by the neural cell adhesion molecule involves multiple immunoglobulin domains. *Proc. Natl. Acad. Sci. USA.* 93:4071–4075.
13. Johnson, C. P., I. Fujimoto, C. Perrin-Tricaud, and D. Leckband. 2004. Mechanism of homophilic adhesion by the neural cell adhesion molecule: use of multiple domains and flexibility. *Proc. Natl. Acad. Sci. USA.* 101:6963–6968.
14. Beck, J. V., and K. J. Arnold. 1977. Parameter Estimation in Engineering and Science. Wiley, New York.
15. Zhou, J. L., A. L. Tits, and C. T. Lawrence. 1997. 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 Dept. and Institute for Systems Research, University of Maryland, College Park, MD.
16. Wieland, J. A., A. Gewirth, and D. E. Leckband. 2005. Single molecule measurements of the impact of lipid phase behavior on anchor strengths. *J. Phys. Chem. B.* 109:5985–5993.
17. Rao, Y., X.-F. Wu, J. Gariepy, U. Rutishauser, and C.-H. Siu. 1992. Identification of a peptide sequence involved in homophilic binding in the neural cell adhesion molecule NCAM. *J. Cell Biol.* 118:937–949.
18. Hukkanen, E. J. 2004. A systems approach to the modeling and control of molecular, microparticle, and biological distributions. Ph.D. thesis. University of Illinois at Urbana-Champaign, IL.