Single-Molecule Force-Clamp Spectroscopy: Dwell Time Analysis and Practical Considerations†

Yi Cao† and Hongbin Li*

Department of Chemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada.
†Present address: Department of Physics, Nanjing University, Nanjing, Jiangsu Province, PR China

Received October 13, 2010. Revised Manuscript Received November 8, 2010

Single-molecule force-clamp spectroscopy has become a powerful tool for studying protein folding/unfolding, bond rupture, and enzymatic reactions. Different methods have been developed to analyze force-clamp spectroscopy data on polyproteins to obtain kinetic parameters characterizing the mechanical unfolding of proteins, which are often modeled as a two-state process (a Poisson process). However, because of the finite number of domains in polyproteins, the statistical analysis of the single-molecule force-clamp spectroscopy data is different from that of a classical Poisson process, and the equivalency of different analysis methods remains to be proven. In this article, we show that these methods are equivalent and lead to accurate measurements of the unfolding rate constant. We also demonstrate that distinct double-constant-pulling-velocity experiments, in which the unfolding rate extracted from the data is dependent on the number of protein domains in the polyproteins (the N effect), force-clamp experiments do not show any N effect. Using a simulated data set, we also highlighted important practical considerations that one needs to take into account when using the single-molecule force-clamp spectroscopy technique to characterize the unfolding energy landscape of proteins.

Introduction

Atomic force microscopy (AFM)-based single-molecule force spectroscopy has evolved into a powerful tool for studying the dynamics and the involved weak noncovalent interactions of many processes, including conformational changes of polymers (including synthetic polymers and biomacromolecules1−5 and DNA6−11), ligand binding,12−18 chemical/enzymatic reactions,19−24 and protein folding and unfolding.25−40 Over the past decade, the use of AFM to investigate the unfolding and folding dynamics of proteins has provided a tremendous amount of new information, opening up new avenues for studying protein folding and unfolding. To identify single-molecule stretching events unambiguously, polyprotein approaches are widely used in typical protein folding/unfolding experiments41−43 in which the protein of interest is constructed into a polymer consisting of identical tandem repeats of the protein. Stretching such polyproteins at a constant velocity gives rise to unique sawtooth-like force−extension curves with each force peak corresponding to the unfolding of a protein domain in the polypeptide chain. Such polyprotein approaches allow for single-molecule force

Figure 1. Schematics of force-clamp experiments and the force-clamp spectrum (red trace). (i) A polyprotein is picked up from the solid substrate by an AFM cantilever and held at a constant stretching force. (ii) The unfolding of a protein domain gives rise to the stepwise elongation of the polyprotein molecule. (iii) The length of the polyprotein molecule will remain constant after all of the domains are unfolded. The length—time profiles of polyproteins show a characteristic staircaselike appearance, where each step corresponds to the mechanical unfolding of an individual domain in the polyprotein chain. As shown in the schematic length—time curve, we define the time interval between consecutive unfolding events as the dwell time ($t_d$) and the time between zero and the unfolding event as the survival time ($t_s$).

The development of force-clamp and force-ramp spectroscopy has made it possible to stretch proteins under a predefined constant force or a force increasing at a constant rate. These new methods use a force feedback loop to precisely control the force being applied to a protein molecule and monitor its length changes over time. Stretching a polyprotein in force-clamp mode results in a staircase-like length—time profile where each step corresponds to a mechanical unfolding event of a protein domain (Figure 1). Because all of the domains are independent and experience the same stretching force, the unfolding traces of a polyprotein can be regarded as the superposition of individual unfolding events, as suggested by Fernandez and co-workers. Therefore, the unfolding probability or unfolding probability density can be directly measured, and the unfolding rate constant can be readily determined at a given stretching force. This led to two widely used data-analysis methods: length average and survival time. In these two methods, the number of domains in the polyproteins is not considered.

Recently, Fernandez and co-workers introduced a dwell-time analysis method to test the two-state Markovian process, which considers the total number of domains in the polyproteins and the order of unfolding events. In this method, the dwell time is defined as the survival time. In a different effort, we proposed a pseudo-dwell-time method, where the dwell time is defined as the time interval between consecutive unfolding events in force-clamp experiments. Apparently, these two methods explicitly considered the number of domains in polyproteins, and the results showed that the number of domains in the polyprotein affects the distribution of each dwell time that is categorized with respect to the order of unfolding events and the total number of domains, raising the question of whether there is a similar $N$ effect in force-clamp experiments on polyproteins and whether these two methods are equivalent to the length-average and survival-time methods. Similar questions also exist for force-ramp experiments.

In this article, we reconcile these issues by showing that these different data-analysis methods are equivalent and measure the same unfolding kinetics. We show that the unfolding kinetics of proteins in force-clamp experiments is not affected by the number of domains in the polyprotein and there is no $N$ effect in force-clamp or force-ramp experiments. Using simulated data sets, we also compared the accuracy and versatility of these different methods in some representative experimental situations (e.g., for different numbers of protein domains, small data pools, the absence of some initial unfolding events because of the presence of nonspecific interactions at the beginning of the traces, and the absence of late unfolding events because of the detachment of molecules from the cantilever or substrate). These practical considerations will help to optimize the experimental conditions in single-molecule force spectroscopy experiments to characterize the unfolding energy landscape of proteins.

Materials and Methods

Monte Carlo Simulation. The simulated data set for the mechanical unfolding of polyproteins using force-clamp spectroscopy was generated using a Monte Carlo procedure as described. The protein unfolding is modeled as a two-state process, and the unfolding rate constant depends on the stretching force following the Bell–Evans equation, where $\alpha(F)$ is the unfolding rate constant for a force $F$, $\alpha_0$ is the spontaneous unfolding rate constant in the absence of a stretching force, $\Delta x_u$ is the unfolding distance between the native state and the mechanical unfolding transition state, and $T$ is the temperature. The parameters

\[ \alpha(F) = \alpha_0 \exp(\frac{F \Delta x_u}{k_B T}) \]

were used to simulate the unfolding process of polyproteins.

References

used in the Monte Carlo simulations are $\alpha_0 = 0.0004 \text{s}^{-1}$, $\Delta x_0 = 0.25 \text{nm}$, $T = 298 \text{K}$, $\Delta t = 0.0001 \text{s}$, and $F = 140 \text{pN}$. The total number of domains is 10, and the total number of traces generated is 2000. This data pool was then used for all of the data analysis in this article. To evaluate the effect of the total number of domains on the unfolding rate, we used different numbers for the total number of domains and kept the remaining parameters unchanged.

**Data Processing and Fitting.** All of the data analysis was done using custom-written procedures in Igor 5.1 (WaveMetrics). For the length-average method, the unfolding traces were first normalized against the total length changes ($\text{length}_{\text{final}} - \text{length}_{\text{initial}}$). Then the normalized curves were averaged to obtain the final length-average traces. Exponential fitting to the length-average traces gave rise to the unfolding rate of the protein. The survival time and pseudo dwell time were calculated using the method specified in the Results section. Then, the histograms of both the survival time and pseudo dwell time were calculated with a bin size of 0.08 s. These histograms were fit to a single-exponential function to obtain the unfolding rate constant. For the force-ramp experiments and constant-velocity experiments, the unfolding probability densities were calculated using a bin size of 5 pN.

**Estimation of Fitting Errors.** To estimate the fitting errors of different sample sizes, we first randomly select a desired number of traces from the data pool with a total of 2000 traces. Then the selected subdata set was analyzed using length-average, survival-time, and pseudo-dwell-time methods to extract the unfolding rate. This procedure was repeated 50 times to get a set of unfolding rates with the same data size. The average and standard deviation of the 50 unfolding rates were used to estimate the quality of the fitting.

**Results and Discussion**

**Brief Introduction to Different Analysis Methods.** AFM-based single-molecule force-clamp spectroscopy uses a force feedback system to maintain the constant force at the preset value by varying the length of the proteins being stretched.\(^{51}\) As shown in Figure 1, the unfolding of a polyprotein in force-clamp mode gives rise to a staircaselike unfolding pattern with each step corresponding to the unfolding of a protein domain in the polyprotein. Here we define the time from the beginning ($t_0$) to the $k$th unfolding event as the survival time of the $k$th domain ($t_{s,N,k}$), where $N$ is the total number of domains being stretched. We define the dwell time ($t_{d,N,k}$) of the $k$th domain as the time interval between the $(k-1)$th and the $k$th unfolding events.

Different methods have been developed to analyze force-clamp data on polyproteins, and we focus on four of them: length-average, survival-time, dwell-time, and pseudo-dwell-time methods.

The length-average method is based on the reasoning that because individual domains in the polyprotein are identical and independent, the averaged normalized length $L_n(t)$ is equivalent to the unfolding probability $P_s(t)$.\(^{51}\) For two-state processes,

$$L_n(t) = P_s(t) = 1 - e^{-\alpha t}$$

Similarly, the distribution of the survival time equals the probability density of unfolding at time $t$.

$$f_s(t) = \alpha e^{-\alpha t}$$

The survival-time method fits the distribution of survival time to eq 2 to extract the unfolding rate at a given stretching force.

For the dwell-time analysis proposed by Fernandez et al.,\(^{53}\) the dwell time of the $k$th domain is defined as the survival time of the $k$th domain. It is of note that this definition is different from the arrival time used in a classical Poisson process.\(^{58}\) The unfolding probability of the $k$th domain to unfold at time $t$ is

$$P_s(t, N, k) = \frac{N}{k} \left((P_s(t))^k (1 - P_s(t))^{N-k}\right)^k$$

$$= \frac{N!}{k! (N-k)!} \left((P_s(t))^k (1 - P_s(t))^{N-k}\right)^k$$

The probability density of the unfolding of the $k$th domain is the probability of having exactly $k-1$ domains unfolded at time $t$ multiplied by the unfolding probability of the $k$th domain in the $[t, t + dt]$ time interval, which equals $(N-k+1)\alpha dt$.

$$f_d(t, N, k) = P_s(t, N, k-1)(N-k+1)\alpha$$

$$= \frac{N}{k-1} \left((P_s(t))^k (1 - P_s(t))^{N-k+1}\right)^k (N-k+1)\alpha$$

$$= \frac{N!}{(k-1)! (N-k)!} \left(1 - e^{-\alpha t}\right)^k e^{-\alpha t}$$

$$= \frac{N!}{(N-k-1)!} (1 - e^{-\alpha t})^{k-1} e^{-\alpha t}$$

$$= \frac{N!}{(N-k-1)!} (1 - e^{-\alpha t})^{k-1} e^{-\alpha t}$$

This equation is normalized from zero to infinity over $t$. In the unfolding probability density of the $k$th domain out of $N$ domains, $N$ is explicitly considered in eq 3. This is different from the gamma function used to describe the classical Poisson processes.\(^{58}\)

For pseudo-dwell-time analysis, the dwell time of the $k$th domain is defined as the time interval between the $k$th and $(k+1)$th domain,\(^{55}\) which is the same definition as the arrival time in a classical Poisson process.\(^{58}\) When considering the number of domains in the polyprotein, the unfolding probability of the $k$th domain of a polyprotein with a total of $N$ domains is

$$P_d(t, N, k) = 1 - e^{-\alpha (N-k+1)\alpha t}$$

and the probability density function is

$$f_d(t, N, K) = (N-k+1)\alpha e^{-\alpha (N-k+1)\alpha t}$$

The pseudo dwell time $t'$ is then defined as $t' = (N-k+1)t$, and eq 6 can be rewritten as

$$f_d(t', N) = \alpha e^{-\alpha t'}$$

The unfolding rate constant can then be obtained by fitting eq 7 to the distribution of the pseudo dwell time.

In all four methods, only the length-average method uses the probability function to fit the experimental data. The other three methods extract the unfolding kinetics by fitting the probability density function to the distribution of time.

As the first attempt to determine whether these methods are equivalent and lead to the correct estimate of unfolding rate constant, we synthesized a data pool of 2000 unfolding traces (10 unfolding events in each traces) using a standard Monte Carlo simulation as described in the Materials and Methods section. As shown in Figure S1, all four methods can adequately recover the input unfolding rate constant of 2.04 s\(^{-1}\) at a stretching force of 140 pN. Moreover, the equations can also reproduce the distribution of the experimental data, which is a validation of these data-analysis methods.

**No N Effect in Force-Clamp Experiments.** In all four data-analysis methods, dwell-time and pseudo-dwell-time methods

---

explicitly consider the number of domains in the polyprotein. The probability density functions (eqs 4 and 6) contain $N$ and are different from the corresponding functions in a classical Poisson process. In contrast, length-average and survival-time methods are based on the assumption that the unfolding traces of a polyprotein can be regarded as the superposition of individual unfolding events, and $N$ is not in the probability and probability distribution functions.

It is of note that the survival time of the $k$th domain out of $N$ domains follows eq 4. If the dwell-time and survival-time methods are equivalent, then summing all the survival times (from 1 to $N$) for eq 4 should lead to the same unfolding probability density function, $f_d(t) = \alpha \exp(-\alpha t)$. Here we provide a mathematical proof for this equivalency.

$$f_s(t, N) = \sum_{k=1}^{N} f_s(t, N, k)$$

$$= \sum_{k=1}^{N} \left( \sum_{j=0}^{N-k} \frac{N}{j} (P_x(t))^{j-1} (1 - P_x(t))^{N-j} (N-k+1) \alpha \right)$$

$$= \sum_{j=0}^{N} \left( \sum_{k=1}^{N-j} \frac{N}{k} (P_x(t))^{k-1} (1 - P_x(t))^{N-k} (N-k+1) \alpha \right)$$

$$f_s(t, N) = \sum_{j=0}^{N} \left( \sum_{k=1}^{N-j} \frac{N}{j} (P_x(t))^{j-1} (1 - P_x(t))^{N-j} \alpha \right)$$

$$= N \alpha - N \alpha P_x(t) \left( \sum_{j=1}^{N-1} \frac{N-j}{j} (P_x(t))^{j-1} (1 - P_x(t))^{(N-1)-(j-1)} \right)$$

$$= N \alpha - N \alpha P_x(t) \left( \sum_{j=1}^{N-1} \frac{N-j}{j} (P_x(t))^{-1} (1 - P_x(t))^{(N-1)-(j-1)} \right)$$

Because $P_x(t) = 1 - \exp(-t)$, we can write $f_s(t, N) = N \alpha \exp(-t)$. The above derivation is based on the following equation:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} b^k a^{n-k}$$

Let $a = 1 - P_x(t)$ and $b = P_x(t)$, hence

$$\sum_{j=0}^{N} \frac{N}{j} (P_x(t))^{j} (1 - P_x(t))^{N-j} = 1$$

$$\sum_{j=1}^{N-1} \frac{N-j}{j} (P_x(t))^{j-1} (1 - P_x(t))^{(N-1)-(j-1)} = 1$$

Because $f_d(t) = \alpha \exp(-\alpha t)$, after normalization we get $f_d(t, N) = \alpha \exp(-\alpha t)$. This result clearly indicated that the dwell-time method is equivalent to the survival-time and length-average methods, and the number of domains $N$ in the polyprotein does not affect the unfolding kinetics. Hence, there is no $N$ effect in force-clamp experiments on polypeptides.

For pseudo-dwell-time analysis, the definition of dwell time is the same as that in classical Poisson processes. However, because of finite $N$, different dwell times no longer follow the same exponential distribution as for classical Poisson processes. Instead, by correcting the effect of a finite number of folded domains in a polyprotein, the pseudo dwell time follows an identical distribution and there is no $N$ effect.

The results clearly show that these methods are equivalent. By choosing the appropriate experimental parameter for analysis, there is no $N$ effect. However, it is of note that a finite number of domains in the polyprotein does make the unfolding of the polyprotein slightly different from that in classical Poisson processes in that both the survival time and dwell time are now dependent upon $N$.

On the basis of the proof of equivalency of the methods, we can immediately draw the conclusion that we can take into account of all of the complete unfolding trajectories of different numbers of domains in the data pool for the analysis using length-average, survival-time and pseudo-dwell-time methods. Analysis of the simulated data set generated by Monte Carlo simulations confirmed this point (Figure 2). However, for a dwell-time analysis that relies on $N$, only trajectories of the same $N$ can be grouped together for analysis. This can significantly limit the size of the data pool and makes this method impractical in most experiments. In the rest of the discussion, we focus on length-average, survival-time, and pseudo-dwell time methods.

**No $N$ Effect in Force-Ramp Experiments.** Different from force-clamp experiments, which measure the dwell time under constant force, in force-ramp experiments the force increases...
linearly and the unfolding force at which each individual domain unfolds is measured to extract the unfolding kinetics. If the unfolding trace of the force-ramp experiments can be considered to be the superposition of $N$ independent single-molecule unfolding events, then the most probable unfolding force should be independent of the number of domains $N$. Following the same approach for force-clamp experiments, we define the unfolding probability density of the $k$th domain out of $N$ domains to unfold at force $F$ as $f(k, N, F)$. The sum of $f(k, N, F)$ should be equal to the unfolding probability density of a single domain, which is of the form \(^{57,60}\)

$$f(F) = \frac{a}{kT} e^{\frac{kF}{a} x_u} - a e^{\frac{kF}{a} x_u} (e^{\frac{kF}{a} x_u} - 1)$$

(8)

where $a$ is the force-ramp rate and the other symbols are as defined as before. We also attempted to prove this point mathematically. However, such a derivation cannot be done analytically because the unfolding probability density function of the $k$th event is conditional on that of the $(k - 1)$th event and $f(1, N, F)$ has a complex expression similar to eq 8.

$$f(1, N, F) = \frac{N a}{a} e^{\frac{kF}{a} x_u} - \frac{N a}{a} (e^{\frac{kF}{a} x_u} - 1)$$

(9)

Instead, we used Monte Carlo simulation to show that indeed the results are consistent with this statement. We generated 2000 force-ramp unfolding traces using Monte Carlo simulation and constructed the unfolding-force distribution for the $i$th (from 1 to $N$) domain in the data set. It is clear that the sum of the unfolding-force distribution from the first domain to the $N$th domain is the same as the unfolding-force distribution of a single domain measured in force-clamp experiments (Figure 3A), suggesting that unfolding trajectories measured from force-ramp experiments can be considered to be the superposition of $N$ unfolding trajectories of a single domain under the same force-ramp condition. Therefore, we can readily combine force-ramp traces of a different number of domains to construct the unfolding-force histogram (the unfolding probability density function). The most probable unfolding force at the given ramp rate is independent of the number of domains in the polypeptide chain, and there is no $N$ effect in the force-ramp experiment. This is in sharp contrast to the constant-velocity experiments (Figure 3B). For example, the average unfolding forces decrease from \(\sim 300\) pN at $N = 1$ to \(\sim 200\) pN at $N = 10$ in constant-velocity experiments. Mixing the data with different $N$ significantly broadens the unfolding-force distribution and may lead to an incorrect estimation of the unfolding distance because the width of the unfolding-force distribution is directly related to the unfolding distance between the native state and the mechanical-unfolding transition state.

What is the origin of the $N$ effect? One intuitive thought might be that the total number of domains in the polypeptides speeds up the unfolding rate of the $k$th unfolding events by a factor of $(N - k + 1)$ because there are $(N - k + 1)$ domains folded in the polypeptide. However, the mathematical proof and Monte Carlo simulation showed that this is not the case. As correctly pointed out by Smith and co-workers,\(^{59}\) the unfolding force is directly related to the effective loading rate at the point of rupture. It is the difference in the loading rate that causes the $N$ effect. In constant-velocity experiments, the force increases with the length of the molecule and the length of the molecule depends on the total number of domains ($N$) and the number of unfolded domains in the polypeptide ($k$). Therefore, the loading rate is a function of $N$. A larger value of $N$ results in a smaller average loading rate. However, in force-ramp and force-clamp experiments, the loading rates of all of the domains are identical and independent of $N$. To extend this point, as long as the force feedback system is used, any waveform of force, linear or nonlinear, can be used in force-clamp experiments and no $N$ effect will be present.

This point is of critical importance in defining the mechanical stability of a protein. Mechanical stability is often defined as the average unfolding force at a given pulling velocity in constant-velocity experiments. However, because the effective loading rate in constant-velocity experiments is affected by many factors such as the polypeptide construct (with different values of $N$ and different linker lengths) and the cantilever spring constant, the mechanical stability of the same protein measured in different laboratories is often quite different, making the comparison of protein mechanical stability difficult. However, force-ramp spectroscopy experiments can effectively circumvent this problem, making these the ideal choice in single-molecule force spectroscopy experiments. Hence, we propose that the mechanical stability of proteins should be reported as the average unfolding force at a given ramp rate in force-ramp spectroscopy experiments to facilitate the comparison of results from different laboratories.

---

Length-Average Method May Be More Suitable for Extracting the Unfolding Rate from Small Data Pools. Because of the difficulties in single-molecule experiments, it is challenging to get an experimental data pool as large as the simulated data pools. How much data is sufficient to give an accurate estimation of the unfolding rate constant? Are the fitting errors for all three data-analysis methods similar? To address these issues, we estimated the fitting errors from simulated data pools of different sizes. We randomly selected the desired number of traces from a total of 2000 traces and obtained the unfolding rate constant using the three different methods. This process was repeated 50 times to compare the mean and standard deviation of the unfolding rate constant from different analysis methods. As shown in Figure 4A, the average unfolding rate constants from the 50 trials are around the input number of 2.04 s\(^{-1}\) for all three methods, which again indicates the correctness of these analysis methods and the accuracy of the fittings. We used the standard deviation of the unfolding rate constant from 50 trials as the indicator for the precision of the fit. The length-average method gave much smaller deviations than the other two methods (Figure 4B). Less than 5% error in the unfolding rate constant was obtained when using around 100 unfolding events whereas for the survival-time and pseudo-dwell-time methods obtaining less than 10% error in the unfolding rate constant required the use of around 400 unfolding events (dotted line). This finding is somehow unexpected but reasonable. Among these three methods, only the length-average method calculates the unfolding probability at different times instead of the probability density. In the length-average method, the data at each \(t\) is essentially integrated from time 0 to \(t\), which increases the effective data pool and reduces the error in fitting.

Choice of the Starting Time Does Not Affect the Estimated Unfolding Rate Constant. In single-molecule AFM experiments, the beginning parts of the force spectroscopy traces are often “contaminated” by features associated with nonspecific interactions, making it difficult to determine the starting time, \(t_0\), for each trace. To examine whether the choice of starting time will lead to the deviation of the estimated unfolding rate constant from the actual value, we changed the \(t_0\) for all 2000 simulated traces to \(t_0' = 0.2\) s. Such a change makes some unfolding traces have less than 10 unfolding events because the unfolding events in between \(t_0\) and \(t_0'\) are missed. Then, these new data were analyzed using all three methods. Such changes did not have any effect on the estimated unfolding rate constant for all three methods, as shown in Figure 5. This reflects important features of a Poisson
process: no memory and history-independent. This is important for experiments because we can effectively include trajectories contaminated by the nonspecific interactions at the beginning of the unfolding trajectories. However, such convenience may not be applicable to unfolding processes that are non-Poisson in nature.

**Missing Initial Fast Unfolding Events Affect the Estimated Unfolding Rate Constant.**

We may also face a difficult situation in choosing the starting point of the traces for real experimental data. Because of the limited response of the force-feedback loop, the first few fast unfolding events may be missed. This leads to the wrong estimation of the starting point for those traces, as shown in Figure 6A. In this situation, the wrong \( t_0 \) does not affect all traces but only those with fast initial unfolding events.\(^{54,55}\) We used the simulated data to test whether this biases the estimated unfolding rate constant. We artificially changed the starting points of all traces to the position of the first unfolding event with a dwell time longer than 0.03 s. Although the force feedback for state-of-the-art AFM instruments can be as fast as 1–3 ms,\(^{54}\) we choose 0.03 s as the threshold just to make the effect obvious. As shown in Figure 6B–D, because of missing fast unfolding events, the data for all three analysis methods deviated from a single-exponential distribution. Simply applying a single-exponential fit to the data will cause significant bias in the input unfolding rate constant (red line in Figure 6B and gray lines in Figure 6C,D). However for both survival-time and pseudo-dwell-time methods, if we omit the first few bins of the distribution we can still get the correct unfolding rate constant from the data (red lines in Figure 6C,D). For force-clamp experiments carried out at a high clamp force, the unfolding of protein domains can be very fast at the beginning of the traces. Therefore, choosing a suitable data-analysis method is very important. It is worth noting that missing short unfolding events have been reported in many force-clamp experiments.\(^{54,55}\) To obtain unbiased data, the response

![Figure 6](image1.png)

**Figure 6.** Missing fast unfolding events may bias the data analysis. (A) Initial fast unfolding events may be missed if an improper \( t_0 \) is chosen in the analysis. (B) Length-average, (C) survival-time, and (D) pseudo-dwell-time methods; black lines represent the data with missing fast unfolding events (\( t = 0.02 \) s is the cutoff), and red lines and gray lines correspond to single-exponential fitting. (C, D) Gray lines are fits to all data, and red lines omit the first few data points in the fitting.

![Figure 7](image2.png)

**Figure 7.** Missing slow unfolding events due to a short observation time window or detachment of the polyprotein molecule either from the cantilever or the substrate may bias the obtained unfolding rate.
speed of the force-feedback loop may determine the limit of the clamped force in force-clamp experiments used to obtain unbiased data.

**Including Unfinished Traces Significantly Affects the Estimated Unfolding Rate.** In single-molecule AFM experiments, another potential error comes from including unfinished traces in the data analysis. This may occur quite often in low-force force-clamp experiments that require a long time for all of the domains in the polyprotein to unfold. However, if the polyprotein is detached from either the cantilever or the substrate before all domains in the polyprotein unfold, then the total number of domains in the polyprotein is also unknown. In these scenarios, the total $N$ in the polyprotein is unknown. Because we have proven that there is no $N$ effect in force-clamp experiments, does missing some unfolding events affect the estimated unfolding rate? Because the missed events mainly occur at long survival times, from intuition we may expect that including unfinished traces may bias the unfolding rate to higher value. We then used the simulated data to demonstrate how significant this effect could be for all three data-analysis methods. We randomly detached half of the unfolding traces at time $t = 1.5$ s to mimic the real experimental condition. As shown in Figure 7, including unfinished traces significantly made the estimated unfolding rate $s$ for all three methods faster than the real value. However, among all three methods, the survival time can still provide a reasonably good estimation of the unfolding rate with an error of less than 10%. It is worth noting that if the number of domains in the polyproteins is different, then including unfinished traces may bias the data toward a single-exponential model and show an artificial multiple-exponential distribution. This may result in a wrong interpretation of the unfolding mechanism of proteins. Therefore, excluding the unfinished traces in the data analysis is critical for single-molecule force-clamp experiments. Moreover, for the pseudo-dwell-time method, if the total number of domains in the polyprotein being stretched is known (e.g., in a refolding experiment), then the unfinished traces can still be used for the analysis, as we demonstrated previously.

**Conclusions**

Force-clamp (including force-ramp) spectroscopy has played an increasingly important role in the study of protein unfolding dynamics. The development of different data-analysis methods greatly facilitates the extraction of the unfolding kinetics of proteins from force-clamp experiments. Here, by combining mathematical derivation and Monte Carlo simulation, we demonstrated that these commonly used methods of analyzing force-clamp data are equivalent and provide an accurate estimation of the unfolding rate of proteins. Moreover, we showed that the force-clamp and force-ramp experiments do not have any $N$ effect, making the extraction of unfolding kinetics from them easier than that from constant-pulling-velocity experiments. Then, we showed that the accuracy and precision of these methods under some practical conditions can vary because of the different mathematical treatment of the experimental data. This comparison will help to optimize the experimental conditions in single-molecule force spectroscopy experiments conducted to characterize the unfolding energy landscape of proteins. It is of note that the data-analysis methods available for force-clamp experiments are not limited to the ones discussed in this article. Other methods have also been developed. Further developments of new analysis methods will be of great importance in extending the use of force-clamp techniques to analyze more complex unfolding kinetics of proteins that deviate from simple two-state unfolding kinetics.

**Acknowledgment.** This work is supported by the Natural Sciences and Engineering Research Council of Canada, the Canada Foundation of Innovation, and the Canada Research Chairs Program.

**Supporting Information Available:** Plots of data from four different analysis methods. This material is available free of charge via the Internet at http://pubs.acs.org.