VIP Single-Molecule Kinetics

DOI: 10.1002/anie.201104033

Signal-Pair Correlation Analysis of Single-Molecule Trajectories**

*Armin Hoffmann and Michael T. Woodside**

Single-molecule (SM) methods have dramatically expanded the study of dynamic processes in biomolecules. Since the first studies of single ion channels,[1] techniques such as Förster Resonance Energy Transfer (FRET) and force spectroscopy[2] have emerged which can be used to study a broad range of phenomena. By directly observing structural changes in a single molecule over time, the states occupied by the molecule can be identified, the possible transitions between them mapped out, and the transition rates measured. Such information has been used to build uniquely detailed pictures of various macromolecular processes, from ion-channel function[1] to molecular-motor motion,[3] enzymatic activity,[4] RNA-based regulation,[5] and biomolecular folding.[6,7]

A key feature of SM approaches is the ability to observe and characterize sub-populations, especially rare or transient states. Kinetic analysis of these states can, however, be challenging. Often the measured signal trajectories are mapped onto a set of discrete states,[8] for example by using thresholding or step-finding algorithms,[9] or more-sophisticated maximum-likelihood methods,[10,11] especially in combination with hidden Markov modeling (HMM).[12,13] Dwell-time distributions are then used to obtain kinetic information from the state trajectories.[14] A drawback of this approach is that any factors hindering state identification (e.g. noisy or overlapping signals, large differences in lifetimes) can introduce biases for which it is difficult to control.

An alternate strategy is to extract kinetics directly by analyzing correlations in the signal record, which has a number of advantages: it is more robust against noise and filtering artifacts, correlation fitting functions are easily calculated, and the fits can be used to test kinetic models directly.[15] Signal–intensity correlations have been used previously to analyze phenomena such as photo-physical processes[16,17] and two-state folding,[18] but such analysis has proven challenging for multi-state systems or processes having similar timescales, because the rates for all the transitions in the system are folded into a single correlation function from which they are difficult to recover separately.[19]

Herein we present a new type of correlation analysis, based not on the entire signal but rather on discrete ranges of the signal associated with different states, which allows transition rates and kinetic schemes to be determined even in multi-state systems. Only part of the signal associated with a given state is needed for correlation analysis, hence the ranges can be chosen to minimize overlap between states in noisy data. A related approach was recently applied to protein diffusion measured by SM FRET.[20] This signal-pair correlation method works even for trajectories with states that overlap because of noise, states with low occupancy, and rates that are very similar or differ by orders of magnitude, all issues that can hinder other approaches.

The method involves two steps: First, the signal (extension, FRET efficiency, current …) is divided into discrete ranges and the time correlations between all pairs of ranges ("signal pairs") are calculated. Next, specific kinetic models are tested by assigning a certain signal range to each state and fitting all cross-correlations between them with the functions derived for a given kinetic scheme. By repeating the fits for all possible schemes, the correct scheme can be validated empirically and the associated rates determined. The selection of signal ranges is simplified by using signal-pair histograms. These contain valuable information about the states present and the transitions between them, without the need to identify transitions or states.

To demonstrate this signal-pair correlation analysis, we used force-spectroscopy measurements of folding in three different molecules: 1) a two-state-folding DNA hairpin, for comparing the correlation analysis to the thresholding method;[21] 2) a three-state-folding DNA hairpin with a known kinetic scheme and overlapping states;[7] and 3) the hamster prion protein (HaPrP), which can fold into non-native structures.[22] Folding trajectories consisting of records of the molecular extension as a function of time were measured at equilibrium for all molecules using dual-beam optical tweezers with a passive force clamp (Figure 1a).[23]

We first analysed the two-state-folding DNA hairpin (Figure 1b, inset). Two states are very clearly observed both in the trajectory itself (Figure 1b) and in the extension histogram (Figure 1c): one at 548 nm representing the folded hairpin (F), the other at 559 nm for the unfolded hairpin (U). We also calculated 2D signal-pair histograms indicating how often an initial extension e_1 at time t led to an extension e_2 after time $t+\tau$; three different delay times τ are shown (Figure 1d–f). The number of events N in each 2D bins is given by Equation (1)

$$N(\tau, e_1, e_2) = \{\, e(t) \in e_1 \pm \Delta e \,|\, e(t+\tau) \in e_2 \pm \Delta e \,\} \quad (1)$$

where Δe is the bin half-width (in this case, 0.25 nm). Note that these signal-pair histograms are significantly different

[*] Dr. A. Hoffmann, Prof. Dr. M. T. Woodside
Department of Physics, University of Alberta, and National Institute of Nanotechnology, NRC
Edmonton AB T6G 2M9 (Canada)
E-mail: michael.woodside@nrc-cnrc.gc.ca
Homepage: http://www.ualberta.ca/~mwoodsid/Home.html

[**] We thank H. Yu and Dr. X. Liu for the prion protein data, and A. Brigley, A. Solanki, and Dr. I. Sosova for the prion protein constructs. We thank PrioNet Canada, Alberta Prion Research Institute, nanoWorks (Alberta Innovates), and the National Institute for Nanotechnology for funding.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201104033.

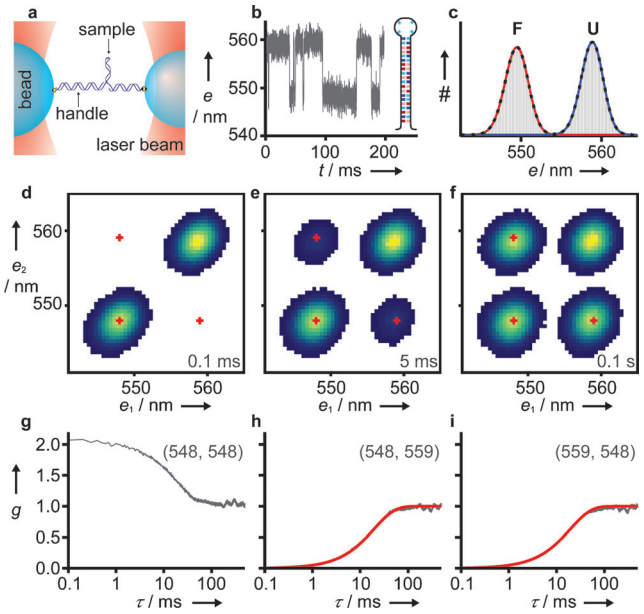

Figure 1. Signal-pair correlation analysis of DNA hairpin folding measured by force spectroscopy. a) Measurement scheme: the hairpin is held under tension between two optical traps, and the extension measured as a function of time. b) Extension trajectory showing two states. Inset: scheme of the hairpin with color-coded sequence (A/T: dark/light blue, G/C: dark/light red). c) The extension histogram shows two distinct peaks, folded (F) and unfolded (U). d)–f) Histograms of signal pairs at different time delays reveal dynamic information. Color scale: dark blue to yellow (2% to ≥50% of largest category ("bin")). Red crosses mark signal pairs (e_1, e_2) for correlation functions in (g)–(i). g) Autocorrelation of the folded state. h),i) Cross-correlations from F to U (h) and U to F (i) fit well to a two-state model (red).

from the similar-appearing and commonly-used transition maps,[24] since we only plot signal pairs for a given τ, without assigning states or transitions. For $\tau = 0.1$ ms, two peaks are seen (Figure 1 d) along the identity line $e_1 = e_2$, representing the two states. The width of the peaks arises from dynamic processes that are fast on the timescale of τ but do not significantly change the molecular extension: in this case, the diffusion of the beads attached to the hairpin.[23] As τ increases, two peaks emerge, (Figure 1 e) and grow in magnitude (Figure 1 f), reflecting a slower, dynamic process that changes the extension, namely the folding/unfolding of the hairpin.

To analyze the kinetics quantitatively, we calculated signal-pair correlation functions, $g(\tau,e_1,e_2)$, derived from Equation (1), varying τ instead of e_1 and e_2 [Eq. (2)].

$$g(\tau, e_1, e_2) = \frac{p(\{e_1, t\} \,|\, \{e_2, t+\tau\})}{p(e_1)\, p(e_2)} = \frac{N(\tau, e_1, e_2)\, T^2}{(T-\tau) N(e_1) N(e_2)} \qquad (2)$$

Here, $p(\{e_1,t\}\,|\,\{e_2,t+\tau\})$ is the joint probability to measure extension e_1 at time t and extension e_2 at time $t+\tau$, $p(e_i)$ is the probability to measure extension e_i, $N(e_i)$ is the number of time bins measuring extension e_i, and T is the total number of time bins. Correlation functions were calculated for all signal pairs (1 nm wide ranges) with sufficient counts; three are shown in Figure 1 g–i with signal pairs near the peak centers in Figure 1 d–f. As expected, the autocorrelation ($e_1 = e_2$: Figure 1 g) shows a positively correlated signal, whereas the cross-correlations ($e_1 \neq e_2$: Figure 1 h,i) are negatively correlated and start from 0. All correlations change on the same time scale (about 10 ms).

To determine the microscopic folding rates, we fitted the cross-correlation signals to functions derived by matrix methods as described previously for fluorescence correlation analysis[25] (see Supporting Information). We reduced the number of fit parameters to a single parameter by using the fractional occupancies of the states obtained from the 1D histogram (Figure 1 c) to relate folding and unfolding rates, according to the principle of detailed balance: if f_i is the fraction of state i and $k_{i,j}$ the rate of the transition from state i to j, then $k_{ij} = k_{ji}\,(f_i/f_j)$. Fitting 25 extension-pair combinations close to the ones shown in Figure 1 h,i, we obtain $k_{F,U} = k_{U,F} = (25 \pm 2)\ s^{-1}$, where the uncertainty is the standard error on the mean. These values agree with the results of the thresholding method ($k_{F,U} = (27 \pm 2)\ s^{-1}$ and $k_{U,F} = (25 \pm 2)\ s^{-1}$).

Having demonstrated that signal-pair correlation analysis works for a simple two-state system, we next applied it to a DNA hairpin whose stem sequence was designed to produce a partially folded, on-pathway intermediate[7] (Figure 2 a, inset). This intermediate (I) is seen in the trajectory (Figure 2 a) and extension histogram (Figure 2 b) at 521 nm; F is at 508 nm and U at 528 nm. Note that around 30% of I overlaps U, making these states difficult to resolve well by thresholding analysis (Supporting Information, Figure S1–S3). As with the two-state hairpin, the signal-pair histogram for $\tau = 0.1$ ms shows peaks only along the identity line $e_1 = e_2$: in this case, there are three peaks (Figure 2 c). At $\tau = 1$ ms, small cross-peaks appear between F and I as well as between I and U (Figure 2 d). By $\tau = 10$ ms, significant cross-peaks also arise between F and U (Figure 2 e). The three cross-correlations shown in Figure 2 f–h correspond to the signal pairs (ranges 1 nm wide) centered on the red crosses in Figure 2 c–e and reflect folding dynamic processes at approximately 1 and 10 ms.

To extract the microscopic rate constants and confirm empirically the expected sequential folding pathway of this hairpin, we chose three signal ranges corresponding to F, I, and U, balancing maximal signal with minimal overlap between states within each range. We then globally fitted the six cross-correlations between the chosen ranges (e.g. as in Figure 2 f–h and Supporting Information, Figure S5) to functions derived for each of the three different kinetic schemes in Figure 2 i. Again, we used the fractional state occupancies from the extension histogram (Figure 2 b) and detailed balance to reduce the number of fitting parameters for all six correlations to $k_{I,F}$ and $k_{U,I}$ (see the Supporting Information). Only the sequential model (Figure 2 f–h, red) fits the data well, confirming that the intermediate is on-pathway. Fits to 125 extension pairs close to those shown in Figure 2 c–e yield $k_{I,F} = (0.21 \pm 0.01)\ ms^{-1}$, $k_{U,I} = (79 \pm 4)\ s^{-1}$, $k_{F,I} = (34 \pm 1)\ s^{-1}$, and $k_{I,U} = (0.43 \pm 0.03)\ ms^{-1}$. We note that neither these four rates nor the kinetic model could be extracted from a standard correlation analysis (Supporting Information, Figure S4). By repeating the signal-pair correlation analysis for trajectories measured at different forces (Supporting

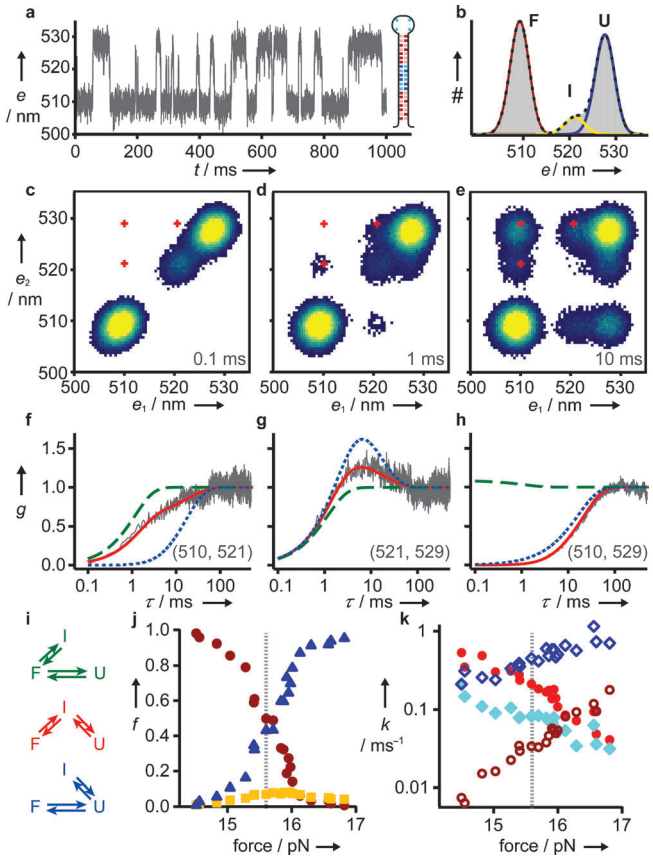

Figure 2. Analysis of a three-state folding DNA hairpin. a) The extension trajectory reveals 3 states. Inset: scheme of the hairpin with color-coded sequence as in Figure 1. b) The extension histogram shows an intermediate state (I) overlapping the unfolded state (U). c)–e) Signal-pair histograms at different time delays. f)–h) The cross-correlations for signal pairs (e_1,e_2) were fitted globally to three different kinetic schemes i) off-pathway (green from F, blue from U) or on-pathway (red). The fits indicate the intermediate is on-pathway. j) Force-dependent state occupancies from extension histograms: F (red circles), I (orange squares), U (blue triangles). k) Microscopic rates for the on-pathway mechanism: $k_{F,I}$ (dark red open circles), $k_{I,F}$ (light red filled circles), $k_{I,U}$ (dark blue open diamonds), and $k_{U,I}$ (light blue filled diamonds). Dotted gray lines in (j) and (k) mark the force for plots (a)–(h).

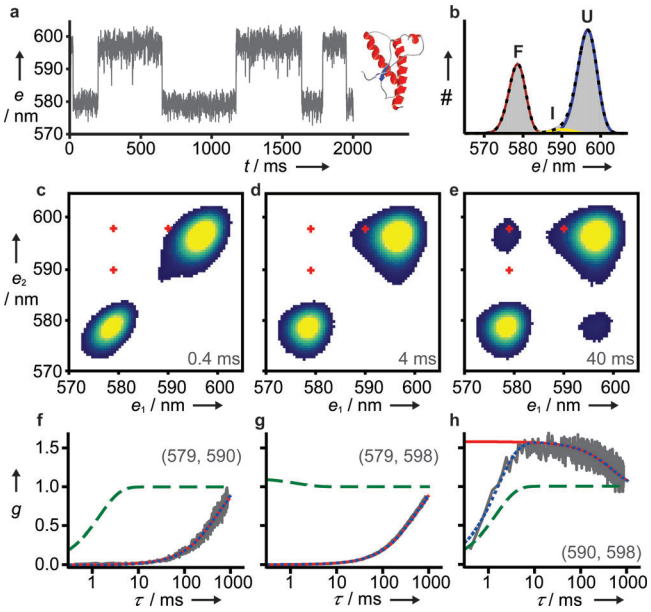

Figure 3. Analysis of HaPrP folding. a) The folding trajectory shows two major states. Inset: protein structure (Protein data bank (PDB) code: 1B10). b) The extension histogram reveals a rarely-occupied intermediate state I (yellow). c)–e) Signal-pair histograms at different time delays. Red crosses: signal pairs used for correlation functions. f)–h) Cross-correlations for the signal pairs centered at (e_1,e_2), fit globally to the same models as in Figure 2i, reveal that I is off-pathway, entered exclusively from U.

Information, Figure S5), where the state occupancies change from mostly folded to mostly unfolded as the force is increased (Figure 2j), the force-dependence of each of the microscopic rates was determined (Figure 2k). The method clearly works well even for low-occupancy states (Supporting Information, Figure S5). Knowing the force-dependent rates allows properties such as the height and location of the folding energy barrier, to be determined for each transition.[26]

As a final example, we applied the method to a folding trajectory of the protease-resistant fragment of the C179A/C214A mutant of HaPrP (Figure 3a, inset). The trajectory (Figure 3a) and extension histogram (Figure 3b) show F at 579 nm, U at 598 nm, and I at 591 nm. The low occupancy and short lifetime of I, compounded with the strong overlap between I and U (about 75% of I overlaps U), make these data very difficult to analyze with standard methods. The signal-pair histogram at $\tau = 0.4$ ms (Figure 3c) shows two major peaks for F and U, and a small peak for I. Due to its short lifetime, this peak for I vanishes by $\tau = 4$ ms (Figure 3d), but cross-peaks between I and U remain. By $\tau = 40$ ms (Figure 3e), additional cross-peaks appear between F and U.

Cross-correlations calculated for the signal pairs (ranges 1 nm wide) shown in Figure 3c (red crosses) reveal that F changes into the U and I after hundreds of ms (Figure 3f,g). The I to U correlation (Figure 3h) also has a rapid component at approximately 1 ms. Fitting all six signal-pair cross-correlations globally to the kinetic schemes in Figure 2i (Supporting Information Figure S6), we find that sequential folding is inconsistent with the data. Instead, the fits indicate that I is off-pathway, reached only from U (Figure 2i, blue). Fitting 125 signal-pair combinations close to those shown in Figure 3c yielded $k_{F,U} = (1.43 \pm 0.03)\ s^{-1}$, $k_{U,F} = (0.83 \pm 0.02)\ s^{-1}$, $k_{I,U} = (0.6 \pm 0.1)\ ms^{-1}$, and $k_{U,I} = (26 \pm 4)\ s^{-1}$. This example shows that the method can identify rates and kinetic schemes even with states that are difficult to distinguish and rarely populated, and with processes occurring on timescales differing by several orders of magnitude.

The method we have demonstrated herein is clearly applicable to a broad range of systems, measurement modalities, and simulations, beyond force spectroscopy of folding. It is also readily generalized to combine different types of signal, for example, from multi-parameter fluorescence spectroscopy, and can be extended beyond signal-pairs to signal-multiples to improve the state selection. A key feature of the method is that the assignment of states, while intuitive

(and simplified using the signal-pair histograms), is done at the end of the analysis—when choosing correlations for fitting—thereby avoiding many of the biases that arise when converting signals into state trajectories, especially with noisy data or short-lived states. Because of this, the method is also useful as a complementary tool for validating the results from other analytical approaches, such as hidden Markov models based on specific assumptions about the system under study. Signal-pair correlation analysis thus provides a powerful approach for extending the reach of single-molecule data.

Received: June 12, 2011
Revised: September 20, 2011
Published online: November 4, 2011

Keywords: biophysics · kinetics · single-molecule studies

[1] E. Neher, B. Sakmann, *Nature* **1976**, *260*, 799 – 802.
[2] W. J. Greenleaf, M. T. Woodside, S. M. Block, *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 171 – 190.
[3] R. D. Vale, T. Funatsu, D. W. Pierce, L. Romberg, Y. Harada, T. Yanagida, *Nature* **1996**, *380*, 451 – 453.
[4] H. P. Lu, L. Xun, X. S. Xie, *Science* **1998**, *282*, 1877 – 1882.
[5] K. Neupane, H. Yu, D. A. N. Foster, F. Wang, M. T. Woodside, *Nucleic Acids Research* **2011**, DOI: 10.1093/nar/gkr305.
[6] A. Borgia, P. M. Williams, J. Clarke, *Annu. Rev. Biochem.* **2008**, *77*, 101 – 125.
[7] M. T. Woodside, C. García-García, S. M. Block, *Curr. Opin. Chem. Biol.* **2008**, *12*, 640 – 646.
[8] A. E. Knight, C. Veigel, C. Chambers, J. E. Molloy, *Prog. Biophys. Mol. Biol.* **2001**, *77*, 45 – 72.
[9] L. P. Watkins, H. Yang, *J. Phys. Chem. B* **2005**, *109*, 617 – 628.
[10] R. Horn, K. Lange, *Biophys. J.* **1983**, *43*, 207 – 223.
[11] I. V. Gopich, A. Szabo, *J. Phys. Chem. B* **2009**, *113*, 10965 – 10973.
[12] F. G. Ball, J. A. Rice, *Math. Biosci.* **1992**, *112*, 189 – 206.
[13] M. Blanco, N. G. Walter, *Methods Enzymol.* **2010**, *472*, 153 – 178.
[14] J. R. Moffitt, Y. R. Chemla, C. Bustamante, *Methods Enzymol.* **2010**, *475*, 221 – 257.
[15] P. Labarca, J. A. Rice, D. R. Fredkin, M. Montal, *Biophys. J.* **1985**, *47*, 469 – 478.
[16] R. M. Dickson, A. B. Cubitt, R. Y. Tsien, W. E. Moerner, *Nature* **1997**, *388*, 355 – 358.
[17] C. Eggeling, J. Widengren, L. Brand, J. Schaffer, S. Felekyan, C. A. M. Seidel, *J. Phys. Chem. A* **2006**, *110*, 2979 – 2995.
[18] H. S. Chung, I. V. Gopich, K. McHale, T. Cellmer, J. M. Louis, W. A. Eaton, *J. Phys. Chem. A* **2010**, DOI: 10.1021/jp1009669.
[19] J. B. Witkoskie, J. Cao, *J. Chem. Phys.* **2004**, *121*, 6361 – 6372.
[20] A. Hoffmann, D. Nettels, J. Clark, A. Borgia, S. E. Radford, J. Clarke, B. Schuler, *Phys. Chem. Chem. Phys.* **2011**, *13*, 1857 – 1871.
[21] M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, S. M. Block, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 6190 – 6195.
[22] H. Yu, X. Liu, K. Neupane, A. N. Gupta, A. Brigley, A. Solanki, I. Sosova, M. T. Woodside, unpublished results.
[23] W. J. Greenleaf, M. T. Woodside, E. A. Abbondanzieri, S. M. Block, *Phys. Rev. Lett.* **2005**, *95*, 208102.
[24] H. S. Chung, J. M. Louis, W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11837 – 11844.
[25] I. V. Gopich, D. Nettels, B. Schuler, A. Szabo, *J. Chem. Phys.* **2009**, *131*, 095102.
[26] O. K. Dudko, G. Hummer, A. Szabo, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 15755 – 15760.

