

Signalpaar-Korrelationsanalyse von Einzelmolekültrajektorien**

Armin Hoffmann und Michael T. Woodside*

Einzelmolekülmethoden haben die Untersuchung dynamischer Prozesse von Biomolekülen enorm erweitert. Nach den ersten Studien an einzelnen Ionenkanälen^[1] wurden Methoden mit breiten Anwendungsspektren entwickelt, wie Förster-Resonanzenergietransfer (FRET) und Kraftspektroskopie.^[2] Die zeitaufgelöste Beobachtung von Strukturänderungen einzelner Moleküle ermöglicht die Bestimmung des jeweiligen Zustandes des Moleküls, der möglichen Übergänge zwischen diesen Zuständen und den entsprechenden Geschwindigkeiten. Solche Informationen haben es ermöglicht, sehr detaillierte Modelle zahlreicher makromolekularer Prozesse zu erstellen, zum Beispiel von der Funktion von Ionenkanälen,^[1] von Bewegungen molekularer Motoren,^[3] enzymatischer Aktivität,^[4] RNS-basierter Regulation^[5] und der Faltung von Biomolekülen.^[6,7]

Eine zentrale Eigenschaft von Einzelmolekülmethoden ist es, Unterpopulation – insbesondere seltene oder Übergangs-Zustände – zu beobachten und zu charakterisieren. Eine kinetische Analyse kann allerdings sehr anspruchsvoll sein. Oft wird die gemessene Signaltrajektorie auf eine diskrete Zustandstrajektorie übertragen,^[8] zum Beispiel mithilfe von Schwellwerten oder Algorithmen zur Sprungdetektion,^[9] oder mit Maximum-Likelihood-Methoden,^[10,11] insbesondere in Kombination mit Hidden-Markov-Modellen (HMM).^[12,13] Die kinetische Information kann anschließend anhand von Verweilzeitverteilungen innerhalb der Zustandstrajektorie ermittelt werden.^[14] Ein Nachteil dieses Vorgehens ist, dass jedwede Faktoren, die die Zustandszuordnung beeinträchtigen (z.B. Rauschen, Signalüberschneidung oder große Unterschiede der Lebensdauern), zu systematischen Fehlern führen können, die schwer zu bestimmen sind.

Eine alternative Strategie besteht darin, die kinetische Information direkt durch Korrelationsanalyse der Signaltrajektorie zu extrahieren. Der Vorteil hierbei ist, dass die Methode gegen Rauschen und Filterartefakte robuster ist, die Modellfunktionen einfach ermittelt werden können, und die Parameteranpassungsfunktionen als Test für das verwendete Modell dienen können.^[15] Signalintensitätskorrelationen wurden verwendet, um z.B. photophysikalische Prozesse^[16,17]

und Zweizustandsfaltung^[18] zu untersuchen. Die Analyse von Mehrzustandssystemen oder Prozessen auf ähnlichen Zeitskalen ist allerdings schwierig, da die Geschwindigkeiten aller Übergänge in einer einzelnen Korrelationsfunktion gefaltet sind.^[19]

Wir präsentieren eine neue Art von Korrelationsanalyse, die nicht auf dem gesamten Signal, sondern auf diskreten Signalebereichen basiert; diese können den verschiedenen Zuständen zugeordnet werden, womit die Übergangsgeschwindigkeiten und das kinetische Modell auch in Mehrzustandssystemen ermittelt werden können. Da nur ein Teil des Signals eines jeweiligen Zustands benötigt wird, können Bereiche mit minimaler Überschneidung zwischen den Zuständen verwendet werden. Ein ähnlicher Ansatz wurde kürzlich für Einzelmolekül-FRET-Messungen diffundierender Moleküle verwendet.^[20] Die Signalpaar-Korrelationsanalyse funktioniert bei Zuständen mit großer Überschneidung (z.B. wegen Rauschens), Zuständen, die nur gering populiert sind, sowie Geschwindigkeiten, die sehr ähnlich sind oder sich um mehrere Größenordnungen unterscheiden – Umstände, die andere Analysemethoden behindern können.

Die Methode besteht aus zwei Schritten: Zunächst wird das Signal (z.B. Streckung, FRET-Effizienz, Strom, ...) in diskrete Bereiche aufgeteilt, und die Zeitkorrelation zwischen allen Paaren von Bereichen („Signalpaare“) wird berechnet. Im zweiten Schritt werden verschiedene kinetische Modelle getestet, indem zunächst jedem Zustand ein Signalebereich zugeordnet wird und anschließend die Kreuzkorrelationen der Bereiche an die Parameter von modellspezifischen Funktionen angepasst werden. Nach Wiederholung des zweiten Schrittes mit allen möglichen kinetischen Modellen kann das richtige Modell ermittelt und die entsprechenden Geschwindigkeiten bestimmt werden. Die Wahl der Signalebereiche wird vereinfacht durch die Verwendung von Signalpaar-Histogrammen. Diese enthalten nützliche Informationen über die vorliegenden Zustände und deren Übergänge, ohne dass die Übergänge oder Zustände selbst definiert werden müssen.

Wir demonstrieren die Signalpaar-Korrelationsanalyse anhand von kraftspektroskopischen Messungen der Faltung von drei verschiedenen Molekülen: 1) einer Zweizustands-DNS-Haarnadel für den Vergleich mit der Schwellwertmethode;^[21] 2) einer Dreizustands-DNS-Haarnadel mit bekanntem kinetischen Modell und Überschneidung der Zustände;^[7] und 3) dem Hamster-Prionenprotein (HaPrP), das in nichtnative Strukturen falten kann.^[22] Die Faltungstrajektorien der molekularen Streckung wurden im Gleichgewicht gemessen mithilfe einer optischen Pinzette mit zwei Fallen und einer passiven Kraftklemme (Abbildung 1a).^[23]

Als erstes analysieren wir die Zweizustands-DNS-Haarnadel (Abbildung 1b, Einschub). Sowohl in der Trajektorie (Abbildung 1b) als auch im Streckungs-Histogramm (Abbil-

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

[**] Wir danken H. Yu und Dr. X. Liu für die Prionendaten und A. Brigley, A. Solanki und Dr. I. Sosova für die Prionenkonstrukte. Wir danken PrioNet Canada, Alberta Prion Research Institute, nanoWorks (Alberta Innovates) und dem National Institute for Nanotechnology für finanzielle Förderung.

Hintergrundinformationen zu diesem Beitrag sind im WWW unter <http://dx.doi.org/10.1002/ange.201104033> zu finden.

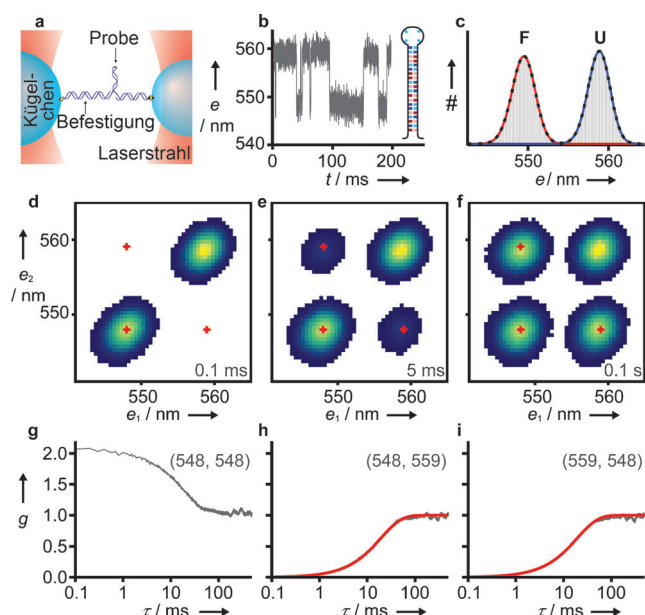


Abbildung 1. Signalpaar-Korrelationsanalyse einer kraftspektroskopischen Messung der Faltung einer DNS-Haarnadel. a) Messaufbau: Die Haarnadel wird zwischen zwei optischen Fallen unter Spannung gehalten, und die Streckung wird zeitaufgelöst gemessen. b) Streckungstrajektorie mit zwei Zuständen. Einschub: Schema der Haarnadel mit farbkodierter Sequenz (A/T: dunkel-/hellblau, G/C: dunkel-/hellrot). c) Das Streckungs-Histogramm zeigt zwei Peaks für den gefalteten (F) und ungefalteten Zustand (U). d)–f) Signalpaar-Histogramme bei verschiedenen Verzögerungszeiten. Farbskala: dunkelblau nach gelb (2% nach $\geq 50\%$ der häufigsten Klasse). Rote Kreuze markieren die Signalleistungen (e_1, e_2) der Korrelationen in (g)–(i). g) Autokorrelation von F. h), i) Kreuzkorrelationen von F nach U (h) und U nach F (i) mit globaler Parameteranpassung für ein Zweizustandsmodell an die Daten.

Abbildung 1c) können der gefaltete Zustand (F) um 548 nm und der ungefaltete Zustand (U) um 559 nm deutlich unterschieden werden. Weitere Informationen enthalten 2D-Signalpaar-Histogramme, die anzeigen, wie oft eine Streckung e_1 zum Zeitpunkt t zu einer Streckung e_2 zum Zeitpunkt $t+\tau$ führt; Abbildung 1d–f zeigt drei Verzögerungen τ . Die Anzahl N pro Klasse ist gegeben durch Gleichung (1).

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

Δe entspricht der Klassenbreite (hier: 0.25 nm für beide Dimensionen). Diese Signalpaar-Histogramme unterscheiden sich wesentlich von den ähnlich erscheinenden und häufig verwendeten „transition maps“ (Übergangsabbildungen),^[24] da hier keine Zustände oder Übergänge zugeordnet werden. Bei $\tau = 0.1$ ms erscheinen die beiden Zustände als zwei Peaks um die Linie $e_1 = e_2$; die Peakbreite wird durch dynamische Prozesse verursacht, die schneller sind als τ , aber die molekulare Streckung nicht stark verändern: Im vorliegenden Fall ist es die Diffusion der Kügelchen, die an der DNS befestigt sind.^[23] Mit steigendem τ wachsen zwei weitere Peaks an (Abbildung 1e und f), die einen langsameren Prozess anzeigen: die Faltung und Entfaltung der Haarnadel.

Für eine quantitative Beschreibung berechnen wir Signalpaar-Korrelationen, $g(\tau, e_1, e_2)$, basierend auf Gleichung (1), mit τ als Variable statt e_1 und e_2 in Gleichung (2).

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

$p(\{e_1, t\} \mid \{e_2, t+\tau\})$ ist die bedingte Wahrscheinlichkeit, e_1 zur Zeit t und e_2 zur Zeit $t+\tau$ zu messen; $p(e_i)$ ist die Wahrscheinlichkeit, e_i zu messen; $N(e_i)$ ist die Anzahl der Zeitfenster, in denen e_i gemessen wurde, und T ist die Gesamtzahl der Zeitfenster. Korrelationsfunktionen wurden für alle Signalpaare mit ausreichendem N berechnet (mit 1 nm Klassenbreite); drei Beispiele sind in Abbildung 1g–i gezeigt. Erwartungsgemäß sind die Autokorrelationen ($e_1 = e_2$: Abbildung 1g) positiv korreliert und die Kreuzkorrelationen negativ ($e_1 \neq e_2$: Abbildung 1h,i). Alle Korrelationen klingen auf derselben Zeitskala ab (bei ca. 10 ms).

Zur Bestimmung der Faltungsgeschwindigkeiten erfolgte eine Parameteranpassung von Funktionen, die mit Matrixmethoden^[25] für das Zweizustandsmodell hergeleitet wurden, an die Korrelationen. Die Anzahl der Parameter wurde reduziert durch Berücksichtigung der Anteile der jeweiligen Zustände im 1D-Histogramm (Abbildung 1c) unter Annahme eines „detaillierten Gleichgewichts“: $k_{ij} = k_{ji}(f_j/f_i)$, mit der Übergangsgeschwindigkeitskonstante k_{ij} zwischen den Zuständen i und j und dem Anteil f_i für Zustand i . Parameteranpassungen an 25 Signalpaar-Kombinationen nahe derjenigen in Abbildung 1h–i ergaben die Geschwindigkeitskonstanten $k_{FU} = k_{UF} = (25 \pm 2) \text{ s}^{-1}$ mit den Standardfehlern des Mittelwertes. Diese Geschwindigkeitskonstanten stimmen mit denen der Schwellwertmethode überein ($k_{FU} = (27 \pm 2) \text{ s}^{-1}$ und $k_{UF} = (25 \pm 2) \text{ s}^{-1}$).

Die nächste Anwendung erfolgte an einer Haarnadel, deren Sequenz für ein teilweise gefaltetes Intermediat (I) entworfen wurde, das zwingend während der Faltung und Entfaltung populiert wird^[7] (Abbildung 2a, Einschub). Das Intermediat kann sowohl in der Trajektorie (Abbildung 2a) als auch im 1D-Histogramm (Abbildung 2b) um 521 nm beobachtet werden, während F um 508 nm und U um 528 nm zu sehen sind. Da sich ca. 30% von I mit U überschneidet, ist dieser Zustand sehr schwer mit der Schwellwertmethode zu identifizieren (Abbildungen S1–S3). Ähnlich der Zweizustands-Haarnadel zeigen die Signalpaar-Histogramme bei 0.1 ms für jeden Zustand einen Peak um die Linie $e_1 = e_2$ (Abbildung 2c). Bei $\tau = 1$ ms erscheinen zunächst Kreuzpeaks sowohl zwischen U und I als auch zwischen F und I (Abbildung 2d) und bei $\tau = 1$ ms schließlich auch zwischen U und F (Abbildung 2e). Die drei Kreuzkorrelationen in Abbildung 2f–h entsprechen den mit roten Kreuzen markierten Signalleistungen (mit Klassenbreite 1 nm) in Abbildung 2c–e und zeigen dynamische Prozesse um 1 ms und um 10 ms an.

Zur Bestimmung der Geschwindigkeiten und zur Bestätigung des kinetischen Modells wurden zunächst drei Signalleistungen den Zuständen F, I und U zugeordnet, dabei wurden Bereiche mit starker Überschneidung und geringem N vermieden. Anschließend erfolgte eine globale Parameteranpassung von kinetischen Modellfunktionen an jeweils sechs Kreuzkorrelationen (wie in den Abbildungen 2f–h und S5).

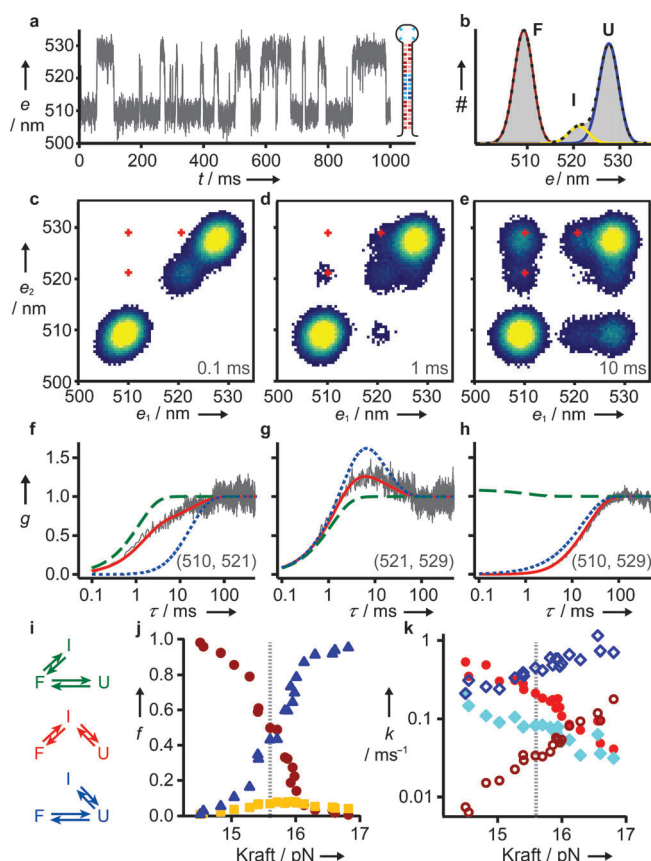


Abbildung 2. Analyse einer Dreizustands-DNS-Haarnadel. a) Die Trajektorie zeigt drei Zustände. Einschub: Schema der Haarnadelstruktur mit Farbkodierung wie in Abbildung 1. b) Das Streckungs-Histogramm zeigt die Überschneidung des Intermediates (I) mit U. c)–e) Signalpaar-Histogramme bei verschiedenen Verzögerungszeiten. f)–h) Globale Parameteranpassung dreier Modelle an die Kreuzkorrelationen von Signalpaaren (e_1, e_2). i) Verwendete kinetische Modelle. Die Anpassung deutet auf das Modell in der Mitte. j)–k) Kraftabhängige Zustandsanteile (j) (F: rote Kreise, I: orangefarbene Quadrate, U: blaue Dreiecke) und mikroskopische Geschwindigkeiten (k) ($k_{F,I}$: dunkelrote offene Kreise, $k_{I,F}$: hellrote gefüllte Kreise, $k_{I,U}$: dunkelblaue offene Rauten, $k_{U,I}$: hellblaue gefüllte Rauten). Die gestrichelte graue Linie entspricht der Kraft für (a)–(h).

Getestet wurden die Modellfunktionen entsprechend den Schemata in Abbildung 2i. Wie zuvor wurden die Zustandsanteile des 1D-Histogramms und „detailliertes Gleichgewicht“ berücksichtigt, um die anzupassenden Parameter auf $k_{I,F}$ und $k_{U,I}$ zu reduzieren (siehe die Hintergrundinformationen). Nur die Parameteranpassung des Modells mit zwingendem Intermediat zwischen U und F (rot in Abbildung 2 f–i) ergab eine befriedigende Anpassung. Die Wiederholung mit 125 Signalbereichskombinationen nahe derer in Abbildung 2c ergab $k_{I,F} = (0.21 \pm 0.01) \text{ ms}^{-1}$, $k_{U,I} = (79 \pm 4) \text{ s}^{-1}$, $k_{F,I} = (34 \pm 1) \text{ s}^{-1}$, und $k_{I,U} = (0.43 \pm 0.03) \text{ ms}^{-1}$. Weder das kinetische Modell noch die Geschwindigkeiten konnten mit der Standard-Korrelationsanalyse ermittelt werden (Abbildung S4). Die Analyse für Trajektorien bei verschiedenen Kräften (Abbildung S5) – bei denen die Zustandsanteile von fast ausschließlich gefaltet bis zu fast ausschließlich entfaltet variieren (Abbildung 2j) – ergab die Kraftabhängigkeit der

mikroskopischen Geschwindigkeiten (Abbildung 2k). Diese Information kann genutzt werden, um z.B. für jeden Übergang die Höhe und die Position der Faltungsenergiebarriere zu bestimmen.^[26]

Als letztes Beispiel wurde die Methode auf eine Faltungstrajektorie des proteaseresistenten Teils der Mutante C179A/C214A von HaPrP (Abbildung 3 a, Einschub) ange-

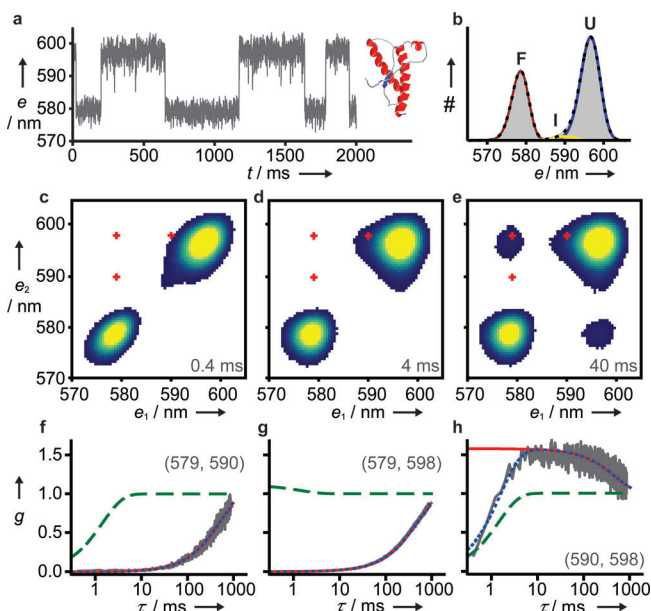


Abbildung 3. Analyse von HaPrP. a) Die Trajektorie zeigt zwei Hauptzustände. Einschub: HaPrP-Struktur (PDB-Code: 1B10). b) Streckungs-Histogramm mit schwach populierte I (gelb). c)–e) Signalpaar-Histogramme verschiedener Verzögerungszeiten. Rote Kreuze markieren Signalpaare der folgenden Korrelation. f)–h) Kreuzkorrelationen bei indizierten Signalpaaren (e_1, e_2) und globale Parameteranpassung mit den Modellen aus Abbildung 2i zeigen, dass I ausschließlich von U aus populiert wird.

wendet. Die Trajektorie in Abbildung 3 a und das 1D-Histogramm in Abbildung 3 b zeigen F bei 579 nm, U bei 598 nm und I bei 591 nm. Der geringe Anteil und die kurze Lebensdauer von I sowie die Überschneidung von ca. 75 % von I mit U erschweren die Verwendung von Standardmethoden enorm. Das Signalpaar-Histogramm bei 0.4 ms zeigt zwei ausgeprägte Peaks für U und F und einen kleinen für I (Abbildung 3 c). Aufgrund seiner kurzen Lebensdauer verschwindet der Peak von I bei 4 ms, und nur Kreuzpeaks von I bleiben zu sehen (Abbildung 3 d). Bei 40 ms erscheinen auch Kreuzpeaks zwischen U und F (Abbildung 3 e).

Die Kreuzkorrelationen in den Abbildung 3 f–h zeigen, dass F sowohl zu I als auch zu U innerhalb von ca. 100 ms relaxiert. Die Korrelation zwischen U und I hat eine zusätzliche Zeitkomponente bei ca. 1 ms. Die Parameteranpassung für die drei Modellfunktionen (entsprechend den Schemata in Abbildung 2i) an sechs Kreuzkorrelationen ergab das Modell, in dem I ausschließlich von U erreicht wird. Die Anpassungen an 125 verschiedene Kombinationen von Signalbereichen nahe der in Abbildung 3 c gezeigten ergaben $k_{F,U} = (1.43 \pm 0.03) \text{ s}^{-1}$, $k_{U,F} = (0.83 \pm 0.02) \text{ s}^{-1}$, $k_{I,U} = (0.6 \pm 0.1) \text{ ms}^{-1}$ und

$k_{U,I} = (26 \pm 4) \text{ s}^{-1}$. Dieses Beispiel verdeutlicht die Fähigkeit der vorgestellten Methode, Geschwindigkeiten und kinetische Modelle auch bei Zuständen zu bestimmen, die schwer zu unterscheiden und sehr kurzlebig sind.

Die Signalpaar-Korrelationsanalyse ist auch für andere Systeme, Messmethoden und Simulationen anwendbar, weit über die hier verwendete Kraftspektroskopie von Faltungsprozessen hinaus. Desweiteren können verschiedene Informationen, z. B. aus einer Multiparameter-Fluoreszenzanalyse, kombiniert werden, um die Signalpaare besser zu definieren und die Zustandszuordnung weiter zu verbessern. Ein zentraler Vorteil der Methode ist es, dass die Zustandszuordnung sowohl intuitiv (und erleichtert mithilfe von 2D-Histogrammen) als auch am Ende der Analyse durchgeführt wird und somit systematische Fehler der Übertragung auf Zustands-trajektorien vermieden werden. Dies macht sie zu einem nützlichen Hilfsmittel um andere, aufwendigere Methoden auch unter schwierigen Bedingungen zu verifizieren (z. B. HMM). Die Signalpaar-Korrelationsanalyse ist damit besonders geeignet, die Reichweite von Einzelmolekülmethoden zu erweitern.

Eingegangen am 12. Juni 2011,

veränderte Fassung am 20. September 2011

Online veröffentlicht am 4. November 2011

Stichwörter: Biophysik · Einzelmolekülstudien · Kinetik

- [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, **2011**, eingereicht.
- [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.