Single-molecule high-resolution imaging with photobleaching

Matthew P. Gordon*, Taekjip Ha*†, and Paul R. Selvin*†‡

*Department of Physics and †Center for Biophysics and Computational Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801

Communicated by Gordon A. Baym, University of Illinois at Urbana–Champaign, Urbana, IL, March 8, 2004 (received for review January 29, 2004)

Conventional light microscopy is limited in its resolving power by the Rayleigh limit to length scales on the order of 200 nm. On the other hand, spectroscopic techniques such as fluorescence resonance energy transfer cannot be used to measure distances >10 nm, leaving a "gap" in the ability of optical techniques to measure distances on the 10- to 100-nm scale. We have previously demonstrated the ability to localize single dye molecules to a precision of 1.5 nm with subsecond time resolution. Here we locate the position of two dyes and determine their separation with 5-nm precision, using the quantal photobleaching behavior of single fluorescent dye molecules. By fitting images both before and after photobleaching of one of the dyes, we may localize both dyes simultaneously and compute their separation. Hence, we have circumvented the Rayleigh limit and achieved nanometer-scale resolution. Specifically, we demonstrate the technique by measuring the distance between single fluorophores separated by 10–20 nm via attachment to the ends of double-stranded DNA molecules immobilized on a surface. In addition to bridging the gap in optical resolution, this technique may be useful for biophysical or genomic applications, including the generation of super-high-density maps of single-nucleotide polymorphisms.

The advent of single-molecule imaging has enabled a revolution in the measurement of the physical parameters underlying biological processes (1). However, there are limitations on the length scales over which single-molecule imaging can be used to measure distances. Conventional far-field microscopy techniques are limited by the Rayleigh criterion (2) to resolving distances greater than ~200 nm. Some recent techniques have been developed to circumvent this limitation (3, 4), but they are technically demanding and at this point of limited applicability to biological systems. Fluorescence resonance energy transfer (FRET) can be used to measure distances much smaller, on the order of a few nanometers. However, because of the strong distance dependence of energy transfer, FRET is limited to measuring distances less than ~10 nm (5). As a result, there is a "gap" in the resolution attainable by single-color optical spectroscopy, making it difficult to measure separations between 10 and 200 nm. Many biological objects of interest are on this scale, including DNA structures, macromolecular complexes, and motor proteins. Although it is possible to measure distances on these scales by using two or more dyes of different colors (6, 7), this technique presents its own problems, such as how to achieve heterogeneous labeling, and calibration of the distance registration between different color wide-field images.

Recently, with the introduction of low-noise high-quantum-yield charge-coupled device (CCD) cameras, it has become practical to localize individual fluorescent dyes at subwavelength scales (8–10). We recently demonstrated the ability to determine the positions of single molecules at room temperature with a precision of 1.5 nm (11) by using total internal reflection (TIR) microscopy and a cooled CCD. This technique has been used to probe the motility of Myosin V and kinesin (11, 12). Here we present a technique, single-molecule high-resolution imaging with photobleaching (SHRImP), that takes advantage of the quantal photobleaching of fluorescent molecules to *resolve* two identical fluorophores in the *x–y* plane separated by distances as small as 10 nm, with a precision for single measurements on the order of 5 nm. In SHRImP, two molecules, m_1 and m_2, with overlapping point-spread functions (PSFs) are imaged continuously, creating a series of images that forms a "movie." A plot of the integrated intensity of the combined PSF shows a two-step signal reduction (Fig. 1*a*), each step corresponding to the photobleaching of an individual molecule. After the first photobleaching event, only m_2 is left, and its PSF can be fit to a Gaussian distribution, allowing us to localize m_2 to within 1.5 nm. Now, by selecting one image from before the photobleaching event (I_{pre}) and one from after (I_{post}), we may compute the distribution of m_1 alone, $I_{pre'}$, by taking $I_{pre} - I_{post}$ (Fig. 1*b*). We may then use this distribution to localize m_1, and hence compute the separation between m_1 and m_2. Although this is a conceptually simple procedure, a more technically involved procedure involves localizing the PSFs simultaneously, by performing a global analysis involving both I_{pre} and I_{post}, fitting I_{post} to a single Gaussian, and fitting I_{pre} to two Gaussians, as discussed below. We used the latter method for the data presented but found that, in practice, both procedures produced nearly identical results.

Materials and Methods

Imaging. Imaging was performed as in ref. 11. Briefly, samples were mounted on an Olympus (Melville, NY) IX-70 microscope and 60× objective (Olympus PLApo 60×/1.45 oil) and excited by using objective-type TIR. Excitation was provided by a 532-nm diode-pumped crystal laser (CrystaLaser, Reno, NV). Images were captured by using a slow-scan back-thinned MicroMax CCD camera (Roper Scientific, Trenton, NJ), allowing continuous imaging with no interframe dead time. The image capture frequency was 2 Hz.

Sample Preparation. Biotinylated Cy3-DNA samples were prepared as in ref. 11. Briefly, coverslips were cleaned by sonication in acetone, then in KOH, and rinsed between baths. Coverslips were incorporated into flow chambers, and Cy3-DNA was attached via a biotin-streptavidin linkage. The singly labeled DNA sequence was the same as reported in ref. 11. Flow chambers for doubly labeled DNA samples (Integrated DNA Technologies, Coralville, IA) were prepared similarly, but coverslips were silanized by using a procedure derived from ref. 14. Coverslips were cleaned as above and then incubated in 6 mM 3-aminopropyltriethoxysilane (APTES, Sigma-Aldrich), pH ≈3, at 60°C for 17.5 h. They were then rinsed with water and stored in distilled ethanol. Doubly labeled DNA samples were diluted to ~100 pM in T50 buffer (10 mM Tris/1 mM EDTA/50 mM NaCl, pH 7) and flowed directly over the silanized surface. They were allowed to sit for 5 min then flushed with T50 buffer, followed by a flush with deoxygenating imaging buffer. The imaging buffer consisted of 1% by volume gloxy, 1% by volume 2-mercaptoethanol, 0.4% by mass glucose, and 15 mM $MgCl_2$ in

Abbreviations: CCD, charge-coupled device; PSF, point-spread functions; SHRImP, single-molecule high-resolution imaging with photobleaching.

‡To whom correspondence should be addressed. E-mail: selvin@uiuc.edu.

www.pnas.org/cgi/doi/10.1073/pnas.0401638101

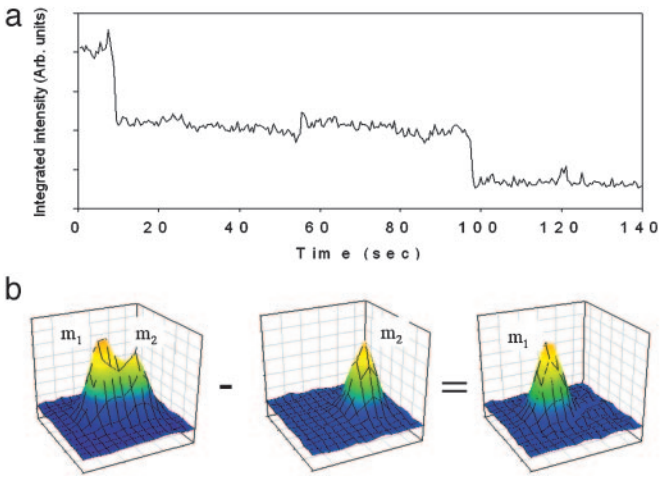

Fig. 1. (*a*) A plot of total integrated intensity versus time for two closely spaced Cy3 molecules, showing a two-step photobleaching behavior. (*b*) I_{pre}, as calculated from I_{pre} and I_{post}.

T50. Gloxy consisted of 30 mg of glucose oxidase (G-7016, Sigma-Aldrich) and 60 μl of catalase solution (106810, Roche Diagnostics) in 200 μl of T50 buffer, spun down and filtered twice with syringe filters. Doubly labeled DNA consisted of two hybridized complementary DNA strands, each covalently labeled on the 5′ end with a Cy3 phosphoramidite dye (Amersham Biosciences). The doubly labeled 51-mer DNA sequence was 5′-TGG ACT CAG AGC ATT CAA GAT GGT CGG TGG ACG GTT GAC CTA CGG CTG CCA, hybridized with its complement. The 40- and 30-mer sequences were the same but truncated from the 3′ end and modified so that the last three bases were CCA.

Data Analysis. Biotinylated Cy3-DNA images were analyzed by performing a global fit with SIGMAPLOT 8.0 (SPSS, Chicago), using single pre- and postbleach images selected by hand. Pre- and postbleach images I_{pre} and I_{post} were analyzed by globally fitting both distributions simultaneously to two-dimensional Gaussians. The underlying distribution is expected to be an Airy disk, but it is common to model it with a Gaussian, because it is more mathematically tractable, and because the differences are expected to be minor in practice (10, 11). For fitting, we used the following functions:

$$I_{pre} = A \exp\left(-\left[(x-x_0)^2/w_{x_1}^2 + (y-y_0)^2/w_{y_1}^2\right]\right) + B \exp\left(-\left[(x-x_0+\delta_x)^2/w_{x_2}^2 + (y-y_0+\delta_y)^2/w_{y_2}^2\right]\right) + z_0 \quad \textbf{[1a]}$$

and

$$I_{post} = A \exp\left(-\left[(x-x_0)^2/w_{x_1}^2 + (y-y_0)^2/w_{y_1}^2\right]\right) + z_1, \quad \textbf{[1b]}$$

where A and B are the heights of the two Gaussians, x_0 and y_0 are the coordinates of the centroid of the first Gaussian, δ_x and δ_y are the x and y components of the vector separating the centroid of the second Gaussian from the first, w_{x_1}, w_{y_1}, w_{x_2}, and w_{y_2} are the widths of the x and y components of the two Gaussians, and z_0 and z_1 are the z-axis offsets of the distributions, arising from the sum of the background noise, dark noise, and readout noise from the CCD, as well as any background fluorescence. The separation of the two molecules was computed from $d = \sqrt{\delta_x^2 + \delta_y^2}$.

To improve throughput, doubly labeled DNA images were analyzed by using custom-written software in IDL (Research Systems, Boulder, CO). Each image frame consisted of several DNA molecules. The software picked individual punctuate objects, and from these, molecules showing a clean, two-step intensity profile similar to Fig. 1*a* and good fit to a Gaussian were chosen for further analysis. Fluorescence intensity stability was variable among molecules, so it was not possible to determine where photobleaching events occurred for some observed molecules. Some spots also showed many more than two photobleaching events, indicating some "clumping" of the sample. However, overall, 30–40% of the observed spots showed clear two-step photobleaching and could be analyzed further.

Each molecule was analyzed in the following way. For values of N ranging from 1 to 4, $2N$ images from before the first bleach step and $2N$ images from after this step were chosen. N images from each set were summed, resulting in four composite images, and each of the two postbleach images were paired with each of the two prebleach images to generate a set of four different pairs of images. We refer to this set as a "quartet." Four separations (d) were computed from the quartet by performing a global fit of both the measured prebleach distribution (I_{pre}) and the measured postbleach distribution (I_{post}) simultaneously. The pre- and postbleach fit functions were similar to those in Eq. **1**, but each point in the distribution was integrated in a square region over the pixel size of the camera, reflecting the fact that each measured datum corresponds to all of the photons falling on a particular pixel of finite size.

The data quality was evaluated according to the following three criteria. The standard deviation of d within each quartet must be <10 nm, the ellipticity of each PSF must be $<20\%$, and each composite image must have enough photons to theoretically allow localization with at least 1.7-nm precision. The ellipticity of the PSF is $\varepsilon = 1 - w_{minor}/w_{major}$, where w_{minor} and w_{major} are the major and minor axes that form the waist of the Gaussian distribution. From the quartets that passed these tests for each value of N, the one with the smallest standard deviation was chosen, and the mean of the quartet was used as a data point for computing the final histogram.

Of the above cited criteria, the ellipticity requirement is intended to filter out dyes with a high degree of immobility. Our observations of Cy3 dried on coverslips shows that highly immobile dyes have PSFs with much higher ellipticity. It has been demonstrated that very small deviations in the image focus can lead to large systematic errors in the localization of dyes with highly immobile emission dipoles (15). Dyes that change angle rapidly on the time scale of imaging (16), however, are not subject to this systematic error. In practice, however, this criterion was almost never violated when observing dyes in solution.

Dye intensity is variable because of variations in illumination, blinking, and local environment. Allowing up to four sequential images to be summed increases the likelihood of being able to localize the molecule to within the required accuracy. However, the more images that are binned together, the worse the time resolution becomes. Also, the further in time that we are from the photobleaching point, the greater the likelihood is that the dye intensity has changed. But, the SHRImP analysis depends on the assumption that the intensity remains constant over time, because we assume the postphotobleach image of m_2 is a good approximation of its contribution to the prephotobleach image. Hence, using larger bins can decrease the accuracy if the intensity fluctuations are large and occur on the time scale of the camera integration time. Taking two sets of images from both before and after the photobleaching event and using them to compute four different distances allows us to use the standard deviation of the quartet distances as a measure of the degree to which intensity fluctuations impact the measurement. Quartets showing a large standard deviation may be due to intensity fluctuations which are too large to yield good measurements.

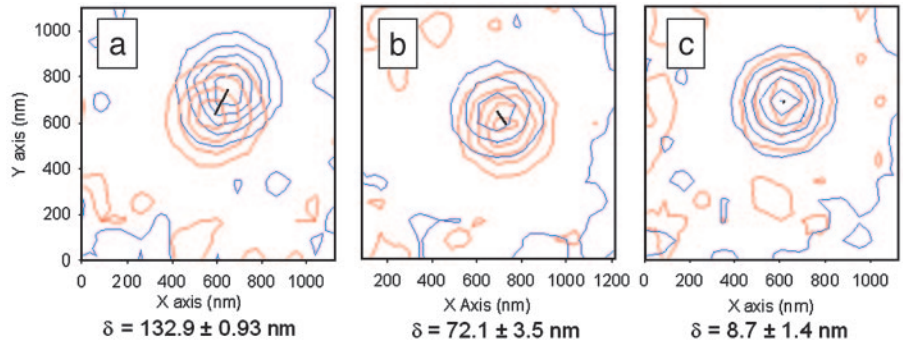

Fig. 2. Three examples of resolved, overlapping molecules with different separations. The black lines indicate the computed center-to-center separations. The errors are computed from the standard error of the mean (σ_μ) of the fit. (*a*) Separation = 132.9 ± 0.93 nm. (*b*) 72.1 ± 3.5 nm. (*c*) 8.7 ± 1.4 nm.

After all molecules were analyzed, the computed separations were used to generate a histogram. The histogram was fit to a Gaussian plus a constant offset, of the form

$$y = y_0 + a \exp\left[-0.5\left(\frac{x - x_0}{w}\right)^2\right]. \qquad [2]$$

The centroid of this Gaussian is the measured value of the separation, with a margin of error given by the standard error of the mean as determined by the fit. The constant offset allows for "background" due to closely spaced but disconnected molecules.

Results and Discussion

In the first set of experiments, we attached DNA molecules singly labeled with Cy3 to a coverslip by using a biotin-streptavidin linker. We picked fluorescent spots with multistep bleaching behavior arising from chance proximity and then analyzed the last two bleaching steps. Fig. 1*a* shows a typical example of such an intensity profile, with the two dyes in Fig. 1*b* separated far enough (330 nm) that they can be resolved without using SHRImP. Results for more closely spaced examples are shown in Fig. 2. Fig. 2 shows contour plots of individually resolved dyes spaced much closer than the Rayleigh criterion minimum (133, 72, and 9 nm). The black lines indicate the separation between the centroids of the distributions. We were able to localize the dyes very precisely (better than 1.5 nm), and the computed separations hence have standard errors of the mean (σ_μ) of 1–3.5 nm. σ_μ was determined by SIGMAPLOT based on the computed fit to the fit functions.

To test whether the separation measured in SHRImP represents a true distance, we performed a second set of experiments, using double-stranded DNA molecules labeled at both of the 5′ ends with Cy3. The DNA oligos were 51, 40, and 30 bp, much less than the persistence length of DNA [~150 bp (17)], and hence are expected to be relatively rigid and straight.

We found that, in practice, a large percentage of the observed molecules (~70%) did not show good two-step photobleaching behavior. In some cases, this was because they showed only a single photobleaching event or showed more than two photobleaching events, or because the photobleaching transitions were obscured by intensity fluctuations. However, because wide-field illumination and CCD imaging allows highly paralleled data collection, we were easily able to generate enough data for analysis.

Of the molecules showing good two-step behavior, more than two thirds were found to pass all three of the filtering criteria. Almost all of the data passed the ellipticity criterion, and the quartet standard deviation was ≈3 nm for those that passed the criteria.

Fig. 3 shows histograms of the distance measurements for the

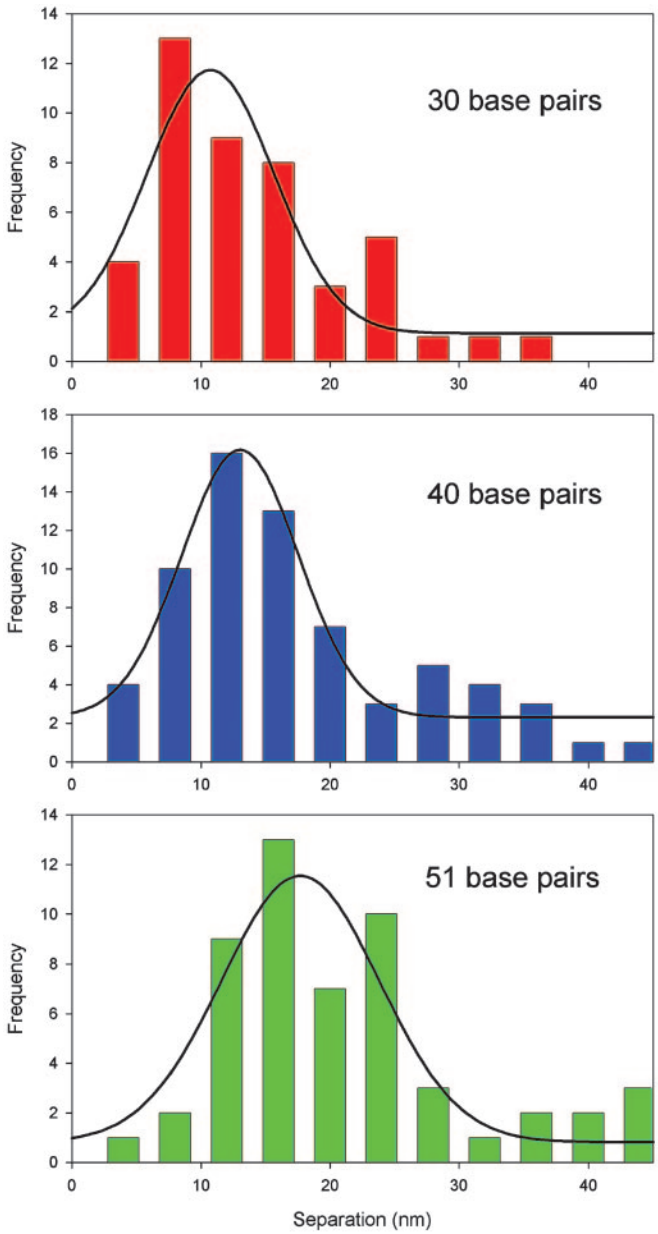

Fig. 3. Histograms of measured end-to-end separations of 30-, 40-, and 51-bp DNA oligos, with fits (black lines). The estimated separations are 17.7 ± 0.7 nm, 13.0 ± 0.5 nm, and 10.7 ± 1.0 nm.

three DNA lengths. A single peak is seen in each histogram, and the peak distance increases for increasing DNA length. The histograms were each fit to a one-dimensional Gaussian with an offset (Eq. **2**). The offset allows for a "background" due to closely spaced but disconnected molecules, because of the fact that there is no *a priori* way to differentiate between two covalently linked dye molecules and two that simply lie close to each other. The density of observed molecules was $\approx 0.3\ \mu m^{-2}$, low enough that we would expect a small incidence of such observations. That such measurements were an observable fraction may be due in part to the clumping discussed in *Materials and Methods*. It also may be due to individual molecules that were deformed because of interaction with the surface, or to data that were not fit well. The Gaussian fits to the histograms shown in Fig. 3 have widths ranging from 4 to 6 nm, where the "width" refers to the parameter w in Eq. **2**. This parameter would be the standard deviation of the measurement in the absence of the "background." Our data filtering is intended to discard molecules that cannot be localized to at least 1.7 nm, and therefore, ideally, we would expect to see a width close to $\sqrt{2}\sigma$, or 2.4 nm. The larger observed width may reflect the impact of two factors: first, the experimentally achieved accuracy is often worse than 1.7 nm because of aberrations in the optical system, and is sometimes as high as 2.5 nm. Second, there is a contribution to the width due to the uncertainty in the intensity of the m_1 dye because of shot noise and "blinking" (18). As noted above, the average quartet standard deviation was ≈ 3 nm. This standard deviation is probably a reasonable measure of the contribution of dye intensity fluctuation to the overall histogram width, and these numbers taken together may account for the magnitude of the observed width.

In Fig. 4, the centroids of the Gaussian fits are plotted against the length of the DNA in base pairs, with error bars showing the standard error of the mean based on the fit. The solid line is a fit to the function $y = ax$. The fit predicts a separation of 3.4 Å per base pair, which agrees exactly with the expected value based on the structure of DNA in solution (19). Thus, it appears likely that the DNA is immobilized with its long axis lying on the surface and its overall structure is not greatly compromised. We suggest that positively charged amino groups on the silanized surface may trap the DNA in a conformation close to its native B form.

Our results demonstrate the feasibility of using SHRImP to measure point-to-point separations as small as 10 nm, and possibly smaller, thus improving on the Rayleigh limit by a factor of 20. We note that SHRImP is not a true imaging technique because it can only be used to measure distances with well separated pairs of molecules. In addition, SHRImP can measure only the two-dimensional projection of distances. However, SHRImP has distinct advantages over two-color localization, because it requires only a single type of dye, can be done with existing optics and hardware designed for single-color localization, and does not present the usual problems of registration between separate two-color images. There are also many potential extensions to this technique, including localizing of three or more points and measuring slowly changing distances by making allowances for small, continuous changes of the fit parameters as one moves further away from the bleach point in time. In addition, the ability to selectively label individual DNA molecules at single-nucleotide polymorphisms (SNPs) (13) presents the possibility of building ultra-high-resolution SNP maps.

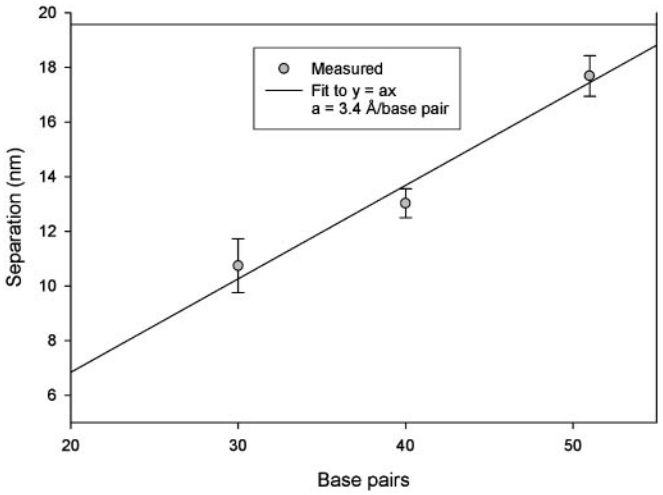

Fig. 4. Centroid values of the histograms in Fig. 3, with the best fit for the line $y = ax$ drawn through them. The error bars are derived from the standard deviation of the mean of the histograms in Fig. 3. The parameter a is estimated by the fit to be 3.4 Å per base pair.

We thank Dr. Gordon Baym for his valuable input on the writing of this paper and Ahmet Yildiz for his instruction regarding the sample preparation and imaging. This work was supported by National Institutes of Health Grants AR44420 and GM65367 and National Science Foundation Grants DBI-02-15869, 9984841, and 0134916. M.P.G. was partially supported by a National Research Service Award in Molecular Biophysics through National Institutes of Health Training Grant PHS 5 T32 GM08276.

1. Weiss, S. (1999) *Science* **283,** 1676–1683.
2. Strutt, J. W. (1874) *Philos. Mag.* **47,** 81–93, 193–205.
3. Dyba, M. & Hell, S. W. (2002) *Phys. Rev. Lett.* **88,** 163901–163904.
4. van Oijen, A. M., Köhler, J., Schmidt, J., Müller, M. & Brakenhoff, G. J. (1999) *J. Opt. Soc. Am. A* **16,** 909–915.
5. Selvin, P. R. (2000) *Nat. Struct. Biol.* **7,** 730–734.
6. Lacoste, T. D., Michalet, X., Pinaud, F., Chemla, D. S., Alivisatos, A. P. & Weiss, S. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 9461–9466.
7. Betzig, E. (1994) *Opt. Lett.* **20,** 237–239.
8. Schmidt, T., Schutz, G. J., Baumgartner, W., Gruber, H. J. & Schindler, H. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 2926–2929.
9. Kubitscheck, U., Kuckmann, O., Kues, T. & Peters, R. (2000) *Biophys. J.* **78,** 2170–2179.
10. Thompson, R. E., Larson, D. R. & Webb, W. W. (2002) *Biophys. J.* **82,** 2775–2783.
11. Yildiz, A., Forkey, J. N., McKinney, S. A., Ha, T., Goldman, Y. E. & Selvin, P. R. (2003) *Science* **300,** 2061–2065.
12. Yildiz, A., Tomishige, M., Vale, R. D. & Selvin, P. R. (2004) *Science* **303,** 676–678.
13. Banér, J., Nilsson, M., Isaksson, A., Mendel-Hartvig, M., Antson, D. O. & Landegren, U. (2001) *Curr. Opin. Biotechnol.* **12,** 11–15.
14. Jing, J., Reed, J., Huang, J., Hu, X., Clarke, V., Edington, J., Housman, D., Anantharaman, T. S., Huff, E. J., Mishra, B., *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95,** 8046–8051.
15. Bartko, A. P. & Dickson, R. M. (1999) *J. Phys. Chem. B* **103,** 11237–11241.
16. Ha, T., Glass, J., Enderle, T., Chemla, D. S. & Weiss, S. (1998) *Phys. Rev. Lett.* **80,** 2093–2096.
17. Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. (1994) *Science* **265,** 1599–1600.
18. Moerner, W. E. (1997) *Science* **277,** 1059–1060.
19. Stryer, L. (1995) *Biochemistry* (Freeman, San Francisco).