# Fluorescent probes and bioconjugation chemistries for single-molecule fluorescence analysis of biomolecules

Achillefs N. Kapanidis[a] and Shimon Weiss
*Department of Chemistry and Biochemistry, and Department of Physiology, University of California at Los Angeles (UCLA), Los Angeles, California 90095-1569*

(Received 7 March 2002; accepted 18 September 2002)

Fluorescence-based detection of single biomolecules in solution and at room temperature has opened new avenues for understanding biological mechanisms. Single-molecule fluorescence spectroscopy (SMFS) of biomolecules requires careful selection of fluorophores, sites of incorporation, and labeling chemistries. SMFS-compatible fluorophores should permit extended, uninterrupted observations of fluorescence with high signal-to-noise ratios; more stringent considerations apply for specific methodologies, such as fluorescence resonance energy transfer and fluorescence anisotropy. Strategies for site-specific *in vitro* labeling of small proteins exploit the reactivity of the amino acid cysteine (Cys), allowing incorporation of one or more fluorophores; labeling of closely spaced Cys residues using bis-functionalized fluorophores allows probing of the orientation of individual protein domains. For *in vitro* labeling of large proteins, the options include peptide ligation, intein-mediated labeling, puromycin-based labeling, unnatural amino acid mutagenesis, and reconstitution from individual subunits or subunit fragments. For *in vivo* analysis, one can use proteins that are labeled *in vitro* and then incorporated in cells; genetic encoding of specific protein sequences can also lead to *in vivo* labeling, either by *in vivo* targeting by fluorophores or by biosynthesis of protein fusions with natural fluorophores such as the green fluorescent protein. The existing methods, along with others under development, will bring SMFS to the mainstream and advance significantly our understanding of vital biological processes.  [DOI: 10.1063/1.1521158]

## I. INTRODUCTION

The past decade marked one of the most exciting scientific developments born of the interaction of physics, chemistry, and biology: the ability to perform real-time, room-temperature observations of individual biomolecules by using optical microscopy techniques described as single-molecule detection (SMD).[1] SMD eliminates the ensemble- and time-averaging present in conventional biochemical and biophysical assays. SMD can uncover and analyze subpopulations and distributions hidden in heterogeneous systems; probe conformational dynamics of biomolecules under equilibrium conditions, an otherwise impossible task due to lack of synchronization; and explore complex reaction kinetics and biochemical pathways under nonequilibrium conditions. Several excellent reviews exist on the topic.[2–8]

A popular method for probing single molecules is single-molecule fluorescence spectroscopy (SMFS). SMFS monitors the structure, dynamics, and interactions of biochemical systems (e.g., intramolecular and intermolecular distances, domain orientations, binding stoichiometry, equilibrium binding constants) through measurements of fluorescence properties of individual species, such as fluorescence emission intensity, lifetime, polarization, and quantum yield.[6,9] SMFS has allowed striking visualizations of single biomolecules at work, including the rotation of the rotor subunit of the rotary stepper motor protein F1-ATPase,[10] the sliding of kinesin on microtubules,[11,12] the translocation of RNA polymerase (RNAP) molecules on double-stranded DNA,[13] the formation of ribozyme intermediates,[14] and the real-time infection of a cell by single virus particles.[15]

In this article, we describe the properties of SMFS-compatible fluorophores, list considerations for popular assays, provide examples of site-specific labeling strategies that enabled SMFS observations, and conclude with a look at fluorophores and labeling strategies that can answer future challenges.

## II. PROPERTIES OF SMFS-COMPATIBLE FLUOROPHORES

Optical detection of single fluorescently labeled biomolecules using SMFS is based on thousands to millions of laser-induced excitation–deexcitation cycles of the fluorophore between the ground state and the first excited electronic state.[3,4,7] A list of fluorophores often used for SMFS, along with characteristic properties, is provided in Table I. Fluorophores compatible with the demanding nature of SMFS fulfill most of the photophysical and photochemical criteria set for ensemble fluorescence spectroscopy and microscopy measurements[16,17] (high extinction coefficient, high quantum yield, large Stokes shift, high photostability), as well as additional properties specific to SMFS.[18] Desirable fluorophore properties include the following.

[a]Author to whom all correspondence should be addressed. Electronic mail: kapanidi@chem.ucla.edu

 

TABLE I. Characteristic properties of selected fluorophores used for SMFS.[a]

| Fluorophore | Absorbance maximum (nm) | Extinction coefficient ($cm^{-1}\,M^{-1}$) | Emission maximum (nm) | Quantum yield | Laser line (nm) | Photostability |
|---|---|---|---|---|---|---|
| GFP | 395 | 30 000 | 509 | 0.8 | 457 | Good |
| | 470 | 7 000 | | | | |
| Fluorescein | 490 | 67 000 | 520 | 0.5–0.9 | 488 | Poor |
| Alexa-488 | 495 | 80 000 | 520 | 0.5–0.9 | 488 | Moderate |
| Tetramethyl-rhodamine | 554 | 85 000 | 585 | 0.2–0.5 | 514 532 | Good |
| Cy3 | 550 | 150 000 | 585 | 0.2 | 514 532 | Good |
| Texas Red | 596 | 85 000 | 620 | 0.5 | 514 532 | Good |
| Cy5 | 650 | 250 000 | 670 | 0.15 | 633 | Moderate |

[a]Values were taken from Refs. 16 and 17, manufacturers of reactive forms of the fluorophores (Molecular Probes, Amersham Pharmacia Biotech), or measured in our laboratory (see Ref. 80).

### A. Brightness

Brightness is a measure of the rate of fluorescence photons emitted by the fluorophore, and is represented by the product of the extinction coefficient of the fluorophore at the wavelength of excitation, $\epsilon_{\mathrm{exc}}$, and the fluorescence quantum yield of the fluorophore at the wavelength of emission, $Q_F$;[16] for typical SMFS-compatible fluorophores, $\epsilon_{\mathrm{exc}} > 20\,000\ \mathrm{cm}^{-1}\,\mathrm{M}^{-1}$ and $Q_F > 0.1$. Although less bright fluorophores can be detected using SMFS, the above-stated parameters set limits for the extraction of reliable biological information from SMFS experiments. These limits currently eliminate fluorophores that are useful for ensemble fluorescence or phosphorescence, such as the fluorescent amino acid tryptophan, the environmental probes dansyl and IAEDANS, and the phosphorescent probes eosin and erythrosin. Brightness can be affected significantly by local environment and buffer conditions ($p$H, viscosity, concentration of organic solvents); accurate evaluation of brightness should be performed with the fluorophore incorporated at the site of interest and in the appropriate assay buffer.

### B. Singlet-state saturation

During the excitation–deexcitation cycle of a single fluorophore, a transition from ground to excited state occurs due to matching of the excitation wavelength to a fluorophore transition, whereas a transition from excited to ground state occurs due to photon emission (fluorescence) or thermal deactivation. Since in SMFS the prerequisite for reexcitation of a *single* fluorophore is the relaxation to the ground state, the maximum number of emitted photons is limited by the rate of fluorescence, $k_F$.[3] Therefore, fluorophores with relatively short fluorescence lifetimes (0.5–5 ns) are preferable for SMFS.

### C. Triplet-state saturation

Intersystem crossing from the singlet excited state of the fluorophore to the triplet state (an excited state lasting up to several milliseconds that is characterized by a spin-forbidden transition to the ground state) can lead to saturation that is even more severe than singlet-state saturation. For some fluorophores, efficient intersystem crossing leads to a large decrease in the signal intensity, rendering them unsuitable for SMFS. The lifetime of a triplet state can be shortened dramatically (by a factor of $10^3$–$10^5$) through quenching by the ground triplet state of molecular oxygen ($^3O_2$);[19] however, since the oxygen concentration is often minimized to delay photodestruction of the fluorophore, a compromise has to be reached between removing oxygen and quenching the triplet state.[19] This dilemma can be addressed by using reagents that act as both oxygen scavengers and triplet-state quenchers for a specific fluorophore.

### D. Signal stability and fluorescence intermittency ("blinking")

Depending on their rotational freedom, fluorophores can interact transiently with their local environment to yield states associated with variable emission spectra ("spectral jumps") or quantum yield. Although such transitions are efficient reporters of the local environment, they can obscure more interesting, dynamic processes. Moreover, some of the states adopted by the fluorophore can be long (up to seconds) and exhibit low quantum yield ("dark" states), resulting in fluorescence intermittency that reduces the detectable signal and obscures time trajectories of fluorescence.

### E. Resistance to photodestruction

Every organic fluorophore is prone to irreversible photodestruction (due to elusive photochemical reactions, mainly photo-oxidation[20]), limiting the total number of emitted photons;[3] SMFS-compatible fluorophores emit $10^5$–$10^6$ photons before photodestruction.[4] The quantum yield for photodestruction (or "photobleaching") varies greatly among fluorophores,[21] with the series fluorescein < Cy5 < Cy3 < tetramethylrhodamine being a general rule for the photostability of common fluorophores. Depending on the time scales of the observation and the processes studied, resistance to photobleaching ranges from desirable to essential. For example, surface-immobilized fluorophores that report on long time trajectories of complex biochemical reactions should possess correspondingly long photosurvival times that allow probing of several time scales along the reaction coordinate. Photosurvival can be extended with

various methods that minimize triplet-state oxygen and its highly reactive metabolites (such as singlet oxygen $[^1O_2]$, hydroxyl-radical $[OH^{\bullet}]$ and hydrogen peroxide $[H_2O_2]$); this is performed through deoxygenation,[19] use of molecular-oxygen scavengers (such as *n*-propyl gallate, *p*-phenylenediamine, or the glucose oxidase/catalase system), singlet-oxygen scavengers (such as carotenoids), oxygen-metabolite scavengers (such as ascorbic acid, cysteine, and imidazole), and reducing agents (such as $\beta$-mercaptoethanol).[4,17]

### F. Spectral resolution from background signal and noise

Reduction of background and its associated noise is the central theme of all SMFS methods.[3,4,6,7] The signal-to-background ratio (SBR) and the signal-to-noise ratio (SNR, as it refers to background-associated noise) signify the ability of the optical system to detect single molecules with high statistical accuracy; high SBR and SNR allow reliable determination of several fluorescence parameters, such as fluorescence lifetime and anisotropy. To increase SBR and SNR for SMFS, one should select excitation sources, fluorophores, and optics that discriminate fluorescence from background arising from Rayleigh scattering, Raman scattering of water, impurities in reaction buffers and reaction chambers, and, in the case of cells, autofluorescence of cellular components. Fluorophores emitting at the far-red portion of the spectrum can be distinguished easily from most sources of background, and are preferable for studies with cells and turbid samples. Also, fluorophores with large Stokes shifts allow easy discrimination of fluorescence from Rayleigh scattering.

### G. Size of fluorophore and linker

Fluorophores with small and short rigid linkers are preferable because they tend to be less perturbative to their local environment and to contain less ''fluorophore noise'' (fluctuations of the fluorescent properties of a single fluorophore due to unknown and uncontrolled changes of its local environment[6]). Moreover, use of hydrophilic fluorophores reduces hydrophobicity-driven aggregation and unpredictable binding to hydrophobic protein surfaces, thus reducing the uncertainty in the determination of the location of the fluorophores and improving the accuracy of fluorescence resonance energy transfer (FRET)-based distance measurements.

### H. Compatibility with laser excitation sources, optical components, and detectors

Matching the properties of the optical system to the properties of the fluorophore(s) used for SFMS is a critical consideration. The large variety of existing cw lasers in the 400–700 nm range provides flexibility in choosing an excitation source; recently introduced all-solid-state pulsed lasers at the popular wavelengths of 532 and 635 nm are convenient and powerful light sources for time-resolved analysis on single molecules. Customized optics supplement the long list of available microscopy filters and allow efficient discrimination of fluorescence from samples containing two or more fluorophores; this ability has been extended by multi-channel detectors that allow simultaneous detection of photons of multiple wavelengths, generating spectral information for a single molecule as a function of time.[22]

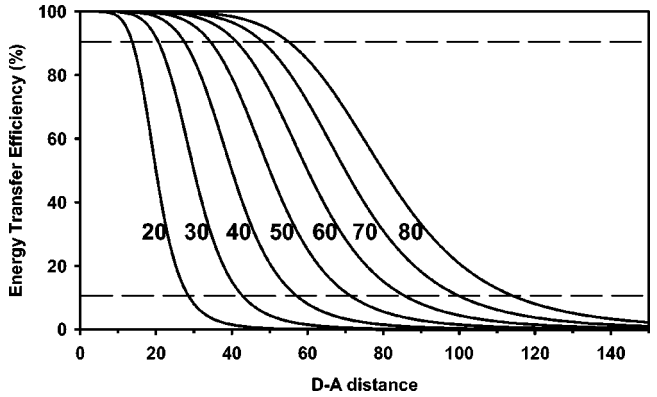


FIG. 1. Dependence of the dynamic range of fluorescence resonance energy transfer (FRET) on the Förster radius $R_0$. The numbers present at the left of each curve correspond to the $R_0$ of each curve (in Å). The dotted lines delineate the regime of maximum sensitivity for each pair with different $R_0$. $D-A$, donor–acceptor.

## III. CHOOSING THE RIGHT FLUOROPHORE FOR THE RIGHT ASSAY

In the case where distances within single biomolecules have to be monitored, the technique of FRET can be used. FRET is a nonradiative physical process that occurs due to weak interactions between the transition dipoles of two complementary fluorophores, a donor ($D$) and an acceptor ($A$). FRET allows measurements of distances between sites spaced by 20–100 Å.[18,23] The efficiency of energy transfer ($E$) is inversely proportional to the sixth power of the distance ($R$) between $D$ and $A$:

$$E=[1+(R/R_0)^6]^{-1},$$

where $R_0$ is the Förster radius, characteristic of the $D-A$ pair and the conditions of the solution; for $R=R_0$, $E=50\%$. $R_0$ depends, in part, on the relative orientation of $D$ and $A$ dipoles, compounded in the orientation factor $\kappa^2$; since $\kappa^2$ cannot be measured experimentally for a single $D-A$ pair, extreme values of $\kappa^2$ (which range from 0 to 4, with 2/3 being the value for an isotropic distribution of dipole orientations) can complicate the distance measurement severely. Therefore, FRET-based distance measurements benefit from rapid reorientation of the fluorophores since it minimizes uncertainties associated with dipole orientation. The rotational freedom of the fluorophores can be evaluated using steady-state and time-resolved fluorescence anisotropy,[24,25] and uncertainty limits can be set for each $D-A$ pair.[26,27]

To maximize the FRET sensitivity in detecting static and dynamic heterogeneity, the $R_0$ values for a $D-A$ pair should lie in the linear range of the $E-R$ dependence curve, with the two extreme-distance end points lying in the range $0.7R_0 < R < 1.5R_0$. This is a consequence of the steep ($1/R^6$) distance-dependence of FRET (Fig. 1). For example, for a $D-A$ pair with $R_0=40$ Å, $E$ changes dramatically for $R$ values bordering $R_0$: $E$ is 85% for $R=30$ Å, 50% at $R$

$=40$ Å, and 21% at $R=50$ Å; thus, the dynamic range is $\sim 28-60$ Å (note: measurements of $R_{D-A}>60$ Å are, in part, hindered due to the tendency of several acceptors (such as the cyanine fluorophore Cy5) to exist in nonabsorbing, nonfluorescent states[28]). Furthermore, a $D-A$ pair of $R_0=70$ Å has no discriminating ability for distances in the 25–45 Å range since for the entire range, $E$ of 100%. Extensive tables with $R_0$ values for $D-A$ FRET pairs are available.[29]

If no high-resolution structural information is available for evaluation of possible distances, one can consider the diameter of the protein of interest (treating it as a globular protein, unless it shows dimensional anisotropy) to estimate a range of distances. Moreover, one can prepare several $D-A$ pairs, with each pair featuring the same donor but different acceptor (or vice versa), thereby exploring a wide ranges of $R_0$ values. Considering the high extinction coefficients and good quantum yields required for efficient detection of single fluorophores, the range of measurable distances extends to $\sim 100$ Å; however, due to the special photophysical properties of SMFS-compatible fluorophores, it becomes challenging to measure *short* distances ($R<30$ Å). For example, the $R_0$ for the $D-A$ pair fluorescein-Cy5 (two fluorophores with apparently small spectral overlap) is $\sim 44$ Å.[30] The inability to probe the short-distance regime (1–10 Å), however, might soon be addressed by use of electron-transfer molecular rulers.[8]

In contrast to FRET, fluorescence anisotropy (FA) explores conformational dynamics and relative orientation of protein elements by probing the transition dipole of rigidly attached fluorophores. To achieve this, a fluorophore with short fluorescence lifetimes (to minimize fluorescence depolarization during the time scale of fluorescence emission) and considerable hydrophobicity (to favor hydrophobic interactions with protein functionalities instead of solvation) is introduced at each of various sites on proteins, and the ensemble FA is used to determine the site with the greatest restriction of fluorophore movement due to noncovalent interactions with the local protein surfaces. Typical anisotropy values used for single-molecule FA range from 0.25 to 0.32 (with limiting anisotropy for a single-dipole fluorophore being in the 0.36–0.4 range). Fluorophore immobilization can also be achieved in a rational and predictable fashion by using intramolecular crosslinking of bis-functional derivatives of fluorophores.

Other assays also have special requirements. Fluorescence-lifetime-based assays, which are useful for ''gating'' the short-lived cellular autofluorescence and for separating mixture components with different fluorescence lifetimes, benefit from fluorophores with long fluorescence lifetimes; however, the lifetimes should not be too long, since this will shift the onset of singlet-state saturation to lower excitation intensities. Several fluorophores can also serve as sensors of the local environment, measuring exposure to solvent, *p*H, and ion concentration; a combination of sensors with FRET or FA probes allows multiparameter analysis of complex systems.

## IV. CHEMISTRIES FOR SITE-SPECIFIC LABELING WITH FLUORESCENT PROBES

Currently, most SMFS experiments are performed *in vitro*, using fluorophores introduced extrinsically after biosynthesis and purification.[16] It is possible to use SMFS to study biomolecules without protein labeling, as has been performed in enzymatic-turnover studies of cholesterol oxidase[31] (where fluorescence arose from the oxidized form of the enzyme cofactor), and in nucleotide-binding studies on the molecular motor myosin[32] (where fluorescent derivatives of nucleotides were used). These cases are protein specific and often require complementary and orthogonal labeling of additional components to extract information. *In vivo* labeling with genetically encoded fluorescent probes is also possible, and its anticipated use in SMFS will allow unprecedented mechanistic investigations in the natural milieu of the cell.

Although the exact site or the high site-specificity of labeling is not required for some imaging applications (provided that labeling does not affect functionality), site-specificity becomes an important issue when precise distance or orientation information is sought. Stochastic (or nonspecific) labeling is inadequate to extract reliable biological information and should be restricted to the preliminary stages of SMFS assay development. This was the case during FRET-based analysis of conformational dynamics of the enzyme staphylococcal nuclease (SNase),[33] where the donor was incorporated in a site-specific fashion, and the acceptor was incorporated nonspecifically on lysine residues of the enzyme. Since SNase has $>10$ surface-exposed lysines, the exact sites of incorporation are unknown. Moreover, the final labeled product contained several species: $D$-only SNase (major species), $D-A$ SNase (with the acceptor in each of several sites), $D-A-A$ SNase (with the acceptors in two of several sites), and minor species with higher degree of labeling. The labeled proteins allowed the development of groundbreaking methodology; however, no reliable biological information could be extracted from these studies.

Site-specific labeling of biomolecules with fluorophores requires careful choice of labeling chemistry, optimization of the labeling reaction and rigorous characterization of the labeled biomolecules for labeling efficiency, site-specificity, and retention of functionality. Labeling nucleic acids is easy, and several fluorophores or reactive groups can be introduced using automated solid-phase synthesis. In this article, we will focus on strategies for site-specific labeling of proteins, multiprotein complexes, and protein–nucleic acid complexes with one or more fluorophores.

The most common method for site-specific labeling of proteins with fluorophores is the *cysteine-specific labeling with thiol-reactive reagents*. During this reaction, proteins with surface-exposed cysteine (Cys) residues are covalently modified by maleimide, iodoacetamide, or other reactive conjugates of fluorophores.[16,34,35] This is the method of choice for small proteins ($<500$ residues), since Cys is a rare amino acid, and can be substituted easily with other amino acids using site-directed mutagenesis,[36] a standard molecular biology technique.

If the protein of interest has no Cys residues, the exact

site of incorporation is selected after inspection of a high-resolution three-dimensional structure (generated using x-ray crystallography or nuclear magnetic resonance) and consideration of the question to be addressed. Labeling should not perturb the enzymatic activity or the accurate spatial arrangement of the protein sequence (also known as the ''protein fold''). Subsequently, an existing amino acid (preferably having a side chain of charge, size, and hydrophobicity similar to that of Cys) at the site of choice is substituted by a Cys using site-directed mutagenesis.[36]

If the unmodified protein has a single preexisting Cys, structural information, along with measurements of the surface accessibility of Cys side chain,[24] will determine whether the existing Cys can be used; otherwise, the preexisting Cys can be converted to the structurally similar amino acid serine, and the procedure for Cys-free proteins can be followed.

SMFS experiments using single-surface Cys proteins labeled using thiol-specific reagents include imaging of single myosin molecules using low-background epifluorescence microscopy;[37] imaging of single myosin molecules, as well as individual cycles of adenosino-triphosphate (ATP, which serves as a biological energy source) binding and hydrolysis (''ATP turnovers'') by total internal reflection fluorescence microscopy (TIRFM);[37] and imaging of single kinesin molecules as they move processively along microtubules.[11] Interestingly, labeling efficiencies as low as 10% were instrumental in establishing new biological information using SMFS.[38]

Fluorophores incorporated on surface Cys residues can also become efficient reporters of the orientation of local protein structure. Tetramethylrhodamine (TMR)-labeled myosin featuring high anisotropy allowed the differentiation between conformational states of myosin.[39,40] In an extension of the seminal experiments that demonstrated the rotation of F1-ATPase $\gamma$-subunit, Kinosita and co-workers used a single fluorophore (instead of a long and extremely bulky fluorescent actin filament) to probe the rotation process by single-fluorophore angular imaging.[10] In this case, the site of fluorophore incorporation was chosen after several cycles of trial and error that resulted in maximization of the fluorescence anisotropy in the labeled protein.

The orientation of a single dipole can be fixed not only through noncovalent interactions of the fluorophore with the local environment, but also by *intramolecular cross-linking of two appropriately spaced Cys residues* with a bis-functional Cys-reactive fluorophore. This approach has been used only for proteins with an existing high-resolution structure. Intramolecular cross-linking restricts the local rotational mobility of the fluorophore significantly, allowing monitoring of the orientation and dynamics of protein domains or other protein structural elements.[41] More importantly, knowledge of the protein structure permits precise design of the *exact* orientation of the fluorophore relative to the protein or to an interacting molecule; such a selection is not possible for immobile fluorophores attached to a single Cys. This approach was introduced in studies of the myosin light-chain orientation relative to actin filaments by ensemble fluorescence polarization[41] and was employed recently in studies of

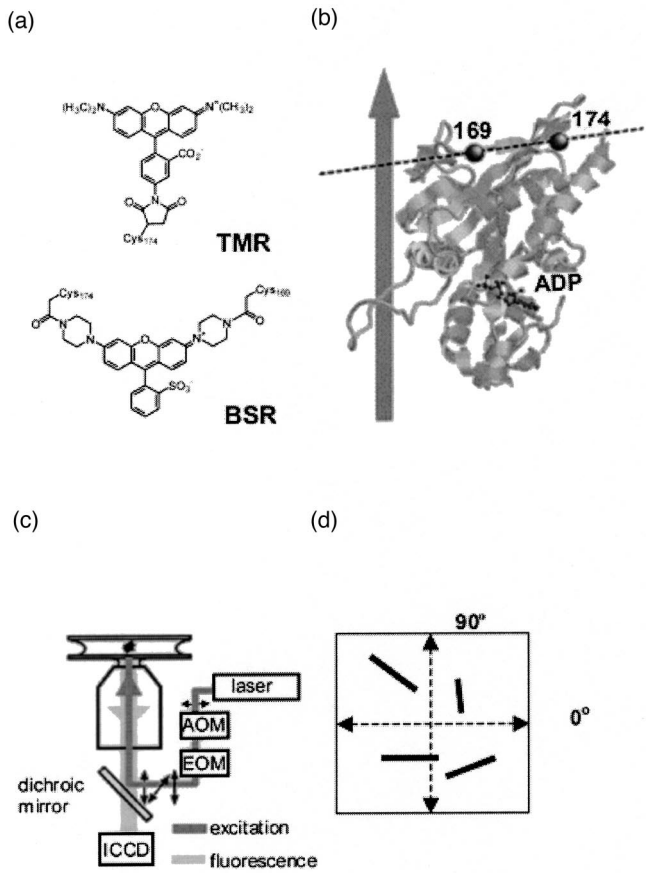


FIG. 2. Strategy for fluorescent labeling and observation of kinesin molecules prepared using intramolecular cross-linking of two appropriately spaced Cys residues. (a) Structures of the fluorophores used to label a single cysteine (Cys174) and a pair of proximal cysteines (Cys169 and Cys174). TMR, tetramethylrhodamine; BSR, bis-[(*N*-iodoacetyl)-piperazinyl]-sulfonerhodamine. (b) Model of kinesin motor domain showing the positions of the selected proximal cysteines; the arrow represents the orientation of microtubules. ADP, adenosine diphosphate. (c) Schematic of the epifluorescence microscope used for the polarization measurements. AOM, acousto-optical modulator; EOM, electro-optical modulator; ICCD, intensified charged-coupled device camera. (d) Schematic showing the random orientation of axonemes on the *xy* plane of the microscope. The dotted arrow lines of 0° and 90° denote the two excitation polarizations used to probe the orientation of the axonemes and associated kinesin molecules. (Reprinted from Ref. 43 with permission from the Biophysical Society).

kinesin intermediates during translocation[42,43](Fig. 2). Sosa and co-workers used the crystallographic structure of kinesin[44] [Fig. 2(b)] to design a kinesin derivative with Cys residues spaced by $\sim 16$ Å, a distance suitable for cross-linking by the bis-functional reactive fluorophore bis-[(*N*-iodoacetyl)-piperazinyl]-sulfonerhodamine [BSR; Fig. 2(a); compare with the mono-functional TMR fluorophore]. Site-directed mutagenesis was used to prepare the protein derivative by substituting eight of the nine preexisting Cys residues of kinesin with alanine residues (keeping only Cys174) and by introducing a second Cys at position Cys169. The two Cys lie on either side of a short $\beta$-sheet [Fig. 2(b)], orienting the transition dipole of BSR to be almost perpendicular (75°) to the long axis of microtubules, and tangential to a radius centered on the microtubule cylinder.[42] The double-Cys kinesin was labeled using a low reactive fluorophore-to-protein molar ratio (1:1; which is

10–100-fold lower than typical reactive fluorophore-to-protein molar ratios used for labeling of single-Cys proteins) to ensure stoichiometric labeling of the site; initial labeling of the first Cys increased the local concentration of the second reactive group of BSR around the second, unreacted Cys, and allowed BSR cross-linking to generate BSR-kinesin (labeling efficiency: 60%–75%). Mass spectrometry of protein digestion products verified the presence of cross-links, although it is not clear what was the exact fraction of cross-linked kinesin in the labeled kinesin product. The spacing and exact orientation of Cys side chains was critical, and the margin for error was small; a similar kinesin derivative with the two Cys spaced by six residues (Cys168 and Cys174) did not result in intramolecular cross-linking. BSR-kinesin was functional and polarization-TIRFM [Fig. 2(c)] shows that it binds to microtubules with the expected orientation [Fig. 2(d)]. Analysis of microtubule-bound BSR-kinesin dynamics uncovered a previously unknown disordered state of kinesin in the presence of adenosine-diphosphate (ADP), and led to new structural models that describe the translocation mechanism of kinesin and the chemical–mechanical coupling during each step.

Analysis of static or dynamic heterogeneity in protein samples (due to the presence of distinct subpopulations or to conformational dynamics) using single-pair FRET requires the incorporation of two complementary, SMFS-compatible fluorophores within a single protein. Initially, this seems easily attainable by generating a pair of appropriately spaced surface-exposed Cys residues ($Cys_1-Cys_2-$protein), and statistical labeling of the two sites with two distinct fluorophores. However, this is true only for dual-Cys proteins with sites of incorporation of the same environment, making SMFS an excellent option for studying homodimeric proteins (symmetric dimers of two identical subunits or protomers) with one surface-exposed Cys *per protomer* ($Cys_1-Cys_{1'}-$protein). Statistical labeling of $Cys_1-Cys_{1'}-$protein will generate the following species: $Cys_1^D-Cys_{1'}^A-$protein, the equivalent $Cys_1^A-Cys_{1'}^D-$protein, the donor-only species $Cys_1^D-Cys_{1'}^D-$protein (which can serve as the perfect donor-only control for FRET measurements, especially after photodestruction of one of the two donor fluorophores), and the acceptor-only species $Cys_1^A-Cys_{1'}^A-$protein (which can serve as an acceptor-only control). SMFS differentiates easily among $D-A$, $D$-only, and $A$-only species, and can use the $D-A$ species for analysis of structure and dynamics (with $D$-only and $A$-only present as internal controls). However, homodimeric proteins constitute a special case; the local environment for two randomly chosen Cys in other proteins is different, and statistical labeling results in the nonequivalent species $Cys_1^D-Cys_2^A-$protein and $Cys_1^A-Cys_2^D-$protein, which have to be resolved, either by purification before data acquisition (e.g., using liquid chromatography[25,45]) or by ''optical'' purification during data analysis. Both of the purification procedures are very difficult or protein specific; to our knowledge, this approach has not been used in SMFS.

An elegant general method to incorporate one or more distinct fluorophores within a single protein involves *peptide ligation*, the assembly of full-length proteins from synthetic or biosynthetic peptide fragments.[46–50] Peptide ligation encompasses various techniques that allow chemical or enzymatic covalent linking of peptides in aqueous solutions. Use of peptides carrying fluorescent probes or other reporter groups and modifications yields protein molecules that are site-specifically labeled at multiple sites. The most established version of peptide ligation is the native chemical ligation, a two-step reaction that occurs between a first peptide containing a *C*-terminal thioester and a second peptide containing an *N*-terminal Cys [Fig. 3(a)]. In the first, reversible step, the *N*-terminal thiol of the second peptide attacks the electrophilic center of the thioester of the first peptide, linking the two peptides. In the second, irreversible step, an intramolecular transfer reaction generates a peptide bond $[R_1-(C=O)-NH-R_2]$ between the two fragments. The striking property of this system is that only the *N*-terminal Cys participates in the second, spontaneous reaction, allowing any number of Cys to be present in virtually any position of the two peptides *without interference*.

If both peptides are prepared using solid-phase peptide synthesis, small proteins or protein domains (up to $\sim$100 residues, due to the 50-residue limit for reliable peptide synthesis) can be assembled. Using the conformationally assisted version of peptide ligation,[50,51] which allows rapid and efficient linking of peptides under conditions that favor the *folded* conformations of the assembled peptides, Deniz and co-workers prepared a doubly labeled version of chymotrypsin inhibitor 2 (CI2), a 64-residue protein that serves as a model for the folding of single protein domains.[52] Fragment (1-39)CI2 featuring a *C*-terminal thioester [functionality $R-(C=O)-SR'$] was synthesized and labeled at the *N*-terminus (position 1) using TMR-succinimidyl ester (an amine-reactive fluorophore). Fragment (40-64)CI2 was synthesized carrying an *N*-terminal Cys (Cys40), a prerequisite for reaction with the *C*-terminal thioester of the (1-39)CI2 fragment. The peptides were ligated under folding conditions, and Cy5 (an efficient FRET acceptor for TMR) was incorporated into the resulting (1-64)CI2 using disulfide exchange (a typical thiol-specific reaction[16]) of CI2 with an activated Cy5-disulfide. Donor–acceptor-labeled CI2 was used to observe folded and unfolded subpopulations by single-pair FRET.

Larger proteins ($>$100 residues) can be labeled using peptide ligation to link short, fluorescently labeled synthetic peptides to larger ''recombinant'' (i.e., prepared by biosynthesis in bacteria) proteins. To achieve this, as in fully synthetic peptide ligation, one should generate proteins with *C*-terminal thioesters and peptides with *N*-terminal Cys. To generate a protein with a *C*-terminal thioester, chemists borrowed a few ideas from the biological process of protein splicing, the remarkable self-cleaving activity of several naturally occurring proteins. In a process resembling an ''inverted'' peptide ligation, certain proteins go through a series of isomerizations and group-transfer reactions to excise an internal protein fragment (an ''intein'') that is nested between two protein fragments (''exteins''); ligation of the exteins yields the final protein product [Fig. 3(b)]. Modulation of the cleavage reaction through protein engineering[53–56] has resulted in inteins that promote only the first step of splicing

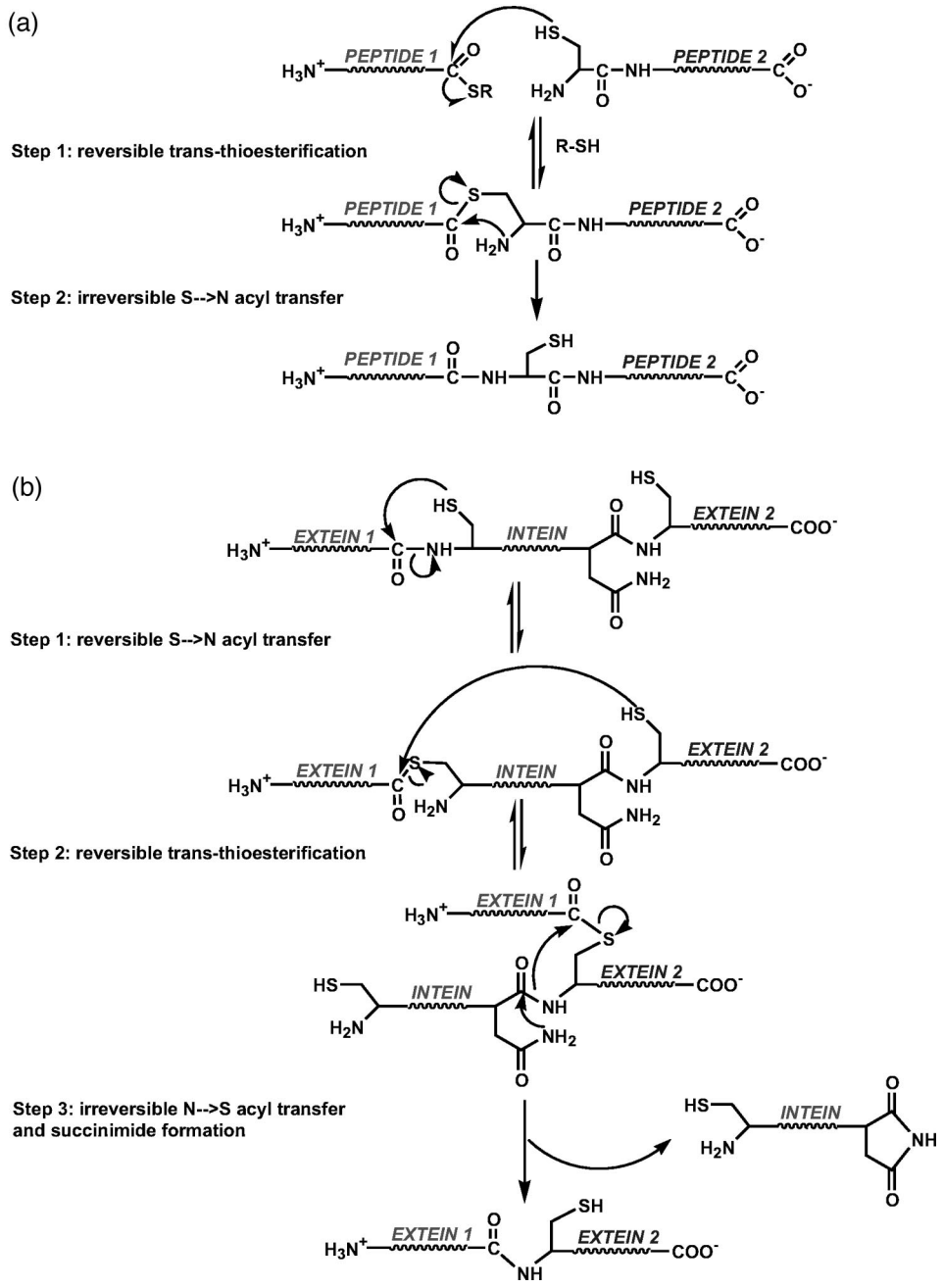


FIG. 3. Principles of peptide ligation and protein splicing. (a) Example of a peptide ligation reaction that yields a long peptide product through the initial reaction of a *C*-terminal thioester (in peptide 1) with the thiol of an *N*-terminal cysteine (in peptide 2) and subsequent isomerization. The peptides can carry various internal functional groups, since such groups do not interfere with the ligation reaction. (b) Example of protein splicing that yields a shorter, processed peptide through a multistep reaction that leads to the excision of the central portion (''intein'') of the initial peptide or protein. (Adapted from Ref. 46 with permission from Elsevier Science.)

and allow formation of *C*-terminal thioesters through intermolecular reactions with excess of thiols. Intein-mediated labeling was used for site-specific labeling of *Escherichia coli* RNAP for systematic, FRET-based, structural analysis of RNAP and RNAP–DNA complexes.[57–59] Each of the $\alpha$ and $\beta'$ subunits were labeled after *in vivo* assembly of RNAP core enzymes containing the modified $\alpha$ or $\beta'$ subunits; the modified subunits had an intein module for labeling, and an affinity tag (CBD: chitin binding domain) for capturing the modified RNAP on a solid support [Fig. 4(a)]. While immobilized, a reaction of the transient thioester moiety of the modified RNAP with Cys-fluorescein [Fig. 4(a), middle panel] simultaneously cleaved the *C*-terminal intein, incorporated a *C*-terminal fluorescein group, and eluted the labeled RNAP from the solid support [Fig. 4(a), left-hand panel]. Fluorescein-RNAP was fully functional, and the efficiency and site-specificity of labeling were outstanding [$>95\%$; Fig. 4(b)]; the protein yield was relatively low ($\sim 0.1$ mg/L of cultured cells), but adequate for several ensemble fluorescence and SMFS experiments. Fluorescein-RNAP was used to show that, contrary to the held belief, the initiation factor $\sigma^{70}$ can remain associated with RNAP beyond the transition of the enzyme from the initiation to the elongation mode of transcription.

Proteins can also be labeled using cell-free RNA translation systems that covalently attach *fluorescent derivatives of the antibiotic puromycin* to the *C*-terminus of the coded protein (Fig. 5). Puromycin inhibits protein synthesis by binding to the *A* site (the site where the incoming aminoacyl-transfer RNA [tRNA] binds) of the ribosome (the protein-synthesis factory of the cell); it can also covalently link the translated protein and the messenger RNA (mRNA).[60] How-

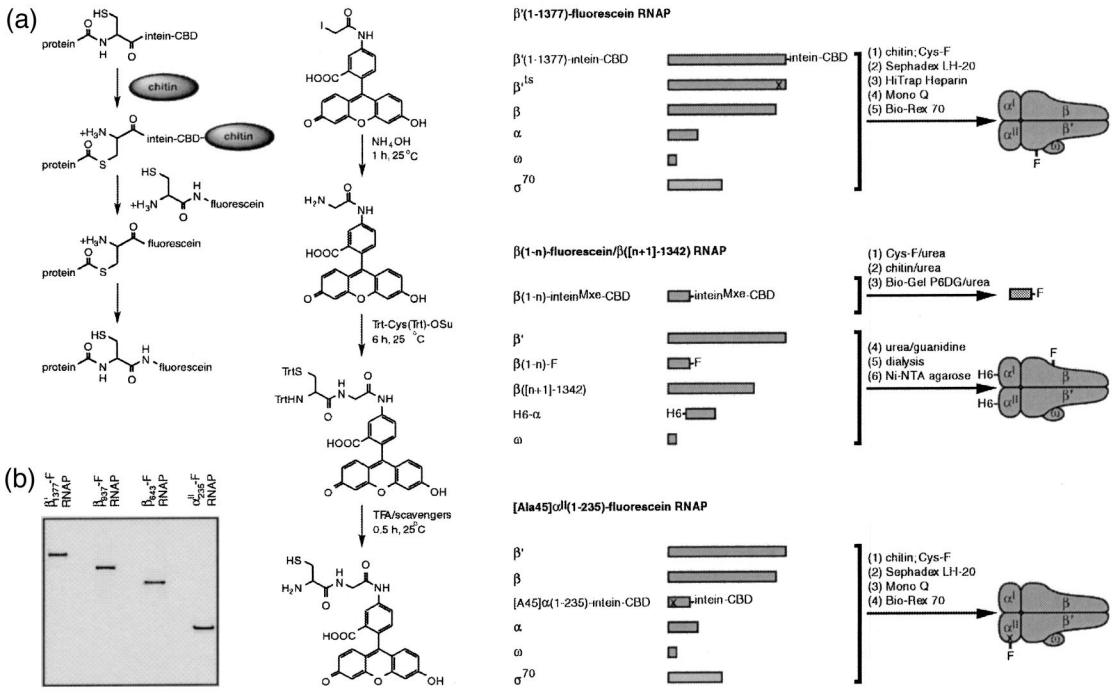


FIG. 4. Intein-mediated labeling of RNA polymerase (RNAP). (a) Incorporation of fluorescein in RNAP core. Left-hand side; Intein-mediated *C*-terminal incorporation of fluorescein using Cys-fluorescein. Center; synthesis of Cys-fluorescein. Right-hand side site-specific incorporation of fluorescein into RNAP core at $\beta'$ 1377, $\beta$ 643 or $\beta$ 937, and $\alpha^{II}$ 235. CBD, chitin binding domain; F, fluorescein; $H_6$, hexahistidine. (b) Labeled RNAP core. Denaturing polyacrylamide electrophoresis gel of RNAP core derivatives labeled at $\beta'$ 1377, $\beta$ 643, $\beta$ 937, or $\alpha^{II}$ 235. Single bands demonstrate the high specificity of labeling. (Reprinted from Ref. 57 with permission from Elsevier Science.)

ever, for mRNA *without* a stop codon (the triplet of nucleotides in mRNA that provides the signal for termination of protein synthesis and release of the complete polypeptide chain), the ribosome stalls and allows covalent modification of the nascent polypeptide chain by puromycin.[61,62] If the puromycin is supplied as a conjugate with a fluorophore, then the targeted protein becomes fluorescent. This method was used to incorporate fluorescein[61] and Cy5[62] [Fig. 5(a)] in proteins, with efficiencies of up to $\sim$90%.[62] SMFS assays of kinesin movement on microtubules showed that Cy5-labeled kinesin prepared using puromycin behaves similarly to unlabeled kinesin,[62] reinforcing the notion of the nonperturbing character of *C*-terminal labeling. Although the yield of labeled protein is lower compared to the intein-mediated protein yields, puromycin-mediated labeling is well-suited for SMFS analysis of proteins, since the high sensitivity of SMFS allows several assays with a simple, 2 h preparation of labeled protein (Fig. 5); however, protein characterization using conventional assays becomes complicated due to the low protein yield. The generality of the method depends partly on the availability of various fluorescent derivatives of puromycin; controlled-pore glass with puromycin groups, convenient for solid-phase synthesis of fluorescent puromycin derivatives, is commercially available (Glen Research, Sterling, VA).

*Unnatural amino acid mutagenesis* also uses *in vitro* transcription/translation for site-specific labeling of proteins.[63–65] Amino acids with side chains carrying reporter or functional groups can be incorporated at any position within a protein by using synthetic amino-acylated tRNA and complementary sequences in the protein-coding DNA;[63] extensions of this method allowed protein modification *in vivo*.[66,67] Incorporation of unnatural amino acids permits direct or indirect introduction of fluorophores. Direct incorporation is performed by using unnatural amino acids with fluorescent side chains; however, the introduced fluorophores are restricted to derivatives of intrinsic protein fluorophores (such as 7-aza-tryptophan, a tryptophan derivative), and are not SMFS-compatible. Indirect incorporation is performed by using unnatural amino acids with nonproteinogenic (i.e., absent in naturally occurring proteins) reactive groups, such as the ketone carbonyl group; such groups allow selective, site-specific labeling in a postsynthetic fashion. This was demonstrated by the site-specific labeling of a single-ketone protein with fluorescein hydrazide.[65] Despite the power of the approach, it has not been used widely for preparing labeled proteins mainly due to the very low yields of the *in vitro* transcription/translation reaction (e.g., large-scale *in vitro* reactions to generate a keto-protein resulted in $\sim$10 $\mu$g protein,[64] $>$100-fold lower than the typical amount of recombinant proteins that are prepared from a small-scale bacterial culture).

A popular way to label large multiprotein or nucleoprotein complexes is to label components of the complex and to assemble the complex by *in vitro reconstitution from purified components*. For example, a hypothetical four-subunit protein with subunit composition ABCD can be site-specifically labeled at subunits A and B by site-specific labeling of subunit A with fluorophore F1 (to prepare $A^{F1}$); site-specific labeling of subunit B with fluorophore F2 (to prepare $B^{F2}$);

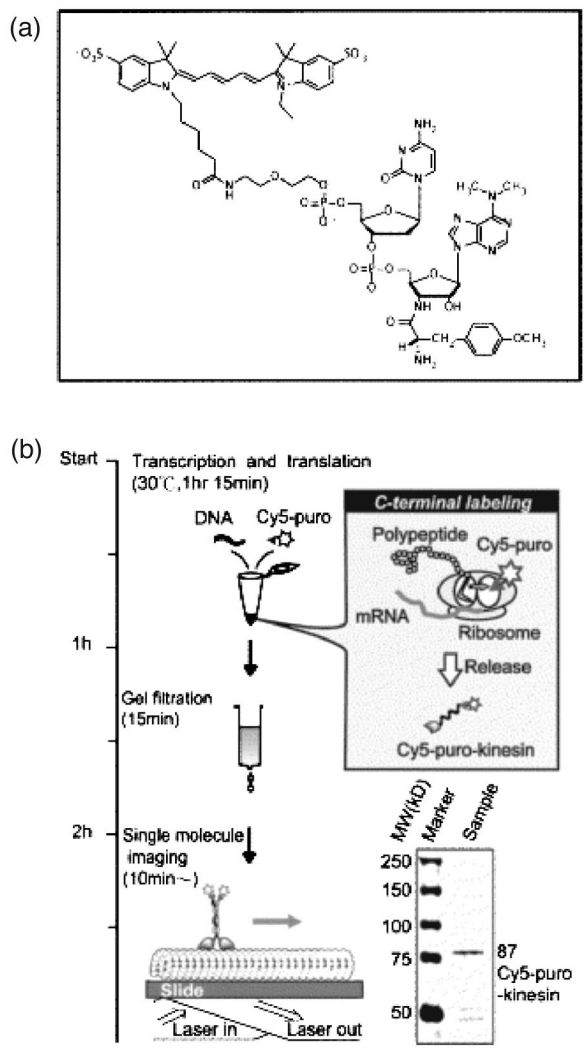


FIG. 5. Preparation of puromycin-labeled kinesin for single-molecule analysis of motility. (a) Structure of Cy5-puromycin conjugate. (b) Schematic of the procedure. Messenger RNA without a stop codon is prepared using an *in vitro* transcription system containing the gene for kinesin; addition of an *in vitro* translation mix and Cy5-puromycin and incubation allow formation of *C*-terminally-labeled kinesin. After purification, the enzyme is diluted and used for motility assays performed on surface-immobilized microtubules. (c) Fluorescent kinesin molecules moving along microtubules (top panel) are visualized using TIRFM. Scale bar, 5 $\mu$m. [Reprinted from Ref. 62 with permission from the FEBS society.]

and mixing labeled subunits $A^{F1}$ and $B^{F2}$ with purified, unlabeled subunits C and D (to prepare the desired $A^{F1}B^{F2}CD$). This can also be achieved by subunit exchange of an unlabeled component with excess of a labeled component (e.g., $ABCD+A^{F1}\rightarrow A^{F1}BCD+A$). Once more, it is critical to establish that fluorescent labeling does not adversely affect the assembly of the complex. This method is especially useful for large proteins, such as myosin[39,40] and RNAP.[57,58] For myosin, the regulatory light chain (RLC) was site-specifically labeled and subsequently exchanged with unlabeled RLC, allowing probing of myosin conformational states. For RNAP, subunit $\sigma^{70}$ was labeled with thiol-reactive fluorophores and incubated with unlabeled RNAP core (subunit composition $\beta'\beta\alpha_2\omega$) to form RNAP holoenzyme (for $\sigma\rightarrow$DNA FRET experiments within RNAP–DNA complexes[57,58]), or with fluorescein-labeled RNAP core (for RNAP core$\rightarrow\sigma$ FRET experiments within RNAP holoenzyme or RNAP–DNA complexes[57,58]). The method has revealed the existence of subpopulations of RNAP–DNA complexes with different behavior relative to promoter escape[59] and allowed investigation of asynchronous processes in transcription.

The reconstitution approach is not limited to assembly of complexes using labeled individual subunits, but extends to assembly of complexes using labeled *fragments* of individual subunits (''split subunits''), thus allowing *internal* labeling *of very large protein subunits*. In an illustration of this concept (Fig. 5), the $\beta$ subunit of RNAP was internally labeled at each of two positions (643 and 937) by preparing $\beta$(1-643)-F/ and $\beta$(1-937)-F (using intein-mediated labeling), and by reconstituting [$\beta$(1-643)-F/$\beta$(644-1342)]-RNAP core and [$\beta$(1-937)-F/(938-1342)]-RNAP core *in vitro*.[57] This complemented a panoply of site-specific labeling methodologies that resulted in incorporation of fluorophores at 4 sites in RNAP core, 20 sites in $\sigma^{70}$, and 3 sites on DNA, allowing measurement of a large set of interprobe distances by ensemble FRET and yielding a detailed picture of RNAP holoenzyme and the RNAP–DNA open complex in solution.

## V. SITE-SPECIFIC LABELING OF PROTEINS IN LIVING CELLS

*In vitro* analysis of biological processes has led to formulation of elaborate models about the mechanisms operating in the living cell. *In vivo* analysis of such processes is essential for the validation of the proposed models, and for the understanding of additional processes that cannot be replicated *in vitro* due to inability to account for all interacting components. Observation of SMFS-compatible fluorophores in living cells is more complicated than *in vitro*, primarily due to the cellular autofluorescence (fluorescence background in the cell due to naturally occurring fluorophores), the difficult control of the fluorophore concentration, and the multitude of interactions that can interfere with the fluorophore detection. There are three major strategies for using site-specifically labeled proteins to perform SMFS in living cells.

In the first and most popular strategy, the labeled proteins are prepared *in vitro*, and are introduced into the cell

through a variety of cell-loading techniques, such as endocytosis, permeabilization, or microinjection. This strategy has allowed SMFS observations of the epidermal growth factor (EGF) dimerization by monitoring FRET between Cy3-EGF and Cy5-EGF;[68] imaging of single $\beta$-galactosidase molecules (with $\sim 5$ Alexa-488 fluorophores per molecule) in the cell nucleus;[69] and detection of diffusing proteins and nucleic acids (as fluorescence bursts) within living cells.[70] The *in vitro*-labeled biomolecules can also be large structures, such as virus particles. Labeling of adeno-associated viruses with a single far-red fluorophore (Cy5) allowed real-time TIRFM trajectories of the viral infection pathway in living cells, starting with the initial contact of virus with the cell surface, continuing with its journey through the cytoplasm, and concluding with invasion of the nucleus.[15]

In the second strategy, a high-affinity protein–ligand interaction is used for *in situ* labeling of specific sequences or structures in the cell. This task is much more difficult than *in vitro* labeling, since the site-specificity of the reagent should be exceptional to avoid nonspecific binding or incorporation to the tens of thousands of different biomolecules found in the cell; the labeling reagent should also be nontoxic and cell-permeant. This strategy was used for labeling of single ion channels on the membrane of Jurkat cells, where Cy5-labeled toxin targeted individual channels.[71] A variation of this strategy allows covalent labeling of proteins in living cells by targeting a 4-Cys $\alpha$-helical motif incorporated to the targeted protein.[72] The validity of this strategy has been demonstrated with the fluorescein-related reagent FLASH (fluorescein arsenical helix binder), and has been extended recently to a second fluorophore. FLASH-based labeling can be used in conjunction with genetically-encoded fluorescent proteins for real-time FRET measurements in living cells. However, this strategy is limited to two fluorophores, and, in some cell lines, is not applicable due to nonspecific interactions with other Cys-rich proteins.[73] Another variation, termed ''epitope tagging,'' uses expression of single-chain antibodies at specific intracellular locations (as receptor-antibody fusion proteins) and labeling of specific sites through high-affinity interactions of the antibody with fluorescent conjugates of its ligand.[74] The last two variations of the *in situ* labeling strategy have not been used yet for SMFS.

In the third strategy, the fluorophore is a genetically encoded fluorescent protein, such as the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria*.[75] According to this method, the DNA sequence coding for GFP is placed immediately adjacent to the sequence coding for the protein of interest. During biosynthesis, the protein will be prepared as a ''GFP-fusion'' protein (i.e., a hybrid protein containing GFP and the protein of interest), and thus, it will be fluorescent. After the DNA is introduced to the cells, the cells synthesize the GFP fusion, which then travels to the cellular structures or compartments that it normally occupies. Variants of GFP (blue, cyan, and green/yellow) or other genetically encoded fluorescent proteins can also be used, allowing flexibility in terms of excitation and emission wavelengths, and generating genetically encoded $D-A$ pairs for *in vivo* FRET. However, use of GFP has its limitations;[72] it can only be introduced at the protein termini and can cause perturbations due to its large size (238 amino acids in length). Furthermore, its limited brightness and its spectral overlap with cellular autofluorescence complicates SMFS. Nonetheless, the first observation of single GFP-fusion proteins in living cells was reported recently[76] (Fig. 6). Using TIRFM [Fig. 6(a)], single copies of a GFP fusion of the adhesion protein

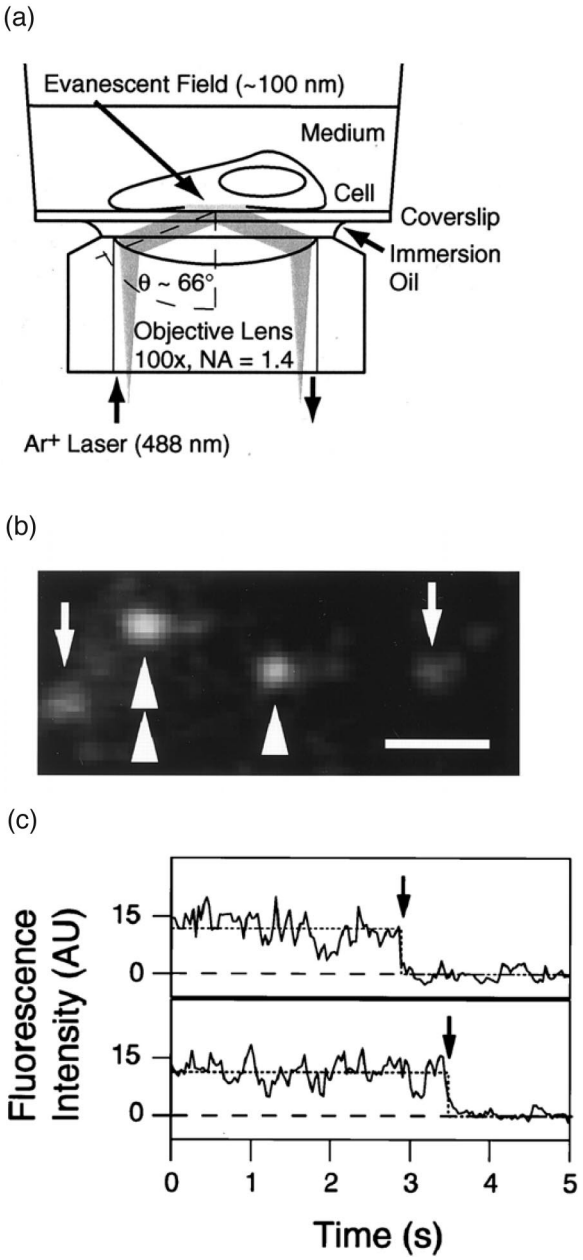


FIG. 6. Single-molecule imaging of GFP-cadherin in living cells. (a) Schematic of the chamber-objective assembly used for the objective-type TIRFM measurements. The cell membrane facing the coverslip is illuminated by the evanescent wave generated by the total internal reflection of the 488 nm laser excitation beam. (b) Image of membrane showing immobile and diffusing molecules exhibiting GFP fluorescence. The single arrowhead points to a possible monomeric GFP-cadherin species, whereas the double arrowhead points to a possible dimeric GFP-cadherin species. The dimmer spots are attributed to diffusing species Scale bar, 1 $\mu$m. (c) Single-step photobleaching of monomeric GFP-cadherin species. The arrow points to the exact timing of photobleaching. (Reprinted from Ref. 76 with permission from the Biophysical Society.)

cadherin (E-cad-GFP) were observed on the membrane of fibroblast cells, and their oligomerization status was investigated. To allow the observation of single copies of E-cad-GFP, the cell line was optimized to minimize the concentration of biosynthesized E-cad-GFP, and the coverslip-proximal membrane was exposed to the evanescent wave for a few seconds to reduce the concentration of the fluorescent species (by photobleaching), thus allowing observation of *individual*, well-separated fluorescent spots [Fig. 6(b)]. Histograms of fluorescence intensity for individual spots showed distributions of quantized fluorescence intensity; spots exhibiting the basic quantized intensity (presumably monomers of E-cad-GFP) also display single-step photobleaching, which serves as a diagnostic test for observing a single fluorophore. Based on these results, an oligomerization-induced trapping model was proposed for the mechanism of cell-adhesion structure formation.

## VI. NEW FLUOROPHORES AND NEW CHEMISTRIES FOR *IN VITRO* AND *IN VIVO* LABELING

Despite the battery of available strategies, site-specific labeling with SMFS-compatible fluorophores is still a limiting step in establishing SMFS assays. Since the popularity of SMFS for analysis of biomolecules is growing rapidly,[1] SMFS is likely to be used extensively for the analysis of numerous human proteins. However, the vast number of proteins found in the human proteome (30 000–40 000), the large number of multiprotein and other multicomponent complexes, and the intricate network of transient protein–protein interactions involved in the structural, regulatory, and communication network of the cell, underlines the need for new fluorophores and new labeling strategies.

Fluorescent semiconductor nanocrystals (NC; or quantum dots) comprise a novel family of fluorophores for labeling biomolecules.[77,78] NC possess several advantages over conventional organic fluorophores, such as broad excitation spectra, narrow and tunable emission spectra, long fluorescence lifetimes, and high resistance to photodestruction. However, bioconjugation and cell-targeting chemistries for NC are still immature, and the intermittency of their fluorescence complicates their use. However, since NC define a particularly active area of research, it is certain that exciting future developments are on the horizon.

Another challenge for both physicists and chemists is the exploration of the UV wavelength range for SMFS. Ability to operate in the UV will allow smaller, less perturbative fluorophores to be used, and will extend the available emission range for SMFS, allowing *simultaneous* probing of a larger set of fluorophores. Currently, SMFS uses only very bright fluorophores with emissions confined in the 450–700 nm spectral range. SMFS in the UV requires the development of brighter UV fluorophores with large Stokes shifts, as well as reliable, intense, and inexpensive UV laser sources. Innovative ways to amplify and to detect the faint signal of intrinsic protein fluorophores (such as tryptophan and its red-shifted analogs), or of DNA-structure-sensitive fluorophores (such as 2-aminopurine [a reporter of DNA double-helicity] or intercalating dyes) will create further opportunities for multiparametric analysis of complex systems.

It is not an exaggeration to say that the best probes for SMFS might already be on the shelves, since thousands of fluorophores have never been screened for performance in SMFS (although they have been evaluated for use in ensemble fluorescence assays). High-throughput characterization of the available libraries of fluorophores for compatibility with SMFS could generate a new category, the ''SMFS-optimal'' probes. Such a comprehensive search could also generate rules that will allow the rational design of SMFS-compatible fluorophores. This effort should be accompanied by a quest for reagents that both extend photobleaching lifetime of fluorophores and quench triplet states.

Site-specific labeling strategies of greater convenience and speed should also be pursued. One way to label recombinant proteins *in vitro* or *in vivo* is to target ''affinity tags'' (short sequences introduced usually at protein termini to allow rapid purification), similar to the FLASH-based site-specific labeling strategy. Another sequence that can be targeted is the amino acid sequence $His_6$ (''hexahistidine tag'') that binds tightly to transition-metal complexes. Targeting of $His_6$ by $(Ni^{2+}$-nitrilotriacetic acid$)_2$-fluorophore conjugates was reported recently;[30] however, the interaction of $His_6$ with the fluorophore conjugates had dissociation constants of $>300$ nM, requiring a large increase in the affinity to make the approach useful for SMFS. This might be possible by using the concept outlined in the construction of antibodies with ''infinite affinity.'' [79] During this remarkable feat of protein engineering, the high-affinity, noncovalent antibody–ligand interaction was used to increase the effective local concentrations of complementary reactive groups placed on each component, resulting in covalent linkage of antibody and ligand and preventing dissociation. In general, to use ''affinity-labeling'' for labeling of living cells, it is necessary to devise chemistries that render the reagent fluorescent only *after* incorporation to the site of interest, thus minimizing the background of unincorporated reagent that will otherwise overwhelm the SMFS signal.

We conclude with a note on the potential of the genetic code expansion (by unnatural amino acid mutagenesis) to revolutionize *in vivo* site-specific labeling. Wang and coworkers have recently developed a sophisticated strategy that combines novel nucleic acids (tRNA molecules) and engineered enzymes (amino-acyl tRNA synthetases) for *in vivo* site-specific incorporation of two unnatural amino acids (L-3-[2-napthyl]-alanine[66] or O-methyl-L-tyrosine[67]) in proteins. If this breakthrough work is extended to encompass side chains with SMFS-compatible fluorophores, it will essentially miniaturize the genetically encoded fluorophores, by reducing the number of additional amino acid residues required for *in vivo* fluorescence from 238 (for GFP) to 1 (the single unnatural amino acid). This advance will help unravel the deep secrets of the cell and open new avenues for exciting research.

## ACKNOWLEDGMENTS

The authors thank Sören Doose for the critical reading of the manuscript, Dr. Richard H. Ebright for helpful discussion and access to unpublished data, and Joanne Tang for editorial assistance. This work was funded by the National Institutes

of Health, Grant No. 1 R01 GM65382:01, and by the U.S. Department of Energy, Grant No. DE-FG03-02ER63339.

[1] A. M. Kelley, X. Michalet, and S. Weiss, Science **292**, 1671 (2001).
[2] Y. Harada, T. Funatsu, M. Tokunaga, K. Saito, H. Higuchi, Y. Ishii, and T. Yanagida, Methods Cell Biol. **55**, 117 (1998).
[3] W. P. Ambrose, P. M. Goodwin, J. H. Jett, A. Van Orden, J. H. Werner, and R. A. Keller, Chem. Rev. **99**, 2929 (1999).
[4] S. Nie and R. N. Zare, Annu. Rev. Biophys. Biomol. Struct. **26**, 567 (1997).
[5] S. Weiss, Nat. Struct. Biol. **7**, 724 (2000).
[6] S. Weiss, Science **283**, 1676 (1999).
[7] W. E. Moerner and M. Orrit, Science **283**, 1670 (1999).
[8] S. Xie, Single Molecules **2**, 229 (2001).
[9] R. Kuhnemuth and C. A. M. Seidel, Single Molecules **2**, 251 (2001).
[10] K. Adachi, R. Yasuda, H. Noji, H. Itoh, Y. Harada, M. Yoshida, and K. Kinosita, Proc. Natl. Acad. Sci. U.S.A. **97**, 7243 (2000).
[11] R. D. Vale, T. Funatsu, D. W. Pierce, L. Romberg, Y. Harada, and T. Yanagida, Nature (London) **380**, 451 (1996).
[12] D. W. Pierce, N. Hom-Booher, and R. D. Vale, Nature (London) **388**, 338 (1997).
[13] Y. Harada, T. Funatsu, K. Murakami, Y. Nonoyama, A. Ishihama, and T. Yanagida, Biophys. J. **76**, 709 (1999).
[14] X. Zhuang, L. E. Bartley, H. P. Babcock, R. Russell, T. Ha, D. Herschlag, and S. Chu, Science **288**, 2048 (2000).
[15] G. Seisenberger, M. U. Ried, T. Endress, H. Buning, M. Hallek, and C. Brauchle, Science **294**, 1929 (2001).
[16] A. Waggoner, Methods Enzymol. **246**, 362 (1995).
[17] R. Tsien and A. Waggoner, in *Handbook of Biological Confocal Microscopy*, edited by J. D. Pawley (Plenum, New York, 1995), pp. 267–279.
[18] T. Ha, Methods **25**, 78 (2001).
[19] C. G. Hubner, A. Renn, I. Renge, and U. P. Wild, J. Chem. Phys. **115**, 9619 (2001).
[20] F. Wilkinson, D. J. McGarvey, and O. F. Olea, J. Phys. Chem. **98**, 3762 (1994).
[21] C. Eggeling, J. Widengren, R. Rigler, and C. A. M. Seidel, Anal. Chem. **70**, 2651 (1998).
[22] M. E. Dickinson, G. Bearman, S. Tilie, R. Lansford, and S. E. Fraser, BioTechniques **31**, 1272 (2001); **31**, 1274 (2001).
[23] P. R. Selvin, Nat. Struct. Biol. **7**, 730 (2000).
[24] A. N. Kapanidis, Y. W. Ebright, R. D. Ludescher, S. Chan, and R. H. Ebright, J. Mol. Biol. **312**, 453 (2001).
[25] M. P. Lillo, J. M. Beechem, B. K. Szpikowska, M. A. Sherman, and M. T. Mas, Biochemistry **36**, 11261 (1997).
[26] R. E. Dale and J. Eisinger, Proc. Natl. Acad. Sci. U.S.A. **73**, 271 (1976).
[27] R. E. Dale, J. Eisinger, and W. E. Blumberg, Biophys. J. **26**, 161 (1979).
[28] T. J. Ha, A. Y. Ting, Liang, A. A. Deniz, D. S. Chemla, P. G. Schultz, and S. Weiss, Chem. Phys. **247**, 107 (1999).
[29] P. Wu and L. Brand, Anal. Biochem. **218**, 1 (1994).
[30] A. N. Kapanidis, Y. W. Ebright, and R. H. Ebright, J. Am. Chem. Soc. **123**, 12123 (2001).
[31] H. P. Lu, L. Xun, and X. S. Xie, Science **282**, 1877 (1998).
[32] A. Ishijima, H. Kojima, T. Funatsu, M. Tokunaga, H. Higuchi, H. Tanaka, and T. Yanagida, Cell **92**, 161 (1998).
[33] T. Ha, A. Y. Ting, J. Liang, W. B. Caldwell, A. A. Deniz, D. S. Chemla, P. G. Schultz, and S. Weiss, Proc. Natl. Acad. Sci. U.S.A. **96**, 893 (1999).
[34] R. P. Haugland, *Handbook of Fluorescent Probes and Research Products*, 8th ed. (2002).
[35] P. R. Selvin, Methods Enzymol. **246**, 300 (1995).
[36] T. A. Kunkel, K. Bebenek, and J. McClary, Methods Enzymol. **204**, 125 (1991).
[37] T. Funatsu, Y. Harada, M. Tokunaga, K. Saito, and T. Yanagida, Nature (London) **374**, 555 (1995).
[38] R. Yamasaki, M. Hoshino, T. Wazawa, Y. Ishii, T. Yanagida, Y. Kawata, T. Higurashi, K. Sakai, J. Nagai, and Y. Goto, J. Mol. Biol. **292**, 965 (1999).
[39] D. M. Warshaw, E. Hayes, D. Gaffney, A. M. Lauzon, J. Wu, G. Kennedy, K. Trybus, S. Lowey, and C. Berger, Proc. Natl. Acad. Sci. U.S.A. **95**, 8034 (1998).
[40] I. Sase, H. Miyata, S. Ishiwata, and K. Kinosita, Jr., Proc. Natl. Acad. Sci. U.S.A. **94**, 5646 (1997).
[41] J. E. Corrie, B. D. Brandmeier, R. E. Ferguson *et al.*, Nature (London) **400**, 425 (1999).
[42] H. Sosa, E. J. Peterman, W. E. Moerner, and L. S. Goldstein, Nat. Struct. Biol. **8**, 540 (2001).
[43] E. J. Peterman, H. Sosa, L. S. Goldstein, and W. E. Moerner, Biophys. J. **81**, 2851 (2001).
[44] F. J. Kull, E. P. Sablin, R. Lau, R. J. Fletterick, and R. D. Vale, Nature (London) **380**, 550 (1996).
[45] M. Sinev, P. Landsmann, E. Sineva, V. Ittah, and E. Haas, Bioconjugate Chem. **11**, 352 (2000).
[46] G. J. Cotton and T. W. Muir, Chem. Biol. **6**, R247 (1999).
[47] T. W. Muir, D. Sondhi, and P. A. Cole, Proc. Natl. Acad. Sci. U.S.A. **95**, 6705 (1998).
[48] K. Severinov and T. W. Muir, J. Biol. Chem. **273**, 16205 (1998).
[49] J. P. Tam, J. Xu, and K. D. Eom, Biopolymers **60**, 194 (2001).
[50] P. E. Dawson and S. B. Kent, Annu. Rev. Biochem. **69**, 923 (2000).
[51] P. E. Dawson, T. W. Muir, I. Clark-Lewis, and S. B. Kent, Science **266**, 776 (1994).
[52] A. A. Deniz, T. A. Laurence, G. S. Beligere, M. Dahan, A. B. Martin, D. S. Chemla, P. E. Dawson, P. G. Schultz, and S. Weiss, Proc. Natl. Acad. Sci. U.S.A. **97**, 5179 (2000).
[53] M. Q. Xu and F. B. Perler, EMBO J. **15**, 5146 (1996).
[54] S. Mathys, T. C. Evans, I. C. Chute, H. Wu, S. Chong, J. Benner, X. Q. Liu, and M. Q. Xu, Gene **231**, 1 (1999).
[55] T. C. Evans, Jr., J. Benner, and M. Q. Xu, J. Biol. Chem. **274**, 3923 (1999).
[56] S. Chong, F. B. Mersha, D. G. Comb *et al.*, Gene **192**, 271 (1997).
[57] V. Mekler, E. Kortkhonjia, J. Mukhopadhyay, J. Knight, A. Revyakin, A. Kapanidis, W. Niu, Y. W. Ebright, R. Levy, and R. H. Ebright, Cell **108**, 1 (2002).
[58] J. Mukhopadhyay, A. N. Kapanidis, V. Mekler, E. Kortkhonjia, Y. W. Ebright, and R. H. Ebright, Cell **106**, 453 (2001).
[59] A. N. Kapanidis, T. Lacoste, V. Mekler, J. Mukhopadhyay, S. Weiss, and R. H. Ebright, Biophys. J. **80**, 478A (2001).
[60] N. Nemoto, E. Miyamoto-Sato, Y. Husimi, and H. Yanagawa, FEBS Lett. **414**, 405 (1997).
[61] N. Nemoto, E. Miyamoto-Sato, and H. Yanagawa, FEBS Lett. **462**, 43 (1999).
[62] J. Yamaguchi, N. Nemoto, T. Sasaki, A. Tokumasu, Y. Mimori-Kiyosue, T. Yagi, and T. Funatsu, FEBS Lett. **502**, 79 (2001).
[63] D. Mendel, V. W. Cornish, and P. G. Schultz, Annu. Rev. Biophys. Biomol. Struct. **24**, 435 (1995).
[64] V. Cornish, K. M. Hahn, and P. G. Schultz, J. Am. Chem. Soc. **118**, 8150 (1996).
[65] V. W. Cornish, D. R. Benson, C. A. Altenbach, K. Hideg, W. L. Hubbell, and P. G. Schultz, Proc. Natl. Acad. Sci. U.S.A. **91**, 2910 (1994).
[66] L. Wang, A. Brock, and P. G. Schultz, J. Am. Chem. Soc. **124**, 1836 (2002).
[67] L. Wang, A. Brock, B. Herberich, and P. G. Schultz, Science **292**, 498 (2001).
[68] Y. Sako, S. Minoghchi, and T. Yanagida, Nat. Cell Biol. **2**, 168 (2000).
[69] T. Kues, R. Peters, and U. Kubitscheck, Biophys. J. **80**, 2954 (2001).
[70] T. A. Byassee, W. C. Chan, and S. Nie, Anal. Chem. **72**, 5606 (2000).
[71] G. J. Schutz, V. P. Pastushenko, H. J. Gruber, H-G. Knaus, B. Pragl, and H. Schindler, Single Molecules **1**, 25 (2000).
[72] B. A. Griffin, S. R. Adams, and R. Y. Tsien, Science **281**, 269 (1998).
[73] K. Stroffekova, C. Proenza, and K. G. Beam, Pfluegers Arch. **442**, 859 (2001).
[74] J. Farinas and A. S. Verkman, J. Biol. Chem. **274**, 7603 (1999).
[75] R. Y. Tsien, Annu. Rev. Biochem. **67**, 509 (1998).
[76] R. Iino, I. Koyama, and A. Kusumi, Biophys. J. **80**, 2667 (2001).
[77] M. Bruchez, Jr., M. Moronne, P. Gin, S. Weiss, and A. P. Alivisatos, Science **281**, 2013 (1998).
[78] X. Michalet, F. Pinaud, T. D. Lacoste, M. Dahan, M. P. Bruchez, A. P. Alivisatos, and S. Weiss, Single Molecules **2**, 261 (2001).
[79] A. J. Chmura, M. S. Orton, and C. F. Meares, Proc. Natl. Acad. Sci. U.S.A. **98**, 8480 (2001).
[80] A. N. Kapanidis, R. H. Ebright, and S. Weiss (unpublished).