Measuring Size Distribution in Highly Heterogeneous Systems with Fluorescence Correlation Spectroscopy

Parijat Sengupta, K. Garai, J. Balaji, N. Periasamy, and S. Maiti
Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai 400005, India

ABSTRACT Fluorescence correlation spectroscopy (FCS) is a sensitive and widely used technique for measuring diffusion. FCS data are conventionally modeled with a finite number of diffusing components and fit with a least-square fitting algorithm. This approach is inadequate for analyzing data obtained from highly heterogeneous systems. We introduce a Maximum Entropy Method based fitting routine (MEMFCS) that analyzes FCS data in terms of a quasicontinuous distribution of diffusing components, and also guarantees a maximally wide distribution that is consistent with the data. We verify that for a homogeneous specimen (green fluorescent protein in dilute aqueous solution), both MEMFCS and conventional fitting yield similar results. Further, we incorporate an appropriate goodness of fit criterion in MEMFCS. We show that for errors estimated from a large number of repeated measurements, the reduced χ^2 value in MEMFCS analysis does approach unity. We find that the theoretical prediction for errors in FCS experiments overestimates the actual error, but can be empirically modified to serve as a guide for estimating the goodness of the fit where reliable error estimates are unavailable. Finally, we compare the performance of MEMFCS with that of a conventional fitting routine for analyzing simulated data describing a highly heterogeneous distribution containing 41 diffusing species. Both methods fit the data well. However, the conventional fit fails to reproduce the essential features of the input distribution, whereas MEMFCS yields a distribution close to the actual input.

INTRODUCTION

Fluorescence correlation spectroscopy (FCS) is one of the most powerful and sensitive techniques for measuring diffusion constants (and therefore the size) of particles in solution (Magde et al., 1972; Elson and Magde, 1974; Thompson, 1991; Eigen and Rigler, 1994; Maiti et al., 1997). It tracks spontaneous concentration fluctuations occurring in a small open volume of a dilute solution using sensitive fluorescence detection. The temporal autocorrelation of these fluctuations can be interpreted in terms of the diffusion constants of the particles and chemical kinetic rate constants of interconversion between them. FCS has been the method of choice for a variety of experimental problems, such as, for measuring diffusion and binding of small fluorescent molecules to larger substrates (Schwille et al., 1996), for investigating the spontaneous chemical kinetics of protein molecules (Haupts et al., 1998; Kummer et al., 2000), and for obtaining the intracellular viscosity of live cells (Berland et al., 1995; Schwille et al., 1999).

Parameters such as diffusion constants are obtained from FCS data by fitting it to an appropriate model. The typically used conventional models assume a small number of discrete diffusing species and are adequate for describing simple systems with limited heterogeneity. However, FCS is being increasingly used to measure dynamics in highly heterogeneous biological systems, e.g., to follow the functionally important oligomerization of receptors on cell membranes (Olsson et al., 2001) and to examine aggregation of prionlike proteins implicated in physiological disorders (Post et al., 1998; Tjernberg et al., 1999; Sengupta et al., 2002). The conventional model is inadequate for describing such situations. Even if the data can be adequately fit by a small number of diffusing components, this may lead to an unphysical description of the real system under study.

An additional problem with FCS data analysis is the lack of a convenient way to estimate the goodness of a fit. Typical FCS data processing hardware modules do not preserve the raw data (photon counts with time), but only provide the autocorrelation averaged over time, without any quantitative information on the noise. Because relative errors of data points remain unknown, they are all given equal weights, and a measure of the goodness of fit such as the reduced χ^2 becomes meaningless. It has recently been shown that for repeated averaged autocorrelation measurements the standard error of the mean does provide a good description of the error (Wohland et al., 2001). However, in most practical situations, such large number of repeats is not practicable and an analytical method for estimating the relative errors of data points is necessary. Koppel provided such a formulation under certain assumptions (Koppel, 1974), and it has been shown that a modified version of Koppel's analytical derivation does provide a reasonable description of the errors, but only at short enough timescales (Wohland et al., 2001). It is desirable that any fitting routine for FCS takes into account errors of individual points, whether actually measured or analytically estimated, and yields a value for the reduced χ^2 that can serve as a meaningful measure for the goodness of fit.

Here we present a data-fitting algorithm for FCS based on the Maximum Entropy Method (henceforth called MEMFCS). The Maximum Entropy Method was first

Submitted November 11, 2001, and accepted for publication July 22, 2002.

Address reprint requests to N. Periasamy or S. Maiti, E-mail: peri@tifr.res.in or maiti@tifr.res.in.

0006-3495/03/03/1977/08 $2.00

proposed for the reconstruction of astronomical images and has the virtue of preserving the maximum uncertainty in the estimation of parameters that is consistent with the data (Skilling and Bryan, 1984). In our implementation in the context of FCS, it provides a bias-free fitting of the data with a quasicontinuous distribution of a large number of diffusing components.

We test this algorithm with data obtained from experimental FCS measurements of diffusion in a simple well-characterized system, viz. a dilute aqueous solution of green fluorescent protein (EGFP mutant). The conventional fitting routine incorporating a single diffusing component (with diffusion time τ_D) provides a reliable description of this system, and thus it presents an opportunity to verify the MEMFCS method. Subsequently we address the goodness of fit issue by taking into account the uncertainties associated with individual points. This is done with errors estimated in three different ways: i), from multiple repeats of the same experiment, ii), from an analytical computation of errors, and iii), by ascribing equal errors to all the points. We test the efficacy of the reduced χ^2 estimation of the goodness of fit in each of the cases by obtaining data from a dilute solution of rhodamine B molecules. We finally apply this algorithm to data that simulate a highly heterogeneous specimen containing a large number of diffusing species. We compare the performance of MEMFCS versus that of a conventional fitting routine with a small number of diffusing components in the analysis of this simulated data.

THEORY

Modeling diffusion in FCS

The autocorrelation function $G(\tau)$ of the concentration $C(t)$ of solute molecules in a small open volume of a dilute solution is defined as

$$G(\tau) = \frac{\langle \delta C(t) \delta C(t+\tau) \rangle}{\langle C(t) \rangle^2}, \tag{1}$$

where angular brackets denote average over time t, and $\delta C(t) = C(t) - \langle C(t) \rangle$. In most FCS experiments the probe volume is approximately described as a three-dimensional Gaussian function with half axes r and l. It can be shown that for such a volume (Eigen and Rigler, 1994),

$$G(\tau) = \frac{1}{N}\left(\frac{1}{1+\frac{\tau}{\tau_D}}\right)\left(\frac{1}{1+\left(\frac{r}{l}\right)^2\frac{\tau}{\tau_D}}\right)^{1/2}, \tag{2}$$

where N is the number of particles in the volume (where the volume is defined in the sense of Mertz et al., 1995) and $\tau_D = r^2/4D$ is the time taken to diffuse through a distance r in two dimensions by a molecule with diffusion constant D. For a solution with n noninteracting fluorescent species with diffusion times τ_{Di}, $G(\tau)$ can be modeled as

$$G(\tau) = \sum_{i=1}^{n} b_i \left(\frac{1}{1+\frac{\tau}{\tau_{Di}}}\right)\left(\frac{1}{1+\left(\frac{r}{l}\right)^2\frac{\tau}{\tau_{Di}}}\right)^{1/2}, \tag{3}$$

where b_i are the relative amplitudes of the components. This amplitude is not simply proportional to the relative concentration of the individual species, but is also related to its relative brightness (Maiti et al., 1997). We note that Eq. 3 is valid under the assumption that the contribution of chemical kinetic processes (e.g., triplet state formation or protonation) can be neglected. In biological diffusion studies, the contribution of such processes can usually be either minimized or separated out in time, and thus this assumption is not too restrictive in practice.

MEM analysis of FCS data

In the model with a continuous distribution of diffusion times, modifying Eq. 3, $G(\tau)$ can be formally related to the diffusion time τ_D by the following relationship:

$$G(\tau) = \int_{\tau_D^L}^{\tau_D^U} a(\tau_D)\left(\frac{1}{1+\frac{\tau}{\tau_D}}\right)\left(\frac{1}{1+\left(\frac{r}{l}\right)^2\frac{\tau}{\tau_D}}\right)^{1/2} d\tau_D. \tag{4}$$

In Eq. 4, the diffusion time τ_D is considered to be a variable and $a(\tau_D)$ is the amplitude associated with τ_D. ${\tau_D}^L$ and ${\tau_D}^U$ are the lower and upper limits for diffusion times appropriate for the sample. The above equation for $G(\tau)$ is consistent with the usual definition of $G(\tau)$ for discrete τ_D values by defining $G(0)$ as in Eq. 5.

$$G(0) = \int a(\tau_D) d\tau_D. \tag{5}$$

The upper and lower limits of τ_D and quantitative implementation of the integral would depend upon prior knowledge of the sample. For example, if the range of τ_D values is approximately known, then a narrow linear distribution of τ_D may be preferred for fitting. However, in multi component samples it is reasonable to expect that the upper and lower limits of τ_D may differ by several orders of magnitude. Then it is desirable for computational purposes that Eq. 4 is considered as an integral over τ_D in logarithmic space. That is,

$$a(\tau_D) d\tau_D = \alpha(\tau_D) d(\ln \tau_D), \tag{6}$$

where $\alpha(\tau_D) = \tau_D a(\tau_D)$.

The distribution of diffusion times ($\alpha(\tau_D)$ vs. τ_D) is obtained by the Maximum Entropy Method. The method is based on the algorithm described by Skilling and Bryan (1984). The algorithm has been used to obtain a distribution of fluorescence lifetimes that fits the fluorescence decay data (Livesey and Brochon, 1987; Swaminathan and Periasamy, 1996) and a distribution of diffusion coefficients that fits fluorescence recovery after photobleaching data (Periasamy and Verkman, 1998). For computational purposes, the integral equation (Eq. 4) is written as a sum:

$$G(\tau) = \sum_{i=1}^{n} \alpha_i \left(\frac{1}{1+\frac{\tau}{\tau_{D_i}}}\right)\left(\frac{1}{1+\left(\frac{r}{l}\right)^2\frac{\tau}{\tau_{D_i}}}\right)^{1/2}. \tag{7}$$

The τ_{Di} in Eq. 7 are logarithmically spaced and not varied. The amplitude α_i have to satisfy the condition that the experimental data $G(\tau)$ is correctly fitted, i.e., the value of $G^c(\tau_i)$ calculated using Eq. 7 and the experimental value $G^e(\tau_i)$ are in agreement for all data. Because the experimental data are noisy, standard methods for evaluating the goodness of fit are preferred.

One of the standard methods is to calculate the weighted residual for each data and examine the residuals (r_i) qualitatively and quantitatively.

$$r_i = \frac{G^c(\tau_i) - G^e(\tau_i)}{\sigma_i} \tag{8}$$

σ_i is the inverse of weight for the i^{th} data. Qualitatively, a random distribution of residuals about the mean value of zero is a useful criterion of good fit even if the weights are not properly estimated. When weights are properly known then the quantitative parameter χ^2 is useful, where

$$\chi^2 = \frac{1}{M}\sum_{i=1}^{M} r_i^2. \quad (9)$$

In Eq. 9, M is the number of FCS data points. For a good fit, χ^2 is approximately equal to unity when M is sufficiently large.

It is often possible that the good fit criterion is satisfied for many different distributions of α_i, especially when the data is noisy. Such distributions may also include solutions for specific models, such as, one or more fixed value for τ_D. The experimental data is then consistent with any of these distributions and thus any model that predicts such a distribution is acceptable. According to the maximum entropy principle, the acceptable distribution is the one for which the value of entropy S is maximum. S is defined as

$$S = -\sum p_i \ln p_i, \quad (10)$$

where $p_i = \alpha_i/\sum\alpha_i$. According to this principle, a discrete solution for τ_D is the least acceptable solution for noisy data because S is the lowest for such a distribution. The widely used MEM algorithm (Skilling and Bryan, 1984) seeks a distribution for which S is maximum and χ^2 is minimum. The important features of the algorithm used in this paper for MEM analysis of FCS data are as follows.

The analysis begins with equal values for all α_i at all τ_{Di}. The distribution is improved in successive iterations: $\alpha_i(new) = \alpha_i(old) + x\Delta\alpha_i$. The correction factor $\Delta\alpha_i$ (an n-dimensional vector, see Eq. 7) is determined by the optimization procedure that uses three search directions constructed using the derivatives, $\nabla\chi^2$, ∇S and $\nabla\nabla\chi^2$. The procedure ensures that χ^2 is minimized in successive iterations and S is maximum for that χ^2. The multiplication factor x is determined by the 'α-chop and p-chop' technique (Skilling and Bryan, 1984) to achieve an aimed value of χ^2. Care is taken to avoid negative value for α_i, by using only a fraction of x and by equating negative values to zero. Successive iterations give distributions with reduced χ^2. The analysis is terminated when χ^2 does not change in successive iterations.

It has been shown (Narayan and Nityananda, 1986) that the Maximum Entropy Method works equally well for other definitions (Eqs. 11 and 12) of entropy, which are called regularization functions.

$$S_1 = -\sum \ln p_i \quad (11)$$

$$S_2 = -\sum\sqrt{p_i} \quad (12)$$

Maximizing these functions have the same effect as maximizing S as defined by Eq. 10. The usefulness of the above regularizing functions was also examined for FCS data analysis.

MATERIALS AND METHODS

The instrument

FCS measurements are performed with both single photon and two photon excitation. For single photon FCS, a green He:Ne laser (wavelength 543.5 nm, Jain Lasertech, Mumbai, India) is used as the light source. The laser beam is focused on the sample contained in a coverslip-bottomed petri dish using a high numerical aperture (NA), oil immersion microscope objective lens (1.3 NA, Carl Zeiss, Jena, Germany). The fluorescence from the sample is collected using the same objective lens and is separated from the excitation laser light using a dichroic mirror (560DCLP, Chroma Tech. Corp., Brattleboro, VT). This signal is then filtered with a band-pass filter (575DF30, Omega Optical, Brattleboro, VT) and focused onto a multimode fiber (50-μm core diameter from Newport, Irvine, CA) using a 15-cm achromat lens (Newport, Irvine, CA). The fiber is coupled to a single photon counting module (SPCM-AQ-140, EG&G, Vaudreuil, Canada), which detects the signal. The detector output is analyzed by a digital signal processing autocorrelator card (ALV5000e, ALV Laser GmbH, Langen, Germany) in a personal computer. For two photon FCS measurements, a Nd-vanadate (VERDI V10, Coherent, Santa Clara, CA) pumped femtosecond, tunable Ti:Sapphire laser (MIRA900, Coherent) is used as the light source. An inverted fluorescence microscope (TE300, Nikon, Tokyo, Japan) is modified to accommodate an external detector (PMT). The laser beam (893 nm, ~100-fs pulse width) is focused into the sample using a 60× water immersion microscope objective lens (1.2 NA, Nikon, Tokyo, Japan). The sample is kept in coverslip-bottomed petri dishes on the microscope sample stage. The fluorescence signal is collected by the same objective lens and EGFP fluorescence is selected using a dichroic mirror (535DCLP, Chroma Tech. Corp.) and a band-pass filter (500DF30, Omega Optical). A saturated $CuSO_4$ solution filter is used in front of the photomultiplier tube detector (Electron Tubes Ltd. Middlesex, UK) to block the infrared laser light. The data analysis is performed as before.

Materials

EGFP is purchased from Clontech, Palo Alto, CA. A 20-nM solution made in 20-mM phosphate buffer, pH 7.4, is used for the experiment. Buffer salts (from SD Fine Chemicals, Mumbai, India) are recrystallized twice to make them free of fluorescent impurities. Rhodamine B is purchased from Sigma-Aldrich Corp., St. Louis, MO. HPLC grade water (from E. Merck (India) Ltd., Mumbai, India) is used after distilling it twice for preparation of all solutions.

RESULTS

Verification of MEMFCS for a single diffusing species

Two-photon FCS experiments are performed with dilute solutions of EGFP in aqueous buffer. There is no photobleaching evident in the fluorescence intensity traces (data not shown). Two different sets of FCS data with different signal to noise (S/N) ratios are analyzed, with all data points assigned equal weights. For the data set with better S/N ratio (Fig. 1 *a*, *open circles*), a single component fit (Fig. 1 *a*, *solid line*) using a conventional fitting routine yields a value of 0.113 (±0.002) ms for the diffusion time τ_D (Fig. 1 *a* inset, *black solid line*; the width of the bar indicates the uncertainty). The data set is also fit with MEMFCS (not shown) using 101 components, logarithmically spaced between 10^{-3} ms to 100 ms, with 20 components per decade of time. The peak of the resulting distribution is at 0.117 ms (Fig. 1 *a inset*, *black dashed line*). The full width at half maximum (FWHM) of this distribution spans a range of 0.06–0.18 ms. When MEMFCS is used to fit the same data set with 101 components distributed linearly in the range of 0.09–0.15 ms (not shown), it produces a narrower distribution of τ_{Di} with a peak at 0.117 ms (Fig. 1 *a inset*, *gray shaded* region) and with a FWHM of 0.002 ms (range 0.116–0.118 ms). The residuals obtained for these three fits are shown in Fig. 1, *b–d*, respectively.

For the poor S/N data (Fig. 1 *e*, *open circles*), the conventional fit incorporating one component (Fig. 1 *e*, *solid line*) yields a peak at 0.112 ms and an uncertainty of ±0.004

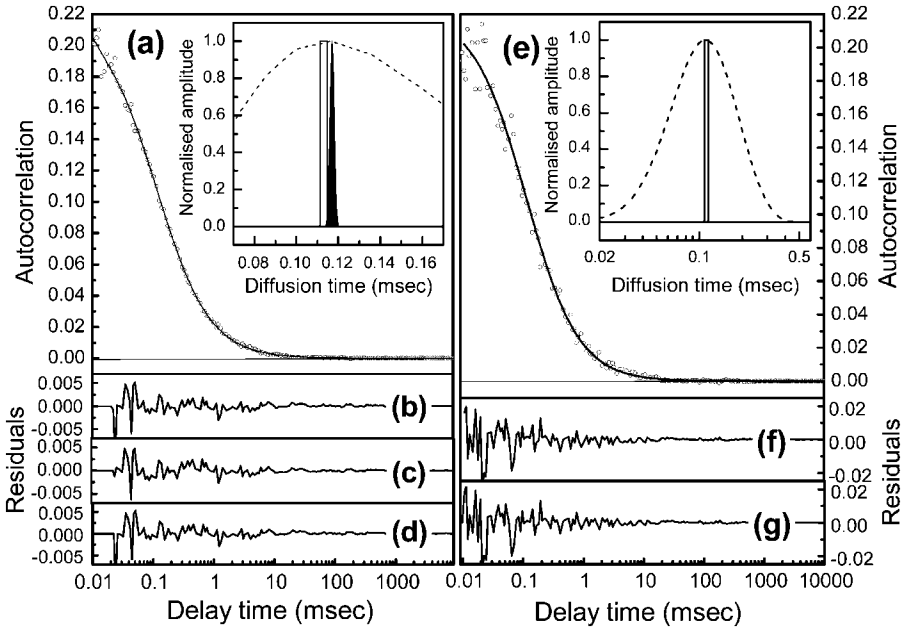

FIGURE 1 Comparison of MEMFCS and conventional fitting for a single component system. (*a*) Autocorrelation data obtained from EGFP in pH 7.4 buffer (*open circles*) with a good *S*/*N* and conventional fit with one diffusion time τ_D (*solid line*). Fit with MEMFCS with a logarithmic distribution of the τ_{Di} (not shown) and fit with MEMFCS with a narrow, linear distribution of the τ_{Di} (not shown) overlap with the conventional fit (*solid line*) except for initial values of delay time τ. Inset of (*a*) shows the distributions of τ_{Di} obtained with the three fitting methods: with single component conventional fit (*black solid line*), with MEMFCS with a logarithmic distribution of the τ_{Di} (*black dashed line*), and fit with MEMFCS with a narrow, linear distribution of the τ_{Di} (*gray shaded region*). For the single component conventional fit (*black solid line*), distribution width depicts the uncertainty of the determined value. (*b*), (*c*), and (*d*) show the residuals for the fits, respectively. (*e*) Data (*open circles*) with a lower *S*/*N* ratio and conventional fit to the data with one component (*solid line*). A MEMFCS fit with a logarithmic distribution of τ_{Di} overlaps this fit (not shown). Inset of (*e*) shows the distribution of τ_{Di} obtained with the two methods: single component conventional fit (*solid line*) and MEMFCS fit (*dashed line*). (*f*) and (*g*) show the residuals for the conventional fit of lower *S*/*N* data with one τ_D and MEMFCS fit with a logarithmic distribution of τ_{Di}, respectively.

ms (Fig. 1 *e inset*, *solid line* shows the distribution). MEMFCS fit with a logarithmic distribution (Fig. 1 *e* inset, *dashed line*) of τ_{Di} yields a distribution with a peak at 0.109 ms and a FWHM 0.144 ms. Corresponding residuals are shown in Fig. 1, *f* and *g*, respectively.

Estimating the goodness of fit in MEMFCS

Thirty separate FCS measurements are performed on a dilute aqueous solution (~4 nM) of rhodamine B with one photon excitation at 543 nm. Each measurement is an average of two autocorrelation traces, individually collected for 45 s. The normalized mean autocorrelation trace (Fig. 2 *a*, *open circles*) is fit as above, but the error information is now incorporated in the analysis. The weights ($1/\sigma_i$) (as defined in Eq. 8) are put into the data analysis in two different ways: i) σ_i are taken as the standard error of the mean calculated from the thirty individual measurements (Fig. 3, *filled circles*); and ii) σ_i set to a constant for all the data points. In the second case, any arbitrary value of σ_i could have been chosen. We choose the average error of $G(\tau)$ at long τ as constant σ_i. The χ^2 values obtained using these two types of error information are 0.92 and 10.2, respectively. In both cases, the FCS data fit well (Fig. 2 *a*, *solid line* and *dashed line*, respectively) and the distributions of τ_{Di} are nearly identical (curves not shown). However, it is observed that the residuals are more uniformly distributed in the former case (Fig. 2 *b*). In the latter case, absolute values for the residuals are larger at short τ compared to the values at long τ (Fig. 2 *c*).

Next, we attempt to estimate the goodness of fit for data obtained from single autocorrelation measurements. One set of data from the above mentioned 30 data sets (Fig. 3 *a*, *open circles*) is fit using MEMFCS. To estimate σ_i, we use the Koppel error expression as modified by Wohland et al. (2001) (Fig. 4, *dashed line*). We further modify it by

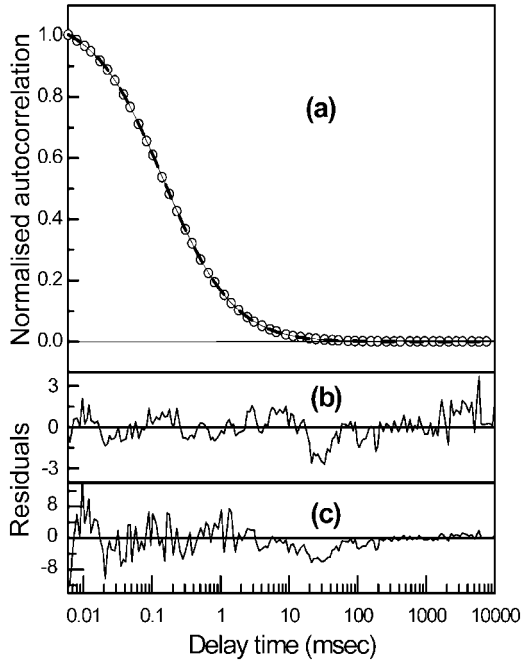

FIGURE 2 Estimating errors using multiple measurements. (*a*) Normalized mean autocorrelation function obtained from 30 repeated measurements of rhodamine B in water (*open circles*), MEMFCS fit using error estimated from the standard error of the mean (*solid line*), and MEMFCS fit assuming equal error (1×10^{-4}) at all delay points (*dashed line*). (*b*) and (*c*) show the residuals for these two fits respectively.

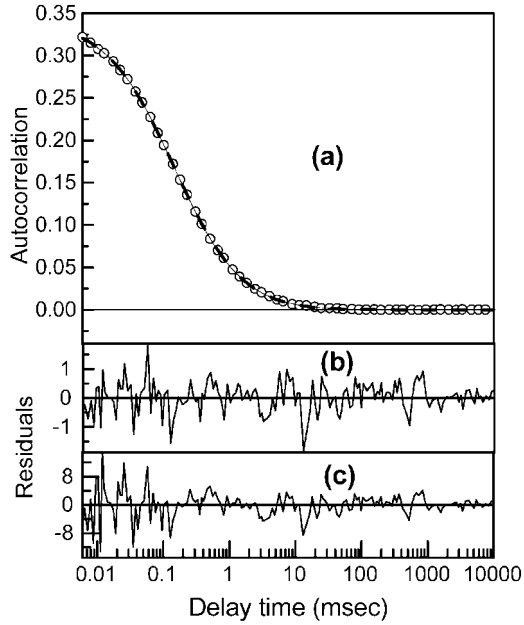

FIGURE 3 Error estimation for a single measurement. (*a*) Autocorrelation measured from rhodamine B in water (*open circles*), MEMFCS fit incorporating noise calculated from Koppel's equation modified as described in the text (*solid line*), and MEMFCS fit with equal noise (1×10^{-4}) at all points (*dashed line*). (*b*) and (*c*) show the residuals for these two cases, respectively.

neglecting the predicted rise in error at long delay times τ, holding errors to a constant value once it reaches a minimum (Fig. 4, *solid line*). The error measured from multiple data sets (Fig. 4, *filled circles*) fail to show the rise in errors at long delay times predicted by Koppel. For comparison, we also fit the data set with σ_i set to a constant. The FCS data are fit well by both measures of σ_i (*solid* and *dashed line,* respectively in Fig. 3 *a*) with near-identical distributions of τ_{Di} (data not shown). The residuals for these fits are shown in Fig. 3, *b* and *c*. The distribution of the residuals is more uniform in the former case. The χ^2 values obtained using these two types of error information are 0.23 and 11.2, respectively.

Comparison of results with different entropy definitions

We have analyzed the autocorrelation of rhodamine B in water (single measurement) using three different definitions of the entropy S (Eqs. 10–12). All other parameters are held constant for the analysis and σ_i estimated from modified Koppel's equation are used as the weights. The distributions obtained for all cases have nearly identical peak positions (Fig. 5). The distributions obtained with definition of S as in Eq. 11 (Fig. 5, *dotted line*) and Eq. 12 (Fig. 5, *dashed line*) tend to have sharper cutoff at either end of the peak, whereas that obtained with definition of S as in Eq. 10 (Fig. 5, *solid line*) has a rather smooth, Gaussianlike appearance. The

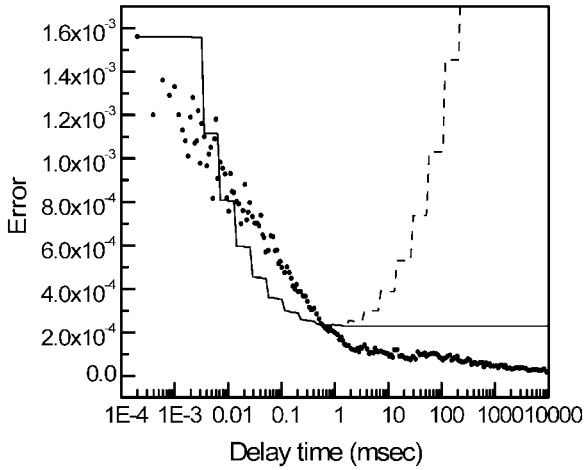

FIGURE 4 Experimental and analytical estimates of errors. Standard error of the mean calculated from 30 repeated measurements (*filled circles*), normalized error calculated from Koppel's equation (*dashed line*), normalized Koppel's error modified at longer delay times (*solid line*). Dashed and solid lines, by definition, overlap until the error reaches its minimum value.

weighted residuals (Fig. 5 *inset*, *a–c*, respectively) look nearly identical. The χ^2 values vary marginally.

Analysis of diffusion data from a simulated heterogeneous system

We simulate FCS data for a highly heterogeneous system diffusing in two dimensions (Fig. 6 *a*, *open circles*). The data is generated using the functional form:

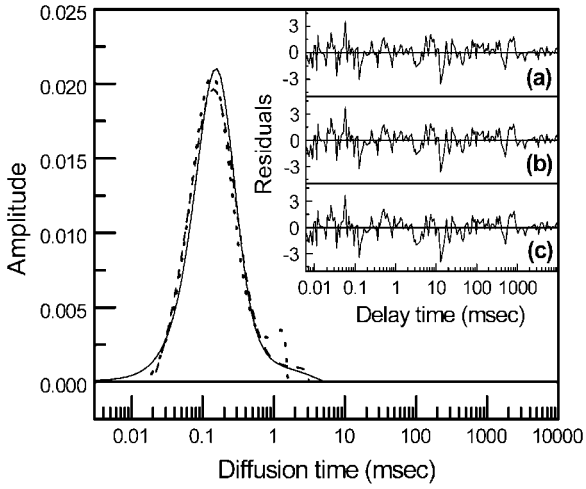

FIGURE 5 Comparison of different definitions of entropy S used for fitting the autocorrelation data obtained from 4 nM rhodamine B in water. Fits using $S = -\Sigma\, p_i \ln p_i$ (*solid line*), $S = -\Sigma\, (p_i)^{1/2}$ (*dashed line*), and $S = -\Sigma \ln p_i$ (*dotted line*). The insets (*a*), (*b*), and (*c*) show the MEMFCS fit residuals obtained from these fits respectively.

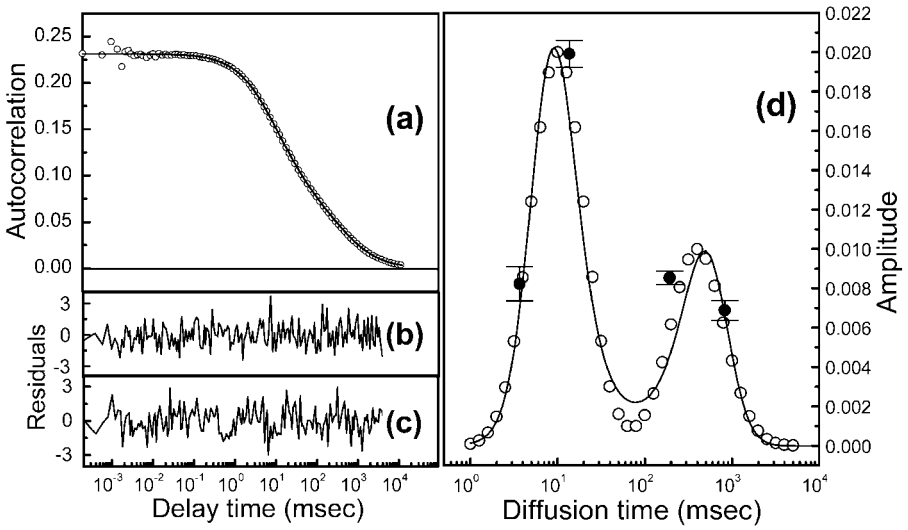

FIGURE 6 Comparison of MEMFCS and conventional fitting for simulated data representing a highly heterogeneous system. (*a*) Simulated data for a 41-component system (*open circles*) and MEMFCS fit (*solid line*) to this data. A conventional fit with four components overlaps this fit (not shown). (*b*) and (*c*) show the residuals for MEMFCS fit and conventional fit with four components, respectively. (*d*) The distributions of τ_{Di}: the input distribution used for simulation (*open circles*), distribution obtained from MEMFCS analysis (*solid line*), and parameters obtained from a four-component conventional fit (*filled circles*) with error bars.

$$G_{sim}(\tau) = G_{theoretical}(\tau) + G_{noise}(\tau) = \sum_{i=1}^{41} a_i \left(\frac{1}{1 + \frac{\tau}{\tau_{D_i}}} \right) + G_{noise}(\tau) \quad (13)$$

with a total of 41 τ_{Di} components logarithmically spaced from 1–5000 ms and a_i are chosen from the following distribution:

$$a_i(\tau_{D_i}) = A_1 \exp\left(-\left(\frac{\ln \tau_{D_i} - \ln \tau_1}{\sigma_1} \right)^2 \right) + A_2 \exp\left(-\left(\frac{\ln \tau_{D_i} - \ln \tau_2}{\sigma_2} \right)^2 \right) \quad (14)$$

with $A_1 = 0.02$, $A_2 = 0.01$, $\sigma_1 = \sigma_2 = 1.0$, $\tau_1 = 10$ ms, and $\tau_2 = 400$ ms.

$G_{noise}(\tau)$ is calculated at each τ by generating a random number from a Gaussian distribution with its mean equal to $G_{theoretical}(\tau)$ and standard deviation (*SD*) proportional to the modified Koppel error value. The magnitude of *SD* at $\tau = 0$ is set at a fraction f of $G_{theoretical}$ at time 0, i.e.,

$$SD(\tau = 0) = f \times G_{theoretical}(0) = f \times \sum_{i=1}^{41} a_i. \quad (15)$$

For this simulation $f = 0.04$. $G_{sim}(\tau)$ (Fig. 6 *a*, *open circles*) is then analyzed with MEMFCS, using 101 diffusing components with τ_{Di} distributed logarithmically between 0.1 ms and 10 s (a rather different distribution from the input τ_{Di} of Eq. 13) and α_i set to a constant value for all τ_{Di} initially. The fit is shown in Fig. 6 *a* (*solid line*). The input (*open circles*) and the obtained distribution (*solid line*) are shown in Fig. 6 *d*. A conventional fit with four components also provides an excellent fit to the data (fit not shown), with component values (Fig. 6 *d*, *filled circles* with *error bars*) near the two peaks of the input distribution. The residuals obtained with MEMFCS fit and the conventional fit with four components are shown in Fig. 6, *b* and *c*, respectively. A conventional fit with six components forces two components to converge with very large values of uncertainty. (Best fit parameters obtained with a conventional fit incorporating six components: $\tau_{D1} = 1129 \pm 196$ ms, $A_1 = 0.019 \pm 0.007$; $\tau_{D2} = 330 \pm 63$ ms, $A_2 = 0.053 \pm 0.004$; $\tau_{D3} = 1.17 \pm 1.1$ ms, $A_3 = 0.006 \pm 0.01$; $\tau_{D4} = 29 \pm 20$ ms, $A_4 = 0.05 \pm 0.05$; $\tau_{D5} = 7 \pm 2084$ ms, $A_5 = 0.05 \pm 358$; and $\tau_{D6} = 7 \pm 2087$ ms, $A_6 = 0.05 \pm 358$). These values are inconsistent with the input distribution profile and crowd near the peaks of the input distribution.

DISCUSSION

The conventional fitting routine with a single diffusing component and MEMFCS fitting with 101 components yield very similar diffusion times (considering the peak values) for EGFP in dilute solution. This verifies that in the limit where a conventional single component fit is expected to provide a correct description, the MEMFCS fitting routine agrees with it. Although the width of the MEMFCS distribution and the uncertainty in τ_D from a conventional single component fit are not exactly equivalent quantities, it is interesting to compare the two. For MEMFCS to provide an unbiased analysis, the width of the obtained distribution must be ultimately dictated by the inherent noise in the data. An essential feature of MEMFCS is that the user at the initiation of the program fixes the value of the component τ_{Di} and only the amplitude α_i can vary. Consequently, for a wide range of input distribution spanning several decades, adequate density of components may not be available at the precise value of τ_D that corresponds to the actual diffusion constant of the diffusing species. This is the case for the logarithmic distribution of τ_{Di} initially used to fit the data (Fig. 1 *a inset*, *black dashed line*). This fit, thus, yields a wide distribution

of τ_{Di}. However, when there exists a priori knowledge that there is a single component, as in the present case, MEMFCS can be used in a second set of iterations with a linear distribution of τ_{Di} closely spaced around the peak value of the distribution. This yields a narrow distribution (Fig. 1 *a inset, gray shaded region*) whose width is comparable to the uncertainty (shown as the *width* of the *black bar* in Fig. 1 *a inset*) of the conventional single component fit.

However, when the data has poor *S*/*N*, MEMFCS yields a much wider distribution (in this case, because the spacing between the components is already small compared to the overall width of the distribution, MEMFCS fit with a linearly spaced τ_{Di} was not attempted). This width is inherent in the data due to the noise present in it. At this stage, the sums of square deviations of the two fits from the data are nearly identical (4.81×10^{-3} for the single component conventional fit and 4.84×10^{-3} for MEMFCS). This implies that both the descriptions fit the data equally well. This indicates that although a priori knowledge of the solution may imply a single peak, the noisy data is only good enough to limit inference to the width obtained. This provides a safe limit for interpreting data from specimens about which such a priori knowledge may not exist.

To have a check on the goodness of fit, it is desirable to compute the reduced χ^2 value for the fit. We have obtained an experimental measure of the weight by repeating the FCS measurement 30 times on a dilute aqueous solution of rhodamine B and taking the standard error of the mean (following Wohland et al. (2001)) of these measurements, for each time point. The reduced χ^2 value calculated from the fit is close to unity (0.92), indicating that the obtained distribution does provide a good description of the data.

For estimating errors in single data sets, we have used the analytical formulation of errors provided by Koppel and modified by us as described in this paper (Fig. 4, *solid line*). Uniform distribution of weighted residuals for all the data points suggests that this approach gives an acceptable estimate of relative errors for all delay times. However, a low value of reduced χ^2 ($\chi^2 = 0.23$) indicates that the error is still overestimated by a factor of $\sim$2. On the other hand, providing equal weights to all points yields a nonuniform distribution of residuals. A satisfactory analysis of data must yield uniformly distributed weighted residuals and hence we recommend Koppel's equation, truncated at its minimum value and then held constant for longer τ values, as the best option to assign weights to the data points for single FCS data sets.

We have compared the effect of using different definitions of entropy *S* in MEMFCS fitting. All the three definitions used yield very similar fits and residuals (Fig. 5 *inset*, *a–c*), and any of them can be used in practice. However, the definition provided by Eq. 10 yields a distribution that is smooth throughout (Fig. 5, *solid line*), and we prefer to use this definition.

The critical test of the usefulness of the MEMFCS routine comes from an analysis of the simulated data that represents a highly heterogeneous system. Such results are expected while investigating membrane protein diffusion or aggregating protein solutions. For the simulated data, both MEMFCS and a conventional fitting routine incorporating four diffusing species fit the data well. The χ^2 values of the fits are nearly identical at this stage (1.24 for four component conventional fit and 1.27 for MEMFCS). The fit parameters obtained with the conventional fit with four components, however, do not represent the input distribution well. Even an increase in the number of components to six in the conventional fitting routine does not yield a better distribution. An attempt to fit the data with six components merges two components near the primary peak of the distribution, and results in extremely large uncertainties for the two extra components. On the other hand, the entropy maximization inherent in MEMFCS ensures that the wide input distribution of τ_{Di} is well-represented by this analysis.

It is evident that a conventional fit with a small number of components does not yield a reliable description of a highly heterogeneous system. MEMFCS on the other hand represents components with a wide range of amplitudes rather well. Thus, if there are reasons to believe that the specimen under investigation contains a distribution of different species, it is essential to use an analysis algorithm, such as MEMFCS, that can not only represent a continuous distribution of species, but also avoids an unnecessarily narrow interpretation of the data.

This work was supported by a Wellcome Trust Senior Overseas Research Fellowship in Biomedical Sciences in India to S.M. (05995/Z/99/Z/HH/KO). P.S. and J.B. acknowledge partial support from the Kanwal Rekhi Scholarship of the Tata Institute of Fundamental Research Endowment Fund.

REFERENCES

Berland, K. M., P. T. So, and E. Gratton. 1995. Two-photon fluorescence correlation spectroscopy: method and application to the intracellular environment. *Biophys. J.* 68:694–701.

Eigen, M., and R. Rigler. 1994. Sorting single molecules: application to diagnostics and evolutionary biotechnology. *Proc. Natl. Acad. Sci. USA.* 91:5740–5747.

Elson, E., and D. Magde. 1974. Fluorescence correlation spectroscopy. I. Conceptual basics and theory. *Biopolymers.* 13:1–27.

Haupts, U., S. Maiti, P. Schwille, and W. W. Webb. 1998. Dynamics of fluorescence fluctuations in green fluorescenct protein observed by fluorescence correlation spectroscopy. *Proc. Natl. Acad. Sci. USA.* 95:13573–13578.

Koppel, D. E. 1974. Statistical accuracy in fluorescence correlation spectroscopy. *Phys. Rev. A.* 10:1938–1945.

Kummer, S., P. Schwille, A. A. Heikal, W. E. Moerner, and W. W. Webb. 2000. Fluorescence correlation spectroscopy reveals fast optical excitation-driven intramolecular dynamics of yellow fluorescent proteins. *Proc. Natl. Acad. Sci. USA.* 97:151–156.

Livesey, A. K., and J. C. Brochon. 1987. Analyzing the distribution of decay constants in pulse fluorimetry using the maximum entropy method. *Biophys. J.* 52:693–706.

Magde, D., E. Elson, and W. W. Webb. 1972. Thermodynamic fluctuation in a reaction system - measurement by fluorescence correlation spectroscopy. *Phys. Rev. Lett.* 29:705–708.

Maiti, S., U. Haupts, and W. W. Webb. 1997. Fluorescence correlation spectroscopy: diagnostics for sparse molecules. *Proc. Natl. Acad. Sci. USA*. 94:11753–11757.

Mertz, J., C. Xu, and W. W. Webb. 1995. Single-molecule detection by two-photon-excited-fluorescence. *Opt. Lett*. 20:2532–2534.

Narayan, R., and R. Nityananda. 1986. Maximum entropy image restoration in astronomy. *Ann. Rev. Astron. Astrophys*. 24:127–170.

Olsson, M., A. Pramanik, U. Langel, T. Bartfai, and R. Rigler. 2001. Fluorescence correlation spectroscopy detects galanin receptor diversity on insulinoma cells. *Biochemistry*. 40:10839–10845.

Periasamy, N., and A. S. Verkman. 1998. Analysis of fluorophore diffusion by continuous distribution of diffusion coefficients: application to photobleaching measurements of multicomponent and anomalous diffusion. *Biophys. J*. 75:557–567.

Post, K., M. Pitschke, O. Schafer, H. Wille, T. R. Appel, D. Kirsch, I. Mehlhorn, H. Serban, S. B. Prusiner, and D. Riesner. 1998. Rapid acquisition of β-sheet structure in the prion protein prior to multimer formation. *Biol. Chem*. 379:1307–1317.

Schwille, P., U. Haupts, S. Maiti, and W. W. Webb. 1999. Molecular dynamics in living cells observed by fluorescence correlation spectroscopy with one- and two-photon excitation. *Biophys. J*. 77:2251–2265.

Schwille, P., F. Oehlenschlager, and M. Eigen. 1996. Detection of HIV-1 RNA by nucleic acid sequence-based amplification combined with fluorescence correlation spectroscopy. *Proc. Natl. Acad. Sci. USA*. 93:12811–12816.

Sengupta, P., K. Garai, D. J. Callaway, and S. Maiti. 2002. Kinetics of aggregation of Amyloid beta peptide. *Biophys. J*. 82:505a (Abstr.)

Skilling, J., and R. K. Bryan. 1984. Maximum entropy image reconstruction: general algorithm. *Mon. Not. R. Astr. Soc*. 211: 111–124.

Swaminathan, R., and N. Periasamy. 1996. Analysis of fluorescence decay by the maximum entropy method: influence of noise and analysis parameters on the width of the distribution of lifetime. *Proc. Indian. Acad. Sci. (Chem. Sci.)*. 108:39–49.

Thompson, N. 1991. Fluorescence correlation spectroscopy. *In* Topics in Fluorescence Spectroscopy. J. R. Lakowicz, editor. Plenum Press, New York. 337–378.

Tjernberg, L. O., A. Pramanik, S. Bjorling, P. Thyberg, J. Thyberg, C. Nordstedt, K. D. Berndt, L. Terenius, and R. Rigler. 1999. Amyloid beta-peptide polymerization studied using fluorescence correlation spectroscopy. *Chem. Biol*. 6:53–62.

Wohland, T., R. Rigler, and H. Vogel. 2001. The standard deviation in fluorescence correlation spectroscopy. *Biophys. J*. 80:2987–2999.