

# Bayesian-Inference-Based Fluorescence Correlation Spectroscopy and Single-Molecule Burst Analysis Reveal the Influence of Dye Selection on DNA Hairpin Dynamics

Wolfgang Kügel,<sup>[a]</sup> Adam Muschielok,<sup>[a]</sup> and Jens Michaelis<sup>\*[a, b, c]</sup>

Fluorescence correlation spectroscopy (FCS) is a powerful tool to gain information about dynamics of biomolecules. However, the key problem is to extract the rates hidden in the FCS data by fitting the data to a meaningful model. A number of different fitting approaches have been described in recent years but the extraction of relevant information to date has still been limited by numerous experimental problems and the fact that the set of starting parameter values chosen could often predefine the result. We establish a new way to globally analyze FCS data based on Bayesian inference to overcome these issues. Moreover, the influence of other remaining experimental error

sources, for example, photophysics, is excluded by additional means. Using this approach in combination with the results from single-molecule burst analysis, we investigate the kinetics of DNA hairpins labeled with a variety of different fluorescent probes as a function of the salt concentration. We find that the rates of hairpin opening and closing as well as the equilibrium constant of the transition depend on the characteristics of the dye molecules used to label the hairpin. Thus, great caution has to be used when utilizing dye molecules as reporters for the kinetics of dynamic macromolecular structures.

## 1. Introduction

Throughout the last years, with the growing interest in the quantitative understanding of dynamic processes in living systems, also the studies of conformational dynamics of biological molecules have seen an increasing attention. Prominent examples include investigations of the dynamics of DNA hairpins,<sup>[1–6]</sup> RNA secondary structures,<sup>[7–12]</sup> nucleosomes,<sup>[13–16]</sup> motor proteins,<sup>[17]</sup> DNA translocases<sup>[18]</sup> as well as a large variety of other proteins.<sup>[19–26]</sup>

To study conformational dynamics, oftentimes a pair of local probes is attached to the molecule of interest, in particular a pair consisting of a dye molecule and a quencher. The term “excited-state quencher” is commonly applied to a molecule that accepts energy from an excited dye. Quenchers are either non-fluorescent<sup>[1,3]</sup> or fluoresce themselves.<sup>[2,16,27]</sup> Dynamic processes which bring about changes in the distance between the two probes cause fluctuations in the brightness of the emitted fluorescent light. Brightness fluctuations of the quenched probe can be monitored for both fluorescent and non-fluorescent quenchers while additional brightness fluctuations of the quencher are observed only if a fluorescent acceptor is used.<sup>[2]</sup>

There is a large variety of different experimental techniques which make use of the attachment of local probes for the investigation of dynamics of a biomolecule, including proximity ratio auto-correlation,<sup>[28,29]</sup> melting temperature analysis,<sup>[1,6]</sup> photon counting histogram<sup>[3,4,6]</sup> and FCS;<sup>[1–6]</sup> however, they all suffer from the same experimental complication. Does the attachment of probes reporting about the conformational dynamics lead to changes in the underlying energy landscape? In particular for investigations of DNA or RNA molecules, where single nucleotide changes are known to influence the dynamics,<sup>[1]</sup> extreme caution has to be used. While the influ-

ence of different dye–quencher pairs on the melting temperature of DNA hybrids has been investigated previously,<sup>[30]</sup> little is known about their influence on the steady-state distribution and on the transition rates of dynamic processes.

Amongst the different biomolecules that have been investigated in dynamic studies, nucleic-acid hairpins have attracted large interest, since understanding their dynamics is required for understanding important biological processes, such as the dynamic nature of four-way junctions or processes driven by ribozymes.<sup>[31,32]</sup> Moreover, commercially available single DNA or RNA strands capable of forming hairpin structures are readily available with a large variety of different labels and are thus a good model system to study dynamics. Previous studies have observed repeated opening and closing of hairpins with rates depending on the sequence, stem length and salt concentration.<sup>[1,3–6,11,12,33,34]</sup> For the proper interpretation of such experiments it is important to determine the influence of the attached probes on these transitions.

[a] W. Kügel, Dr. A. Muschielok, Prof. Dr. J. Michaelis  
Department of Chemistry  
Ludwig-Maximilians-University Munich  
81377 Munich (Germany)  
Fax: (+49) 73-150-23059  
E-mail: jens.michaelis@uni-ulm.de

[b] Prof. Dr. J. Michaelis  
Physics Department, Ulm University  
89081 Ulm (Germany)

[c] Prof. Dr. J. Michaelis  
Center for Integrated Protein Science Munich (CIPSM)  
81377 Munich (Germany)

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/cphc.201100720>.

A common technique for studying dynamic processes from the sub-microseconds to the hundreds-of-milliseconds time-scale is FCS.<sup>[1–4,14–16]</sup> In FCS, the computed correlation functions contain information about all dynamic processes, in particular diffusion, photophysics as well as conformational dynamics. It is therefore important that the respective contributions can be separated from each other. Unfortunately, diffusion and kinetics oftentimes occur on similar timescales, making it difficult to extract the parameter of interest with high accuracy.

Conformational kinetics and diffusion can be distinguished by comparing measurements of double-labeled samples (with a quencher molecule and a corresponding quencher) to those with only a single label (the quencher molecule of the double-labeled samples).<sup>[1,14]</sup> Only the double-labeled sample is sensitive to dye–quencher distance changes, while all other fluctuations are common to both samples. Thus, by dividing the auto-correlation function of the double-labeled sample by that of the single-labeled one, one obtains the desired information about the conformational dynamics.<sup>[1]</sup>

A drawback of this method is that other dynamic processes are removed entirely only if the concentration of molecules as well as the excitation and detection volumes are identical for both measurements. In contrast, identical concentrations are intrinsic when performing pulsed interleaved excitation (PIE)<sup>[35]</sup> experiments which is also known as nanosecond alternating laser excitation (nsALEX).<sup>[2,35–37]</sup> In PIE, one directly excites the quencher and calculates its auto-correlation function. By dividing the donor auto-correlation function containing the conformational dynamics signal by the auto-correlation function of the acceptor after direct excitation, diffusion kinetics can be removed. There are, however, other challenges that remain such as intrinsic photo physics of the labels (causing a “false” kinetic signal) and the requirement that the confocal volumes of the two different colors need to be identical.

While FCS is an excellent tool for the investigation of transition rates, the information about the underlying state distributions is very limited. Single-molecule burst analysis,<sup>[37–39]</sup> in contrast, allows for a direct measurement of the FRET states and thus gives access to intra-molecular distances between donor and quencher. Thus, it has become a standard method for investigations of distributions of conformational states of fluorescently labeled biomolecules.

Herein, we present confocal microscopy of DNA hairpins in combination with FCS<sup>[40–43]</sup> and burst analysis<sup>[37–39]</sup> techniques and investigate the influence of different dye molecules on the distribution of states and rate constants. We present a novel approach for extracting kinetic rates by globally analyzing a set of five correlation curves including a separate donor-only measurement using Bayesian inference.

## 2. Results

### 2.1. FCS Global Analysis

DNA hairpins consisting of a 21-base loop and a five-base-pair stem were investigated in a custom-built PIE–FCS setup. The dye labels were positioned on opposite ends of the DNA con-

struct allowing for a maximum FRET efficiency change upon hairpin opening and closing (see Methods section for details).

We globally analyzed the auto-correlation function of the green photon stream of a donor-only sample after donor excitation (*GG*) together with the correlation functions obtained from a double-labeled sample, namely, the auto-correlation of the donor photons after donor excitation (*GG*), acceptor photons after acceptor excitation (*RR*), acceptor photons after donor excitation (*FF*) and the cross-correlation of donor and acceptor photons after donor excitation (*FG*) using Bayesian parameter estimation (Figure 1 B–F).

All correlation functions have the general form [Eq. (1)]:

$$G_{xy}(\tau) = M_{xy}(\tau)X_{xy}(\tau) \quad (1)$$

where the indices *x* and *y* depict the photon streams that are correlated,  $X_{xy}$  describes the conformational kinetics of interest and  $M_{xy}$  is the diffusion term given by Equation (2):<sup>[44]</sup>

$$M_{xy}(\tau) = \frac{\alpha}{\langle N_{xy} \rangle} \left( \frac{1}{1 + \frac{4D\tau}{\omega_{xy}^2}} \right) \left( \frac{1}{1 + \frac{4D\tau}{\omega_{zy}^2}} \right)^{\frac{1}{2}} \quad (2)$$

Here,  $\langle N_{xy} \rangle$  is the average number of observed molecules in the focal volume,  $D$  denotes the diffusion constant,  $\omega_{xy}$  and  $\omega_{zy}$  are the lateral and axial dimensions of the focal volume and  $\alpha$  is a correction factor depending on the focal volume geometry (here  $\alpha = 2^{-3/2}$  since a three-dimensional Gaussian is assumed).

A two-state model as depicted in Figure 1 A was used to describe the conformational dynamics of the DNA hairpin. The transition between states of FRET efficiency  $E_1$  and  $E_2$  lead to a kinetic contribution to the FCS signal described by Equations (3), (4):<sup>[2]</sup>

$$X_{GG}(\tau) = 1 + A_{GG}(\tau) = 1 + \frac{k_{21}k_{12}(E_1 - E_2)^2}{[k_{21}(1 - E_1) + k_{12}(1 - E_2)]^2} e^{-(k_{12}+k_{21})\tau} \quad (3)$$

$$X_{FF}(\tau) = 1 + A_{FF}(\tau) = 1 + \frac{k_{21}k_{12}(E_1 - E_2)^2}{(k_{21}E_1 + k_{12}E_2)^2} e^{-(k_{12}+k_{21})\tau} \quad (4)$$

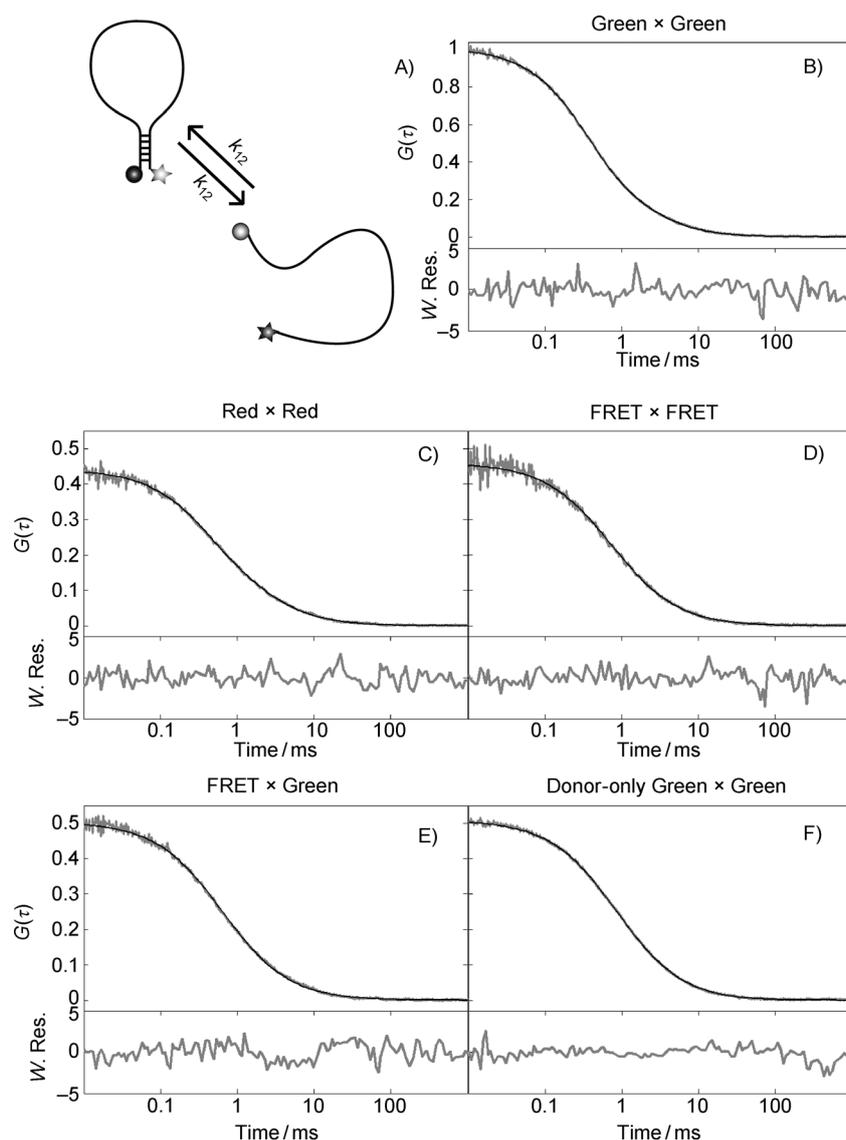
and (5):

$$X_{FG}(\tau) = 1 - A_{FG}(\tau) = 1 - \frac{k_{21}k_{12}(E_1 - E_2)^2}{[k_{21}(1 - E_1) + k_{12}(1 - E_2)](k_{21}E_1 + k_{12}E_2)} e^{-(k_{12}+k_{21})\tau} \quad (5)$$

where  $k_{12}$  and  $k_{21}$  are the transition rates between the two conformational states and  $A_{XY}$  is a measure for the actual conformational kinetic signal amplitude.

In contrast, the auto-correlation function of photons emitted by the acceptor dye after direct excitation (*RR*) is insensitive to the hairpin dynamics and can be described by Equation (6):

$$X_{RR}(\tau) = 1 + \frac{f_{isc}}{1 - f_{isc}} e^{-\frac{\tau}{\tau_{isc}}} \quad (6)$$



**Figure 1.** Hairpin-opening dynamics identified by global analysis of correlation functions: A) Schematic drawing of the two-state model used to describe the dynamics observed in the five correlation functions that were used for a global analysis of the recorded data (B–F). The forward and backward rates are denoted as  $k_{12}$  and  $k_{21}$ , the red and green dye positions are depicted by a star and a circle. Exemplary set of auto-correlation functions (B: donor–donor, C: acceptor–acceptor, D: FRET–FRET, and F: donor-only sample) as well as cross-correlation (E) data for DNA hairpins labeled with Alexa532–Atto647N containing 10 mM NaCl measured at 16  $\mu$ W laser power (grey). The corresponding results of the global analysis are shown in black and the weighted residuals [W. Res. (a.u.)] are given below each correlation curve.

with  $f_{isc}$  being the probability for a molecule to enter the triplet state upon excitation and  $\tau_{isc}$  describing the triplet lifetime. To minimize computational expense, we did not fit an additional triplet term to the  $GG$ ,  $FF$  and  $GGd$  correlation curves; instead, we limited the experimental data to a time range not showing donor photophysics. Since the donor-only labeled samples are insensitive to the conformational dynamics (assuming no changes in brightness), there is no kinetic term in these data sets [i.e.  $X_{GGd}(\tau) = 1$ ].

To globally fit all the correlation functions [Eqs. (1)–(6)] to the experimental data, 20 model parameters have to be determined (a summary of the variables and their status in the

global analysis is given in Scheme 1). Commonly, one aims at finding the global optimum of an objective function, which is typically the  $\chi^2$  function in a least-squares fit or the likelihood function in maximum likelihood estimation (MLE).<sup>[51,52]</sup> The confidence bounds of the estimate are calculated from a local property of the objective function evaluated at the found optimum, namely, the second derivative matrix of  $\chi^2$  or the logarithm of the likelihood, respectively. While this approach is usually reasonable for a few model parameters, it can become inappropriate if the data is described by a more complex model with considerably more, possibly interdependent, parameters. In this case, the local approximation of the objective function used to compute the errors of the estimate breaks down, rendering the determined error bars inappropriate. Indications of such failure are slow convergence of the optimization and several equally “good” optima found in separate optimization runs started at different initial positions in the parameter space, both of which were observed for our data set. A solution to these problems is Bayesian inference, a general data analysis method (which contains least squares and MLE as special cases).<sup>[44,49,53,54]</sup> Bayesian inference does not rely on the optimization and local approximation of an objective function. Furthermore, it allows to directly

include “prior information” in the analysis, for example the result of calibration experiments, simplifying error propagation.

Therefore, we used Bayesian inference (see Methods section) to globally determine all the parameters and experimental uncertainties. In brief, to calculate the likelihood of the data we assumed uncorrelated experimental noise on top of the theoretical autocorrelation functions [Eqs. (1)–(6)]. We assumed normally distributed noise with zero mean and a known, time-dependent amplitude, so that the likelihood is proportional to  $\exp(-\chi^2/2)$ , where  $\chi^2$  is the weighted sum of the squared residuals. An example for the result of a global Bayesian analysis of the  $GG$ ,  $RR$ ,  $FF$ ,  $FG$  and  $GGd$  correlation functions is shown in

Parameter	FF	GF	GG	GGd	RR
$N$					
$D$					
$\omega_R$					
$\omega_z$					
$\tau_{isc}$					
$f_{isc}$					
$k_{12}$					
$k_{21}$					
$E_1$					
$E_2$					

**Scheme 1.** Overview of the parameters used in the Bayesian analysis of the FCS data: Parameter overview indicating globally optimized variables in black and individually optimized variables in grey. The white fields indicate parameters not included in the analysis of the respective correlation function.

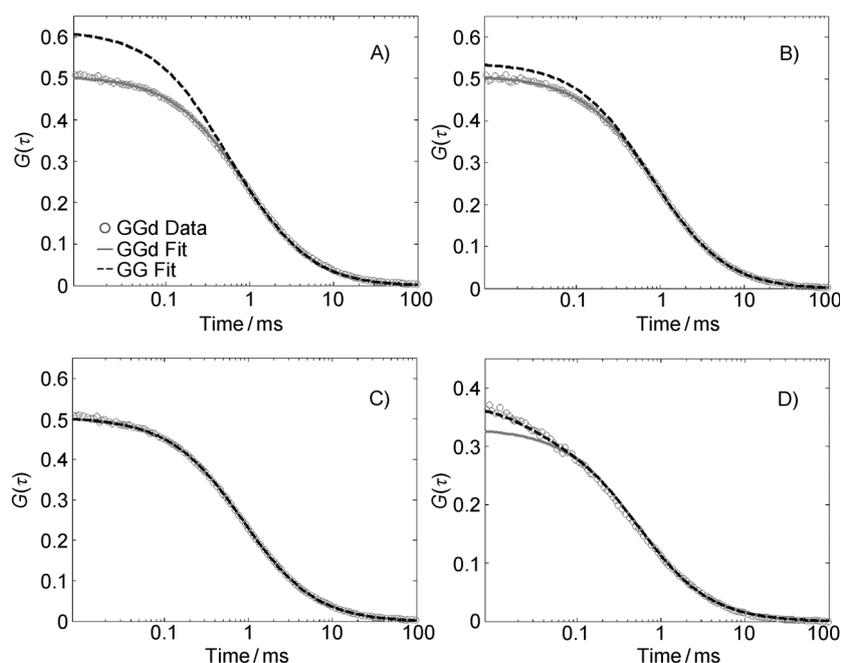
Figure 1B–F. Importantly, the residuals for all correlation functions show no apparent correlation.

As our model does not account for triplet kinetics of the donor, which could be mistaken for conformational kinetics of the hairpin, we compared the data of the donor-only hairpin ( $GGd$ ), its fit function  $G_{GGd}(\tau)$  and the fit function  $G_{GG}(\tau)$  of the FRET sample (Figure 2). Since the concentration for single- and double-labeled measurements is not required to be identical, the number of molecules in the confocal volume  $\langle N_{yy} \rangle$  of the FRET sample was set to the respective value of the donor only sample for this overlay. Examples of a successful analysis of samples not affected by donor triplet kinetics are given in Figures 2A,B, where a good fit of both correlation curves was obtained and a clear difference between donor-only (grey) and FRET (dashed black) maximum posterior results can be observed. Here, the donor-only data (grey circles) can be described by the simple diffusion model, while a clearly visible kinetic contribution is contained in the FRET result. For different datasets (Figure 2C), the data from the donor-only and the FRET sample are indistinguishable over the complete accessible correlation time range. Thus, there is no apparent signal of conformational dynamics in the  $GG$  correlation function. A

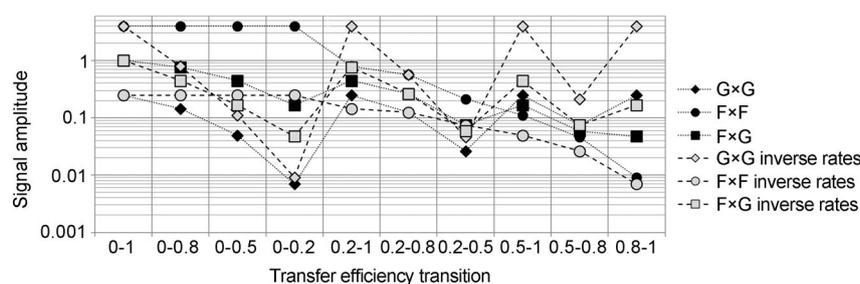
completely different behavior can be seen in Figure 2D. In this example, using a Cy3–Cy5 dye pair, the global analysis yields parameters that cannot describe the fast dynamics observed in the donor-only sample which instead is described by the kinetic parameters obtained in the global analysis for the FRET sample. This can be understood if photophysical effects of the donor have a stronger effect on the correlation function than the conformational dynamics. In such cases, during the global analysis, the kinetic term in Equation (3) is misused to describe the photophysical signal instead of the conformational dynamics. Moreover, Cy3 is known to be sensitive to the local environment, resulting in lifetime and brightness variations.<sup>[30,45]</sup> Therefore, the Cy3–Cy5 dye pair was excluded from further FCS Analysis.

It is important to note that not all correlation functions are equally sensitive to kinetics of a particular FRET transition (Figure 3). Thus, by globally analyzing all correlation functions that contain a kinetic term, one avoids missing apparent transitions, an important advantage of the global analysis.

By performing a global analysis using Bayesian inference, all necessary parameters can be determined even without in-detail prior knowledge; however, some of the parameters are

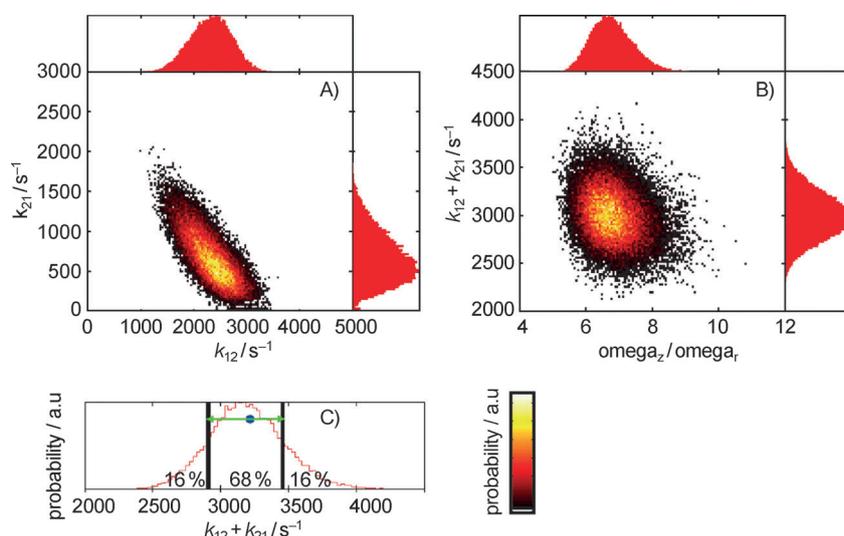


**Figure 2.** Comparison of conformational kinetics and photophysics. The experimental auto-correlation for the: A) Atto 532 and B) Alexa 532 donor-only sample  $G_{GGd}$  is shown (grey circles) together with the analysis results calculated using Equation (2) (solid grey line). Overlaid are the kinetic results obtained from the global analysis of the A) Atto532–Atto647N and B) Alexa532–Alexa647-labeled hairpins (dashed black line). All four samples were measured at 10 mM NaCl and 16  $\mu$ W laser power. For the overlay, the number of molecules in the confocal volume ( $N$ ) of the FRET sample was set to the respective value of the donor-only sample, since concentrations of single- and double-labeled measurements were not precisely identical. For short times, one clearly observes a difference between donor-only and FRET results, indicating the presence of additional kinetics. C) Atto532–Alexa647 hairpin sample (measured under the same conditions) showing a different behavior. Here, the results from the donor-only and the FRET sample are indistinguishable over the complete accessible time range. Thus, there is no apparent signal of conformational dynamics. D) The Cy3–Cy5 data measured at 100 mM NaCl and 25  $\mu$ W laser power shows a situation where the global analysis yields parameters that cannot describe the short-time behavior observed in the donor-only sample. In contrast, the kinetic parameters obtained in the global analysis for the FRET sample describe well the observed signal of the donor-only sample (again using  $N$  from the donor-only sample).



**Figure 3.** Comparison of expected kinetic signal amplitudes for the three available channels as a function of the two FRET states and rates. Simulated amplitudes  $A$  [Eq. (3)–(5)] of the three correlations containing a kinetic term, namely, the donor (GG) and acceptor (FF) auto-correlations and the cross-correlation (FG) calculated for the time point  $\tau = 1e^{-8}$  s, which is chosen to be significantly smaller than the timescale of molecular dynamics encoded in the correlation functions. Results are shown for different FRET transitions at rates of  $k_{12} = 4000$  and  $k_{21} = 1000$  s $^{-1}$ , and vice versa (inverse rates).

correlated and thus there is not enough information in the recorded data to accurately determine these parameters individually, most notably the rates  $k_{12}$  and  $k_{21}$  (Figure 4A). Here, the uncertainties in the amplitude of the experimental correlation function caused by for example, poor statistics make it impossible to resolve parameters relying on small changes in this amplitude (this is problematic also for  $E_1$  and  $E_2$ ). Instead, the sum of the rates  $k_{12} + k_{21}$  [rate constant of the exponential decay in Equations (3)–(5)] can be determined with high accuracy (Figure 4B). Likewise, the structural parameter defined as the ratio of length  $\omega_z$  to width  $\omega_r$  of the focal volume (see Figure 4B and Figure S2 of the Supporting Information) can be determined accurately.



**Figure 4.** Correlations of experimental parameters: A) Two-dimensional histogram of the rate constants  $k_{12}$  and  $k_{21}$  for an Atto532–Atto647N sample containing 10 mM NaCl measured at 60  $\mu$ W laser power. The top and side show the corresponding one-dimensional projection. The two parameters appear strongly correlated. B) Two-dimensional histogram and one-dimensional projections of the sum of transition rates versus the structural parameter ( $\omega_z/\omega_r$ ). In contrast to the individual rates, the sum of rates as well as the structural parameter can be determined accurately. C) Schematic representation of the fit-error determination. Errors are given as the distance of the best fitting sample (blue point) to the lower and upper boarders that mark  $1\sigma$  of the distribution (black lines). Since this best fitting sample is not required to show an equal distance to both boarders, the greater of the two distances was used as fit error for weighted averaging of the results.

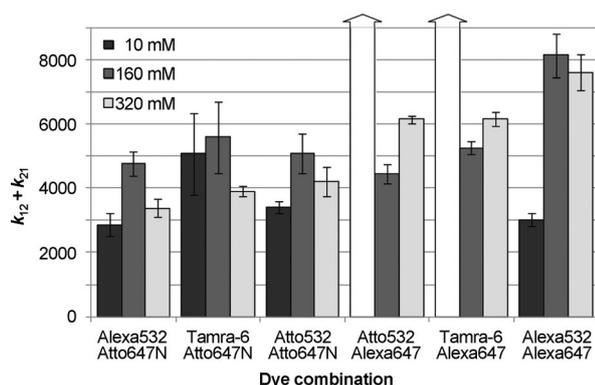
To summarize the accuracy of the parameter estimation, the fit error was always defined as the distance of the maximum posterior sample to the points of the distribution including 68% of the probability as shown in Figure 4C. Note that for this type of multi-dimensional analysis, oftentimes the best fitting sample (i.e. maximum posterior) is not positioned exactly in the middle of that interval; in fact, it does not even need to lie inside it.

Both the choice of dye molecules and the salt buffer conditions (i.e. the salt concentration) have an effect on the equilibrium constant, as well as on the kinetics between open and closed states for DNA hairpins. For the particular hairpins used in this work, assuming a perfect two-state (open-closed) system, one expects transitions from the open state with  $\sim 10$ – $20\%$  FRET (depending on the Förster distance of the respective dye pair) to the closed state with  $\sim 100\%$  FRET.

Herein, we determined the kinetics of hairpins for different dye pairs and different salt concentrations, summarized in Figure 5. The total rate shows almost no dependence on the dye pair at 10 mM salt. At 160 mM NaCl, only the Alexa532–Alexa647 dye combination has a significantly higher rate whereas at 320 mM NaCl, more distinct influences of the dye molecules are observed. Hairpins labeled with Atto647N show

slow transitions but no differences between the donor dyes (6-Tamra, Alexa532, Atto532). Changing the red dye to Alexa647 leads, however, to a significant rate increase. In this case, Atto532 and 6-Tamra labeled hairpins show equal rates while hairpins with Alexa532 open and close much faster but at rates similar to the ones found for 160 mM NaCl.

While the FCS data presented in Figure 1 show the typical situation for hairpins containing Alexa647, which can be described using a simple two-state model, we find that using Atto647N instead causes unexpectedly high residuals of the fits to the correlation functions (Figure S3A of the Supporting Information). The observed anti-correlations of the residuals from the fits to the donor auto-correlation and to the donor-FRET cross-correlation show an anti-



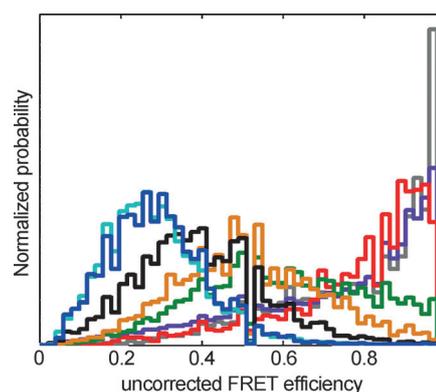
**Figure 5.** Summary of the kinetic results obtained from the FCS experiments: Overview of the determined hairpin kinetic rates for a variety of dye combinations and salt amounts; determined sum of kinetic rates  $k_{12}$  and  $k_{21}$  for 10 mM (black), 160 mM (dark grey) and 320 mM (light grey) NaCl. Experiments were performed in triplicates and error bars were calculated for the fit-error-weighted averages. Samples not showing useful kinetic information are marked in white (dominating photophysical effects: Atto532 Atto647N 10 mM, no kinetic signal detected: Tamra-6 Alexa647 10 mM).

correlation on a timescale of a few 100  $\mu$ s (Figure S3 B, Supporting Information), which has previously been described as an indicator for the presence of additional kinetics.<sup>[2]</sup> More than two states have been reported for a five-base-pair  $T_{21}$  DNA hairpin, but until now only one transition was accessible by conventional FCS.<sup>[3]</sup>

## 2.2. FRET Analysis

Due to the discussed difficulties in extracting  $E_1$  and  $E_2$  from the FCS datasets, additional measurements with reduced DNA concentrations were performed, so that single-molecule fluorescence bursts were observed. With the analysis of these bursts we were able to gain insight into the distribution of states for each dye pair. To this end, the FRET efficiency ( $E$ ) and stoichiometry ( $S$ )<sup>[37]</sup> were calculated for each burst (Methods section). Data was collected for the double-labeled DNA hairpins at salt concentrations of 10, 160, and 320 mM NaCl.

Assuming a two-state kinetic system, one expects a gradual change of the equilibrium between open and closed DNA hairpins from a completely open conformation at low salt (low-FRET-state-populated) to a completely closed conformation at high salt concentrations (high-FRET-state-populated). From the kinetic rates determined in the FCS analysis we estimate 3–8 transitions during the average burst duration of  $\sim 1$  ms. Therefore, the mean FRET efficiencies per burst are likely to be averaged out and one expects only one distinct peak at any salt concentration, with the position of the peak determined at the respective equilibrium value.<sup>[38]</sup> The hairpins labeled with the Cy3–Cy5 dye combination match this expected behavior, reaching almost 100% population of the closed state at  $\sim 100$  mM NaCl (Figure 6). In contrast, for a variety of combinations of 6-Tamra, Atto and Alexa dyes, different FRET histograms were observed (Figure 7 A–F). While at low salt concentrations the histograms for all dye combinations show the expected shape, this is no longer true for elevated salt concen-



**Figure 6.** FRET-efficiency distributions for hairpins labeled with the Cy3–Cy5 dye pair obtained from single-molecule burst analysis experiments. Comparison of the FRET efficiency distributions for the Cy3–Cy5 double-labeled hairpins recorded at 0 mM (dark blue), 10 mM (light blue), 20 mM (black), 40 mM (gold), 80 mM (green), 100 mM (red), 160 mM (purple), and 200 mM (grey) NaCl concentration. Note that these FRET efficiencies have not been corrected for crosstalk, direct excitation, differences in fluorescence quantum yield or differences in detection efficiency.

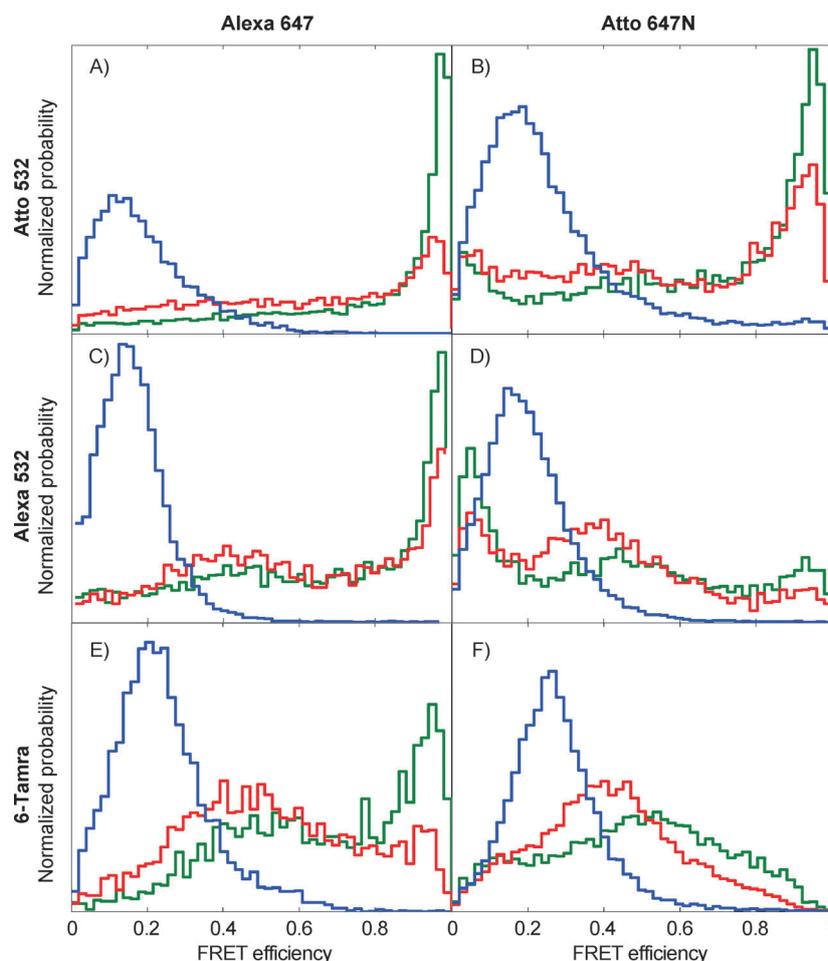
trations. For high salt concentrations (160 mM), the different dye pairs lead to distinct histograms. Under such conditions (at least) two distinct peaks are observed. These distributions cannot be explained using a simple two-state model, in agreement with deviations from the simplistic picture in the FCS fit residuals discussed above.

## 3. Discussion and Conclusion

FCS and single-molecule burst analysis experiments of DNA hairpins labeled with different dye molecules were performed to understand the effect of the probe molecules on DNA hairpin dynamics. For the analysis of the FCS datasets, a global analysis approach using Bayesian-parameter estimation was developed to accurately extract kinetic information. With this approach, significant differences between hairpins labeled with different probes were observed, not only for the distribution of states but also for the kinetics.

A related approach to globally fit FCS data has been published recently,<sup>[46]</sup> where the number of fit parameters was reduced by assuming a transition between 0% and 100% FRET efficiency as well as by pre-determining several parameters such as the triplet kinetics, focal volume and diffusion constants. By globally fitting a set of correlation functions with the remaining variables, it was possible to determine the kinetic rate.<sup>[46]</sup> This approach has its limitations since it requires a perfect overlap of the green and red detection volumes and is only applicable to systems known to switch between 0% and 100% FRET. In contrast, our approach uses a global analysis of the correlation functions without any restrictions. Even in the absence of a reliable calibration of the lateral and axial focal dimensions, it is possible to extract information such as kinetic rates and the structural parameter.

With this approach—and in combination with burst analysis—we could then show that the opening and closing of DNA hairpins is influenced by the choice of dye molecules. Our re-



**Figure 7.** FRET distributions of DNA hairpins as a function of salt and choice of dye pair obtained from single-molecule burst analysis experiments. FRET-efficiency distributions for different FRET pair combinations measured at 100  $\mu$ W laser power with NaCl concentrations of 10 mM (blue), 160 mM (red) and 320 mM (green). Shown are burst analysis data for Atto532–Alexa647 (A), Atto532–Atto647N (B), Alexa532–Alexa647 (C), Alexa532–Atto647N (D), 6-Tamra–Alexa647 (E), and 6-Tamra–Atto647N (F).

sults show a significant influence of the dye pair on transition rates as well on the equilibrium constant for the DNA hairpins investigated. At low salt concentrations, where the hairpins are expected to exist predominantly in the open conformation and therefore the dyes are expected to be far away from each other, all hairpins behave in a similar fashion. In contrast, a change to higher salt concentrations leads to distinct differences depending on the dyes selected. No stable hairpin is formed for Alexa532–Atto647N as well as 6-Tamra–Atto647N, even for the highest investigated salt concentrations. Possible reasons include a sterical hindrance of hairpin formation or a repulsive interaction of the dye molecules. In contrast, other combinations of dye molecules, especially Cy3–Cy5 (and to a lower extent Alexa532–Alexa647, Atto532–Alexa647 and Atto532–Atto647N) stabilize the closed-loop conformation already at 160 mM NaCl while the 6-Tamra–Alexa647-labeled hairpins require 320 mM NaCl to show a distinct high FRET peak. Moreover, at higher NaCl concentrations, the FCS data show that also the kinetics are influenced by the dye pair. We can classify the investigated hairpins into two groups accord-

ing to their rates at 320 mM NaCl. The first group shows 3000–4000 transitions per second and contains all molecules labeled with Atto647N. In contrast, all Alexa647-labeled hairpins show significantly higher rates (6000–7500 per second). Moreover, samples with Atto647N showed a decrease in the rates while Alexa647 led to a rate increase when going from 160 to 320 mM NaCl, indicating that the molecular reason for the observed effects is quite complicated and has to include several different physical properties of the dye molecule, including hydrophobicity, charge, rigidity and size. Similar results come from the burst analysis data, where we find comparatively high populations of low FRET states (open hairpin) for Atto647N, even at elevated salt concentrations, but not for Alexa647. These data suggests that Atto647N strongly interacts with the DNA bases, thus hindering the DNA hairpin closure.

More insights into the effects of the dye combinations are gained by comparing the number of states that are required to describe the data for the different hairpins. The FRET distributions observed for Cy3–Cy5-labeled hairpins match the shape expected for a direct transition from an open to a closed state at a rate faster than the observation time (Figure 6).<sup>[38]</sup> To describe the behavior of the hairpins labeled with different dye molecules, at least one additional intermediate state is required. Previously, a three-state model was required to describe the dynamics detected using FCS for a hairpin labeled with Rhodamine6G and Dabcyl and containing a four-base-pair stem. In contrast, the same hairpin containing a five-base-pair stem could be described by a two-state model.<sup>[3,6]</sup> It was found that for the five-base-pair system, fluctuations were too slow to be detected using FCS while photon counting histogram experiments revealed the existence of a third state. Our data supports this interpretation, since we find no additional correlations in the residuals of the fits to the correlation functions for the 160 and 320 mM NaCl FCS datasets for the Alexa647 dye combinations, while the Atto647N dye combinations show weak additional fluctuations in the range of a few 100  $\mu$ s (see Figure S3 of the Supporting Information). The latter data, as well as the burst analysis data of all Atto647N and Alexa647 dye combinations, can only be

explained if at least three states are used, as suggested previously.<sup>[3]</sup> A possible model is to introduce a semi-closed intermediate state consisting of numerous partially or mis-folded hairpins. In such a model, the slow process can be attributed to the formation and disruption of the fully closed stem while the fast process describes the transitions between an intermediate and a fully open structure.

While simple repulsion of the dye molecules would lead to a highly dynamic interchange between the three possible states, a sterical hindrance to hairpin closure, for example, caused by a dye molecule stacking to a base of the DNA, could yield a more long-lived state. DNA interactions have been reported for different dye molecules such as Tamra<sup>[47,48]</sup> or TexasRed.<sup>[48]</sup> For such a case, one expects a static heterogeneity, as observed in the burst analysis data. This hypothetical model is supported by in-gel burst analysis experiments, in which, depending on the position along a band, distinct FRET distributions can be identified (Figure S4 of the Supporting Information). One should note that there are other experimental techniques that can also be used for testing the influence of a dye molecule on the equilibrium value of the open versus closed state. For example, one could exploit hypochromicity to compare the salt dependence of DNA hairpin closure in unlabeled and labeled samples. In exemplary experiments we found that while the Cy3-labeled and the unlabeled hairpins behave quite similar, 6-Tamra-labeled samples show a deviation (Figure S5 of the Supporting Information), consistent with our interpretation that in these samples the hairpin closure is hindered.

In summary, DNA hairpins are known to be highly sensitive to changes in the stem or even the loop sequence.<sup>[1]</sup> Due to their similarity to RNA structures in ribozymes and t-RNA, understanding the effects of external reporter systems on the energy landscape is essential for numerous applications. Even if the dye influence on bigger systems such as proteins is less pronounced than the influences reported here, it is important to realize that the results will depend on the choice of dye molecules. Thus, trying and comparing several dye combinations whenever working with dynamic systems is recommended. Based on our findings we suggest the use of Alexa647 as an acceptor and, if photophysics are properly accounted for, Cy3 as the donor as good (initial) candidates for studies of molecular dynamics.

## Experimental Section

### Experimental Setup

Measurements were performed on a custom-built confocal microscope setup based on a Nikon TE3000 (Nikon) inverted microscope. Dye molecules were excited by short pulses of two lasers with wavelengths of 532 nm (Pico-TA 530 Picoquant) and 640 nm (LDH-D-C-640, Picoquant) at a repetition rate of 26.66 MHz using pulsed interleaved excitation (PIE).<sup>[35]</sup> The lasers were fiber-coupled to single-mode fibers (Schäfter+Kirchhoff), combined using a wavelength division multiplexer (AMS Technologies) and collimated using a fiber collimator (60FC-4-RGB11-47, Schäfter+Kirchhoff). Laser powers were 25  $\mu$ W (measured directly before the beams enter the objective lens) for FCS and 100  $\mu$ W for the burst measurements

if not stated otherwise. Fluorescence was collected using a Nikon water immersion objective (Nikon Plan Apo VC 60x/1.20 WI). The confocal volume was defined by focusing onto a  $d=50$   $\mu$ m pinhole using a  $f=200$  mm focal-length lens. Fluorescence was split by polarization (PBS3, Thorlabs) and color (laser beam-splitter 640DCXR, AHF analysetechnik AG) and cleaned up by a set of filters (green detection channels: Brightline HC582/75; red detection channels: Brightline HQ700/75, AHF analysetechnik AG). Four single photon counting avalanche photodiodes (PerkinElmer) were used for fluorescence detection and connected to four single-photon-counting modules (Becker&Hickl SPC-154).

### Nucleic-Acid Hairpins

DNA hairpins (sequence CCCAA-(T)<sub>21</sub>-TTGGG) containing the acceptor at the 3'-end and the donor at the 5'-end attached via a C6 linker (see schematic in Figure 1A) were purchased from IBA GmbH. A separate set of DNA molecules with identical sequence but containing only dye molecules at the 5'-end were used for donor-only reference measurements. As fluorescence donors, Atto532, Alexa532, 6-Tamra and Cy3 were used, while Atto647N, Alexa647 and Cy5 (NHS) served as acceptors. Experiments were performed at 21 °C in 1x TE buffer (100 mM Tris, 1 mM EDTA, pH 7.6) at varying NaCl concentrations of 10, 160 or 320 mM within Lab-Tek II chambers (Nalge Nunc In. Corp. Naperville, USA). Nucleic-acid concentrations were 100  $\mu$ M for the burst analysis and 1–10 nM for the FCS measurements.

### Data Processing

Data were analyzed using custom software written in MATLAB (MathWorks) and C++. The recorded photon streams of one color but different polarization were merged resulting in a total of two detection channels, since polarization was not investigated in the present study. Photons in the acceptor channel were further split into two channels according to their arrival times, that is, after the 640 nm laser pulse (acceptor after acceptor excitation, *R*) or after the 532 nm laser pulse (acceptor after donor excitation, *F*). Photon-arrival-time correlations were computed using a fast correlation algorithm.<sup>[42]</sup> For the donor-only sample, only photons after donor excitation (*GGd*) were auto-correlated while for the double-labeled samples, donor photons after donor excitation (*GG*), acceptor photons after acceptor excitation (*RR*) and acceptor photons after donor excitation (*FF*) were auto-correlated. In addition, the cross-correlation function of donor and acceptor photons after donor excitation (*FG*) was computed. FCS error bars were determined by segmenting the total measurement time for each hairpin into eight parts, performing independent correlations for each of these segments and calculating the mean and the standard error. All correlation functions containing conformational dynamics were limited to timescales of 10  $\mu$ s–1 s in order to be insensitive to photophysical effects of the labels.

For the burst analysis, photon bursts were selected using an all-photons burst search (APBS)<sup>[38]</sup> with thresholds of at least 30 photons within a time interval of 1.2 ms and a total burst size of at least 60 photons.

FRET efficiency (*E*) and stoichiometry (*S*) were calculated for each burst using Equations (7) and (8):

$$E = \frac{F - d * R - c * G}{F - d * R - c * G + \gamma * G} \quad (7)$$

$$S = \frac{F - d * R - c * G + \gamma * G}{F - d * R - c * G + \gamma * G + R} \quad (8)$$

where  $F$  denotes the number of red photons after green excitation,  $R$  the number of red photons after red excitation, and  $G$  the number of green photons after green excitation. Direct excitation of the acceptor is accounted for by the correction factor  $d$ ,  $c$  accounts for crosstalk and  $\gamma$  is a factor correcting for the different quantum yields of the dyes and the detection efficiencies of the red and green detection channels, respectively.<sup>[39]</sup> These correction factors were determined following standard procedures.<sup>[36,37]</sup> Single molecule bursts were separated from background events using a Stoichiometry threshold depending on the respective dye pair.

## Bayesian Inference

In Bayesian data analysis, the information about the expected data given a model of the physical process is encoded in the *likelihood*,  $p(d|\theta, I)$ . This probability density connects the measured data,  $d$  (here the correlation curves), with the parameters,  $\theta$ , that quantify the model (here the kinetic model including diffusion, dynamics and beam geometry). The so-called background information,  $I$ , summarizes all information available to the experimenter (such as FRET theory, photon counting statistics etc.).

The state of information of the experimenter before the experiment was performed,  $p(\theta|I)$ , is called *prior*. It contains all knowledge about the model parameters, for instance constraints imposed by physical laws or information from previous measurements.

The information contained in the data can be combined with the prior information by using Bayes' theorem [Eq. (9)]:

$$p(\theta|d, I) \propto p(d|\theta, I)p(\theta|I) \quad (9)$$

The probability density  $p(\theta|d, I)$  denotes the updated information after data analysis and is called *posterior*.

When the model is described by many parameters, it is oftentimes infeasible to evaluate the posterior probability density systematically, for example, on an evenly spaced grid, to account for every combination of model parameter values. In most cases, the computed densities would be close to zero, indicating that the particular values of the model parameters are very unlikely. Instead, one chooses to represent the posterior by a large-enough set of samples in the parameter space drawn from the posterior distribution, that is, occurring with the posterior probability.

Bayesian data analysis of problems in moderate- and high-dimensional parameter spaces oftentimes requires advanced numerical tools. Herein, we used nested sampling,<sup>[49]</sup> an algorithm that searches the region of the parameter space that contains the major part of the posterior probability. This numerical approach for Bayesian analysis was used until recently, mainly in astrophysics, but has also been applied to the localization of unknown molecule parts,<sup>[50]</sup> instrument noise evaluation,<sup>[51]</sup> and model selection.<sup>[52]</sup>

Our custom implementation of the nested sampling algorithm is based on Metropolis Markov chain Monte Carlo written in MATLAB (The MathWorks) and C. The algorithm was previously developed to compute the position of unknown domains in macromolecular complexes in an analysis termed Nano Positioning System NPS.<sup>[50]</sup>

In order to be insensitive to experimental parameters that could vary when changing the samples, we chose to use a prior where none of the model parameters were pre-determined, but instead limited the parameters to a reasonable range (see Table S1 of the Supporting Information). A Jeffreys prior<sup>[53]</sup> was used for all variables (see Scheme 1), except for  $N$ ,  $E$  and  $f_{isc}$ ,<sup>[53]</sup> where we used a flat

prior to reduce the computational expense since feasible values can vary only in a narrow range.

Since nested sampling produces samples with varying weight<sup>[49]</sup> (Figure S1 of the Supporting Information), a subset of samples is chosen by drawing from this set of samples, with a probability proportional to the sample weight under the constraint that each sample can only be drawn once.<sup>[54]</sup> All samples obtained in this way are used to represent the posterior in scatter plots and histograms.

## Acknowledgments

We thank C. Röcker for fruitful discussions, M. Höller for providing the software correlation algorithm, A.-L. Coast for help with control measurements and P. Braun and H. Langhals for their help with the UV spectra. This work was supported by the nano initiative munich (NIM), the Deutsche Forschungsgemeinschaft (SFB646), and by a Starting Investigator Grant of the European research council (ERC).

**Keywords:** Bayesian inference · DNA · dynamics · fluorescence correlation spectroscopy · single-molecule studies

- [1] G. Bonnet, O. Krichinsky, A. Libchaber, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 8602–8606.
- [2] T. Torres, M. Levitus, *J. Phys. Chem. B* **2007**, *111*, 7392–7400.
- [3] J. Jung, R. Ihly, E. Scott, M. Yu, A. Van Orden, *J. Phys. Chem. B* **2008**, *112*, 127–133.
- [4] J. Jung, A. Van Orden, *J. Am. Chem. Soc.* **2006**, *128*, 1240–1249.
- [5] J. Jung, A. Van Orden, *J. Phys. Chem. B* **2005**, *109*, 3648–3657.
- [6] A. V. Orden, J. Jung, *Biopolymers* **2008**, *89*, 1–16.
- [7] S. Liu, G. Bokinsky, N. G. Walter, X. Zhuang, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 12634–12639.
- [8] D. Rueda, G. Bokinsky, M. M. Rhodes, M. J. Rust, X. Zhuang, N. G. Walter, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 10066–10071.
- [9] R. Russell, X. Zhuang, H. P. Babcock, I. S. Millett, S. Doniach, S. Chu, D. Herschlag, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 155–160.
- [10] L. E. Bartley, X. Zhuang, R. Das, S. Chu, D. Herschlag, *J. Mol. Biol.* **2003**, *328*, 1011–1026.
- [11] G. Bokinsky, D. Rueda, V. K. Misra, M. M. Rhodes, A. Gordus, H. P. Babcock, N. G. Walter, X. Zhuang, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 9302–9307.
- [12] G. Bokinsky, X. Zhuang, *Acc. Chem. Res.* **2005**, *38*, 566–573.
- [13] T. R. Blosser, J. G. Yang, M. D. Stone, G. J. Narlikar, X. Zhuang, *Nature* **2009**, *462*, 1022–1027.
- [14] K. Gurunathan, M. Levitus, *Prog. Nucleic Acid Res. Mol. Biol.* **2008**, *82*, 33–69.
- [15] K. Gurunathan, M. Levitus, *Curr. Pharm. Biotechnol.* **2009**, *10*, 559–568.
- [16] G. Li, M. Levitus, C. Bustamante, J. Widom, *Nat. Struct. Mol. Biol.* **2005**, *12*, 46–53.
- [17] R. D. Vale, *J. Cell Biol.* **2003**, *163*, 445–450.
- [18] T. Hugel, J. Michaelis, C. L. Hetherington, P. J. Jardine, S. Grimes, J. M. Walter, W. Falk, D. L. Anderson, C. Bustamante, *PLoS Biol.* **2007**, *5*, e59.
- [19] X. Michalet, S. Weiss, M. Jager, *Chem. Rev.* **2006**, *106*, 1785–1813.
- [20] B. Okumus, T. Ha, *Methods Mol. Biol.* **2010**, *608*, 81–96.
- [21] D. Nettels, I. V. Gopich, A. Hoffmann, B. Schuler, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2655–2660.
- [22] D. Nettels, A. Hoffmann, B. Schuler, *J. Phys. Chem. B* **2008**, *112*, 6137–6146.
- [23] S. Weiss, *Nat. Struct. Biol.* **2000**, *7*, 724–729.
- [24] D. S. Majumdar, I. Smirnova, V. Kasho, E. Nir, X. Kong, S. Weiss, H. R. Kaback, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 12640–12645.
- [25] Y. Santos, C. M. Joyce, O. Potapova, L. Le Reste, J. Hohlbein, J. P. Torella, N. D. Grindley, A. N. Kapanidis, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 715–720.

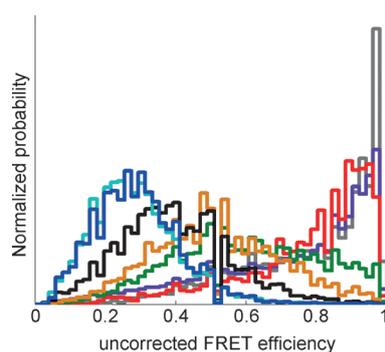
- [26] M. Schuttpelz, J. C. Schoning, S. Doose, H. Neuweiler, E. Peters, D. Staiger, M. Sauer, *J. Am. Chem. Soc.* **2008**, *130*, 9507–9513.
- [27] M. Margittai, J. Widengren, E. Schweinberger, G. F. Schroder, S. Felekyan, E. Hausteil, M. König, D. Fasshauer, H. Grubmüller, R. Jahn, C. A. Seidel, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15516–15521.
- [28] M. I. Wallace, L. Ying, S. Balasubramanian, D. Klenerman, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5584–5589.
- [29] K. Gurunathan, M. Levitus, *J. Phys. Chem. B* **2010**, *114*, 980–986.
- [30] S. A. Marras, F. R. Kramer, S. Tyagi, *Nucleic Acids Res.* **2002**, *30*, 122e.
- [31] X. Zhuang, L. E. Bartley, H. P. Babcock, R. Russell, T. Ha, D. Herschlag, S. Chu, *Science* **2000**, *288*, 2048–2051.
- [32] X. Zhuang, H. Kim, M. J. Pereira, H. P. Babcock, N. G. Walter, S. Chu, *Science* **2002**, *296*, 1473–1476.
- [33] E. Tan, T. J. Wilson, M. K. Nahas, R. M. Clegg, D. M. Lilley, T. Ha, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 9308–9313.
- [34] S. Hohng, T. J. Wilson, E. Tan, R. M. Clegg, D. M. Lilley, T. Ha, *J. Mol. Biol.* **2004**, *336*, 69–79.
- [35] B. K. Müller, E. Zaychikov, C. Brauchle, D. C. Lamb, *Biophys. J.* **2005**, *89*, 3508–3522.
- [36] A. N. Kapanidis, T. A. Laurence, N. K. Lee, E. Margeat, X. Kong, S. Weiss, *Acc. Chem. Res.* **2005**, *38*, 523–533.
- [37] A. N. Kapanidis, N. K. Lee, T. A. Laurence, S. Doose, E. Margeat, S. Weiss, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 8936–8941.
- [38] E. Nir, X. Michalet, K. M. Hamadani, T. A. Laurence, D. Neuhauser, Y. Kovchegov, S. Weiss, *J. Phys. Chem. B* **2006**, *110*, 22103–22124.
- [39] N. K. Lee, A. N. Kapanidis, Y. Wang, X. Michalet, J. Mukhopadhyay, R. H. Ebright, S. Weiss, *Biophys. J.* **2005**, *88*, 2939–2953.
- [40] D. Magde, E. L. Elson, W. W. Webb, *Biopolymers* **1974**, *13*, 29–61.
- [41] M. Böhmer, M. Wahl, H.-J. Rahn, R. Erdmann, J. Enderlein, *Chem. Phys. Lett.* **2002**, *353*, 439–445.
- [42] M. Wahl, I. Gregor, M. Patting, J. Enderlein, *Opt. Express* **2003**, *11*, 3583–3591.
- [43] M. Wahl, H. J. Rahn, I. Gregor, R. Erdmann, J. Enderlein, *Rev. Sci. Instrum.* **2007**, *78*, 033106.
- [44] C. A. M. Seidel, J. Widengren, *Phys. Chem. Chem. Phys.* **2000**, *2*, 3435–3441.
- [45] M. E. Sanborn, B. K. Connolly, K. Gurunathan, M. Levitus, *J. Phys. Chem. B* **2007**, *111*, 11064–11074.
- [46] E. S. Price, M. S. DeVore, C. K. Johnson, *J. Phys. Chem. B* **2010**, *114*, 5895–5902.
- [47] C. Eggeling, J. R. Fries, L. Brand, R. Gunther, C. A. Seidel, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 1556–1561.
- [48] J. R. Unruh, G. Gokulrangan, G. H. Lushington, C. K. Johnson, G. S. Wilson, *Biophys. J.* **2005**, *88*, 3455–3465.
- [49] J. Skilling, *Bayesian Anal.* **2006**, *1*, 833–860.
- [50] A. Muschielok, J. Michaelis, *J. Phys. Chem. B* **2011**, *115*, 11927–11937.
- [51] P. Maragakis, F. Ritort, C. Bustamante, M. Karplus, G. E. Crooks, *J. Chem. Phys.* **2008**, *129*, 024102.
- [52] J. E. Bronson, J. Fei, J. M. Hofman, R. L. Gonzalez, Jr., C. H. Wiggins, *Biophys. J.* **2009**, *97*, 3196–3205.
- [53] H. Jeffreys, *Theory of probability*, Clarendon Press, Oxford, **1939**.
- [54] D. Sivia, *Data Analysis, 2nd ed.*, Oxford University Press, Oxford, UK, **2006**.

Received: September 20, 2011

Published online on ■ ■ ■, 2012

## ARTICLES

**Dynamic studies:** Fluorescence correlation spectroscopy (FCS) data are globally analyzed using the Bayesian inference. This approach is applied, in combination with the results from single-molecule burst analysis, to investigate the influence of various labels on DNA hairpin kinetics as a function of the salt concentration.



W. Kügel, A. Muschielok, J. Michaelis\*



**Bayesian-Inference-Based  
Fluorescence Correlation Spectroscopy  
and Single-Molecule Burst Analysis  
Reveal the Influence of Dye Selection  
on DNA Hairpin Dynamics**

