THE JOURNAL OF CHEMICAL PHYSICS **147**, 152708 (2017)

Combining short- and long-range fluorescence reporters with simulations to explore the intramolecular dynamics of an intrinsically disordered protein

Franziska Zosel,[1,a)] Dominik Haenni,[1,2,a)] Andrea Soranno,[1,3] Daniel Nettels,[1]
and Benjamin Schuler[1,4,b)]
[1]*Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland*
[2]*Center for Microscopy and Image Analysis, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland*
[3]*Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110, USA*
[4]*Department of Physics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland*

(Received 12 May 2017; accepted 23 June 2017; published online 21 July 2017)

Intrinsically disordered proteins (IDPs) are increasingly recognized as a class of molecules that can exert essential biological functions even in the absence of a well-defined three-dimensional structure. Understanding the conformational distributions and dynamics of these highly flexible proteins is thus essential for explaining the molecular mechanisms underlying their function. Single-molecule fluorescence spectroscopy in combination with Förster resonance energy transfer (FRET) is a powerful tool for probing intramolecular distances and the rapid long-range distance dynamics in IDPs. To complement the information from FRET, we combine it with photoinduced electron transfer (PET) quenching to monitor local loop-closure kinetics at the same time and in the same molecule. Here we employed this combination to investigate the intrinsically disordered N-terminal domain of HIV-1 integrase. The results show that both long-range dynamics and loop closure kinetics on the sub-microsecond time scale can be obtained reliably from a single set of measurements by the analysis with a comprehensive model of the underlying photon statistics including both FRET and PET. A more detailed molecular interpretation of the results is enabled by direct comparison with a recent extensive atomistic molecular dynamics simulation of integrase. The simulations are in good agreement with experiment and can explain the deviation from simple models of chain dynamics by the formation of persistent local secondary structure. The results illustrate the power of a close combination of single-molecule spectroscopy and simulations for advancing our understanding of the dynamics and detailed mechanisms in unfolded and intrinsically disordered proteins. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4992800]

I. INTRODUCTION

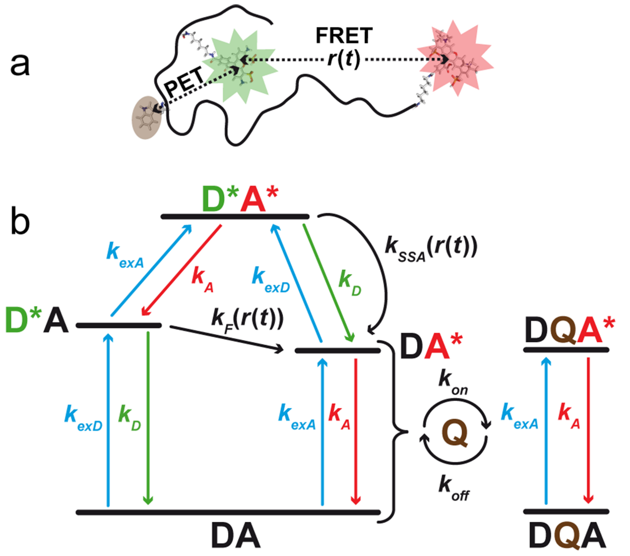
Proteins perform a vast range of functions at the molecular level in all domains of life. Until recently, this functionality was considered to invariably require a well-defined three-dimensional structure, a notion shaped by the immense success of crystallography at solving protein structures with atomic resolution. Over the past decade, however, it has become increasingly clear that many proteins lack a defined 3D-structure under physiological conditions but nevertheless are functional.[1,2] Bioinformatics analyses based on whole-genome DNA sequences suggest that at least 30% of all mammalian proteins contain large unstructured regions or are entirely unfolded.[1] Many of these "intrinsically disordered proteins" (IDPs) are involved in regulatory processes such as signal transduction and transcription[3] and in the formation of complex cellular interaction networks.[4] As a result, many IDPs are of great medical relevance.[5] Understanding the dynamics of these flexible molecules is an important aspect in explaining their biological mode of action, and new types of approaches are required for quantifying and describing their properties.

Single-molecule fluorescence spectroscopy in combination with Förster resonance energy transfer (FRET)[6,7] and nanosecond fluorescence correlation spectroscopy (ns-FCS) is a powerful tool for probing the distance distributions and dynamics in unfolded proteins and IDPs by enabling the observation of long-range distances and their fluctuations.[8–11] Complementary information on short-range dynamics can be obtained by loop-closure experiments that require contact formation between two reporter groups, e.g., utilizing triplet-triplet energy transfer,[12] triplet-state quenching,[13] or photo-induced electron transfer (PET).[14,15] Recently, PET and FRET experiments were used to obtain complementary information on the short- and long-range chain dynamics of an unfolded protein.[16] Even though these experiments were performed on the same protein, they were obtained in independent measurements and on different variants of the protein required for FRET and PET. Ideally, long- and short-range

---

[a)] F. Zosel and D. Haenni contributed equally to this work.
[b)] Author to whom correspondence should be addressed: schuler@bioc.uzh.ch

distance dynamics would be measured simultaneously within one experiment. This approach would eliminate uncertainties due to possible differences between the variants[17] and may enable the analysis of correlations between dynamics on different length scales.[18] Moreover, a potential complication of single-molecule FRET experiments on proteins is that many of the commonly used fluorophores are quenched by the amino acid tryptophan (Trp),[19,20] and quenching can thus be an undesired side-effect that can modulate both the observed photon count rates (and therefore transfer efficiencies)[21] and the observed fluorescence correlation functions.[16,20] A quantitative analysis of such experiments would thus be highly desirable, but it requires a detailed model combining the photophysics of both FRET and PET with the dynamics of the chain.



c

$$\mathbf{K_0} = \begin{pmatrix} K_{11} & k_D & k_A & 0 & k_{off} & 0 \\ k_{exD} & K_{22} & 0 & k_A & 0 & 0 \\ k_{exA} & k_F(r(t)) & K_{33} & k_D + k_{SSA} & 0 & k_{off} \\ 0 & k_{exA} & k_{exD} & K_{44} & 0 & 0 \\ k_{on} & k_{on} & 0 & 0 & K_{55} & k_A \\ 0 & 0 & k_{on} & k_{on} & k_{exA} & K_{66} \end{pmatrix}$$

$$\tilde{\mathbf{V}}_\mathbf{D} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \tilde{\mathbf{V}}_\mathbf{A} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

FIG. 1. Kinetic model of the electronic and conformational transitions in the combined FRET and PET system.[21] (a) Schematic illustration of the investigated flexible IDP to which the donor and acceptor dyes (Alexa 488, green, and Alexa 594, red) are attached by flexible linkers. The Trp residue, which acts as a quencher, is depicted in brown. In the case shown here, the quenching of the acceptor is negligible in comparison to donor quenching. (b) Kinetic scheme describing the photophysical transitions. DA, D*A, DA*, and D*A* denote the electronic states of the donor (D) and acceptor (A) in the absence of quenching by Trp. The asterisks refer to the first exited singlet states S1 of the dyes. In contrast to the static FRET model used previously,[21] the energy transfer rate coefficient k_F is now allowed to fluctuate in accordance with the expected inter-dye-distance dynamics, which are modeled by diffusive motion in a potential of mean force (see text). DQA and DQA* refer to the states where donor and Trp form a dark complex DQ but the acceptor is still active. (c) Representation of the rate matrix $\mathbf{K_0}$ and the detection matrices $\tilde{\mathbf{V}}_\mathbf{D}$ and $\tilde{\mathbf{V}}_\mathbf{A}$ that are used to calculate FCS curves and mean photon rates (see main text). The diagonal elements K_{ij} are calculated from the requirement that the sums of the elements must be zero for each column of the matrix.

We previously[21] laid the groundwork for such an analysis by characterizing loop closure dynamics combined with a static FRET system in the same peptide. The two dyes Alexa 488 and 594 as the FRET donor and acceptor, respectively, were separated by a stiff polyproline peptide, and one dye was additionally subject to static PET-mediated quenching by a tryptophan residue connected via a flexible linker. Of all the naturally occurring amino acids, Trp has the largest quenching effect on Alexa 488 and Alexa 594.[19] A global analysis of ns-FCS curves and transfer efficiency histograms using a detailed model of the photon statistics allowed us to simultaneously extract inter-dye distances from FRET efficiencies and nanosecond loop formation dynamics from the on- and off-rates of Trp-induced dye quenching.[21]

Here we extend this approach to the integrated analysis of the long- and short-range dynamics of an intrinsically disordered protein that was site-specifically labeled with FRET donor and acceptor dyes and additionally contains the PET quencher Trp in close proximity to one dye [see Fig. 1(a)]. Compared to the polyproline model system, additional fluorescence intensity fluctuations are caused by the distance dynamics between the FRET dyes present in the IDP. Both processes, inter-dye distance dynamics (probed by FRET) and loop closure dynamics (probed by PET), occur on similar time scales in the sub-microsecond range. Hence, a robust analysis is required to disentangle both contributions. We show that reliable intra-chain diffusion coefficients and on- and off-rates for PET quenching can be obtained from a single protein variant, in one experiment. Finally, we compare our findings with recent atomistic simulations of integrase using an improved water model,[22] which show remarkably good agreement with the measurements and enable us to interpret the experimental findings on a detailed structural level.

II. RESULTS

A. Modelling combined dynamics from FRET and PET

The aim of this study is a full quantitative understanding of the ns-FCS data recorded on highly dynamic IDPs, which are labeled with a FRET dye pair to probe long-range distance dynamics and which additionally contain a static quencher (Trp) that quenches (predominantly) one of the dyes, e.g., the donor, to probe short-distance dynamics [Fig. 1(a)]. The analysis of such FCS curves requires extending the previously described model of a static polyproline system[21] to include protein dynamics.[9,23] We use the mathematical formalism introduced by Gopich and Szabo,[24] in which the combined effect of inter-dye distance dynamics and photophysical kinetics is modeled by a set of partial differential equations

$$\partial \mathbf{p}/\partial t = \mathbf{Kp} \tag{1}$$

that describes the evolution of the state population vector $\mathbf{p} = \mathbf{p}(r,t)$. The components of $\mathbf{p}$ are the interdye-distance probability density distributions $p_s = p_s(r,t)$ for each of the six states of the photophysical model depicted in Fig. 1(b) so that $\mathbf{p}$ is of the form $\mathbf{p} = (p_{DA}\ p_{D^*A}\ p_{DA^*}\ p_{D^*A^*}\ p_{DQA}\ p_{DQA^*})^T$, with $\sum_s \int p_s(r,t)dr = 1$. The operator $\mathbf{K}$ has two terms

$$\mathbf{K} = \mathcal{L}(r)\mathbf{I} + \mathbf{K_0}(r). \tag{2}$$

The first term contains the diffusion operator $\mathcal{L} = D\frac{\partial}{\partial r}p_{eq}(r)\frac{\partial}{\partial r}(p_{eq}(r))^{-1}$, which describes the diffusion of the inter-dye distance r in a potential of mean force with the equilibrium distribution $p_{eq}(r)$ and diffusion coefficient D. $\mathbf{I}$ is the identity matrix. We approximate $p_{eq}(r)$ by the distance distribution of a Gaussian chain[9,23,24]

$$p_{eq}(r) = 4\pi r^2\left(2\pi\left\langle r^2\right\rangle/3\right)^{-3/2}\exp\left(-3r^2/2\left\langle r^2\right\rangle\right), \tag{3}$$

where $\langle r^2\rangle$ is the mean squared dye-to-dye distance. The second term in Eq. (2) is the rate matrix $\mathbf{K_0}(r)$ that describes the transitions between electronic states of the fluorophores as well as the dynamics of dye-quencher complex formation. Its mathematical form and the underlying photophysical model are depicted in Figs. 1(c) and 1(b), respectively. Starting from the ground state, DA, the donor (D) or acceptor (A) can be excited by irradiation with laser light with the rate coefficients k_{exD} ($k_{exD} = 0.02\ \mathrm{ns}^{-1}$)[21] or $k_{exA} = \alpha k_{exD}$ to reach states D*A or DA*, respectively ($\alpha = 0.05$ is the fraction of acceptor direct excitation[21]). Absorption of another laser photon can lead to the double excited state D*A*. Donor and acceptor excited states (D* and A*) are depopulated with the rate coefficients k_D and k_A. Excitation energy is transferred from the donor to the acceptor with the rate coefficient $k_F(r) = k_D(R_0/r)^6$, where R_0 is the Förster radius of the dye pair. The double-excited state can be depopulated by singlet-singlet-annihilation.[21,25,26] This process is described by the rate coefficient $k_{SSA}(r) = k_D(R_0^{SSA}/r)^6$. Here, R_0^{SSA} is the Förster radius of the energy transfer that leads to singlet-singlet-annihilation. We used a value of $k_{SSA}(r) = 0.98k_F(r)$, which was determined for this dye pair in a recent study.[26] k_{on} and k_{off} are the association and dissociation rate coefficients of the π-stacked donor-quencher complex DQ (in the case described here, quenching of the acceptor dye is negligible because of the large sequence separation, see Table I and Sec. II D). Since the acceptor is still able to absorb and emit photons while the donor is quenched, we also account for the additional states DQA and DQA*. Complex formation populates DQA by transitions from DA and D*A, whereas DQA* is populated from DA* and D*A*. The complexed states DQA and DQA* dissociate to DA and DA*, respectively. In the formalism of Gopich and Szabo, fluorescence detection is modeled by radiative transition matrices given here as $\mathbf{V_D} = \xi_D Q_D k_D \tilde{\mathbf{V}}_\mathbf{D}$ and $\mathbf{V_A} = \xi_A Q_A k_A \tilde{\mathbf{V}}_\mathbf{A} + \beta\mathbf{V_D}$, where ξ_i and Q_i are the respective detection efficiencies and fluorescence quantum yields. The matrices $\tilde{\mathbf{V}}_\mathbf{D}$ and $\tilde{\mathbf{V}}_A$ indicate which transitions are monitored by the detectors. $\beta = 0.08$ is the fraction of donor fluorescence detected in the acceptor channel of our instrument.[21]

For the calculations used here, the continuous variable r was discretized as described previously[9] such that $\mathbf{K}$, $\tilde{\mathbf{V}}_\mathbf{D}$, and $\tilde{\mathbf{V}}_\mathbf{A}$ become square matrices of dimension $NM \times NM$. Here, $N = 40$ is the number of discrete distances $r_1 \ldots r_N$, distributed uniformly between $r_1 = 0.01$ nm and $r_N = 3\left\langle r^2\right\rangle^{1/2}$,[9] and $M = 6$ is the number of photophysical states represented in Fig. 1. Correspondingly, $\mathbf{p}$ becomes a vector of length NM. With the discretization, $\partial\mathbf{p}/\partial t = \mathbf{Kp}$ is reduced to a rate equation, for which the steady state solution $\mathbf{p_{ss}}$ (with $\mathbf{Kp_{ss}} = 0$) is easily found numerically. The theoretical fluorescence intensity correlation function $g_{ij}(\tau)$ between detection channels i and j for a single donor-acceptor-labeled molecule is then obtained from[27]

$$g_{ij}(\tau) = \frac{\mathbf{1}^T\mathbf{V}_j e^{\mathbf{K}\tau}\mathbf{V}_i\mathbf{p_{ss}}}{(\mathbf{1}^T\mathbf{V}_i\mathbf{p_{ss}})(\mathbf{1}^T\mathbf{V}_j\mathbf{p_{ss}})}, \tag{4}$$

where $i, j = A, D$, $\mathbf{1}^T = (11\ldots)$ is the transposed unity vector, τ is the lag time, and $e^{\mathbf{K}\tau}$ is the matrix exponential of $\mathbf{K}\tau$.

We measured labeled protein molecules freely diffusing through the confocal volume of the single-molecule instrument. Besides the donor-acceptor-labeled species, a subpopulation of proteins lacking an active acceptor fluorophore (donor-only population) was taken into account. The resulting final model functions, $G_{DD}(\tau)$, $G_{AA}(\tau)$, and $G_{DA}(\tau)$ for fitting the measured FCS data, which also include terms describing triplet state blinking and take into account differences in relative concentrations of the labeled species, molecular brightnesses, and background signals, are given in Eqs. (3)–(6) of Haenni *et al.*[21] The global fitting procedure was implemented as described previously.[21] The time constants of triplet blinking were determined with control protein variants lacking the Trp residue (see Table I) and set to the resulting values of 2.4 μs (G_{DD}), 2.2 μs (G_{AA}), and 3.5 μs (G_{DA}) for all fits. The triplet amplitudes were quantified based on the fits of the correlation curves in the time range between 0 and 4 μs.

B. Simplified kinetic models for fitting the control constructs

To test the validity of the approach described above, we prepared control constructs that either exhibited only FRET

TABLE I. Protein variants used in this study. Alexa 488-labeled Cys11, Alexa 594-labeled Cys60, and Trp 23/Phe 23 are indicated in bold.

Name	Sequence and labeling positions					
WDA	GSHMFLDGID	**C**AQEEHEKAH	SN**W**RAMASDF	NLPPVVAKEI	VASCDKCQLK	GEAMHGQVD**C**
FDA	GSHMFLDGID	**C**AQEEHEKAH	SN**F**RAMASDF	NLPPVVAKEI	VASCDKCQLK	GEAMHGQVD**C**
WD	GSHMFLDGID	**C**AQEEHEKAH	SN**W**RAMASDF	NLPPVVAKEI	VASCDKCQLK	GEAMHGQVD**C**
FD	GSHMFLDGID	**C**AQEEHEKAH	SN**F**RAMASDF	NLPPVVAKEI	VASCDKCQLK	GEAMHGQVD**C**

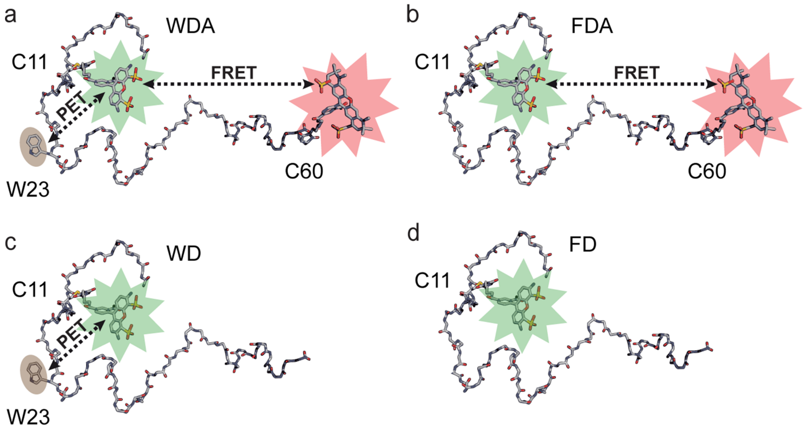


FIG. 2. Backbone representations of disordered integrase constructs with the relevant amino acids and the fluorophores shown explicitly. In the variant WDA (a), the donor dye Alexa 488 (green) can transfer energy to the acceptor dye Alexa 594 (red) and also undergo PET quenching by complex formation with Trp (brown). In the variant FDA (b), Trp was replaced by Phe to create a FRET-only control. The variant WD (c) carries only Alexa 488 at position 11 and serves as a PET-only control. In the variant FD (d), there is neither FRET nor PET quenching.

(by replacing Trp by phenylalanine (Phe), FDA) or only PET quenching (by omitting the FRET acceptor, WD), as depicted in Fig. 2. Simpler kinetic schemes than the one in Fig. 1 are sufficient to fit the FCS data of these constructs. A four-state model, corresponding to the left part of the kinetic scheme in Fig. 1(b), was used for a global to fit the FCS data of the FRET-only variant FDA. A detailed outline of its application to the study of unfolded protein dynamics is given in Nettels *et al.*[9] The FCS data of the PET-only construct WD was fitted with a three-state model with the population vector $\mathbf{p} = (p_D\ p_{D^*}\ p_{DQ})^T$. In this case, the fluorescence emission of the dye is only modulated by contact formation with the quencher Trp, so no diffusive dynamics have to be included in the kinetic model. The corresponding rate matrix is analogous to Eq. (15) in the work of Haenni *et al.*[21]

C. Protein and labeling variants

We investigated the N-terminal domain of the HIV-1 integrase, a 60-residue protein that is intrinsically disordered in the absence of Zn^{2+} ions.[28,29] We site-specifically labeled cysteine (Cys) residues introduced at positions 11 and 60 with Alexa 488 and 594 as the FRET donor and acceptor, respectively (Table I; see the Appendix for details of sample preparation). The only naturally occurring Trp residue in the sequence is located at position 23 and is thus expected to lead to pronounced quenching of Alexa 488 at position 11 and to have little effect on Alexa 594 at position 60. This construct will be subsequently referred to as WDA. We also prepared a control construct (FDA), where Trp at position 23 was replaced by phenylalanine (Phe) to eliminate quenching.[19] Additionally, two Alexa 488-only controls lacking an acceptor dye were prepared (WD/FD). The constructs are depicted schematically in Fig. 2. The sequences and labeling positions of all variants are compiled in Table I.

D. Extracting short-range and long-range dynamics from a combined PET/FRET experiment

The aim of this work is to extract information on short-range and long-range polypeptide chain dynamics from FCS measurements of a single construct labeled with a FRET dye pair and carrying a PET quencher at the same time. More specifically, we want to determine the reconfiguration time of the chain as monitored by FRET, reporting on long-range dye-to-dye distance dynamics, and the association rates between one of the dyes and the quencher Trp. To test the robustness of the approach outlined in Sec. II A, we recorded FCS curves at different concentrations of the denaturant guanidinium chloride (GdmCl). The influence of GdmCl on the dynamics of unfolded proteins has been characterized previously.[29,30] Additionally, GdmCl affects PET quenching by decreasing the stability of the dye-quencher complex.[20] Hence, the results from our analysis can be validated by comparison with already known effects of GdmCl on protein dimensions, dynamics, and PET quenching.

Before discussing the results, we briefly outline the fitting procedure and the requirements for obtaining robust estimates for long- and short range dynamics. To be able to apply the six-state photon-statistics model (Sec. II A, Fig. 1) for fitting the experimental data, we first need to ensure that only the donor dye is quenched by Trp. Upon direct excitation of Alexa 594 in WDA and FDA, only slight differences between the variants are observed: both the fluorescence lifetimes [which would indicate dynamic quenching, Fig. 3(b)] and the static quenching equilibrium constants probed with FCS are very similar (Fig. S5 of the supplementary material). In contrast, the fluorescence lifetime of Alexa 488 [Fig. 3(b)] and its static quenching equilibrium constant (Fig. S5) exhibit pronounced differences for the two variants, as expected for efficient quenching by the Trp residue. Thus, a photophysical model without acceptor quenching is a reasonable approximation.

To extract long- and short-range dynamics simultaneously, we record the correlation curves G_{DD} (donor autocorrelation), G_{AA} (acceptor autocorrelation), and G_{DA} (donor-acceptor cross correlation) after direct excitation of the donor dye in WDA. These correlation curves are then fitted globally with the six-state photon-statistics model (Sec. II A, Fig. 1). We aim to extract the three parameters of the model that describe the dynamics of the chain: the intra-chain diffusion coefficient D (reflecting the distance fluctuations between the FRET dyes), as well as the on- and off-rates for PET quenching, k_{on} and k_{off}. Almost all other model parameters are either previously known (k_{exD}, k_{exA}, k_{SSA}, α, β, γ, see Sec. II A) or

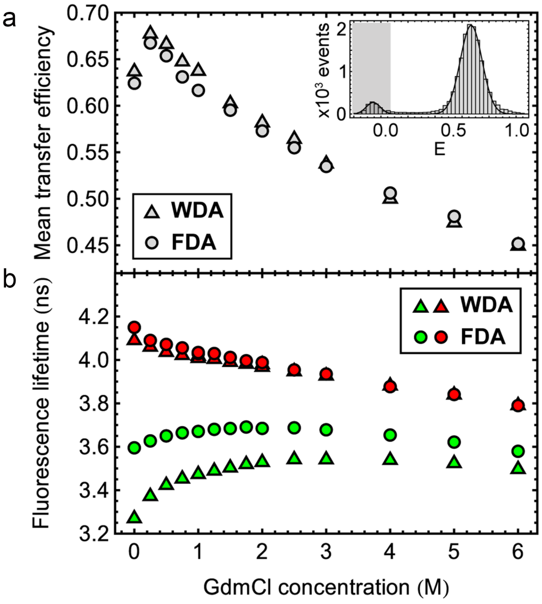


FIG. 3. (a) Mean transfer efficiencies obtained from single-molecule measurements of WDA and FDA at different GdmCl concentrations. The inset shows a transfer efficiency histogram of FDA recorded at 0.25M GdmCl. The small peak at zero transfer efficiency (shaded) originates from molecules without an active acceptor dye. The transfer efficiencies of FDA were used to calculate the root mean squared dye-to-dye distance R_{DA}; the WDA transfer efficiencies are shown for comparison. The systematic error of the measured transfer efficiency is ±0.03.[21,31] (b) Alexa 488 (green) and Alexa 594 (red) fluorescence lifetimes measured at different GdmCl concentrations for the WDA and FDA variants. Fluorescence lifetimes were obtained from single-exponential fits to the decay histograms (Fig. S3 of the supplementary material). Donor-only constructs (WD/FD) were used to obtain the Alexa 488 fluorescence lifetimes. For Alexa 594 fluorescence lifetimes, the acceptor was excited directly in double-labeled constructs (WDA/FDA). The statistical error estimated from four independent measurements is within the size of the symbols. The decrease in fluorescence lifetime at high GdmCl concentrations is partly caused by dynamic quenching of the dyes by GdmCl (Fig. S4 of the supplementary material).

obtained by independent measurements: the root mean squared dye-to-dye distance R_{DA} ($\sqrt{\langle r^2 \rangle}$, where r is the dye-to-dye distance) is calculated from transfer efficiency histograms of the FDA variant [Fig. 3(a), Figs. S1 and S2 of the supplementary material], where quenching of the FRET dyes by Trp is absent, as described in the Appendix (from the histograms, we also determine the fraction of molecules without an active acceptor). Note the slight compaction of the chain upon the addition of small GdmCl concentrations due to charge screening, followed by chain expansion due to preferential interactions of the chain with denaturant at higher GdmCl concentrations.[29] The excited-state relaxation rates k_D and k_A, which are modulated by dynamic quenching of the dyes by Trp, are measured in ensemble fluorescence lifetime decays of the respective constructs [Fig. 3(b), Fig. S3]. Triplet decay times were determined for the FDA variant and fixed to the same values for all measurements and protein variants (see Sec. II A; note that triplet dynamics occur on a time scale that is well-separated from the dynamics of interest). These independent constraints and the simultaneous fit of all three correlation curves are essential for the robustness of the fitting procedure. The only free fit parameters besides D, k_{on}, and k_{off} are thus the absolute FCS amplitudes (corresponding to the inverse of the average number of molecules in the confocal volume) and the triplet amplitudes.

Figure 4 shows the FCS curves recorded for all three integrase variants (WDA, FDA, and WD) at 0M and 6M GdmCl as examples and illustrates the dynamics occurring on the time scale of ∼10 to 100 ns. For WDA at 0M GdmCl, we observe both FRET dynamics and PET quenching in the three correlation functions G_{DD}, G_{AA}, and G_{DA}. The blue lines show the global fit of the data with the 6-state photon statistics model (Sec. II A, Fig. 1). From the fit, we obtain the inter-dye diffusion coefficient D and the on- and off-rate coefficients k_{on} and k_{off} for PET quenching of Alexa 488 by Trp. To highlight the effect of PET quenching on the FCS curves, we additionally show the curves expected in the absence of PET quenching by calculating the model functions of the fit with k_{on} and k set to zero (black dashed lines). This comparison illustrates that both PET quenching and FRET dynamics generate a correlated component (i.e., with positive amplitude) in the autocorrelation curves (G_{DD} and G_{AA}). However, they contribute with opposite signs to the cross correlation curve (G_{DA}): PET quenching is correlated, whereas FRET dynamics are anticorrelated (negative amplitude). At 6M GdmCl, where the IDP chain is more expanded (see Fig. 3) and the complex between Trp and Alexa 488 is strongly destabilized (Fig. S6 of the supplementary material), we observe no positive component in the cross correlation curve, which indicates weak or absent Trp-quenching; the blue and the black dashed lines including and excluding PET quenching, respectively, are virtually indistinguishable.

In the absence of Trp (FDA variant), the FCS data can be fitted with a simpler model with four states accounting for FRET dynamics only [Fig. 4(b)],[9,23] leaving D as the only relevant fit parameter reflecting IDP dynamics. At 6M GdmCl, the FDA correlation data are indeed very well described by the 4-state model. At 0M GdmCl, the fit is still reasonable but deviates slightly from the measured data. However, including PET quenching of the donor dye in the model does not improve the fit (see Fig. S7 of the supplementary material). A small contribution to the observed deviation is dye-dye quenching,[32,33] which we investigated with a proteolytically cleaved FDA construct (Fig. S8). Cleavage leads to a slightly reduced amplitude in G_{AA} measured upon direct excitation of the acceptor dye, which indicates the presence of a small component of dye-dye quenching in the double-labeled construct. Interestingly, however, we observe qualitatively similar deviations between the fit and the data when the fluorescence correlation functions are calculated based on a 40 μs atomistic simulation of integrase[22] (Fig. S7). This indicates that the inter-dye distance dynamics are not well described by a single time scale but contain an additional component. In the absence of FRET but in the presence of Trp-quenching [WD variant, see Fig. 4(c)], the donor autocorrelation can be fitted with an even simpler 3-state model that includes only the electronic ground and excited states of the donor dye and its quenched state. In this case, k_{on} and k_{off} remain as free fit parameters. We observe a pronounced positive amplitude at 0M GdmCl that vanishes almost completely at 6M GdmCl, where the Alexa 488-Trp complex is destabilized so much that PET quenching becomes virtually undetectable (see Fig. S6).

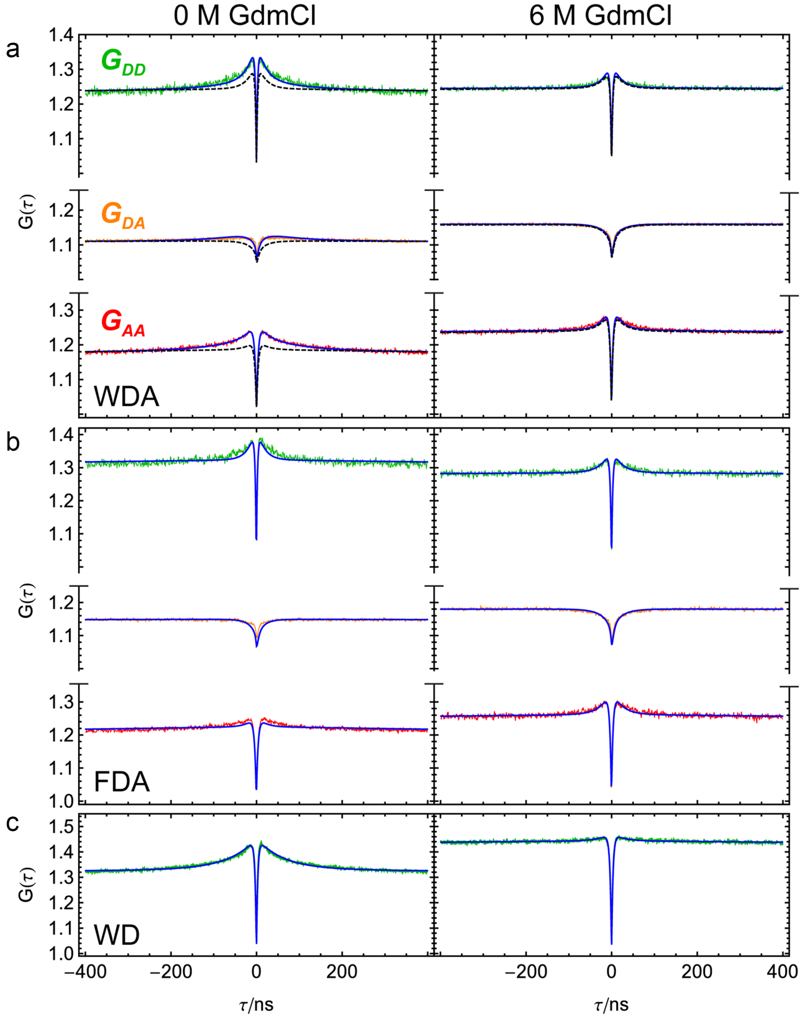


FIG. 4. FCS curves of the three integrase variants measured at 0M and 6M GdmCl illustrating the dynamics in the range of ∼10–100 ns. (a) The three correlation curves of the WDA variant (G_{DD}, G_{DA}, and G_{AA}) were fitted globally (blue lines) with a six-state photon statistics model (see Sec. II A) to obtain D, k_{on}, and k_{off} for both measurements. To illustrate how the data would appear in the absence of Trp-quenching, we also show curves calculated from the model with k_{on} set to zero, but all other parameters kept as obtained from the global fit (dashed lines). (b) The correlation curves of the FDA variant were fitted globally with a reduced photon statistics model without Trp-quenching, where D is the only fit parameter that describes chain dynamics. (c) The donor autocorrelation of the WD variant was fitted with a three-state photon statistics model to obtain the rates k_{on} and k_{off} for PET quenching of Alexa 488 by Trp. The rapid rise in the time range of the fluorescence lifetime (a few nanoseconds) reflects photon antibunching.[25]

E. Long- and short-range dynamics can be accurately quantified over a wide range of denaturant concentrations

The comparison of the best-fit parameters over the whole range of GdmCl concentrations between 0 and 6M is detailed in Fig. 5. Fitting of the FCS curves was performed as described above, the raw data and fits are compiled in Figs. S8-S14 of the supplementary material. To assess the reliability of extracting long-range dynamics with the 6-state model (Sec. II A, Fig. 1), we compare in Fig. 5(a) the inter-dye diffusion coefficient D obtained from fitting the data of the WDA variant with the 6-state model and the FDA variant with the 4-state model without PET. In both variants, the absolute values of D deviate by less than 25% at all denaturant concentrations. Likewise, very similar trends are observed for both WDA and FDA: up to 0.5M GdmCl, D decreases, as expected from its compaction due to charge screening;[29] above that concentration, an increase in D is observed, which levels off at high GdmCl concentrations.

To evaluate how well the PET-quenching parameters k_{on} and k_{off} are fitted from the WDA FCS data, the appropriate control is the WD variant, where no FRET acceptor is present. A robust measure of PET quenching is the equilibrium constant, $K_{eq} = k_{on}/k_{off}$ [Fig. 5(b)], as it is obtained from the amplitude of the FCS curves.[20] WDA and WD are in remarkable agreement for this parameter. The individual quenching rate coefficients also agree quite well and exhibit the same trends in their dependence on the GdmCl concentration. k_{on} [Fig. 5(c)] decreases with increasing denaturant concentration, as contact formation between the dye and Trp becomes less frequent with increasing viscosity of the solution and expansion of the IDP [Fig. 3(a)]. k_{off} [Fig. 5(d)] instead increases—in accordance to the observations made for the quenching of the oxazine dye MR121 by free Trp.[16,20] Note that the values of k_{on} and k_{off} become increasingly uncertain at GdmCl concentrations above 3M, when the PET-quenching component of the FCS curves becomes too small to be reliably discriminated from noise (cf. Fig. S11 of the supplementary material).[16]

In conclusion, we demonstrate that reliable estimates of D, k_{on}, and k_{off} can be obtained from fitting FCS curves of the WDA variant over a wide range of denaturant concentrations. Apart from these parameters, only the absolute FCS amplitudes and the amplitudes for the triplet-blinking component were left unconstrained in the fits. k_D, k_A, k_{exD}, k_{exA}, and R_{DA} were obtained from independent measurements. In particular, we used the R_{DA} values obtained from single molecule FRET efficiency histograms [Fig. 3(a)] for fitting the correlation curves. Surprisingly, however, the value of R_{DA} has

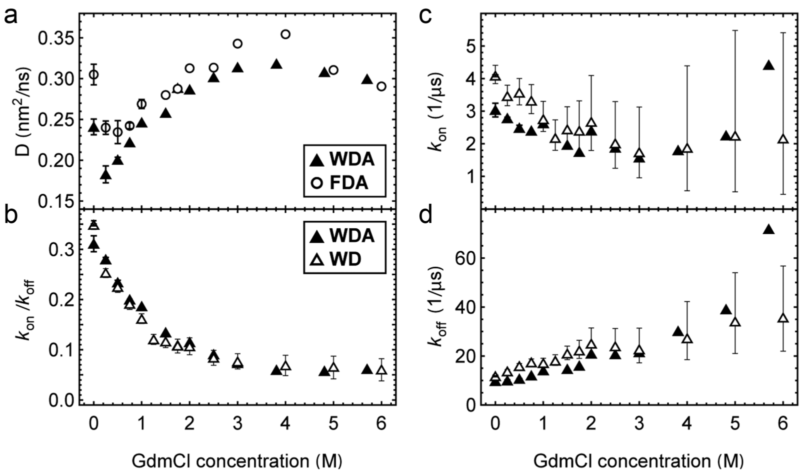


FIG. 5. Comparison of best-fit parameters (D, k_{on}, and k_{off}) obtained for the WDA variant (full triangles) with the corresponding values for the control variants FDA (open circles) and WD (open triangles). (a) Inter-dye diffusion coefficients, (b) equilibrium constants for quenching, (c) on-, and (d) off-rate coefficients. The parameters for WDA were obtained from global fits of the correlation functions with the 6-state model (Fig. 1). For FDA, a 4-state model without quenching was employed, and for WD, a three-state model accounting only for PET quenching was used. Thick error bars indicate the standard deviation of 2-3 independent measurements (cf. Figs. S9-S11 of the supplementary material), where available. Thin error bars show the standard deviation of 1000 fits of the WD data when the input parameters were varied by their estimated uncertainty of ±5%.

a sufficiently strong influence on the shape of the FCS curves that it is even possible to treat R_{DA} as an additional fit parameter and still obtain virtually the same results for D, k_{on}, and k_{off} (see Fig. S15 of the supplementary material). Consistent with this finding, we retrieve R_{DA} values that are in remarkable agreement with those derived from the mean transfer efficiencies (Fig. 6). Even the slight collapse of integrase between 0 and 0.25M GdmCl[29] can be recovered with this approach, attesting to the robustness of the fit.

F. Intramolecular contact rates between dye and quencher

The remaining step is the conversion of the association rate k_{on} between the dye and quencher into an intramolecular contact rate. The observed rate of 3-4 μs^{-1} (from the fit of the WDA and the WD construct) constitutes a lower limit for this process, since not all collisions between Alexa 488 and Trp result in the formation of the statically quenched complex. The quenching efficiency amounts to 11%-16% when both the dye and quencher are attached to a flexible polypeptide linker.[21] Using this correction factor, we arrive at an intramolecular contact rate of 18-36 μs^{-1}. The contact rate can be quantified independently by comparing the donor fluorescence lifetime τ_D of the WD and the FD constructs ($\tau_{D,WD}$ = 3.27 ns and $\tau_{D,FD}$ = 3.59 ns, respectively). The reduction of τ_D in the WD construct is caused by dynamic quenching of Alexa 488 by Trp, which is commonly assumed to be diffusion-limited.[20,21] The contact rate between the dye and Trp can thus be obtained from $k_{contact} = (\tau_{D,WD})^{-1} - (\tau_{D,FD})^{-1}$. The resulting value of 27 ± 2 μs^{-1} is in agreement with the contact rate estimated from PET quenching, attesting to the reliability of the approach.

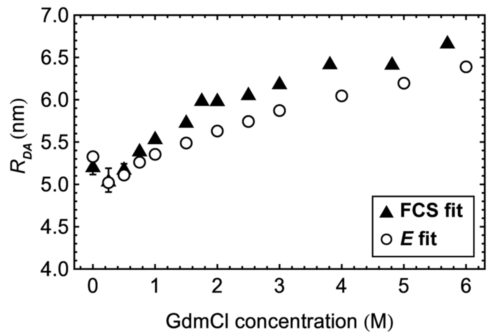


FIG. 6. Root mean squared dye-to-dye distances R_{DA} obtained from fits of the 6-state model to the WDA FCS curves, where in addition to D, k_{on}, and k_{off}, R_{DA} is a free fit parameter (triangles). For comparison, the values determined from transfer efficiency histograms of FDA (Fig. S2 of the supplementary material) are shown (circles). The best-fit parameters of D, k_{on}, and k_{off} are within the error of those in Fig. 5 (see Fig. S15). Error bars indicate the standard deviation of 2-3 independent measurements, where available.

G. Indications for local structure formation in molecular dynamics simulations and experiment

So far, we have analyzed the data assuming an idealized unfolded protein, whose dynamics are modeled by diffusion in a potential of mean force corresponding to a Gaussian chain, a course-grained description based on polymer physics that is very useful for quantifying the global, long-range properties of unfolded and intrinsically disordered proteins.[34] However, the deviations between the FDA correlation functions and a fit with this model point towards a contribution of more complex dynamics in the real protein and call for a more detailed model. The most detailed models of proteins currently available for describing their dynamics on the time scales relevant for our experiments are all-atom explicit solvent molecular dynamics (MD) simulations. Here, we take advantage of extensive recent MD simulations of integrase, which use the Amber 12 force field together with the TIP4P-D water model, a combination that has been very successful in reproducing experimentally observed dimensions and dynamics of unfolded and intrinsically disordered proteins.[16,22] We calculated several experimentally relevant observables from the 40-μs MD trajectory (provided by Shaw Research): the distance distributions and average distances between the amino acid residues used in the measurements, the distance correlation function between residues Lys11 and Cys60 (to extract the intrachain diffusion coefficient D monitored by FRET), as well as the contact rates between Lys11 and Trp23 and between Cys60 and Trp23 (reflecting dynamic quenching between the fluorophores and Trp). Note that the simulation does not include the FRET dyes, so the most distal non-hydrogen atoms of the respective side chains of Lys11 (N^6) and Cys60 (S^3) were used for the analysis. The corresponding average distance between Lys11 and Cys60 in the simulation is within ∼12% of the value determined from

TABLE II. Comparison of distances and dynamic parameters derived from experiment and MD simulation for the residue pairs specified.

Part of MD trajectory analyzed	$D_{11,60}$ (nm^2/ns)	$R_{11,60}$ (nm)	Contact rate 11,~~60~~ (μs^{-1})	Contact rate 23,60 (μs^{-1})
0-40 μs	0.29 ± 0.13[a,b]	4.43 ± 0.25[b]	4.8 (−1.5/+1.8)[c]	0.7 (−0.2/+0.6)[c]
10.2-26 μs	0.42 ± 0.10[a,d]	4.33 ± 0.24[b]	32 (−21/+23)[c]	0.4 (−0.1/+0.4)[c]
0-10.2 and 26-40 μs	0.23 ± 0.05[a,d]	4.51 ± 0.36[b]	3.1 (−0.7/+0.9)[c]	1.6 (−0.5/+1.0)[c]
Experiment	0.30 ± 0.02[e]	5.0 ± 0.2[f]	27 ± 2[g]; 18-36[h]	3.0 ± 2.0[g]

[a] Intrachain diffusion coefficient D was obtained by globally fitting the correlation curves G_{DD}, G_{AA}, and G_{DA} (calculated from the MD trajectory as described in the Appendix) with a 4-state photon statistics model (see Sec. II B).

[b] Errors correspond to the standard deviation estimated from the analysis of four 10-μs segments obtained by splitting the MD trajectory.

[c] Errors reflect a 10% variation of the contact radius (1.0 ± 0.1 nm).

[d] Errors report on the spread of the values estimated from the analysis based on splitting the MD trajectory into two segments of equal duration.

[e] Standard deviation estimated from 3 independent experiments.

[f] The experimental root-mean-squared dye-to-dye distance R_{DA} was rescaled according to Gaussian chain statistics to account for the absence of dyes in the simulation (see the Appendix). The error reflects the systematic uncertainty in FRET efficiency as estimated from instrument-to-instrument variation.[31]

[g] Contact rates determined from fluorescence lifetimes according to $k_{contact} = (\tau_{D,WD})^{-1} - (\tau_{D,FD})^{-1}$ (see the main text). The error corresponds to the standard deviation of 0.02 ns estimated from four independent fluorescence lifetime decays.

[h] Contact rate estimated from the PET quenching association rate k_{on}. The range encompasses the uncertainty associated with k_{on} (3-4 μs^{-1}) and the quenching efficiency (11%-16%). For contact formation between Trp23 and Alexa 594 at position C60, an estimate of the rate is only possible using fluorescence lifetimes, since the FCS amplitudes associated with PET quenching are too small (Figs. S5, S12, and S13 of the supplementary material).

the FRET efficiency (Table II), indicating a suitable balance between intra-protein and protein-solvent interactions in the simulation.[22,35] The correlation functions and contact rates were calculated from the MD trajectory as described in the Appendix; a contact radius of 1.0 ± 0.1 nm was used, in line with previous estimates.[15,16,36] The results are summarized in Fig. 7.

Already from the inter-residue distance trajectories and distributions [Fig. 7(b)], deviations from the simple picture of a polymer devoid of specific intra-chain interactions are apparent, especially for the shortest segment probed experimentally: Whereas the distance distributions between residues 11 and 60 [$P(r_{11,60})$] and between residues 23 and 60 [$P(r_{23,60})$] are reasonably well approximated by $P(r)$ of a Gaussian chain, $P(r_{11,23})$ is clearly bimodal, with the average distance for the largest part of the simulation being much shorter than predicted by simple rescaling of a Gaussian chain, indicative of structure formation in this region of the IDP. Indeed, the analysis of secondary structure in the MD simulation indicates the formation of a long-lived α-helix between residues 16 and 25 [Fig. 7(a)]. Since the helix persists for up to ∼25 μs in the simulation, even this 40-μs trajectory (which is at the limit of current feasibility) is thus clearly insufficient for reaching full convergence, but intrachain dynamics faster than helix formation can be extracted from the analysis of selected segments of the simulation: When the helix is formed and residues 11 and 23 are in close proximity, their contact rate of 32 μs^{-1} is correspondingly high, remarkably similar to the experimentally observed contact rate of 27 μs^{-1} for contact quenching between these sites (Table II). When no helix is formed, the contact rate is 10 times lower. This difference is also reflected by the decays of the survival probability of the unquenched state for the different trajectory segments [Fig. 7(d)], which closely correspond to the measured correlation curves due to PET.[16] The similarity of time scales of the fast decay calculated from the segment of the MD trajectory where the helix is formed and the experimentally observed rate suggests that the helix is predominantly formed under the experimental conditions.

As expected, the contact rate between the more sequence-distant residues 23 and 60 is less affected by structure formation [Table II, Fig. 7(d)]. The contact rates calculated from the MD trajectory are lower than the experimentally observed contact rate, but considering the large experimental error due to the small change in fluorescence lifetimes for this long segment [Fig. 3(b)], the agreement is still reasonable. The difference between MD trajectory segments with and without the helix present is even less pronounced for the donor fluorescence correlation function calculated assuming FRET between Lys11 and Cys60 [Fig. 7(c)]; correspondingly, all values of the inter-residue diffusion are remarkably close to the experimental result (Table II). Interestingly, a global fit of all correlation functions (G$_{DD}$, G$_{AA}$, and G$_{DA}$) calculated from the MD trajectory (Fig. S7 of the supplementary material) shows qualitatively the same deviations from the photon statistics model as the measured FCS curves of the FDA construct.

These observations have several important implications: First, the agreement between experiment and simulation is surprisingly good, reflecting the recent progress in MD force fields for unfolded and intrinsically disordered proteins.[37] Second, the effect of local structure formation is less pronounced for FRET-based FCS measurements of chain dynamics than for contact-based experiments because the relatively long chain segments accessible to single-molecule FRET are less affected by local structure than the contact-formation measurements, which are usually performed on shorter segments and most sensitive to the short-distance tail of the underlying distance distribution.[38] Third, the combination of both types of measurements with simulations

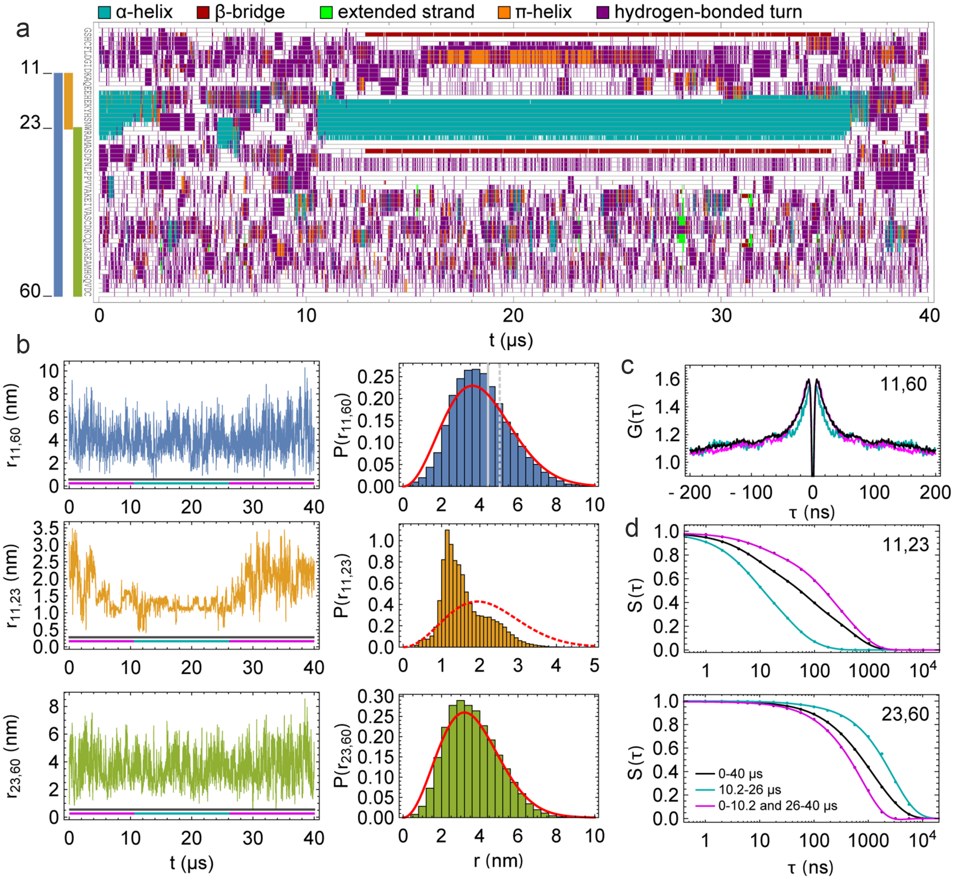


FIG. 7. Local structure formation and dynamics from a 40-μs MD simulation of integrase.[22] (a) Assignment of secondary according to the STRIDE algorithm[39] showing that residues 16-25 are in an α-helical conformation for tens of microseconds. On the left, the dye labeling/quenching positions and experimentally probed sequence stretches are indicated using the color code from (b). (b) Inter-atomic distance trajectories (left; average of 10 consecutive 1 ns time steps) and resulting distance distributions (right, 1 ns time steps) corresponding to the experimentally probed positions (N^6 of Lys11; center of mass of Trp23 indole ring; S^3 of Cys60). Whereas the distributions of $r_{11,60}$ and $r_{23,60}$ are reasonably well approximated by $P(r)$ for a Gaussian chain (red lines), $r_{11,23}$ shows clear deviations due to local structure formation (the dashed red line shows the prediction from rescaling $R_{11,60}$ according to Gaussian chain statistics). For $P(r_{11,60})$, the root mean squared inter-residue distance $R_{11,60}$ $\left(\sqrt{\left\langle r_{11,60}^2\right\rangle}\right)$ and the experimentally derived distance R_{DA} are depicted as full and dashed gray vertical lines, respectively. For further analysis, the MD trajectory was separately analyzed in the segment from 10.2 to 26 μs (corresponding to the compact state with helix 16-25 formed and $r_{11,23} < 1.5$ nm) and the combined segments from 0 to 10.2 μs and from 26 to 40 μs, as indicated below the distance trajectories. (c) Donor-donor fluorescence intensity autocorrelation G_{DD} calculated based on the whole MD trajectory (black), the 10.2-26 μs segment (cyan) and the 0-10.2 and 26-40 μs segments (magenta) (see the Appendix for details of simulation of fluorescence emission). (d) Survival probability $S(\tau)$ of the unquenched state for the three trajectory segments for contact formation between Lys11 and Trp23 (top) and Cys60 and Trp23 (bottom). For 11 and 23, the decay of the survival probability calculated from the whole MD trajectory (black) is biphasic. Analyzing the trajectory separately for the segments with and without helix 16-26 formed separates the time scales. For the 10.2 to 26 μs segment (cyan), contact formation is very fast, whereas it is much slower for the 0 to 10.2 μs and 26 to 40 μs segments (magenta). These differences are less pronounced for contact formation between Cys60 and Trp23 [see the Appendix for details of calculating $S(\tau)$].

enables a much more detailed picture of the properties of unfolded proteins than previously available. In the case of integrase investigated here, a complete structural interpretation of the experimental results would not have been possible without the simulations, and at the same time, the experimental results provide a stringent benchmark for the simulations.

III. DISCUSSION

In this work, we describe the analysis of fluorescence fluctuations caused by long-range and short-range distance dynamics in the same protein molecule. For probing long-range dynamics, we labeled the IDP integrase with the FRET pair Alexa 488 and Alexa 594. An intrinsic Trp residue of the protein close in sequence to Alexa 488 can additionally form non-fluorescent complexes with the donor dye and thus reports on short-range contact dynamics. By applying a photon statistics model encompassing both FRET and PET, we are able to analyze the resulting FCS curves and determine the underlying kinetic parameters of both long-range dynamics and contact formation occurring simultaneously. Even without constraining the transfer efficiency based on independent measurements, we are able to quantify the distances between the FRET donor and acceptor of the IDP within the same framework from the analysis of the FCS curves. Our results thus demonstrate that information on distances and both FRET- and PET-based dynamics can be extracted in a robust fashion even if they occur within the same molecule. This approach hence also enables a quantitative treatment of ns-FCS data in the case of undesired quenching of FRET dyes by Trp residues in a protein.

We utilize the effects of GdmCl on integrase to assess the feasibility of detecting systematic changes in intramolecular distances and dynamics expected with changes in denaturant concentration.[29,34] The results of GdmCl addition on chain expansion and dynamics we observe align well with previous investigations on chain dynamics in unfolded proteins and IDPs.[40,41] At high denaturant concentrations, integrase is highly expanded owing to favorable interactions of the protein with GdmCl. With decreasing denaturant concentration, the IDP continuously collapses, down to 0.25M GdmCl, followed by a slight re-expansion in the GdmCl-free buffer due to the negative net charge of the protein and the resulting intra-chain electrostatic repulsion in the absence of charge screening.[29] Importantly, this behavior is recovered even when treating R_{DA} as a free fit parameter in the analysis of the FCS curves, which corroborates the robustness of the analysis.

The intrachain diffusion coefficient D, reflecting the reconfiguration dynamics of the IDP, follows the same trend as the dimensions of the IDP, in keeping with the previously observed correlation between chain compaction and internal friction that slows down dynamics in unfolded proteins.[8,16,34] Similarly, k_{on} for PET quenching decreases by 30%-40% between 0 and 2M GdmCl (the range where it can be determined reliably from the FCS data) (Fig. 5), and the fluorescence lifetime of Alexa 488, which reports on dynamic quenching, increases from 0 to 2M GdmCl (Fig. 3). A more expanded chain thus also leads to less contact formation, as expected.

However, a more detailed structural interpretation of the results, including the deviations from dynamics based on simple polymer reconfiguration or two-state contact kinetics observed here, benefits from a more detailed molecular model. The traditional focus on folded proteins for the parameterization of the energy functions for MD simulations has often resulted in unfolded states that were much more compact than observed experimentally.[22,35] The increasing interest in unfolded and intrinsically disordered proteins, however, has triggered developments to remedy this deficiency,[37] in particular, the optimization of suitable implicit[42,43] and explicit water models.[22,35,44,45] The increasing availability of realistic simulations of unfolded proteins in atomic detail at experimentally relevant time scales[37] provides an exciting opportunity for the direct comparison of experiment and simulation.

Diffusion in a potential of mean force based on analytical polymer models is a good description for unfolded proteins without persistent structure: in the present case, such a model describes the dynamics of integrase in 6M GdmCl and of the very charged and structure-less IDP ProTα well (Fig. S7 of the supplementary material), as in other cases investigated previously.[34] In the case of integrase in the absence of denaturant, however, deviations between such a simple model and the correlation functions become apparent, raising the question of a possible influence of local structure in the protein. Remarkably, the experimentally observed characteristics of the correlation functions are recapitulated very well by the MD simulation and can be explained by slow dynamics and a persistent α-helix in the segment of integrase probed by PET. We thus conclude that fluorescence correlation curves of a system with both short- and long-range dynamics are able to report on the presence of slow protein dynamics and long-lived secondary structure.

IV. CONCLUSIONS

Our results show that the direct combination of FRET and PET in the same molecule can be analyzed quantitatively and that it provides a useful tool for studying the dimensions and dynamics of unfolded and intrinsically disordered proteins by offering complementary information about protein dynamics within the very same system. The increasing availability of realistic MD simulations of IDPs on the same time scales accessible in such experiments enables a more detailed interpretation of the results in the present case in terms of specific structure formation in the protein. Our findings also suggest that the technique is a very promising and stringent benchmark for atomistic simulations and will thus aid in the development of force fields. Altogether, the close combination of experiments of increasing sensitivity, simple theoretical models that enable a quantitative analysis of the data, and molecular simulations revealing structural detail is a promising route towards a better understanding of the behavior of IDPs, whose broad spectrum of biological functions and underlying mechanisms we are only starting to recognize.

SUPPLEMENTARY MATERIAL

See supplementary material for the compiled raw data (transfer efficiency histograms, FCS curves, fluorescence lifetime decays); control experiments for acceptor and dye-dye quenching; the influence of GdmCl on the lifetime of Alexa 488 and static quenching between Alexa 488 and Trp; a fit of simulated fluorescence correlation curves; and best-fit parameters obtained with R_{DA} left unconstrained.

ACKNOWLEDGMENTS

We thank D. E. Shaw Research, especially Stefano Piana-Agostinetti and Rebecca Bish-Cornelissen, for providing access to their MD trajectories. Mass spectrometry was performed by Serge Chesnov at the Functional Genomics Center Zurich. This work was supported by the Swiss National Science Foundation (B.S.).

APPENDIX: MATERIALS AND METHODS

1. Protein preparation and labeling

Four different integrase (IN) variants were prepared by site-directed mutagenesis of the IN DNA sequence cloned into a pET15b vector. For labeling with fluorescent dyes, we substituted Lys11 by Cys and added a cysteine at the C-terminus of the protein sequence (Cys60). Additionally, Tyr19 from the wild-type sequence was replaced by Ala[46] to minimize any additional quenching of Alexa 488.[19] The resulting sequence was GSHMFLDGID C_{11}AQEEHEKAH SNW_{23}RAMASDF NLPPVVAKEI VASCDKCQLK $GEAMHGQVDC_{60}$ (variant WDA), where the residues with subscripts indicate the labeling positions and the quenching Trp. Furthermore, a control construct was produced with Trp23 exchanged to Phe (variant FDA).

IN protein constructs were expressed in *E. coli* BL21(DE3) and purified on a Ni-NTA agarose column (Ni-NTA Agarose, Qiagen) using a cleavable N-terminal hexahistidine tag (His-Tag). The eluate was dialyzed against 50 mM $NaHCO_3$ pH 9.3, 0.5M Na_2SO_4, 0.1M NaCl, 5 mM EDTA, 1 mM DTT, and the His-Tag was cleaved off with thrombin (Thrombin from bovine plasma, SERVA Electrophoresis). A second Ni-NTA affinity chromatography was performed to remove uncleaved IN and the His-Tag. In a final step, the eluate was reduced with 10 mM TCEP and purified with reversed-phase chromatography on a C18 column (XTerra RP18 Column, 5 μm, 4.6 × 250 mm, Waters), using a ddH_2O + 0.1% trifluoroacetic acid/acetonitrile gradient. Proteins were lyophilized and dissolved just prior to labeling. The molecular mass was confirmed by mass spectrometry.

The thiol groups of Cys11 and Cys60 were labeled with maleimide-functionalized fluorescent dyes (Alexa Fluor 488 C_5 and Alexa Fluor 594 C_5, Thermo Scientific). All labeling reactions were conducted in 50 mM Tris pH 7.5, 0.5 mM $ZnCl_2$, and 0.5M arginine (labeling buffer). $ZnCl_2$ was included to avoid labeling of the two intrinsic cysteines C44 and C47, which form a high-affinity complex with Zn^{2+} ions under these conditions and thus strongly decrease their reactivity towards maleimide. Site-specific labeling was achieved by exploiting the different reactivities of Cys11 and Cys60 in combination with reverse phase chromatography purification. (Cys60 is less reactive than Cys11 presumably because its pK_a has increased as a result of two neighboring negative charges from Asp59 and the C-terminal carboxylate.)[47] In the first step, 50-100 nmol of lyophilized protein was dissolved in labeling buffer to a concentration of 200 μM and reacted with Alexa 488 designated for the Cys11 position in a 1:0.8 molar ratio. The single-labeled protein was reduced with 10 mM TCEP, purified with reverse-phase chromatography as described above and lyophilized. In the second step, the procedure was repeated analogously with Alexa 594, which was coupled to Cys60. In this way, the site-specifically labeled WDA and FDA variants were generated with purities of >95%. For the control constructs WD and FD, only Alexa 488 was coupled to Cys11. The identity of the labeling positions was verified with tryptic digest and subsequent electrospray mass spectrometry of the fragments.

All fluorescence measurements were conducted in 50 mM Tris-HCl buffer (pH 7.5) containing 0.5 mM EDTA, 143 mM 2-mercaptoethanol, and 0.01% (v/v) Tween 20 (referred to as "measurement buffer"). GdmCl concentrations were adjusted by adding an appropriate volume of 7.2M GdmCl (Pierce), 50 mM Tris-HCl, pH 7.5 to the measurement buffer.

2. Factor Xa protease digest

To test the influence of dye-dye quenching, 10 nM FDA in 50 mM Tris pH 7.5, supplied with 5 μM EDTA and 0.01% Tween 20 (total volume 50 μl), was digested by the addition of 4 U Factor Xa protease (Novagen). After 30 min at room temperature, a sample was diluted 100fold in Tris buffer to confirm complete digestion with single-molecule spectroscopy. As expected, no bursts with a transfer efficiency >0.1 could be detected after Alexa 488 excitation. Subsequently, a sample of the digest was diluted tenfold in measurement buffer to record FCS curves after direct acceptor excitation using a helium-neon (He-Ne) laser at 594 nm. Inactivation of Factor Xa protease during the measurements was ensured by 2-mercaptoethanol contained in the measurement buffer.

3. Fluorescence lifetime measurements

Fluorescence lifetime decays were measured with a custom-built ensemble time-correlated single-photon counting instrument described previously.[30] The system was equipped with a 20 MHz pulsed laser (SCF450-2, Fianium) and the appropriate excitation wavelengths for Alexa 488 and 594 were selected with HQ 470/40 and z582/15 (Chroma) filters, respectively. Dye and protein samples were measured at a concentration of 50 nM in measurement buffer at the appropriate GdmCl concentration. The instrument was used in magic-angle configuration for all measurements. Donor-only labeled constructs (WD/FD) were used to measure the fluorescence lifetime of Alexa 488; the fluorescence lifetime of Alexa 594 was measured on donor-acceptor labeled samples by direct excitation of the acceptor dye. The recorded fluorescence lifetime decays were fitted with single exponential decays convolved with the instrument response function obtained from the measurement of scattered laser light.

4. Single molecule instrumentation and measurements

Single-molecule fluorescence data were recorded using a MicroTime 200 (PicoQuant) confocal microscope equipped with four detectors in a configuration described previously.[8,21] Transfer efficiency histograms and FCS data were recorded at sample concentrations of 100 pM and 1-3 nM, respectively, in measurement buffer at the appropriate GdmCl concentration. FRET constructs and donor-only controls were excited with a diode laser (LDH-D-C-485, PicoQuant) at a power of 100 μW (measured at the back aperture of the objective). Acceptor dye direct excitation was carried out with a He-Ne laser at a power of 40 μW (594 nm, CWI Melles Griot). Transfer efficiency histograms were recorded for 1 h, FCS curves for 12-16 h. The presence of a folded subpopulation of IN could be excluded based on the absence of a high-transfer efficiency peak in the E histograms.[29] The calculation of the FCS curves from photon arrival times on the four detectors is detailed in the work of Haenni *et al.*[21] The FCS data were analyzed with a time binning of 1 ns.

5. Analysis of transfer efficiency histograms and mean squared dye-to-dye distances

Transfer efficiency histograms were built from transfer efficiencies E calculated from the measured photon bursts (burst detection criterion: 50 consecutive photons with an interphoton time of less than 150 μs) according to

$$E = \frac{n_A}{n_A + \gamma n_D},$$

where n_D and n_A are the number of donor and acceptor photons of the individual bursts corrected for background, cross talk between the detection channels, and acceptor direct excitation.[21] It is important to note that the value of the correction factor γ (defined as $\gamma = \xi_A Q_A / \xi_D Q_D$) needed adaptation for each data point at different GdmCl concentrations due to changes of the quantum efficiencies of the dyes with denaturant concentration. As described previously,[21] we used the relation $\gamma = \gamma_0(\tau_{D0}/\tau_D)(\tau_A/\tau_{A0})$, where γ_0 and the mean donor and acceptor fluorescence lifetimes τ_{D0} and τ_{A0} were determined for the free dyes in measurement buffer.[21] $\tau_D = k_D^{-1}$ and $\tau_A = k_A^{-1}$, the mean fluorescence lifetimes of the dyes bound to the IN variants, were measured under the same conditions as the transfer efficiency histograms (see Subsection 3 of the Appendix). Transfer efficiency histograms were fitted with two Gaussian peak functions (one for the donor-only peak and one for the FRET peak, respectively) to extract mean transfer efficiencies and the fraction of molecules lacking an active acceptor dye. We converted those values to root mean squared dye-to-dye distances assuming the probability distribution function of a Gaussian chain.[30] The Förster radius R_0 was calculated to be 5.6 nm for the free dye pair Alexa 488 and Alexa 594 at 0M GdmCl. For each data point, R_0 was corrected to account for refractive index changes upon adding GdmCl as well as changes in Q_D caused by dynamic quenching of Alexa 488 when bound to the protein ($Q_D = \tau_D/\tau_{D0}$, where τ_D is the mean fluorescence donor lifetime measured under the same conditions as the transfer efficiency histograms).

6. Analysis of MD simulations

A 40-μs trajectory of unfolded integrase simulated with the Amber 12 force field and the TIP4P-D water model[22] was kindly provided by Shaw Research as atomic coordinates saved with a time resolution of 1 ns. Side chain atoms (N^6 of Lys11, S^3 of Cys60, the center of mass of the indole group of Trp23) were used to calculate distances within the chain and contact formation times. The contact formation time t_c was calculated as outlined in the work of Soranno *et al.*,[16] by integrating the survival probability of the unquenched state $S(\tau)$: $t_c = \int_0^\infty S(\tau)d\tau$, with $S(\tau)$ given by

$$S(\tau) = \left\langle \exp\left(- \int_{\tau_0}^{\tau_0+\tau} k_0 \theta\left(R_c - r_{i,j}(t)\right) dt \right) f_q(\tau_0) \right\rangle_{\tau_0} \Big/ \left\langle f_q(\tau_0) \right\rangle_{\tau_0}.$$

Here, the average over time origins, τ_0, is defined as $\langle \cdots \rangle_{\tau_0} = \frac{1}{T}\int_0^T \cdots d\tau_0$, with T denoting the length of the trajectory. $k_0\theta\left(R_c - r_{i,j}(t)\right)$ describes the distance dependence of the quenching rate (approximated by the Heaviside step function θ) between residues $i, j = 11, 23$ and $i, j = 23, 60$, respectively. The term $f_q(\tau_0) = \theta\left(r_{i,j}(\tau_0) - R_c\right)$ eliminates from $S(\tau)$ the small fraction of conformations that are already in contact at τ_0; $\left\langle f_q(\tau_0) \right\rangle_{\tau_0}$ is required for normalization. k_0 is chosen sufficiently large in order to recover t_c in the diffusion limit. The contact radius R_c remains as the only adjustable parameter.

Fluorescence intensity correlation curves were obtained from the MD trajectory by simulating and correlating donor and acceptor photon time-trajectories. For this purpose, the time evolution of the electronic state was simulated in each time step t_i according to the time-dependent rate matrix $\mathbf{K_0}(r(t_i))$. $r(t_i)$ was taken from the MD simulations as the distance between N^6 of Lys11 and S^3 of Cys60. Donor or acceptor photons were emitted each time a radiative emission by the donor or acceptor (defined by $\tilde{\mathbf{V}}_{\mathbf{D}}$ or $\tilde{\mathbf{V}}_A$, respectively) occurred. A constant time step $t_i - t_{i-1} = 0.1$ ns was chosen to be able to reproduce details of the photophysics on this time scale, such as photon anti-bunching. As the time resolution of the MD trajectory was only 1 ns, we interpolated for intermediate time steps. Ideal correlation functions are shown, without explicit treatment of spectral cross talk and the presence of a donor-only species.

7. Rescaling of root mean squared distances

Assuming Gaussian chain statistics, the root mean squared distance $R_n = \sqrt{\langle r^2 \rangle} \propto \sqrt{n}$ between two points in a chain separated by n segments was rescaled to R_m of a chain of m segments with the relation $R_m = R_n\sqrt{m/n}$. We approximate the number of segments as the number of peptide bond equivalents between the two points in the chain probed. In the simulation, the number of segments thus is ~51.3 (N^6 of Lys11 to S^3 of Cys60) and ~14.7 (N^6 of Lys11 to indole of Trp23). For the experimental data, the number of peptide segments between the dyes was estimated to be 58, using the value of 4.5 peptide bonds between C^α and the center of mass of the dye previously estimated from a global analysis of a large data set of different segment lengths labeled with the same fluorophores used here.[48]

[1] A. K. Dunker, I. Silman, V. N. Uversky, and J. L. Sussman, Curr. Opin. Struct. Biol. **18**(6), 756 (2008).

[2] H. J. Dyson and P. E. Wright, Nat. Rev. Mol. Cell Biol. **6**(3), 197 (2005).

[3] T. Mittag, L. E. Kay, and J. D. Forman-Kay, J. Mol. Recognit. **23**(2), 105 (2010).

[4] A. K. Dunker, M. S. Cortese, P. Romero, L. M. Iakoucheva, and V. N. Uversky, FEBS J. **272**(20), 5129 (2005).

[5] V. N. Uversky, C. J. Oldfield, and A. K. Dunker, Annu. Rev. Biophys. **37**, 215 (2008).

[6] Th. Förster, Ann. Phys. **437**(1-2), 55 (1948).

[7] C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, and T. Ha, Annu. Rev. Biochem. **77**, 51 (2008).

[8] A. Soranno, B. Buchli, D. Nettels, R. R. Cheng, S. Muller-Spath, S. H. Pfeil, A. Hoffmann, E. A. Lipman, D. E. Makarov, and B. Schuler, Proc. Natl. Acad. Sci. U. S. A. **109**(44), 17800 (2012).

[9] D. Nettels, I. V. Gopich, A. Hoffmann, and B. Schuler, Proc. Natl. Acad. Sci. U. S. A. **104**(8), 2655 (2007).

[10] B. Schuler, S. Müller-Späth, A. Soranno and D. Nettels, in *Intrinsically Disordered Protein Analysis: Volume 2, Methods and Experimental Tools*, edited by V. N. Uversky and A. Keith Dunker (Springer, New York, 2012), Vol. 896, p. 21.

[11] J. A. Lamboy, H. Kim, K. S. Lee, T. Ha, and E. A. Komives, Proc. Natl. Acad. Sci. U. S. A. **108**(25), 10178 (2011).

[12] O. Bieri, J. Wirz, B. Hellrung, M. Schutkowski, M. Drewello, and T. Kiefhaber, Proc. Natl. Acad. Sci. U. S. A. **96**(17), 9597 (1999).

[13] L. J. Lapidus, W. A. Eaton, and J. Hofrichter, Proc. Natl. Acad. Sci. U. S. A. **97**(13), 7220 (2000).

[14] H. Neuweiler, A. Schulz, M. Bohmer, J. Enderlein, and M. Sauer, J. Am. Chem. Soc. **125**(18), 5324 (2003).

[15] S. Doose, H. Neuweiler, and M. Sauer, ChemPhysChem **10**(9-10), 1389 (2009).

[16] A. Soranno, A. Holla, F. Dingfelder, D. Nettels, D. E. Makarov, and B. Schuler, Proc. Natl. Acad. Sci. U. S. A. **114**(10), E1833 (2017).

[17] G. Rahamim, D. Amir, and E. Haas, Biophys. J. **112**(9), 1786 (2017).

[18] M. Gotz, P. Wortmann, S. Schmid, and T. Hugel, Methods Enzymol. **581**, 487 (2016).

[19] H. M. Chen, S. S. Ahsan, M. B. Santiago-Berrios, H. D. Abruna, and W. W. Webb, J. Am. Chem. Soc. **132**(21), 7244 (2010).

[20] S. Doose, H. Neuweiler, and M. Sauer, ChemPhysChem **6**(11), 2277 (2005).

[21] D. Haenni, F. Zosel, L. Reymond, D. Nettels, and B. Schuler, J. Phys. Chem. B **117**(42), 13015 (2013).

[22] S. Piana, A. G. Donchev, P. Robustelli, and D. E. Shaw, J. Phys. Chem. B **119**(16), 5113 (2015).

[23] D. Nettels, A. Hoffmann, and B. Schuler, J. Phys. Chem. B **112**(19), 6137 (2008).

[24] I. V. Gopich, D. Nettels, B. Schuler, and A. Szabo, J. Chem. Phys. **131**(9), 095102 (2009).

[25] C. G. Hübner, G. Zumofen, A. Renn, A. Herrmann, K. Müllen, and T. Basché, Phys. Rev. Lett. **91**(9), 093903 (2003).

[26] D. Nettels, D. Haenni, S. Maillot, M. Gueye, A. Barth, V. Hirschfeld, C. G. Hubner, J. Leonard, and B. Schuler, Phys. Chem. Chem. Phys. **17**(48), 32304 (2015).

[27] I. V. Gopich and A. Szabo, J. Chem. Phys. **124**(15), 154712 (2006).

[28] R. Zheng, T. M. Jenkins, and R. Craigie, Proc. Natl. Acad. Sci. U. S. A. **93**(24), 13659 (1996).

[29] S. Müller-Späth, A. Soranno, V. Hirschfeld, H. Hofmann, S. Rüegger, L. Reymond, D. Nettels, and B. Schuler, Proc. Natl. Acad. Sci. U. S. A. **107**(33), 14609 (2010).

[30] A. Hoffmann, A. Kane, D. Nettels, D. E. Hertzog, P. Baumgärtel, J. Lengefeld, G. Reichardt, D. A. Horsley, R. Seckler, O. Bakajin, and B. Schuler, Proc. Natl. Acad. Sci. U. S. A. **104**(1), 105 (2007).

[31] A. Borgia, W. Zheng, K. Buholzer, M. B. Borgia, A. Schuler, H. Hofmann, A. Soranno, D. Nettels, K. Gast, A. Grishaev, R. B. Best, and B. Schuler, J. Am. Chem. Soc. **138**(36), 11714 (2016).

[32] N. Di Fiori and A. Meller, Biophys. J. **98**(10), 2265 (2010).

[33] H. S. Chung, J. M. Louis, and W. A. Eaton, Biophys. J. **98**(4), 696 (2010).

[34] B. Schuler, H. Hofmann, A. Soranno, and D. Nettels, Annu. Rev. Biophys. **45**, 207 (2016).

[35] R. B. Best, W. Zheng, and J. Mittal, J. Chem. Theory Comput. **10**(11), 5113 (2014).

[36] A. C. Vaiana, H. Neuweiler, A. Schulz, J. Wolfrum, M. Sauer, and J. C. Smith, J. Am. Chem. Soc. **125**(47), 14564 (2003).

[37] R. B. Best, Curr. Opin. Struct. Biol. **42**, 147 (2017).

[38] A. Szabo, K. Schulten, and Z. Schulten, J. Chem. Phys. **72**(8), 4350 (1980).

[39] D. Frishman and P. Argos, Proteins **23**(4), 566 (1995).

[40] B. Schuler and H. Hofmann, Curr. Opin. Struct. Biol. **23**(1), 36 (2013).

[41] G. Haran, Curr. Opin. Struct. Biol. **22**(1), 14 (2012).

[42] A. Vitalis and R. V. Pappu, J. Comput. Chem. **30**(5), 673 (2009).

[43] R. Wuttke, H. Hofmann, D. Nettels, M. B. Borgia, J. Mittal, R. B. Best, and B. Schuler, Proc. Natl. Acad. Sci. U. S. A. **111**(14), 5213 (2014).

[44] R. B. Best and J. Mittal, J. Phys. Chem. B **114**(46), 14916 (2010).

[45] P. S. Nerenberg, B. Jo, C. So, A. Tripathy, and T. Head-Gordon, J. Phys. Chem. B **116**(15), 4524 (2012).

[46] Y. Nomura, T. Masuda, and G. Kawai, J. Biochem. **139**(4), 753 (2006).

[47] M. P. Lutolf, N. Tirelli, S. Cerritelli, L. Cavalli, and J. A. Hubbell, Bioconjugate Chem. **12**(6), 1051 (2001).

[48] M. Aznauryan, L. Delgado, A. Soranno, D. Nettels, J. R. Huang, A. M. Labhardt, S. Grzesiek, and B. Schuler, Proc. Natl. Acad. Sci. U. S. A. **113**(37), E5389 (2016).

Supplementary Information Zosel *et al.*

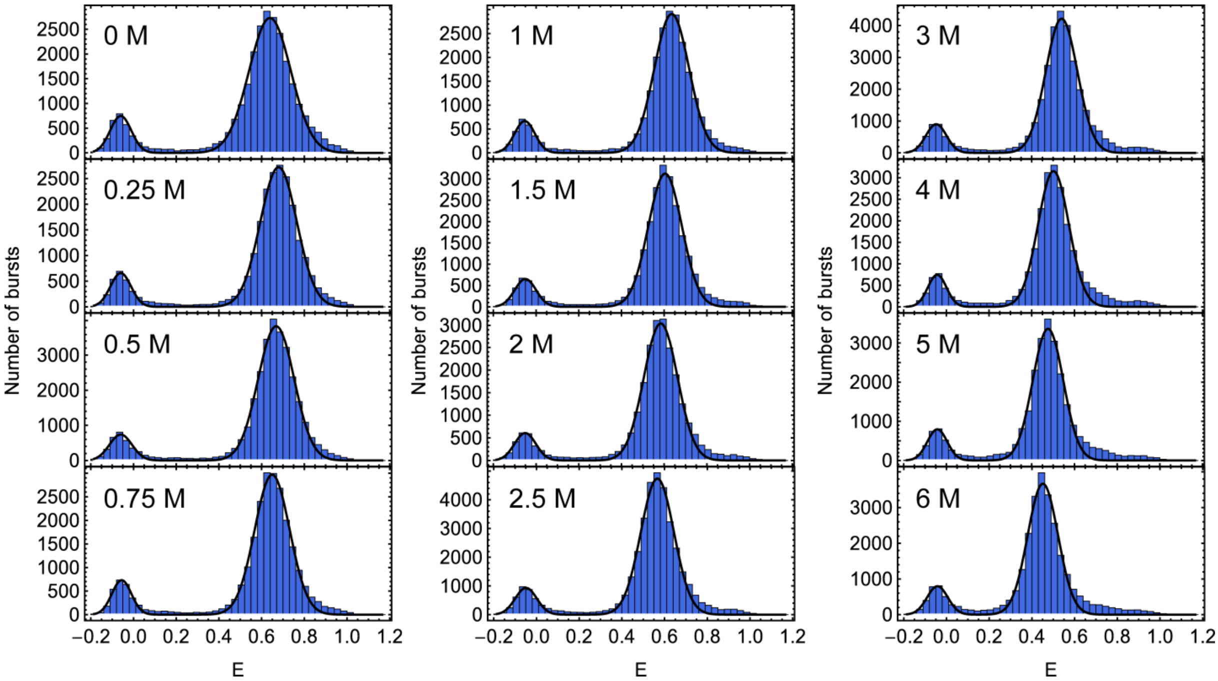


Figure S1. Transfer efficiency histograms of WDA at the indicated GdmCl concentrations. The histograms are fitted with two Gaussian peak functions. The peak at low transfer efficiency originates from molecules without an active acceptor dye (11.3±0.5% of all detected bursts).

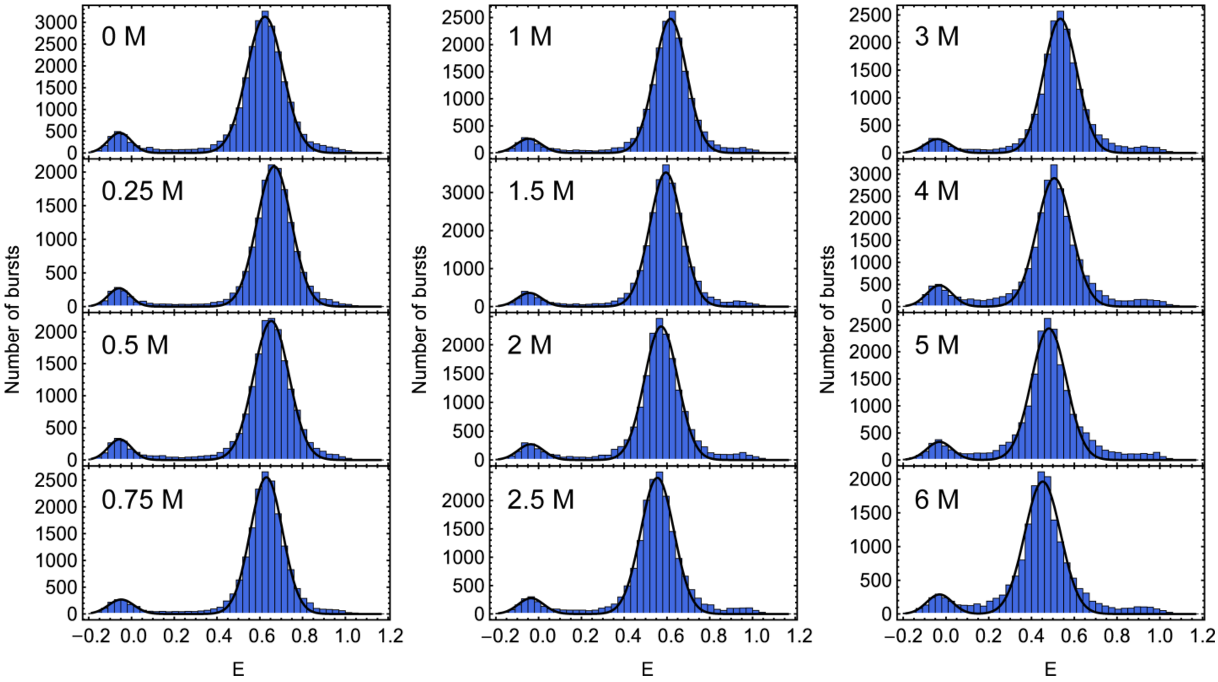


Figure S2. Transfer efficiency histograms of FDA at the indicated GdmCl concentrations. The histograms are fitted with two Gaussian peak functions. The peak at low transfer efficiency originates from molecules without an active acceptor dye (7.5±0.7% of all detected bursts).

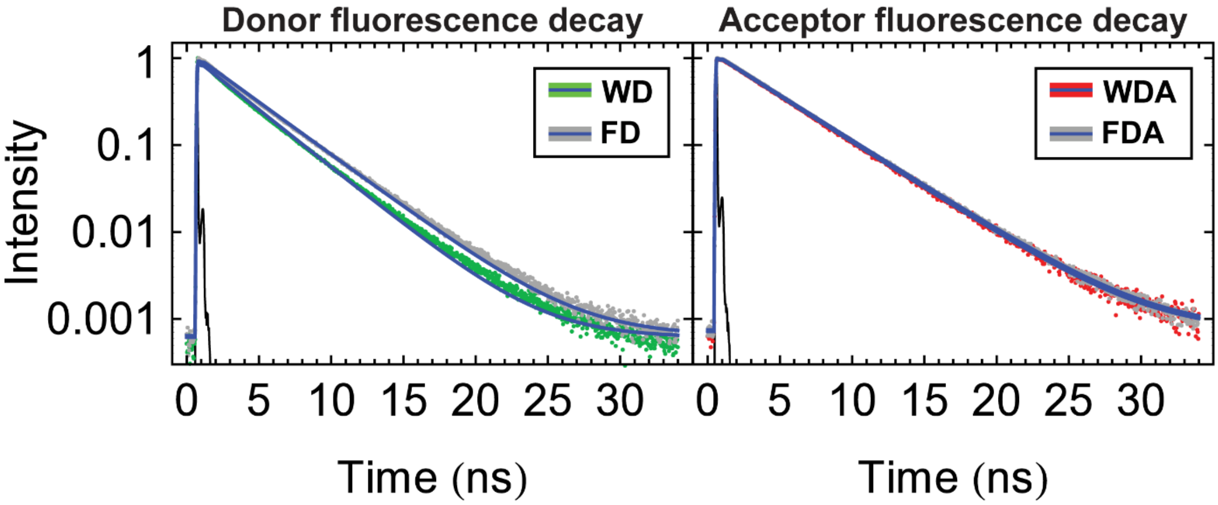


Figure S3. Donor and acceptor fluorescence lifetime decays of the quenched (WD/WDA) and un-quenched (FD/FDA) IN variants measured in buffer (0 M GdmCl). The donor fluorescence decays were recorded with the donor-only variants WD and FD. The acceptor fluorescence decays were recorded after direct excitation of the acceptor in the double-labeled variants WDA and FDA. The data were fitted with a convolution of the instrumental response function (black) with a single-exponential decay (blue lines).

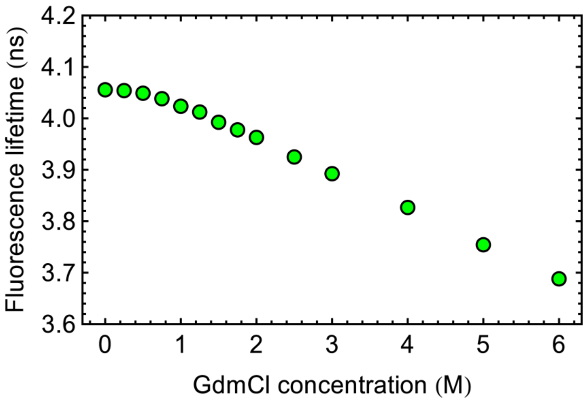


Figure S4. Fluorescence lifetime of free Alexa 488 as a function of GdmCl concentration. Fluorescence lifetime decays were fitted with a convolution of the instrumental response function with a single-exponential decay.

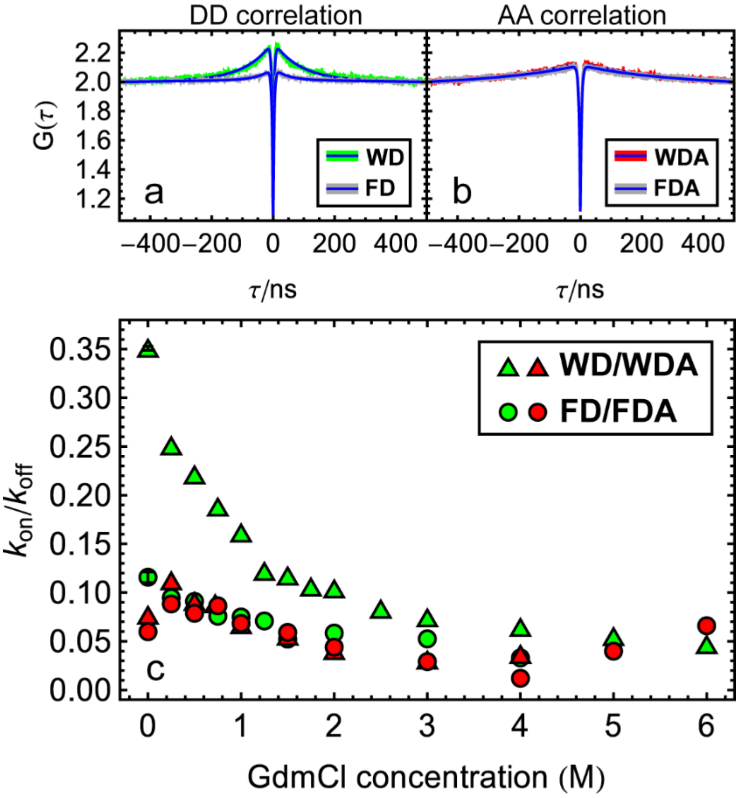


Figure S5. PET quenching of Alexa 488 and Alexa 594 in the absence of FRET. (a) Comparison of FCS data obtained from exciting Alexa 488 in WD and FD variants in 0.25 M GdmCl. (b) Comparison of FCS data obtained from exciting Alexa 594 in WDA and FDA variants in 0.25 M GdmCl. The data were fitted with a three-state photon statistics model[1] (blue lines) to obtain equilibrium constants (k_{on}/k_{off}) for quenching, which are shown in (c) as a function of the GdmCl concentration. Donor quenching equilibrium constants are shown in green, acceptor quenching equilibrium constants in red. Error bars (within the size of the symbols) are shown where multiple measurements were carried out. All FCS curves and fits of the measurements are shown in Figures S11-S14. Only Alexa 488 in the WDA variant experiences significant PET quenching by Trp in the constructs used here.

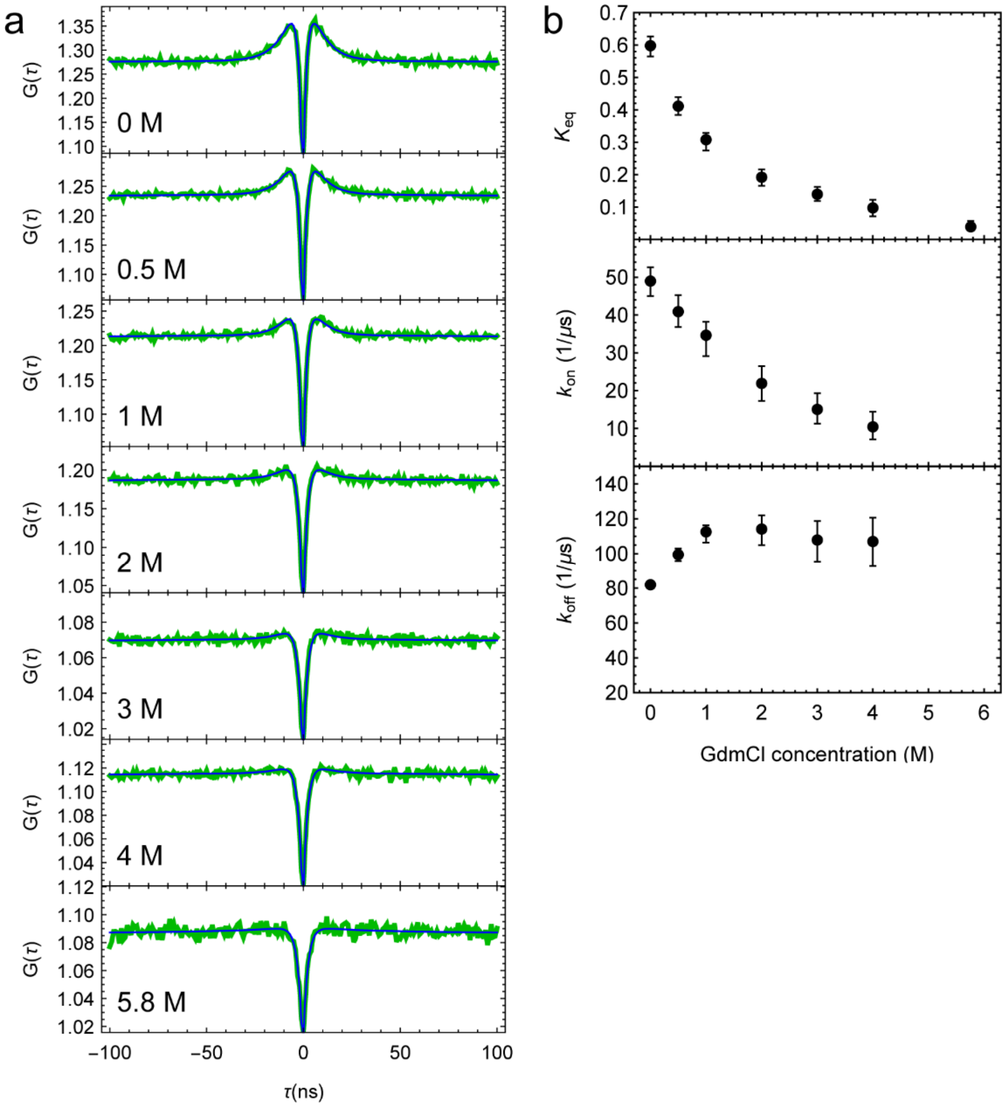


Figure S6. Bimolecular PET quenching of Alexa 488 by Trp (50 mM) as a function of the GdmCl concentration. (a) FCS measurements (green) were obtained at the GdmCl concentrations indicated. For each measurement, a model correlation curve $G_{DD}(\tau)$ (blue line) was fitted to the data, assuming a simplified 3-state model with parameters k_{on} and k_{off} for PET quenching (see "Modelling combined dynamics from FRET and PET"). The fluorescence emission rate coefficient k_D was set to 1/2.37 ns (the fluorescence lifetime of Alexa 488 in presence of 50 mM Trp, assuming a Stern-Volmer constant for dynamic quenching of 14.2 M^{-1}).[1] (b) Best-fit parameters for K_{eq}, k_{on} and k_{off}. Error bars indicate the standard deviation of 100 independent fits where the input parameters that were kept fixed in the fit were varied by ±5%, our estimate of their uncertainty. Above 4 M GdmCl, K_{eq} is too small for a reliable fit of k_{on} and k_{off}. The formation of the bimolecular complex between Alexa 488 and Trp becomes less efficient with increasing GdmCl concentration, which is largely caused by a decrease in the association rate. At 0 M GdmCl, the determined association and dissociation rates (k_{on} =1.0·10^9 $M^{-1}s^{-1}$, k_{off} =0.08 ns^{-1}) agree well with previously published values.[1]

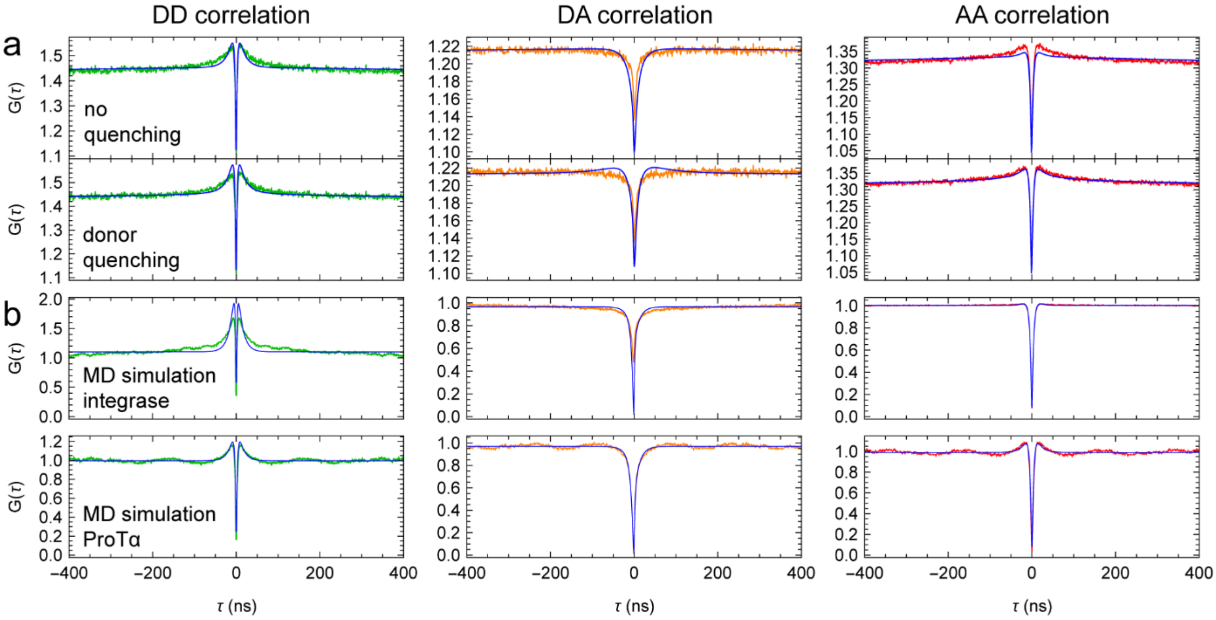


Figure S7. (a) FCS measurements of the FDA variant at 0 M GdmCl. The donor-donor (G_{DD}; green), donor-acceptor (G_{DA}; orange), and acceptor-acceptor (G_{AA}; red) FCS curves are shown. In the upper panels, the DD, DA, and AA data were fitted globally with intrachain diffusion coefficient D as a shared fit parameter (blue lines) with a 4-state-model without quenching (as in Fig. 3b). In the lower panels, the same data were fitted globally (blue lines) with a six-state-model (shared D, k_{on}, and k_{off}), as for the WDA variant. Although the 6-state model describes the acceptor-acceptor correlation better, it gives rise to a positive amplitude in the donor-acceptor crosscorrelation, which conflicts with the data. (b) Correlation curves calculated from MD simulations of integrase[2], shown with global fits with a 4-state-model without quenching (blue lines). Upper panels: The FCS curves calculated based on the integrase simulation are not well described by the 4-state model. As for the measured data, the fit does not account for a slow decay phase. Lower panel: the 4-state model describes the correlation curves of unfolded proteins without residual structure well, as shown for the completely disordered, highly charged N-terminal half (residues 1 to 56) of ProTα[2, 3].

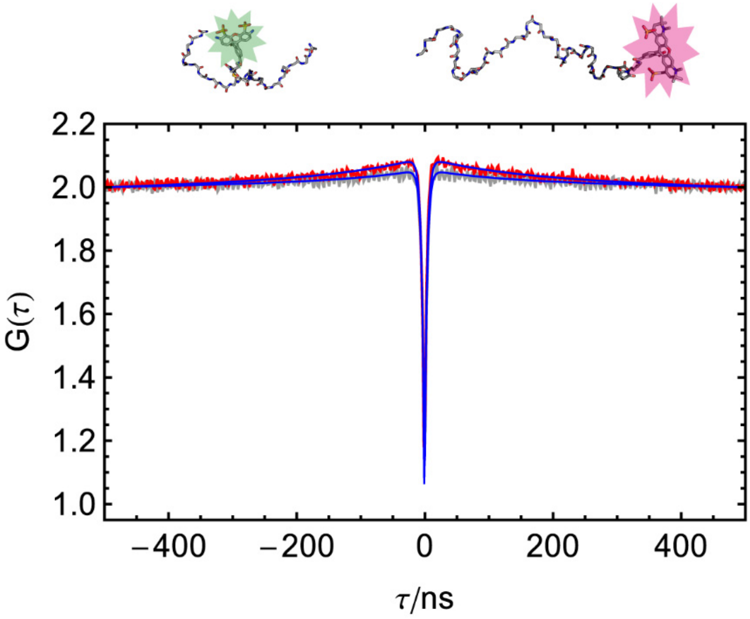


Figure S8. To test for dye-dye quenching, the acceptor dye Alexa 594 was excited directly in the FDA variant at 0 M GdmCl (red curve). A small positive amplitude is observed in the acceptor autocorrelation curve. A fit with a three-state model (blue line) yields an equilibrium constant for quenching of 0.06. FDA was then cleaved after position 24 with Factor Xa protease. The cleavage separates the two dyes at positions 11 and 60 without changing the local environment of the acceptor dye (see structural representation on top). We indeed observe a reduced FCS amplitude (equilibrium constant for quenching of 0.03) in cleaved FDA (gray curve), which indicates the presence of a small contribution of dye-dye quenching in the double-labeled construct. The FCS data were normalized to facilitate the comparison.

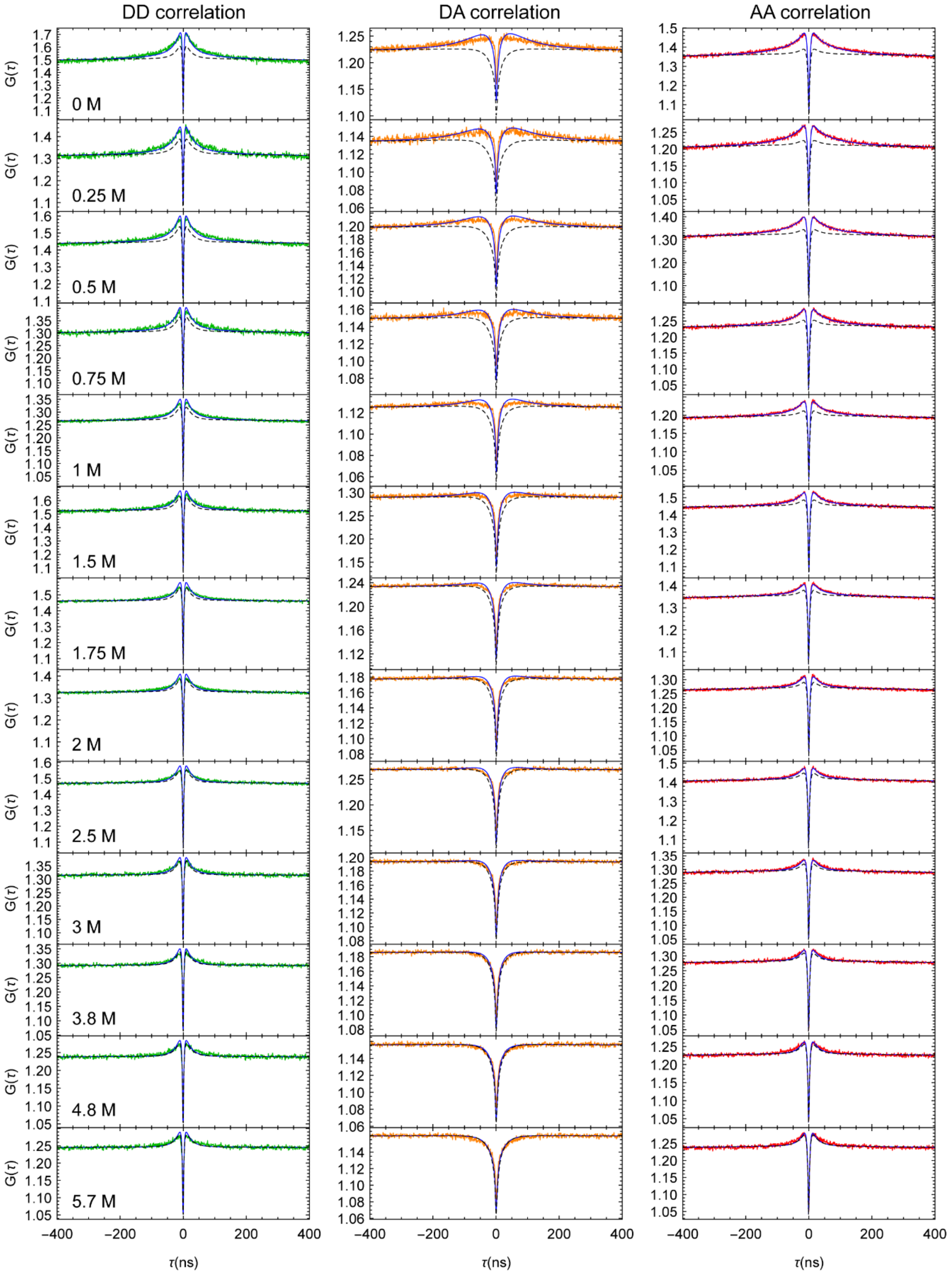


Figure S9. FCS curves of the WDA variant at increasing GdmCl concentrations from 0 M to 5.7 M (as indicated). For all measurements, the donor-donor (green), donor-acceptor (orange), and acceptor-acceptor (red) FCS curves are shown. For each set of measurements, model correlation curves $G_{DD}(\tau)$, $G_{DA}(\tau)$ and $G_{AA}(\tau)$ (blue lines; 6-state model) were fitted globally (shared D, k_{on}, and k_{off}) to the three correlations. To illustrate the magnitude of the quenching effect and show the FCS contribution of chain diffusion, the dashed lines show the fit results with k_{on} set to zero.

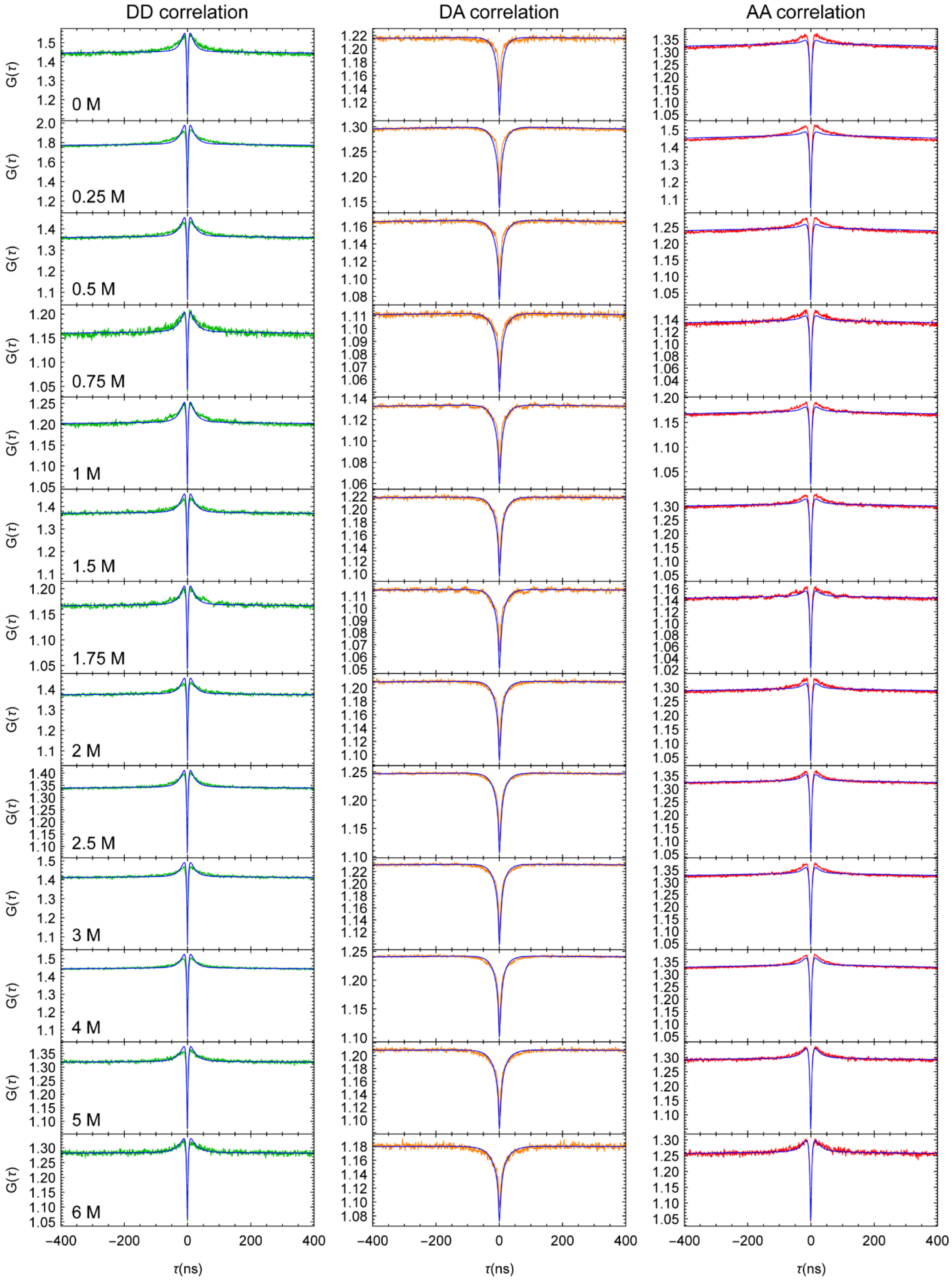


Figure S10. FCS curves of the FDA variant at GdmCl concentrations from 0 M to 6 M (as indicated). For all measurements, the donor-donor (green), donor-acceptor (orange), and acceptor-acceptor (red) FCS curves are shown. For each set of curves, model correlation curves $G_{DD}(\tau)$, $G_{DA}(\tau)$ and $G_{AA}(\tau)$ (blue lines; calculated from a 4-state model) were fitted globally (shared *D*) to the three correlations.

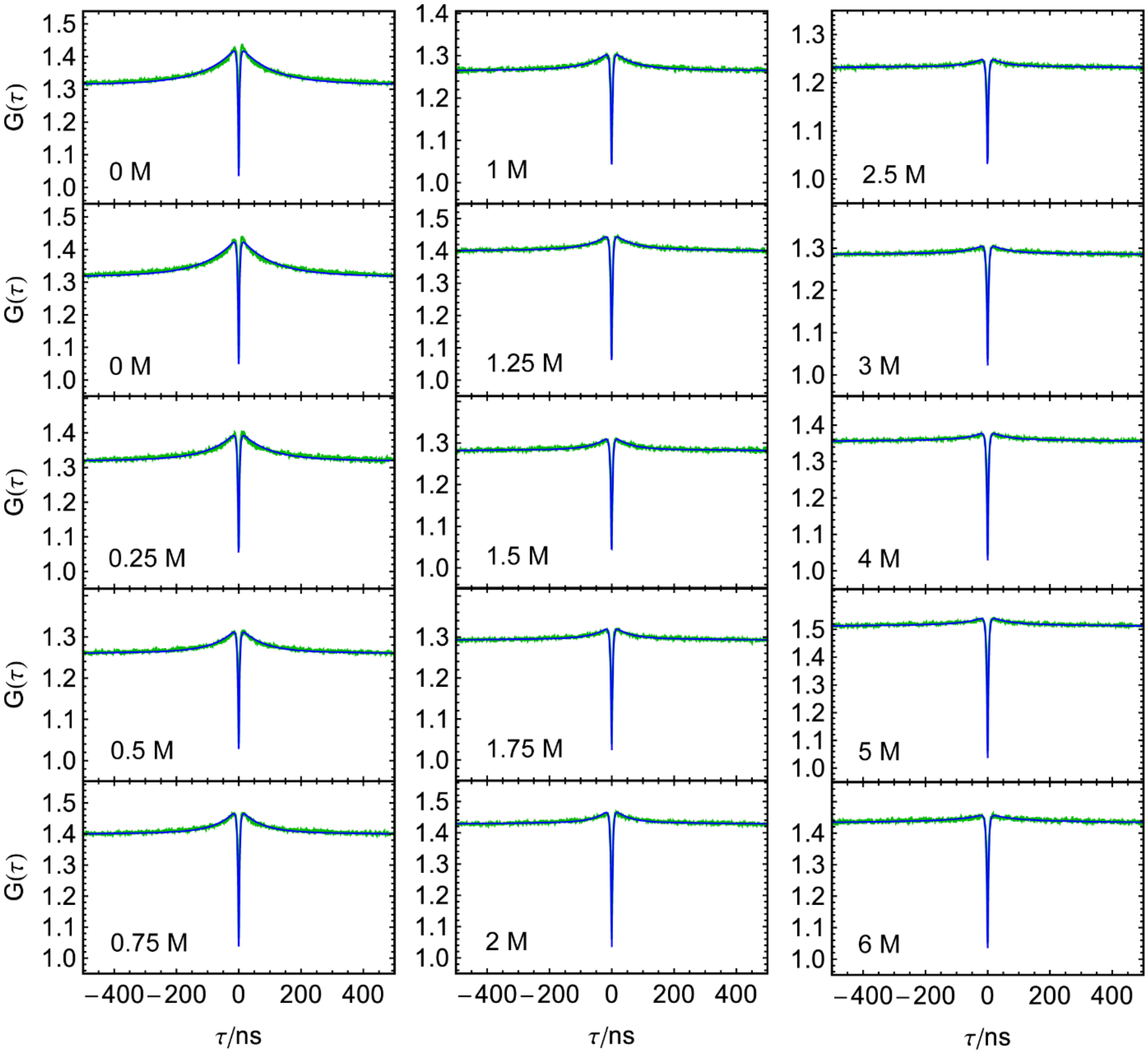


Figure S11. Donor FCS data for the WD variant. FCS measurements (green) were obtained at the GdmCl concentrations indicated. For each measurement, a model correlation curve $G_{DD}(\tau)$ (blue line) was fitted to the data, assuming a simplified 3-state model with parameters k_{on} and k_{off} for PET quenching (see “Modelling combined dynamics from FRET and PET”).

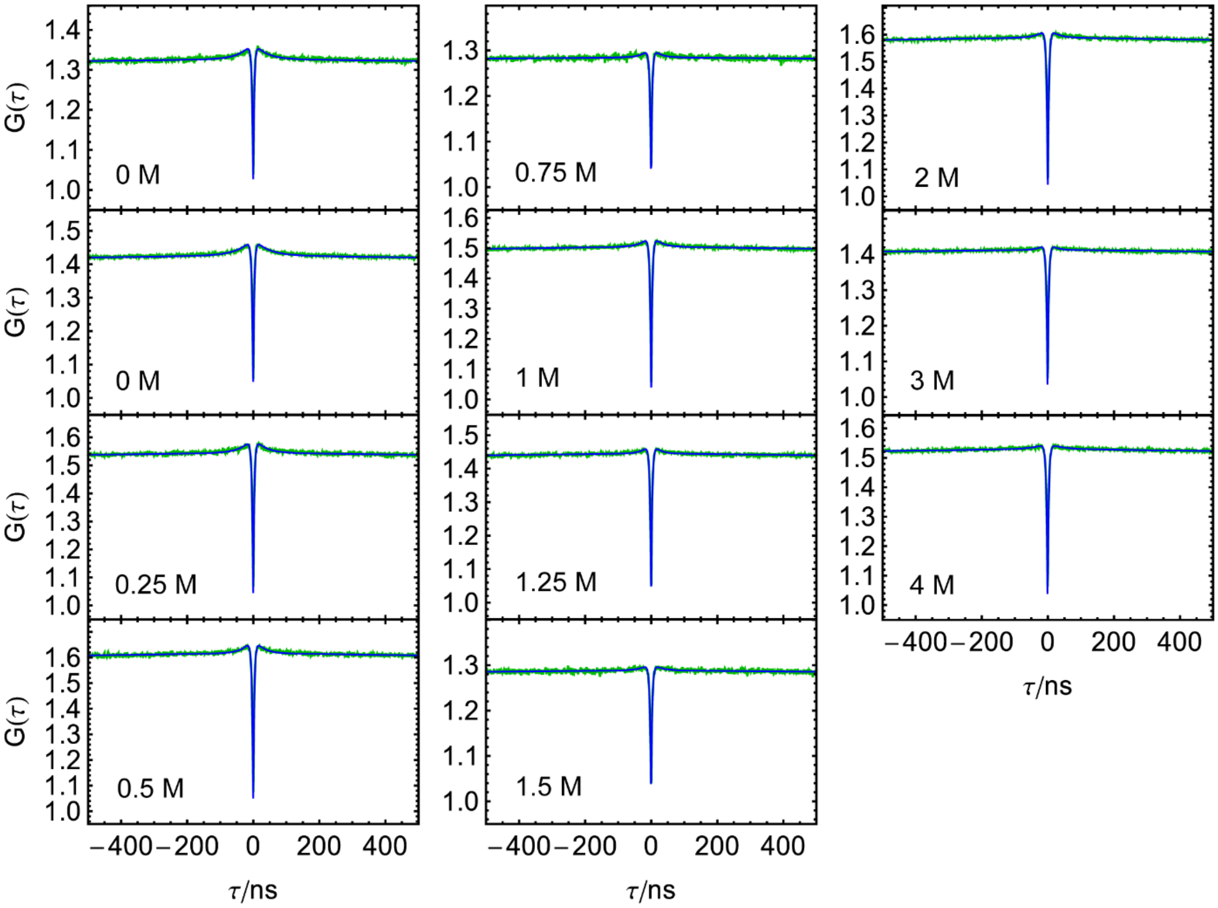


Figure S12. Donor FCS data for the FD variant. FCS measurements (green) were obtained from a sample at the indicated GdmCl concentrations. For each measurement, a model correlation curve $G_{DD}(\tau)$ (blue line) was fitted to the data, assuming a simplified 3-state model with parameters k_{on} and k_{off} for quenching (see "Modelling combined dynamics from FRET and PET").

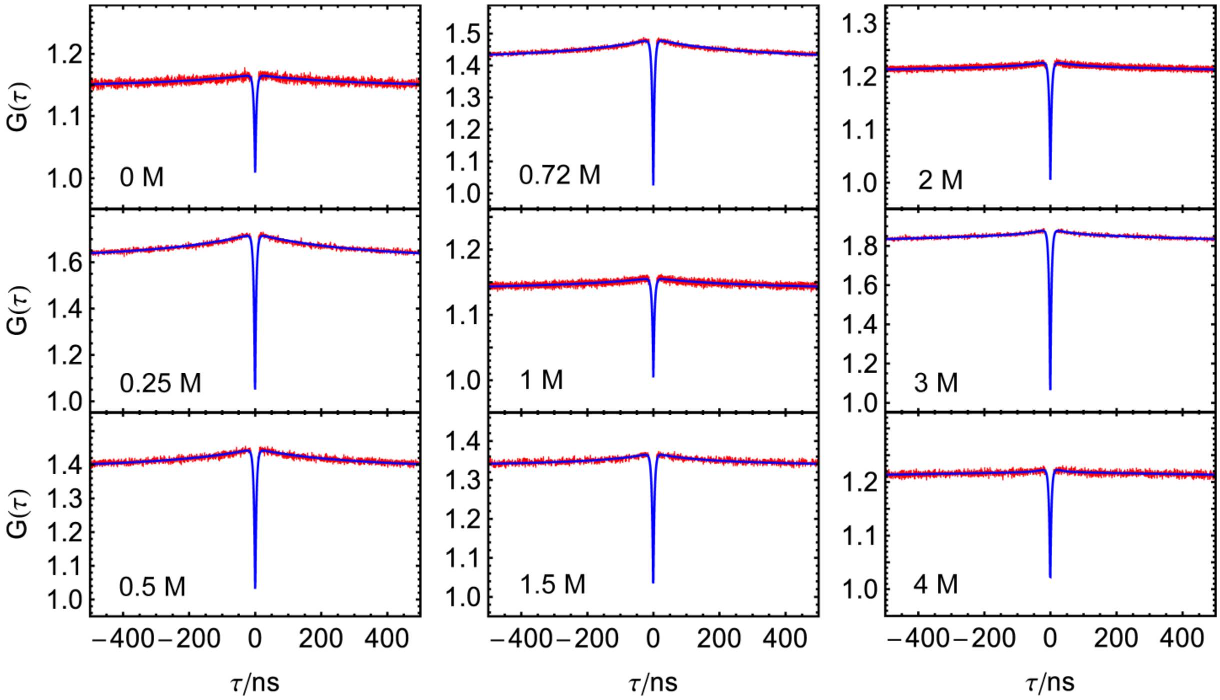


Figure S13. Acceptor FCS data for the WDA variant. FCS curves (red) were measured with excitation at 594 nm at the GdmCl concentrations indicated. For each measurement, a model correlation curve $G_{AA}(\tau)$ (blue line) was fitted to the data. Fitting was performed analogous to the WD/FD variants, assuming a simplified 3-state model with parameters k_{on} and k_{off} for quenching (see "Modelling combined dynamics from FRET and PET").

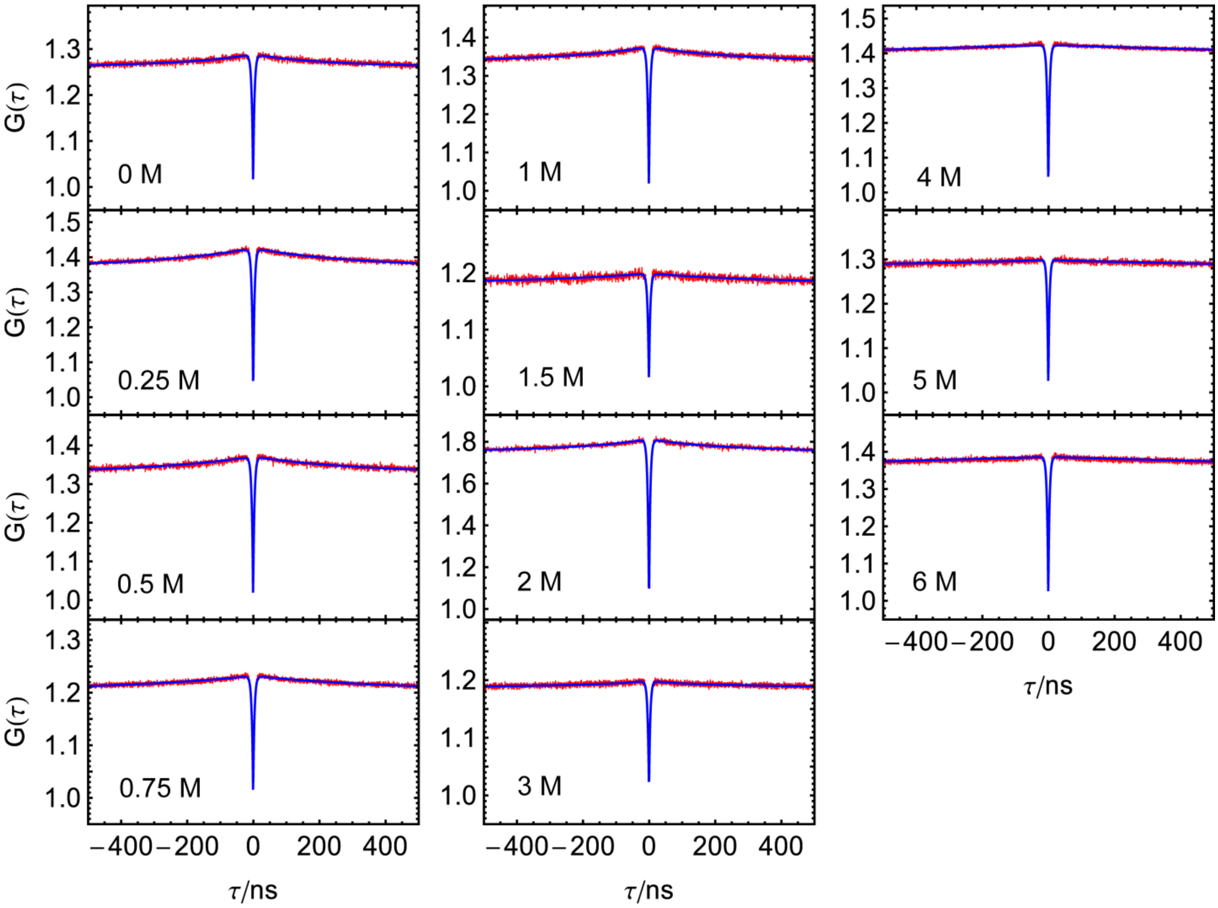


Figure S14. Acceptor FCS data for the FDA variant. FCS curves (red) were measured with excitation at 594 nm at the indicated GdmCl concentrations. For each measurement, a model correlation curve $G_{AA}(\tau)$ (blue line) was fitted to the data. Fitting was performed analogous to the WD/FD variants, assuming a simplified 3-state model with parameters k_{on} and k_{off} for quenching (see "Modelling combined dynamics from FRET and PET").

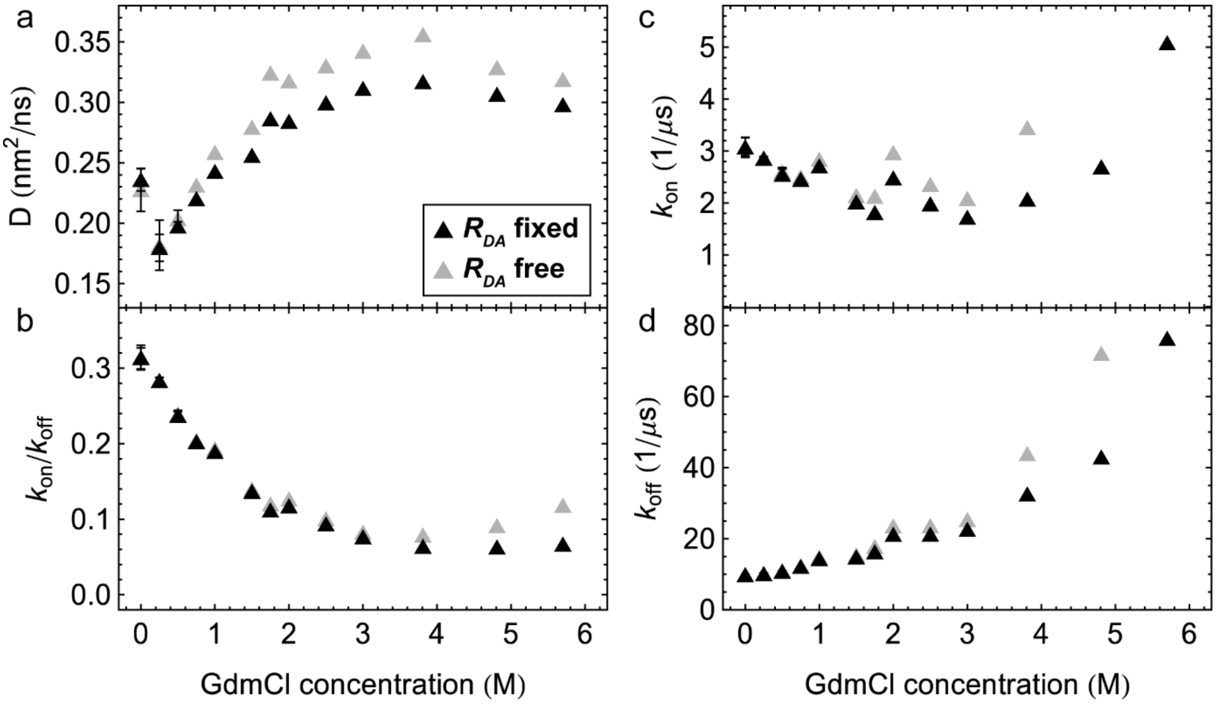


Figure S15. Comparison of best-fit parameters (D, k_{on}, and k_{off}) for the WDA variant, fitted with the six-state model, where R_{DA} was either obtained independently from transfer efficiency histograms (black triangles) or used as a free fit parameter (gray triangles). (a) Inter-dye diffusion coefficients, (b) equilibrium constants for quenching, (c) on- and (d) off-rate coefficients. The parameters for WDA were obtained from global fits with the 6-state model (Fig. 1). Where available, error bars indicate the standard deviation of 2-3 independent measurements.

References

1. D. Haenni, F. Zosel, L. Reymond, D. Nettels and B. Schuler, J. Phys. Chem. B **117** (42), 13015 (2013).
2. S. Piana, A. G. Donchev, P. Robustelli and D. E. Shaw, J. Phys. Chem. B **119** (16), 5113 (2015).
3. A. Soranno, A. Holla, F. Dingfelder, D. Nettels, D. E. Makarov and B. Schuler, Proc. Natl. Acad. Sci. U.S.A. **114** (10), E1833 (2017).