Biophysical Journal, Volume *115*

Supplemental Information

Analyzing the Dynamics of Single TBP-DNA-NC2 Complexes Using Hidden Markov Models

Nawid Zarrabi, Peter Schluesche, Michael Meisterernst, Michael Börsch, and Don C. Lamb

Supporting Information

"Analyzing the Dynamics of Single TBP-DNA-NC2 complexes by Hidden Markov Models"

Nawid Zarrabi[1], Peter Schluesche[2], Micheal Meisterernst[3,4], Micheal Börsch[1], Don C. Lamb[2]

1) 3. Physikalisches Institut, Universität Stuttgart, Pfaffenwaldring 57, 70550 Stuttgart, Germany.

2) Department Chemie, Center for Nano Science (CENS), Center for Integrated Protein Science (CIPSM) and Nanosystems Initiative München (NIM), Ludwig Maximilians-Universität Munich, Butenandstr. 5-13, 81377 Munich, Germany

3) National Research Center for Environment and Health - GSF, Gene Expression, Marchionini-Str. 25, 81377 Munich, Germany

4) Institute for Tumor Biology, Department of Medicine, University of Muenster, 48149 Muenster

Table of contents **Page**

Estimation of the Number of Detected Photons and the Uncertainty using an EMCCD detector

To calculate the uncertainty in the measured fluorescence intensities in the single-pair FRET traces, it is necessary to know the number of photons that have been detected. At the low intensities available in single molecule experiments, the uncertainty is dominated by shot-noise. Shot-noise is Poisson distributed and hence, when using photon-counting detectors, the uncertainty (i.e. standard deviation) is given by the square root of the number of detected photons.

Although current EMCCD cameras detect single photons, the readout is not given in photons and additional noise-sources are present. Each detected photon generates an electron in the CCD chip. However, to readout the number of electrons, the charge has to be shifted and is often amplified. The number of detected photons can be estimated from the camera counts using:

$$N_{\text{Photons}} = \frac{N_{\text{Camera Counts}} \times N_{e^-/\text{Camera Counts}}}{Gain} \tag{S1}$$

where the number of electrons per camera count is given for each individual camera by the manufacture and the *Gain* is either provided directly, as in the case of a calibrated linear gain, or is determined independently. In our analysis, photons detected from the same molecule on different pixels are summed together to give an estimate of the total intensity of the molecule in the individual frame. The total fluorescence signal is coming from the molecule of interest and from the background. The background intensity is calculated from the pixels surrounding the individual molecules. The average background intensity is determined and subtracted from the total intensity of the individual molecule. Usually, one can select a large enough region of pixels for calculating the average background intensity that no additional error is brought in via background subtraction. When the background intensity changes slowly with time, the background can also be averaged with a sliding time window to provide a very accurate average value. Thus, the uncertainty in the fluorescence signal is given by the shot noise, i.e. the total number of detected photons (signal plus background). The fluorescence signal and uncertainty in each frame is then given by:

$$\begin{aligned} I_D &= N_D - I_D^{background}; & \sigma_D^2 &= N_D \\ I_A &= N_A - I_A^{background} - I_D^{crosstalk}; & \sigma_A^2 &= N_A \end{aligned} \tag{S2}$$

where N_i represents the total number of detected photons in the ith channel and I_D and I_A the corrected intensities for the donor and acceptor channels respectively. We neglect direct excitation of the acceptor molecule with donor excitation, although it could be easily

incorporated if desired. Additional noise sources include thermal noise or dark counts, clock-induced charge, multiplicative noise and readout noise. See references (1, 2) for a detailed description of camera noise. For our EMCCD camera (DV887-BV iXon$^+$, Andor Technology), the thermal noise and clock-induced charge are negligible. In addition, the readout noise is significantly smaller than one photon after amplification. If one could directly determine the number of electrons per pixel, then a CCD camera could be used as a photon-counting device. However, due to the different noise sources involved in reading the output of the camera, it is not possible to uniquely determine the number of detected photons. Therefore, we approximate the number of detected photons and thereby estimate the uncertainty of the measurement.

Comparison of HMM analyses with and without incorporation of the camera noise.
We simulated 20 single molecule FRET trajectories using a four-state Markov model. The parameters for the four-states were chosen with FRET values of {0.25, 0.45, 0.65, 0.85}, a variance and dwell time for each state of 0.002 and 100 ms, and an equal probability for transitions between all other states. The twenty state sequences (i.e. 20 molecules) were generated each with a length of 800 data points (one data point representing a 5 ms frame). Starting from the state sequence, the fluorescence signals were created. Each molecule was assigned an average total intensity from a random number between 10 and 110 counts per frame to generate bright and dark molecules. The average intensity of the donor and acceptor signal were calculated from the FRET state occupied at that time point. Noise was than generated by producing independent random numbers with a Gaussian distribution with a variance given by mean intensity (i.e. the variance of a Poissonian distribution) multiplied by a factor of 2 (i.e. to simulate the addition noise of the EMCCD-technology). This leads to 20 pairs of donor and acceptor traces of 800 frames duration each, which were than analyzed with the different HMM approaches. The results are summarized in Supporting Table S1.

Model selection criteria
For all four different promotors, the loglikelihood values increased with rising model complexity (Figure S5, first row). Every additional state for the parameter room led to a better description of the data and resulted in a higher loglikelihood value. To handle this problem, the Bayesian Information Criterion (BIC) has been established by Schwarz in 1978 to give a Bayesian argument for model selections by accounting for the model complexity (3). The BIC, which is still under scientific debate (4), is defined as:

$$BIC = -2\ln L + \theta \ln T \tag{S3}$$

with ln *L,* loglikelihood of the data given the complete set of free parameters; *T,* the number of data points, and θ, the number of free parameters.

The second term is called the "penalty term", because it increases with a rising number of states and, therefore, acts against the first term, which decreases with expanded models. In principle, this leads to a minimum of the BIC for a certain number of free parameters indicating an appropriate choice of the number of hidden states. However, in our case, the BIC always decreased with an increasing number of states (see Figure S5, second row) implying at least ten different hidden states.

Alternatively, we obtained a reasonable model size by investigating the variances of the steps delivered by the model by calculated the χ^2 values. The correct number of states should yield similar results for the steady traces (without NC2) and dynamic case (in the presence of NC2) (Figure 3, third row):

$$\chi^2 = \frac{1}{T}\sum_{step}\sum_{t\in T_{step}}(x_t - \mu_{step})^2, \text{ with } \mu_{step} = \frac{1}{T_{step}}\sum_{t\in T_{step}} x_t \qquad \text{(S4)}$$

For the static FRET efficiency case in the absence of NC2, all ten hidden Markov models were expected to predominantly mark one step per molecule. Calculating the variances for these steps defined a lower bound, which was basically independent of the model. In the dynamic case of TBP-NC2-DNA, only suitable models were capable to assign the hidden states correctly. Incorrectly assigned steps included either missing jumps in the FRET efficiency, leading to raised variances, or falsely divided steps that fit the noise, resulting in underestimated variances. The results are shown in Figure S5 for the four different promotors investigated, both before (static in red) and after (dynamic in blue) addition of NC2. The χ^2-curves of the dynamic cases suggested that already a two-state model should resolve the main behavior of the TBP-NC2-complex, i.e. jumps between the steady FRET-level (~0.4) and a second higher FRET-level (~0.8). Additional HMM states refined this picture and resolved more and more short-living intermediate states around ~0.65 between the two major states. A fourth additional low-FRET-state was found at E_{FRET} ~0.2. We assumed that a sufficient model complexity should be achieved when the χ^2 values of the dynamic cases reached the χ^2 value for the static conditions. Adding more than four states did not lead to any significant improvement of the χ^2 values for the AdML promoter and provided already reasonable information for the promoter H2B (Figure S5). As a result, a global HMM with a minimal number of four states was chosen to describe the main features of the TBP-NC2-DNA dynamics.

The validity of the usual BIC analysis can be increased by taking into account a potential deviance of the gamma-factors between the molecules. Therefore, an additional BIC was calculated from the modified χ^2-value (Eq. S4) from the Loglikelihood:

$$\mathrm{BIC} = T\log(\chi^2) + \theta\log(T) \tag{S5}$$

where T denotes the total number of analyzed frames, χ^2 is the modified residual sum of squares according to Eq. S4, and θ counts the number of free parameters.

Like the usual BIC, the first summand decreases with increasing model complexity whereas the second increases acting like a "penalty term". Both summands together should indicate the right model size by a minimum of the BIC value. The result is shown in Figure S5, forth row. The usual BIC decreases without interruption with increasing model complexity whereas the decrease of the modified BIC stagnates for all samples at four hidden states.

The gain in validity of the χ^2-BIC is given by the modified calculation of the χ^2-value. The χ^2-value usually describes the squared difference between the fit and the data. Note, that the fit here was replaced by the mean value of a marked step instead of using the resulting FRET-efficiency of the corresponding hidden state. Small deviances in the gamma-factors across the molecules yield to shifts of the FRET efficiencies of all hidden states. This usually increases the mismatch between the model and the data resulting in higher χ^2-value or lower loglikelihood values. With the modified χ^2-value, the influence of state shifts can be suppressed.

In summary, the best model should have a maximum likelihood and a minimal BIC value. However, the likelihood always increased, as more states were included in the model. The usual BIC (second row) does not show any minimum, neither for the steady (red) nor for the dynamic (blue) FRET-trajectories. This suggests that the penalty term in the BIC for the complexity of the model was too small. For both the χ^2-value and the χ2-BIC, the static traces (without NC2, red), favoring a one-state model. For dynamic traces (after addition of NC2, blue) both values decrease with increasing model complexity up to four hidden states. Afterwards the χ^2-value and χ^2-BIC stagnates indicating that a hidden Markov model with four hidden states is sufficient to describe the main feature of the dynamic traces.

Transition-Density Plots

The transition density plots (TDPs) are 2D-density plots and are usually obtained by performing a 2D-binning of the data. The larger the binning size, the more counts there are per bin but with a reduction in the resolution of the maxima. We developed an alternative method to create the 2D-histograms without a loss of resolution. Starting with a picture of an arbitrary size, e.g. 300×300 pixels, the usual binning procedure is performed. This results in a 2D-histogram, where only a few pixels have more than one count. This 2D-histogram is convoluted once with a 2D-Gaussian. Its standard deviation is a free parameter analogue to the binning size of the standard procedure. A higher standard deviation creates maxima with higher counts without loss of the position accuracy. A sharper 2D-Gaussian yields more maxima with lower count rates analogue to a higher binning in the standard procedure.
The transition density plots (Figure 5b and Figure S7, right panel) where created with a resolution of 300×300 pixels and a standard deviation of 4 pixels.

Trace-wise versus global HMM

For a comparison of the four promoters, the transition rates between the two main states, state 2 and state 4, were expected to unravel potential differences between the promoters. However, the intermediate states, only poorly represented by a global HMM, were involved in the determination of that transition rate. Therefore, one new HMM analysis with 10 states for each molecule was independently performed and optimized. The subsequent calculation of the individual Viterbi paths was performed and used to assign different steps. The number of 10 states was sufficient to allocate all rarely appearing intermediate states in each molecule. The inordinate number of states used in the analysis did not disturb the results as states that were not needed were not occupied (5). The results are summarized as 2D-histograms according to their mean FRET-efficiencies and their dwell time in Figure 4.
For the H2B promoter constructs, the steps derived of the molecule-wise HMM around the global HMM states 1 and 3 did not cluster into a clear peak in the histogram shown in Figure 4 in comparison to the two main states. This could support the existence of many different intermediate states rather than a single broad one. In the latter case, a global HMM would describe this region by allocating a single hidden state with a broadened emission function. In contrast, in our case, the HMM tried to sample this area by putting more and more intermediate states with slightly different FRET values. The same was true for the HMM state at $E_{FRET} = 0.2$.

In order to derive global transition rates from these individual states from each molecule (Figure 6), we re-assigned these states back into the four global HMM states of interest, namely states 1, 2 and 4. For this purpose, the most likely global state i_k was determined out of the subset $\tilde{q}$ of those global states for every molecule-wise obtained step k individually:

$$i_k = \arg\max\left(f_{\tilde{q}}(m_k \mid \mu_{\tilde{q}}, \sigma_{\tilde{q}}^2 + \sigma_k^2)\right) \tag{S6}$$

Afterwards, adjacent steps of identical states were merged. This procedure led to local correctly assigned steps, whereas the short-living intermediate states were in this manner transferred to their best fitting neighboring main state. This enabled the determination of the transition rate $r_{i\to j}$ between all states by counting the number of corresponding transitions N_{ij} related to the summed duration of all steps k with assigned state i:

$$r_{i\to j} = N_{ij} / \sum_{k=1}^{N_i} T_{i,k} \tag{S7}$$

$T_{i,k}$ denotes the duration of the k^{th} step assigned to the main state i, N_i counts their number, N_{ij} is the number of all transitions from step i to step j.

Distribution of the FRET efficiency from normal distributed fluorescence intensities

The probability density functions of both fluorescence channels can be, with the assumption of sufficient count rates, approximated by a normal distribution.

The joint density function of both channels is then the product of both normal distributions $f_{I_D} = \mathcal{N}(\mu_{I_D}, \sigma_{i_D}^2)$ and $f_{I_A} = \mathcal{N}(\mu_{I_A}, \sigma_{i_A}^2)$ and is again a normal distribution:

$$f_{I_D} \otimes f_{I_A} = \frac{1}{2\pi\sigma_{I_D}\sigma_{I_A}} \exp\left(-\frac{(I_D - \mu_{I_D})^2}{2\sigma_{I_D}^2} - \frac{(I_A - \mu_{I_A})^2}{2\sigma_{I_A}^2} \right) \tag{S8}$$

I_D and I_A are the time dependent photon count rates of the donor- respectively acceptor channel with means μ_{I_D} and μ_{I_A} and variances $\sigma_{I_D}^2$ and $\sigma_{I_A}^2$. The variances are directly related to their corresponding means due to the intrinsic Poissonian characteristics of fluorescence count rates:

$$\sigma_{I_D}^2 = k_D\ \mu_{I_D} \text{ and } \sigma_{I_A}^2 = k_A\ \mu_{I_A} \tag{S9}$$

where k_D and k_A are one in the case of an ideal Poisson distribution. Due to the induced noise by the amplification from the EMCCD-camera used here, these parameters have a value of two (1) and increase further with an increasing level of subtracted background photons I_D^{noise} respectively I_A^{noise} and crosstalk I_D^{cross}:

$$\begin{aligned} \sigma_{I_D}^2 &= 2\,(\mu_{I_D} + \mu_{I_D}^{noise}) \quad &\Rightarrow \quad k_D &= 2\,(1 + \mu_{I_D}^{noise} / \mu_{I_D}) \\ \sigma_{I_A}^2 &= 2\,(\mu_{I_A} + \mu_{I_A}^{noise} + \mu_{I_D}^{cross}) \quad &\Rightarrow \quad k_A &= 2\,(1 + \mu_{I_A}^{noise} / \mu_{I_A} + \mu_{I_D}^{cross} / \mu_{I_A}) \end{aligned} \tag{S10}$$

The following coordination transformation leads to the distribution functions for the measured FRET efficiency *E* and total intensity *I*:

$$\begin{aligned} I_D &= h_1(E, I) = (1 - E)\, I \\ I_A &= h_2(E, I) = E\, I \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} E &= g_1(I_D, I_A) = \frac{I_A}{I_D + I_A} \\ I &= g_2(I_D, I_A) = I_D + I_A \end{aligned} \tag{S11}$$

The Jacobi determinant for the coordinate transformation is given by:

$$|\mathbf{J}| = \begin{vmatrix} \partial h_1/\partial E & \partial h_1/\partial I \\ \partial h_2/\partial E & \partial h_2/\partial I \end{vmatrix} = \begin{vmatrix} -I & 1-E \\ I & E \end{vmatrix} = |I| \tag{S12}$$

The dependence of the new joint density functions of the new variables, E and I, have the following form:

$$f_{E,I}(E,I) = |\mathbf{J}|\, f_{I_D,I_A}(I_D,I_A)$$

$$f_{E,I}(E,I) = \frac{|I|}{2\pi\sigma_{I_D}\sigma_{I_A}} \exp\left(-\frac{((1-E)I-\mu_{I_D})^2}{2\sigma_{I_D}^2} - \frac{(EI-\mu_{I_A})^2}{2\sigma_{I_A}^2} \right) \tag{S13}$$

The final distributions in FRET efficiency and intensity can be determined by summing over the intensity or FRET efficiency respectively. For generality, we integrate from -∞ to +∞, although the density functions should only have amplitudes where the intensity is positive and for FRET efficiencies are between 0 and 1.

$$f_E(E) = \int_{-\infty}^{\infty} f_{E,I}(E,I)\, dI$$

and

$$f_I(I) = \int_{-\infty}^{\infty} f_{E,I}(E,I)\, dE$$

We first consider $f_I(I)$:

$$f_I(I) = \int_{-\infty}^{\infty} \frac{|I|}{2\pi\sigma_{I_D}\sigma_{I_A}} \exp\left(-\frac{((1-E)I-\mu_{I_D})^2}{2\sigma_{I_D}^2} - \frac{(EI-\mu_{I_A})^2}{2\sigma_{I_A}^2} \right) dE = c|I| \int_{-\infty}^{\infty} \exp\left(-O(E^2)\right) dE \tag{S14}$$

where c represents a constant and $O(E^2)$ the exponent, which is a function of E^2. The Gaussian integral converges. As $I > 0$, the absolute value operation can be ignored and the result is a normal distribution:

$$f_I(I \mid \mu_I, \sigma_I^2) = \frac{1}{\sigma_I\sqrt{2\pi}} \exp\left(-\frac{(I-\mu_I)^2}{2\sigma_I^2} \right) \text{ with } \mu_I = \mu_{I_D} + \mu_{I_A} \text{ and } \sigma_I^2 = \sigma_{I_D}^2 + \sigma_{I_A}^2 \tag{S15}$$

The mean values and variances of the fluorescence intensities just sum up to the sum intensity, as one would expect.

Solving the integral for $f_E(E)$ is more involved:

$$f_E(E) = \int_{-\infty}^{\infty} \frac{|I|}{2\pi\sigma_{I_D}\sigma_{I_A}} exp\left(-\frac{((1-E)I-\mu_{I_D})^2}{2\sigma^2_{I_D}} - \frac{(EI-\mu_{I_A})^2}{2\sigma^2_{I_A}}\right) dI$$
$$= \int_{-\infty}^{\infty} |I| exp\left(-O(I^2)\right) dI \quad \text{(S16)}$$

where $O(I^2)$ represents the exponent, which is a function of I^2. The absolute value of the sum intensity $|I|$ can be replaced by I, because both Gaussians correspond to approximations of Poissonian distributions with μ_{I_D} or μ_{I_A}. The impact of negative I-values is therefore negligible.

Assuming $\mu_{I_D} > 0$ and $\mu_{I_A} > 0$, the integral is soluble and given by:

$$f_E\left(E|\mu_E,\sigma^2_E\right) = \frac{v(E)}{u(E)} \frac{1}{\sqrt{2\pi\sigma^2_E(E)}} exp\left(-\frac{\left(E-\mu_E\right)^2}{2\sigma^2_E(E)}\right) \quad \text{(S17)}$$

where we have used Eqn (S9) to convert the variances $\sigma^2_{I_D}$ and $\sigma^2_{I_A}$ into μ_{I_D} and μ_{I_A} and introduced additional functions $u(E)$, $v(E)$ and $\sigma^2_E(E)$ to simplify the representation of the formula;

$$\mu_E = \frac{\mu_{I_A}}{\mu_{I_D}+\mu_{I_A}},\ \mu_I = \mu_{I_D}+\mu_{I_A}$$

$$\sigma^2_E(E) = \frac{u(E)}{\mu_I}$$

and

$$u(E) = k_A\ \mu_E\ (1-E)^2 + k_D\ E^2(1-\mu_E)$$

$$v(E) = \mu_E(1-\mu_E)(k_A(1-E)+k_D\ E)$$

To obtain the integral, we rewrote Eqn S16 as the intensity multiplied by an exponential function with a quadratic function as an exponent:

$$f_E(E) = \int_{-\infty}^{\infty} I\ exp\left(-\left\{\left[\frac{(1-E)^2}{2\sigma^2_{I_D}} + \frac{E^2}{2\sigma^2_{I_A}}\right]I^2 - \left[\frac{\mu_{I_D}(1-E)}{\sigma^2_{I_D}} + \frac{\mu_{I_A}E}{\sigma^2_{I_A}}\right]I + \left[\frac{\mu^2_{I_D}}{2\sigma^2_{I_D}} + \frac{\mu^2_{I_A}}{2\sigma^2_{I_A}} + ln\left(2\pi\sigma_{I_D}\sigma_{I_A}\right)\right]\right\}\right) dI \quad \text{(S18)}$$

This can be equated to the expectation value of a variable x that is distributed with a normal distribution multiplied by a constant c

$$c\langle x\rangle = \int_{-\infty}^{\infty} x \frac{k}{\sqrt{2\pi\sigma_x^2}} exp\left(-\frac{\left(x-\mu_x\right)^2}{2\sigma_x^2}\right) dx$$
$$= \int_{-\infty}^{\infty} x\, exp\left(-\left\{\frac{1}{2\sigma_x^2}x^2 - \frac{\mu_x}{\sigma_x^2}x + \left[\frac{\mu_x^2}{2\sigma_x^2} + \frac{ln\left(2\pi\sigma_x^2\right)}{2} - ln\left(c\right)\right]\right\}\right) dx = c\mu_x \quad \text{(S19)}$$

By setting the coefficients of the quadratic exponent equal and solving for μ_x, σ_x and c yields Eqn S17 from the product of $c\mu_x$.

Eqn S17 approximates a normal distribution multiplied by a prefactor, $\frac{v(E)}{u(E)}$ and a variance, $\sigma_E^2\left(E\right)$, that depends on E. We can approximate the prefactor and variance using a Taylor series expansion about $E = \mu_E$, which is the crucial region of the formula. For the prefactor:

$$Taylor\left(\frac{v(E)}{u(E)}, E=\mu_E\right) = 1 - \frac{k_A - k_D}{k_A(1-\mu_E)+k_D\mu_E}(E-\mu_E) + O((E-\mu_E)^2) \quad \text{(S20)}$$

and for the variance:

$$Taylor\left(\frac{1}{\mu_I}\left(k_A\mu_E(1-E)^2 + k_D(1-\mu_E)E^2\right), E=\mu_E\right) =$$
$$\frac{\mu_E(1-\mu_E)}{\mu_I}\left(k_A(1-\mu_E)+k_D\mu_E\right)+\ldots \quad \text{(S21)}$$
$$\frac{2\mu_E(1-\mu_E)}{\mu_I}\left(k_D-k_A\right)(E-\mu_E)+O((E-\mu_E)^2)$$

Replacing the arguments that depend on E with the zeroth terms of their Taylor-approximations, the pre-factor becomes one and variance is replaced by the zeroth term:

$$\sigma_E^2 = \left(k_A(1-\mu_E)+k_D\mu_E\right)\frac{\mu_E(1-\mu_E)}{\mu_I} \quad \text{(S22)}$$

Substituting this into Eqn (S17) yields:

$$f_E(E \mid \mu_E, \sigma_E^2) = \frac{1}{\sqrt{2\pi}\sigma_E}\exp\left(-\frac{\left(E-\mu_E\right)^2}{2\sigma_E^2}\right) \quad \text{(S23)}$$

Thus, the distribution in FRET efficiency can be well approximated by a Gaussian distribution. In practice, k_A and k_D are equal when there is sufficient signal. In this case, the linear terms in the Taylor series expansions (Eqns (S20) and (S21)) become zero, which further increases the quality of this approximation.

Supplementary Figures

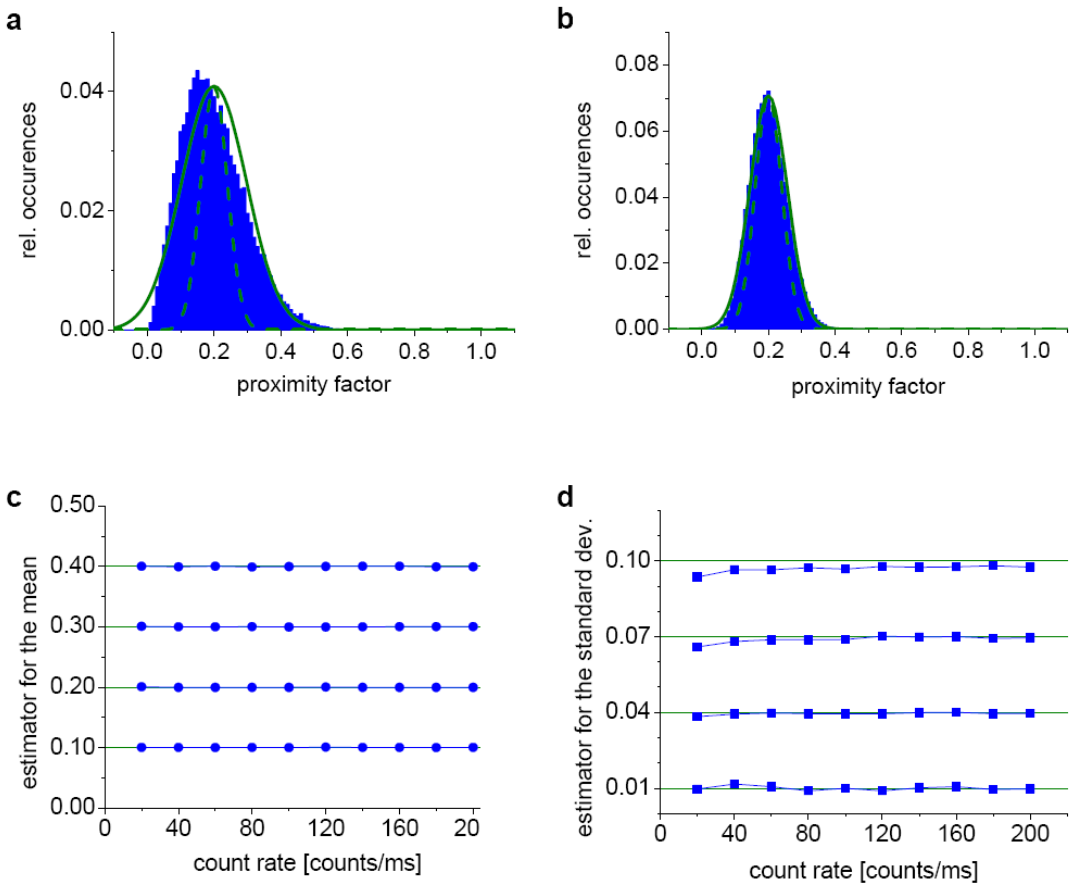

Figure S1: Monte Carlo Simulations and Extended HMM analysis of spFRET histograms. (**a**,**b**) Histogram of the spFRET proximity ratio (blue) with 50,000 simulated data points with an average FRET efficiency of 0.20 and a standard deviation of 0.04 for (**a**) 20 counts/ms and (**b**) 100 counts/ms. Dotted green lines: the underlying hidden FRET distribution. Solid green lines: broadening due to the limited number of measured photons. (**c**) The estimator tested for mean FRET efficiencies of 0.1, 0.2, 0.3 and 0.4 together with an inherent standard deviation of 0.05. The HMM works reliably over a wide range of values even when the underlying Gaussian distribution falls outside of the boarders of 0 and 1 and the appearing asymmetry of the data is not accounted by the estimator. (**d**) The estimation of the variances at standard deviations of 0.01, 0.04, 0.07 and 0.10 together for a mean FRET efficiency of 0.2. A slight systematic deviation at lower count rates and higher standard deviations is observed.

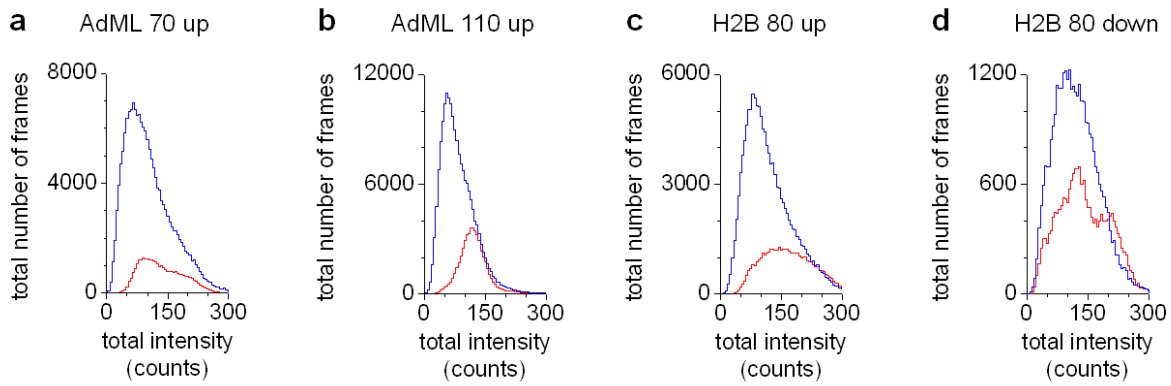

Figure S2: Experimental distribution of the total intensity per frame. Histograms of the total photons detected in the donor and acceptor channels per 5 ms frame are shown for TBP-DNA complexes in the absence (red) and presence (blue) of NC2. The different DNA sequences investigate are (**a**) a 70 bp upstream-labeled DNA containing the AdML TATA box, (**b**) a 110 bp upstream-labeled DNA containing the AdML TATA box, (**c**) an 80 bp upstream labeled DNA containing the H2B TATA box and (**d**) an 80 bp downstream labeled DNA containing the H2B TATA box.

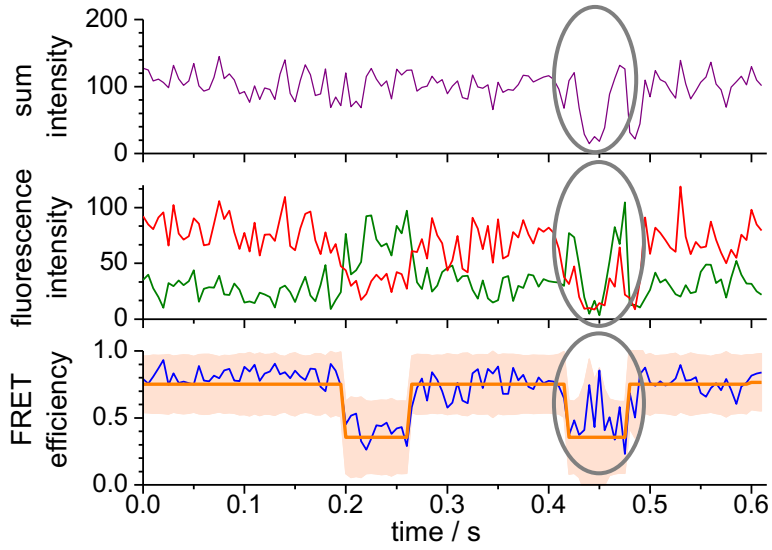

Figure S3: SpFRET trace with donor-quenching. An exemplary spFRET trace of the promoter AdML 110 bp upstream-labeled construct after addition of NC2 where the donor is transiently quenched. Purple: Total intensity per 5 ms, green: donor fluorescence counts per 5 ms, red: acceptor fluorescence counts per 5 ms, blue: FRET efficiency, orange: Viterbi path of the two-state model HMM analysis with the standard deviation due to shot noise shown as the envelope about the Viterbi path. The sudden drop in fluorescence intensity around 0.45 s due to transient quenching of the donor molecule leads to high fluctuations in the FRET efficiency trace. The new estimators are able to account for this effect: the density function (orange) is locally broadened when the total intensity drops, which ensures a correct assignment of the FRET conformation at this time to the low-FRET state.

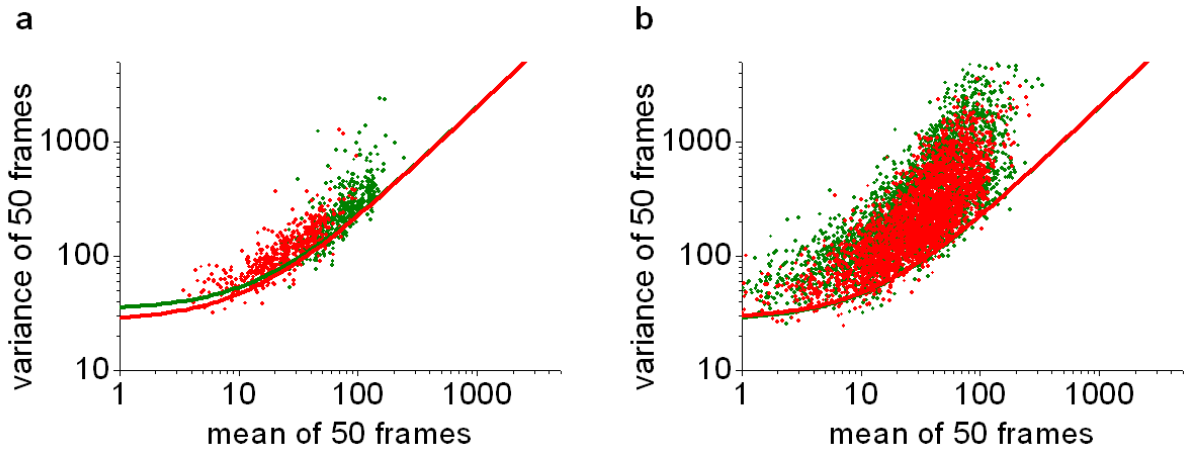

Figure S4: Plots of the experimentally determined variances versus mean intensity. The variance determined from 50 consecutive frames is plotted as a function of the corresponding background-corrected mean value for the donor (green) and acceptor (red) channels in the (**a**) absence (525 points) and (**b**) presence of NC2 (1519 points). The theoretical dependence expected from shot noise:

$$\sigma^2_{D/A,theo}(\mu_{D/A}) = 2\left(\mu_{D/A} + \left\langle I^{background}_{D/A0} \right\rangle\right),$$

are shown as a solid lines for the respective channels. For a Poissonian distribution, the variance is equal to the mean of the detected photons (signal and background). The factor of two accounts for the additional noise generated by the on-chip gain of the EMCCD camera. Before addition of NC2 (a), the data points lie close to the theoretical curve whereas, after addition of NC2 (b), the variances from the data points are much higher than the theoretical curve revealing the presence of additional conformational dynamics.

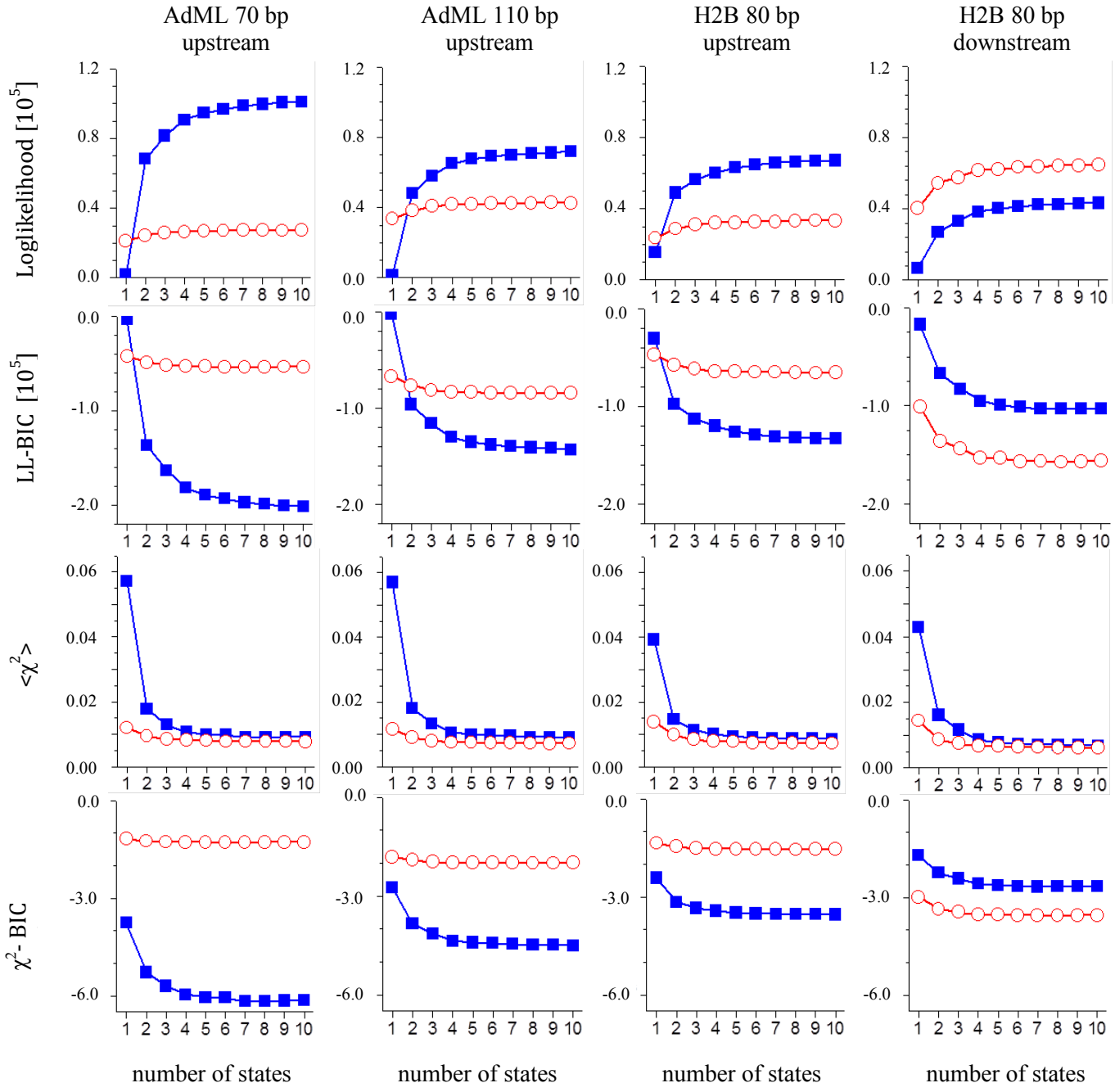

Figure S5: Determination of the number of relevant FRET states. Various criteria were tried to determine the number of distinguishable FRET states in the spFRET measurements. The Loglikelihood, χ^2-value and their corresponding Bayesian Information Criterion (BIC) are shown for the four different samples in the absence (red) and presence (blue) of NC2. First row: The Loglikelihood is plotted as a function of the number of states. Second row: The BIC calculated according to the obtained Loglikelihood-value. Third row: A plot of the χ^2 for the comparison of the Viterbi path to the spFRET data is plotted as a function of the number of states in the HMM. Fourth row: The BIC calculated according to the modified χ^2-value.

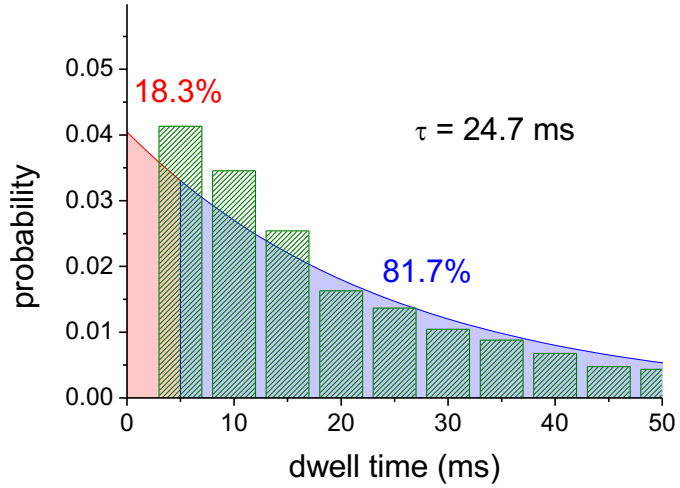

Figure S6: Fraction of Missed Transitions. The survival probability of the 0.64 FRET efficiency state for TBP bound to the 70 bp upstream-labeled DNA containing the AdML TATA box in the presence of NC2 determined from the HMM analysis is shown. An exponential function with a lifetime of 24.7 ms is shown for comparison. Assuming a minimum dwell time of 5 ms for a transition to be detected with the HMM analysis, ~ 18 % of the transitions from $E = 0.40$ to $E = 0.64$ on to $E = 0.83$ would be detected as a direct transition between the $E = 0.40$ and $E = 0.83$ FRET efficiency states.

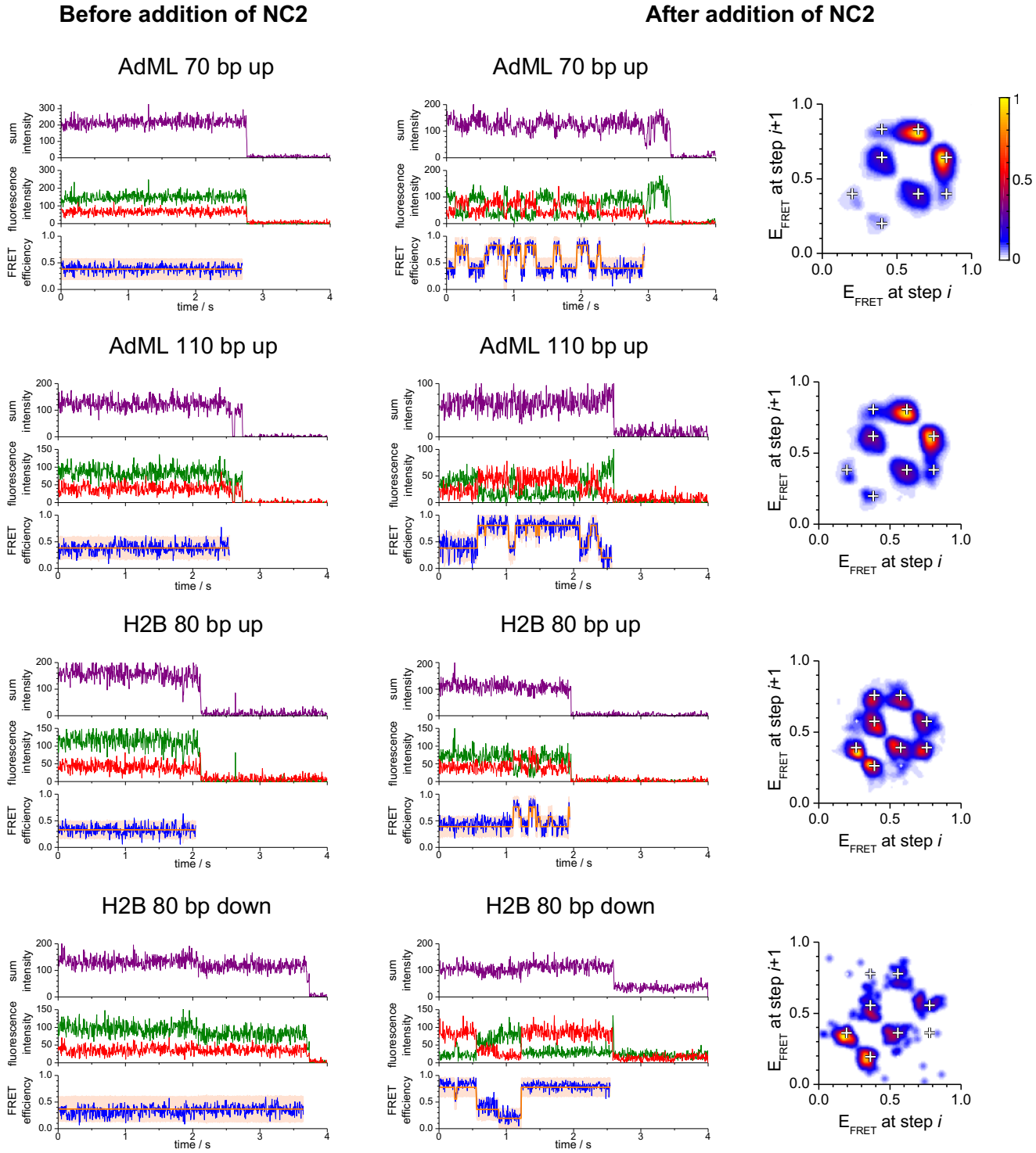

Figure S7: SpFRET traces and TDPs of all four sample preparations. Representative spFRET traces are shown for the four constructs investigated in this work in the absence (*left*) and presence (*middle*) of NC2. The total intensity is shown in purple, the intensity of the donor fluorophore is shown in green, the intensity of the acceptor fluorophore is shown in red, the frame-wise FRET efficiency is shown in blue and the Viterbi path and uncertainty due to shot-noise are shown in orange. (*right*) The TDPs are shown for the different complexes. The optimized Viterbi path from the global four-well HMM analysis was calculated for the individual traces and average FRET efficiency plotted as a Gaussian with a width of 2% for each level. The plots are normalized to the maximum number of transitions and indicates how often the transitions were observed with rare transitions given in blue and more frequent transitions highlighted in yellow. The corresponding color bar is shown to the right. The white pluses represent the values returned from the global four-well HMM analysis.

Supplementary Tables

State	FRET Efficiency μ			Width σ			Dwell Time (ms)		
	Model	Standard HMM	Extended HMM	Model	Standard HMM	Extended HMM	Model	Standard HMM	Extended HMM
s_1	0.250	0.246	0.251	0.002	0.011	0.002	100.0	80.3	106.1
s_2	0.450	0.449	0.450	0.002	0.015	0.002	100.0	90.3	108.8
s_3	0.650	0.651	0.648	0.002	0.014	0.002	100.0	73.9	92.5
s_4	0.850	0.850	0.849	0.002	0.008	0.002	100.0	87.5	110.8

Table S1: Comparison of the Standard HMM and the Extended HMM. Results from a simulation of 20 molecules using either the standard HMM or the HMM where the camera noise is incorporated into the analysis.

	AdML 70 bp up stream	AdML 110 bp up stream	H2B 80 bp up stream	H2B 80 bp down stream
without NC2	103	141	132	62
with NC2	432	315	279	55

Table S2: Number of molecules used by the hidden Markov analysis. Data were collected at 5 ms/frame or 200 Hz. From each molecule, the donor and acceptor intensities where extracted and a FRET trajectory was calculated.

	AdML 70 bp up stream		AdML 110 bp up stream		H2B 80 bp up stream		H2B 80 bp down stream	
	–NC2	+NC2	–NC2	+NC2	–NC2	+NC2	–NC2	+NC2
1 hidden state	103	432	141	315	132	279	62	55
2 hidden states	147	3643	217	2610	252	1113	82	128
3 hidden states	173	8146	241	4128	330	1829	123	184
4 hidden states	160	9546	252	6416	304	1966	112	291
5 hidden states	176	9786	256	6381	323	1984	124	340
6 hidden states	181	9299	234	6008	354	2031	129	366
7 hidden states	183	11797	239	6126	378	1885	140	411
8 hidden states	150	11746	262	7002	315	1853	137	358
9 hidden states	200	12688	263	6489	273	1807	128	367
10 hidden states	180	12582	296	8235	313	1930	132	371

Table S3: Transitions found in Various HMM Analyses. The number of detected transitions for the different hidden Markov models for each sample before (steady) and after (dynamic) the addition of NC2.

Number of hidden states	AdML 70 bp up stream			AdML 110 bp up stream			H2B 80 bp up stream			H2B 80 bp down stream		
	μ	σ	*f*	μ	σ	*f*	μ	σ	*f*	μ	σ	*f*
1	0.61	0.22	100%	0.57	0.22	100%	0.46	0.18	100%	0.45	0.19	100%
2	0.39	0.09	43%	0.36	0.12	45%	0.36	0.09	72%	0.33	0.10	66%
	0.78	0.12	57%	0.75	0.09	55%	0.71	0.09	28%	0.68	0.11	34%
3	0.34	0.10	32%	0.26	0.09	21%	0.30	0.08	39%	0.31	0.09	56%
	0.60	0.08	23%	0.48	0.08	30%	0.44	0.06	36%	0.52	0.07	25%
	0.82	0.05	44%	0.77	0.07	49%	0.73	0.08	25%	0.76	0.07	19%
4	0.20	0.08	7%	0.20	0.07	10%	0.26	0.07	19%	0.19	0.06	15%
	0.40	0.07	29%	0.38	0.06	29%	0.39	0.06	47%	0.36	0.06	47%
	0.64	0.07	23%	0.62	0.06	24%	0.58	0.06	14%	0.56	0.06	22%
	0.83	0.04	41%	0.81	0.04	37%	0.76	0.07	19%	0.78	0.06	17%
5	0.18	0.08	6%	0.18	0.07	9%	0.23	0.07	10%	0.19	0.06	13%
	0.36	0.06	21%	0.35	0.05	23%	0.36	0.06	43%	0.35	0.05	42%
	0.50	0.06	15%	0.50	0.05	15%	0.47	0.06	22%	0.49	0.05	18%
	0.70	0.07	24%	0.68	0.06	26%	0.69	0.05	18%	0.63	0.03	12%
	0.85	0.03	34%	0.83	0.03	28%	0.82	0.06	7%	0.79	0.05	14%
6	0.15	0.07	4%	0.16	0.06	6%	0.20	0.07	7%	0.17	0.06	10%
	0.32	0.06	12%	0.30	0.05	13%	0.33	0.06	30%	0.30	0.05	22%
	0.42	0.05	19%	0.41	0.05	20%	0.41	0.05	30%	0.40	0.04	29%
	0.57	0.06	11%	0.57	0.06	13%	0.55	0.06	11%	0.54	0.05	18%
	0.73	0.06	27%	0.71	0.05	26%	0.72	0.04	18%	0.71	0.04	11%
	0.86	0.03	27%	0.84	0.02	23%	0.86	0.05	4%	0.83±	0.04	9%
7	0.14	0.06	4%	0.13	0.06	4%	0.17	0.06	3%	0.15	0.05	8%
	0.28	0.06	7%	0.26	0.05	9%	0.29	0.06	16%	0.27	0.05	16%
	0.39	0.05	20%	0.37	0.05	20%	0.37	0.05	35%	0.37	0.05	34%
	0.51	0.05	10%	0.49	0.05	10%	0.47	0.05	18%	0.49	0.05	16%
	0.65	0.06	16%	0.62	0.06	14%	0.63	0.05	10%	0.62	0.03	11%
	0.80	0.03	31%	0.75	0.04	25%	0.74	0.04	14%	0.75	0.03	11%
	0.89	0.02	13%	0.85	0.02	18%	0.88	0.05	3%	0.86	0.03	5%
8	0.13	0.06	3%	0.11	0.06	3%	0.16	0.06	3%	0.14	0.05	6%
	0.25	0.06	4%	0.22	0.05	7%	0.27	0.06	12%	0.25	0.05	12%
	0.36	0.05	15%	0.33	0.05	14%	0.35	0.05	32%	0.35	0.05	32%
	0.44	0.05	14%	0.42	0.04	15%	0.43	0.05	21%	0.43	0.05	14%
	0.57	0.05	7%	0.55	0.05	9%	0.54	0.05	9%	0.52	0.04	12%
	0.67	0.06	16%	0.66	0.06	16%	0.67	0.04	10%	0.64	0.03	9%
	0.81	0.02	29%	0.78	0.02	22%	0.75	0.03	11%	0.75	0.02	10%
	0.89	0.02	12%	0.86	0.02	14%	0.88	0.04	3%	0.86	0.03	5%

Number of hidden states	AdML 70 bp up stream			AdML 110 bp up stream			H2B 80 bp up stream			H2B 80 bp down stream		
	μ	σ	*f*	μ	σ	*f*	μ	σ	*f*	μ	σ	*f*
9	0.12	0.06	3%	0.11	0.05	2%	0.15	0.06	3%	0.14	0.05	5%
	0.24	0.06	4%	0.21	0.05	6%	0.25	0.06	6%	0.22	0.05	9%
	0.34	0.05	11%	0.32	00.05	12%	0.32	0.05	19%	0.31	0.05	19%
	0.42	0.05	14%	0.40	0.04	16%	0.38	0.05	27%	0.38	0.04	26%
	0.52	0.05	9%	0.51	0.05	7%	0.46	0.05	15%	0.49	0.05	16%
	0.66	0.05	11%	0.61	0.06	12%	0.57	0.05	7%	0.62	0.03	10%
	0.68	0.07	9%	0.71	0.04	13%	0.68	0.04	10%	0.73	0.03	8%
	0.82	0.02	28%	0.80	0.02	25%	0.76	0.03	9%	0.80	0.01	6%
	0.89	0.02	11%	0.88	0.01	6%	0.88	0.04	3%	0.87	0.03	3%
10	0.11	0.05	2%	0.08	0.05	1%	0.14	0.06	2%	0.13	0.05	4%
	0.21	0.06	3%	0.19	0.05	5%	0.23	0.05	5%	0.20	0.05	7%
	0.30	0.06	7%	0.29	0.05	10%	0.30	0.06	14%	0.30	0.04	17%
	0.38	0.05	15%	0.38	0.04	15%	0.37	0.05	30%	0.37	0.04	24%
	0.45	0.05	10%	0.47	0.05	10%	0.44	0.05	18%	0.45	0.05	12%
	0.56	0.05	7%	0.59	0.05	9%	0.54	0.05	8%	0.53	0.04	10%
	0.69	0.05	12%	0.63	0.06	6%	0.66	0.04	8%	0.63	0.03	9%
	0.68	0.07	8%	0.71	0.03	13%	0.73	0.03	9%	0.74	0.02	7%
	0.82	0.01	26%	0.81	0.01	25%	0.79	0.03	5%	0.81	0.02	7%
	0.89	0.02	10%	0.88	0.01	5%	0.90	0.03	2%	0.88	0.02	2%

Table S4: Results from various HMM Analyses. The FRET efficiencies μ with their residual standard deviations σ beyond shot-noise broadening and the relative occurrences for a global hidden Markov model analysis with 1 to 10 states for all four samples after addition of NC2.

	Results from Global four-well HMM				**Results from a Global HMM to a linear four-well model**			
Transition Probability Matrix	0.9659	0.0313	0.0022	0.0005	0.9658	0.0342	0.0000	0.0000
	0.0087	0.9459	0.0405	0.0049	0.0093	0.9431	0.0476	0.0000
	0.0007	0.0509	0.8170	0.1314	0.0000	0.0575	0.8050	0.1375
	0.0001	0.0034	0.0789	0.0005	0.0000	0.0000	0.0870	0.9130
Rates (s^{-1}) $k_{row \rightarrow column}$	-	6.26	0.44	.01	-	6.84	0.00	0.00
	1.74	-	8.10	0.98	1.86	-	9.52	0.00
	0.14	10.18	-	26.28	0.00	11.50	-	27.5
	.02	0.68	15.78	-	0.00	0.00	17.4	-
Dwell times (ms)	144.3	89.9	24.7	58.1	143.6	85.3	23.1	55.0
Log-Likelihood	9.0771e+004				9.0670e+004			
BIC	1.8140e+005				1.8110e+005			
χ^2	0.008742				0.010625			

Table S5: Transition Rates Matrices. The transition rate matrices and rate matrix for the four state HMM analysis and for a linear four-well model plotted in the presence and absence of direct transitions between the E = 0.40 and E = 0.83 FRET efficiency states.

References

1. Hirsch, M., R. J. Wareham, M. L. Martin-Fernandez, M. P. Hobson, and D. J. Rolfe. 2013. A stochastic model for electron multiplication charge-coupled devices--from theory to practice. PloS one 8:e53671.
2. Borner, R., D. Kowerko, M. Hadzic, S. L. B. Konig, M. Ritter, and R. K. O. Sigel. 2018. Simulations of camera-based single-molecule fluorescence experiments. PloS one 13:e0195277.
3. Schwarz, G. 1978. Estimating the dimension of a model. The annals of Statistic 6:461-464.
4. Irizarry, R. A. 2001. Information and Posterior Probability Criteria for Model Selection in Local Likelihood Estimation. Journal of the American Statistical Association 96:303-315.
5. McKinney, S. A., C. Joo, and T. Ha. 2006. Analysis of single-molecule FRET trajectories using hidden Markov modeling. Biophysical journal 91:1941-1951.