Video Article

Using Three-color Single-molecule FRET to Study the Correlation of Protein Interactions

Markus Götz[1], Philipp Wortmann[1], Sonja Schmid[1,2], Thorsten Hugel[1]

[1]Institute of Physical Chemistry, University of Freiburg

[2]Department of Bionanoscience, Kavli Institute of Nanoscience Delft, Delft University of Technology

Correspondence to: Thorsten Hugel at thorsten.hugel@pc.uni-freiburg.de

URL: https://www.jove.com/video/56896
DOI: doi:10.3791/56896

Keywords: Biochemistry, Issue 131, Biophysics, protein kinetics, multi-color smFRET, TIRF, HMM, Hsp90, cooperativity, single-molecule FRET

Date Published: 1/30/2018

Citation: Götz, M., Wortmann, P., Schmid, S., Hugel, T. Using Three-color Single-molecule FRET to Study the Correlation of Protein Interactions. *J. Vis. Exp.* (131), e56896, doi:10.3791/56896 (2018).

Abstract

Single-molecule Förster resonance energy transfer (smFRET) has become a widely used biophysical technique to study the dynamics of biomolecules. For many molecular machines in a cell proteins have to act together with interaction partners in a functional cycle to fulfill their task. The extension of two-color to multi-color smFRET makes it possible to simultaneously probe more than one interaction or conformational change. This not only adds a new dimension to smFRET experiments but it also offers the unique possibility to directly study the sequence of events and to detect correlated interactions when using an immobilized sample and a total internal reflection fluorescence microscope (TIRFM). Therefore, multi-color smFRET is a versatile tool for studying biomolecular complexes in a quantitative manner and in a previously unachievable detail.

Here, we demonstrate how to overcome the special challenges of multi-color smFRET experiments on proteins. We present detailed protocols for obtaining the data and for extracting kinetic information. This includes trace selection criteria, state separation, and the recovery of state trajectories from the noisy data using a 3D ensemble Hidden Markov Model (HMM). Compared to other methods, the kinetic information is not recovered from dwell time histograms but directly from the HMM. The maximum likelihood framework allows us to critically evaluate the kinetic model and to provide meaningful uncertainties for the rates.

By applying our method to the heat shock protein 90 (Hsp90), we are able to disentangle the nucleotide binding and the global conformational changes of the protein. This allows us to directly observe the cooperativity between the two nucleotide binding pockets of the Hsp90 dimer.

Video Link

The video component of this article can be found at https://www.jove.com/video/56896/

Introduction

Many proteins fulfill their function in dynamic complexes with other molecules, mediated by conformational changes and transient associations on a broad range of timescales[1,2,3]. Coupled to an external energy source (*e.g.,* ATP) these dynamic interactions can lead to directionality in a functional cycle and ultimately maintain the non-equilibrium steady-state in a cell, the prerequisite for life.

In order to fully understand these molecular machines, a static description guided by structural studies is not sufficient. In addition, it is essential to have knowledge of the underlying kinetic model and to determine the kinetic rate constants. Several existing methods allow researchers to study the dynamics of binary interactions between two molecules of interest, *e.g.,* surface plasmon resonance, relaxation methods with a spectroscopic readout (*e.g.,* jump or stopped-flow techniques), and nuclear magnetic resonance. However, their applicability is in most cases limited to simple two-state systems (*e.g.,* one bound and one unbound state) due to the averaging inherent to bulk experiments. In cases where more states or intermediates are involved, they yield only a complex mixture of the rate constants. Single-molecule methods such as optical or magnetic tweezers or two-color smFRET, *i.e.,* one donor and one acceptor fluorophore, with a surface-immobilized sample can recover the rate constants for all observed conformational changes. However, when it comes to interactions affecting more than one binding site, these methods remain limited and the information on the possible correlation of the two (or more) interactions will only be accessible via indirect conclusions from a set of experiments.

Multi-color smFRET[4,5,6,7,8,9] offers the opportunity to study the interaction between these components directly, at real time and under near-physiological conditions[10]. This permits one to investigate for example, the conformation-dependent binding of a ligand or another protein[8,9,11]. The overall approach presented here is to label the protein(s) of interest at specific positions, to attach one protein to the surface of the measurement chamber, and to track the fluorescence intensity over time on a prism-type TIRFM (for details see [9,12]). The spatial proximity of the different dyes can then be determined from the energy transfer between them. Labeling strategies may vary from protein to protein (reviewed in [13]) and guidelines to avoid artifacts in smFRET measurements exist[14].

Since a donor dye may transfer energy to different acceptor dyes in a multi-color smFRET experiment, the relative position of all dyes is not accessible from excitation of one dye alone[15,16]. But in combination with alternating laser excitation (ALEX[17], and reviewed in [18]) this method provides all spatio-temporal information at sub-second and sub-nanometer resolution.

In principal, high resolution structural information can be achieved by using the inter-dye distances calculated from the combination of all fluorescence intensities in a multi-color smFRET experiment with ALEX. However, here we focus on state identification and separation as well as the extraction of kinetic models, where multi-color smFRET is indispensable. When "only" structure determination by triangulation is desired, a set of simpler two-color smFRET experiments with high signal-to-noise ratio can be performed[12,19].

We use the partial fluorescence (PF_{em}^{ex}) as a proxy for the energy transfer between two fluorophores[7]. The *PF* is calculated from the fluorescence intensity analogous to the FRET efficiency of a two-color experiment:

$$PF_b^a = \frac{I_b^a}{\sum_{i=a}^{c} I_i^a}$$

Where, I_{em}^{ex} is the intensity in emission channel *em* after excitation with color *ex*, and *c* is the acceptor with the longest wavelength. Detection channels represent the same position in the sample chamber but record different spectral ranges of the fluorescence light. The same identifier for excitation and emission are used in this protocol (*i.e.,* "blue," "green," and "red").

Because of experimental shortcomings the measured fluorescence intensities depend not only on the energy transfer but also on fluorophore and setup properties. In order to obtain the true energy transfer efficiency between two fluorophores, the measured intensities have to be corrected. The following procedure is based on reference[9]. Correction factors for apparent leakage (*lk*, *i.e.,* the detection of photons from a fluorophore in a channel designated for another dye) and apparent gamma (*ag*, *i.e.,* the fluorescence quantum yield of the dye and the detection efficiency of the channel) are obtained from single-molecule traces that show an acceptor bleaching event.

The leakage of the donor dye into every possible acceptor channel is calculated from all data points in the recorded fluorescence traces where the acceptor dye bleached but the donor is still fluorescent ($^{\dagger}I$):

$$lk_b^a = {}^{\dagger}I_b^a / {}^{\dagger}I_a^a$$

The median of the leakage histogram is used as the apparent leakage factor. After correction for leakage, the apparent gamma factor is determined from the same set of traces. It is calculated by dividing the change of fluorescence in the acceptor channel by the change of fluorescence in the donor channel upon bleaching of the acceptor dye:

$$ag_a = \frac{\Delta I_c^a}{\Delta I_a^a} = \frac{I_c^a - {}^{\dagger}I_c^a}{{}^{\dagger}I_a^a - I_a^a}$$

Where *c* again is the detection channel for the acceptor with the longest wavelength. The median of the resulting distribution is used as the apparent correction factor.

The corrected intensities in each channel are obtained by:

$${}^{corr}I_b^a = \left(I_b^a - \sum_{i=a}^{b-1} I_i^a \cdot lk_b^i \right) \cdot ag_b$$

The *PF* is then calculated according to:

$$PF_b^a = \frac{{}^{corr}I_b^a}{\sum_{i=a}^{c} {}^{corr}I_i^a}$$

Different populations can be separated in the multi-dimensional space spanned by the *PF*s. The position and width of each state is determined by fitting the data with multi-dimensional Gaussian functions. Subsequent optimization of one global HMM based on all *PF* traces provides a quantitative description of the observed kinetics. Even small changes of the rates are detectable.

HMMs provide a way of inferring a state model from a collection of noisy time traces. The system is considered to be in one of a set of discrete, hidden states at any given time and the actual observation (*i.e.,* the emission) is a probabilistic function of this hidden state[20]. In the case of TIRFM smFRET data, the emission probabilities b_i per state *i* can be modeled by continuous Gaussian probability density functions. At regularly spaced discrete time points, transitions from one to another state can occur according to the transition probability that is time-invariant and only depends on the current state. The transition matrix *A* contains these transition probabilities a_{ij} between all hidden states. The initial state distribution Π gives the state-specific probabilities π_i for the first time point of a time trace. Using a maximum-likelihood approach, these parameters can be optimized to best describe the data with the Forward-Backward and Baum-Welch algorithms[20,21]. This yields the maximum likelihood estimators (MLE). Finally, the state sequence that most likely produced the trajectory of observations can be inferred with the Viterbi algorithm. In contrast to other HMM analyses of smFRET data[24,25,26] we do not use the HMM as a mere "smoothing" of the data but extract the kinetic state model from the data set without the need for fitting dwell time histograms[27]. HMM analysis is done with in-house scripts using

Igor Pro. Implementation of the code is based on reference[21]. We provide a software kit and exemplary data on our webpage in order to follow sections 5 and 6 of this protocol (https://www.singlemolecule.uni-freiburg.de/software/3d-fret). Full software is available upon request.

Time points in the data with *PF* <-1 or *PF* >2 in any detection channel are assigned the minimal emission probability for all states (10^{-200}). This prevents artificial transitions at these data points.

The parameters for the emission probabilities are obtained from the fit of the 3D *PF* histogram with Gaussian functions as described in step 5.7. These parameters are kept fixed during the optimization of the HMM.

In the presented approach, the initial state distribution vector and the transition matrix are used globally to describe the entire ensemble of traces. They are updated based on all *N* molecules from the data set according to reference[27].

Start parameters for the initial state distribution are determined from 2D projections of the *PF* histogram (step 5.3) and the transition probabilities are set to 0.05 with the exception of the probabilities to stay in the same state, which are chosen such that the probability to leave a certain state is normalized to unity.

A likelihood profiling method is used to give confidence intervals (CIs) for all transition rates[21,22], which serve as meaningful estimates for their uncertainty. To calculate the bounds of the CI for a specific rate, the transition probability of interest is fixed to a value other than the MLE. This yields the test model λ'. A likelihood ratio (*LR*) test of the likelihood $\mathcal{L}(\lambda'|O)$ given the data set *O* is performed according to:

$$LR = 2 \cdot (\ln \mathcal{L}(\lambda^{\mathrm{MLE}}|O) - \ln \mathcal{L}(\lambda'|O))$$

The 95% confidence bound for the parameter is reached when *LR* exceeds 3.841, the 95% quantile of a x^2-distribution with one degree of freedom[22,23].

The power of the method is demonstrated using the Hsp90. This abundant protein is found in bacteria and eukaryotes and is part of the cellular stress response[28]. It is a promising drug target in cancer treatment[29]. Hsp90 is a homodimer with one nucleotide binding pocket in the N-terminal domain of each subunit[30]. It can undergo transitions between at least two globally distinct conformations, one closed and one N-terminal open, V-shaped conformation[19,31,32]. The dimeric nature directly raises the question of the interplay between the two nucleotide binding sites in Hsp90.

In the following, we provide a step-by-step protocol for the data acquisition and analysis of a three-color smFRET experiment on yeast Hsp90 and nucleotide. The conformation-dependent binding of fluorescently labeled AMP-PNP (AMP-PNP*, a non-hydrolyzable analog of ATP) is analyzed. The application of the described procedure permits the study of the nucleotide binding and at the same time the conformational changes of Hsp90 and thereby reveals the cooperativity between the two nucleotide binding pockets of Hsp90.

Protocol

1. Setup and Prerequisites

1. Perform the multi-color smFRET measurements on a prism-type TIRFM. A description of a two-color setup as a JoVE publication is given in reference[12].
    1. Construct a multi-color TIRFM. A general layout is detailed in [9].
    2. Use switchable, diode pumped solid state continuous wave lasers, which render the use of mechanical shutters in the excitation paths unnecessary.
    3. Employ an asymmetric, elongated prism that prevents the back reflection of the excitation beam from the rear side to enter the objective.
    4. Use 2-inch achromatic aspheric fused silica lenses in the detection paths that collect as much light as possible and prevent auto-fluorescence and aberrations, *e.g.,* distortions in off-center regions of the image.
    5. Focus each detection path on the chip of the EMCCD with a separate lens. This allows optimal focusing of each detection channel. **Caution:** Class 3B lasers are used in the TIRFM. This means they are hazardous if the eye is exposed directly, but diffuse reflections are not harmful. Ensure compliance with laser safety precautions according to local government regulations before the system is operated.

2. Determine the correction factors for setup and fluorophore properties beforehand using dsDNA samples.
    1. Use one high-FRET dsDNA sample for each dye in combination with the acceptor having the longest excitation wavelength (for the presented setup: Atto488-Atto647N, Atto550-Atto647N, Atto594-Atto647N). Make sure that the DNA is additionally modified with a biotin.
    2. Dilute the sample to 5 nM with TNM buffer (5 mM Tris pH 7.5, 5 mM NaCl, 20 mM $MgCl_2$) and 2 mM Trolox (use this buffer also for measurement).
    3. Immobilize the dsDNA as described for Hsp90 in steps 2.5 and 2.7.
    4. Calculate the correction factors for apparent leakage (*lk*) and apparent gamma (*ag*) from single-molecule traces that show an acceptor bleaching event.

3. Construct a flow chamber that is a sandwich of a PEG/biotin-PEG passivated quartz slide, a thin film that is adhesive on both sides, and a cover slip. For detail protocols for the cleaning of the quartz slides and passivation see reference[9].
    1. Use thick (3 mm) quartz slides to geometrically prevent the collection of laser light scattered at the quartz-glycerol-quartz interface between the prism and the functionalized quartz slide.
    2. Use a thin (40 µm) sealing film that is sprayed with adhesive on the non-adhesive side. The thin film reduces the distance between the surface-attached molecules and objective. Heat to 80 °C and press on.

3. Place a cover slip on the top. Heat to 80 °C and press on. Use a drop of glycerol when placing the flow chamber over the prism.
   NOTE: The materials of the prism and the quartz slides as well as the glycerol are index-matched.

4. Express Hsp90 from *Saccharomyces cerevisiae* in the form of two single cysteine point mutants at the positions D61 or Q385. Add a C-terminal coiled-coil motif to prevent dimer dissociation at picomolar concentrations. Label the mutant proteins separately and exchange the monomers to obtain heterodimers labeled with Atto488 at amino acid position 61 and Atto550 at amino acid position 385[9].

2. Measurement

1. Start the camera software and set the imaging parameters as specified below:
   1. Set the temperature for the sensor cooling as low as possible (-95 °C with external water cooling) to decrease the dark current noise.
   2. Use camera settings that are optimized for single-molecule recording: 3.3 µs vertical shift speed, normal vertical clock voltage, 17 MHz 16-bit horizontal read out, pre-amplification gain 3, gain of electron multiplier 1,000.
   3. Set the triggering of the acquisition to "External" and the exposure time to 70 ms. Record movies with a length of 750 acquisition cycles.
      NOTE: Turn off the room light when the camera is acquiring to prevent saturation of the EMCCD sensor.

2. Create a folder on the local SSD for the measurement. In the software settings go to the <Auto-Save> rider, enable <Auto-save>, and choose file format "Tiff" for movie acquisition. Select the folder on the SSD as auto-save location.
3. Start the software that controls the acousto-optical tunable filter (AOTF), the software that controls the operation of the lasers and the trigger software that synchronizes the lasers, AOTF, shutters in the detection path, and cameras. Adjust the laser power with the AOTF (ca. 3 mW before entering the prism) and load the correct triggering pattern.
4. Mount the sample holder with the prism and the flow chamber, attach tubing, place the inlet tubing in a microcentrifuge cup, and connect the outlet tubing to a syringe pump. Flush the chamber with approximately 150 µL buffer, align the excitation beam, and focus. Bleach any fluorescent contaminants on the surface by slowly moving along the complete detection range of the slide with a laser power of about 10 mW for all lasers. This takes about 1 h.
   NOTE: If not stated otherwise, the used buffer contains 40 mM HEPES pH 7.5, 150 mM KCl, and 10 mM $MgCl_2$.
5. Flush approximately 300 µL of a NeutrAvidin solution (0.25 mg/mL in buffer) into the chamber and incubate for 1 min. Flush out unbound NeutrAvidin with buffer and flush approximately 300 µL of a BSA solution (0.5 mg/mL in buffer) through the chamber.
   NOTE: This block remaining surface functionalization defects by unspecific adsorption of BSA to the surface.
6. Immobilize the sample by loading the flow chamber with approximately 150 µL of the biotinylated and labeled Hsp90 at rising concentrations (diluted in buffer + 0.5 mg/mL BSA) until a sufficient surface density is reached, which is usually the case at a concentration of 5 - 10 pM. Wash out unbound protein with approximately 300 µL buffer + 0.5 mg/mL BSA.
7. Flush in 150 µL of 25 nM AMP-PNP* in buffer + 0.5 mg/mL BSA. Let it incubate for 5 min and repeat this step once to ensure correct nucleotide concentration.
   NOTE: For experiments in the presence of additional, unlabeled AMP-PNP, also add 250 µM AMP-PNP.
8. Begin with the data acquisition. For an appropriate amount of data acquire approximately 20 movies which takes about 1.5 h.
   1. Move the position of the sample chamber perpendicular to the excitation beam with a piezo stepper to change the field of view.
   2. Adjust the focus with the z-piezo that controls the height of the objective if needed. This should not be necessary too often when the measurement chamber is mounted without inclination.
   3. Prepare the recording of the cameras by pressing <Take Signal> in the camera software and start the excitation/acquisition cycles using the <Start> button in the trigger software. This starts the acquisition of the fluorescence intensity.

9. Perform a channel registration by first recording a movie with fluorescent beads that show fluorescence emission in the spectral range of all detection channels of the setup. Then, detect bead positions in the calibration movie by searching for the brightest spots and determine the central position from a Gaussian fit to the intensity profile. Save the coordinates of beads that are found in all channels and fit both the mapping offset in x- and in y-direction with a 2D polynomial of degree three[9].

3. Selection of Single-molecule Traces

1. Data analysis is done with in-house scripts using Igor Pro. Load all necessary scripts by opening "iniTIRF.ipf" and select the correct type of experiment.
   NOTE: In the following, <Button> specifies clickable elements in the menu or the user interface. Function calls are indicated in quotes, *e.g.,* "Print "Hello World"". These commands can be pasted to the command line of Igor Pro (without the enclosing quotes).
2. Make sure the parameters for the detection channel registration are loaded.
3. Start the GUI by clicking <smFRET new|Analysis GUI>.
4. Load the movies (*i.e.,* the sequence of frames with 512 x 512 pixels stored as 16 bit TIFF stacks) that hold the intensity in the respective channels. Do this by pressing the <Load Film> button and selecting the files from the so-called "master" and "slave" cameras one after the other.
5. Identify the positions of potential single molecules by searching for the brightest spots in the sum of the first five frames in a certain detection channel. Calculate the corresponding positions in all other detection channels from the channel mapping. To obtain the fluorescence intensity trace, sum the intensity of a pixel square around the central position for each frame. Do this by pressing the button <Find Traces> in the GUI.
   NOTE: The side length of the square (in pixel) is given by: 2 * <Sum Pxs> + 1.
6. For each molecule, calculate a joint raw intensity trace as the sum of all traces of this spot with the same excitation color. Evaluate the intensity profile of the molecule in all channels for the following criteria:
   1. A roughly flat plateau in the joint raw intensity and a single bleaching step for all excitation colors, anticorrelated behavior in the appropriate detection channels, detection of red fluorescence (indicating FRET to a bound AMP-PNP*) at least once within the trace, and no multiple steps in the red fluorescence, which would indicate the presence of two AMP-PNP* bound to one Hsp90 dimer.

7. Save the fluorescence traces of the spot for further analysis if these criteria are fulfilled. Do this by selecting the trace with the cursor and then pressing the <Save> button in the "timelines" graph. Manually inspect the intensity traces in all three detection channels after blue excitation for about 200 molecules per movie.

4. Calculation of the Partial Fluorescence Traces

1. Display all fluorescence intensity traces for one of the saved molecules. Use the cursors in the graph to select time ranges.
2. Select a time interval where all fluorophores are bleached already. The mean background intensity is calculated from this range and subtracted from the intensity trace in each channel. Do this by pressing the <Background> button.
3. Select the FRET efficiency range, where at least both the dyes attached to Hsp90 (Atto488 and Atto550) are present. Make sure to exclude traces that contain a blinking event (see **Figure 3B**). These events are characterized by a drop-in fluorescence intensity in one channel without an accompanying increase in any other channel.
4. Calculate the *PF* traces. Do this by pressing the <PF Calc> button. The predefined correction factors for apparent leakage (*lk*) and apparent gamma (*ag*) are applied to the raw intensity in order to correct for photo-physical and setup properties.

5. Population Selection and 3D Histogram Fitting

1. Remove molecules that show a low signal-to-noise ratio in the *PF* traces. Molecules that exceed the interval [-1;2] in any *PF* trace for more than 10% of the frames are removed from the data set. Do this by executing "RemoveTracesLowSNR()" from the Command Window.
2. Calculate binned 2D projections of the *PF* data. Plot PF_{red}^{green} over PF_{red}^{blue} and PF_{green}^{blue} over PF_{red}^{blue} in the range [-0.5; 1.5] with a resolution of 100 x 100 bins. To do so, execute:
    1. "HistFret2D("r_b", "r_g", binHist=100)"
    2. "HistFret2D("r_b", "g_b", binHist=100); MoveWindow 553.5, 42.5, 1055.25, 508.25"
3. Determine the relative population of each distinguishable state in the 2D projections.
    1. Bring the appropriate graph to the front and execute "panelHist2DCount()".
    2. Press the <Init> button and draw a free-hand polygon around the peak.
    3. Click the <Count> button. The number of data points in the polygon and the total number of data points in the projection are printed in the Command Window.
4. Prepare a 3D histogram of the *PF* data by executing "HistFret3D("g_b", "r_b", "r_g")".
5. Normalize the 3D histogram to an integral of 1. Execute the following:
    1. "NewDataFolder/S fit0"
    2. "Duplicate/O ::FRET:Hist3D, Hist3D"
    3. "Variable /G div = sum(Hist3D)*(DimDelta(Hist3D,0))^3"
    4. "Hist3D /= div; Print div"
6. Provide initial parameters for the 3D Gaussian fit and prepare necessary data structures.
    1. Execute "Gauss3D_initParam(); edit W_coef_old."
    2. Add the state populations to the end of the parameter vector.
    3. Execute "Gauss3D_prepareFit()."
    Note: W_coef_old is a vector that holds the initial parameters for the fit. Per state this means $\{\mu_x, \mu_y, \mu_z, \sigma_x^2, \sigma_{xy}, \sigma_{xz}, \sigma_{yx}, \sigma_y^2, \sigma_{yz}, \sigma_{zx}, \sigma_{zy}, \sigma_z^2\}$ and the state population, which is concatenated at the end of the vector. Make sure that the covariance matrix is symmetric.
7. Fit the sum of *S* 3D Gaussian functions to the 3D *PF* histogram, with *S* being the number of distinguishable states.
    1. Execute "do3D()." This may take an hour or more on a normal office PC, depending on the quality of the initial parameters.
    2. Execute "postprocessFitMultiGauss3D(); evalFitMultiGauss3D(); edit W_coef."
8. Display the fit result. For each of the two 2D projections, use the following commands:
    1. "contourPF3D_new(0); contourPF3D_new(1); contourPF3D_new(2); contourPF3D_new(3); contourPF3D_new(4)"
    2. "contourPF3D_colorize()"

6. Kinetic Analysis with 3D Ensemble HMM

1. Prepare an ensemble HMM run to extract the kinetic information. One HMM is optimized from all molecules in the data set. Use the information obtained in the previous step to define the position and width of each state in the 3D *PF* space.
    1. Initialize the HMM user interface (<HMM|Init HMM>) and choose the appropriate number of states (in the case of the Hsp90 data this means <NumStates> = 5), number of dimensions of the input signal (<NumDims> = 3), and the type of the input (<Input Type> = "FRET 3D bgr").
2. Optimize the parameters of the HMM by letting the software converge the likelihood of the HMM (execute "prepENS_CONVERGE_gB(GetDataFolder(1), -14)") until the change in the transition matrix compared to the previous iteration falls below a threshold (10^{-14} for the sum of the absolute change for each transition probability). This yields the MLE for the transition probabilities in about an hour on a normal office PC.

3. Repeat the population selection, Gaussian fitting, and the HMM optimization for subsets of the data (*e.g.,* 75% of the full data set). If the full data set was merged from different experiments, repeat the optimization also for each of the single experiments. Analysis of subsets permits to estimate the uncertainty of the manual population selection and the variability within the data set.
4. Calculate the CI for the transition probabilities, which report on the data set heterogeneity and the precision of the HMM.
    1. Get a rough estimate of the CI bounds by executing "cd $(root:path3Dimport + "HMM"); loop_getCI_estimate_limits()."
    2. Calculate the exact limits of the CI by executing:
    3. "loop_getCI_HMM_converge(1)"
    4. "CIresults_conv_new()"
    5. "cd ::HMM_CIresult; reportCI_conv()"
    6. "cd ::cmp_CI_conv; CI_plot2("HMM", doAppend=0)"

5. Condense the available kinetic information to simplify the interpretation.
    1. Collect information about the time a labeled nucleotide stays bound to Hsp90. Do this by collapsing the states that are bound to AMP-PNP* (*i.e.,* S_1, S_2 and S_3, see also **Figure 2**) and compile the dwell time histogram. Execute "cd $(root:path3Dimport +"HMM"); collapse_states_get_DT({0,1,1,1,0})", which combines dwells of states S_0 and S_4 as well as S_1, S_2, and S_3.
    2. Extract the dwell times from the Viterbi path for each state of interest and compare the resulting dwell time histogram for different experimental conditions. Execute "plot_collapsed_DT_Hist (wDTo_01110_record1)."
    3. For a more detailed picture, collapse states that have distinct *PF* but are functionally identical to ease the further data analysis, *e.g.*, in the case of a three-color experiment with labeled Hsp90 and nucleotide, states S_2 and S_3 can be collapsed. Execute "collapse_states_get_DT({0,1,2,2,4})."

Representative Results

Multi-color smFRET measurements allow the direct detection of correlation between two or more distinct interaction sites. This renders the technique unique to investigate multi-component systems, such as protein complexes. We focus on the presentation of a three-color smFRET experiment here, which serves as an illustrative example.

The general workflow of the method is shown in **Figure 1**. The first part comprises the recording of multi-color smFRET data on a prism-type TIRF microscope. The surface attachment strategy and the schematics of the setup are depicted in **Figure 2A**. For a more detailed description of the setup, refer to reference[9]. The second part of the presented method focuses on the data analysis. Exemplary fluorescence intensity traces are shown in **Figure 2B**. Suitable time traces show: (i) a clear bleaching step for both fluorophores attached to Hsp90, (ii) flat intensity plateaus, (iii) anti-correlated behavior in the corresponding channels, and (iv) at least one binding event of AMP-PNP* (**Figure 3**). More than 400 molecules that meet the selection criteria were selected to yield reliable statistics. In the studied system, five states can be distinguished by the fluorescence intensities, with four states being functionally distinct (**Figure 2C**).

From the fluorescence intensity traces the partial fluorescence (*PF*, the extension of the FRET efficiency for multi-color smFRET experiments) can be calculated (**Figure 4A**). The *PF* is related to the proximity of the dyes. In a three-color smFRET experiment, the data spans a 3D space (**Figure 5B**). 2D projections of the 3D histogram have proven to be useful for state separation (**Figure 4B**, C). In a successful experiment, all states that are theoretically expected under the applied experimental conditions are distinguishable by their *PF* in the 2D projections.

The relative populations of the states are determined from the 2D projections by drawing free-hand polygons that enclose the peak (**Figure 5C**). This approach was found to be accurate and reliable[9]. The emission probabilities for the HMM are obtained by fitting the 3D *PF* histogram with the sum of 3D Gaussian functions, where *S* is the number of distinguishable states (five in the presented case; **Figure 5D**). For this fit to converge properly, only the relative population p_i of each state *i* has to be held constant while the position and width of the Gaussians are free.

One ensemble HMM is optimized over the full data set with the emission probabilities fixed to the parameters obtained from the Gaussian fit (**Figure 6**). In order to get a measure for the uncertainty of the extracted transition probabilities, the 95% CI for each transition is determined (**Figure 7**).

In addition, the average time that the labeled reporter nucleotide AMP-PNP* remains bound to Hsp90 can be extracted under different experimental conditions (**Figure 8A**). This helps to further reduce the complexity of the results. To do so, states that represent AMP-PNP* bound and unbound conformations are collapsed in the state trajectories, respectively. From this the average dwell time for the AMP-PNP* dissociation can be calculated (**Figure 8B**).

By comparing the kinetics observed in the absence and the presence of additional, unlabeled AMP-PNP, unique information on the correlation between the conformational changes of Hsp90 and the nucleotide state can be gained. This makes it possible to directly study the cooperativity between the two nucleotide binding pockets of Hsp90. Besides, it circumvents the need for titration experiments that measure the binding site occupation as a function of the substrate concentration (*e.g.,* Hill plots). For highly dynamic protein systems such as Hsp90, this approach is also sensitive to small changes in the rates[11].

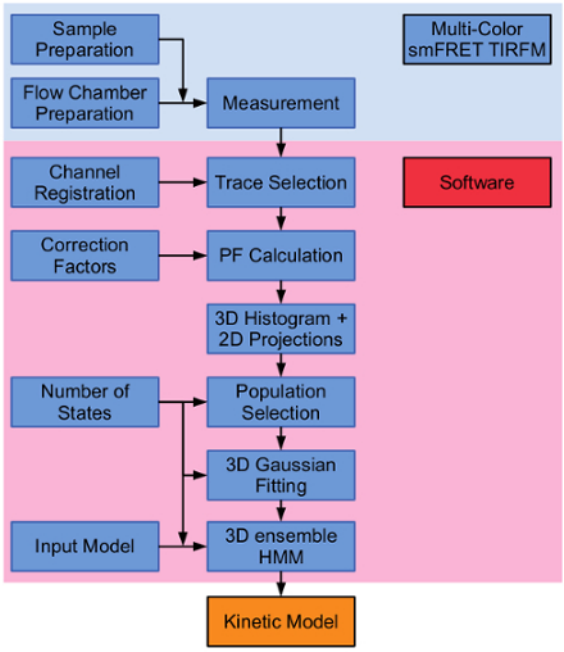

Figure 1: General workflow of the data acquisition and analysis. Data is acquired on a multi-color smFRET total internal reflection fluorescence microscope (TIRFM). After trace selection and calculation of the partial fluorescence (*PF*), the 3D *PF* histogram and 2D projections thereof are compiled. Using the 2D projections, the population of all distinguishable states can be determined. These are used as a constraint for a 3D Gaussian fit to the *PF* histogram. The Gaussian probability density functions serve as emission probabilities for the subsequent 3D ensemble Hidden Markov Modeling (HMM). This yields the kinetic description of the system. Please click here to view a larger version of this figure.

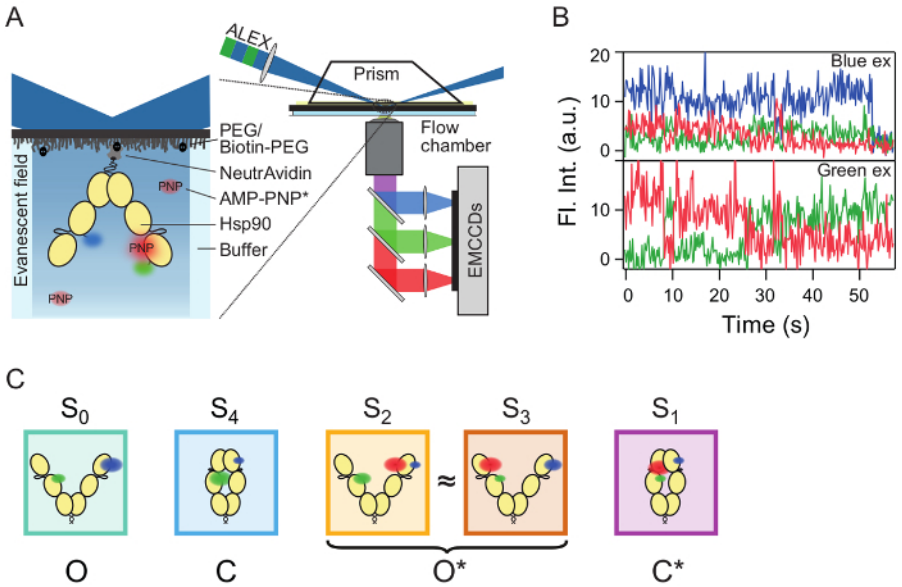

Figure 2: Scheme of the data acquisition. (**A**) Pictogram of the studied system consisting of an Hsp90 dimer (yellow ovals represent the domain structure) attached to the surface with the labels Atto488 (blue) and Atto550 (green) and the reporter nucleotide AMP-PNP* in solution, labeled with Atto647N (red). Data are recorded on a prism-type TIRF microscope with alternating laser excitation (ALEX). (**B**) Exemplary fluorescence intensity (Fl. Int.) traces after blue and green excitation. (**C**) Pictograms of the distinguishable conformational states of Hsp90 (S_0, S_1, S_2, S_3, S_4) and their respective identifier for the functional state (O, C, O*, C*) used in this work. Fluorophore positions are indicated in blue, green, and red. Two populations represent the same functional state, namely open Hsp90 with AMP-PNP* bound (S_2 and S_3). Please click here to view a larger version of this figure.

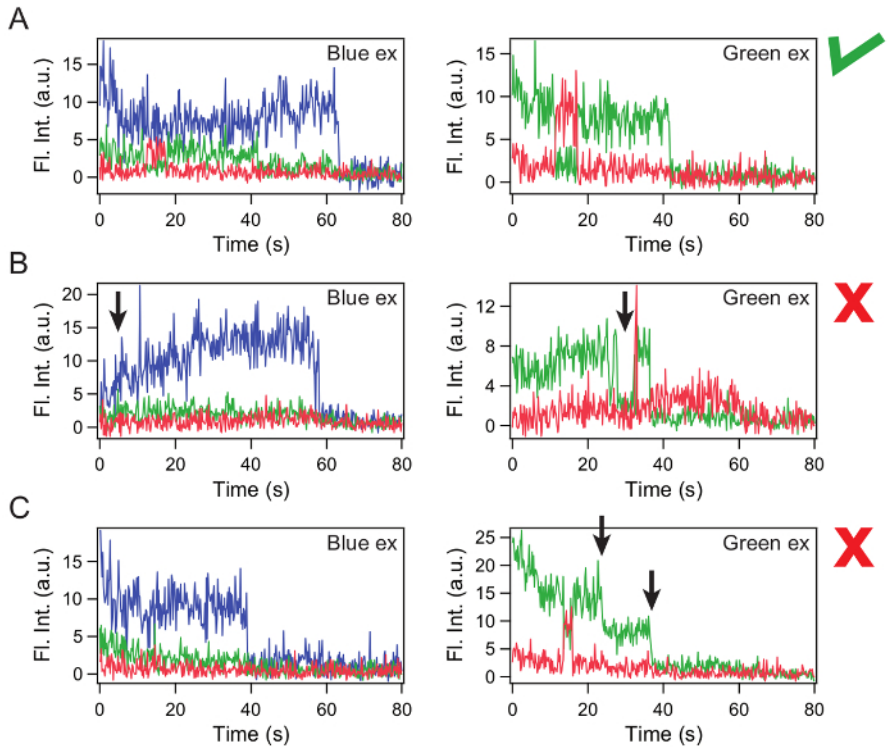

Figure 3: Selection criteria. (**A**) A molecule selected for further analysis. (**B**, **C**) Molecules not selected for further analysis. (**B**) No flat plateaus and Atto550 is in a dark state around 30 s after green excitation (indicated by arrows). (**C**) Multiple bleaching steps after green excitation (indicated by arrows). Fl. Int.: fluorescence intensity, blue: Atto488, green: Atto550, red Atto647N, ex: excitation. Please click here to view a larger version of this figure.

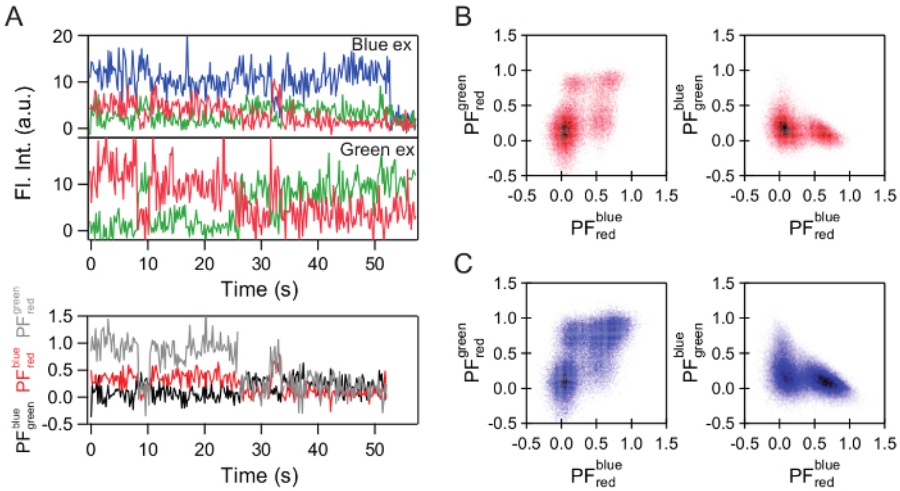

Figure 4: Calculation of the *PF* traces and representative histograms. (**A**) Representative fluorescence intensity (Fl. Int.) traces and the corresponding partial fluorescence (*PF*) traces. (**B**) Two 2D projections of the 3D *PF* histogram for the measurement in absence of additional, unlabeled nucleotide. (**C**) The same projections for the experiment in the presence of additional 250 µM AMP-PNP. Please click here to view a larger version of this figure.

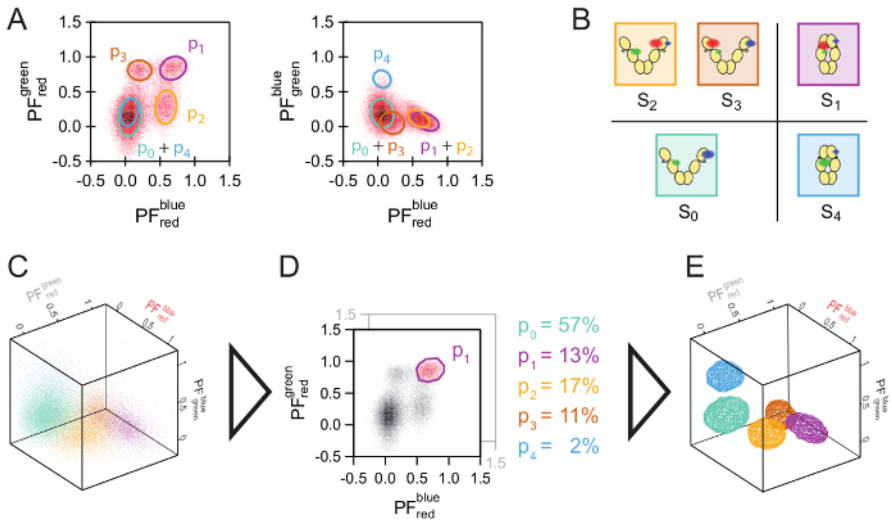

Figure 5: Population selection process. (**A**) The position of the five distinguishable populations in the two 2D projections. (**B**) Copy of the pictograms in **Figure 2C**. (**C**) Representative 3D scatter plot of the *PF* data. Coloration of the data points is only for visualization. The same color code as in B is used. (**D**) Determination of the relative populations is done by drawing free-hand polygons around the peaks in the 2D histogram. Using a combination of the two projections depicted in **Figure 3**, all five populations can be distinguished. (**E**) Results of the 3D Gaussian fit to the histogram of the *PF* data shown in A. Depicted are the isosurfaces at FWHM, which represent the five different populations. Please click here to view a larger version of this figure.

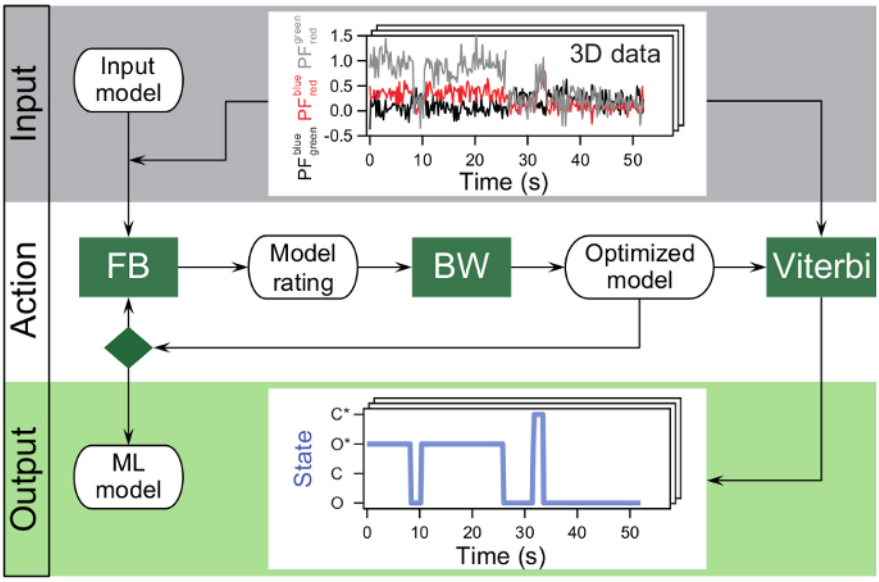

Figure 6: Flow chart of the ensemble 3D HMM optimization. One model is optimized for all molecules from the data set. The start values are given by the input model (with a pre-defined number of states). The likelihood of the model given the data (the 3D *PF* traces) is rated with the Forward-Backward (FB) algorithm. The Baum-Welch (BW) algorithm yields a local maximum likelihood estimate (MLE) of the parameters. The global MLE then can be found iteratively. The Viterbi algorithm computes the most likely state trajectory given a model. Please click here to view a larger version of this figure.

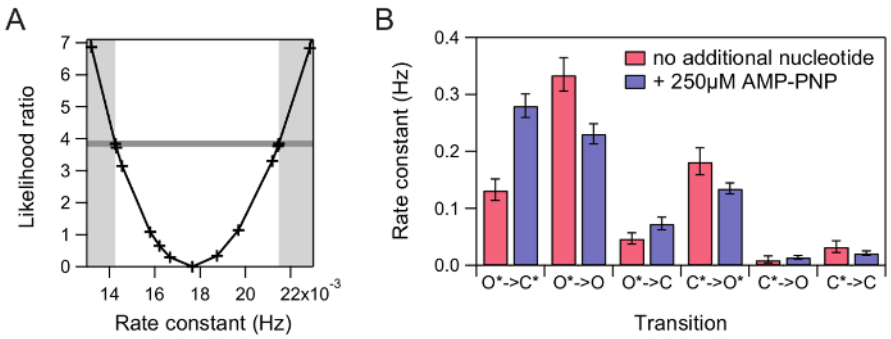

Figure 7: Meaningful uncertainty estimation with CI. (**A**) Determination of the CI for one exemplary rate constant. The likelihood ratio (LR) of the test model compared to the maximum likelihood estimator (MLE) model is calculated in the region around the MLE for the rate constant. The 95% confidence bound is reached when the LR exceeds 3.841 (horizontal dark grey line). (**B**) The extracted rate constants without additional nucleotide (red) and with AMP-PNP (blue) and their 95% CI. Please click here to view a larger version of this figure.

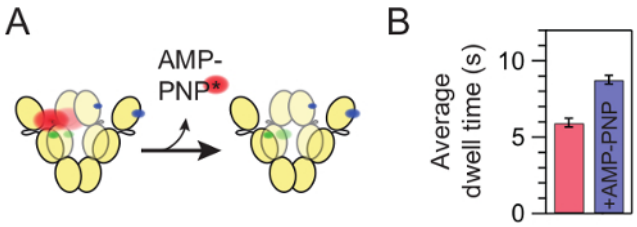

Figure 8: The average dwell time of the reporter nucleotide AMP-PNP* bound to Hsp90 is prolonged in presence of additional nucleotide. (**A**) Pictogram of the observed dissociation of labeled AMP-PNP* from the Hsp90 dimer, averaging over all conformations of Hsp90. The domain structure of Hsp90 is depicted by yellow ovals and the conformational flexibility is indicated by overlaying an open and closed dimer. (**B**) Average dwell time of AMP-PNP* bound to Hsp90 in absence of additional nucleotide (red) and in presence of 250 µM unlabeled AMP-PNP (blue). Please click here to view a larger version of this figure.

Discussion

We present the experimental procedure to obtain three-color smFRET data for a complex protein system and a step-by-step description of the analysis of these measurements. This approach offers the unique possibility to directly assess the correlation between multiple interaction sites or conformational changes.

In order to obtain suitable multi-color single-molecule data on proteins it is important to perform reproducible measurements at a low noise level. This can be achieved by using an efficient and reliable surface passivation protocol for the flow chamber[9]. A sufficient density of molecules immobilized to the surface should be used to increase the yield of molecules per recorded movie. This means that fluorescence light from single molecules should fall on about 5 - 10% of the pixels in the recorded images. At the same time overloading should be prevented, because it would result in an overlap of neighboring molecules. When a labeled species is free in solution (AMP-PNP* in the presented study), make sure the concentration is low enough to not interfere with the measurement by direct excitation from lasers with shorter wavelength (*i.e.,* typically well below a micromolar concentration). Also, ensure that the concentration of these compounds is not lower than intended due to unspecific binding to interfaces. Additionally, the optimal excitation power represents a trade-off between the signal-to-noise ratio and the photobleaching of the fluorophores.

No oxygen scavenging system is employed in the presented protocol. The used Atto-Tec dyes do not show significant blinking on the time scale of the experiment (70 ms illumination per excitation cycle) and no decrease of the photo bleach rate was observed by a scavenger system consisting of glucose oxidase, catalase, and Trolox[33,34]. Also, absence of oxygen scavenger avoids artifacts due to unspecific protein interactions, as oxygen scavenging systems usually contain protein components up to a micromolar concentration[35].

Another critical step is the trace selection. The size of the area for summing the intensity of a single fluorophore is chosen to be as small as possible but still larger than the point spread function of the setup. Be aware that a flat plateau in the joint intensity traces can only be expected if the detection channels have similar apparent gamma factors, since the traces are not corrected at this stage of the analysis. Only molecules that meet well-defined criteria for the fluorescence intensity traces are included in the analysis (Section 3). Molecules with a poor signal-to-noise ratio in the *PF* traces are excluded from further analysis and do not affect data evaluation as clear selection criteria are used.

Depending on the data quality, alternative approaches provide optimal state allocation. A free multi-dimensional ensemble HMM based either on the fluorescence intensity traces[27] or the *PF* traces could be applied to allocate the underlying states by optimizing the corresponding emission probabilities (*e.g.,* Gaussian PDFs) while at the same time optimizing the kinetic model. Both approaches have their advantages and disadvantages. *PF* traces may contain unfavorable spikes while fluorescence intensity levels vary from molecule to molecule. The presented protocol solves this issue by preventing artificial transitions when spikes occur and by estimating the emission probabilities from a 3D Gaussian fit to the *PF* histogram and only optimizing the transition probabilities in the subsequent ensemble HMM run.

The presented approach is limited by the need for a labeled interaction partner that binds with a comparably high affinity in order to be compatible with the low concentrations necessary for conventional single-molecule measurements. This may be overcome by strategies to increase the local concentration (tethering of the interaction partners[36], encapsulation in vesicles[37,38]) or methods to decrease the excited volume (*e.g.,* zero-mode waveguides[39]). In addition, labeling positions have to be chosen with care and control experiments are necessary to exclude severe side-effects of the fluorophores. In case of Hsp90, this is usually checked by an ATPase assay, which proves the enzymatic activity[40]. If available, structural information from crystal or NMR structure should be considered. Preferred positions are in surface-exposed loops, away from interaction interfaces and not buried in a protein surface pocket. This ensures a large accessible volume for the dye.

The framework is suited for a multi-color smFRET experiment with an arbitrary number of colors. We focus on three-color measurements here, as these already provide the feature that renders them distinct from other single-molecule experiments, namely the possibility to directly observe correlation between two processes (*e.g.,* binding of an interaction partner and conformational changes). This information is inaccessible in a standard two-color smFRET experiment but is of fundamental interest for understanding the function of any molecular machine.

The presented experiment and analysis is capable of characterizing protein systems, which might display dynamics on a wide range of time scales between states that are highly flexible themselves[3]. This contrasts previously published multi-color smFRET experiments, which focused on DNA or RNA model systems, such as Holliday junctions[4], exhibiting transitions between (mostly two) well-defined states. In protein systems, the signal-to-noise ratio is lower and extracting the complete set of transition probabilities is difficult because of the limited temporal bandwidth in smFRET experiments due to photobleaching. Here, we show that - if applied carefully and with reasonable constraints - the obstacles in measuring protein systems can be overcome with multi-color smFRET using multi-dimensional state allocation and ensemble hidden Markov analysis[9].

Disclosures

The authors declare no conflicts of interest.

Acknowledgements

This work is funded by the German Research Foundation (INST 39/969-1) and the European Research Council through the ERC Grant Agreement n. 681891.

References

1. Nooren, I.M.A., & Thornton, J.M. Diversity of protein-protein interactions. *EMBO J.* **22** (14), 3486-92 (2003).
2. Marsh, J.A., & Teichmann, S.A. Structure, dynamics, assembly, and evolution of protein complexes. *Annu Rev Biochem.* **84**, 551-75 (2015).
3. Henzler-Wildman, K., & Kern, D. Dynamic personalities of proteins. *Nature.* **450** (7172), 964-72 (2007).
4. Hohng, S., Joo, C., & Ha, T. Single-Molecule Three-Color FRET. *Biophys J.* **87** (2), 1328-37 (2004).
5. Person, B., Stein, I.H., Steinhauer, C., Vogelsang, J., & Tinnefeld, P. Correlated movement and bending of nucleic acid structures visualized by multicolor single-molecule spectroscopy. *ChemPhysChem.* **10** (9-10), 1455-60 (2009).
6. Lee, J., Lee, S., Ragunathan, K., Joo, C., Ha, T., & Hohng, S. Single-molecule four-color FRET. *Angew Chem Int Ed.* **49** (51), 9922-5 (2010).
7. Ratzke, C., Berkemeier, F., & Hugel, T. Heat shock protein 90's mechanochemical cycle is dominated by thermal fluctuations. *Proc Natl Acad Sci U S A.* **109** (1), 161-6 (2012).
8. Ratzke, C., Hellenkamp, B., & Hugel, T. Four-colour FRET reveals directionality in the Hsp90 multicomponent machinery. *Nat Commun.* **5**, 4192 (2014).
9. Götz, M., Wortmann, P., Schmid, S., & Hugel, T. A Multicolor Single-Molecule FRET Approach to Study Protein Dynamics and Interactions Simultaneously. *Methods Enzymol.* **581**, 487-516 (2016).
10. Yengo, C.M., & Berger, C.L. Fluorescence anisotropy and resonance energy transfer: Powerful tools for measuring real time protein dynamics in a physiological environment. *Curr Opin Pharmacol.* **10** (6), 731-7 (2010).
11. Wortmann, P., Götz, M., & Hugel, T. Cooperative Nucleotide Binding in Hsp90 and Its Regulation by Aha1. *Biophys J.* **113**, 1711-1718 (2017).
12. Dörfler, T., Eilert, T., Röcker, C., Nagy, J., & Michaelis, J. Structural Information from Single-molecule FRET Experiments Using the Fast Nano-positioning System. *J Vis Exp.* (120), e54782-e54782 (2017).
13. Stephanopoulos, N., & Francis, M.B. Choosing an effective protein bioconjugation strategy. *Nature chemical biology.* **7** (12), 876-84 (2011).
14. Sánchez-Rico, C., Voith von Voithenberg, L., Warner, L., Lamb, D.C., & Sattler, M. Effects of Fluorophore Attachment on Protein Conformation and Dynamics Studied by spFRET and NMR Spectroscopy. *Chemistry (Weinheim an der Bergstrasse, Germany).* (2017).
15. Roy, R., Hohng, S., & Ha, T. A practical guide to single-molecule FRET. *Nat Methods.* **5** (6), 507-16 (2008).
16. Lee, N.K., *et al.* Three-color alternating-laser excitation of single molecules: monitoring multiple interactions and distances. *Biophys J.* **92** (1), 303-12 (2007).
17. Kapanidis, A.N., Lee, N.K., Laurence, T.A., Doose, S., Margeat, E., & Weiss, S. Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc Natl Acad Sci U S A.* **101** (24), 8936-41 (2004).
18. Hohlbein, J., Craggs, T.D., & Cordes, T. Alternating-laser excitation: single-molecule FRET and beyond. *Chem Soc Rev.* **43** (4), 1156-71 (2014).
19. Hellenkamp, B., Wortmann, P., Kandzia, F., Zacharias, M., & Hugel, T. Multidomain structure and correlated dynamics determined by self-consistent FRET networks. *Nat Methods.* **14**, 174-80 (2017).
20. Rabiner, L.R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE.* **77** (2), 257-86 (1989).
21. Fink, G.A. *Markov Models for Pattern Recognition.* Springer London, London (2014).
22. Giudici, P., Ryden, T., & Vandekerkhove, P. Likelihood-Ratio Tests for Hidden Markov Models. *Biometrics.* **56** (3), 742-7 (2000).
23. Visser, I., Raijmakers, M.E.J., & Molenaar, P.C.M. Confidence intervals for hidden Markov model parameters. *Br J Math Stat Psychol.* **53** (2), 317-27 (2000).

24. McKinney, S.A., Joo, C., & Ha, T. Analysis of Single-Molecule FRET Trajectories Using Hidden Markov Modeling. *Biophys J.* **91** (5), 1941-51 (2006).
25. Bronson, J.E., Fei, J., Hofman, J.M., Gonzalez, R.L., & Wiggins, C.H. Learning Rates and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-Molecule FRET Data. *Biophys J.* **97** (12), 3196-205 (2009).
26. Greenfeld, M., Pavlichin, D.S., Mabuchi, H., & Herschlag, D. Single Molecule Analysis Research Tool (SMART): an integrated approach for analyzing single molecule data. *PLoS ONE.* **7** (2), e30024 (2012).
27. Schmid, S., Götz, M., & Hugel, T. Single-Molecule Analysis beyond Dwell Times: Demonstration and Assessment in and out of Equilibrium. *Biophys J.* **111** (7), 1375-84 (2016).
28. Taipale, M., Jarosz, D.F., & Lindquist, S. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat Rev Mol Cell Biol.* **11** (7), 515-28 (2010).
29. Trepel, J., Mollapour, M., Giaccone, G., & Neckers, L. Targeting the dynamic HSP90 complex in cancer. *Nat Rev Cancer.* **10** (8), 537-49 (2010).
30. Wayne, N., & Bolon, D.N. Dimerization of Hsp90 is required for in vivo function. Design and analysis of monomers and dimers. *J Biol Chem.* **282** (48), 35386-95 (2007).
31. Ali, M.M.U*., et al.* Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature.* **440** (7087), 1013-7 (2006).
32. Southworth, D.R., & Agard, D.A. Species-dependent ensembles of conserved conformational states define the Hsp90 chaperone ATPase cycle. *Mol Cell.* **32** (5), 631-40 (2008).
33. Blanchard, S.C., Kim, H.D., Gonzalez, R.L., Puglisi, J.D., & Chu, S. tRNA dynamics on the ribosome during translation. *Proc Natl Acad Sci U S A.* **101** (35), 12893-8 (2004).
34. Aitken, C.E., Marshall, R.A., & Puglisi, J.D. An Oxygen Scavenging System for Improvement of Dye Stability in Single-Molecule Fluorescence Experiments. *Biophys J.* **94** (5), 1826-35 (2008).
35. Swoboda, M*., et al.* Enzymatic oxygen scavenging for photostability without pH drop in single-molecule experiments. *ACS Nano.* **6** (7), 6364-9 (2012).
36. Rognoni, L., Stigler, J., Pelz, B., Ylänne, J., & Rief, M. Dynamic force sensing of filamin revealed in single-molecule experiments. *Proc Natl Acad Sci U S A.* **109** (48), 19679-84 (2012).
37. Okumus, B., Wilson, T.J., Lilley, D.M.J., & Ha, T. Vesicle encapsulation studies reveal that single molecule ribozyme heterogeneities are intrinsic. *Biophys J.* **87** (4), 2798-806 (2004).
38. Boukobza, E., Sonnenfeld, A., & Haran, G. Immobilization in Surface-Tethered Lipid Vesicles as a New Tool for Single Biomolecule Spectroscopy. *J Phys Chem B.* **105** (48), 12165-70 (2001).
39. Levene, M.J., Korlach, J., Turner, S.W., Foquet, M., Craighead, H.G., & Webb, W.W. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science.* **299** (5607), 682-6 (2003).
40. Panaretou, B*., et al.* ATP binding and hydrolysis are essential to the function of the Hsp90 molecular chaperone in vivo. *EMBO J.* **17** (16), 4829-36 (1998).

