Multi-parameter photon-by-photon hidden Markov modeling

Paul David Harris*[1], Shimon Weiss[2,3], and Eitan Lerner†[1,4]

[1]Department of Biological Chemistry, The Alexander Silberman Institute of Life Sciences, Faculty of Mathematics & Science, The Edmond J. Safra Campus, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel
[2]Department of Chemistry and Biochemistry, and Department of Physiology, University of California, Los Angeles, California, United States
[3]Department of Physiology, CaliforniaNanoSystems Institute, University of California, Los Angeles, California, United States
[4]The Center for Nanoscience and Nanotechnology, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel

April 8, 2021

Abstract

Single molecule FRET (smFRET) is a useful tool for studying biomolecular sub-populations and their dynamics. Advanced smFRET-based techniques often track multiple parameters simultaneously, increasing the information content of the measurement. Photon-by-photon hidden Markov modelling (H^2MM) is a smFRET analysis tool that quantifies FRET dynamics of single biomolecules, even if they occur in sub-milliseconds. However, sub-populations can be characterized by additional experimentally-derived parameters other than the FRET efficiency. We introduce multi-parameter H^2MM (mpH^2MM) that identifies sub-populations and their transition dynamics based on multiple experimentally-derived parameters, simultaneously. We show the use of this tool in deciphering the number of underlying sub-populations, their mean characteristics and the rate constants of their transitions for a DNA hairpin exhibiting milliseconds FRET dynamics, and for the RNA polymerase promoter open complex exhibiting sub-millisecond FRET dynamics of the transcription bubble. Overall, we show that using mpH^2MM facilitates the identification and

*harripd@gmail.com
†eitan.lerner@mail.huji.ac.il

quantification of biomolecular sub-populations in smFRET measurements that are otherwise difficult to identify. Finally we provide the means to use mpH^{2}MM in analyzing FRET dynamics in advanced multi-color smFRET-based measurements.

1 Introduction

The role of structural dynamics in biomolecular function has come to the forefront of biophysical research[1, 2]. Biomolecules in solution exhibit structural dynamics at a hierarchy of timescales and modes, from bond rotations to movements of entire globular domains, occurring at times from picoseconds to seconds and slower[3]. In many cases, the stages in the functions that biomolecules carry are promoted by different sub-populations of closely-related structures, better known as conformations. This is indeed the case in the coupling of catalysis with substrate binding domain dynamics in some enzymes[4, 5], the dynamics of the DNA bubble in transcription initiation to support transcription start site selection[6, 7], involved in DNA mismatch repair[8], to facilitate protein translocation across membranes[9], assisting chaperone action[10], involved in the allosteric regulation of the AAA+ disaggregase[11], influences the selectivity in ABC importers[12], and many other important bio-processes, in which structural dynamics is coupled to biological function or influences it[1, 2]. Thus, methods capable of identifying and characterizing distinctly time-separated structural sub-populations of biomolecules are of great interest in biomolecular sciences and in dynamic structural biology.

NMR- and EPR-based methods[13–16] and single-molecule methods[17–21] have come to the forefront, each with their own advantages and disadvantages. Single-molecule methods allow measuring one biomolecule at a time and tracking different experimental parameters simultaneously. One of the most popular single-molecule approaches is single-molecule FRET (smFRET)[22], where the bio-molecule of interest is site-specifically labeled at two strategic points with two fluorescent dyes, which can exhibit transfer of the excitation energy from the donor dye to the acceptor dye with a probability (or efficiency; E) that is inversely proportional to the sixth power of the distance between them according to the phenomenon of Fluorescence- (or Förster) resonance energy transfer (FRET)[23–25]. In the absence of factors correcting for differences in donor and acceptor fluorescence quantum yields and detection efficiencies (better known as the γ factor), acceptor direct excitation, and leakage of donor photons into the acceptor channel, the raw ratio of acceptor photons to donor and acceptor photons, better known as the proximity ratio (PR), serves as the apparent or uncorrected FRET efficiency[26].

smFRET has proven a powerful tool to disentangle sub-populations of biomacromolecules undergoing dynamic transitions over a range of time scales, but remains limited by the time resolution of the apparatus, which in the case of freely diffusing molecules, is often on the order of a few milliseconds, and down to the sub-milliseconds using advanced analyses of photon statis-

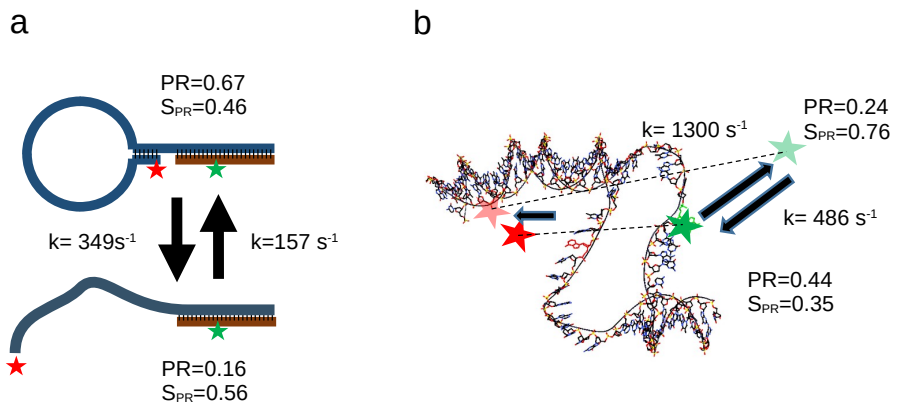

Figure 1: **Cartoon representations of a) DNA hairpin and b) transcription bubble $\mathbf{RP_o}$** conformational changes, along with mpH^{2}MM derived PR, and S_{PR} values and the accompanying transition rates.

tics within single-molecule photon bursts, summarized in recent reviews of the field[1, 2]. Photon-by-photon hidden Markov modelling (H^{2}MM)[27] can extract the number of sub-populations involved in the dynamics, their proximity ratio values and transition rate constants. Nevertheless, while advanced smFRET setups often detect multiple parameters beyond the simple proximity ratio, such as in alternating-laser excitation (ALEX; also known as pulsed-interleaved excitation, PIE)[28, 29] or in multi-color smFRET-based measurements[30–37], H^{2}MM in its current iteration only reports on the proximity ratio of a single donor-acceptor pair of dyes.

Here,we introduce multi-parameter H^{2}MM (mpH^{2}MM), which enables incorporation of multiple parameters, through additional photon streams in nanosecond ALEX (nsALEX)(a.k.a. PIE)[28, 29] experiments. We apply this approach to 1) a DNA hairpin exhibiting well-defined two-state opening and closing dynamics[38], and 2) with the transcription bubble in the RNA polymerase (RNAP) promoter open (RP_o) complex, exhibiting sub-millisecond dynamics (figure 1)[6, 7]. We show that mpH^{2}MM is able to extract sub-populations based on both the nsALEX-relevant mean parameters, PR and the stoichiometry, S_{PR}, as well as their transition rate constants, demonstrating FRET-relevant conformational transitions, as well as FRET-irrelevant photo-physical transitions.

2 Results

2.1 DNA hairpin exhibiting millisecond dynamics

Tsukanov *et.al.* introduced a DNA hairpin with a loop containing 31 adenines and a six base-pair stem[38]. The opening and closing rate constants of the hairpin varied as a function of the GC content of the stem as well as the sodium chloride concentration. When appropriately labeled with a FRET donor and acceptor pair (ATTO 550 and ATTO 647N, respectively), the open and closed hairpin sub-populations resulted in low and high mean PR values, respectively. The hairpin containing two GCs out of the six stem bases, exhibited opening and closing rates of a few milliseconds. Taken together with the knowledge of the opening and closing transition rates at each sodium chloride concentration, this DNA construct serves as an ideal model system to test the performance of H^2MM.

We perform two-color nsALEX (2c-nsALEX) experiments[29] with this construct at a concentration of 300 mM sodium chloride, in which the open and closed hairpin sub-populations have been shown as equally populated[38]. The burst variance analysis (BVA)[39] plot shows clear evidence of within-burst dynamics occurring while the molecule crosses the confocal volume (figure 2a). However, it is not necessarily clear how many distinct sub-populations are actually involved in the within-burst dynamics leading to the outcome shown in the BVA plot. In visual examination of the 2D E-S plot, three sub-populations are apparent: 1) an open hairpin sub-population with mean PR ∼0.2 and 2) a closed hairpin sub-population, with a PR ∼0.65, both of which have mean S_{PR} values ∼0.5, and 3) a third sub-population with a mean PR ∼0, and mean S_{PR} ∼1, where the acceptor is either in a dark state, possibly a triplet state (i.e., dark acceptor sub-population), or missing altogether (Figure 2b). The 2D E-S plot also exhibits bursts with intermediate PR values bridging between the open and closed hairpin sub-populations, and corresponding to bursts where the hairpin is undergoing opening and closing transitions while crossing the confocal volume.

Analysis with spH^2MM and mpH^2MM can be performed using any given state model. However, we first must select the ideal state model from several ones, where the models differ in the number of states present, and hence also the number of mean PR values and number of transition rate constants. Discriminating over- and under-fit state models from the ideal model has been difficult using H^2MM[7, 40]. Lerner *et al.* have previously introduced the modified Bayes information criterion (BIC'), that does not provide an extremum-based decision on the most likely state-model[7]. We now implement the integrated complete likelihood (ICL)[41, 42], which gets a minimum value for the most-likely state-model. While it should not be viewed as an absolute discriminator, using smFRET simulated data, where the ground truth is known, we find that the ICL consistently chooses the correct number of states (see jupyter notebooks in supplementary dataset[43]). Based on the ICL, we find that for spH^2MM analysis of the DNA hairpin data, the two-state model is the state-model that best

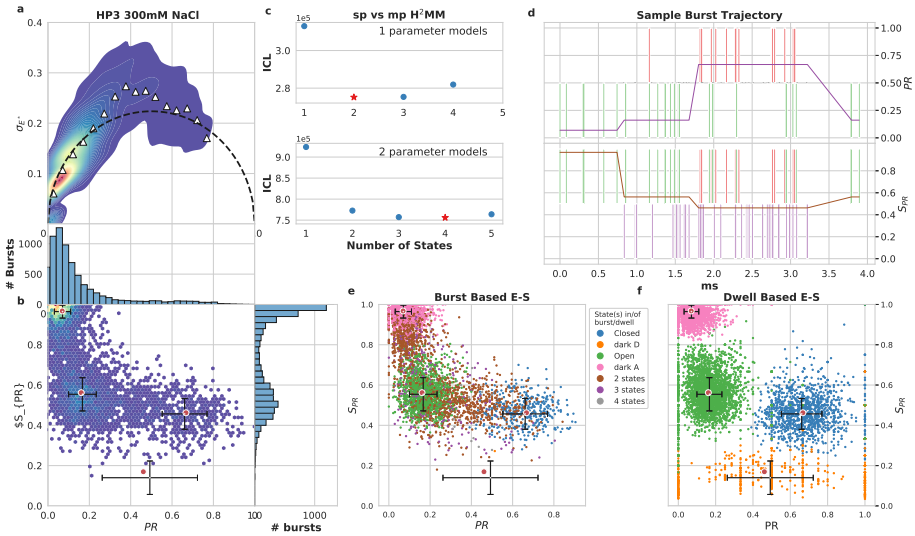

Figure 2: mpH^{2}MM results for DNA hairpin at 300 mM NaCl. a) Burst variance analysis (BVA), the standard deviation of *PR* values of bursts is displayed versus their *PR* values. Bursts with standard deviations higher than expected solely from shot noise (semicircle), are ones that include dynamic heterogeneity, such as within-burst FRET dynamics. Triangles indicate the average of standard deviation values per *PR* bin. b) 2D histogram of *PR* and S_{PR} (E-S plots, colloquially) of bursts. The *PR* and S_{PR} values of sub-populations derived from mpH^{2}MM are marked by red circles, and the standard deviation of these values, derived from the *Viterbi* dwell time analysis, are marked by black crosses. c) Comparison of values of the integrated complete likelihood (ICL) of spH^{2}MM (top panel) and mpH^{2}MM (bottom panel) of optimized models with different state-models. The ideal state-model is marked in red. d) A sample burst trajectory, with photons represented as colored vertical bars, with donor excitation photons colored green or red for donor and acceptor, respectively, and acceptor excitation photons colored purple. *PR* (top panel) and S_{PR} (bottom panel) of sub-populations determined from dwells using the *Viterbi* algorithm, are overlayed on the photon bars. e,f) E-S scatter plots of data processed by the *Viterbi* algorithm. mpH^{2}MM sub-populations and *viterbi*-derived standard deviations are overlayed as red circles and black crosses, respectively. e,f) E-S scatter plot of bursts (e) or dwells within bursts (f), color coded by which states are present in the bursts (e) or according to the state of the dwell (f), according to *Viterbi* algorithm.

explains the donor-excitation data. However, after using mpH^{2}MM, it turns out that the four-state model is the state-model that best explains all of the data, including the acceptor-excitation data.
Visual inspection of the one-dimensional projection of burst data onto the *PR* parameter immediately suggests an explanation for this discrepancy, as it appears as only two sub-populations. The dark acceptor sub-population and the open hairpin FRET sub-population exhibit similar low *PR* values, and are hard to distinguish. This projection reflects the data accessible to spH^{2}MM, the donor excitation streams, and thus the open hairpin and dark acceptor sub-populations are expected to have nearly identical signatures with regard to the streams accessible to spH^{2}MM, thus leading to the false inference of only two states. The open hairpin FRET sub-population and the dark-acceptor sub-population are, however, very much distinct with regard to the acceptor excitation stream, which is accessible to mpH^{2}MM. In the ICL-selected four-state model retrieved by mpH^{2}MM, two out of the four sub-populations match nicely with the sub-populations in the ICL-selected spH^{2}MM model, having similar *PR* values. The S_{PR} values are $\sim$0.5 (figure 2b red circles), as expected for molecules undergoing FRET. The third and fourth sub-populations in the model can be matched to dark acceptor and dark donor sub-populaions, respectively. The third state has a *PR* value $\sim$0 and a S_{PR} value $\sim$1 (figure 2b, top left red circle). This state has a clear sub-population of bursts associated with it in the 2D E-S plot. The fourth sub-population has an intermediate *PR* value, and a very low S_{PR} value of 0.17 (figure 2b, bottom red circle). There is no obvious sub-population visually observed in the E-S plots to which this would correspond, but the *PR* and S_{PR} values are consistent with this dark donor sub-population. More importantly, comparing the parameters of the state models retrieved by spH^{2}MM and mpH^{2}MM, we find that the transition rate constants derived using mpH^{2}MM are closer to those found by Tsukanov *et. al.*[38] than those extracted using spH^{2}MM (table S2, and supplementary csv files of all state-models found by H^{2}MM analysis[43]). The transition rate constants provide a clue as to why the fourth sub-population does not show up in the E-S plots, as they predict rare transitions to it, and fast transitions away. Thus, populating the fourth sub-population occurs only briefly and rarely in bursts undergoing rapid dynamics, such that it does not appear as a clear sub-population in the E-S plots (table S2, and supplementary csv file [43]).
The *Viterbi* algorithm finds the most-likely state path through each burst, given a state-model and its parameter values (figure 2d). We use this to classify bursts by which sub-populations are present within each burst (figure 2e), and separate photons into dwells, for which *PR* and S_{PR} can be defined (figure 2f). Visual examination of this plot shows that the *Viterbi* algorithm reasonably classifies most bursts that have *PR* and S_{PR} values close to the predicted value of a given sub-population as only having that sub-population present, as well as bursts with intermediate *PR* and S_{PR} that are predicted to include dwells of multiple sub-populations. Notably, there are only a few bursts classified as having dwells solely in the dark donor sub-population (figure 2e), keeping with what is predicted by the transition rates, and indeed, few dwells are even found in

this sub-population (figure 2f). In his study of photo-bleaching pathways in smFRET experiments, Kong *et. al.*[44] found that transitions of the donor to dark states are rare, while such transitions of the acceptor are frequent. The scarcity of the donor dark state in the *Viterbi* analysis serves to both confirm this observation, and prove the sensitivity of mpH^{2}MM at the same time.

2.2 RNAP-promoter open complex

Using two-color microsecond ALEX experiments, Robb *et. al.* has shown that the transcription bubble in the RNA polymerase (RNAP)-promoter open complex (RP_o) undergoes sub-millisecond FRET dynamics[6]. Later, using several double labeled RP_o constructs, Lerner *et. al.* used spH^{2}MM to show that the transcription bubble exhibits two FRET sub-populations associated with two transcription initiation complex forms, exhibiting sub-millisecond transitions, as well as free DNA and dark acceptor sub-populations[7]. We analyze 2c-nsALEX data using mpH^{2}MM on one of the constructs previously used by Lerner *et. al.*, where an ATTO 647N acceptor dye labels the base at register (-8) in the template strand, (-8)TA, and a Cy3B donor dye labels the base at register (-6) in the non-template strand, (-6)NTD, in a lacCONS promoter[45, 46]. Accordingly, we term this construct (-8)TA-(-6)NTD. The results of the analysis show that the four-state model minimizes the ICL (figure 3c). Based on the mean PR and S_{PR} values, the sub-populations resolved by mpH^{2}MM are similar to the ones previously observed by Lerner *et al.*[7]: 1) a dark acceptor sub-population, 2) a high mean PR free DNA sub-population, 3) an RNAP-bound open-bubble sub-population with intermediate mean PR (RP_o), and 4) the previously-observed RNAP-bound scrunched-bubble sub-population with a lower mean PR(scrunched RP_o) (figure 3b). These results compare quite well with the ones previously observed by Lerner *et al.*, but with the added benefit that the S_{PR} values are also recovered. The S_{PR} values are particularly interesting, as the RNAP-bound open-bubble sub-population exhibits an anomalously low S_{PR}, relative to that of the scrunched-bubble sub-population (figure 3e,f orange points). The half-life of the RNAP-bound open-bubble sub-population is shorter than the burst duration, and thus only a few transcription initiation complexes are expected to traverse the confocal volume while being solely in this sub-population. This is born out by the *Viterbi* analysis (figure 3e), and thus the anomalous S_{PR} would not be expected to be obvious from the E-S plot (figure 3b). Because the S_{PR} values are sensitive to the ratio between the donor and acceptor fluorescence quantum yields, this result might show that not only does the distance between the dyes change between open- and scrunched- bubble conformations, but also does the dyes' immediate environment, influencing the fluorescence quantum yields of one of the dyes. In that context, the substantial decrease in the S_{PR} value of the open-bubble sub-population relative to the value of the scrunched-bubble sub-population points to a decrease in the donor fluorescence quantum yield or an increase in the acceptor fluorescence quantum yield, or both. Whether it is the donor or the acceptor that change their fluorescence quantum yield can be assessed by inspecting whether the acceptor fluorescence

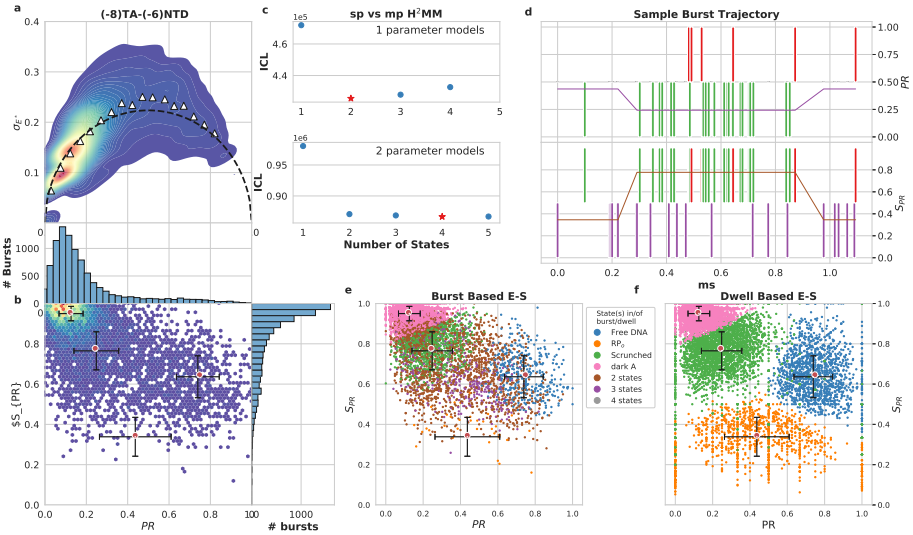

Figure 3: mpH^{2}MM results for the (-8)TA-(-6)NTD lacCONS promoter construct in RP$_o$, donor active burst selection. a) Burst variance analysis (BVA), the standard deviation of PR values of bursts is displayed versus their PR values. Bursts with standard deviations higher than expected solely from shot noise (semicircle), are ones that include dynamic heterogeneity, such as within-burst FRET dynamics. Triangles indicate the average of standard deviation values per PR bin. b) 2D histogram of PR and S_{PR} (E-S plots, colloquially) of bursts. The PR and S_{PR} values of sub-populations derived from mpH^{2}MM are marked by red circles, and the standard deviation of these values, derived from the *Viterbi* dwell time analysis, are marked by black crosses. c) Comparison of values of the integrated complete likelihood (ICL) of spH^{2}MM (top panel) and mpH^{2}MM (bottom panel) of optimized models with different state-models. The ideal state-model is marked in red. d) A sample burst trajectory, with photons represented as colored vertical bars, with donor excitation photons colored green or red for donor and acceptor, respectively, and acceptor excitation photons colored purple. PR (top panel) and S_{PR} (bottom panel) of sub-populations determined from dwells using the *Viterbi* algorithm, are overlayed on the photon bars. e,f) E-S scatter plots of data processed by the *Viterbi* algorithm. mpH^{2}MM sub-populations and *viterbi*-derived standard deviations are overlayed as red circles and black crosses, respectively. e) E-S scatter plot of bursts (e) or dwells within bursts (f), color coded by which states are present in the bursts (e) or according to the state of the dwell (f), according to *Viterbi* algorithm.

decay after acceptor direct excitation has changed or not when comparing between the sub-populations. Visual examination of the fluorescence decays of the open bubble RP_o and scrunched bubble RP_o sub-populations, indicate that the anomalous S_{PR} is certainly not due to an increase in the acceptor fluorescence quantum yield, as the normalized acceptor fluorescence decays after acceptor excitation of the two sub-populations are nearly identical, while there is a small difference in the donor decays (figure S13). Therefore, the anomalously lower S_{PR} value of the open bubble RP_o sub-population is due to a decreased fluorescence quantum yield of the donor labeling register (-6) of the nontemplate strand at the lacCONS promoter. This additional piece of information can be used when modeling the structure of RP_o, since the most common reasons for the change in fluorescence quantum yield are the rate of dynamic quenching of fluorescence by the solvent, and hence solvent exposure, and fluorescence quenching or enhancement by specific groups within RP_o.

3 Conclusions

Multi-parameter H^2MM increases both the information content of the results, and the sensitivity of the H^2MM algorithm to differences in the photon steams that are too subtle when examining only as single parameter. mpH^2MM is able to disentangle dark acceptor molecules from the very low FRET open hairpin conformation, and detect the low in S_{PR} of the RP_o open bubble conformation. With the added analysis of the fluorescent lifetimes, we determine that this is due to a decrease in donor fluorescence quantum yield, showing that the donor has a unique interaction with the protein in this conformation. This finding itself serves as additional information on the immediate microenvironment of the donor dye within the RP_o open bubble conformation, which can then be used in modelling the underlying structure.

As ALEX or PIE setups are now ubiquitous, the acceptor excitation stream is almost always available, and ignoring it is essentially throwing away useful data. mpH^2MM therefore maximizes the use of available data for characterizing rapidly interconverting sub-populations.

We have demonstrated mpH^2MM with 2c-nsALEX measurements, but it is by no means restricted to these setups. The most obvious application of mpH^2MM beyond 2c-nsALEX, is with the multiple photon streams in multi-color smFRET-based measurements[30–37]. Here, three or even four spectrally-distinct dyes are attached to the biomolecule of interest, and each produces a distinct photon stream. This enables the simultaneous observation of multiple inter-dye proximity ratios at once. If qualitative tests indicate that such a system is undergoing within-burst dynamics, mpH^2MM is well-suited to extract the transfer efficiencies relevant to the underlying dynamically-interconverting sub-populations. Applying these methods is as simple as assigning an index to each photon stream. The transition probability matrix will be unaffected, while the emission probability matrix will expand with more photon streams. Conversion of the emission probability matrix into useful $PR_{i,j}$ values can be achieved by taking

the values of the corresponding streams, and treating them as intensities. The supplied Jupyter notebooks provide examples for how to execute mpH^{2}MM using FRETbursts. Experimenters using other platforms must utilize their knowledge of the peculiarities of their data to properly filter and cast their data into the simple and general format that the H2MM_C package[47] accepts.

4 Methods

4.1 Experimental setup

We performed the 2c-nsALEX smFRET measurements of the doubly-labeled DNA hairpin construct[38] in the presence of 300 mM sodium chloride, using a confocal-based setup (ISS^{TM}, USA) assembled on top of an Olympus IX73 inverted microscope stand. We used a pulsed picosecond fiber laser (λ=532 nm, pulse width of 100 ps FWHM, operating at 20 MHz repetition rate and 100 µW measured at the back aperture of the objective lens) for exciting the Cy3B donor dye (FL-532-PICO, CNI, China), and a pulsed picosecond diode laser (λ=642 nm, pulse width of 100 ps FWHM, operating at 20 MHz repetition rate and 60 µW measured at the back aperture of the objective lens) for exciting the ATTO 647N acceptor dye (QuixX® 642-140 PS, Omicron, GmbH), delayed by 25 ns. The laser beams pass through a polarization-maintaining optical fiber and then further shaped by a linear polarizer and a halfwave plate. A dichroic beam splitter with high reflectivity at 532 and 640 nm (ZT532/640rpc, Chroma, USA) reflect the light through the optical path to a high numerical aperture (NA) super apochromatic objective (60X, NA=1.2, water immersion, Olympus, Japan), which focuses the light onto a small confocal volume. The microscope collects the fluorescence from the excited molecules through the same objective, and focuses it with an achromatic lens (f = 100 mm) onto a 100 µm diameter pinhole (variable pinhole, motorized, tunable from 20 µm to 1 mm), and then re-collimates it with an achromatic lens (f = 100 mm). Then, donor and acceptor fluorescence are split between two detection channels using a dichroic mirror with a cutoff wavelength at λ=652 nm (FF652-Di01-25x36, Semrock Rochester NY, USA). We further filter the donor and acceptor fluorescence from other light sources 585/40 nm (FF01-585/40-25, Semrock Rochester NY, USA) and 698/70 nm (FF01-698/70-25, Semrock Rochester NY, USA) band-pass filters, respectively, and detect the donor and acceptor fluorescence signals using two hybrid photomultipliers (Model R10467U-40, Hamamatsu, Japan), routed through a 4-to-1 router to a time-correlated single photon counting (TCSPC) module (SPC-150, Becker & Hickl, GmbH) as its START signal (the STOP signal is routed from the laser controller). We perform data acquisition using the VistaVision software (version 4.2.095, 64-bit, ISS^{TM}, USA) in the time-tagged time-resolved (TTTR) file format. After acquiring the data, we transform it into the photon HDF5 file format[48] for easy dissemination of raw data to the public, and easy input in the FRETbursts analysis software. The 2c-nsALEX measurements of the (-8)TA-(-6)NTD lacCONS promoter construct, in the absence and presence

of the *E. coli* RNAP holoenzyme, was performed using the setup in Ingargiola *et al.*[49].

4.2 Data Processing

All data processing and analysis is performed using Jupyter Notebooks available in supplementary dataset, along with the accompanying photon-HDF5 files containing the raw data[43]. We perform burst search and selection using the FRETbursts analysis software[50]. Background is assessed per each 30 seconds of acquisition, and bursts are identified as time periods were the instantaneous photon count rate of a sliding window of $m = 10$ consecutive photons is at least $F = 6$ times higher than the background rate. Bursts in the normal selection are selected if they include at least 30 photons in total between all streams. Visualizations are performed using FRETbursts' dplot function. Bursts identified by FRETbursts are then converted into a format readable by the H2MM_C software[47], by a simple function supplied in the Jupyter notebooks available in supplementary dataset[43]. In spH^2MM, only photons arising from donor excitation are considered, assigned to either donor or acceptor streams, identified by index 0 or 1 respectively, depending on at which detector they arrived. mpH^2MM also considers photons arriving during acceptor excitation, assigning these photons an index of 2. All H^2MM calculations were performed within the Jupyter notebooks, available in supplmentary dataset[43], using the Python package by Paul David Harris[47]. We use the H_2MM algorithm (both single- and multi-parameter) to test how well 2, 3 or 4 state-models describe the photon data. To choose the best model, we use the ICL[41, 42], where the state-model reaching a minimal ICL is generally considered the one that describes the data best, with minimal free parameters. If the ICL is minimized for the 4 state model, a 5 state model is also tested to confirm minimization of the ICL. The ICL parameter is defined in Eq. 1:

$$ICL(m) = -2\ln(\mathbf{p}(\mathbf{y}, \hat{\mathbf{s}}|m, \hat{\lambda_m})) + K\ln(n) \tag{1}$$

where $\ln(\mathbf{p}(\mathbf{y}, \hat{\mathbf{s}}|m, \hat{\lambda_m}))$ is the posterior probability of the most likely state path, as determined by the *Viterbi* algorithm, K is the number of free parameters in the model, and n is the number of photons in all bursts in the data set. K is calculated as in Eq. 2:

$$K = q^2 + (r-1)q - 1 \tag{2}$$

where q is the number of states the state-model represents, and r is the number of photon streams used for the calculation of all of the parameters that are assessed. For spH^2MM, $r = 2$, while for 2c-nsALEX mpH^2MM, $r = 3$. The ICL is preferable as an extremum-based criterion over the previously proposed threshold based on the modified Bayes Information Criterion (BIC')[7]. See supplementary dataset[43] for Jupyter notebooks testing the reliability of ICL with simulated data sets generated using PyBroMo[51] (https://github.com/OpenSMFS/PyBroMo/releases/tag/0.8.1; was utilized in previous works[7, 48, 52]). We optimize the H^2MM model with state-models having successively greater numbers of states, until the ICL ceases

to improve, and always optimizing two-, three-, and four-state models. We use the *Viterbi* algorithm to find the most likely state path based on the posterior probability. From the state path, photons are separated into dwells, each of which can be assigned a duration, a mean PR, and for mpH^{2}MM, a mean S_{PR}. Dwells with fewer than 5 photons are discarded. This also allows bursts to by classified by which and how many states are present. We use the weighted standard error of the PR and S_{PR} as a proxy for the standard error of the H^{2}MM model (see SI for full derivation). The fluorescent decays of sub-populations are also analyzed based on sorting photons by the *Viterbi* algorithm, see SI for further details, and a proposed method for mapping the non-binomially distributed lifetime to a binomially-distributed parameter ammenable to mpH^{2}MM.

5 Author Contributions

E.L. performed 2c-ALEX measurements. The measurements of the RNA polymerase-promoter open complex were performed by E.L. in the laboratory of S.W. P.D.H. analyzed data and contributed analytical tools. P.D.H. & E.L. designed and performed the research and wrote the manuscript.

6 Acknowledgements

This paper was supported by the National Institutes of Health (NIH, grant R01 GM130942 to S.W. and E.L. as a subaward), the National Science Foundation (NSF, grants 1818147 and 1842951 to S.W.), the Human Frontiers Science Program (HFSP, grant RGP0061/2019 to S.W.), the Israel Science Foundation (ISF, grant 3565/20 to E.L., within the *KillCorona* – Curbing Coronavirus Research Program), the Milner Fund (to E.L.), and the Hebrew University of Jerusalem (start-up funds to E.L.). We would also like to acknowledge the great help of Bill Harris in making the H2MM_C package cross platform.

References

[1] Eitan Lerner et al. "Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer". In: *Science* 359.6373 (Jan. 2018), eaan1133. ISSN: 0036-8075. DOI: 10.1126/science.aan1133. URL: https://www.sciencemag.org/lookup/doi/10.1126/science.aan1133.

[2] Eitan Lerner et al. "FRET-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices". In: *eLife* 10 (2021), e60416. DOI: 10.7554/eLife.60416.

[3] Benjamin Schuler and Hagen Hofmann. "Single-molecule spectroscopy of protein folding dynamics-expanding scope and timescales". In: *Current Opinion in Structural Biology* 23.1 (2013), pp. 36–47. ISSN: 0959440X. DOI: 10.1016/j.sbi.2012.10.008. URL: http://dx.doi.org/10.1016/j.sbi.2012.10.008.

[4] Haim Yuval Aviram et al. "Direct observation of ultrafast large-scale dynamics of an enzyme under turnover conditions". In: *Proceedings of the National Academy of Sciences* 115.13 (Mar. 2018), pp. 3243–3248. ISSN: 0027-8424. DOI: 10.1073/pnas.1720448115. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1720448115.

[5] Hisham Mazal and Gilad Haran. "Single-molecule FRET methods to study the dynamics of proteins at work". In: *Current Opinion in Biomedical Engineering* 12 (2019), pp. 8–17. DOI: 10.1016/j.cobme.2019.08.007.

[6] Nicole C Robb et al. "The Transcription Bubble of the RNA Polymerase – Promoter Open Complex Exhibits Conformational Heterogeneity and Millisecond-Scale Dynamics : Implications for Transcription Start-Site Selection". In: *Journal of Molecular Biology* 425.5 (2013), pp. 875–885. ISSN: 0022-2836. DOI: 10.1016/j.jmb.2012.12.015. URL: http://dx.doi.org/10.1016/j.jmb.2012.12.015.

[7] Eitan Lerner, Antonino Ingargiola, and Shimon Weiss. "Characterizing highly dynamic conformational states: The transcription bubble in RNAP-promoter open complex as an example". In: *The Journal of Chemical Physics* 148.12 (Mar. 2018), p. 123315. ISSN: 0021-9606. DOI: 10.1063/1.5004606. URL: http://dx.doi.org/10.1063/1.5004606%20http://aip.scitation.org/doi/10.1063/1.5004606.

[8] Michele Cristóvão et al. "Single-molecule multiparameter fluorescence spectroscopy reveals directional MutS binding to mismatched bases in DNA". In: *Nucleic Acids Research* 40 (12 2012), pp. 5448–5464. DOI: 10.1093/nar/gks138.

[9] Tomas Fessl et al. "Dynamic action of the Sec machinery during initiation, protein translocation and termination". In: *eLIFE* 7 (2018), e35112. DOI: 10.7554/eLife.35112.

[10] Antonio N. Calabrese et al. "Inter-domain dynamics in the chaperone SurA and multi-site binding to its outer membrane protein clients". In: *Nature Communications* 11 (2020), p. 2155. DOI: 10.1038/s41467-020-15702-1.

[11] Hisham Mazal et al. "Tunable microsecond dynamics of an allosteric switch regulate the activity of a AAA+ disaggregation machine". In: *Nature Communications* 10 (2019), p. 1438. DOI: 10.1038/s41467-019-09474-6.

[12] Marijn de Boer et al. "Conformational and dynamic plasticity in substrate-binding proteins underlies selective transport in ABC importers". In: *eLIFE* 8 (2019), e44652. DOI: 10.7554/eLife.44652.

[13] Nicholas J. Anthis and G. Marius Clore. "Visualizing transient dark states by NMR spectroscopy". In: *Quarterly Reviews of Biophysics* 48.1 (Feb. 2015), pp. 35–116. ISSN: 0033-5835. DOI: 10.1017/S0033583514000122. URL: https://www.cambridge.org/core/product/identifier/S0033583514000122/type/journal_article.

[14] G. Marius Clore and Junji Iwahara. "Theory, Practice, and Applications of Paramagnetic Relaxation Enhancement for the Characterization of Transient Low-Population States of Biological Macromolecules and Their Complexes". In: *Chemical Reviews* 109.9 (Sept. 2009), pp. 4108–4139. ISSN: 0009-2665. DOI: 10.1021/cr900033p. URL: https://pubs.acs.org/doi/10.1021/cr900033p.

[15] Arthur G. Palmer. "NMR Characterization of the Dynamics of Biomacromolecules". In: *Chemical Reviews* 104.8 (Aug. 2004), pp. 3623–3640. ISSN: 0009-2665. DOI: 10.1021/cr030413t. URL: https://pubs.acs.org/doi/10.1021/cr030413t.

[16] Enrico Ravera et al. "Insights into Domain–Domain Motions in Proteins and RNA from Solution NMR". In: *Accounts of Chemical Research* 47.10 (Oct. 2014), pp. 3118–3126. ISSN: 0001-4842. DOI: 10.1021/ar5002318. URL: https://pubs.acs.org/doi/10.1021/ar5002318.

[17] Qian Peter Su and Lining Arnold Ju. "Biophysical nanotools for single-molecule dynamics". In: *Biophysical Reviews* 10.5 (Oct. 2018), pp. 1349–1357. ISSN: 1867-2450. DOI: 10.1007/s12551-018-0447-y. URL: http://link.springer.com/10.1007/s12551-018-0447-y.

[18] Krutika Bavishi and Nikos Hatzakis. "Shedding Light on Protein Folding, Structural and Functional Dynamics by Single Molecule Studies". In: *Molecules* 19.12 (Nov. 2014), pp. 19407–19434. ISSN: 1420-3049. DOI: 10.3390/molecules191219407. URL: http://www.mdpi.com/1420-3049/19/12/19407.

[19] Exequiel Medina, Danielle R. Latham, and Hugo Sanabria. "Unraveling protein's structural dynamics: from configurational dynamics to ensemble switching guides functional mesoscale assemblies". In: *Current Opinion in Structural Biology* 66 (2021), pp. 129–138. ISSN: 1879033X. DOI: 10.1016/j.sbi.2020.10.016. URL: https://doi.org/10.1016/j.sbi.2020.10.016.

[20] Soumit S. Mandal. "Force Spectroscopy on Single Molecules of Life". In: *ACS Omega* 5.20 (May 2020), pp. 11271–11278. ISSN: 2470-1343. DOI: 10.1021/acsomega.0c00814. URL: https://pubs.acs.org/doi/10.1021/acsomega.0c00814.

[21] Mykola Dimura et al. "Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems". In: *Current Opinion in Structural Biology* 40 (Oct. 2016), pp. 163–185. ISSN: 0959440X. DOI: 10.1016/j.sbi.2016.11.012. URL: http://dx.doi.org/10.1016/j.sbi.2016.11.012%20https://linkinghub.elsevier.com/retrieve/pii/S0959440X1630197X.

[22] Taekjip Ha et al. "Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor". In: *Proceedings of the National Academy of Sciences of the United States of America* 93.13 (1996), pp. 6264–6268. ISSN: 0027-8424. DOI: 10.1073/pnas.93.13.6264.

[23] Th. Förster. "Zwischenmolekulare Energiewanderung und Fluoreszenz". In: *Annalen der Physik* 437.1-2 (1948), pp. 55–75. ISSN: 15213889. DOI: 10.1002/andp.19484370105.

[24] Th. Förster. "10th Spiers Memorial Lecture. Transfer mechanisms of electronic excitation". In: *Discussions of the Faraday Society* 27.10 (1959), p. 7. ISSN: 0366-9033. DOI: 10.1039/df9592700007. URL: http://xlink.rsc.org/?DOI=df9592700007.

[25] L Stryer and R P Haugland. "Energy transfer: a spectroscopic ruler." In: *Proceedings of the National Academy of Sciences of the United States of America* 58.2 (1967), pp. 719–726. ISSN: 0066-4154. eprint: arXiv:1011.1669v3.

[26] Nam Ki Lee et al. "Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation". In: *Biophysical Journal* 88.4 (2005), pp. 2939–2953. ISSN: 00063495. DOI: 10.1529/biophysj.104.054114. URL: http://dx.doi.org/10.1529/biophysj.104.054114.

[27] Menahem Pirchi et al. "Photon-by-Photon Hidden Markov Model Analysis for Microsecond Single-Molecule FRET Kinetics". In: *Journal of Physical Chemistry B* 120.51 (2016), pp. 13065–13075. ISSN: 15205207. DOI: 10.1021/acs.jpcb.6b10726.

[28] Barbara K. Müller et al. "Pulsed interleaved excitation". In: *Biophysical Journal* 89.5 (2005), pp. 3508–3522. ISSN: 00063495. DOI: 10.1529/biophysj.105.064766.

[29] Ted A. Laurence et al. "Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins". In: *Proceedings of the National Academy of Sciences* 102.48 (Nov. 2005), pp. 17348–17353. DOI: 10.1073/pnas.0508584102. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.0508584102.

[30] Sungchul Hohng, Chirlmin Joo, and Taekjip Ha. "Single-Molecule Three-Color FRET". In: *Biophysical Journal* 87.2 (Aug. 2004), pp. 1328–1337. ISSN: 00063495. DOI: 10.1529/biophysj.104.043935. URL: https://linkinghub.elsevier.com/retrieve/pii/S0006349504736127.

[31] Jean-Pierre Clamme and Ashok A. Deniz. "Three-Color Single-Molecule Fluorescence Resonance Energy Transfer". In: *ChemPhysChem* 6 (1 2005), pp. 74–77. DOI: 10.1002/cphc.200400261.

[32] Nam Ki Lee, Hye Ran Koh, and Seong Keun Kim. "Folding of 8-17 Deoxyribozyme Studied by Three-Color Alternating-Laser Excitation of Single Molecules". In: *Journal of the American Chemical Society* 129 (50 2007), pp. 15526–15534. DOI: 10.1021/ja0725145.

[33] Nam Ki Lee et al. "Three-Color Alternating-Laser Excitation of Single Molecules: Monitoring Multiple Interactions and Distances". In: *Biophysical Journal* 92 (1 2007), pp. 303–312. DOI: 10.1529/biophysj.106.093211.

[34] Sanghwa Lee, Jinwoo Lee, and Sungchul Hohng. "Single-Molecule Three-Color FRET with Both Negligible Spectral Overlap and Long Observation Time". In: *PLoS One* 5 (2010), e12270. DOI: 10.1371/journal.pone.0012270.

[35] Ingo H. Stein, Christian Steinhauser, and Philip Tinnefeld. "Single-Molecule Four-Color FRET Visualizes Energy-Transfer Paths on DNA Origami". In: *Journal of the American Chemical Society* 133 (12 2011), pp. 719–726.

[36] Seok W. Yim et al. "Four-Color Alternating-Laser Excitation Single-Molecule Fluorescence Spectroscopy for Next-Generation Biodetection Assays". In: *Clinical Chemistry* 58 (4 2012), pp. 707–716. DOI: 10.1373/clinchem.2011.176958.

[37] C. Ratzke, B. Hellenkamp, and Thorsten Hugel. "Four-colour FRET reveals directionality in the Hsp90 multicomponent machinery". In: *Nature Communications* 5 (2014), p. 4192. DOI: 10.1038/ncomms5192.

[38] Roman Tsukanov et al. "Conformational Dynamics of DNA Hairpins at Millisecond Resolution Obtained from Analysis of Single-Molecule FRET Histograms". In: *The Journal of Physical Chemistry B* 117.50 (Dec. 2013), pp. 16105–16109. ISSN: 1520-6106. DOI: 10.1021/jp411280n. URL: https://pubs.acs.org/doi/10.1021/jp411280n.

[39] Joseph P. Torella et al. "Identifying molecular dynamics in single-molecule fret experiments with burst variance analysis". In: *Biophysical Journal* 100.6 (2011), pp. 1568–1577. ISSN: 15420086. DOI: 10.1016/j.bpj.2011.01.066. URL: http://dx.doi.org/10.1016/j.bpj.2011.01.066.

[40] Paul D. Harris, Samir M. Hamdan, and Satoshi Habuchi. "Relative Contributions of Base Stacking and Electrostatic Repulsion on DNA Nicks and Gaps". In: *The Journal of Physical Chemistry B* 124 (47 2020), pp. 10663–10672. DOI: 10.1021/acs.jpcb.0c06941.

[41] C. Biernacki, G. Celeux, and G. Govaert. "Assessing a mixture model for clustering with the integrated completed likelihood". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.7 (July 2000), pp. 719–725. ISSN: 01628828. DOI: 10.1109/34.865189. URL: http://journals.sagepub.com/doi/10.1177/003591577406700922%20http://ieeexplore.ieee.org/document/865189/.

[42] Gilles Celeux and Jean-Baptiste Durand. "Selecting hidden Markov model state number with cross-validated likelihood". In: *Computational Statistics* 23.4 (Oct. 2008), pp. 541–564. ISSN: 0943-4062. DOI: 10.1007/s00180-007-0097-1. URL: http://link.springer.com/10.1007/s00180-007-0097-1.

[43] Paul David Harris, Shimon Weiss, and Eitan Lerner. *Multi-parameter photon-byphoton hidden Markov modeling dataset*. Apr. 2021. DOI: 10.5281/ZENODO.4671393. URL: https://zenodo.org/record/4671393.

[44] Xiangxu Kong et al. "Photobleaching pathways in single-molecule FRET experiments". In: *Journal of the American Chemical Society* 129.7 (2007), pp. 4643–4654. ISSN: 00027863. DOI: 10.1021/ja068002s.

[45] Jayanta Mukhopadhyay et al. "Translocation of σ^{70} with RNA Polymerase during Transcription: Fluorescence Resonance Energy Transfer Assay for Movement Relative to DNA". In: *Cell* 106 (4 2001), pp. 453–463. DOI: 10.1016/S0092-8674(01)00464-0.

[46] Soohong Kim et al. "High-throughput single-molecule optofluidic analysis". In: *Nature Methods* 8 (3 2011), pp. 242–245. DOI: 10.1038/nmeth.1569.

[47] Paul David Harris. *H2MMpythonlib: H2MM_C*. Version 2.0.1. Apr. 8, 2021. DOI: 10.5281/zenodo.4671440. URL: https://zenodo.org/record/4671440.

[48] Antonino Ingargiola et al. "Photon-HDF5". In: *Biophysical Journal* 110 (2016), pp. 25–33. ISSN: 00063495.

[49] Antonino Ingargiola, Shimon Weiss, and Eitan Lerner. "Monte Carlo Diffusion-Enhanced Photon Inference: Distance Distributions and Conformational Dynamics in Single-Molecule FRET". In: *The Journal of Physical Chemistry B* 122 (49 2018), pp. 11598–11615. DOI: 10.1021/acs.jpcb.8b07608.

[50] Antonino Ingargiola et al. "FRETBursts: An open source toolkit for analysis of freely-diffusing Single-molecule FRET". In: *PLoS ONE* 11.8 (2016), pp. 1–27. ISSN: 19326203. DOI: 10.1371/journal.pone.0160716.

[51] Antonino Ingargiola. "FOpenSMFS/PyBroMo: Version 0.8.1". In: *zenodo.org* (2019). DOI: 10.5281/zenodo.3256149.

[52] Dolev Hagai and Eitan Lerner. "Systematic assessment of burst impurity in confocal-based single-molecule fluorescence detection using Brownian motion simulations". In: *Molecules* 24.14 (2019). ISSN: 14203049. DOI: 10.3390/molecules24142557.