



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

Doctoral Thesis

Quantifying Nonequilibrium Thermodynamics of Proteins Using Single Molecule Spectroscopy

Saurabh Talele

Department of Biomedical Engineering

Ulsan National Institute of Science and Technology

2023

Quantifying Nonequilibrium Thermodynamics of Proteins Using Single Molecule Spectroscopy

Saurabh Talele

Department of Biomedical Engineering

Ulsan National Institute of Science and Technology

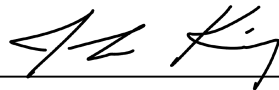
Quantifying Nonequilibrium Thermodynamics of Proteins Using Single Molecule Spectroscopy

A thesis/dissertation submitted to
Ulsan National Institute of Science and Technology
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Saurabh Talele

05.23.2023 of submission

Approved by



Advisor

John King

Quantifying Nonequilibrium Thermodynamics of Proteins Using Single Molecule Spectroscopy

Saurabh Talele

This certifies that the thesis/dissertation of Saurabh Talele is approved.

05.23.2023 of submission

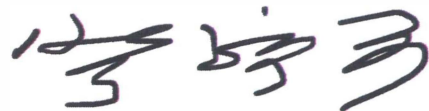
Signature



Advisor: John King



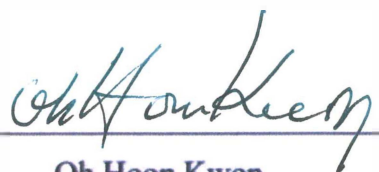
Francois Amblard



Hyuk Kyu Pak



Yoon Kyoung Cho



Oh Hoon Kwon

Abstract

Proteins are dynamic molecules, and their structural fluctuations play a crucial role in their function. Quantification of these motions is hence central to our understanding of a protein's function. The timescales of these fluctuations span over 12 orders of magnitude from ultrafast picoseconds vibrational motions to slow seconds long global changes. While the ultrafast and slow motions have been well characterized, the intermediate conformational changes occurring at microsecond to millisecond timescales remain to be clearly understood. This thesis explores the use of single molecule multidimensional fluorescence spectroscopy to investigate protein conformational changes occurring at microsecond to millisecond timescales. The information of conformational transition rates gives insight into their operating principles and the mechanisms behind their biological functions.

Single-molecule Two-Dimensional Fluorescence Lifetime Correlation Spectroscopy (sm-2DFLCS) is used to resolve distinct conformations of protein molecules based on different fluorescence lifetimes. Time resolved fluorescence photon traces are converted to two-dimensional emission delay correlation maps which are then converted to lifetime correlation spectra using a nonlinear optimization algorithm. Key advancements in the technique were achieved by adapting analytical strategies from 2D NMR analysis protocols which convert the underlying optimization problem from constrained to unconstrained optimization resulting in a fast and robust 2D Inverse Laplace Transformation algorithm (2D-ILT). This enables monitoring of forward and reverse conformational transitions in a protein independently.

Nonequilibrium thermodynamics of pump protein *Bacteriorhodopsin* is characterized using the improvised 2D-ILT algorithm. Independent measurement of forward and reverse conformational transitions reveals microscopically irreversible transitions in the reaction cycle of the proton pump. Nonequilibrium thermodynamic properties such as entropy production, flux and affinity are quantified through portions of the reaction cycle. It is shown that the rate of irreversible transitions shows an inverse dependence on temperature and fitting the trend with Gibbs-Helmholtz relation yields experimentally determined enthalpy of transition.

Coupling of proteins internal reaction coordinate and surrounding solvent is investigated by monitoring microsecond equilibrium fluctuations in the chromophore pocket of eGFP. It is observed that the dynamics of local rearrangements around the chromophore are coupled to the bulk viscosity of the solvent. The dependence is observed to deviate from Kramer's scaling and the deviation is attributed to proteins' internal friction.

Table of Contents

Acknowledgements	(7)
Abstract	(8)
List of Figures	(12)
List of Tables	(18)
List of Abbreviations	(19)
Chapter 1. Introduction	(21)
1.1 Motivation	(21)
1.2 Chemical exchange and irreversibility in proteins	(22)
1.3 Fluorescence	(23)
1.4 Thesis plan	(26)
Chapter 2. Technical Background	(31)
2.1. Fluorescence microscopy	(31)
2.2. Single molecule detection	(34)
2.3. Time Correlated Single Photon Counting (TCSPC)	(36)
2.3.1 Advanced TCSPC	(39)
2.4. Analysis of fluorescence decays	(40)
2.4.1 Maximum Entropy Method	(44)
Chapter 3. Fast and robust two-dimensional inverse Laplace transformation of single-molecule fluorescence lifetime data	(48)
3.1. Introduction	(48)
3.2. Methods	(51)
3.2.1. Generation of 2D emission delay histograms	(53)
3.2.2. Tow-Dimensional Inverse Laplace Transform (2D-ILT)	(54)
3.2.3. 2D-ILT by Tikhonov Regularization	(56)
3.2.3.1. Regularization function	(56)
3.2.3.2. Kernel compression	(57)
3.2.3.3. Tikhonov regularization with BRD	(63)
3.2.4. Monte carlo simulations	(66)

3.3. Results and Discussions	(68)
3.3.1. Implementation	(68)
3.3.2. Effect of compression	(72)
3.3.3. Dependence on total photon counts	(73)
3.3.4. Application to experimental data	(77)
3.3.5. Comparison to existing methods	(78)
3.4. Conclusion	(80)
Chapter 4. Reaction cycle of operating pump protein studied with single molecule spectroscopy ...	(84)
4.1. Introduction	(84)
4.2. Materials and methods	(85)
4.2.1. Sample preparation	(86)
4.2.2. Sm-2D-FLCS	(88)
4.3. Results	(91)
4.4. Discussion	(95)
4.4.1 Exchange kinetics of D96N mutant	(98)
4.4.2. Influence of continuous laser illumination on the photocycle	(99)
4.4.3. Lapping effect in the photocycle	(100)
4.5. Conclusions	(100)
Chapter 5. Protein-Bath coupling at intermediate timescales	(105)
5.1. Introduction	(105)
5.2. Methods	(106)
5.2.1. Sample preparation	(106)
5.2.2. Single molecule spectroscopy	(107)
5.3. Results	(107)
5.4. Discussion	(110)
5.5 Conclusion	(112)
Chapter 6. Conclusions	(116)
6.1. General conclusions	(116)
6.2. Future outlook	(117)

List of Figures

Figure 1.1: The Jablonski diagram. Horizontal grey lines represent the electronic states where S_0 is the ground state, S_1 and S_2 are electronic excited states. Arrows represent energy transfer. Photon absorption is marked blue, emission of fluorescent photon is marked green, and phosphorescence is marked red. The non-radiative decay processes are denoted with dashed arrows.(17)

Figure 1.2. Schematic comparison of steady-state and time-resolved fluorescence measurements. A and C are the Temporal profiles of excitation intensities for steady-state and time-resolved measurements respectively. B shows the Emission spectra obtained using a continuous illumination. D shows a histogram of the time-resolved fluorescence delays measured using pulsed excitation shown in C.....(19)

Figure 2.1 Schematic of a widefield microscope. The light beam coming from the laser is focused on the back focal plane of the objective lens which causes a parallel beam to be incident on the sample in focus. The fluorescence from the illuminated area is collected by the objective and passes through the dichroic followed by an emission filter after which the image is formed on the CCD camera.....(24)

Figure 2.2. Schematic of a confocal microscope. The light coming from the laser source is focused inside the sample using an objective lens. The fluorescence coming from the diffraction limited spot is collected by the objective and passes through the dichroic mirror onto a confocal pinhole which filters out the fluorescence coming from out of focus planes. This light is then passed through an emission filter and is detected using a point detector such as APD or PMT.....(25)

Figure 2.3 Fluorescence decay of Alexa 488 in water. The trace in blue shows the fluorescence emission delay histogram of Alexa 488 solution. The experimentally observed histogram is the result of the true decay convoluted by the Instrument Response Function (IRF) shown in orange. The decay can be fit to obtain the fluorescence lifetime of Alexa 488 which in this case is 4.8 ns..... (29)

Figure 3.4 Components of a TCSPC setup. The fluorescence photons from the sample arrive on the detector. The signal is relayed to the Constant Fraction Discriminator (CFD) which determines the timing of the pulse. The sync pulse from the laser driver is also connected to another CFD to determine its timing. Outputs of both CFDs are connected to the Time to Amplitude Converter (TAC) which, then is connected to the Analog to Digital Converter (ADC) and this timing information of the incident photon is stored in the memory (MEM) and is further sent to the computer.(31)

Figure 2.5. Working of a Constant Fraction Discriminator (CFD). The input pulse is split into two components and one of them is inverted. A delay circuit produces a time lag in one of the pulse. Both pulses are

then added together, and the zero-crossing point of the resultant pulse is determined which represents the peak time of the input pulse.(31)

Figure 2.6 Working of a Time to Amplitude Converter (TAC). This is a linear ramp generator circuit which has a start trigger and a stop trigger. The resultant analog voltage is proportional to the time delay between both triggers. This voltage is held till the ADC processes it and is then reset to the default value to be ready for accepting the next start pulse.(33)

Figure 2.7 Inverse Laplace Transform. A shows a simulated bi-exponential fluoresce decay and B shows its corresponding Inverse Laplace Transform revealing the component lifetimes at 0.6 ns and 2.5 ns.(35)

Figure 2.8. ILT on experimental data using Maximum Entropy Method (MEM). Panel A shows the experimentally obtained fluorescence decay from eGFP protein solution (black). The observed decay is the result of true exponential decay convoluted by the IRF (shown in red). Panel B shows the resultant ILT obtained using Maximum entropy method which shows the lifetime components at 1.5 ns and 2.7 ns.(37)

Figure 3.1. Schematic of 2D-FLCS analysis. a) excitation of a model 2-state system with a pulsed laser source results in a bi-exponential fluorescence decay. b) typical TCSPC measurement shows individual photon detection events following pulsed laser excitation. The emission delays between the excitation and emission are denoted by t . c) A 2D emission delay correlation map is constructed by measuring the delays between photon pairs separated by a preset time lag ΔT . d) The 2D Inverse Laplace Transform of the emission delay correlation map at given ΔT gives a 2D fluorescence lifetime correlation spectra where, individual species are represented by peaks along the diagonal and their chemical exchange is represented on the off-diagonal positions.(45)

Figure 3.2 (a) An example fluorescence decay data exhibiting a biexponential decay. (b) Singular values of the kernel matrix K in Eq.3.9 comprising of exponential decays with 100 distinct lifetimes in range (0.1 ns–10 ns). The singular values decrease rapidly, and K can be represented accurately by 30 highest singular values. (c) Compressed data $\sim m$ obtained by Eq.3.19 represented using 30 point.(50)

Figure 3.3. Kernel compression for data analysis. (a) The distribution of singular values of the kernels with size 100 and various ranges. (b) The distribution of singular values of the kernels constructed using basis in the range [0.1, 10ns] using various sizes. (c) Least square error in the compression where the kernels are constructed using the number of singular values on the x-axis. (d) The difference between the primary kernel and the compressed kernel constructed using 30 highest singular values..... (52)

Figure 3.4. Monte carlo simulation for data with IRF included. (a) The 1D emission delay histogram. (b) Emission delay histogram (black) convoluted by the Instrument response function (red). (c) The 1D ILT spectra at various zero-time shifts of the IRF. (d) Evaluated least-square error in the fit and convoluted histogram. The

ILT spectra at the minima of the plot in (d) is selected as the 1D lifetime spectra.(55)

Figure 3.5 Noise estimation for stopping criterion. (a) Normalized fluorescence data superimposed with \tilde{m} , projection of compressed data onto the range space of kernel given by Eq.3.23. It is observed that \tilde{m} serves as a smooth approximation to input data referring to which we infer the corresponding noise level given by Eq.3.35. (b) Shows the difference $\tilde{m} - m$, which is dependent on the photon counts. We normalize both traces using square root of photon counts which eliminates the intensity dependence. (c) Shows the difference after normalization given by $\sqrt{\tilde{m}} - \sqrt{m}$(57)

Figure 3.6. Laplace inversions by Tikhonov Regularization. Monte Carlo simulations generate an artificial photon stream of a two-component system with fluorescence lifetimes of $\tau_1 = 1\text{ns}$ and $\tau_2 = 3\text{ns}$ undergoing equilibrium chemical exchange at a rate of $1 \times 10^3\text{s}^{-1}$. (a) Raw 1D-photon histogram (blue) and the obtained fit (red) from 1D-ILT. (b) The 1D lifetime spectrum shows two peaks with relaxation rates of 1 and 3 ns, consistent with fluorescence lifetime of the two components in the system. (c) Residual between the 1D photon histogram and the fit obtained by 1D-ILT. (d) 2D photon correlation histogram and (e) 2D relaxation rate spectrum computed at a waiting time of $\Delta T = 2\text{ms}$ obtained from a 2D-ILT. The spectrum consists of diagonal peaks at 1 and 3 ns, as well as cross-peaks indicating chemical exchange between the two species. (f) Residual between the 2D photon correlation histogram and 2D-ILT fit.(64)

Figure 3.7 Chemical Exchange kinetics of a two-state system. 2D-FLCS spectra computed at (a) $\Delta T = 10 - 20\mu\text{s}$, (b) $\Delta T = 100 - 200\mu\text{s}$, and (c) $\Delta T = 1 - 2\text{ms}$. At $\Delta T < \tau_{\text{exchange}}$, no cross-peaks are observed between the two states, indicating no exchange has occurred. As ΔT approaches τ_{exchange} , cross-peaks emerge indicating both forward and reverse transitions between the two states. (d) The kinetics of the exchange process are reflected in the autocorrelation and cross-correlation functions of the diagonal and off-diagonal peaks, respectively. Fitting either the autocorrelation or cross-correlation functions provide a direct measure of the exchange kinetics in the system.(65)

Figure 3.8 Dependence of 2D spectrum on total photon counts. (a) The estimated $\hat{\sigma}$ with respect to total photon counts in the fluorescence data. (b) Shows the optimum regularization parameter obtained with respect to the total photon counts in the fluorescence data. Error bars denote the standard deviation using 30 instances of simulated data. Higher α_{opt} due to lower photon counts results in smoother fits and vice versa as demonstrated in (c), which showing 1D-ILT of three trajectories with varying total photons. (d-f) show the normalized 2D-ILT obtained at $\Delta T = 2\text{ms}$ using the simulated trajectories from (c).(68)

Figure 3.9 (a) shows the dependence of peak-widths of obtained 1D ILTs with respect to the total photon counts in the units of basis points. (b) shows the dependence of peak widths with respect to the size of the basis of lifetimes chosen for the 1D ILT.(69)

Figure 3.10. Baseline correction for background noise. (a) shows the effect of added background noise on

the ILT as an observed “tail” shown in blue. On implementing baseline correction on the same data, this artefact is eliminated as shown in red. (b) at lower signal to background ratios, minor but systematic shifts are observed in the peak positions of obtained 1D-ILT which are also eliminated by applying baseline correction for the background noise. (c) The “tail” artefact is observed to have systematic dependence on the SBR as shown in green and is reduced on applying the baseline correction..... (71)

Figure 3.11. 2D-ILT on Bacteriorhodopsin single-molecule fluorescence data. (a) Crystal structure of bR, with the retinal chromophore highlighted in green. (b) Absorption spectrum of the bR in the ground state, with $\lambda_{max}=560\text{ nm}$. (c and d) 2D-FLCS spectra computed for $\Delta T = 10 - 50\mu\text{s}$ and $\Delta T = 10 - 50\text{ms}$. The three diagonal peaks can be assigned to the K, L, and N intermediate formed during the photoinduced reaction cycle of bR. Cross-peaks evolve because of chemical exchange between the intermediate species.(73)

Figure 3.12. Comparison of TK and MEM. (a–f) Show 1D-ILT of same simulated data set comprising 2×10^6 total photons with lifetimes 1 and 3 ns. (g–l) show the 2D-ILT of the same data set evaluated at $\Delta T = 2\text{ms}$. The computational times for the simulations are given in Tables 2 and 3.(74)

Figure 4.1. Single-molecule spectroscopy of Bacteriorhodopsin. Crystal structure of ground state bR showing a) the protein and b) the retinal chromophore and the amino acids involved in the H^+ transport chain of the reaction cycle (PDB: 1KBG). c) Photo-induced reaction cycle of bR, which involves 6 intermediates in addition to the ground state (gs). The subscripts on the labels denote the absorption maximum of the retinal chromophore for the given conformation. d) Absorption spectrum of retinal embedded in bR shows strong absorption at $\lambda=568\text{ nm}$ at $\text{pH}=6$. e) The fluorescence lifetime histogram measured from a single protein shows multi-component relaxation following 532 nm excitation. A 1D-ILT of the lifetime histogram reveals three relaxation times of $\tau_1' \sim 0.08\text{ ns}$, $\tau_2' \sim 1.0\text{ ns}$, and $\tau_3' \sim 5.0\text{ ns}$ (inset). f) sm-2D-FLCS spectrum measured at a waiting time of $\Delta t=10\ \mu\text{s}$(83)

Figure 4.2. Single-molecule imaging and photo-physics. (a) Confocal image of single Bacteriorhodopsin proteins immobilized on a glass surface. (b) Diffraction-limited cross-section of fluorescence signal measured in confocal image. (c) Single-molecule fluorescence trace showing single-step photo-physics. (d) Fluorescence lifetime histogram from single wt-bR protein showing multicomponent relaxation.....(85)

Figure 4.3. Experimental procedure for measuring sm-2D-FLCS data. (a) Excitation pulse train and photon stream measured with resolved photon emission time t_{em} (standard TCSPC measurement in Time-Tagged-Time-Resolved mode). (b) 1D histogram of photon emission time and the fluorescence lifetime spectrum computed by a 1D ILT for a simple two state system. (c) 2D histogram of photon emission measured at a set waiting time Δt and 2D-FLCS spectra computed by a 2D ILT. At early waiting times ($\Delta t < \text{exchange time}$) peaks are only measured along the diagonal, indicating species that remained in their initial state during the waiting time. At later times ($\Delta t > \text{exchange time}$), cross-peaks arise as species that were in an initial state exchange to the other state during the waiting time..... (86)

Figure 4.4. Temperature dependence of bR state occupation. (a) 1D fluorescence emission time spectrum and (b) relative state occupancies of bR as a function of temperature..... (88)

Figure 4.5. Microscopic irreversibility of bR catalytic cycle. a) sm-2D-FLCS spectra of bR shown for waiting times ranging from 0.01 to 50 ms. Exchange kinetics at 291 K for waiting times $\Delta t=0.01-200$ ms between b) K intermediate to L intermediate and c) L intermediate to N intermediate. Forward transitions, represented by solid symbols and solid lines, are measured from the upper quadrant of the 2D spectra. Reverse transitions, represented by open symbols and dashed lines, are measured from the lower quadrant of the 2D spectra. The forward (kij) and reverse (kji) transition rates are measured by fitting the cross-correlation functions to exponential growth functions (solid and dashed lines). The experiments were performed at $T=291$ K.(89)

Figure 4.6. Thermal stability of the bR reaction cycle. Arrhenius plots for the a) K-L and b) L-N transition rates. The k_{ij} and k_{ji} values are shown in green and blue, respectively. c) The affinity A of the K-L transition plotted vs T^{-1} . As the transition is microscopically reversible, the A value is ~ 0 kJmol $^{-1}$ at all temperatures measured. d) The A of the L-N transition plotted vs T^{-1} . The microscopically irreversible transition has an A value of ~ 10 kJmol $^{-1}$ at 291 K that decreases linearly towards 0 kJmol $^{-1}$ with increasing temperature. The temperature dependencies of A are fit to the Gibbs-Helmholtz equation [red line, Equation (6)] (92)

Figure 4.7. Entropy production rate and variance of flux. a) Entropy production for the K-L and L-N transitions computed from Equation (7). The entropy production rate σ associated with the reversible K-L transition is 0 kJmol $^{-1}$ K $^{-1}$ s $^{-1}$ at all temperatures measured, while the irreversible L-N transition has significant entropy production rate (~ 50 kJmol $^{-1}$ K $^{-1}$ s $^{-1}$) that decreases exponentially with increasing temperature.(94)

Figure 4.8. Exchange dynamics of D96N mutant. (a) Chemical structure of reprotonation site at Asp96 residue and the structure of the D96N mutant. (b) Chemical exchange from L to N (solid symbols) and N to L states (open symbols), showing near reversible exchange in the mutant protein.(95)

Figure 5.1. Structure and internal reaction of eGFP. (a) Crystal structure of eGFP (top) with the chromophore pocket highlighted (bottom). (b) Internal reaction between two bright conformations of the chromophore pocket. Dihedral rotation of the Glu222 residue drives a rearrangement of the local hydrogen bonding network around the chromophore..... (102)

Figure 5.2. eGFP single-molecule fluorescence lifetime data. (a) Fluorescence emission delay histogram for eGFP (red data), shown with the instrument response function (grey data). (b) Corresponding 1D-ILT of the relaxation data, showing two peaks at $t_1 \sim 1.8$ and $t_2 \sim 3.2$ ns, corresponding to the two conformations of the chromophore pocket.(104)

Figure 5.3. Reaction kinetics of eGFP chromophore pocket. (a) sm-2D-FLCS spectra series ranging from Dt of 0.005 ms to 5 ms. Forward exchange between the two bright states is observed as signal in the upper-left quadrant of the spectrum (highlighted in solid box), and reverse exchange is observed as signal in the lower-right quadrant of the spectrum (highlighted in dashed box). Exchange dynamics extracted from sm-2D-FLCS spectra

cross-peak amplitudes for (b) forward and (c) reverse exchange as a function of solvent viscosity. The h_0 value is given by the viscosity of PBS buffer at 20°C. The exchange times decrease as the solvent viscosity increases, consistent with the internal reaction coordinate being coupled to the fluctuations in the bath.(106)

Figure 5.4. Forward and reverse exchange dynamics of eGFP in buffer. The forward and reverse reaction kinetics are extracted from opposite regions of the 2D spectrum. Under conditions of microscopic reversibility, the signal amplitudes are equivalent and the 2D spectrum is symmetric along the diagonal. As expected for equilibrium dynamics, the forward and reverse reaction kinetics are equivalent for the reaction occurring in the eGFP chromophore pocket(107)

Figure 5.5. Deviation of protein reaction from Kramers' theory. Exchange times for forward (solid symbols) and reverse (open symbols) for the isomerization reaction in chromophore pocket of eGFP as a function of solvent viscosity. The dashed line shows the scaling expected from Kramers' theory. The exchange time deviates significantly at high viscosities, consistent with the picture of internal friction contributing to the reaction dynamics. The red line shows a power-law fit to the data..... (109)

List of Tables

<i>Table 2.1 Common regularizing functions and their functional forms</i>	<i>(40)</i>
<i>Table 3.1 Monte Carlo simulation parameters.....</i>	<i>(64)</i>
<i>Table 3.2 Computation times in seconds for methods for 1D ILT of same set of data comprising 2×10^6 total photons</i>	<i>(76)</i>
<i>Table 3.3 Computation times in seconds for methods for 2D ILT of same set of data with the 2D emission delay histograms evaluated at $\Delta T = 2ms$.....</i>	<i>(76)</i>
<i>Table 3.4 CPU time required for execution of the fits for various sizes of basis set.....</i>	<i>(77)</i>

List of Abbreviations

APD – Avalanche Photo Diode

PMT – Photo Multiplier Tube

CCD – Charged Coupled Device

TCSPC – Time Correlated Single Photon Counting

PSF – Point Spread Function

FWHM – Full Width Half Maximum

sm-FRET- Single Molecule Forster Resonance Energy Transfer

FCS – Fluorescence Correlation Spectroscopy

PALM – Photo Activated Localization Microscopy

STORM – Stochastic Optical Reconstruction Microscopy

R6G – Rhodamine-6G fluorophore

IRF – Instrument Response Function

CFD – Constant Fraction Discriminator

TAC – Time to Amplitude Converter

ADC – Analogue to Digital Converter

TTTR – Time-Tagged Time-Resolved mode in TCSPC setup

ILT – Inverse Laplace Transform

MEM – Maximum Entropy Method

TK – Tikhonov Regularization

SVD – Singular Valued Decomposition

List of Mathematical Constants and Scientific Notations

S_0 – electronic ground state of a molecule

S_1 – first excited electronic state

Γ – rate of radiative relaxation from excited state to ground state

k_{nr} – rate of non-radiative relaxation.

Q – Fluorescence quantum yield

τ – Fluorescence lifetime

ε – Molar extinction coefficient

σ – Absorption cross-section

N_A – Avogadro's Number = $6.02214076 \times 10^{23} \text{ mol}^{-1}$.

\hbar - Plank's Constant = $6.62607015 \times 10^{-34} \text{ J}\cdot\text{Hz}^{-1}$

K_{ij} – Kernel with pre-defined lifetime decays

A_{ij} – Column vector of amplitudes of fluorescence decays

$M(t_{1,i}, t_{2,j}, \Delta T)$ – 2D emission delay histogram

m – 1D vector lexicographically ordered M matrix.

$F(\tau_1, \tau_2, \Delta T)$ - joint probability distribution of occurrence of a photon with lifetime τ_2 occurring ΔT on registering a photon with lifetime τ_1 .

f - 1D vector lexicographically ordered F matrix.

$Q(f)$ – objective function to be minimized

α – Regularization constant

U_i – left singular values of kernel K_i

Σ_i – Eigen values of kernel K_i

V_i – right singular values of kernel K_i

$\hat{\sigma}$ – standard deviation of the noise

Chapter 1. Introduction

1.1. Motivation

Proteins are dynamic and not static systems (1, 2) and the dynamics play a crucial role in the function of proteins (3). Proteins are known as the functional unit of life because they are responsible for performing vast arrays of functions within living organisms. Structurally, a protein molecule is a long sequence of amino acid residues which folds in a well-defined 3-dimensional shape. The function of a protein is believed to be intimately connected to its 3D structure. Thus, the structure-function relationship in proteins has been an important area of study particularly for biophysicists. This has been greatly facilitated by techniques such as X-ray crystallography (4, 5), NMR spectroscopy (6-8) and circular dichroism spectroscopy (9), which help resolve the structure of a protein molecule.

Traditionally, protein function has been examined via chemical kinetic measurements where the chemical rate constants of catalytic enzymes were probed and their dependence with respect to environmental parameters such as temperature, pH and salinity was explored (10, 11). These approaches still did not provide the complete understanding of how the structure played a role in the proteins function. It was shown that large scale conformational motions were responsible for altering the local electronic environment and thus acting as catalysts (1, 12). Recently, the focus has shifted towards studying the structure-dynamics-function relationship in proteins. This exploration into protein dynamics has been greatly facilitated by modern techniques such as high-resolution X-ray crystallography, Cryo-electron microscopy, molecular dynamics simulations, *etc.* Single-molecule fluorescence techniques such as Forster Resonance Energy Transfer (FRET) and fluorescence correlation spectroscopy (FCS), amongst other techniques, have provided direct quantitative measurements of conformational motions in proteins. The work presented in this thesis outlines the development of emerging single-molecule multi-dimensional fluorescence lifetime correlation spectroscopy (sm-2D-FLCS) and its application to the study of proteins functioning both at equilibrium and far-from-equilibrium.

1.2. Chemical exchange and irreversibility in proteins

Protein molecules exist in an interesting range of length scales and energetics. Systems at such colloidal length scale are not large enough to be influenced by inertial forces which govern the motion of macroscopic objects and not small enough to be described efficiently by our knowledge of atomic scales and chemical bonds. This results in a complex system where the interactions are dominated by entropic factors given the large degrees of freedom rather than the enthalpic factors that are relevant at atomic level. Thus, one starts to build a picture of a protein as a highly complex machine closely coupled with its surrounding thermal bath and, hence exhibiting stochastic fluctuations.

Protein dynamics is understood in terms of motion on a rugged free energy landscape with respect to a suitable reaction coordinate. This is a statistical description of a protein's potential surface where each conformational intermediate is represented by a local minimum (valleys) in the free energy surface and the energy required to transition within the intermediates is captured by local maxima (peaks) (2, 12, 13). Despite its abstractness, the energy landscape is a useful representation of a protein because it captures both the thermodynamic as well as the kinetic information of the system. Using this representation, Frauenfelder et.al explained the dynamics and function of myoglobin and predicted the heights of energy barriers between distinct conformational states.

That is however not the complete picture describing protein dynamics because some functions inside the cell require energies over a magnitude higher than what can be harnessed from thermal fluctuations to achieve a net directional transport (e.g., Kinesin, Transmembrane channels) or to perform work against a chemical potential (e.g., F-ATPase). To accomplish work against the randomizing thermal forces, a molecular machine must consume external energy to access modes that are practically improbable to reach using surrounding thermal fluctuations and hence should function out of equilibrium. The dissipation of energy along a series of conformational changes can cause a temporal asymmetry and thus ensure a net directional action of the protein. The dissipative nature of transitions is important in maintaining asymmetry in the forward and reverse transition probabilities (14-17). It is

not yet established how one can incorporate a generalized free-energy landscape for out-of-equilibrium protein systems. However, the first steps towards experimentally probing it would be to design a technique capable of detecting distinct conformations, measuring the rate of conformational changes, and measuring the forward and reverse transitions independently.

To accurately model the protein conformation changes, one also needs to consider how proteins surroundings affect its dynamics. The extent of coupling between a protein's motions determines its transition kinetics. When residues in the protein are exposed to solvent, it is observed that large scale motions such as domain rearrangement are strongly coupled to the surrounding solvent as these require significant displacement of the solvent. On the other hand, fast side chain fluctuations that require insignificant solvent rearrangement are uncoupled from the bath. Little is known about the hydrodynamic coupling of conformational changes that take place in the interior of the protein. Quantifying such coupling with the internal reaction coordinate would require one to measure conformational transition rates with respect to varying physical properties of the solvent.

This thesis focuses on application of one such technique known as single molecule multi-dimensional fluorescence lifetime correlation spectroscopy using which we quantify the thermodynamic parameters relevant to understanding of non-equilibrium dynamics of protein systems. This technique is based on 2-Dimensional Fluorescence Lifetime Correlation Spectroscopy introduced by Tahara and coworkers (18, 19). Adaptation of optimization protocols from multi-dimensional NMR relaxation spectroscopy (20-22) results into fast and robust data fitting enabling us to measure asymmetric kinetics in protein conformational changes. This is a time resolved fluorescence technique that utilizes changes in fluorescence emission lifetimes as a measure for changes in microscopic environment of the fluorophore and thus can detect distinct protein conformations. Following section covers the prerequisite introduction to experimental aspects of time resolved fluorescence.

1.3. Fluorescence

Fluorescence is the process by which a molecule in an electronically excited state decays to ground state via emission of a photon. Molecules exhibiting this phenomenon are called chromophores or

fluorophores. Excitation can be achieved by absorption of a single or multiple photons. On an average, the emitted photon has lower energy than that of the absorbed photon and hence higher wavelength. This change in wavelength is known as Stokes' shift. This is helpfully illustrated using the Jablonski diagram (Figure 1.1), where the electronic states of a molecule are represented according to their energy levels and the transitions between them are denoted by directional arrows (23, 24). Excitation is followed by rapid internal conversion involving relaxation from higher energy excited states to first excited singlet state (S_1). From here, the molecule can return to ground state via spontaneous radiative decay (fluorescence) or non-radiative decay.

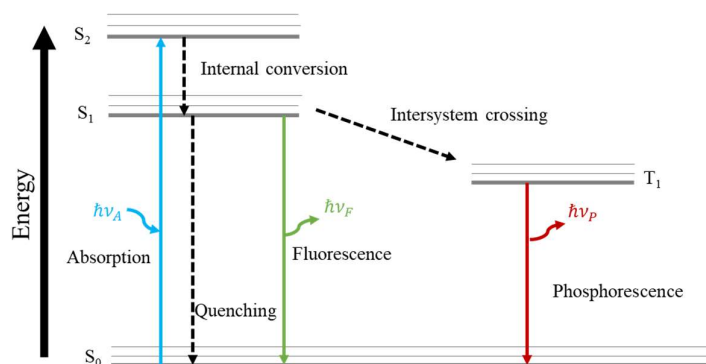


Figure 4.1: The Jablonski diagram. Horizontal grey lines represent the electronic states where S_0 is the ground state, S_1 and S_2 are electronic excited states. Arrows represent energy transfer: Photon absorption is marked blue, emission of fluorescent photon is marked green, and phosphorescence is marked red. The non-radiative decay processes are denoted with dashed arrows.

The rate of relaxation to ground state is a function of the molecular orbital overlap of the excited state with the ground state (S_0) and the electronic microenvironment of the chromophore. The rate of radiative relaxation also known as the emissive rate is given by Γ and the non-radiative decay rate is given by k_{nr} which comprises of various processes such as intersystem crossing to triplet state and dissipation via collisions. The fluorescence quantum yield (Q) is the probability that the excited state of a molecule will deactivate to the ground state with emission of a photon which is given by,

$$Q = \frac{P_r}{P_r + P_{nr}} \quad 1.1$$

Where P_r is the probability of relaxation by radiative decay and P_{nr} is the probability of relaxation by a non-radiative decay. Fluorescence emission is a spontaneous process and individual emissions are independent of each other. This results in a single exponential decay statistics with a characteristic rate known as fluorescence lifetime (τ) which can be given by

$$\tau = \frac{1}{P_r + P_{nr}} \quad 1.2$$

Alternatively, it can also be referred to as the average time between excitation and fluorescence emission. This process has typical timescales of nanoseconds (10^{-9} s) and the wavelengths associated typically lie on the visible spectrum (400-800nm) of the electromagnetic radiation. Fluorescence emission from a system can be characterized by physical observables such as intensity, wavelength (energy), polarization and lifetime with very high sensitivity and specificity. Since the emission properties are directly linked to the molecular structure and the electronic microenvironment of the chromophore, measurement of these properties gives us the ability to directly monitor chemical processes at molecular level. This is the essence of fluorescence spectroscopy.

Time-resolved fluorescence measurements:

We can classify fluorescence measurements in two categories: steady-state and time-resolved. Steady-state measurements are the most widely used and the technique consists of irradiating the sample with a continuous illumination of light. The intensity or the emission spectrum is then recorded. Time-resolved measurement includes illumination of the sample with a pulsed source of light. Such measurements yield information about the intensity decay or the anisotropy decay of the sample. The pulse width of light should be much shorter than the decay timescales and the detector should have a sufficiently high time resolution. Steady-state measurements can be thought of as an average of the time-resolved processes. Although it is practically difficult and expensive to setup time-resolved fluorescence measurements, most of the temporal information about the molecule and its dynamics is lost during the steady-state measurements. For instance, a molecule can exhibit two conformations with

different intensities and switch from one state to another. A steady-state measurement of such system will yield an intensity which is average of intensity of two states however, a time-resolved experiment will not only detect two different intensities, but one can also probe the rate of conformational transitions. Time-resolved fluorescence measurements can thus provide the dynamic information about the chromophore and its environment.

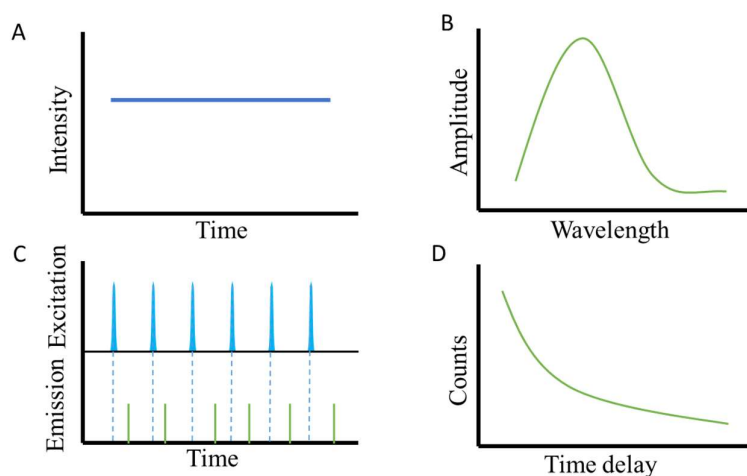


Figure 1.5. Schematic comparison of steady-state and time-resolved fluorescence measurements. *A* and *C* are the *C* are the Temporal profiles of excitation intensities for steady-state and time-resolved measurements respectively. *B* shows the Emission spectra obtained using a continuous illumination. *D* shows a histogram of the time-resolved fluorescence delays measured using pulsed excitation shown in *C*.

1.4. Thesis plan

The following chapters focus on development and implementation of a time resolved fluorescence technique known as Two-Dimensional Fluorescence Lifetime Correlation Spectroscopy. Using this technique, we monitor the chemical exchange dynamics of single protein molecules undergoing conformational change and infer the reaction dynamics and effect of local microenvironment.

Chapter 2 provides the prerequisite information for the experimental and analytical techniques used this this thesis. It introduces the basic techniques of fluorescence microscopy, single molecule fluorescence detection, and principles of time resolved fluorescence measurements using Time Correlated Single Photon Counting (TCSPC). Then we discuss the computational technique used to

analyze time-resolved fluorescence data known as Inverse Laplace Transform (ILT) using the Maximum Entropy Method (MEM). This chapter is written with broader scope for readers who are not familiar to fluorescence microscopy techniques.

Chapter 3 covers the detailed procedure for 2D-FLCS analysis of the time resolved fluorescence time traces. The core of the technique is based around regularized optimization to obtain the Two-Dimensional Inverse Laplace Transform (2D-ILT) of the fluorescence data. We introduce a fast and robust alternative to the conventional MEM technique which has been adapted from 2D NMR relaxometry techniques. There are three key advancements which makes our technique more efficient namely, data compression using Singular Value Decomposition (SVD), Tikhonov Regularization instead of the MEM regularizer and finally, the Butler-Reeds-Dawson algorithm which converts the constrained optimization in the Tikhonov regularization to an unconstrained problem which results in tremendous speedup. We demonstrate the working of this technique using discrete time markovian monte carlo simulations.

Chapter 4 demonstrates the implementation of single molecule 2D-FLCS to investigate nonequilibrium thermodynamics of a trans-membrane proton pump, Bacteriorhodopsin. We independently quantify the forward and reverse conformational transitions through portions of the catalytic cycle during operation and show that microscopic irreversibility is the fundamental operating principle of optically driven nonequilibrium pump proteins. We quantify the reaction flux, affinity, associated entropy production and enthalpy as a function of temperature and infer that the degree of irreversibility decreases with increasing temperature following the Gibbs-Helmholtz relation.

Chapter 5 explores the coupling of internal reaction coordinate with the surrounding solvent in the protein eGFP at microsecond timescales. Using 2D-FLCS, we monitor the equilibrium fluctuations inside the chromophore pocket of eGFP protein and quantify transition rates with respect to solvent viscosity. We show that the internal dynamics are coupled to the solvent viscosity however the dependence deviates from predications from Kramer's theory and is attributed to proteins internal

friction.

Chapter 6 concludes the thesis summarizing the advantages of the technique, its implementation cases, limitations, and possibilities of its development in the future.

REFERENCES

1. Frauenfelder, H., G. Chen, J. Berendzen, P. W. Fenimore, H. Jansson, B. H. McMahon, I. R. Stroe, J. Swenson, and R. D. Young. 2009. A unified model of protein dynamics. *Proceedings of the National Academy of Sciences* 106:5129-5134.
2. Frauenfelder, H., S. G. Sligar, and P. G. Wolynes. 1991. The energy landscapes and motions of proteins. *Science* 254:1598-1603.
3. Henzler-Wildman, K., and D. Kern. 2007. Dynamic personalities of proteins. *Nature* 450:964-972.
4. Kendrew, J. C., G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips. 1958. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* 181:662-666.
5. Kim, Y., L. Bigelow, M. Borovilos, I. Dementieva, E. Duggan, W. eschenfeldt, C. Hatzos, G. Joachimiak, H. Li, N. Maltseva, R. Mulligan, P. Quartey, A. Sather, L. Stols, L. Volkart, R. Wu, M. Zhou, and A. Joachimiak. 2008. High-Throughput Protein Purification for X-Ray Crystallography and NMR. *Advances in Protein Chemistry and Structural Biology*. A. Joachimiak, editor. Academic Press, pp. 85-105.
6. Clore, G. M., and A. M. Gronenborn. 1989. Determination of Three-Dimensional Structures of Proteins and Nucleic Acids in Solution by Nuclear Magnetic Resonance Spectroscopy. *Critical Reviews in Biochemistry and Molecular Biology* 24:479-564.
7. Wüthrich, K. 1990. Protein structure determination in solution by NMR spectroscopy. *Journal of Biological Chemistry* 265:22059-22062.
8. Wüthrich, K. 2001. The way to NMR structures of proteins. *Nature Structural Biology* 8:923-925.
9. Kelly, S. M., T. J. Jess, and N. C. Price. 2005. How to study proteins by circular dichroism. *Biochim Biophys Acta* 1751:119-139.

10. Ebbing, D. D., and S. D. Gammon. 1998. General chemistry 6th ed. New York: Houghton Mifflin Company.
11. Michaelis, L., and M. L. Menten. 1913. Die kinetik der invertinwirkung. *Biochem. z* 49:352.
12. Frauenfelder, H., and P. G. Wolynes. 1985. Rate theories and puzzles of heme protein kinetics. *Science* 229:337-345.
13. Portman, J. J., S. Takada, and P. G. Wolynes. 2001. Microscopic theory of protein folding rates. II. Local reaction coordinates and chain dynamics. *The Journal of Chemical Physics* 114:5082-5096.
14. Crooks, G. E. 1999. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E* 60:2721.
15. Crooks, G. E. 1998. Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *Journal of Statistical Physics* 90:1481-1487.
16. De Groot, S. R., and P. Mazur. 2013. *Non-equilibrium thermodynamics*. Courier Corporation.
17. Fang, X. A., Jin, B. Journal Name: *Annual Review of Biophysics*, J. V. 49, and J. I. 1. 2020. Nonequilibrium Thermodynamics in Cell Biology: Extending Equilibrium Formalism to Cover Living Systems. Journal Name: *Annual Review of Biophysics*; Journal Volume: 49; Journal Issue: 1; Medium: X; Size: 227 to 246.
18. Ishii, K., and T. Tahara. 2013. Two-Dimensional Fluorescence Lifetime Correlation Spectroscopy. 1. Principle. *The Journal of Physical Chemistry B* 117:11414-11422.
19. Ishii, K., and T. Tahara. 2013. Two-Dimensional Fluorescence Lifetime Correlation Spectroscopy. 2. Application. *The Journal of Physical Chemistry B* 117:11423-11432.
20. Chouzenoux, E., S. Moussaoui, J. Idier, and F. Mariette. 2010. Efficient maximum entropy reconstruction of nuclear magnetic resonance T1-T2 spectra. *IEEE Transactions on Signal Processing* 58:6040-6051.

21. Su, G., X. Zhou, L. Wang, X. Wang, P. Yang, S. Nie, and Y. Zhang. 2019. Improved Butler–Reeds–Dawson algorithm for the inversion of two-dimensional NMR relaxometry data. *Mathematical Problems in Engineering* 2019.
22. Venkataramanan, L., Y.-Q. Song, and M. D. Hurlimann. 2002. Solving Fredholm integrals of the first kind with tensor product structure in 2 and 2.5 dimensions. *IEEE Transactions on signal processing* 50:1017-1026.
23. Albrecht, C. 2008. Joseph R. Lakowicz: Principles of fluorescence spectroscopy. Springer.
24. Jablonski, A. 1933. Efficiency of Anti-Stokes Fluorescence in Dyes. *Nature* 131:839-840.

Chapter 2. Technical background

This chapter describes the preliminary experimental and analytical techniques which constitute the building blocks of the work presented in this thesis. Fluorescence microscopy has become increasingly beneficial for studying biological systems since it provides high sensitivity down to single photon detection and high specificity since the signal only originates from the target fluorescent molecules. It has also gained popularity due to its ease of use, recent advances in fast and sensitive detectors, newly developed fluorophores, and methods of tagging biomolecules with fluorophores. A fluorescence microscope thus combines the magnifying properties of a light microscope with the emission properties of fluorescent molecules.

2.1 Fluorescence microscopy.

Basic components of a fluorescence microscope are an intense light source such as Xe-arc lamp or laser, a magnifying objective, a dichroic filter, and a detector such as Avalanche Photo Diode (APD), Photo Multiplier Tube (PMT) or a CCD camera. The most common arrangement of these components is epifluorescence microscopy where the illumination and the collection of fluorescence is done using the same objective. Based on the geometry of illumination and image generation, there are two widely used configurations namely, wide-field, and confocal microscope. In the case of wide-field microscope, a parallel beam excites all the fluorophores in the field of view at once and the resulting fluorescence is captured on a camera. This simultaneous illumination allows for faster imaging however it suffers from lower contrast due to fluorescence from sample out of focus Fig.2.1. To circumvent this problem, the idea of rejecting the out-of-focus was proposed by Minsky in 1957 (1) and the confocal microscope was developed in 1977 by Sheppard et al (2). Confocal microscopy employs a focused illumination at a point in the sample and collects the fluorescence with a point detector through a pinhole to reject the background from out-of-focus light. The pinhole is placed at the conjugate focal plane such that only the fluorescence originating from the part of the sample in the focal volume of the objective lens, can

pass through. This acts as a special filter to reduce the incidence of out-of-focus light on the detector and hence improves the image contrast. The beam is then raster scanned along a grid usually using galvo mirrors to generate the point-by-point image of the sample. Fig.2.2 shows a schematic of a confocal microscope setup. This also provides an optical sectioning capability which can be used to obtain a 3-dimensional reconstruction of the fluorescent sample.

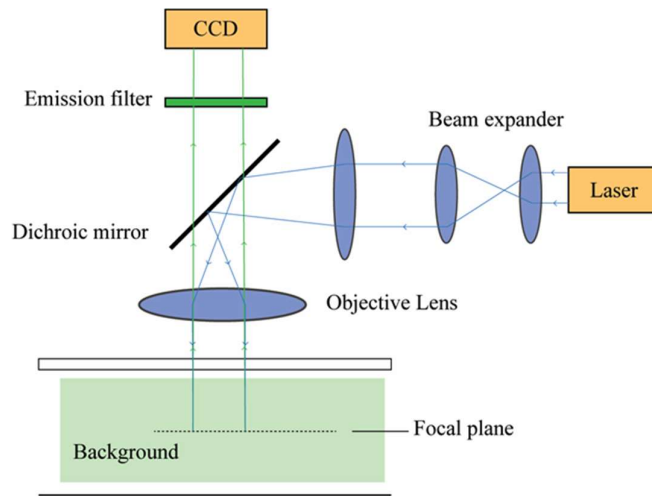


Figure 2.1 Schematic of a widefield microscope. The light beam coming from the laser is focused on the back focal plane of the objective lens which causes a parallel beam to be incident on the sample in focus. The fluorescence from the illuminated area is collected by the objective and passes through the dichroic followed by an emission filter after which the image is formed on the CCD camera.

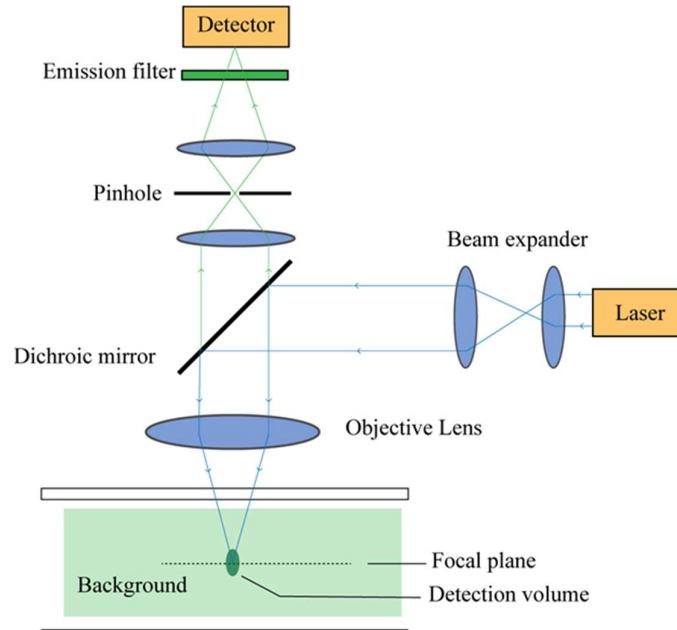


Figure 2.2. Schematic of a confocal microscope. The light coming from the laser source is focused inside the sample using an objective lens. The fluorescence coming from the diffraction limited spot is collected by the objective and passes through the dichroic mirror onto a confocal pinhole which filters out the fluorescence coming from out of focus planes. This light is then passed through an emission filter and is detected using a point detector such as APD or PMT.

Sharpest image contrast is obtained when the pinhole size and the excitation spot size is the smallest. The minimum size of excitation volume is limited by numerical aperture (NA) of the objective lens, refractive index of the sample, and the wavelength of the light used. The limit of resolution for a light microscope was introduced by Ernst Abbe in 1873 which states that the minimum resolvable distance between two points using light of wavelength λ , in a medium of refractive index n and the convergence half-angle θ is given by.

$$d = \frac{\lambda}{2 n \sin\theta} = \frac{\lambda}{2NA} \quad 2.3$$

The diffraction limited excitation spot has a shape of a 3-dimensional Gaussian function and is known as the point spread function (PSF) where, the full width half maximum (FWHM) in the x-y plane is given by Abbe's limit and the along the z-axis is typically 3-5 times the lateral resolution.

2.2 Single molecule detection

The power to detect a bright signal against a dark background within a diffraction limited volume and with high specificity allows for the ability to detect fluorescence from even a single molecule. It is appealing to measure single molecule fluorescence because the measured signal now represents the absolute quantities rather than relative intensity values. The most important reason to use single molecule detection is to avoid ensemble averaging. This permits us to probe time varying processes of the molecule directly if the photophysical properties change too. For instance, consider a protein that fluctuates between an open and a closed conformation such that the fluorescence intensity is reduced in a closed state and increased in open state. In a bulk spectroscopic measurement, we would only be able to measure the average fluorescence intensity which will be a function of the probability distribution of the conformations, and this yield no information about the rate of switching between the two states. On the other hand, if we probe the same system at the single molecule level, the measured signal will consist of a time trace consisting regions of high and low intensity detected directly from one molecule. One can then obtain the information about frequency of 'flips' and construct a frequency histogram and obtain the rate of transition between the states. A distribution provides more information than the average value alone especially in the case of a heterogeneous sample. Single molecule fluorescence is also a local reporter of the molecule's microenvironment can reveal properties of the surroundings of a fluorophore. Another advantage of single molecule measurement is that it removes the need of bulk synchronization which typically involves complex instrumentation (such as NMR). If the sample has multiple states, one can only probe the properties till the states remain correlated even after synchronization. Thus, we establish the motivations and advantages of using single molecule fluorescence detection.

The earliest accounts of single molecule fluorescence measurements are from 1970s when Hirschfeld observed antibodies labelled with multiple fluorescein molecules (3). In 1989, Moerner et.al detected single dye molecule embedded in organic crystals at liquid helium temperatures (4). Keller and coworkers were the first ones to report single molecule measurement at room temperature in 1990 where, they detected single rhodamine-6G molecules flowing through a small detection volume (5). In 2014, the Nobel Prize in Chemistry was awarded to Stefan Hell, W.E. Moerner and Eric Betzig for development of single molecule localization and super resolution techniques. Since then, the field of single molecule fluorescence-based techniques has advanced rapidly introducing powerful techniques such as sm-FRET, PALM, STORM, etc.

To get an idea of experimentally measured quantities in case of single molecule detection, we can consider a typical fluorophore Alexa-488 illuminated using a 532 nm laser with intensity of 100W/cm². We can estimate the number of photons which can be detected from a single Alexa molecule dissolved in water. The absorption cross-section σ is defined as the effective area presented by a molecule to the incident radiation.

$$\sigma = \frac{2.303\varepsilon}{N_A} = 2.7 \text{ \AA}^2 \quad 2.2$$

Where N_A is the Avogadro's number. From these values, we can estimate and compare the molar extinction coefficient of Alexa488 determined from bulk absorption spectroscopy;

$\varepsilon = 73,000 \text{ L/mol/cm}$. To estimate the number of photons illuminating this Alexa molecule, we first calculate energy per photon using.

$$E = \hbar c / \lambda \quad 2.4$$

Where \hbar is the Plank's constant, c is speed of light and λ is the wavelength of the laser which yields $E = 3.7 \times 10^{-19} \text{ J/photon}$. Hence with given illumination intensity, we obtain a photon flux of

$q_p = 2.7 \times 10^{20}$ photons/s/cm². Thus, the number of photons absorbed by the molecule is the product of incident flux and the extinction coefficient.

$$q_a = q_p \cdot \sigma_{Alexa} \approx 7.3 \times 10^4 \text{ photons} \quad 2.5$$

The quantum yield of Alexa 488 is $Q = 0.92$. Hence, the amount of fluorescently emitted photons is.

$$q_a \cdot Q_{Alexa} = 6.7 \times 10^4 \text{ photons} \quad 2.6$$

We can estimate the upper bound of the number of detectable photons using an objective lens of a high numerical aperture (N.A. = 1.4) and refractive index of media ($n = 1.52$ for immersion oil). Using the relation $N.A. = n \sin \theta$, we get $\theta \approx 67^\circ$ which leads to only 12% of the solid angle accessible from the fluorescence emission. Thus, we have $\approx 8 \times 10^3$ detectable photons from a single Alexa 488 molecule which can be easily detected by modern detectors.

For freely diffusing fluorophores into the detection volume, the detected signal takes the form of emission bursts which originates from molecules passing through the confocal volume. One can then evaluate the correlation function of these fluorescence traces and determine properties of the fluorophores such as diffusion coefficient, and anisotropy. The measurement is temporally limited by the average residence time of the diffusing molecules in the detection volume ($\approx \mu\text{s}$). With the advent of surface immobilization techniques, one can covalently fix single molecules on the coverslip surface and localize single molecules and perform measurement over prolonged time scales limited by photo bleaching (ms-s).

2.3 Time Correlated Single Photon Counting (TCSPC)

Fluorescence microscopy when combined with time resolved measurement capability can be a powerful tool to obtain temporal information at microscopic scales. The most common apparatus for time resolved fluorescence experiments is Time Correlated Single Photon Counting (TCSPC). This technique uses a pulsed laser with high repetition rate and pulse width of order of femtoseconds to

picoseconds which is typically the output of a mode-locked laser. In principle, TCSPC detects single photons and measures the time of their arrival with respect to a reference signal from the light source. Using high repetition rate, one can accumulate sufficient photons and get a precise statistical distribution of their arrival times. The photon detection rate is usually set to a very low value such that coincidence of collecting two photons in the same signal period is extremely unlikely. There are numerous advantages of using TCSPC mainly, it has a near-ideal photon efficiency meaning that every photon sensed by the detector ends up in the resultant photon distribution and secondly, it has an extremely high time resolution which is limited by the instrument response function (IRF) of the detector (typically ~ 50 ps). (6-8)

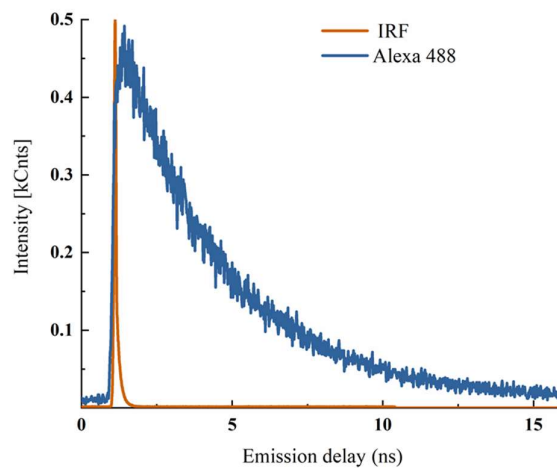


Figure 2.3 Fluorescence decay of Alexa 488 in water. The trace in blue shows the fluorescence emission delay histogram of Alexa 488 solution. The experimentally observed histogram is the result of the true decay convoluted by the Instrument Response Function (IRF) shown in orange. The decay can be fit to obtain the fluorescence lifetime of Alexa 488 which in this case is 4.8 ns.

The working of TCSPC can be understood as a fast stopwatch controlled by a START signal and a STOP signal. An internal clock is initiated with the START signal coming from the laser excitation pulse and the STOP signal is sent by the detector on arrival of a photon. The time delay between the

START and the STOP pulse is then saved as one data point for the arrival time. With large number of repetitions, a histogram of the stored arrival times is given as output. Specifically, the components of TCSPC electronics consist of Constant Fraction Discriminators (CFD), Time-to-Amplitude Converter (TAC), Analog to Digital Converter (ADC) and digital memory. Fig.2.4.

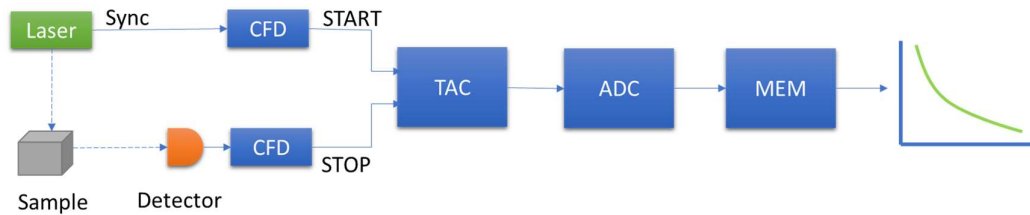


Figure 6.4 Components of a TCSPC setup. The fluorescence photons from the sample arrive on the detector. The signal is relayed to the Constant Fraction Discriminator (CFD) which determines the timing of the pulse. The sync pulse from the laser driver is also connected to another CFD to determine its timing. Outputs of both CFDs are connected to the Time to Amplitude Converter (TAC) which, then is connected to the Analog to Digital Converter (ADC) and this timing information of the incident photon is stored in the memory (MEM) and is further sent to the computer.

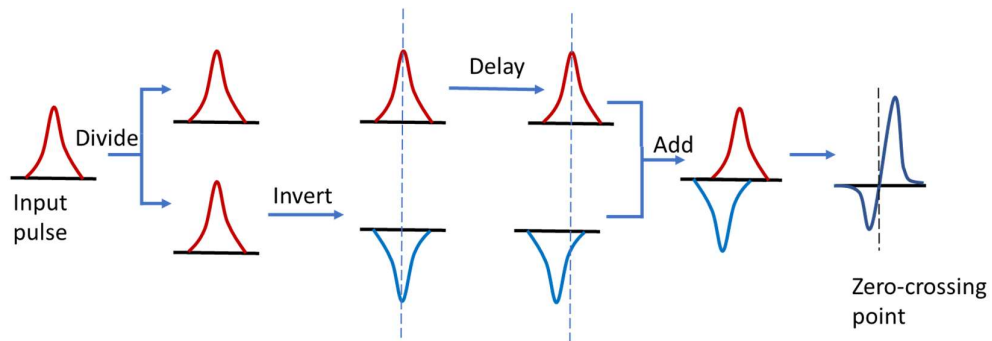


Figure 2.5. Working of a Constant Fraction Discriminator (CFD). The input pulse is split into two components and one of them is inverted. A delay circuit produces a known time lag in one of the pulses. Both pulses are then added together. The zero-crossing point of the resultant pulse is determined. This point represents the time when the input pulse is at its maximum amplitude. Such timing calculation is robust to fluctuations in the absolute amplitudes of the input pulse.

First the incoming signals are threshold filtered to eliminate electronic noise. The CFDs are implemented at both the START and the STOP inputs and are used to extract the precise timing information of an electronic pulse. In a CFD, the input signal is compared to an inverted and delayed version of itself. The comparison signal changes its sign when a constant fraction of the detector pulse height is reached. This zero-crossing point is suitable to derive the timing of the pulse independent of the pulse amplitude as depicted in Fig.2.5. The TAC is a fast clock which initiates counting in the form of a growing voltage ramp signal on the arrival of the start pulse and halts it on the arrival of the stop pulse. The height of this ramp thus corresponds to the timing delay between the START and the STOP pulse Fig.2.4. The ADC then converts this analog voltage to a digital output and stores it in the memory. The resolution of timing detection is given by the number of channels in the ADC and has units picoseconds/channel. Modern TCSPCs (PicoQuant - PicoHarp 300 (9) used in the experiments presented in this work) have resolution as low as 4 ps/channel. The shortest lifetime that can be measured however, is limited by the quadrature sum of all the components that create noise in the timing signal i.e., electronics jitter, the IRF of the detector, and the pulse width of the excitation laser.

$$\tau_{min} = \sqrt{e_{jitter}^2 + e_{IRF}^2 + e_{pw}^2} \quad 2.7$$

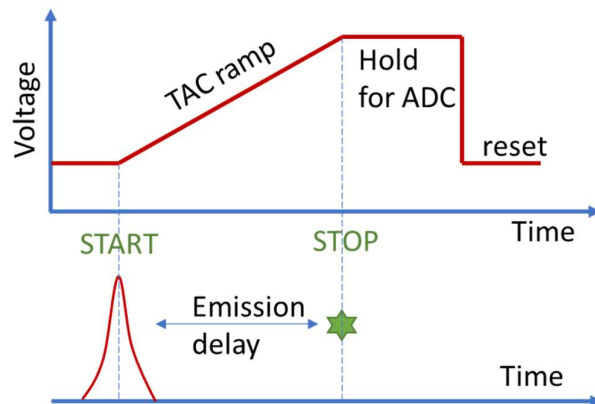
Where e represents the full width half maximum error of the respective component.

Fig.2.3. shows an example fluorescence lifetime data set obtained from TCSPC measurement of Alexa 488 dye in water illuminated using a 488nm laser and a PMT hybrid detector. Black curve represents the instrument response function of the detector. The fluorescence lifetime of A488 can thus be estimated by either fitting the tail of the obtained decay or by using IRF re-convolution and is found to be 4.8 ns.

2.3.1. Advanced TCSPC:

The above explained method is adequate to obtain the fluorescence lifetime histograms.

However, in modern practices, we can implement efficient photon collection and memory storage methods. One such practice is the use of TCSPC in reverse mode. Since the count rate is sufficiently low, most of the pulse intervals have no photons in them. Thus, in reverse mode, the light source is connected to the STOP input of the TAC and the detector is connected to the START. By adjusting a fixed delay in the reference pulse from the laser driver, the TAC is only triggered if a photon is detected. The time axis is hence reversed but this makes the collection process highly efficient (7). Instead of storing the photons in a histogram, a global timing can also be performed separately for each photon which records the time from the start of the photon collection experiment. This is called the Time-Tagged Time-Resolved mode (TTTR) and is useful in photon correlation analysis. There are two modes of operating TTTR mode; T2 – where both TCSPC inputs are connected to detectors and T3 – where detector is connected to STOP, and laser SYNC is connected to the STOP input. (10) The work in this



this thesis utilizes the T3 mode measurements.

Figure 2.6 Working of a Time to Amplitude Converter (TAC). This is a linear ramp generator circuit which has a start trigger and a stop trigger. The resultant analog voltage is proportional to the time delay between both triggers. This voltage is held till the ADC processes it and is then reset to the default value to be ready for accepting the next start pulse.

2.4 Analysis of fluorescence decays

Fluorescence parameters such as the absorption and emission spectra, lifetime, quantum yield,

etc. are extremely sensitive to the local microenvironment of the fluorophore. This same sensitivity, however, makes it difficult to interpret the data if various configurations of the microenvironment exist. Though there exists a method of associating absorption or emission peaks and peak shifts with the properties of the local environment, there is no consensus on interpreting fluorescence decays and associating the parameters with the microenvironment. Time-resolved fluorescence data is thus complex and cannot be analyzed by straightforward curve fitting procedures.

Given a time trace of fluorescence photons, the emission delays follow an exponential distribution and can be represented as

$$I(t) = \sum_{j=1}^n a_j \exp(-t/\tau_j) \quad 2.8$$

Consisting of n components with amplitudes a_j and respective fluorescence lifetimes τ_j . However, experimentally measured TCSPC data is a convolution of the intrinsic fluorescence intensity decay and the instrument response function (*IRF*) given by

$$I^{obs}(t) = \int_0^t IRF(t_i - t_0) I(t - t_i) dt \quad 2.9$$

Where t_0 is the zero-time shift of the detector. In time resolved experiments, one is interested in determining τ 's which are the fluorescence lifetimes and their corresponding amplitudes (a) which must be obtained by fitting the observed intensity decays in Eq.2.8. This process of converting the information from intensity in time-space to the amplitudes in lifetime-space is also known as the Inverse Laplace Transform (ILT). The result of ILT of fluorescence decays yields what is known as the lifetime spectra Fig.2.5. which is more discernable for spectroscopists. However, there is no analytical method for the ILT. In other words, it is clear from Eq.2.8. that given the lifetime and the amplitude information, one can easily obtain the intensity-time distribution but, given for a given intensity-time plot, one cannot

trivially calculate the lifetime spectra. Thus, in order to achieve the ILT, one must use numerical approximation methods.

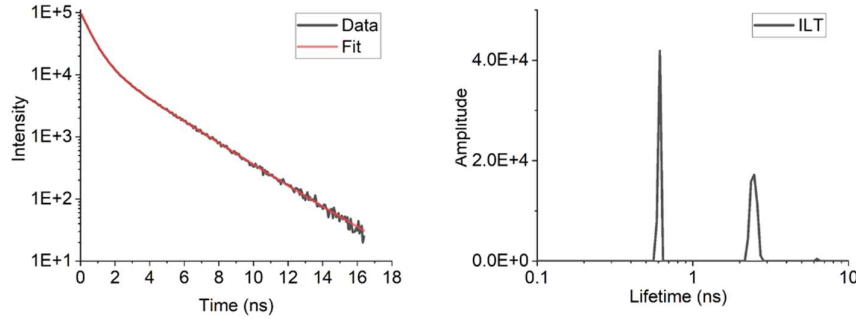


Figure 2.7 Inverse Laplace Transform. A shows a simulated biexponential fluoresce decay and B shows its corresponding Inverse Laplace Transform revealing the component lifetimes at 0.6 ns and 2.5 ns.

Traditional curve fitting procedures use a nonlinear least-square minimization technique to estimate the parameters of the distribution. It consists of defining a χ^2 function which is a measure of the error between the data and the fit and is given by.

$$\chi^2 = \sum_i (c_i - m_i)^2 w_i \tag{2.10}$$

Where c_i is the experimental data, m_i is the fit and w_i is weighting which is typically $w_i = 1/\sigma_i^2$; σ_i being the point-wise standard deviation of the data. One typically begins with an initial guess of the fit parameters, computes the χ^2 using Eq.2.10, and iteratively updates the guess to minimize it. It is observed that standard methods of nonlinear least-square curve fitting are inaccurate because the small number of counts lead to the errors having Poissonian distribution rather than Gaussian. Even for decays consisting of single exponential components, the error in lifetime estimates maybe as high as 15%. There have been approaches to reduce this error by modifying the weighting function (11) however, inaccuracies still exist for fitting multi-exponential decays using. The reason behind this imprecision is

because the ILT is notorious for being an ill-conditioned problem because it is numerically unstable. In other words, there are multiple possible solutions which can be fit to a multi-exponential decay with given confidence and, small fluctuations in the data in Eq.2.8 $I(t)$ can lead to large fluctuations in the ILT given by $a(\tau)$.

A typical approach of solving ill-conditioned problems and prevent overfitting is regularization. This involves adding a regularizing function to the least square error term which acts as a penalty for every new information added to the solution. Thus, it helps in eliminating unfeasible solutions and promotes stable solutions. The combined function is known as the objective function and takes the form.

$$Q(c_i) = \chi^2 + \alpha R(c_i) \quad 2.11$$

Where, α is a regularization constant which determines the weight of the regularizing term with respect to χ^2 term. $R(c_i)$ can take various functional forms however they achieve a common goal. Table.2.1. shows a few examples of commonly used regularization functions.

Regularizing function	Functional form
Maximum Entropy	$R(p_i) = -\sum p_i \log p_i$
Tikhonov / L2 / Ridge regression	$R(p_i) = \sum p_i^2$
L1 / Lasso regression	$R(p_i) = \sum p_i $

Table 2.1 Common regularizing functions and their functional forms

2.4.1 Maximum Entropy Method

First account of regularized fitting for analyzing fluorescence data was published by Livesey et.al. in 1987 (12) where they used the Maximum Entropy Method (MEM) which has been the method of choice for ILT analysis of fluorescence decays. MEM was originally implemented famously for image reconstruction by Bryan and Skilling in 1984 (13). In principle, MEM tries to fit the data without adding any extra correlations and thus yields solutions which follow Bayesian statistics (14) (15). Among all feasible solutions possible, one chooses the solution with maximum information entropy given by:

$$S = \int_0^{\infty} \left(a(\tau) - m(\tau) - a(\tau) \cdot \ln \left(\frac{a(\tau)}{m(\tau)} \right) \right) d\tau \quad 2.12$$

Where, $a(\tau)$ is the amplitude of lifetime component τ and $m(\tau)$ is a prior model of the about the system (typically, it is set to constant at the beginning of fitting indicating no prior information).

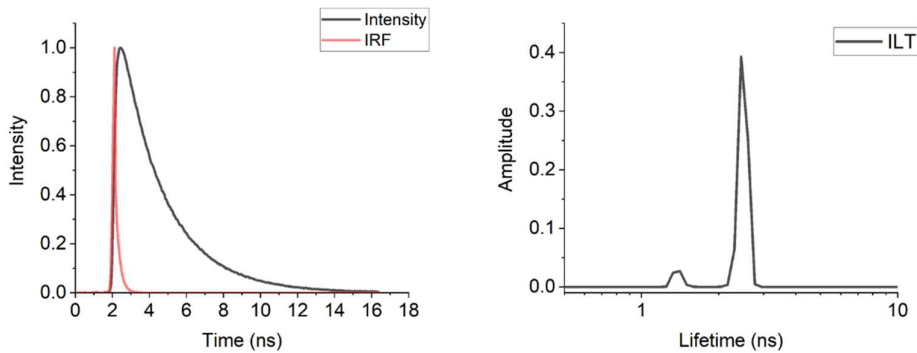


Figure 2.8. ILT on experimental data using Maximum Entropy Method (MEM). Panel A shows the experimentally obtained fluorescence decay from eGFP protein solution (black). The observed decay is the result of true exponential decay convoluted by the IRF (shown in red). Panel B shows the resultant ILT obtained using Maximum entropy method which shows the lifetime components at 1.5 ns and 2.7 ns.

The MEM regularized optimization procedure is as follows:

- i. One defines the Objective function following Eq.2.10 and Eq.2.12

$$Q(c) = \chi^2(c) + \alpha S(a) \quad 2.13$$

Note that in above equation, χ^2 term is summed over time space and the S term is summed over lifetime-space.

- ii. First guess solution is set to a uniform prior multiplied with a small constant.
- iii. χ^2 term is evaluated from the guess solution using the forward function given in Eq.2.10
- iv. α is set to a high value which emphasizes the entropic penalty for the initial iterations.
- v. One then minimizes the objective function Q and stores the optimum solution as m which sets the prior in Eq.2.12 for the next iteration.
- vi. This process is iterated while gradually decreasing the value of α until a suitable stopping criterion is reached.
- vii. The result is a solution which has the least error from the data while having the maximum information entropy. In other words, this procedure yields a solution with the least bias or external correlations.

Various methods have been suggested for the stopping criteria but, they are intended to achieve the same goal (12, 13, 15, 16). Fig.2.6. shows an example experimental fluorescence decay and its ILT obtained using MEM.

References

1. Minsky, M. 1961. Microscopy apparatus US patent 3013467. USP Office, Ed. US.
2. Sheppard, C. J. R., and A. Choudhury. 1977. Image Formation in the Scanning Microscope. *Optica Acta: International Journal of Optics* 24:1051-1073.
3. Hirschfeld, T. 1976. Optical microscopic observation of single small molecules. *Applied optics* 15:2965-2966.
4. Moerner, W. E., and L. Kador. 1989. Finding a single molecule in a haystack. Optical detection and spectroscopy of single absorbers in solids. *Analytical Chemistry* 61:1217A-1223A.
5. Shera, E. B., N. K. Seitzinger, L. M. Davis, R. A. Keller, and S. A. Soper. 1990. Detection of single fluorescent molecules. *Chemical Physics Letters* 174:553-557.
6. Becker-Hickl. Classic TCSPC-Becker-Hickl.
7. Instruments, E. What is TCSPC?
8. PicoQuant. TCSPC- Technical note.
9. PicoQuant. PicoHarp 300.
10. PicoQuant. Time Tagged Time-Resolved Fluorescence Data Collection in Life Sciences.
11. Turton, D. A., G. D. Reid, and G. S. Beddard. 2003. Accurate analysis of fluorescence decays from single molecules in photon counting experiments. *Analytical chemistry* 75:4182-4187.
12. Livesey, A., and J. Brochon. 1987. Analyzing the distribution of decay constants in pulse-fluorimetry using the maximum entropy method. *Biophysical journal* 52:693-706.
13. Skilling, J., and R. Bryan. 1984. Maximum entropy image reconstruction-general algorithm. *Monthly notices of the royal astronomical society* 211:111.
14. Gull, S. F., and J. Skilling. 1984. Maximum entropy method in image processing. In *IEE Proceedings F-Communications, Radar and Signal Processing*. IET. 646-659.
15. Gull, S. F., and J. Skilling. 1999. Quantified maximum entropy MemSys5 users' manual. Maximum Entropy Data Consultants Ltd 42.

16. Esposito, R., C. Altucci, and R. Velotta. 2013. Analysis of simulated fluorescence intensities decays by a new maximum entropy method algorithm. *Journal of Fluorescence* 23:203-211.

Chapter 3: Fast and robust two-dimensional inverse Laplace transformation of single-molecule fluorescence lifetime data

Section 3.1 Introduction

Biological systems viewed from a physical standpoint are non-equilibrium dynamical systems that exchange between metastable states to carry out a certain function. To determine their kinetics, dynamics, efficiency, and working, the first step is to measure these exchange rates. This is especially important in case of biological macromolecules such as proteins and nucleic acids.

The first experimental report on probing the chemical exchange rates was provided by Jeener et.al (1). who used 2D NMR spectroscopy to look at exchange rates in heptamethylbenzenonium ion. Fayer, et.al (2) also demonstrated ultrafast 2D IR vibrational echo exchange spectroscopy. Exchange in larger macromolecules such as proteins is more efficiently probed using fluorescence-based techniques since NMR and IR characterization becomes increasingly difficult with increasing molecular size.

Single molecule fluorescence spectroscopy has provided great insights into dynamics of biomolecules. sm-FRET has been the primary technique of choice since it provides a signal which is highly sensitive to the distance between chromophores. It involves labelling a biomolecule with two chromophores which are FRET pairs; meaning excited electronic state of one (donor) is coupled with another (acceptor) such that the excitation energy is transferred from donor to acceptor in a non-radiative fashion. One measures appropriate optical filters to get signals from both channels separately. Conformational transitions are observed as rapid changes in emission intensity at both donor and acceptor channels depending on the distance between them. One can then change the placement of the chromophores to obtain more details about the conformational transitions. Hoffmann et.al (3) demonstrated exchange between conformational states of polypeptides and proteins labelled with FRET pairs. However, FRET is experimentally challenging in detecting conformation changes in proteins since it requires precise labelling of two chromophores on a single protein macromolecule and it projects

a complex configuration change onto a single reaction coordinate axis corresponding to distance between the chromophores. (4-7).

Single-molecule 2D fluorescence lifetime correlation spectroscopy (sm-2D-FLCS), first developed by Tahara and coworkers (8-10) and later extended into the single-molecule regime by Schlau-Cohen and workers (11), provides an alternative analysis of single-molecule data that does not sacrifice chemical selectivity or temporal resolution. In this approach, time-correlated single photon counting (TCSPC) is used to detect fluorescence intensity and emission delay time from a single chromophore, either freely diffusing in dilute solution (8-10) or surface immobilized (11). The data are recorded in the TTTR mode which generates a real-time photon stream (characterized by a global macro time) with a recorded emission delay time (characterized by a micro-time) for each registered photon (12). Distinct chemical species in the system are distinguished by their fluorescence lifetime. Thus, chemical exchange between two states can be kinetically resolved on the condition that they have different fluorescence lifetimes.

A series of 2D photon histograms is generated by cataloging all photon pairs separated by a systematically varying waiting time ΔT from an experimentally recorded photon stream (8, 9). A 2D inverse Laplace transform (2D-ILT) of the lifetime histogram generates a 2D-FCLS spectrum at the given ΔT . Analogous to 2D-NMR, species that do not undergo any form of exchange during ΔT appear as diagonal peaks in the 2D spectrum, while species that exchange during ΔT appear as cross-peaks (1). Measuring 2D spectra for a series of ΔT allows the overall correlation function to be effectively split into its components comprising auto-correlations for the diagonal components and cross-correlations for the off-diagonal components. Thus, the chemical exchange kinetics among the components can be measured directly through time correlation functions reflecting chemical exchange between two states (8-11). As this technique relies on a single chromophore to resolve conformational dynamics, it can be applied to systems where fluorescence labeling is challenging or to native proteins with an endogenous chromophore. Additionally, for a conformational transition of interest, the forward and reverse kin

tics are resolved as they are recorded in different regions of the 2D spectrum. The ability to measure asymmetric forward and reverse kinetics is relevant for studying systems operating out of equilibrium. Lastly, as the technique requires no additional optical setup or heavy computing hardware, it can be implemented on any setup already used for FCS or FRET.

The challenge of this experimental approach is the need to perform a 2D-ILT, which is an ill-conditioned problem and is therefore numerically unstable. This results in multiple solutions satisfying the same problem when solved with traditional least-square analysis (13). Instead, this class of problems must be solved via regularized least-square analysis, which imposes a penalty on solutions with undesired features and hence promotes smoother solutions. To date, the calculation of the 2D spectra is traditionally performed using the maximum entropy method (MEM) (8, 9, 11), which produces a solution using Bayesian inference. The application of MEM has two severe limitations: First, it involves a constrained optimization since entropy cannot be defined for negative values, so one must check for the positivity constraint during each iteration and modify the fit accordingly. This makes the approach computationally inefficient. Second, the 2D spectra is fit as a lexicographically ordered 1D vector resulting in a fitting kernel matrix whose size is of the order of the 4th power of number of points in the lifetime spectra, leading to a tremendous computational cost. To decrease the computational cost, the data is often binned non-uniformly before fitting, which results in loss of temporal information in the input data (9).

The challenges of performing 2D inverse Laplace transforms (i.e., converting the time dependent relaxation data to a relaxation rate spectrum) are not unique to fluorescence spectroscopy. Indeed, multi-dimensional Laplace inversions are commonly employed in NMR spectroscopy, for instance in the computation of T1-T2 correlation spectrum from relaxometry data. Work by Venkataramanan and coworkers (14) outlined an efficient algorithm for Laplace inversion in NMR relaxometry data that leveraged singular-valued decomposition (SVD) and Tikhonov regularization. Subsequent improvements of this approach have been reported (15, 16).

In this chapter, we adopt this general approach for application to single-molecule fluorescence spectroscopy analysis. This algorithm reduces the computation time to merely a few seconds per 2D spectrum, and hence enables the analysis of large data sets with high resolution. We first elaborate on the methods to generate 2D photon emission delay correlation histograms and later present the algorithm to obtain 2D-ILT from these histograms. Finally, we compare the performance with traditional techniques and test its robustness to noise. The work in this chapter has been published in *Biophysical Journal* (17).

Section 3.2 Methods

A 2D-FLCS spectrum is generated from a 2D-ILT of a lifetime correlation histogram, easily measured using standard time-correlated single-photon counting (TCSPC) techniques (18, 19). We start with a familiar 1D picture to introduce the concept of ILT. Given a time series of recorded photons, the emission delays are distributed exponentially and can be represented as

$$I(t) = \sum_{j=1}^n a_j \exp\left(-\frac{t}{\tau_j}\right) \quad 3.1$$

for n independent components with amplitudes a_j and respective fluorescence lifetimes τ_j . This can be represented in matrix form and generalized for the sake of broader application as

$$I_i = \sum_{j=1}^n K_{ij} \cdot A_j \quad 3.14$$

Where K is a kernel with pre-defined basis of fluorescence lifetime decays $K_{ij} = (1/\tau_j)\exp(-t_i/\tau_j)$, and A_j is a column vector representing the amplitudes corresponding the species τ_j . Kernel K can be defined such that the experimental lifetimes are included in $[t_{min}, t_{max}]$ with sufficient resolution such that any measured decay can be represented by a pre-calculated kernel and an amplitude vector. Under this construction, $A(\tau_j)$ represents a 1D ILT of $I(t)$, or the lifetime spectrum corresponding to $I(t)$, as we convert the time dependent

t decay into its lifetime components and their probability distribution amplitudes.

This can be easily extended to 2D. Cataloging all the photon pairs separated by a global lag time ΔT in the experimentally recorded time series and having emission delays t_1 and t_2 , we can generate a 2D histogram with axes as t_1 and t_2 , where each point $M(t_1, t_2, \Delta T)$ represents the number of coincidences when a photon with delay t_2 is detected ΔT after having detected a photon with delay t_1 . The process is described in the following section and Figure 3.1.

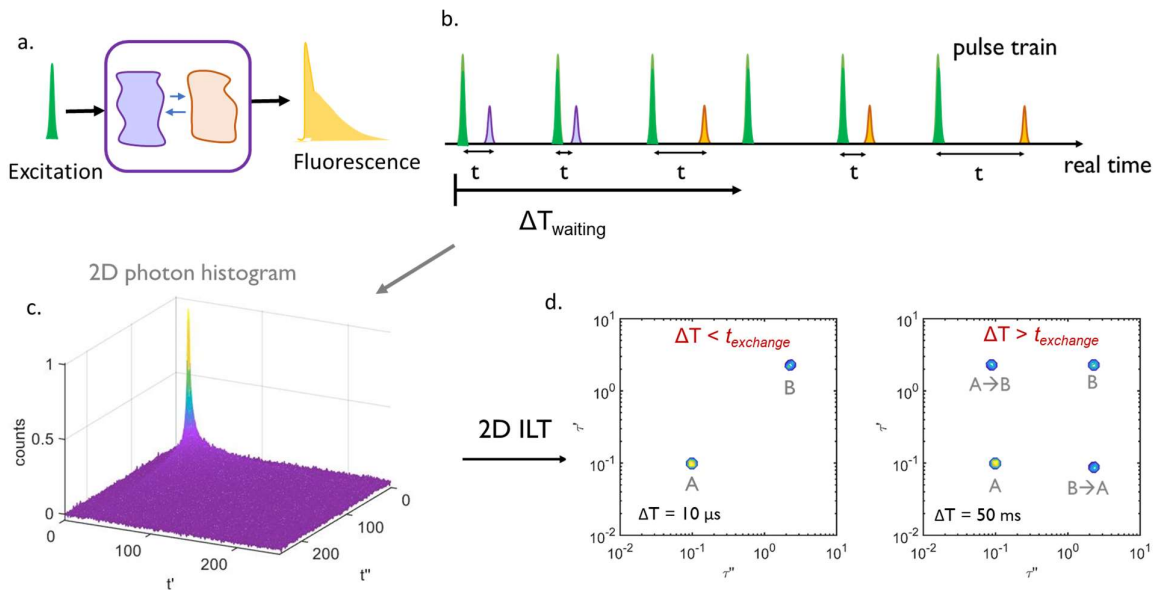


Figure 3.1. Schematic of 2D-FLCS analysis. a) excitation of a model 2-state system with a pulsed laser source results in a bi-exponential fluorescence decay. b) typical TCSPC measurement shows individual photon detection events following pulsed laser excitation. The emission delays between the excitation and emission are denoted by t . c) A 2D emission delay correlation map is constructed by measuring the delays between photon pairs separated by a preset time lag ΔT . d) The 2D Inverse Laplace Transform of the emission delay correlation map at given ΔT gives a 2D fluorescence lifetime correlation spectra where, individual species are represented by peaks along the diagonal and their chemical exchange is represented on the off-diagonal positions.

3.2.1 Generation of 2D emission delay correlation histograms

This technique can be used in a typical confocal fluorescence microscope equipped with TCSPC. The data acquisition is in the form of fluorescence time trace acquired in the TTTR mode. In the experiments performed for the work presented in this thesis, we have used TCSPC hardware made by PicoQuant. The photon streams files are stored with '.ptu' extension. We use an open-source demo code 'Read_PTU.m' to convert the files to .csv to be used in further analysis (20). The data stored is in the form of photon index, its global arrival time (MacroTime) and its emission delay time after excitation (MicroTime).

The photon stream is then partitioned to generate a 2D emission delay histogram at a given ΔT in the protocol described as follows:

- a. For a given ΔT , define a window of size $\Delta\Delta T$ such that its upper bound is $\Delta T + \frac{\Delta\Delta T}{2}$ and the lower bound is $\Delta T - \frac{\Delta\Delta T}{2}$.
- b. Initialize a 2D empty array (M) of size ($N \times N$), where N is the number of TCSPC channels.
- c. For the first photon in the photon trace, store its microtime (as x).
- d. Find all the photons in the trace whose Macrotime lies within the constructed window.

$$\Delta T - \frac{\Delta\Delta T}{2} < T_M < \Delta T + \frac{\Delta\Delta T}{2}$$
- e. Store all the recorded Microtimes as an array.
- f. In the 2D array, in the column denoted by x , increment the elements by 1 for each occurrence of their corresponding microtime in array.
- g. Repeat steps [c] to [f] for all the photons in the trace.
- h. Repeat steps [a] to [g] for all ΔT windows.

3.2.2. Two-Dimensional Inverse Laplace Transform

The 2D emission delay histograms can be represented in the form:

$$M(t_{1,i}, t_{2,j}, \Delta T) = \sum_{kl} F(\tau_{1,k}, \tau_{2,l}, \Delta T) \cdot \exp(-t_{1,i}/\tau_{1,k}) \cdot \exp(-t_{2,j}/\tau_{2,l}) \quad 3.3$$

Using the kernel structure mentioned in Eq.3.2 above, we can write in matrix notation,

$$M(t_1, t_2, \Delta T) = K_1(t_1, \tau_1) F(\tau_1, \tau_2, \Delta T) K_2(t_2, \tau_2)^T \quad 3.4$$

where $F(\tau_1, \tau_2, \Delta T)$ is the joint probability distribution of occurrence of a photon with lifetime τ_1 and τ_2 occurring ΔT after registering a photon with lifetime τ_1 . Each point on $F(\tau_1, \tau_2, \Delta T)$ represents the amplitude of correlation between the components at (τ_1, τ_2) separated by lag time ΔT , the autocorrelations appear along the diagonal of $F(\tau_1 = \tau_2)$ and the cross correlations appear as off-diagonal peaks of $F(\tau_1 \neq \tau_2)$. By varying ΔT , one can determine the separated time correlation functions of the constituent components in any fluorescence emission time series.

Determining $F(\tau_1, \tau_2, \Delta T)$ is equivalent to performing a 2D-ILT on M . For efficient handling, we can convert the 2D form of **Eq.4** to 1D by lexicographically ordering the matrices $m = \text{vec}[M]$, $f = \text{vec}[F]$ and can write the kernel operations K_1 and K_2 by a single operator K_0 given by the Kronecker product.

$$K_0 = K_1 \otimes K_2 \quad 3.5$$

And we represent the equivalent 1D problem as:

$$m = K_0 f \quad 3.6$$

Now, the only task at hand is to solve for f given m and K_0 . As previously mentioned, this equation cannot be solved analytically since it is an ill-conditioned inversion resulting in a solution that is not stable or unique. Instead, regularized least-squared technique must

be used to approximate the inversion. In regularized least square minimization, the least square minimizes the difference between the input data and the fit, while the regularization imposes a penalty on undesired features of the fitted solution. The regularization prevents overfitting or numerical instability. To avoid any bias, we start with a uniform prior guess of f and iteratively find the solution by minimizing the objective function given by

$$Q(f) = \|K_0 f - m\|^2 + \alpha R(f) \quad 3.7$$

where the first term on RHS represents the least square term, $R(f)$ is the regularization function of choice, and α is a regularization constant that weighs the importance of the least square fitting vs regularization. The choice of α is critical to appropriate fitting; too small and the minimization remains unstable (overfitting), too large and fit may not reflect the underlying experimental data (underfitting). Maximum Entropy Method (MEM) has been the traditional method of choice in ILT of fluorescence decays (as discussed in section 2.4.1) as well as for the recently developed 2D-ILT techniques (8, 9, 11). However, for typical 2D-FLCS spectra the data sets are too large to compute ILT using MEM efficiently. Instead, the data is binned logarithmically, which decreases the data size at the cost of resolution. Furthermore, to obtain a meaningful ILT using MEM, one must impose a non-negativity constraint on f after every iteration, since the entropy cannot be defined for negative values. This significantly decreases the efficiency of the calculation.

Tikhonov regularization (TK) is an alternative regularization technique which, although does not give the solution with maximum entropy, is sufficient to perform numerical ILT fitting for fluorescence data. The greatest advantage however is that it enables faster and unconstrained optimization techniques which drastically improve the calculation speeds. Here we highlight the strength of Tikhonov regularization at providing a fast and robust method for performing 2D-ILT.

3.2.3 2D-ILT by Tikhonov Regularization

3.2.3.1 Regularization Function.

An alternative approach has been developed in NMR spectroscopy which uses singular-valued decomposition (SVD) based data compression and Tikhonov regularization (14-16). This approach differs from MEM in several ways. First, the size of the numerical calculation is greatly reduced by compressing the kernels using SVD instead of non-uniform binning of the data. Since the kernels are smooth functions, the elementwise data is vastly redundant and SVD can reduce the data size by roughly a hundred-fold without compromising the quality of the fit or the spectral resolution. Second, it uses Tikhonov regularization on the compressed data, which ensures a unique solution to the optimization due to the quadratic nature of the terms. Third, it employs Butler-Reeds-Dawson (BRD) method to transform the constrained optimization to an unconstrained optimization which is computationally efficient to implement. Lastly, it provides a method for choosing an appropriate regularization constant proportional to variance in the data. The outline of the method follows what is presented in references (14-16). Using Tikhonov regularization, the objective function to be minimized from, Eq.3.7 becomes

$$\hat{f} = \arg \min_{f > 0} Q(f) \quad 3.8$$

Where,

$$Q(f) = \|\tilde{m} - \tilde{K}_0 f\|^2 + \alpha \|f\|^2 \quad 3.9$$

The quadratic nature of the terms ensures a unique solution (14). The second term penalizes solutions that have large norms, characteristic of functions with sharp features. Solutions with smoothly varying features are therefore promoted, and α scales the desired smoothness with respect to the least squares difference. However, since f is the probability distribution, it cannot take negative values which makes this a constrained optimization problem. To converge

rt the inversion to an unconstrained problem the Butler-Reeds-Dawson (BRD) algorithm is employed (14). We do not anticipate significant speed improvement by implementing Tikhonov regularizer alone. Instead, its purpose is to enable the use of BRD algorithm which is incompatible with maximum entropy regularizer.

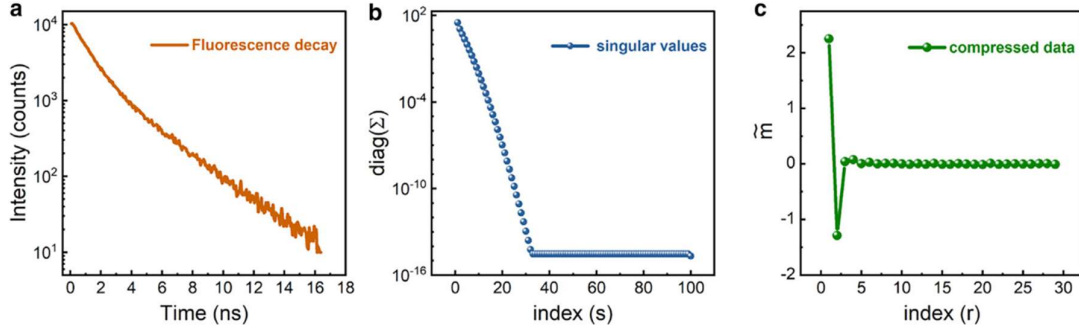


Figure 3.2 (a) An example fluorescence decay data exhibiting a biexponential decay. (b) Singular values of the kernel matrix K in Eq.3.9 comprising of exponential decays with 100 distinct lifetimes in range (0.1 ns–10 ns). The singular values decrease rapidly, and K can be represented accurately by 30 highest singular values. (c) Compressed data $\sim m$ obtained by Eq.3.19 represented using 30 point.

3.2.3.2. Kernel Compression.

Here we demonstrate efficient reduction of computational size of the problem in Eq.3.7 without significant loss of temporal information of the data. To get a sense of magnitude of the computational size, TCSPC experiment performed at the resolution of 64 picoseconds, has data 256 data points **Fig. 3.2.a**. The 2D emission delay histogram generated from the time trace are thus of size 256x256, if we desire to perform 2D ILT with a basis size of 100 lifetime constants, we have the estimated spectrum of the size 100x100 and hence the kernels of size 256x100. Converting the problem to 1D by lexicographic ordering Eq.3.6 gives us $m_{65536 \times 1} = K_{65536 \times 10000} f_{10000 \times 1}$,

Which is a tremendously large problem. Traditional solutions include non-uniform binning of the data and using smaller basis of lifetimes. However, maximum storage space is occupied by the kernel which is precalculated and free of noise.

Here we show how the size of the problem can be reduced by compressing the kernel using singular valued decomposition as demonstrated in (14). For simplicity we illustrate the implementation for 1D ILT Eq.3.2. For the implementation in the case of 2D ILT, the compression is performed before converting the problem to 1D as follows:

The 2D ILT problem at hand is to solve for F were,

$$M = K_1 F K_2^T \quad 3.10$$

M is the 2D emission delay histogram of photon pairs, K_1 and K_2 are the pre-calculated kernels, and F is the 2D lifetime correlation spectrum. We represent the kernels K_1 and K_2 by their singular valued decomposition (SVD), $K_1 = U_1 \Sigma_1 V_1^T$ and $K_2 = U_2 \Sigma_2 V_2^T$, where U and V are unitary matrices and Σ is a diagonal matrix with the singular values of K arranged in descending order.

$$M = U_1 \Sigma_1 V_1^T F V_2 \Sigma_2^T U_2^T \quad 3.11$$

Figure 3.3 shows the singular values of a kernel formed using basis of 100 decays within range of lifetimes [0.1 to 10 ns] and for time [0 to 16 ns]. We can represent the kernel matrix by a compact SVD where we only keep the r singular values greater than 10^{-15} (first 30 values in this case: $\sigma_i < 10^{-15}; i > 30$). The resultant compressed kernels $\tilde{U}_1 \tilde{\Sigma}_1 \tilde{V}_1^T$ is an approximation of the original kernel and has the same dimensions as K_1 and K_2 . The difference between the original and the compressed kernel is shown in **Figure 3.3.d**. We can observe that there is negligible loss of information and that \tilde{K} approximates K well.

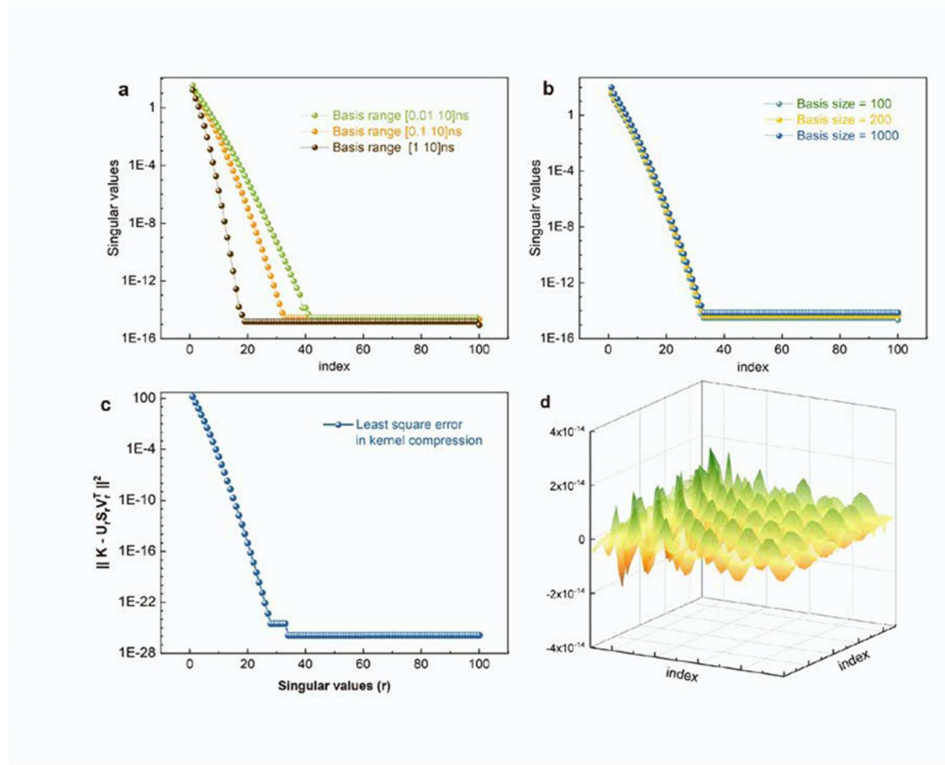


Figure 3.3. Kernel compression for data analysis. (a) The distribution of singular values of the kernels with size 100 and various ranges. (b) The distribution of singular values of the kernels constructed using basis in the range [0.1, 10ns] using various sizes. (c) Least square error in the compression where the kernels are constructed using the number of singular values on the x-axis. (d) The difference between the primary kernel and the compressed kernel constructed using 30 highest singular values.

Using the compact SVD, Eq.3.10 now becomes

$$M = \tilde{U}_1 \tilde{\Sigma}_1 \tilde{V}_1^T F \tilde{V}_2 \tilde{\Sigma}_2^T \tilde{U}_2^T \quad 3.12$$

Left multiply both sides by \tilde{U}_1^T and right multiply both sides by \tilde{U}_2

$$\tilde{U}_1^T M \tilde{U}_2 = \tilde{U}_1^T \tilde{U}_1 \tilde{\Sigma}_1 \tilde{V}_1^T F \tilde{V}_2 \tilde{\Sigma}_2^T \tilde{U}_2^T \tilde{U}_2 \quad 3.13$$

Since \tilde{U}_1 and \tilde{U}_2 are unitary matrices, $\tilde{U}_1^T \tilde{U}_1 = I$ and $\tilde{U}_2^T \tilde{U}_2 = I$

$$\tilde{U}_1^T M \tilde{U}_2 = \tilde{\Sigma}_1 \tilde{V}_1^T F \tilde{V}_2 \tilde{\Sigma}_2^T \quad 3.14$$

$$\tilde{M} = \tilde{K}_1 F \tilde{K}_2^T \quad 3.15$$

Where $\tilde{K}_1 = \tilde{\Sigma}_1 \tilde{V}_1^T$, $\tilde{K}_2 = \tilde{\Sigma}_2 \tilde{V}_2^T$ and $\tilde{M} = \tilde{U}_1^T M \tilde{U}_2$ is referred to as compressed data. The projection of the original data on the range space is given by $\tilde{\tilde{M}} = \tilde{U}_1 \tilde{U}_1^T M \tilde{U}_2 \tilde{U}_2^T = \tilde{U}_1^T \tilde{M} \tilde{U}_2$. $\tilde{\tilde{M}}$ looks like a smoother version or a low pass filtered version of original data M . The reduced 2D problem Eq.3.15 is converted to 1D problem $\tilde{m} = \text{vec}[\tilde{M}]$ and $\tilde{K}_0 = \tilde{K}_1 \otimes \tilde{K}_2$ and

$$\tilde{m} = \tilde{K}_0 f \quad 3.16$$

Comparing the size of the problem in Eq.3.12 to that of Eq.3.16, we observe that the problem size is reduced by a factor of ~ 100 . We can automatically determine r using a lower bound for the singular values which in practice is found to be at $1E-9$ (~ 20 singular values). This reduces the problem size by a factor of ~ 250 without significant loss of information.

Now, the 1D problem at hand is,

$$m_{d \times 1} = K_{d \times s} f_{s \times 1} \quad 3.17$$

where the subscripts denote the dimensions ($d \sim 256, s = 100$). We represent the kernel K by its singular valued decomposition (SVD), $K = U \Sigma V^T$, where U and V are unitary matrices ($U^T U = V^T V = I$) and Σ is a diagonal matrix with the singular values of K arranged in descending order.

$$m_{d \times 1} = U_{d \times s} \Sigma_{s \times s} V_{s \times s}^T f_{s \times 1} \quad 3.18$$

Figure 3.2.b shows the singular values of a kernel formed using basis of 100 decays within range of lifetimes [0.1 to 10 ns] and for time [0 to 16 ns]. We see that the singular values become negligible after approximately 30 values ($\sigma_i < 10^{-15}$; $i > 30$). This is attributed to the smooth nature of the kernels, which implies that we can represent the kernel matrix by a compact SVD where we only keep the r highest singular values greater than 10^{-15} (first 30

values in this case). Thus, the matrix $\Sigma_{s \times s}$ can be compressed to $\tilde{\Sigma}_{r \times r}$ without significant loss of information. $\tilde{U}_{d \times r}$ and $\tilde{V}_{s \times r}$ are formed by selecting the first r columns of U and first r rows of V . The resultant compact kernel $\tilde{U}_{d \times r} \tilde{\Sigma}_{r \times r} \tilde{V}_{r \times s}^T$ is an approximation of the original kernel and has the same dimensions as $K_{d \times s}$. The difference between the original and the compact kernel is shown in **Figure 3.4**. We can observe that there is negligible loss of information and that $\tilde{U}_{d \times r} \tilde{\Sigma}_{r \times r} \tilde{V}_{r \times s}^T$ approximates K well.

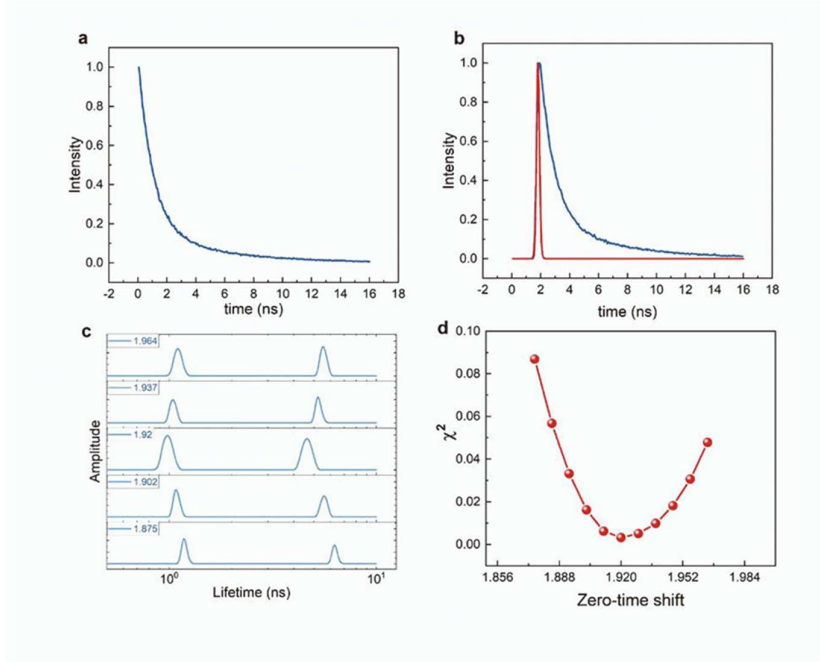


Figure 3.4. Monte Carlo simulation for data with IRF included. (a) The 1D emission delay histogram. (b) Emission delay histogram (black) convoluted by the Instrument response function (red). (c) The 1D ILT spectra at various zero-time shifts of the IRF. (d) Evaluated least-square error in the fit and convoluted histogram. The ILT spectra at the minima of the plot in (d) is selected as the 1D lifetime spectra.

Using the compact SVD, Eq.3.12 now becomes

$$m_{d \times 1} = \tilde{U}_{d \times r} \tilde{\Sigma}_{r \times r} \tilde{V}_{r \times s}^T f_{s \times 1} \quad 3.19$$

$$\tilde{U}_{r \times d}^T m_{d \times 1} = \tilde{\Sigma}_{r \times r} \tilde{V}_{r \times s}^T f_{s \times 1} \quad 3.20$$

$$\tilde{m}_{r \times 1} = \tilde{K}_{r \times s} f_{s \times 1} \quad 3.21$$

Where $\tilde{m} = \tilde{U}^T m$ is referred to as compressed data and $\tilde{K} = \tilde{\Sigma} \tilde{V}^T$. Note that the size of the problem is reduced using the SVD of kernels and not the experimental data m . The fitting is now performed on \tilde{m} that has dimensions of $r \times 1$ **Figure 1c**. For the typical case of $r = 30$,

$$\tilde{m}_{30 \times 1} = \tilde{K}_{30 \times 100} f_{100 \times 1} \quad 3.22$$

This makes it different than the traditional non-uniform binning approach. Venkatramanan et al. have shown that the structure of the problem Eq.3.7 remains unchanged after this transformation (14). i.e., the solution f that satisfies Eq.3.17 is the same one that satisfies Eq.3.22 and again, the solution is unique due to the quadratic nature of the problem.

Comparing the size of the problem in Eq.3.17 to that of Eq.3.22, we observe that the problem size is reduced by a factor of 10. The reduction factor is even greater in the case of 2D problem (250 times as shown above Eq.3.16). This compression is a general technique and can be used alongside the existing MEM approach (**Table 3.1, 3.2**) or broadly, with any other problem involving smooth large kernel functions. The projection of the original data on the range space is given by.

$$\tilde{\tilde{m}} = \tilde{U} \tilde{U}^T m = \tilde{U} \tilde{m} \quad 3.23$$

where \tilde{m} resembles a smoother version or a low pass filtered version of original data m **Fig**

Figure 3.5.a.

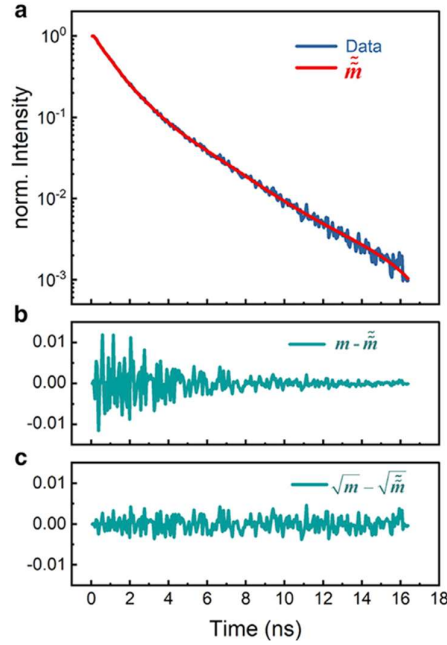


Figure 3.5 Noise estimation for stopping criterion. (a) Normalized fluorescence data superimposed with \tilde{m} , projection of compressed data onto the range space of kernel given by Eq.3.23. It is observed that \tilde{m} serves as a smooth approximation to input data referring to which we infer the corresponding noise level given by Eq.3.35. (b) Shows the difference $\tilde{m} - m$, which is dependent on the photon counts. We normalize both traces using square root of photon counts which eliminates the intensity dependence. (c) Shows the difference after normalization given by $\sqrt{\tilde{m}} - \sqrt{m}$.

3.2.3.3. Tikhonov regularization with BRD.

A brief overview of the BRD approach (14, 21) summarizing is given below. In Eq.3.9, the gradient of Q with respect to f must be zero at its minimum,

$$\frac{\partial Q}{\partial f_i} = -\tilde{K}_{0,i}^T(\tilde{m} - \tilde{K}_0 f) + \alpha f_i = 0 \quad 3.24$$

Where,

$$\alpha f_i = \tilde{K}_{0,i}^T(\tilde{m} - \tilde{K}_0 f) \quad 3.25$$

and,

$$f_i = \tilde{K}_{0,i}^T c \quad 3.26$$

We invoke a vector c which maps f as,

$$c = \frac{\tilde{m} - \tilde{K}_0 f}{\alpha} \quad 3.27$$

Where c has the same dimensions as the compressed data \tilde{m} . The unconstrained problem then becomes (16)

$$\hat{c} = \arg \min \chi(c) \quad 3.28$$

Where, $\arg \min f(x) := \{x \in S: f(s) \geq f(x) \text{ for all } s \in S\}$,

$$\chi(c) = \frac{1}{2} c^T [G(c) + \alpha I] c - c^T \tilde{m} \quad 3.29$$

and

$$G(c) = \tilde{K}_0 \text{diag} \left(H(\tilde{K}_0^T c) \right) \tilde{K}_0^T \quad 3.30$$

Here, $H(\cdot)$ denotes the Heaviside function, which ensures positive semi-definiteness. The minimization of $\chi(c)$ can be carried out via standard unconstrained inverse Newton minimization routines such as *fminunc* in MATLAB. The required gradients and Hessian of Eq.3.28 are easily computed as

$$\nabla \chi(c) = (G(c) + \alpha I) c - \tilde{m} \quad 3.31$$

and,

$$\nabla \nabla \chi(c) = G(c) + \alpha I \quad 3.32$$

Using the optimized vector c , the ordered vector f is calculated as

$$f = \max(0, \tilde{K}_0^T c) \quad 3.33$$

The 2D-ILT spectrum F is then given by reshaping the $f_{1 \times kl}$ vector to a matrix $F_{k \times l}$. We start with a large value of α to estimate c . The recommended optimum α for the following iter

ations is given by,

$$\alpha_{opt} = \frac{\sqrt{k \times l} \cdot \hat{\sigma}}{\|c\|_F} \quad 3.34$$

With,

$$\hat{\sigma} = std(\sqrt{\tilde{m}} - \sqrt{m}) \quad 3.35$$

Eq.3.35 represents the approximation of noise level in the data obtained by first normalizing the data by square root of the photon counts and then comparing it with a smoothed version of itself (**Figure 3.5.a**). The optimized vector c is iteratively evaluated using this α_{opt} . A suitable stopping criterion is achieved when.

$$\|c\|_F > 1/\hat{\sigma} \quad 3.36$$

or when the relative difference between consecutive α to be less than 0.1%.

$$\frac{|\alpha_i - \alpha_{i+1}|}{\alpha_i} < 10^{-3} \quad 3.37$$

Stopping criteria in eqn. 3.36 is based on the comparing the cost function with the noise present in the data to avoid overfitting. Stopping criteria in eqn.3.37 is based on converging values of the regularization constant. Typically, convergence is achieved within 10 iterations for both criteria and the iterations are stopped when either of the above-mentioned criteria are met.

Algorithm.

The 2D-ILT algorithm can thus be summarized in the following step-by-step procedure.

- a. Define a basis with an upper bound lower bound and number of points.
- b. Import data and IRF.
- c. Generate a kernel matrix (whose size is length of histogram x ILT basis size).

$$K_{ij} = \exp(-t_i/\tau_j)$$

3.38

- d. Convolute the kernel with IRF.
e. Compute the SVD of the kernel and stored the compressed U, S, V matrices

the compressed kernels are $\tilde{K}_1 = \tilde{\Sigma}_1 \tilde{V}_1^T$ and $\tilde{K}_2 = \tilde{\Sigma}_2 \tilde{V}_2^T$.

Compressed data is $\tilde{m} = \tilde{U}_1^T m \tilde{U}_2$ and $\tilde{\tilde{m}} = \tilde{U}_1 \tilde{U}_1^T m \tilde{U}_2 \tilde{U}_2^T = \tilde{U}_1 \tilde{m} \tilde{U}_2^T$

Generate the Kronecker product matrix $\tilde{K}_0 = \tilde{K}_1 \otimes \tilde{K}_2$.

- f. Unconstrained optimization: initialize a constant array (c), start with initial.

$$\chi(c) = \frac{1}{2} c^T [G(c) + \alpha I] c - c^T \tilde{m} \quad 3.39$$

$$\nabla \chi(c) = (G(c) + \alpha I) c - \tilde{m} \quad 3.40$$

$$\nabla \nabla \chi(c) = G(c) + \alpha I \quad 3.41$$

Where $G(c) = \tilde{K}_0 \text{diag} \left(H \left(\tilde{K}_0^T c \right) \right) \tilde{K}_0^T$ and H is the Heaviside function.

- g. $\alpha_{opt} = \frac{l \hat{\sigma}}{|c|}$, where l is the size of the basis, and $\hat{\sigma} = \text{std} \left(\sqrt{\tilde{\tilde{m}}} - \sqrt{m} \right)$.

- h. Check convergence condition: $\left| \frac{\alpha_i - \alpha_{i+1}}{\alpha_i} \right| < 10^{-3}$

- i. $F_{ILT} = \max(0, K^T c)$

3.2.4 Monte Carlo Simulations

We demonstrate and compare the performance of both MEM and Tikhonov regularization using Markovian Monte Carlo simulations to generate artificial photon time series for a two-state system with user defined inputs for fluorescence lifetime, emission intensity and transition rate matrix. The Monte Carlo simulation is a 4-step process: Initialize, Reaction, Photon emission and Detection.

Initialization: This involves setting up the system and defining the parameters for the simulation.

ion. An example initialization parameters setup is shown in the following table.

Parameter	Value	Comment
Simulation time step	25 ns	1/Laser rep. rate (40MHz)
No. of TCSPC channels	256	Bin Width = 64 picoseconds
Number of conformations	2	
Brightness	$[10^4, 10^4]$	photons/second
Fluoresce lifetimes	[1, 3]	nanoseconds
Transition probability matrix	$\begin{bmatrix} 0 & 10^{-4} \\ 10^{-4} & 0 \end{bmatrix}$	
Simulation size	50 seconds	2×10^9 time steps

Table 3.1 Monte Carlo simulation parameters

Reaction: The chemical reaction is governed by transition rate matrix. The initial states are randomly chosen based on the steady state probabilities. In case of a 2-state simulation, $P_1 = K_{2 \rightarrow 1} / (K_{1 \rightarrow 2} + K_{2 \rightarrow 1})$ and $P_2 = K_{1 \rightarrow 2} / (K_{1 \rightarrow 2} + K_{2 \rightarrow 1})$. The initial state is stored, and the probability of transition is calculated as $P_{i \rightarrow j} = K_{i \rightarrow j} \cdot \delta t$ where $K_{i \rightarrow j}$ is the rate of transition from state i to j and δt is the time step. A uniform random number R_r is generated between (0,1) ('rand()' function in MATLAB) and compared with $P_{i \rightarrow j}$. The reaction succeeds if $R_r < P_{i \rightarrow j}$ and the state is updated from i to j . Otherwise, the molecule remains in i state. The updated state is stored, and the process is repeated for the length of simulation using the new initial state as the updated state. The result is an array which shows the trajectory of chemical reaction of the molecule.

Photon emission: Each state is assigned a brightness property B_i in units of *photons/second* and

a characteristic fluorescence lifetime τ_i . For every point along the reaction trajectory generated above, the probability of emission is given by the $P_{em} = B_i \cdot \delta t$. P_{em} is compared with another uniform random number R_{em} between (0,1) and photon emission is registered during the time step if $R_{em} < P_{em}$. This results in a photon emission trajectory. For every emitted photon, an emission delay time is generated from using an exponential random number generator ('*exprnd(mu)*' in MATLAB) with the lifetime equal to the lifetime of current state. The result is an emission delay trajectory corresponding to the photon emission trajectory.

Detection: The photon emission trajectory is called the MacroTime (T) and the emission delay trajectory is the MicroTime (t). The microtime is discretized based on the expected TCSPC parameters (channel width = 64 picoseconds and 256 bins. Photon data beyond the bin limit (16 ns = 256 x 64ps) is discarded. The final resultant photon stream is stored in the form of an array with 2 columns – MacroTime and MicroTime (T_i, t_i)

3.3 Result and Discussion

To demonstrate the workings of the proposed algorithm and to compare it with traditional MEM, we generate a simulated photon stream using Markov chain Monte carlo simulations as described above (section 3.2.4). We compare the compression efficiency, tolerance to noise, and the timing of the algorithm for various relevant parameters.

3.3.1 Implementation.

We simulate a two-state system with fluorescence lifetimes of $\tau_1 = 1.0 \text{ ns}$ and $\tau_2 = 3.0 \text{ ns}$, and exchange rates of $k_f = k_r = 1 \times 10^3 \text{ s}^{-1}$ corresponding to an exchange time of 1 ms. The number of photons analyzed is roughly 2.5×10^5 . The corresponding 1D and 2D lifetime spectra were obtained by ILT by Tikhonov regularization as described above, using a basis set of size $L = 100$ and with a singular valued decomposition using 30 greatest values. The 1D photon histogram, resulting fit computed by a 1D-ILT, and residual are shown in **Figure 3.6.b, c**. The 1D relaxation rate spectrum shows two peaks located at $\tau_1 = 1.0 \text{ ns}$ and $\tau_2 = 3.0 \text{ ns}$,

consistent with the input parameters of the simulation **Figure 3.6.d**.

The 2D photon correlation histogram evaluated at $\Delta T = 2 \text{ ms}$, resulting fit computed by a 2D-ILT, and residual are shown in **Figure 3.6.e, f**. A 2D-FLCS spectrum shows two diagonal peaks at $\tau_1 = 1.0 \text{ ns}$ and $\tau_2 = 3.0 \text{ ns}$, and two off-diagonal cross-peaks between these transitions **Figure 3.6.g**. The cross-peaks indicate that chemical exchange has occurred within the timescale $\Delta T = 2 \text{ ms}$, which is consistent with the input transition rate matrix.

Computing the 2D-FLCS spectrum for a series of ΔT delay times allows the chemical exchange kinetics to be directly measured. **Figure 3.7.a-c** shows the 2D-FLCS spectrum measured for delay times ranging from $10 \mu\text{s}$ to 2 s . At early times $\Delta T < \tau_{\text{exchange}}$, only diagonal peaks are observed as no chemical exchange had occurred **Figure 3.7.a**. As ΔT is increased, cross-peaks emerge as chemical exchange occurs **Figure 3.7.b**. The exchange cross-peaks reach a steady state as $\Delta T > \tau_{\text{exchange}}$ **Figure 3.7.c**. Monitoring the amplitude of these diagonal and cross-peaks provides direct information regarding the chemical exchange process. The amplitude of the peaks can be expressed as,

$$C_{ij}(\Delta T) = \langle S_i(t)S_j(t + \Delta T) \rangle \quad 3.42$$

Where i and j represent the state of the system with lifetimes τ_i and τ_j . $C_{ij}(\Delta T)$ thus represents the probability of the system in state i at time t to be observed in state j at later time $t + \Delta T$. The diagonal is given by the condition $i = j$, while cross-peaks are given by the condition $i \neq j$. The two autocorrelation and cross-correlation functions obtained from the 2D-FLCS spectra are shown in **Figure 3.7.d**. As expected, the auto-correlation functions decay on the timescale of τ_{exchange} , while the cross-correlation functions grow on the timescale τ_{exchange} . Fitting the data provides a direct measure of τ_{exchange} for the system.

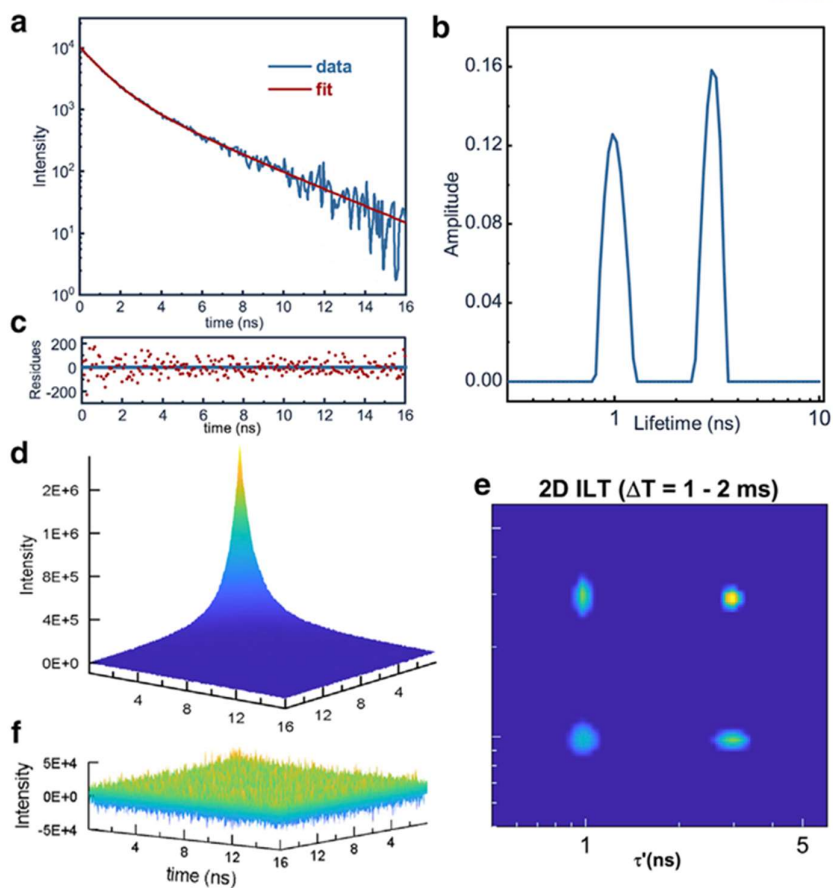


Figure 3.6. Laplace inversions by Tikhonov Regularization. Monte Carlo simulations generate an artificial photon stream of a two-component system with fluorescence lifetimes of $\tau_1 = 1$ ns and $\tau_2 = 3$ ns undergoing equilibrium chemical exchange at a rate of $1 \times 10^3 \text{ s}^{-1}$. (a) Raw 1D-photon histogram (blue) and the obtained fit (red) from 1D-ILT. (b) The 1D lifetime spectrum shows two peaks with relaxation rates of 1 and 3 ns, consistent with fluorescence lifetime of the two components in the system. (c) Residual between the 1D photon histogram and the fit obtained by 1D-ILT. (d) 2D photon correlation histogram and (e) 2D relaxation rate spectrum computed at a waiting time of $\Delta T = 2$ ms obtained from a 2D-ILT. The spectrum consists of diagonal peaks at 1 and 3 ns, as well as cross-peaks indicating chemical exchange between the two species. (f) Residual between the 2D photon correlation histogram and 2D-ILT fit.

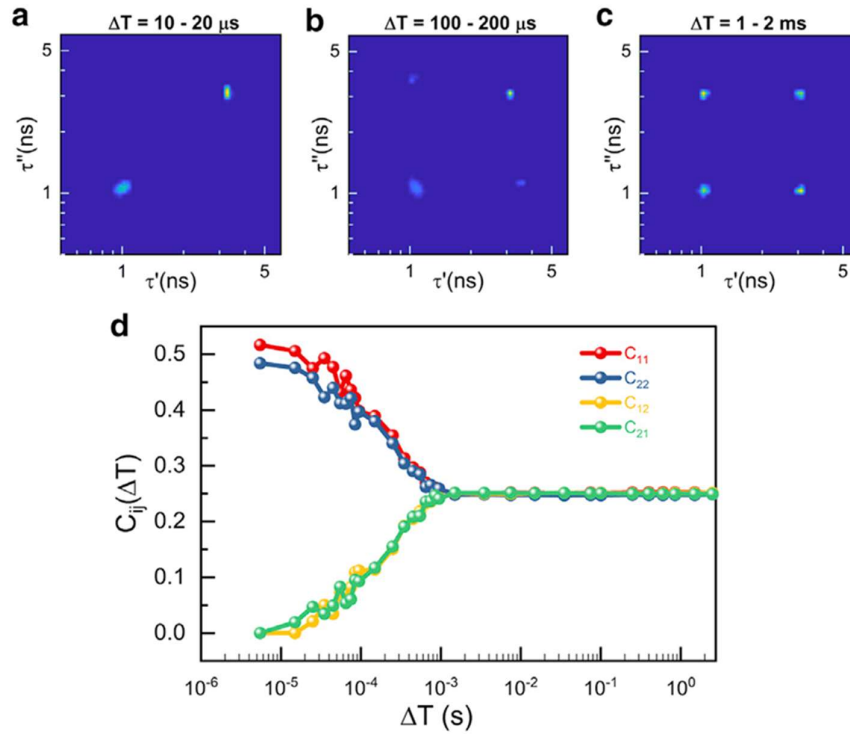


Figure 3.7 Chemical Exchange kinetics of a two-state system. 2D-FLCS spectra computed at (a) $\Delta T = 10 - 20 \mu s$, (b) $\Delta T = 100 - 200 \mu s$, and (c) $\Delta T = 1 - 2 ms$. At $\Delta T < \tau_{exchange}$, no cross-peaks are observed between the two states, indicating no exchange has occurred. As ΔT approaches $\tau_{exchange}$, cross-peaks emerge indicating both forward and reverse transitions between the two states. (d) The kinetics of the exchange process are reflected in the autocorrelation and cross-correlation functions of the diagonal and off-diagonal peaks, respectively. Fitting either the autocorrelation or cross-correlation functions provide a direct measure of the exchange kinetics in the system.

In practice, lifetime histograms are convoluted by a systematic instrument response function IRF and do not manifest as pure exponential decays. We can account for this by fitting the data with kernels convoluted by a known IRF which can be measured experimentally from scattering of the excitation pulse from a colloidal medium. The experimentally observed T CSPC histogram is modelled from Eq.3.1 as

$$I^{obs}(t, \tau) = \sum_i \text{IRF}(t_i - t_0) \exp\left(-\frac{t - t_i}{\tau}\right) \quad 3.43$$

where, t_0 is the unavoidable zero-time shift for the IRF. It is important so estimate the correct value of t_0 as it can lead to undesirable errors in the ILT spectrum. To circumvent this, we calculate the 1D-ILT spectrum at various values of zero-time shifts and use the range over which the fitting least square error χ^2 is minimum. The ILT obtained from this range is then averaged to obtain the final 1D-ILT spectrum to be used for further analysis. 2D spectrum is calculated using the same range of zero-time-shifts. The kernels to be fitted are also convoluted with the IRF similarly **Figure 3.4**.

3.3.2. Effect of compression:

Kernel compression increases the computational speed by excluding redundant or insignificant information and thus reducing the size of the problem. However, there is a tradeoff between the magnitude of compression and loss of information. The measure of compression in our case is given by the number of singular values chosen to represent the kernel of exponential. **Figure 3.2.b** shows the singular values for a kernel comprising of 100 basis exponentials ranging from lifetimes [0.1ns to 10ns]. It is observed that only the greatest 30 singular values are enough to sufficiently describe the kernel. This is demonstrated in **Figure 3.3.d** showing the difference between the true kernel and the approximated kernel constructed by choosing only the 30 greatest singular values and **Figure 3.3.c** showing the least square error (lsq) in the kernel approximation $lsq = \Sigma(K - K_{compact})^2$.

Finally, transforming the problem by operation of U^T , Eq.3.20 reduces the computational size by 10 times for 1D ILT and 250 times for 2D ILT. The number of significant singular values depends largely on the range of lifetimes used to generate the basis and not the number of basis vectors chosen for fitting (**Figures 3.3.a, b**).

3.3.3. Dependence on total photon counts:

A robust algorithm should produce consistent results for various realizations of data obtained from same source. In case of regularized optimization, this is reflected in the choice of the regularization parameter α Eq.3.7. α represents the bias towards the smoothness with respect to the least square errors which manifests as the width of the peaks in case of the inverse Laplace transform. We start with a sufficiently high value of α and gradually decrease it following Eq.3.34 as suggested in (21), where $\hat{\sigma}^2$ is a measure of the variance in data.

Determining $\hat{\sigma}$ is straightforward in experiments influenced by uniform gaussian noise however, for the fluorescence lifetime histogram, the photon counts follow Poisson distribution at each point individually. As a result, the variance at each point is different and is proportional to the total accumulated photons. The best case is to determine the point-wise variance experimentally however, the data used for determination can be simply added together into a fluorescence decay with more counts and hence less relative noise. Thus, we chose to approximate the point-wise variance by comparing the data with a smoothed version of itself. Luckily, $\tilde{\tilde{m}} = \tilde{U}^T \tilde{U} m$ which is the projection of compressed data to the range space of the kernels serves as an approximate smoothed curve corresponding to the data given that the total number of photons is sufficiently high (above $\sim 10^4$ total photons) **Figure 3.5.a**. Then we estimate $\hat{\sigma}$ using Eq.3.35, where the traces m and $\tilde{\tilde{m}}$ are normalized by their square root which effectively removes the intensity dependence of each data point and produces uniformly distributed residues **Figure 3.5.b, c**. Decrease in photon counts leads to higher $\hat{\sigma}$ which in turn leads to a higher α_{opt} and hence smoother fits. **Figure 3.8.a, b** shows dependence of $\hat{\sigma}$ and α_{opt} on the total photon counts. **Figure 3.8.c-f** also shows the normalized 1D and 2D ILT fits for trajectories with various total photon counts. A similar approach of using preliminary fits has been shown to improve the accuracy and precision of estimated parameters in fitting of single molecule fluorescence decays (22). A qualitative analysis of effect of noise on the resulting ILT fits is detailed as follows and demonstrated in **Figures S3, S4**.

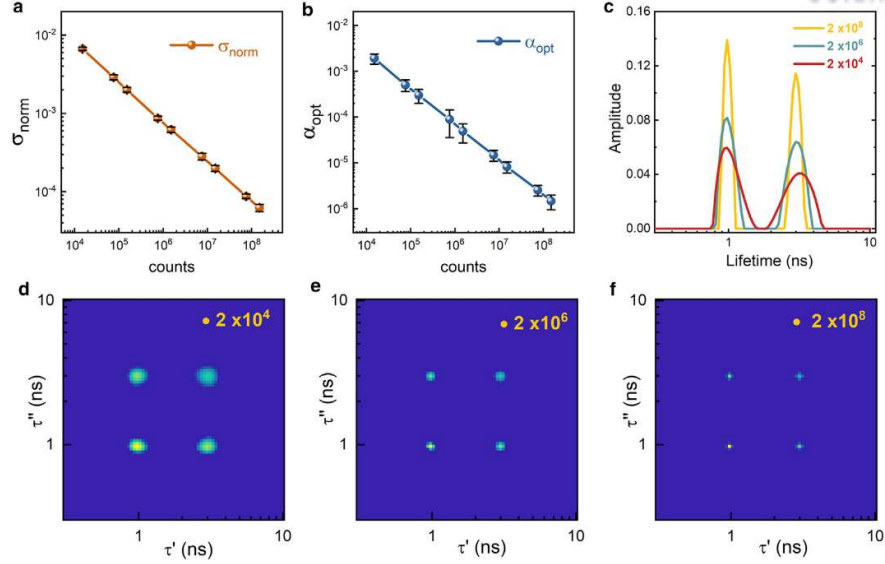


Figure 3.8 *Dependence of 2D spectrum on total photon counts.* (a) The estimated $\hat{\sigma}$ with respect to total photon counts in the fluorescence data. (b) Shows the optimum regularization parameter obtained with respect to the total photon counts in the fluorescence data. Error bars denote the standard deviation using 30 instances of simulated data. Higher α_{opt} due to lower photon counts results in smoother fits and vice versa as demonstrated in (c), which showing 1D-ILT of three trajectories with varying total photons. (d–f) show the normalized 2D-ILT obtained at $\Delta T = 2ms$ using the simulated trajectories from (c).

Examining the effect of total photon counts on the obtained ILT spectrum, it is observed from **Figure 3.8.c** that the peak width systematically decreases with increasing total photon counts and hence proportionally, decreasing the estimated noise given by $\hat{\sigma}$ (**Fig 3.8.a**). **Figure.3.9.a** shows the average peak widths with respect to total photon counts averaged from 30 instances of simulated data. Although it is tempting to characterize the obtained spectra, we must note that the spectra consist of a logarithmically spaced basis of lifetimes to which, the algorithm assigns the most probable amplitudes. It is not clear what the width of the peak would imply in the spectra composed of a logarithmically spaced points since, even if two peaks appear to have similar widths, they span vastly different range. We also noticed that the widths vary inversely with the size of the basis (**Figure.3.9.b**) which might be attributed to larger basis having more resolution or more degrees of freedom.

For cases of lower fluorescence photon counts, Lower signal to background ratios (SBR), might contribute to either artefacts or cause a systematic bias in the ILT spectrum. We tested the background dependence using additive gaussian noise to the simulated fluorescence decays. We added a second source in the simulation which emits photons at experimentally observed rate of 150 cps (counts per second) and is independent of the microtime. Keeping the background levels constant, we varied the brightness of the fluorophore to obtain fluorescence decays. The SBR is quantified as ratio of total fluorescence photons to the total background photons. We notice firstly that the spectrum systematically shifts to prefer the higher lifetimes and secondly, we start to observe a “tail” at highest lifetime in the spectra at lower SBR as shown in **Figure.3.10.a**. This is most likely because at longer microtime, a non-zero background is equivalent to having a component with very large lifetime.

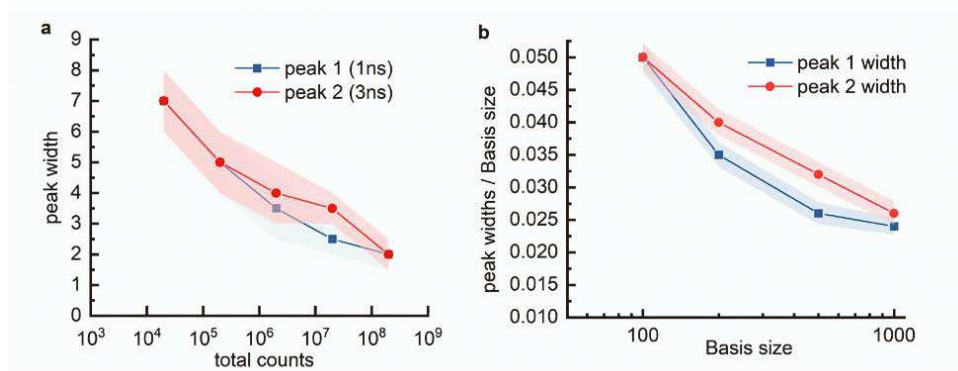


Figure 3.9 (a) shows the dependence of peak-widths of obtained 1D ILTs with respect to the total photon counts in the units of basis points. (b) shows the dependence of peak widths with respect to the size of the basis of lifetimes chosen for the 1D ILT

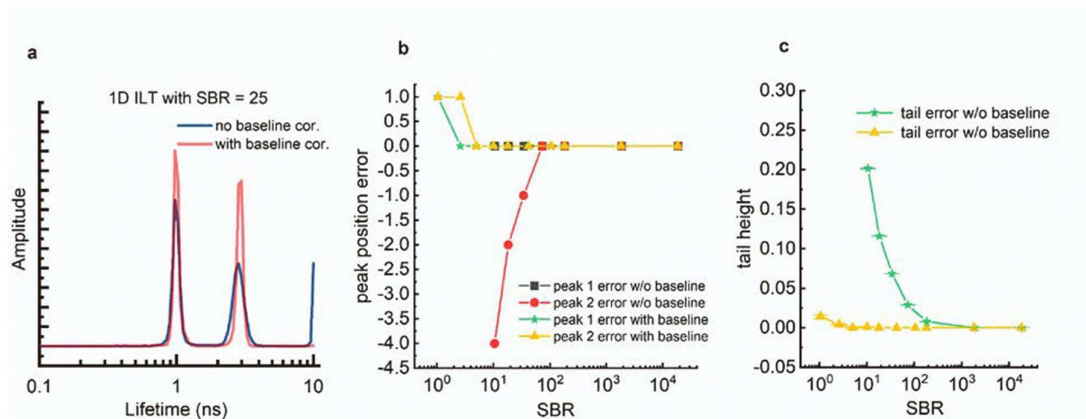


Figure 3.10. Baseline correction for background noise. (a) shows the effect of added background noise on the ILT as an observed “tail” shown in blue. On implementing baseline correction on the same data, this artefact is eliminated as shown in red. (b) at lower signal to background ratios, minor but systematic shifts are observed in the peak positions of obtained 1D-ILT which are also eliminated by applying baseline correction for the background noise. (c) The “tail” artefact is observed to have systematic dependence on the SBR as shown in green and is reduced on applying the baseline correction.

For improving the accuracy of the algorithm, one can parameterize the background and extend the current algorithm to fit for the background parameter. However, during the experiment, one can also measure the background separately by focusing on a dark spot in the sample and running a blank experiment. We found that applying an approximate correction in the form of baseline subtraction reduces the systematic errors. Since the detectors used have a systematic zero-time shift, one can approximate the points on the left-hand side of the IRF as background and use the mean of the counts of that segment of the trace as a baseline. **Figure.3.10.b** shows the errors in estimated peak positions with respect to the signal to background ratio. **Figure.3.10.c** shows the height of the “tail” originating due to higher SBR at low. We thus set a lower bound on the SBR till which we can obtain reliable ILT results which is $SBR=70$ ($1E4$ cps) without baseline correction and $SBR = 2.5$ (400 cps) with baseline correction where the brackets denote the brightness of fluorophore corresponding to the SBR for background fixed at 150 cps.

3.3.4. Application to experimental data.

In this section, we demonstrate the utility of the outlined 2D-ILT algorithm on experimental data. We measure 2D-FLCS for bacteriorhodopsin (bR), a trans-membrane proton pump protein with an endogenous retinal chromophore. Photon absorption by the retinal chromophore initiates a complex reaction cycle that involves a series of conformational transitions that occur on timescales ranging from μs to ms (23). The dynamics of conformational transitions involved in the biochemical cycle can be mapped through changes in the retinal chromophore fluorescence lifetime (24). Note, here we are interested only in demonstrating the use of the outlined approach to obtaining reliable 2D-FLCS spectra for an experimental system. An analysis of the dynamics of bR reaction cycle are reported elsewhere (24).

The crystal structure and absorption spectra of bR is shown in **Figure 3.11.a, b**. The experiments were carried out with 532 nm excitation. 2D-FLCS spectra computed at $\Delta T = 10\text{-}50 \mu\text{s}$ and $\Delta T = 10\text{-}50 \text{ms}$ are shown in **Figure 3.11.c, d** (single-molecule fluorescence traces contained $\sim 5 \times 10^6$ photons). At $\Delta T = 10\text{-}50 \mu\text{s}$, the spectrum contains three diagonal peaks, which can be assigned to the K, L, and N intermediates of the reaction cycle (**Figure 3.11.c**) (23-25). Cross-peaks are observed between the K and L intermediates indicating exchange, consistent with the ms exchange rates measured from bulk spectroscopy. At later times, cross-peaks between the L and N state, as well as the K and N state, appear, consistent with a ms exchange time. (26)

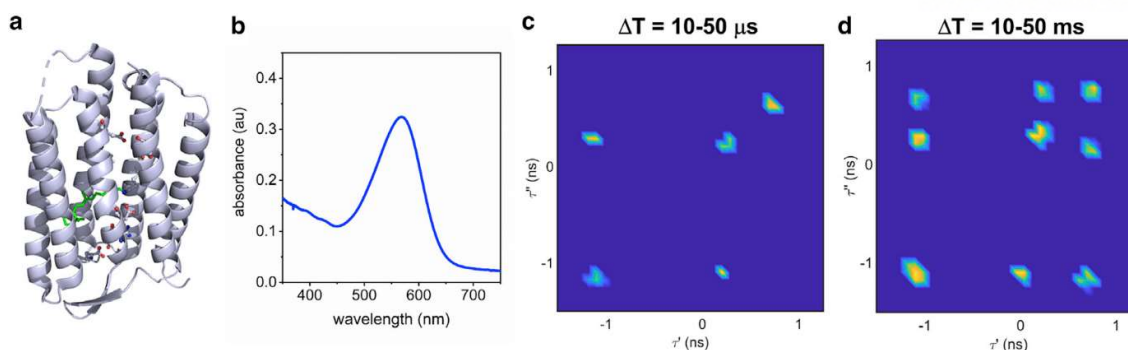


Figure 3.11. 2D-ILT on Bacteriorhodopsin single-molecule fluorescence data. (a) Crystal structure of bR, with the retinal chromophore highlighted in green. (b) Absorption spectrum of the bR in the ground state, with $\lambda_{max}=560$ nm. (c and d) 2D-FLCS spectra computed for $\Delta T = 10 - 50\mu s$ and $\Delta T = 10 - 50ms$. The three diagonal peaks can be assigned to the K, L, and N intermediate formed during the photoinduced reaction cycle of bR. Cross-peaks evolve because of chemical exchange between the intermediate species.

3.3.5. Comparison to existing methods.

We tested the fitting of same sets of simulated data and quantified the performance of each step individually. The key improvements suggested are kernel compression (KC), Tikhonov regularization (TK), and implementation of the BRD algorithm (BRD). **Table 3.2 and 3.3** show the computational time in seconds required to evaluate the 1D and 2D ILT fits using MEM alone, MEM with KC, (TK) alone, TK and KC, TK and BRD, and finally TR with KC and BRD. The corresponding 1D and 2D ILT fits are shown in **Figure 3.12**. We observe that all the above techniques give accurate results. For cases involving constrained optimization without compression, **Figure 3.12.g, i** (MEM and TK) the 2D ILTs appear smoother as compared to the fits obtained using constrained optimization along with kernel compression. However, considering computational efficiency and memory requirements (**Table 3.2** and **Table 3.3**), we observe that the best results are obtained when the techniques (TK, KC, BRD) are used in conjunction **Figure 3.12.e, k**.

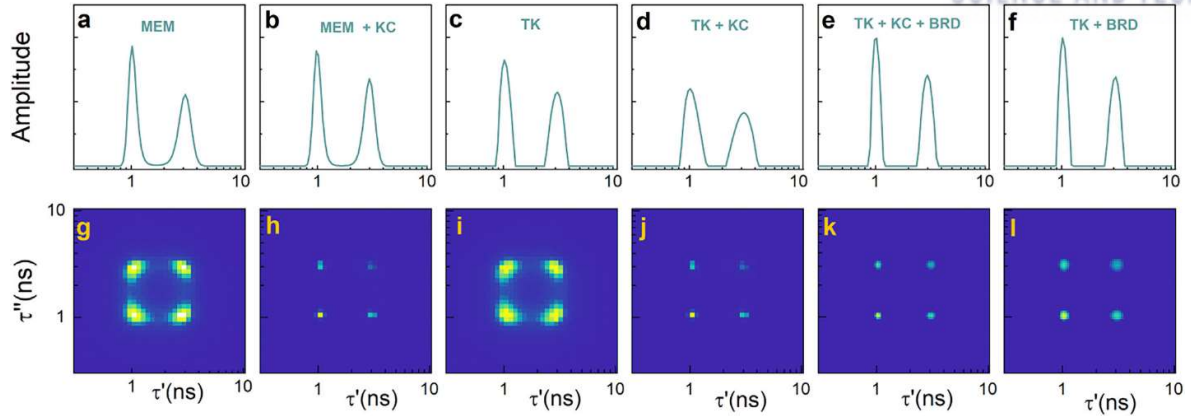


Figure 3.12. Comparison of TK and MEM. (a–f) Show 1D-ILT of same simulated data set comprising 2×10^6 total photons with lifetimes 1 and 3 ns. (g–l) show the 2D-ILT of the same data set evaluated at $\Delta T = 2\text{ms}$. The computational times for the simulations are given in Tables 2 and 3.

Regularizer	No Compression	Compression on	BRD	BRD Only
MEM	32 ± 3	9.1 ± 0.3	-	-
Tikhonov	29 ± 3	8.90 ± 0.25	0.17 ± 0.02	0.22 ± 0.02

Table 3.2 Computation times in seconds for methods for 1D ILT of same set of data comprising 2×10^6 total photons.

Regularizer	No Compression	Compression on	BRD	BRD Only
MEM	9×10^3	4×10^3	-	-
Tikhonov	8×10^3	1.4×10^3	20	500

Table 3.3 Computation times in seconds for methods for 2D ILT of same set of data with the 2D emission delay histograms evaluated at $\Delta T = 2\text{ms}$.

Basis Size	ID ILT Time (s)	2D ILT Time (s)
1000	0.40	1000
100	0.17	20
50	0.05	1

Table 3.4 CPU time required for execution of the fits for various sizes of basis set.

Table 3.4 shows typical computational times required to evaluate 1D and 2D-ILT fits for various sizes of lifetime

basis. The fittings were performed on an Intel-Core i7-4790 CPU and using home-built code written in MATLAB 2020a software. The code is publicly available at (https://github.com/saurabhtauke/2D-FLCS_King_lab).

3.3 Conclusion

Single-molecule 2D-FLCS is a powerful tool for studying dynamics of biological macromolecules, though the difficulty of computing inverse Laplace transforms may be prohibitive to its wide-spread application. Here, we outline a fast and robust method of computing 2D inverse Laplace transforms. The method, based on singular-valued decomposition and Tikhonov regularization, is adopted from NMR spectroscopy for application to single-molecule fluorescence spectroscopy. The approach allows for stable inversions of large, noisy data sets, common in single-molecule spectroscopy, to be carried out efficiently, without sacrificing the resolution of the spectra. Furthermore, using Monte carlo simulations to generate artificial photon streams, we demonstrate that this technique is robust in terms of the spectral resolution, noise tolerance, and computational efficiency. This provides an alternative method (beyond typical maximum entropy methods) for performing Laplace inversions of single-molecule fluorescence data that is easily implemented.

References

1. Jeener, J., B. Meier, P. Bachmann, and R. Ernst. 1979. Investigation of exchange processes by two-dimensional NMR spectroscopy. *The Journal of chemical physics* 71:4546-4553.
2. Moilanen, D. E., D. Wong, D. E. Rosenfeld, E. E. Fenn, and M. D. Fayer. 2009. Ion-water hydrogen-bond switching observed with 2D IR vibrational echo chemical exchange spectroscopy. *Proceedings of the National Academy of Sciences* 106:375-380.
3. Hoffmann, A., D. Nettels, J. Clark, A. Borgia, S. E. Radford, J. Clarke, and B. Schuler. 2011. Quantifying heterogeneity and conformational dynamics from single molecule FRET of diffusing molecules: recurrence analysis of single particles (RASP). *Physical Chemistry Chemical Physics* 13:1857.
4. Ha, T. 2001. Single-molecule fluorescence methods for the study of nucleic acids. *Current opinion in structural biology* 11:287-292.
5. Joo, C., H. Balci, Y. Ishitsuka, C. Buranachai, and T. Ha. 2008. Advances in single-molecule fluorescence methods for molecular biology. *Annu. Rev. Biochem.* 77:51-76.
6. Roy, R., S. Hohng, and T. Ha. 2008. A practical guide to single-molecule FRET. *Nature methods* 5:507-516.
7. Schuler, B., and W. A. Eaton. 2008. Protein folding studied by single-molecule FRET. *Current opinion in structural biology* 18:16-26.
8. Ishii, K., and T. Tahara. 2013. Two-Dimensional Fluorescence Lifetime Correlation Spectroscopy. 1. Principle. *The Journal of Physical Chemistry B* 117:11414-11422.
9. Ishii, K., and T. Tahara. 2013. Two-Dimensional Fluorescence Lifetime Correlation Spectroscopy. 2. Application. *The Journal of Physical Chemistry B* 117:11423-11432.
10. Otsu, T., K. Ishii, and T. Tahara. 2015. Microsecond protein dynamics observed at the single-molecule level. *Nature communications* 6:1-9.

11. Kondo, T., J. B. Gordon, A. Pinnola, L. Dall'Osto, R. Bassi, and G. S. Schlau-Cohen. 2019. Microsecond and millisecond dynamics in the photosynthetic protein LHCSR1 observed by single-molecule correlation spectroscopy. *Proceedings of the National Academy of Sciences* 116:11247-11252.
12. Kapusta, P., M. Wahl, A. Benda, M. Hof, and J. Enderlein. 2007. Fluorescence lifetime correlation spectroscopy. *Journal of fluorescence* 17:43-48.
13. Tikhonov, A. N. 1963. On the solution of ill-posed problems and the method of regularization. In *Doklady Akademii Nauk. Russian Academy of Sciences*. 501-504.
14. Venkataramanan, L., Y.-Q. Song, and M. D. Hurlimann. 2002. Solving Fredholm integrals of the first kind with tensor product structure in 2 and 2.5 dimensions. *IEEE Transactions on signal processing* 50:1017-1026.
15. Chouzenoux, E., S. Moussaoui, J. Idier, and F. Mariette. 2010. Efficient maximum entropy reconstruction of nuclear magnetic resonance T1-T2 spectra. *IEEE Transactions on Signal Processing* 58:6040-6051.
16. Su, G., X. Zhou, L. Wang, X. Wang, P. Yang, S. Nie, and Y. Zhang. 2019. Improved Butler–Reeds–Dawson algorithm for the inversion of two-dimensional NMR relaxometry data. *Mathematical Problems in Engineering* 2019.
17. Talele, S., and J. T. King. 2021. Fast and robust two-dimensional inverse Laplace transformation of single-molecule fluorescence lifetime data. *Biophysical Journal* 120:4590-4599.
18. Ishii, K., and T. Tahara. 2013. Two-dimensional fluorescence lifetime correlation spectroscopy. 1. Principle. *J. Phys. Chem. B* 117:11414-11422.
19. Ishii, K., and T. Tahara. 2013. Two-dimensional fluorescence lifetime correlation spectroscopy. 2. Application. *J. Phys. Chem. B* 117:11423-11432.
20. Read_PTU.m.

21. Butler, J., J. Reeds, and S. Dawson. 1981. Estimating solutions of first kind integral equations with nonnegative constraints and optimal smoothing. *SIAM Journal on Numerical Analysis* 18:381-397.
22. Turton, D. A., G. D. Reid, and G. S. Beddard. 2003. Accurate analysis of fluorescence decays from single molecules in photon counting experiments. *Analytical chemistry* 75:4182-4187.
23. Lanyi, J. K. 2004. Bacteriorhodopsin. *Annu. Rev. Physiol.* 66:665-688.
24. Talele, S., and J. T. King. 2021. Stability of a nonequilibrium biochemical cycle revealed by single-molecule spectroscopy. *BioRxiv*.
25. Lanyi, J. K. 1993. Proton translocation mechanism and energetics in the light-driven pump bacteriorhodopsin. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1183:241-261.
26. Varo, G., and J. K. Lanyi. 1991. Kinetic and spectroscopic evidence for an irreversible step between deprotonation and reprotonation of the Schiff base in the bacteriorhodopsin photocycle. *Biochemistry* 30:5008-5015.

Chapter 4: Reaction cycle of operating pump protein studied with single molecule spectroscopy

4.1 Introduction

Motor and pump proteins operate through non-thermal motions (nonequilibrium fluctuations)(1-4) induced by the input of energy, for example, by ATP hydrolysis or photon absorption. For some pump proteins, the mechanism of action has been adequately described using the principle of microscopic reversibility,(5) where equilibrium mechanical motions of the protein are rectified by an asymmetric potential that biases diffusion in one direction.(6, 7) In contrast, optically driven pump proteins are thought to operate through a ‘power-stroke’ mechanism, where a series of directional conformational transitions follow from a sudden structural change.(3, 8) For this physical mechanism, which is inherently far-from-equilibrium, the validity of equilibrium notions of time-reversal symmetry and detailed balance, as well as the response to temperature, are yet to be established.

In this work, we study the reaction cycle of bacteriorhodopsin (bR), an optically-driven H^+ pump found in Archaea (**Figure.4.1.a,b**),(9) at the single-molecule level. This system was chosen as a model system as the intermediate species involved in the reaction cycle have been well characterized and bulk spectroscopy experiments have provided the timescale for elementary transitions of the cycle.(10) However, ensemble experiments cannot resolve forward and reverse transitions required to quantify nonequilibrium thermodynamic properties.

Photon absorption induces a *trans-cis* isomerization of the retinal chromophore. Steric repulsions between the isomerized chromophore and the protein scaffold initiates a multi-step reaction cycle involving multiple intermediate species (**Figure.4.1c**).(11, 12) Key to our understanding of nonequilibrium reaction cycles are the roles of time-reversal symmetry and the corresponding entropy production.(13) To characterize these properties on a single-molecule level, we leverage multidimensional single-molecule fluorescence lifetime correlation spectroscopy (sm-2D-FLCS),(14-

16) which utilizes multiple light-matter interactions separated by a controlled waiting time to monitor structural or chemical transitions of a molecule. The details of the technique are published elsewhere.(17) Briefly, sm-2D-FLCS takes advantage of the differences in fluorescence lifetimes as a probe for resolving multiple intermediates during the catalytic cycle. The experiment is carried out on a confocal microscope equipped with TCSPC (Time Correlated Single Photon Counting) setup. Time resolved fluorescence photon streams are collected from single molecules and a 2D emission delay histogram is constructed by correlating the observed fluorescence emission delays in the photon stream at a known waiting time (Δt). A 2D Inverse Laplace transform (2D ILT) then yields a 2D fluorescence lifetime correlation spectra where species are represented by peaks corresponding to their fluorescence lifetimes. The exchange dynamics are observed through time-correlation functions (tcf) for forward and reverse transitions, which appear in opposite quadrants of the 2D spectrum(18) and therefore allow the forward and reverse conformational transitions to be measured independently. The violation of time-reversal symmetry within the reaction cycle is observed from the reciprocal tcf,(19) (where, reciprocal refers to the pair of forward and reverse tcf), which are equivalent for reversible transitions (obey time-reversal symmetry) and are not equivalent for irreversible transitions (violate time-reversal symmetry). The extent to which time-reversal symmetry is violated is a measure of entropy production rate in the cycle.(13) Using this approach, we are able to experimentally characterize the nonequilibrium thermodynamics and kinetics of portions of the bR reaction cycle and quantify the stability of the cycle when subject to kinetic perturbation.

4.2 Materials and methods

Purified and lyophilized wt-bR and bR-D96N extracted from *Halobacterium salinarum* were purchased from Halotek biomaterials (Halotek, Germany). The sample was loaded into UV-treated 8-well chambers (Lab-Tek, ThermoFisher, USA) for single-molecule experiments. Single-molecule fluorescence experiments were carried out on a commercial Leica SP8x microscope (Leica Microsystems GmbH, Germany). Pulsed excitation light at 532 nm was provided by a Fianium

Supercontinuum Laser operating at 40 MHz with an average power of $\sim 6 \mu\text{W}$ at the back-focal-plane.

Excitation and collection of fluorescence emission was carried out through a 100X oil-immersion objective. Fluorescence signal was passed through a band-pass filter (Semrock, USA) and was detected by photon avalanche photodiodes (PicoQuant, Germany) equipped with a picoHarp300 (PicoQuant, Germany) time-correlated single-photon-counting (TCSPC) system. Data processing and analysis was performed using in-house MATLAB code.

4.2.1. Sample preparation:

A $1 \mu\text{M}$ solution of membrane-bound bacteriorhodopsin (bR) was prepared in phosphate buffer (100 mM ionic strength, pH = 6). This solution was vortexed and sonicated for 5 minutes in cold water to disperse the purple membrane. The dispersed suspension was centrifuged for 5 minutes at 10K rpm. The supernatant was extracted, and the sonication/centrifuge procedure was repeated 3 times. The third-degree supernatant ($\sim 1 \text{ nM}$ bR) was used for experiments. An 8-well chamber (LabTek, Australia) was UV treated for > 8 hrs prior to use. The bR solution was then added to the chamber and kept at ambient light for 10-15 minutes. Membrane bound bR adsorb to the glass surface and are immobilized.

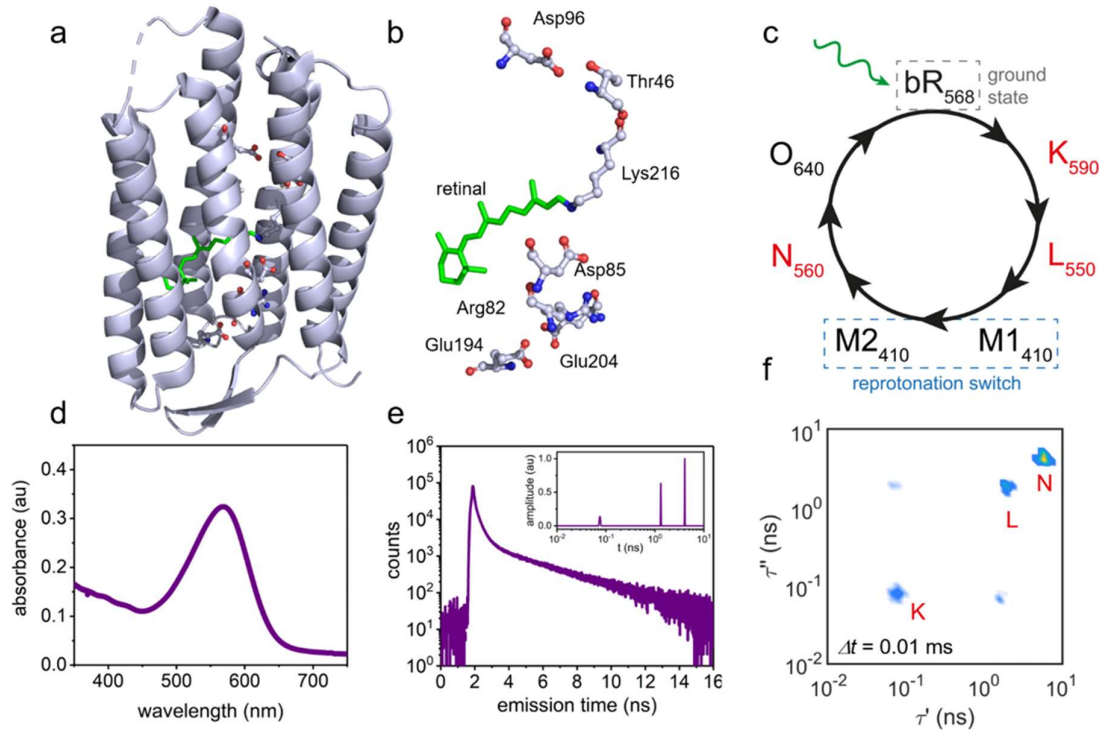


Figure 4.1. Single-molecule spectroscopy of Bacteriorhodopsin. Crystal structure of ground state bR showing a) the protein and b) the retinal chromophore and the amino acids involved in the H^+ transport chain of the reaction cycle (PDB: 1KBG). c) Photo-induced reaction cycle of bR, which involves 6 intermediates in addition to the ground state (gs). The subscripts on the labels denote the absorption maximum of the retinal chromophore for the given conformation. d) Absorption spectrum of retinal embedded in bR shows strong absorption at $\lambda=568$ nm at pH=6. e) The fluorescence lifetime histogram measured from a single protein shows multi-component relaxation following 532 nm excitation. A 1D-ILT of the lifetime histogram reveals three relaxation times of $\tau_1' \sim 0.08$ ns, $\tau_2' \sim 1.0$ ns, and $\tau_3' \sim 5.0$ ns (inset). f) sm-2D-FLCS spectrum measured at a waiting time of $\Delta t = 10 \mu s$.

4.2.2. SM-2D-FLCS:

Single-molecule dispersion and immobilization of bR was confirmed by confocal microscopy and fluorescence time traces (**Figure.4.2.a-c**). Emission from bR showed diffraction limited detection spots (**Figure. 4.2.a, b**) and single-step photo-physics (**Figure. 4.2.c**) expected for single molecules. Excitation at 532 nm overlaps with the absorption spectrum of the retinal chromophore (in ground state, K, L, and N intermediate states) embedded in the protein scaffold. The excitation laser both initiates the catalytic cycle and probes transient intermediates formed during the reaction cycle. Fluorescence relaxation from a single bR protein shows multi-component relaxation (**Figure. 4.2.d**). The emission delay times extracted from a real-time photon stream (**Figure. 4.3.a**) can be binned into a 1D histogram which are used to estimate a relaxation spectrum through an inverse Laplace transformation (**Figure. 4.3.b**). For a two-state system where each state has a distinct fluorescence lifetime the 1D fluorescence lifetime spectrum will show a peak at the relaxation rate of each state. 2D-FLCS spectra are computed by generating a 2D photon histogram of photon pairs separated by a systematic waiting time Δt (**Figure. 4.3.a**). A 2D inverse Laplace transformation of the histogram gives a 2D-FLCS spectrum for a given Δt (**Figure.4.3.c**). For a two-state system that undergoes chemical exchange at a well-defined timescale, the 2D-FLCS spectrum will contain diagonal peaks corresponding to species that did not exchange within Δt and cross-peaks corresponding to species that did exchange within Δt (**Figure. 4.3.c**). Monitoring the kinetics of the cross-peaks provides direct access to the forward and reverse exchange times.

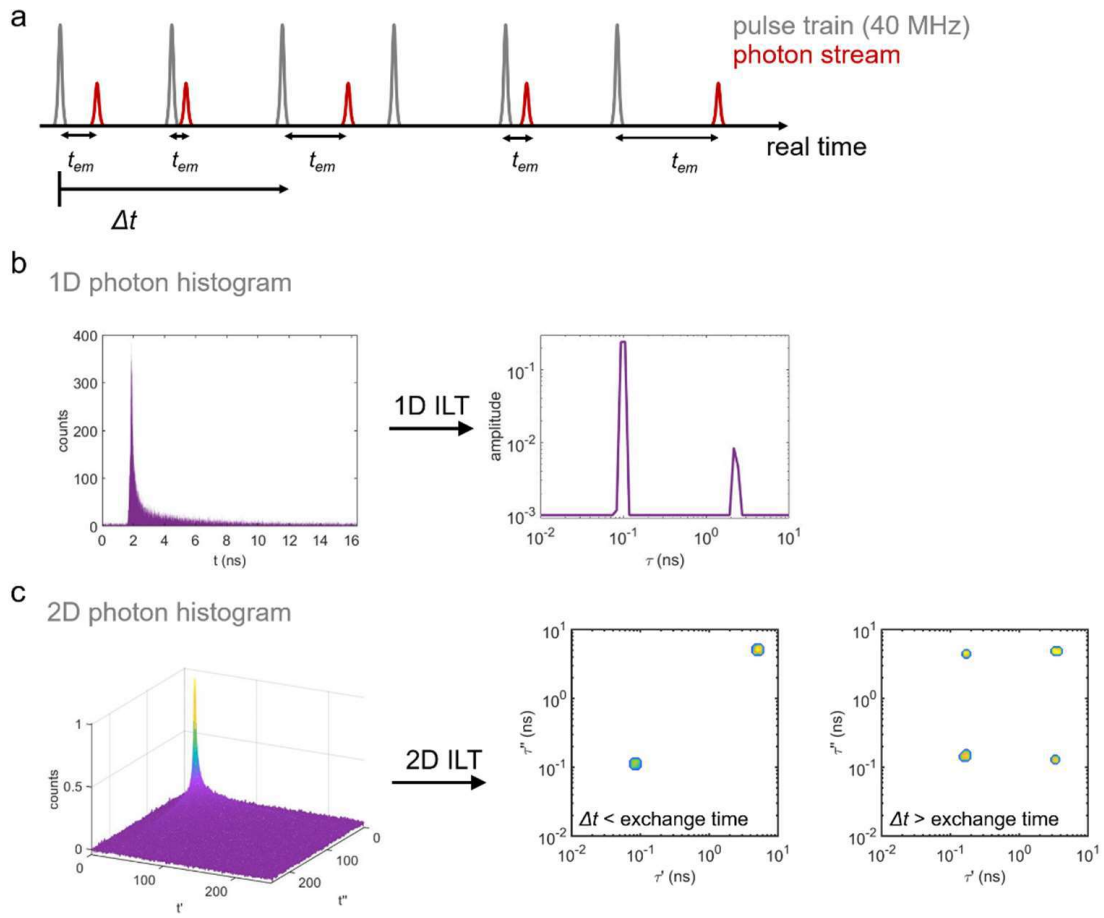


Figure 4.2. Single-molecule imaging and photo-physics. (a) Confocal image of single Bacteriorhodopsin proteins immobilized on a glass surface. (b) Diffraction-limited cross-section of fluorescence signal measured in confocal image. (c) Single-molecule fluorescence trace showing single-step photo-physics. (d) Fluorescence lifetime histogram from single wt-bR protein showing multicomponent relaxation.

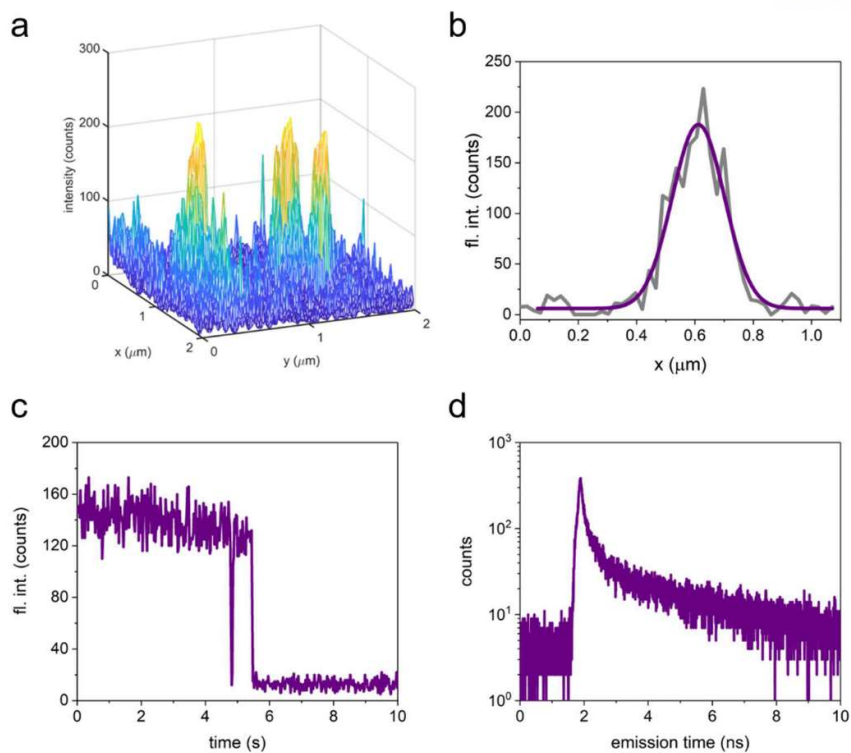


Figure 4.3. Experimental procedure for measuring sm-2D-FLCS data. (a) Excitation pulse train and photon stream measured with resolved photon emission time t_{em} (standard TCSPC measurement in Time-Tagged-Time-Resolved mode). (b) 1D histogram of photon emission time and the fluorescence lifetime spectrum computed by a 1D ILT for a simple two state system. (c) 2D histogram of photon emission measured at a set waiting time Δt and 2D-FLCS spectra computed by a 2D ILT. At early waiting times ($\Delta t < \text{exchange time}$) peaks are only measured along the diagonal, indicating species that remained in their initial state during the waiting time. At later times ($\Delta t > \text{exchange time}$), cross-peaks arise as species that were in an initial state exchange to the other state during the waiting time.

4.3 Results

We first characterize the directional flux through portions of the cycle by monitoring the transition kinetics between several intermediates. The endogenous retinal chromophore serves as the photo-trigger for the reaction cycle as well as a probe of the intermediates during the cycle. Limitations in the experiment prevent transitions between each intermediate to be directly measured. Instead, portions of the reaction cycle are characterized through select intermediates to which our experiments are sensitive. The cycle of bR involves six known intermediate structures that exist for timescales ranging between μs to ms (K, L, M1, M2, N, O, **Figure.4.1.c**).^(11, 12) In the ground state structure, the retinal chromophore has an absorption maximum at $\lambda_{\text{max}} = 568 \text{ nm}$ (**Figure.4.1.d**). Fluorescence emission from a single monomeric bR protein, confirmed by diffraction limited emission spots observed in confocal microscopy and single-step photo-physics (**Figure. 4.2.a-c**), shows multi-component relaxation following a 532 nm excitation (**Figure.4.1.e, Figure. 4.2.d**). An inverse Laplace transform (ILT)⁽¹⁷⁾ of the cumulative photon histogram measured from more than 100 single protein molecules (**Figure.4.1.e**) gives a 1D fluorescence lifetime spectrum that contains three distinct relaxation peaks (**Figure.4.1.e, inset**). We do not anticipate signal from the M1, M2, and O intermediates due to low absorption at the 532 nm excitation.^(11, 12) Furthermore, fluorescence from the ground state structure is unlikely due to the ultrafast isomerization reaction that occurs in the excited state.⁽²⁰⁾ Therefore, the signals arise from the K intermediate, the L intermediate, and the N intermediate.

The occupancies of the K, L, and N conformational states of bR was determined from 1D emission lifetime spectra. The relative amplitude of each state in the spectrum is proportional to the number of time steps in the experiment spent in each state. The occupancies of each state depend on temperature and change slightly over the temperature change used in this study (**Figure 4.4**).

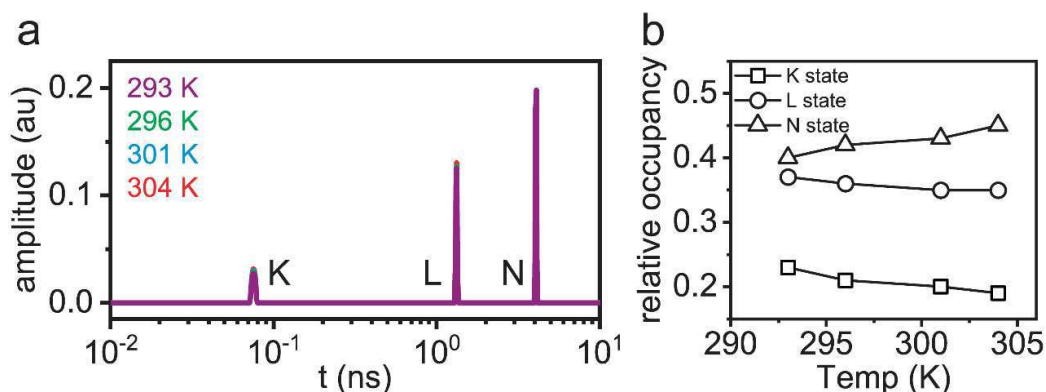


Figure 4.4. Temperature dependence of bR state occupation. (a) 1D fluorescence emission time spectrum and (b) relative state occupancies of bR as a function of temperature.

The 2D-FLCS spectrum generated from a 2D-ILT of the photon histogram(17) calculated at a waiting time of $\Delta t = 10 \mu\text{s}$ shows three distinct diagonal peaks (**Figure.4.1.f**). The cross-peaks between the signals at $\tau' \sim 0.08 \text{ ns}$ and $\tau' 1.0 \text{ ns}$ indicate rapid exchange between the two states represented by these signals. However, no cross-peaks are observed between the diagonal signal at $\tau' \sim 5.0 \text{ ns}$ and the other two states. This is consistent with ensemble experiments which have measured the formation time of the N intermediate to be on the ms timescale.(11, 12) Furthermore, the 1D relaxation spectrum (**Figure. 4.4**) shows the peaks at $\tau' \sim 0.08 \text{ ns}$, $\tau' \sim 1.0 \text{ ns}$ are in a 1:3 ratio, consistent with the populations of the K and L intermediates.(21) From these observations we can assign the diagonal peaks at $\tau' \sim 0.08 \text{ ns}$, $\tau' \sim 1.0 \text{ ns}$, and $\tau' \sim 5.0 \text{ ns}$ to the K state, the L intermediate, and the N intermediates, respectively (**Figure. 4.1.c,f**).

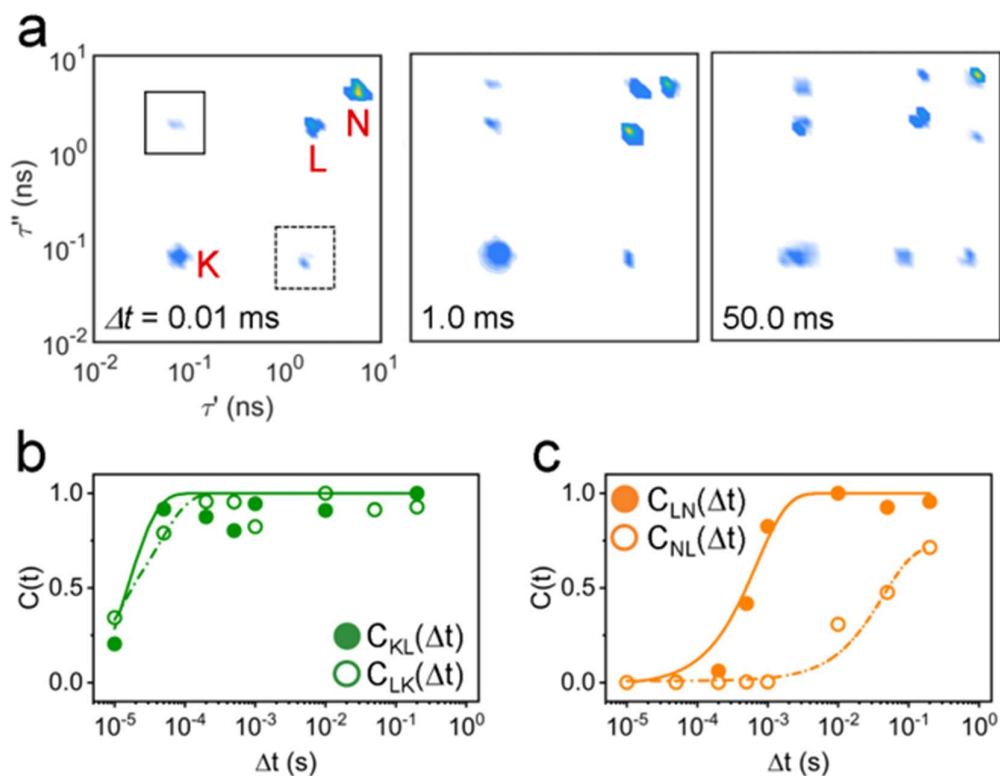


Figure 4.5. Microscopic irreversibility of bR catalytic cycle. a) sm-2D-FLCS spectra of bR shown for waiting times ranging from 0.01 to 50 ms. Exchange kinetics at 291 K for waiting times $\Delta t=0.01$ –200 ms between b) K intermediate to L intermediate and c) L intermediate to N intermediate. Forward transitions, represented by solid symbols and solid lines, are measured from the upper quadrant of the 2D spectra. Reverse transitions, represented by open symbols and dashed lines, are measured from the lower quadrant of the 2D spectra. The forward (k_{ij}) and reverse (k_{ji}) transition rates are measured by fitting the cross-correlation functions to exponential growth functions (solid and dashed lines). The experiments were performed at $T=291$ K.

To characterize the kinetics of exchange between different intermediates of bR reaction cycle, we measure kinetic traces for variable waiting times Δt ranging from 0.01 to 200 ms (**Figure.4.5.a**) at temperature of 291 K. The time-dependent amplitudes of the spectrum are given by the tefs,(14, 15)

$$C_{ij}(\Delta t) = \frac{\langle S_i(t)S_j(t + \Delta\tau) \rangle}{\langle S_i(t)S_j(t + \infty) \rangle} \quad 4.15$$

and

$$C_{ji}(\Delta t) = \frac{\langle S_j(t)S_i(t + \Delta\tau) \rangle}{\langle S_j(t)S_i(t + \infty) \rangle} \quad 4.16$$

where S_i and S_j are the instantaneous probabilities of observing the system in states i and j , at given times respectively, measured in our experiment through unique fluorescence lifetimes τ_i and τ_j of each state. The equivalency of **Eq. 4.1** and **Eq. 4.2** under the condition of microscopic reversibility is given by detailed balance and time-reversal symmetry.(19) The characteristic decay rate of the tcf for the i to j transition, r_{ij} , is related to the transition probability by(19)

$$r_{ij} = p_i k_{ij}(i/j; \Delta\tau) \quad 4.17$$

and

$$r_{ji} = p_j k_{ji}(j/i; \Delta\tau) = p_i k_{ij}(j/i; -\Delta\tau) \quad 4.18$$

where p_i is the average occupancy of state i , and k_{ij} is the probability of observing a transition from state i to j within time Δt . Experimentally measured decay rate, r_{ij} , and state occupancy p_i allows for the transition probability k_{ij} to be determined and an elementary rate law to be defined. Under conditions of microscopic reversibility, the forward and reverse transition probabilities are equivalent, reflecting detailed balance and time-reversal symmetry of the reaction process. The 2D spectra for a reversible process are therefore symmetric along the diagonal, as the reciprocal cross-correlation functions are identical (defined by Eq.4.1 and 4.2). In contrast, for microscopically irreversible transitions, time-reversal symmetry of the reaction process is violated, and the equivalency of cross-correlation functions Eq. 4.3 and Eq. 4.4 is broken. The resulting 2D spectra are asymmetric along the diagonal, indicating the presence of nonequilibrium fluxes through a given transition.(19, 22)

The cross-peak dynamics between the K intermediate and the L intermediate show rapid forward (**Figure.4.5.b**, solid symbols) and reverse reactions (**Figure.4.5.b**, open symbols), with the exchange

rates being of the same order of magnitude ($k_{KL} = 6.9 \times 10^4 \text{ s}^{-1}$ and $k_{LK} = 6.6 \times 10^4 \text{ s}^{-1}$). The near equivalency of the exchange rates suggests rapid equilibration of the transition. At $\Delta t = 0.20 \text{ ms}$, a cross-peak emerges between the L and N intermediates that is asymmetric (no cross-peak for the reverse transition) (**Figure.4.5.a**), indicating a microscopically irreversible step in the reaction cycle. The forward transition occurs on a timescale of $\sim 0.50 \text{ ms}$ ($k_{LN} = 2.0 \times 10^3 \text{ s}^{-1}$), while the reverse transition occurs on a timescale over 200 ms ($k_{NL} = 5.0 \text{ s}^{-1}$) (**Figure.4.5.c**).

Chemically, the L-N transition corresponds to the reprotonation of the Asp96 residue on the cytoplasmic side of the protein (M intermediate) and formation of a Grotthuss-like H^+ wire between the Asp96 residue and the Schiff base of retinal.(23, 24) Proton uptake and reprotonation of the retinal Schiff base (formation of N intermediate) from the cytoplasmic side is thought to be the switch step of the photo-cycle.(25, 26) The properties of the measured tcf demonstrate that the reprotonation of Asp96 is microscopically irreversible and has non-zero entropy production associated with the transition. The presence of microscopically irreversible transitions implies that nonequilibrium transitions are inherent to the protein's reaction cycle and are required to maintain stable fluxes through portions of the cycle.

4.4 Discussion

We analyze the nonequilibrium thermodynamic properties of the bR reaction cycle as a function of temperature. The temperature-dependent occupancy of the K, L, and N states are shown in **Figure.4.6**. The k_{ij} and k_{ji} for an elementary transition are given by Arrhenius expressions,

$$k_{ij} = ae^{\left(\frac{-\Delta G_{ij}^\ddagger}{k_B T}\right)} \quad 4.19$$

and

$$k_{ji} = ae^{\left(\frac{-\Delta G_{ji}^\ddagger}{k_B T}\right)} = ae^{\left(\frac{-[\Delta G_{ij}^\ddagger + \Delta G_{trans}]}{k_B T}\right)} \quad 4.20$$

where a is a pre-exponential factor, ΔG_{ij}^\ddagger and ΔG_{ji}^\ddagger are the Gibbs free energy of activation for the forward and reverse transitions, respectively, and ΔG_{trans} is the Gibbs free energy of the transition.

Assuming the entropic contribution is temperature independent over the range studied, the temperature dependence of k_{ij} and k_{ji} for a given transition in the cycle will be identical on the condition that $\Delta H_{trans} = 0$ and different when $\Delta H_{trans} \neq 0$. Arrhenius plots for the K-L transitions and the L-N transitions are shown in **Figure.4.3.a, b**. The k_{KL} and k_{LK} for the K-L transitions show identical temperature dependence (**Figure.4.6.a**), indicating that the $\Delta H_{trans} = 0$ for the transition. In contrast, the k_{LN} and k_{NL} for the L-N transitions (**Figure.4.6.b**) are significantly different at low temperature, but converge at high temperature, suggesting that $\Delta H_{trans} \neq 0$ and the irreversibility of the transition needed to maintain a net forward flux through the cycle diminishes with increasing temperature.

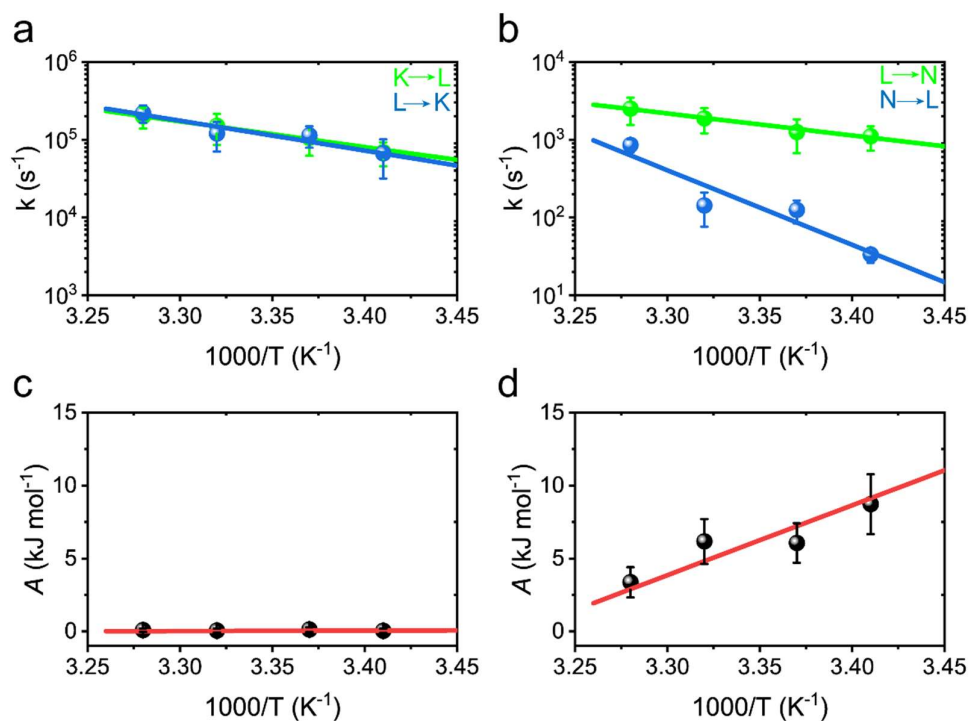


Figure 4.6. Thermal stability of the bR reaction cycle. Arrhenius plots for the a) K-L and b) L-N transition rates. The k_{ij} and k_{ji} values are shown in green and blue, respectively. c) The affinity A of the K-L transition plotted vs T^{-1} . As the transition is microscopically reversible, the A value is ~ 0 kJmol⁻¹ at all temperatures measured. d) The A of the L-N transition plotted vs T^{-1} . The microscopically irreversible transition has an A value of ~ 10 kJmol⁻¹ at 291 K that decreases linearly towards 0 kJmol⁻¹ with increasing temperature. The temperature dependencies of A are fit to the Gibbs-Helmholtz equation [red line, Equation (6)].

The affinity A of a nonequilibrium process quantifies the thermodynamic driving force. de Donder demonstrated that A is related to ΔG^0 for a chemical reaction.(27)

$$A = -\Delta G_{trans}^0 \quad 4.21$$

Using **Eq. 4.3-4.6**, we can rewrite A for a given transition in terms of k_{ij} and k_{ji} , and the state occupancies,

$$A = Nk_B T \ln \left(\frac{p_i \cdot k_{ij}}{p_j \cdot k_{ji}} \right) \quad 4.22$$

where the ΔG^0 values are given by $\Delta G - NK_b T \log(Q)$. Plots of A for the K-L and L-N transitions are shown in **Figure.4.6.c, d**. The A for the K-L transition, and by extension the ΔG_{trans} , is 0 kJ mol^{-1} at all temperatures (**Figure.4.6.c**). In contrast, A for the L-N transition is $\sim 10 \text{ kJ mol}^{-1}$ at 291 K and decreases rapidly towards 0 kJ mol^{-1} with increasing temperature (**Figure. 4.6.d**). According to **Eq. 4.7**, the temperature dependence of A is given by the Gibbs-Helmholtz equation.

$$\frac{\Delta G_{trans}(T_2)}{T_2} - \frac{\Delta G_{trans}(T_1)}{T_1} = \Delta G_{trans} \left(\frac{1}{T_2} - \frac{1}{T_1} \right) \quad 4.23$$

We find that the experimental data for A as a function of T is adequately fit with **Eq. 4.9**, yielding $\Delta H_{trans} \sim 0 \text{ kJ mol}^{-1}$ and $\Delta H_{trans} \sim -30 \text{ kJ mol}^{-1}$ for the K-L and L-N transitions, respectively (**Figure.4.6.c, d**).

The thermodynamics of a nonequilibrium process can be quantified by the rate of entropy production rate, σ , given by (28),

$$\sigma = \frac{J \cdot A}{T} \quad 4.24$$

where J is the net flux defined for an elementary transition as the difference in forward and reverse reaction rates, $r_{ij} - r_{ji}$. Therefore, experimental measurements of the forward and reverse reaction rates (r_{ij} and r_{ji}) and rate constants (k_{ij} and k_{ji}) are sufficient to determine σ of a nonequilibrium chemical process. For K-L and L-N transitions, σ is shown as a function of temperature (**Figure.4.7.a**). As expected, the K-L transition shows negligible σ_{KL} . In contrast, the L-N transition shows significant σ_{LN} , expected for an irreversible transition, that decreases exponentially with increasing temperature.

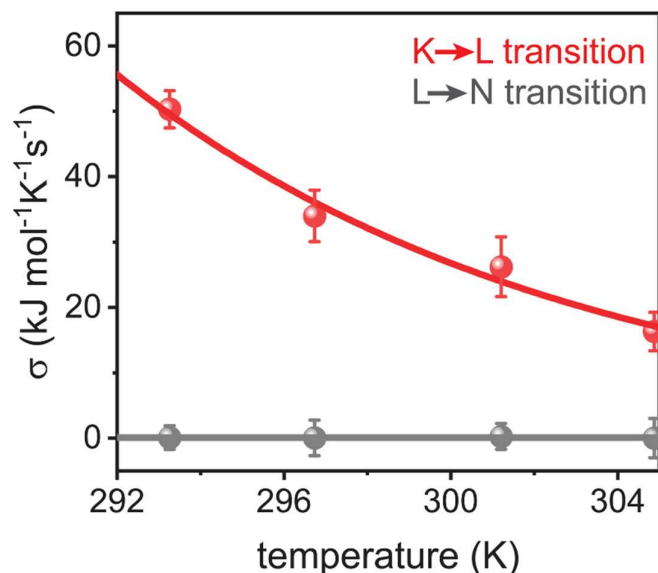


Figure 4.7. Entropy production rate and variance of flux. a) Entropy production for the K-L and L-N transitions computed from Equation (7). The entropy production rate σ associated with the reversible K-L transition is $0 \text{ kJmol}^{-1}\text{K}^{-1}\text{s}^{-1}$ at all temperatures measured, while the irreversible L-N transition has significant entropy production rate ($\sim 50 \text{ kJmol}^{-1}\text{K}^{-1}\text{s}^{-1}$) that decreases exponentially with increasing temperature.

4.4.1. Exchange Kinetics of D96N Mutant.

The L to N transition is observed to consist of an irreversible M1-M2 transition in wt-bR. The D96N mutant protein changes the aspartic acid residue at the D96 site, which is responsible for the proton uptake step, to an asparagine residue that hinders proton uptake (**Figure.4.8.a**). The D96N mutant has previously been observed to have a reversible switch step, with the M1-M2 and thus effectively the L-N transition being significantly hindered [1-2]. Sm-2D-FLCS exchange experiments confirm the reversible nature of the L to N transition in the D96N mutant, without the forward transition timescale

being significantly hindered.

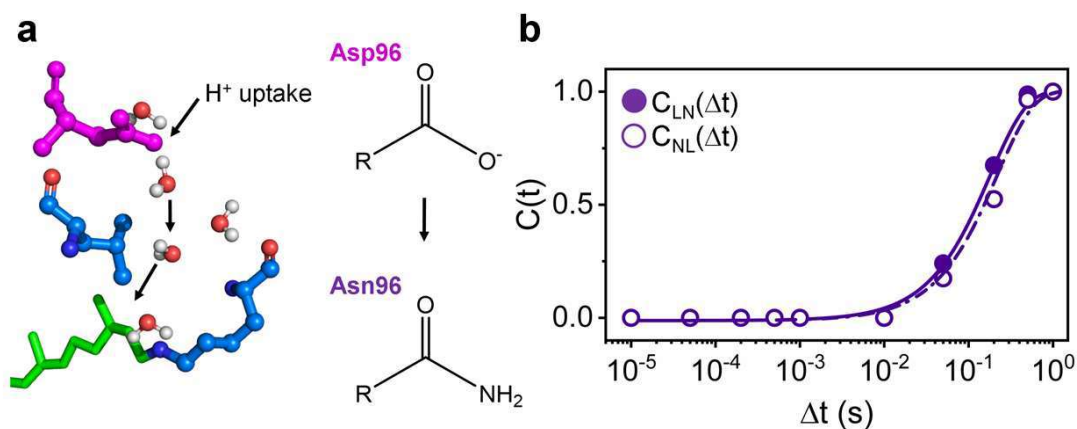


Figure 4.8. Exchange dynamics of D96N mutant. (a) Chemical structure of reprotonation site at Asp96 residue and the structure of the D96N mutant. (b) Chemical exchange from L to N (solid symbols) and N to L states (open symbols), showing near reversible exchange in the mutant protein.

4.4.2. Influence of continuous laser illumination on the photocycle.

Bulk experiments using laser excitation has convincingly demonstrated that the photocycle mechanism of bR is not impacted by excitation [9]. Following photon absorption, a small amount of energy is dissipated as heat as the excited chromophore vibrationally relaxes. For low quantum-yield chromophores, however, the absence of fluorescence necessarily means that all the energy contained in the photon absorption must be thermally dissipated, leading to significant local heating. To demonstrate that this is not a concern in the bR experiments, we measured how the transition rates responded to small changes in external temperature (~18 °C to 30°C). As expected for an unperturbed system, the transition rates showed a strong dependence on changes to external temperature and nearly perfect Arrhenius behavior (**Figure 4.6.a,b**). Such a clear trend would not be expected if the transition rates of the photocycle were strongly influenced from photon absorption during the probe process. This leads us to conclude that the dissipated energy from the probe does not interfere with the bacteriorhodopsin photocycle.

4.4.3. Lapping effect in the photocycle.

We can rule out the lapping effect in the photocycle using sm-2D-FLCS experiments on bR-D96N mutant, which replaces the Asp96 residue responsible for proton uptake with Asn residue. This mutation is known to significantly hinder the L to N transition time (specifically, the M-N transition time), which contains the switch step responsible for the directional pumping action [10]. Hindering M to N with the D96N mutation causes this transition to become reversible. Our sm-2D-FLCS experiments show that the forward and reverse transition times for the L-N and N-L transitions are on the order of ~200 ms and 260 ms, respectively. The data is consistent with bulk experiments showing the reversibility of the switch step in the D96N mutant protein. Our results demonstrate that the reversibility is due to the destabilization of the N state hindering the L-N transition, while the reverse N-L transition is only slightly modulated. In the case of the D96N mutant the forward and reverse reaction times are comparable. Here, the only contribution to the N-L correlation function is through the direct reverse transition between the states and not through the lapping effect. The consistency between the timescale for the reverse transition rate from N-L in the wt-bR and the D96N mutant confirms that the lapping effect does not significantly contribute to the observed kinetics.

4.5 Conclusion

The mechanism by which biological machinery overcomes randomizing thermal forces to achieve directional action are not comprehensively understood, with key questions regarding microscopic reversibility and the applicability of equilibrium thermodynamics still remaining. The results presented here demonstrate that microscopic irreversibility is the fundamental operating principle of directional H⁺ transport by bacteriorhodopsin. Therefore, equilibrium notions of detailed balance and time-reversal symmetry often invoked for chemically driven pumps are not applicable to the function of optically driven pump proteins. Similarly, the thermal destabilization of the reaction cycle of bacteriorhodopsin is likely unique to pumps that operate through a power-stroke mechanism in which the nonequilibrium nature of the cycle is inherent to select conformational transitions of the protein.

REFERENCES

1. Vale, R. D., and R. A. Milligan. 2000. The way things move: Looking under the hood of molecular motor proteins. *Science* 288:88-95.
2. Leibler, S., and D. A. Huse. 1993. Porters versus rowers: A unified stochastic model of motor proteins. *J. Cell Biol.* 121:1357-1368.
3. Hwang, W., and M. Karplus. 2019. Structural basis for power stroke vs. Brownian ratchet mechanisms of motor proteins. *Proc. Natl. Acad. Sci. U.S.A.* 116:19777-19785.
4. Taniguchi, Y., M. Nishiyama, Y. Ishii, and T. Yanagida. 2005. Entropy rectifies the Brownian steps of kinesin. *Nat. Chem. Biol.* 1:342-347.
5. Astumian, R. D. 2012. Microscopic reversibility as the organizing principle of molecular machines. *Nat. Nanotechnol.* 7:684-688.
6. Astumian, R. D. 1997. Thermodynamics and kinetics of a Brownian motor. *Science* 276:917-922.
7. Julicher, F., A. Ajdari, and J. Prost. 1997. Modeling molecular motors. *Reviews of Modern Physics* 69:1269-1281.
8. Astumian, R. D. 2016. Optical vs. chemical driving for molecular machines. *Faraday Discussions* 195:583-597.
9. Luecke, H., B. Schobert, H. T. Richter, J. P. Cartailler, and J. K. Lanyi. 1999. Structure of bacteriorhodopsin at 1.55 angstrom resolution. *Journal of Molecular Biology* 291:899-911.
10. Neutze, R., E. Pebay-Peyroula, K. Edman, A. Royant, J. Navarro, and E. M. Landau. 2002. Bacteriorhodopsin: a high-resolution structural view of vectorial proton transport. *Biochimica Et Biophysica Acta-Biomembranes* 1565:144-167.
11. Lanyi, J. K. 2004. Bacteriorhodopsin. *Annu. Rev. Physiol.* 66:665-688.

12. Ernst, O. P., D. T. Lodowski, M. Elstner, P. Hegemann, L. S. Brown, and H. Kandori. 2014. Microbial and animal rhodopsins: Structures, functions, and molecular mechanisms. *Chemical Reviews* 114:126-163.
13. Maes, C., and K. Netocny. 2003. Time-reversal and entropy. *Journal of Statistical Physics* 110:269-310.
14. Ishii, K., and T. Tahara. 2013. Two-dimensional fluorescence lifetime correlation spectroscopy. 1. Principle. *J. Phys. Chem. B* 117:11414-11422.
15. Ishii, K., and T. Tahara. 2013. Two-dimensional fluorescence lifetime correlation spectroscopy. 2. Application. *J. Phys. Chem. B* 117:11423-11432.
16. Otsu, T., K. Ishii, and T. Tahara. 2015. Microsecond protein dynamics observed at the single-molecule level. *Nat. Commun.* 6:1-9.
17. Talele, S., and J. T. King. 2021. Fast and robust 2D inverse Laplace transformation of single-molecule fluorescence lifetime data. *Biophysical Journal* 120:4590-4599.
18. Aue, W. P., E. Bartholdi, and R. R. Ernst. 1976. 2-dimensional spectroscopy - Application to nuclear magnetic-resonance. *Journal of Chemical Physics* 64:2229-2246.
19. Steinberg, I. Z. 1986. On the time-reversal of noise signals. *Biophysical Journal* 50:171-179.
20. Nogly, P., T. Weinert, D. James, S. Carbajo, D. Ozerov, A. Furrer, D. Gashi, V. Borin, P. Skopintsev, K. Jaeger, K. Nass, P. Bath, R. Bosman, J. Koglin, M. Seaberg, T. Lane, D. Kekilli, S. Brunle, T. Tanaka, W. T. Wu, C. Milne, T. White, A. Barty, U. Weierstall, V. Panneels, E. Nango, S. Iwata, M. Hunter, I. Schapiro, G. Schertler, R. Neutze, and J. Standfuss. 2018. Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser. *Science* 361.
21. Wickstrand, C., P. Nogly, E. Nango, S. Iwata, J. Standfuss, and R. Neutze. 2019. Bacteriorhodopsin: Structural Insights Revealed Using X-Ray Lasers and Synchrotron Radiation. In *Annual Review of Biochemistry*. *Annual Review of Biochemistry*, Vol 88. R. D. Kornberg, editor, pp. 59-83.

22. Qian, H., and E. L. Elson. 2004. Fluorescence correlation spectroscopy with high-order and dual-color correlation to probe nonequilibrium steady states. *Proceedings of the National Academy of Sciences of the United States of America* 101:2828-2833.
23. Garczarek, F., and K. Gerwert. 2006. Functional waters in intraprotein proton transfer monitored by FTIR difference spectroscopy. *Nature* 439:109-112.
24. Weinert, T., P. Skopintsev, D. James, F. Dworkowski, E. Panepucci, D. Kekilli, A. Furrer, S. Brunle, S. Mous, D. Ozerov, P. Nogly, M. T. Wang, and J. Standfuss. 2019. Proton uptake mechanism in bacteriorhodopsin captured by serial synchrotron crystallography. *Science* 365:61-65.
25. Lanyi, J. K. 1993. Proton translocation mechanism and energetics in the light-driven pump bacteriorhodopsin. *Biochim. Biophys. Acta* 1183:241-261.
26. Kataoka, M., H. Kamikubo, F. Tokunaga, L. S. Brown, Y. Yamazaki, A. Maeda, M. Sheves, R. Needleman, and J. K. Lanyi. 1994. Energy coupling in an ion-pump - The reprotonation switch of bacteriorhodopsin. *Journal of Molecular Biology* 243:621-638.
27. de Donder, T., and P. V. Rysselberghe. 1936. *Thermodynamic Theory of Affinity*. Stanford University Press, California.
28. de Groot, S. R., and P. Mazur. 1962. *Non-equilibrium thermodynamics*. Dover, New York.

Chapter 5. Protein-Bath Coupling at Intermediate Timescales

5.1 Introduction.

Diffusive barrier-crossing processes are central to the reaction kinetics of proteins. Early work by Frauenfelder and colleagues (1, 2) on the binding kinetics of carbon monoxide to myoglobin revealed that large scale conformational dynamics of the protein were strongly coupled to the fluctuations of the bath. From this work emerged a framework for understanding protein motions based on a rugged free energy landscape and a hierarchy of dynamic timescales. (3-9). Slow global conformational motions (ms-s) require a significant rearrangement of solvent molecules surrounding the protein and are therefore coupled to the viscosity of the bath (4). In contrast, fast librational motions of sidechains (ps-ns) require little to no rearrangement of surrounding molecules and are therefore largely uncoupled from the bath (4, 9). Less is known about protein-solvent coupling at intermediate timescales.

In this work, we study the microsecond chemical reaction occurring in the chromophore pocket of enhanced green fluorescent protein (eGFP) using single-molecule spectroscopy. The crystal structure of eGFP and chromophore pocket are shown in **Figure 5.1.a** (10). At equilibrium, the chromophore pocket undergoes constant rearrangements of the local hydrogen bonding network driven by the dihedral rotation of the Glu222 residue (**Figure 5.1.b**). The internal reaction results in a change in the fluorescence lifetime of the endogenous chromophore without a significant change in fluorescence intensity. Previous work using fluorescence lifetime correlation spectroscopy (FLCS) resolved the exchange times between the two states to be on the order of μs (11). Here, we are interested in studying the internal reaction kinetics as a function of solvent viscosity.

Section 5.2 Methods

5.2.1. Sample preparation

eGFP (Biovision Inc., United States) was labeled using a biotin labeling kit (Roche, Switzerland) and purified. A solution of biotin-labeled eGFP (~0.5 mM) was incubated on a streptavidin coated coverslip (Microsurface Inc, United States) for 60 minutes. Following incubation, the coverslip was rinsed with DI water to remove unbound eGFP. The coverslip was then mounted to a microscope slide and solvated with a buffer solution (PBS, pH 7.45, Thermofisher, United States). For viscosity dependent studies, mixtures of PBS buffer mixed with glycerol (Sigma Aldrich, United States) at volume ratios ranging from 0% to 30% by volume.

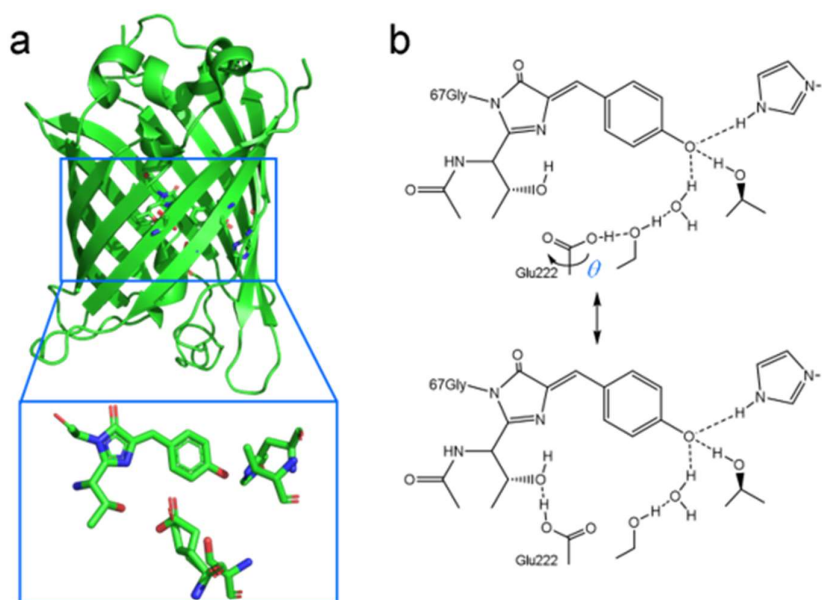


Figure 5.1. Structure and internal reaction of eGFP. (a) Crystal structure of eGFP (top) with the chromophore pocket highlighted (bottom). (b) Internal reaction between two bright conformations of the chromophore pocket. Dihedral rotation of the Glu222 residue drives a rearrangement of the local hydrogen bonding network around the chromophore.

5.2.2. Single-molecule spectroscopy

Single-molecule imaging was carried out on a commercial Leica SP8 system (Leica, Germany). Photon-counting was performed by a SPAD detector equipped with TCSPC unit (PicoQuant, Germany). From the single-molecule photon stream containing roughly 5×10^6 photons, a series of 2D-FLCS spectra were computed using a 2D-inverse Laplace transformation (2D-ILT). The entire data analysis process is outlined in detail elsewhere.(12)

5.3 Results

We use single-molecule 2D-FLCS (sm-2D-FLCS) (13-15) to resolve the forward and reverse exchange times, t_f and t_r , as the solvent composition is systematically changed from PBS buffer only to 30% glycerol-PBS buffer mixtures. The eGFP molecules are immobilized sparsely on a glass substrate using biotin-streptavidin linkage. The experimental details of this approach are reported in detail elsewhere (16, 17). Briefly, a time-resolved fluorescence photon stream is collected from isolated eGFP proteins using 488 nm excitation. From the single-molecule photon stream containing $\sim 5 \times 10^6$ photons, a fluorescence emission histogram shows two component relaxation (**Figure.5.2.a**), consistent with two bright conformations of eGFP. A series of 2D fluorescence lifetime histograms are generated correlating photons recorded at the beginning and the end of a systematically varied waiting time, Δt , (13, 14, 18) which can range from ms to s. A 2D fluorescence emission rate spectrum is then computed at each Δt by a 2D inverse Laplace transformation from the lifetime histogram(18). The signal amplitudes in the 2D spectrum are given by time-correlation functions (TCFs)

$$C_{ij}(\Delta t) = \langle S_i(t) \cdot S_j(t + \Delta t) \rangle \quad 25$$

where S is the probability of the system being found in state i or j . The autocorrelation functions ($i = j$) appear on the diagonal of the spectrum, and cross-correlation functions ($i \neq j$) indicating chemical exchange between two species appear as cross-peaks. This experimental approach has proven productive in studying complex equilibrium (13, 14, 19, 20) and nonequilibrium (15) conformational transitions of biological macromolecules.

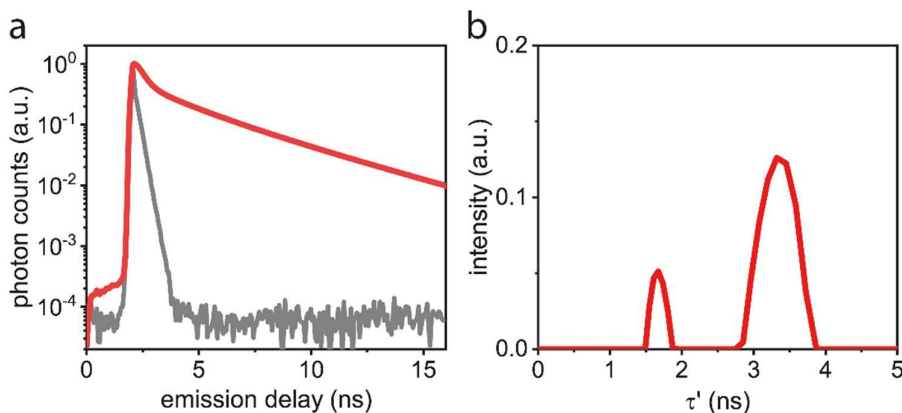


Figure 5.2. eGFP single-molecule fluorescence lifetime data. (a) Fluorescence emission delay histogram for eGFP (red data), shown with the instrument response function (grey data). (b) Corresponding 1D-ILLT of the relaxation data, showing two peaks at $t_1 \sim 1.8$ and $t_2 \sim 3.2$ ns, corresponding to the two conformations of the chromophore pocket.

Sm-2D-FLCS spectra of eGFP as a function of Δt are shown in **Figure 5.3.a**. At $Dt = 0.005$ ms, the 2D-spectrum shows two peaks along the diagonal at $t_1 \sim 1.8$ and $t_2 \sim 3.2$ ns, corresponding to the two conformations of the chromophore pocket (**Figure 5.1.b**). There are also two weak cross-peaks, indicating exchange between these conformations has occurred within Δt . As Δt increases, the cross-peak amplitudes grow until saturating near $\Delta t = 0.05$ ms. Plotting the amplitude of the cross-peaks as a function of Δt shows an exponential growth with a time constant of roughly $\tau_{rxn} = 5$ ms for eGFP in buffer (**Figure 5.3.b, c - grey data**). Note, the forward and reverse exchange between the two conformations appear in the upper-left and lower-right quadrant of the spectra, respectively. Therefore, the forward and reverse reaction times (τ_f and τ_r) are independently resolved. However, as the two

cross-peaks are given by reciprocal TCFs in **Eq. 1**, the two functions are identical on the condition of microscopic reversibility.(18) Thus, for an equilibrium reaction the 2D-spectra are expected to be symmetric along the diagonal of the 2D spectra. This is indeed observed for the eGFP reaction (**Figure.5.4**).

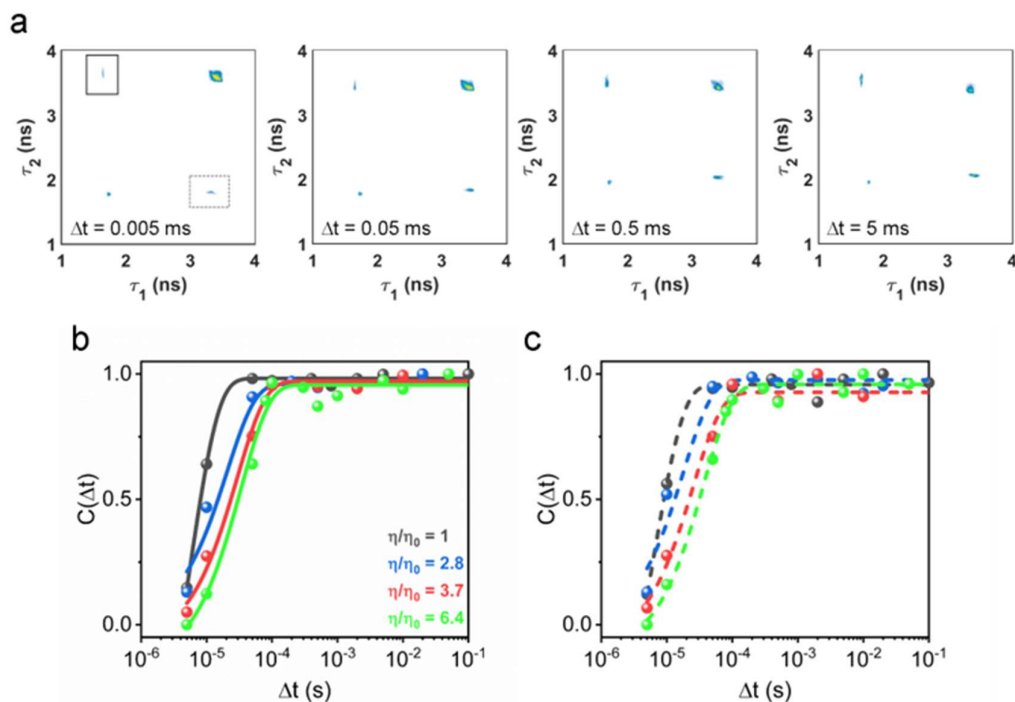


Figure 5.3. Reaction kinetics of eGFP chromophore pocket. (a) sm-2D-FLCS spectra series ranging from Δt of 0.005 ms to 5 ms. Forward exchange between the two bright states is observed as signal in the upper-left quadrant of the spectrum (highlighted in solid box), and reverse exchange is observed as signal in the lower-right quadrant of the spectrum (highlighted in dashed box). Exchange dynamics extracted from sm-2D-FLCS spectra cross-peak amplitudes for (b) forward and (c) reverse exchange as a function of solvent viscosity. The η_0 value is given by the viscosity of PBS buffer at 20°C. The exchange times decrease as the solvent viscosity increases, consistent with the internal reaction coordinate being coupled to the fluctuations in the bath.

We next study the reaction kinetics of eGFP as a function of solvent viscosity. Glycerol-buffer mixtures are used to systematically increase the viscosity of the solvent from 0.8 mPa·s (buffer only)

to 6.4 mPa·s (30% glycerol-buffer by volume). **Figure 5.3.b and 2c** show the exchange TCFs between the two conformations of eGFP's chromophore pocket as a function of solvent viscosity. This clearly shows that the exchange times increase as the solvent viscosity increases, indicating that the ms exchange dynamics are coupled to the environment. **Figure.5.5** shows the τ_{rxn} , determined by fitting the experimental data to exponential growth functions, as a function of normalized solvent viscosity $\frac{\eta}{\eta_0}$, where η_0 is the viscosity of PBS buffer at 20°C. The forward and reverse τ_{rxn} times are shown in solid and open symbols, respectively.

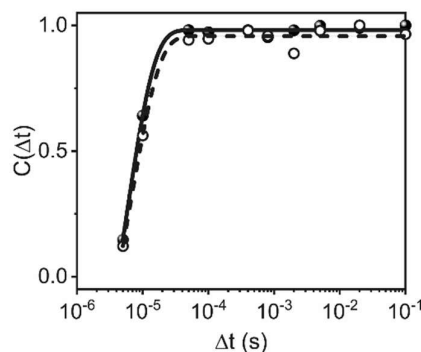


Figure 5.4. Forward and reverse exchange dynamics of eGFP in buffer. The forward and reverse reaction kinetics are extracted from opposite regions of the 2D spectrum. Under conditions of microscopic reversibility, the signal amplitudes are equivalent and the 2D spectrum is symmetric along the diagonal. As expected for equilibrium dynamics, the forward and reverse reaction kinetics are equivalent for the reaction occurring in the eGFP chromophore pocket.

5.4 Discussion

Typically, thermally activated barrier-crossing processes in the overdamped limit are described by Kramers' reaction rate theory (7, 21, 22). While it has not been demonstrated that the interior of a protein can be modelled as a thermal bath, this still provides a reasonable starting point to which we can compare the observed solvent-dependent friction effects (7). In the high friction regime, Kramers'

theory predicts t scales with ξ , where ξ is the friction of the solvent. Assuming a linear relationship between ξ and solvent viscosity η (Stokes' law), this gives a scaling of $\tau \propto \eta$. Expected Kramers' scaling is shown as a dashed line in **Figure.5.5**. Clearly, the observed dependence of the exchange time on the solvent viscosity deviates significantly from expected behavior at high viscosities, instead showing a power-law dependence of η^a with $a = 0.5$. Similar non-Kramers' kinetics have been observed in a range of protein dynamic processes, including folding kinetics (23, 24) and side-chain fluctuations (21). Indeed, in these experiments, a similar power-law scaling of exchange time with solvent viscosity was observed (23-25), where exchange times deviate significantly from Kramers' theory at high viscosity. These studies motivated the introduction of a friction term arising from the protein, termed 'internal friction', which serves to decrease the effective viscosity of the bath. The microscopic origins are considered to be fast, librational motions of sidechains that contribute to the barrier-crossing processes but are not coupled to the bath (23-27).

The molecular origin of internal friction has been attributed to the rough and hierarchical free-energy landscape typically associated with a protein (6, 10, 26, 27). It is common to consider motions along such a free-energy landscape as being partitioned into tiers (6): tier 0 transitions represent global conformational motions, tier 1 transitions represent intermediate motions within a given conformation, such as the rearrangement in the chromophore pocket of eGFP, and tier 2 motions. Tier 1 barrier crossings are thermally activated and are therefore expected to be coupled to the dynamics of the environment. However, in the high-viscosity regime, the system spends increasing amounts of time relaxing in local free-energy basins corresponding to tier 2 motions, which have been shown to be independent of solvent viscosity (4, 9). Deviations from Kramers' theory are therefore expected in the high viscosity regime, as the reaction time becomes more strongly governed by local fluctuations. A central assumption in Kramers' theory is that the bath fluctuations are Markovian, i.e., uncorrelated in time. However, in complex environments such as interior protein, relevant fluctuations may be correlated, and hence non-Markovian on the time scale of the reaction process. Work by Grote and Hynes (28) extended Kramers' rate theory to include non-Markovian fluctuations and suggested

pronounced deviations from Kramers' theory at high solvent viscosity, consistent with our observations.

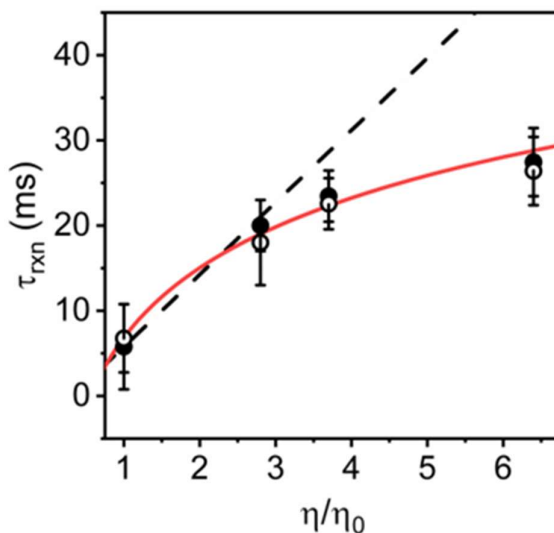


Figure 5.5. Deviation of protein reaction from Kramers' theory. Exchange times for forward (solid symbols) and reverse (open symbols) for the isomerization reaction in chromophore pocket of eGFP as a function of solvent viscosity. The dashed line shows the scaling expected from Kramers' theory. The exchange time deviates significantly at high viscosities, consistent with the picture of internal friction contributing to the reaction dynamics. The red line shows a power-law fit to the data.

5.5 Conclusions

In this work, we observe coupling between the microsecond reaction occurring in the chromophore pocket of eGFP and the surrounding solvent. The coupling is characterized by a power-law dependence of reaction time with solvent viscosity, with a scaling exponent of 0.5. The stark deviation from Kramers' behavior suggest weak coupling between the interior of the protein and the bath. The resulting picture is consistent with a rugged, hierarchical free-energy surface for protein motions where the coupling strength of the reactive process to the bath is proportional to the net displacement on the free energy surface, with weakly coupled local fluctuations and strongly coupled long-range motions.

REFERENCES

1. Austin, R. H., K. Beeson, L. Eisenstein, H. Frauenfelder, and I. Gunsalus. 1975. Dynamics of ligand binding to myoglobin. *Biochemistry* 14:5355-5373.
2. Beece, D., L. Eisenstein, H. Frauenfelder, D. Good, M. Marden, L. Reinisch, A. Reynolds, L. Sorensen, and K. Yue. 1980. Solvent viscosity and protein dynamics. *Biochemistry* 19:5147-5157.
3. Ebbinghaus, S., S. J. Kim, M. Heyden, X. Yu, U. Heugen, M. Gruebele, D. M. Leitner, and M. Havenith. 2007. An extended dynamical hydration shell around proteins. *Proceedings of the National Academy of Sciences* 104:20749-20752.
4. Fenimore, P. W., H. Frauenfelder, B. H. McMahon, and F. G. Parak. 2002. Slaving: solvent fluctuations dominate protein dynamics and functions. *Proceedings of the National Academy of Sciences* 99:16047-16051.
5. Frauenfelder, H., G. Chen, J. Berendzen, P. W. Fenimore, H. Jansson, B. H. McMahon, I. R. Stroe, J. Swenson, and R. D. Young. 2009. A unified model of protein dynamics. *Proceedings of the National Academy of Sciences* 106:5129-5134.
6. Frauenfelder, H., S. G. Sligar, and P. G. Wolynes. 1991. The energy landscapes and motions of proteins. *Science* 254:1598-1603.
7. Frauenfelder, H., and P. G. Wolynes. 1985. Rate theories and puzzles of heme protein kinetics. *Science* 229:337-345.
8. King, J. T., and K. J. Kubarych. 2012. Site-specific coupling of hydration water and protein flexibility studied in solution with ultrafast 2D-IR spectroscopy. *Journal of the American Chemical Society* 134:18705-18712.
9. Marques, B. S., M. A. Stetz, C. Jorge, K. G. Valentine, A. J. Wand, and N. V. Nucci. 2020. Protein conformational entropy is not slaved to water. *Scientific reports* 10:1-8.

10. Zwanzig, R. 1988. Diffusion in a rough potential. *Proceedings of the National Academy of Sciences* 85:2029-2030.
11. Arpino, J. A., P. J. Rizkallah, and D. D. Jones. 2012. Crystal structure of enhanced green fluorescent protein to 1.35 Å resolution reveals alternative conformations for Glu222.
12. Talele, S., and J. T. King. 2021. Fast and Robust 2D Inverse Laplace Transformation of Single-Molecule Fluorescence Lifetime Data. *bioRxiv*.
13. Ishii, K., and T. Tahara. 2013. Two-Dimensional Fluorescence Lifetime Correlation Spectroscopy. 1. Principle. *The Journal of Physical Chemistry B* 117:11414-11422.
14. Ishii, K., and T. Tahara. 2013. Two-Dimensional Fluorescence Lifetime Correlation Spectroscopy. 2. Application. *The Journal of Physical Chemistry B* 117:11423-11432.
15. Talele, S., and J. T. King. 2021. Fast and robust two-dimensional inverse Laplace transformation of single-molecule fluorescence lifetime data. *Biophysical Journal* 120:4590-4599.
16. Ghosh, A., S. Isbaner, M. Veiga-Gutierrez, I. Gregor, J. r. Enderlein, and N. Karedla. 2017. Quantifying microsecond transition times using fluorescence lifetime correlation spectroscopy. *The journal of physical chemistry letters* 8:6022-6028.
17. Royant, A., and M. Noirclerc-Savoie. 2011. Stabilizing role of glutamic acid 222 in the structure of Enhanced Green Fluorescent Protein. *Journal of structural biology* 174:385-390.
18. Talele, S., and J. T. King. 2021. Stability of a nonequilibrium biochemical cycle revealed by single-molecule spectroscopy. *BioRxiv*.
19. Kondo, T., J. B. Gordon, A. Pinnola, L. Dall'Osto, R. Bassi, and G. S. Schlau-Cohen. 2019. Microsecond and millisecond dynamics in the photosynthetic protein LHCSR1 observed by single-molecule correlation spectroscopy. *Proceedings of the National Academy of Sciences* 116:11247-11252.
20. Otsu, T., K. Ishii, and T. Tahara. 2015. Microsecond protein dynamics observed at the single-molecule level. *Nature communications* 6:1-9.

21. Hänggi, P., P. Talkner, and M. Borkovec. 1990. Reaction-rate theory: fifty years after Kramers. *Reviews of modern physics* 62:251.
22. Kramers, H. A. 1940. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* 7:284-304.
23. Ansari, A., C. M. Jones, E. R. Henry, J. Hofrichter, and W. A. Eaton. 1992. The role of solvent viscosity in the dynamics of protein conformational changes. *Science* 256:1796-1798.
24. Chung, H. S., and W. A. Eaton. 2013. Single-molecule fluorescence probes dynamics of barrier crossing. *Nature* 502:685-688.
25. Soranno, A., B. Buchli, D. Nettels, R. R. Cheng, S. Müller-Späh, S. H. Pfeil, A. Hoffmann, E. A. Lipman, D. E. Makarov, and B. Schuler. 2012. Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proceedings of the National Academy of Sciences* 109:17800-17806.
26. De Sancho, D., A. Sirur, and R. B. Best. 2014. Molecular origins of internal friction effects on protein-folding rates. *Nature communications* 5:1-10.
27. Portman, J. J., S. Takada, and P. G. Wolynes. 2001. Microscopic theory of protein folding rates. II. Local reaction coordinates and chain dynamics. *The Journal of Chemical Physics* 114:5082-5096.
28. Grote, R. F., and J. T. Hynes. 1980. The stable states picture of chemical reactions. II. Rate constants for condensed and gas phase reaction models. *The Journal of Chemical Physics* 73:2715-2732.

Chapter 6: Conclusions

6.1. General conclusions

In conclusion, this thesis provides a comprehensive insight into the microsecond to millisecond conformational changes of proteins and their impact on protein function. The use of single molecule multidimensional fluorescence spectroscopy (sm-2DFLCS) and the development of a fast and robust 2D Inverse Laplace Transformation algorithm (2D-ILT) allows for the resolution of distinct conformations of protein molecules based on different fluorescence lifetimes. The results of the study demonstrate that the conformational changes of proteins are crucial for their function and the quantification of these motions provides insight into their operating principles and mechanisms. The study also highlights the importance of considering the coupling between the protein's internal reaction coordinate and the surrounding solvent in understanding protein dynamics. The information obtained from these studies provides valuable insight into the operating principles and mechanisms behind the biological functions of proteins and advances our understanding of protein dynamics.

The study has successfully demonstrated that Single-molecule 2D-FLCS is a powerful tool for investigating the dynamics of biological macromolecules. The challenge of computing inverse Laplace transforms, which may have hindered its widespread use, has been overcome by presenting a fast and robust method based on singular-valued decomposition and Tikhonov regularization adopted from NMR spectroscopy. The approach allows for the efficient inversion of large, noisy datasets, common in single-molecule spectroscopy, while preserving high spectral resolution. Monte Carlo simulations of artificial photon streams confirmed the robustness of the technique in terms of spectral resolution, noise tolerance, and computational efficiency. This provides an alternative and easily implementable method beyond typical maximum entropy methods for performing Laplace inversions of single-molecule fluorescence data. The preliminaries for using this technique are presented in Chapter 2 and the stepwise implementation as well as detailed explanation of the core concepts are presented in Chapter 3.

Through Chapter 4, this thesis provides new insights into the mechanisms behind directional action

of biological machinery, particularly in the case of bacteriorhodopsin. Our findings demonstrate that microscopic irreversibility is a crucial operating principle for directional H⁺ transport in this optically driven pump protein. The results challenge the commonly held view of detailed balance and time-reversal symmetry in chemically driven pumps and highlight the uniqueness of the thermal destabilization of bacteriorhodopsin's reaction cycle, which is inherent to its power-stroke mechanism. This study sheds light on the complex interplay between thermodynamics and protein function and opens up new avenues for further investigation.

The results of 2D-FLCS analysis of EGFP in Chapter 5 indicate a strong coupling between the reaction within the chromophore pocket of eGFP and its surrounding solvent. The observed power-law dependence of reaction time with solvent viscosity, with a scaling exponent of 0.5, is a deviation from Kramers' behavior and suggests a weak coupling between the interior of the protein and the bath. Our findings support the hypothesis of a rugged, hierarchical free-energy surface for protein motions, where the coupling strength of the reactive process to the bath is proportional to the net displacement on the free energy surface, with weakly coupled local fluctuations and strongly coupled long-range motions.

6.2. Future Outlook

The logical steps forward from this study is to further develop the experimental setup such that one can excite at multiple wavelengths. This will enable us to observe the complete photocycle of bacteriorhodopsin and improve the quality of current findings. In this technique of 2D-FLCS one determines physical quantities using lifetime-lifetime correlations however, in principle the same approach can be used to obtain meaningful insights by observing correlations of other photophysical properties of a sample such as emission spectra, fluorescence anisotropy, etc.

In order to further investigate protein bath coupling and to uncover the underlying reason for deviation from Kramer behavior, one can introduce point mutations in EGFP outer structure without interfering the internal chromophore dynamics. We hypothesize that the residues adjacent to the Glu 222 residue might have a role in the protein-bath coupling of internal reaction as they are exposed to the outside solvent while the Glu 222 is a bridge between them but faces towards the chromophore

pocket inside. Thus, we suspect that mutating these residues (L221 and F223) with heavier or lighter amino acid side chains of the same type should provide deeper insights for understating the peculiar deviation from Kramers behavior.

Since the techniques allows us to augment data by simply adding the emission delay correlation histograms, given that sufficient amount of data is collected, one can investigate the presence and abundance of rare conformational states in a system. This technique exploits changes in lifetime of a molecule to predict conformational change. Since the main reason for changing lifetime of a molecule is change in its surrounding electronic structure, in principle, any process that exposes a fluorophore to different environments while undergoing conformational change can be studied using 2D-FLCS. Some of the potential experiments involve examining the dynamics of DNA base flipping in case of base-pair mismatch and its role in DNA repair mechanism, observing thermodynamics of heat shock proteins and their chaperone action at different temperatures.

The work presented here is an example of how development of new techniques opens up new avenues for investigating new and fundamental workings of biological processes. This is especially true for the study of single-molecule dynamics, which is crucial for understanding the underlying mechanisms of biological processes at the molecular level. The fact that this technique is easily implemented, robust in terms of spectral resolution, noise tolerance, and computational efficiency makes it an attractive alternative for performing Laplace inversions of single-molecule fluorescence data. The result of this study shows the potential of obtaining wide range of understanding of biological processes by developing novel experimental and analysis techniques.

Acknowledgements

Undergoing a PhD is a challenging and rewarding journey that would not have been possible without the support and guidance of many people. I would like to express my sincere gratitude to all of them.

First and foremost, I would like to thank my supervisor, Professor John King, for his invaluable advice, encouragement, and feedback throughout all our projects. He has been a great mentor and a source of inspiration for me, given me the freedom to pursue my own ideas and the confidence to overcome the difficulties along the way. I am indebted to my colleagues and friends at IBS CSLM and UNIST for their support and friendship. They have created a stimulating and enjoyable research environment that has fostered my intellectual growth and personal development.

Finally, I would like to dedicate this thesis to my parents and my sister, who have always been there for me with their unconditional love and encouragement. I am grateful for their sacrifices, guidance, and support throughout my life. They have taught me the values of hard work, perseverance and honesty that have shaped me as a person and a researcher.

To all of you, thank you from the bottom of my heart.

