NOVEL TECHNIQUES FOR CHARACTERIZATION OF BIOMOLECULAR BEHAVIOR ACROSS MULTIPLE BIOLOGICAL SCALES

A Thesis

Submitted to the Faculty

of

Purdue University

by

Tyler VanDyk

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

Dec 2020

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF THESIS APPROVAL

Dr. Tamara Kinzer-Ursem, Chair

Weldon School of Biomedical Engineering

Dr. David Umulis

Weldon School of Biomedical Engineering

Dr. Fang Huang

Weldon School of Biomedical Engineering

Approved by:

Dr. George R. Wodicka

Head of the Weldon School of Biomedical Engineering

This work is dedicated to the my friends, lab mates, advisors, and loved ones.

Generic meaningful phrase of appreciation.

ACKNOWLEDGMENTS

I would like to thank first and foremost Dr. Tamara Kinzer-Ursem, for her gracious support, advise, and mentorship during my Purdue career. I would also like to thank my other committee members: Dr. David Umulis for the intellectually stimulating conversations and courses I've shared with him (and the occasional head-nod as I pass by his office), and Dr. Fang Huang for his kindly guidance in my initial efforts in the field of FCS and the expertise in microscopy he offered. I deeply appreciate the effort each of you has invested in my studies, and am proud to have worked with you in my capacity as a graduate student.

I thank also my lab mates and coworkers that have provided a constant supply of scientific and personal guidance throughout my years in the Kinzer-Ursem Lab group. In particular, I would like to thank Dr. Matthew Pharris and Dr. Scott Bolton for their mentorship in science and willingness to go out of their way to teach me new skills. I would also like to specially thank Dr. Aya Saleh for the endless effort she invested in data collection to inform models presented in this work, and for generally being one of the most impressively driven scientists I've had the pleasure to work with. I've also felt very fortunate to work closely alongside Dr. Kaisa Ejendal, Dr. Gunjan Saini, Peter Brumm, Barrett Davis, Patrick Giolando, Neil Patel and Eugene Kim. I've learned so much from each of you and wish you all the best in all future endeavors.

I would also like to thank Dr. Charles Babbs for first inspiring me to pursue a career in medicine. His educational prowess and optimistic attitude are an inspiration to me. I deeply appreciate the lessons and mentorship he has offered, as well as his frequent and friendly invitations to enjoy a game of pickup soccer.

Lastly, I would like to thank Purdue University for a wonderful educational experience including all the facilities, funding, and opportunities that I've had access to as a student at this institution.

TABLE OF CONTENTS

Page

LIST OF TABLES . . . viii

LIST OF FIGURES . . . ix

ABSTRACT . . . xiv

1 INTRODUCTION . . . 1

1.1 Significance of Biomolecular Study . . . 1

1.2 Microscale: Protein-protein Interactions (PPIs) . . . 2

1.2.1 Background and Significance . . . 2

1.2.2 Survey of Current PPI Measurement Techniques . . . 8

1.2.3 Observed Gap in PPI Measurement Technology . . . 10

1.3 Macroscale: Dynamics of the Proteome and Metabolome . . . 11

1.3.1 Background and Significance . . . 11

1.3.2 Survey of Current Metabolic Labeling Techniques . . . 13

1.3.3 Observed Gap in Proteome Labeling Techniques . . . 14

2 FLUORESCENCE ROLLING CORRELATION SPECTROSCOPY (FRCS) 16

2.1 Introduction . . . 16

2.1.1 Background and Significance . . . 16

2.1.2 Principles of Diffusometry . . . 18

2.1.3 Principles of FCS . . . 20

2.2 Methods . . . 32

2.2.1 FRCS Algorithm and GUI Design . . . 32

2.3 Results . . . 33

2.3.1 FRCS Theory . . . 33

2.3.2 Software Correlation . . . 40

2.3.3 Python 3.6 Implementation of FRCS . . . 42

2.3.4 FRCS Data Fitting Algorithm and Workflow . . . 49

2.3.5 Graphical User Interface for FRCS Experiments . . . 59

Page

2.4 Conclusions 60

3 PRELIMINARY SIMULATIONS AND EXPERIMENTAL FINDINGS OF FCS/FRCS 63

3.1 Introduction 63

3.1.1 Background and Significance 63

3.1.2 Fluorescent Labeling of Proteins 64

3.2 Methods 66

3.2.1 Spatial Stochastic Simulations of FRCS Data 66

3.2.2 FCS Measurement of Protein Binding Interaction 66

3.3 Results 71

3.3.1 Spatial Stochastic Simulations of Binding Reactions 71

3.3.2 FCS Studies of CaM-CaN binding 83

3.3.3 Preliminary FRCS of CaM-mCaMKII 86

3.4 Conclusions 86

4 KINETICS OF AZIDOHOMOALANINE BIODISTRIBUTION AND IT'S METABOLIC IMPLICATIONS *IN VIVO* 88

4.1 Introduction 88

4.2 Methods 90

4.2.1 Animal Model 90

4.2.2 Aha Injection, and Plasma and Tissue Collection 90

4.2.3 Sample Preparation for Aha Analysis 91

4.2.4 LC-MS Targeted Analysis of Aha 92

4.2.5 Western Blot Analysis of Aha-labeled Tissues 93

4.2.6 Plasma Sample Preparation for Untargeted Metabolomic Analysis95

4.2.7 LC-MS Untargeted Metabolomic Analysis 95

4.2.8 Metabolomics Data Analysis 96

4.2.9 Kinetic Modelling of Aha distribution 96

4.2.10 Model Validation and Sensitivity Analysis 96

4.3 Results 97

Page

4.3.1 Kinetic Model of Aha Distribution 97
4.3.2 Kinetic Model of Protein Labeling 99
4.3.3 Estimation of Relative Protein Synthesis and Turnover Rates . 101
4.3.4 Model Validation 103
4.3.5 Predictive Simulations of Dosing Paradigms 105
4.4 Conclusions 107
5 RECOMMENDATIONS 108
5.1 Implications of FRCS 108
5.1.1 Applications and Limitations 108
5.1.2 Future Work 111
5.2 Implications of Aha Labeling 114
5.2.1 Applications and Limitations 114
5.2.2 Future Work 115
5.3 Final Thoughts 116
REFERENCES 118
A Appendix: Code 129
A.1 FRCS Scripts and GUI 129
A.2 Aha Models and Scripts 129
B Appendix: Equations 130
B.1 System of ODEs for Aha Model 130
B.2 All Listed Equations 132
C Appendix Figures 141
D Appendix: Tables 144

LIST OF TABLES

Table Page

4.1 Multiple reaction monitoring table for data acquisition. 93

4.2 Tissue relative protein synthesis rates and protein half-lives. *Relative synthesis rate = $k_f k_s$, units include fluoresence response units (RU) and mass pAha/fAha ($\mu g_p/\mu g_f$) 102

4.3 Parameter and goodness-of-fit statistics related to alternative dosing paradigm predictive models 106

5.1 Binding reaction systems that are good targets of future study with FRCS. 112

D.1 FRCS simulation parameters, determined kinetics from FRCS fitting, and goodness of fit statistics for all studies discussed in chapter 3. Seeds 1-39 are represented in the plots shown in figure 3.5 and results from seed 100-110 are displayed in figures 3.2-3.4. $S_{F,x}$ is a standard deviation of the residuals of complex composition fraction (F_{AB}) from the best fit kinetic binding curve and is shown because R^2 is poor fitting statistic for non-linear regression. 144

D.2 Distribution model fitting parameters including best fit values, fitting ranges, standard error of fit, PRCC values and informing literature. Parameters without published literature values are sourced from this study (TS) either from an experimental measurement or best estimate initial value that produced a single time-step Δ[Aha] that was one order of magnitude lower than the maximum recorded experimental value and swept upward and downward 1.5 orders of magnitude for fitting. 145

D.3 Distribution model static parameters including model value, references utilized for scaling and PRCC values. All mass parameters are sourced from literature and experimental measurements from this study (TS) and were held constant to generate an "average" behavior according to our murine animal model. Compartment plasma masses are parameterized from average mass of plasma per unit mass tissue and appropriately scaled to match tissue mass utilized in this model for measurements of Aha. . . 146

D.4 Incorporation model fitting parameters including best fit values, fitting ranges, standard error of fit, PRCC values and informing literature. Parameters without published literature values are sourced from this study (TS) either from an experimental measurement or best estimate initial value that produced a single time-step Δ[Aha] that was one order of magnitude lower than the maximum recorded experimental value and swept upward and downward 1.5 orders of magnitude for fitting. 147

LIST OF FIGURES

Figure Page

2.1 A cartoon showing the reaction of two proteins A and B into complex AB. Each species has a unique size and therefore a unique diffusion. 19

2.2 Left: Experimental setup for FCS. Laser (yellow) is passed up through the objective focused to a point within a desired protein sample. Excited fluorophores emit photons captured by the same objective (red), filtered and passed through a confocal pinhole before being captured by the photon counter. Right: freely diffusing fluorescent particles enter the CV and become excited, emitting photons for counting. 21

2.3 Top: Cartoon showing possible paths of fluorescent particles through an XY cross-section of the CV. For stable diffusion, the particle path determines residency time within the CV. Bottom: example photon count intensity trace, $I(t)$ and autocorrelation, $G(\tau)$. Shown also are possible observed lag times for the color coordinated particles in the top panel (not to scale). 23

2.4 Single labeled binding measurable with one-photon FCS 34

2.5 Expected correlation geometry of a two particle fitting model over the timecourse of a binding reaction between a fluorescently labeled ligand and a larger inert protein. Characteristic residency times are τ_{FAST} and τ_{SLOW} for the free and bound ligand respectively, and similarly the fraction of each species present is indicated by F_{FAST} and F_{SLOW}. Note that for a reversible binding interaction full saturation is unlikely to occur with equivalent concentrations of ligand and receptor. 35

2.6 Simulated binding interactions of a fluorescent ligand fA, to a inert receptor B over a time course of minutes. Right panels show the same time course in log scale, grid lines indicate a 30 second period (as subsequent FCS measurements might capture). Plotted is the fraction of complex fAB using kinetics and concentrations as observed by Schuler et al. [49]. 36

2.7 Simulated binding interactions of a fluorescent ligand fA, to a inert receptor B over a time course of minutes. Right panels show the same time course in log scale, grid lines indicate a 30 second period (as subsequent FCS measurements might capture). Plotted is the fraction of complex species fAB using various kinetics and concentrations reasonable for PPIs 39

2.8 Block Diagram of a multiple-tau scheme that applies the BM Correlation approach. 42

Figure Page

2.9 Benchmarking of FRCS algorithm implemenation in Python 3.6. On the x-axis is the width in seconds of the correlation frame, on the y-axis is the average evaluation time for a frame of each width selected randomly from within several seeds of simulated experimental measurements and normalized by average photon count rate for each simulation (n = 10: 2 frames/seed x 5 seeds/photon count rate; 3 simulated photon count rates). 48

2.10 Example of fluorescent bleaching from a preliminary experimental measurement of freely diffusive Alkyne-Alexafluor 647nm. Shown is the photon count rate, declining between linearly spaced correlation frames (n=20) during a 10 minute measurement. Note that Alexafluors are engineered to exhibit robust bleaching resistance but this timescale may be significantly shorter for other fluorophores. 52

2.11 FRCS Evaluation of a simulated binding interaction between two proteins. In this figure reaction time is defined as the left edge of each correlation frame of width 20 s. Free ligand introduced to system at reaction time t = 0, receptor introduced at t = 30 s. Top: All FRCS generated correlation frames. Middle: Fit correlation traces for each frame using fitting model described in equation 2.27. Bottom: Resulting fractional composition of complex as a function of reaction time. Simulated binding data generated as discussed in chapter 3 for a binding interaction with k_a = 2e6 $M^{-1}{\cdot}s^{-1}$, k_d = 5e−2 s^{-1}, $[A]_0 = [B]_0$ = 10 nM, $D_{c,A}$ = 1.5e−10 $m^2{\cdot}s^{-1}$, and $D_{c,B} = D_{c,AB}$ = 0.75e−10 $m^2{\cdot}s^{-1}$ 54

2.12 GUI Correlation Tab, contains fields for the source/destination directories and desired correlation parameters as well as organizing functionality for experimental measurement files. 60

2.13 GUI Fitting Tab, (a) contains fields for correlation file directory as well as all fitting parameters including initial values, fitting constraints and variance boolean check boxes. Also shows (b) graphical and tabular representation of best fit parameters. 61

3.1 A spatial stochastic simulation of an FCS experiment. Shown are three progressive time points from a simulated binding reaction of a small fluorescent ligand and large inert receptor as in equation 2.20. Small icospheres (blue, grey, yellow) display particle positions, the large green icosphere shows a transparent border defining the CV, and the large wire-mesh show the diffusion boundaries of the simulated solution droplet. 72

Figure Page

3.2 FRCS Evaluation of a simulated binding interaction between two proteins. Shown are data for all seeds from the same simulation (and plotted on the same axes) as displayed in figure 2.11. Free ligand introduced to system at t = 0, receptor introduced at time t = 20 s. Left: FRCS generated correlation frames with frame width of 20 seconds. Middle: Fit correlation traces for each frame using fitting model described in equation 2.27. Right: Resulting fractional composition of complex as a function of reaction time ([fAB](t_{rxn}). Simulated binding data generated for a binding interaction with k_a = 2e6 $M^{-1}{\cdot}s^{-1}$, k_d = 5e−2 s^{-1}, $[A]_0 = [B]_0$ = 10 nM, $D_{c,A}$ = 1.5e−10 $m^2{\cdot}s^{-1}$, and $D_{c,B} = D_{c,AB}$ = 0.75e−10 $m^2{\cdot}s^{-1}$ 74

3.3 FRCS simulation with and without an adjustment $t_{rxn,f}$. Top: [fAB](t_{rxn}) generated from the data in figure 3.2 and best fit binding kinetic curve (k_a = 1.61e6 $M^{-1}{\cdot}s^{-1}$, k_d = 6.99e−2 s^{-1}) of all seeds (n = 5) without a dynamic $t_{rxn,f}$ ($t_{rxn,f} = t_{m,f} + \Delta t_f$), Bottom: [fAB]($t_{rxn}$) and best fit curve ($k_a$ = 1.70e6 $M^{-1}{\cdot}s^{-1}$, k_d = 7.03e−2 s^{-1}) with a dynamic $t_{rxn,f} = t_{m,f} + \lambda_f \Delta t_f$. Error bars show standard deviation of determined [fAB] across seeds. 76

3.4 FRCS simulation with and without experimentally limited reaction start time. Top: Determined [fAB](t_{rxn}) and best fit curve (k_a = 1.70e6 $M^{-1}{\cdot}s^{-1}$, k_d = 7.0e−2 s^{-1}) of all seeds (n = 5) with a simulated release event during the experimental time course. Right: Determined [fAB](t_{rxn}) and best fit curve of all seeds (n = 5) with an experimentally mimicked mixing prior to the experimental time course (k_a = 1.64e6 $M^{-1}{\cdot}s^{-1}$, k_d = 4.8e−2 s^{-1}). In both systems, species B was added to the system at $t_{rxn} = 0s$, and conditions were otherwise identical to the simulation data displayed in figures 3.2/3.3. Error bars show standard deviation of determined [fAB] across seeds. 78

3.5 Points indicate a single frame from a simulated measurement (n=3), solid trace indicates the non-linear regression best fit, and the dashed trace indicates the predicted reaction kinetics given input parameters. Parameters studied, top to bottom: (i) the diffusivity of the complex species fAB in comparison to a fixed diffusivity of fA (left-to-right: $D_{c,AB} = 3.0\text{e}{-10}, 2.0\text{e}{-10}, 1.0\text{e}{-10}$ m^2s^{-1}, $D_{c,A}$ = 4.0e−10 m^2s^{-1}), (ii) the initial concentration of species B ($[B]_0 = 5, 10, 20$ nM) with a fixed initial concentration of fA ($[fA]_0$ = 10 nM) (iii) the binding affinity, K_D, of the reaction by sweeping k_a ($k_a = 5\text{e}7, 5\text{e}6, 5\text{e}5$ $M^{-1}s^{-1}$) with a fixed k_d (k_d = 5e−3 s^{-1}), and (iv) the binding rate k_a ($k_a = 5\text{e}8, 5\text{e}7, 5\text{e}6$ $M^{-1}s^{-1}$) with a fixed K_D ($k_d = 5\text{e}{-1}, 5\text{e}{-2}, 5\text{e}{-3}$ s^{-1}). Unless otherwise specified, other parameters were held constant to standard values as defined in methods. 82

Figure Page

3.6 Diffusivity of particles characterized via FCS using a single particle fitting model. Unligated alexafluor was used as a calibration and fluorescently labeled CaM was measured in the presence and absence of Ca^{2+} and CaN . 84

3.7 Serial dilution of CaN associating with a fixed 40nM [CaM] measured via FCS. Notably, as the [CaN] increases, the diffusion of CaM appears to slow substantially, indicating generation of complex protein 85

3.8 Purified CaM ligated to Alexa Flour 647 and isolated via a 7K MWCO desalting column. Two panels show the same SDS-page gel imaged with fluorescence and coomassie staining. As evident, no significant amounts of protein contaminants are present, however some contaminant fluorescent species with lower MW than CaM (16.8 kDa [98]) become evident when imaged at 647 nm. From left to right, lanes contain a MW ladder, a blank, and 6 lanes with 20 μL dilutions of desalted A647-CaM (2 lanes each 3:1, 2:2, and 1:3 of A647-CaM desalted reaction mixture into CaM-CaN Binding Buffer). 86

4.1 Biodistribution of fAha via transport and exchange. Introduced at a distinct injection site, fAha is allowed to enter the venous circulation (blue). From the venous circulation fAha is driven through the arterial system (red) into distinct tissue compartments where exchange occurs at tissue specific rates. Arrows indicate directional movement of fAha, ellipses indicate distinct compartments of plasma available for surface exchange, and rectangles indicate cumulative intracellular volume for each tissue in the model. Three additional terms exist for elimination of fAha via excretion and metabolization. 98

4.2 Aha distribution and protein labeling given a dosing paradigm identical to the experimental methods (0.1 mg Aha per g body weight, injected subcutaneously) Top: [fAha] in each tissue over time. Bottom: rF signal in each tissue as a function of the [fAHA]. Filled points indicate mean experimental measurement at each time point, error bars indicate experimental standard deviation. Traces indicate best fit of model to each dataset, with darker and lighter shaded regions showing 95% prediction intervals for residual error based off the mean experimental value and all experimental values, respectively. 102

Figure Page

4.3 Model accurately and precisely predicts relative labeling in the brain and liver with multiple injection doses. Left: rF experimental data and fit model for 12hrrd. Right: rF experimental data and fit model for 12hrrd. Color indicates liver (red) and brain (blue) tissues, solid line shows best fit of model from original robust dataset, dashed line shows refit model to repeated dose paradigm. Dots indicate individual experimental measurements and crosses represent experimental means for data collected 6, 18, and 32 hpi. 105

5.1 A cartoon diagram of a chip enabling FRCS with a 'mid-measurement' mixing of reagents. This chip allows simultaneous injection and mixing of two solutions at some time $t = t_{rxn,0}$ (injection flow direction described by blue arrows), into a static reservoir during an active measurement started at $t = 0$. Note the CV and particles are not drawn to scale. 113

C.1 Temperature change of a small 15uL sample as it is moved first from an Ice Bath (4C) to RT, then measured with FCS under a 633 He-Ne laser at RT for 60 s. Temperature recorded using a thermocouple that samples at 1Hz. 142

C.2 Binding of N3CaM-647 to various concentrations of mCaMKII in the presence and absence of Ca^{2+}. Binding reactions were measured in duplicate for 120 s and correlated into 50 frames of width 25 s, spaced linearly over the timecourse of measurement. Correlated frames were fit with according to calibrated measurements of the diffusion time of CaM0/4 (382/437 µs) and CaM-CaMKII (463µs). Notably, these calibrated diffusion times do not differ by a factor of least 1.6 as desired for FCS fitting analysis [26]. This may explain very noisy measurements that are not altogether firmly aligned with the expected fitting model behavior. All binding reactions performed in the same buffer (20 mM Hepes (pH 7.5), 100 mM NaCl, 1 mg/mL BSA, 1mM EGTA, 1mM EDTA, $< 2\%$ Glycerol) supplemented with either 12 mM Ca^{2+} or an additional 2 mM EGTA for CaM4 and CaM0 measurements respectively. 143

ABSTRACT

The development of the central dogma of molecular biology in the late 1950's unlocked a now booming field of medical innovation centered about the study of biomolecules. Of these, proteins play perhaps the most active role, acting as molecular devices to carry out dynamic processes of cellular function and life maintenance including energy metabolism, DNA replication, cellular signaling, and many others. Historically, proteins have been studied in small subsets; however, there is a recent paradigm shift toward studying proteins and their interactions within complex, physiologically relevant conditions. This shift is driven by advances in both sequencing and computing technologies that enable novel approaches in integrative biology. Here, I present two such techniques that combine computational and experimental methods to elucidate and quantify protein behavior across multiple biological scales.

The first is a tool to measure protein-protein interaction (PPI) kinetics. Fluorescence Rolling Correlation Spectroscopy (FRCS) is a novel algorithm and software that applies principles of diffusometry to enable facile characterization of kinetics across the entire dynamic range of typical PPIs [$k_a \epsilon (1e4, 1e6)$ $M^{-1} \cdot s^{-1}$, $k_d \epsilon (1e{-}4, 1e1)$ s^{-1}]. The second is a kinetic model of non-canonical amino acid proteome labeling. I present a system of ordinary differential equations describing Azidohomoalanine (Aha) distribution and selective labeling of nascent proteins *in vivo* for either enrichment or imaging. This model demonstrates accurate predictions of labeling across multiple tissues and timescales. Each of these tools is presented with open source software and models to enable future work in the study of protein behavior.

1. INTRODUCTION

1.1 Significance of Biomolecular Study

The discovery and development of what is now known as the central dogma of molecular biology in the late 1950's unlocked what is now a booming field of biological and medical innovation centered about the study of biological molecules [1–3]. These molecules, deoxyribonucleic acid (DNA), ribonucleic acid (RNA), proteins and carbohydrates, compose the molecular machinery of biology and function together in innumerable patterns of behavior that are required to sustain life. Of these, proteins play perhaps the most active role in the exertion of biological function. Acting as molecular devices, proteins carry out the vast majority of dynamic processes of life maintenance and cellular function including energy metabolism, DNA replication, transcription/translation of RNA, cellular signalling, transport, and many others [4].

Despite this great diversity of function, and the fact that the human genome alone contains genetic coding of somewhere around 20,000 unique proteins, proteins are also remarkable in that they are all composed of some combination of the same 20 canonical amino acids described by the codons of the universal genetic code [4–6]. Each amino acid (AA) is possessed of a carboxylic acid and amine group, linked by a central carbon, and these two reactive groups covalently bind in complimentary pairs to form the amide backbone of the protein. In addition to these common groups, each of the 20 canonical AAs contain a unique side chain, often called the -R group, that imparts distinct physical properties (including mass, charge, hydrophobicity, and shape) to the overall composition of the protein. The cumulative influence of these properties allows each protein to display unique folding, dynamics and interactions with other biomolecules [5]. To truly understand these molecules therefore requires contributions from many classical sciences including biology, physics, chemistry and mathematics as well as from newer fields of computer science and informatics.

Indeed proteins have historically been studied relatively independently. Early efforts in the field of protein biochemistry generally consisted of identifying a coding sequence for a protein of interest, cloning it into an expression vector and purifying a quantity sufficient for analysis. These purified proteins were then studied with a number biochemical assays or structurally imaged with tools such as NMR spectroscopy or X-ray crystallography [7]. In the early 90's however, there was a paradigm shift toward studying larger sets of proteins and the interactions of proteins within more complex, physiologically relevant systems. This shift was in large part driven by advances in sequencing and computing technologies that made the study of entire genomes possible [8]. These same technologies paved the way for the development of many powerful tools to further in-depth investigation of protein behavior, including bioengineering of synthetic proteins, peptides and labels.

With the development of these tools and technologies, came new approaches and techniques in integrative biology that sought to describe protein behavior at both micro- and macro-scopic scales. By combining computational and experimental methodology, proteins can now be studied with resolution previously unattainable, regardless of the scale of study. In this work, I present two such techniques that utilize an integrative approach with elements of both computational and experimental biology to elucidate and quantify protein behavior across multiple biological scales.

1.2 Microscale: Protein-protein Interactions (PPIs)

1.2.1 Background and Significance

When considering the dynamic function of proteins, two sub-classes are most prominent: enzymatic and modulatory proteins. Enzymes are biocatalysts, organic molecules that act upon the activation energy of biochemical reactions, driving them forward at a faster rate. While any given biochemical reaction may be capable of proceeding in the absence of enzymatic activity, enzymes allow these reactions to proceed in complex, life sustaining environments [9, 10]. While not all enzymes are

proteinous - notably some RNA molecules called ribozymes demonstrate important enzmyatic activity during gene expression - proteins compose the vast majority of enzyme function [11]. As catalysts, enzymes are typically not modified during the course of a reaction, and are therefore able to exist in low concentrations while maintaining their key function within the cell. Modulatory proteins, conversely, help to regulate the function of enzymes, by either promoting or inhibiting their processes depending on the status of other cellular systems. This regulation is a key component of cellular function, allowing organisms to perform a variety of tasks effectively and to respond to their environment.

A typical enzyme catalyzed reaction can be represented as 3 distinct events: a reversible binding event, an irreversible catalysis event, and a reversible release event.

$$S + E \rightleftharpoons ES \rightarrow EP \rightleftharpoons E + P \tag{1.1}$$

Here S denotes some substrate, P the reaction product, E the enzyme, and ES/EP represent a complex of the enzyme with the substrate and product respectively. Each arrow represents a reaction step with a unique activation energy, and thus a unique rate at which the step will proceed. Typically, the reaction product has a significantly reduced binding affinity with the enzyme, such that the EP complex is transient and unlikely to persist. Biologically speaking, this is an advantageous reality that optimizes the available concentration of enzyme available for substrate binding. As such, the catalysis and product release events can usually be represented as a single irreversible process.

$$S + E \rightleftharpoons ES \rightarrow E + P \tag{1.2}$$

This model of enzyme kinetics is called the Michaelis-Menten model after the authors who first published it in 1913 [12–14]. In this model, the velocity of product output is modulated according to the laws of mass action; velocity can therefore be

described by the concentrations of each species present in the system, the rates of binding/release of substrate to the enzyme (k_a and k_d respectively), and the lumped catalysis and product release rate (k_{cat}). The rate of change of each species over time can be classically represented with a mass balance system of ordinary differential equations (ODEs) that incorporate these rates.

$$\frac{d[S]}{dt} = k_d[ES] - k_a[E][S] \tag{1.3a}$$

$$\frac{d[E]}{dt} = (k_d + k_{cat})[ES] - k_a[E][S] \tag{1.3b}$$

$$\frac{d[ES]}{dt} = k_a[E][S] - (k_d + k_{cat})[ES] \tag{1.3c}$$

$$\frac{d[P]}{dt} = k_{cat}[ES] \tag{1.3d}$$

Characterizing these rates of reaction can yield valuable information about the roles played by enzymatic processes. For instance, elevated rates may indicate more rapid processes such as signaling, while slower rates may be more prevalent in basal processes such as cellular restructuring [15–19]. In most cases however, the true kinetics of a dynamic system are difficult to capture due subsecond time scales. Typically, these rates are measured as some lumped constants that are easier to characterize with experimental methods. Protein binding for instance is often represented and measured as a binding affinity constant (K_D) rather than as distinct binding association and dissociation parameters. Similarly, the Michaelis-Menten model relies on a few assumptions to relate the velocity of product production as a function of substrate and enzyme initial concentrations, which can be measured by any number of activity assays.

$$\frac{d[P]}{dt} = V_{max}\frac{[S]}{K_M + [S]} = k_{cat}[E]_0\frac{[S]}{K_M + [S]} \tag{1.4}$$

Here the rate of production is defined by the maximum velocity of reaction ($V_{max} = k_{cat}[E]_0$, where $[E]_0$ is the initial enzyme concentration), some current substrate concentration ($[S]$), the rate of catalysis (k_{cat}), and a lumped constant called the Michaelis-Menten constant (K_M) which is a parameter that is similar to the binding affinity of the ES complex (K_D), but additionally accounts for catalysis.

$$K_D = \frac{[E]_{eq}[S]_{eq}}{[ES]_{eq}} = \frac{1}{K_{eq}} = \frac{k_d}{k_a} \tag{1.5a}$$

$$K_M = \frac{k_d + k_{cat}}{k_a} \tag{1.5b}$$

However, this model of enzymatic activity is relatively simple in that it assumes the enzymatic reaction defined in eq. (1.2) exists in isolation. In reality, most *in vivo* reactions occur in a more complex system, including some external modulation of the reaction. These processes are often driven by regulatory proteins, which can either act as a competitor for the substrate binding domain or act allosterically upon enzymes to promote or inhibit their activity. The introduction of such a species complicates the enzymatic reaction. Take for instance, a system with a competitive inhibitor (I) that binds to enzyme E and disallows it's interaction with the substrate by creating a persistent complex obstructing the substrate binding domain.

$$S + E \rightleftharpoons ES \rightarrow E + P \tag{1.2}$$

$$I + E \rightleftharpoons EI \tag{1.6}$$

In this modified model, we can think of the enzyme in three potential states: free for substrate binding (E), unavailable due to complex with competitor (EI) or unavailable due to complex with substrate (ES). This behavior can be relatively easily modeled by adding a few new equations in our non-linear system of ODEs to capture these dynamics.

$$\frac{d[S]}{dt} = k_d[ES] - k_a[E][S] \tag{1.7a}$$
$$\frac{d[I]}{dt} = k_{d,I}[EI] - k_{a,I}[E][I] \tag{1.7b}$$
$$\frac{d[EI]}{dt} = k_{a,I}[E][I] - k_{d,I}[EI] \tag{1.7c}$$
$$\frac{d[E]}{dt} = (k_d + k_{cat})[ES] + k_{d,I}[EI] - k_a[E]([S] + [I]) \tag{1.7d}$$
$$\frac{d[ES]}{dt} = k_a[E][S] - (k_d + k_{cat})[ES] \tag{1.7e}$$
$$\frac{d[P]}{dt} = k_{cat}[ES] \tag{1.7f}$$

Here additional terms have been added to account for binding of I to E with corresponding association and dissociation rates $(k_{a,I}, k_{d,I})$. The Michaelis-Menten model can also be modified to accommodate this behavior by adjusting the Michaelis-Menten constant to account for competitive inhibition.

$$\frac{d[P]}{dt} = k_{cat}[E]_0 \frac{[S]}{K_{M,I} + [S]} \tag{1.8}$$
$$K_{M,I} = K_M \left(\frac{1 + [I]}{k_{d,I}} \right) \tag{1.9}$$

With only a single regulator in the system, it is still reasonable to characterize this system by an experimental assay, particularly if the original K_M has already been characterized. However, this model tends to become less valuable as more regulatory species are added to the system, as it becomes harder to distinguish the effects of unique species on the system by experimental methods.

Indeed, while experimental tools are viable for studying the interactions of small sets of proteins, it is much more difficult to distinguish the influence of single species in more complex environments such as *in vivo*. While certainly it is certainly possible to focus on a single protein species *in vivo*, to do so often requires a robust experimental

design involving engineered controls and a variety of complex measurement tools [20]. One alternative to such methods is to predict the response of a complex system of proteins via *in silico* simulations of the system based upon the well characterized interactions of each protein in isolation [18, 19]. Systems of ODEs like the two simple systems presented above can be used to probe the dynamics of a system with a much finer temporal resolution than typically possible via experimentation, and can be used to probe system response to perturbations of individual components.

More complex models, such as spatial stochastic models of diffusion and collision driven reactions as described in chapter 3, can be further used to introduce spatial effects to these systems. With fine spatiotemporal resolution these models can account for the effects of biophysical phenomena, such localization and structural geometry, upon these protein interaction systems. A well characterized model can be used to make informed predictions of the response of a system to specific parameters, and can be used to guide the design of focused experimental studies.

Such models do have a few notable limitations. Firstly, even these models are limited in complexity by computational and design expenses. Modeling a complex system of interacting proteins requires modeling assumptions and reasonable boundary conditions for the model, which must therefore inherently fail to perfectly capture the true conditions of a biophysical system. It becomes critical to accurately describe protein interactions as closely as possible to allow for physiologically relevant predictions. To that end these models rely heavily on experimental characterization of parameters, many of which can only obtained through a robust examination of pertinent literature. Unfortunately, most experimental measurements are not designed with computational parameterization in mind.

Enzyme catalytic rates tend to be relatively easy to measure via enzyme activity assays and are a commonly discussed talking point when describing enzyme function. As such these k_{cat} parameters tend to widely available. However when examining protein binding behavior, most experimental studies seek to qualitatively describe the relationship between interacting proteins rather than to quantifiably define their

interaction. For instance, in the study of a protein in the presence and absence of an inhibitor, an experiment simply needs to demonstrate the reaction is *relatively* slower in the presence of inhibitor to characterize that relationship. In some cases these changes can be quantified and used to inform parameters, but not always. Similarly there exist innumerable methods of describing the apparent binding affinity $K_D app$ under given conditions, as this requires only some form of measurement of relative composition at equilibrium. In comparison, technologies capable of measure true binding kinetics, binding association and dissociation rates, are much more scarce.

1.2.2 Survey of Current PPI Measurement Techniques

Each of these technologies operates on the principle of measuring different biophysical properties to charactize binding.

Surface plasmon resonance (SPR) measures the diffraction angle of light off of thin metal film. This refraction angle is highly sensitivity to the optical properties of the medium on either side. In 1983, Liedberg et al. showed that by affixing a biomolecule, in their case human immunoglobulin, IgG, to this film they could observe a progressive change in the characteristic diffraction angle as they allowed a solution containing a binding molecule, anti-IgG, to flow across the surface of the chip [21]. Similarly, the diffraction angle relaxed when the surface of the chip was washed with a pure buffer solution. They concluded the rate of change of the diffraction angle was therefore analogous to the binding and release kinetics of the two biomolecules [21]. SPR's limitations are primarily in the complexity of experimental design due to the required adsorption of a biomolecule onto the metal film. There has also been some debate as to whether or not the binding affinities measured by SPR truly reflect the binding dynamics of the native biomolecule. By immoblizing one binding species to the metal film, it is possible that (i) the immobilized species is unable to maintain a native configuration, (ii) the immobilized species may present an orientation limited binding region (the binding domain may be partially or entirely facing the surface of

the film rather than outward into the solution), (iii) the immobilized species may not experience the true concentration of the solution since all binding occurs localized to the surface (localized concentration may be lower than the experimental solution due to local binding), and (iv) nonspecific binding and mass transfer may influence the measured signal [22]. As such, a large portion of SPR kinetic measurements can be reasonably assumed to be inaccurate in the absence of rigorous controls and assumption validations [23, 24]. Furthermore, the experimental setup can be quite expensive and is relatively specialized for this technique, and so is not a tool present in many biochemical labs.

Isothermal titration calorimetry (ITC) is calorimetry measurement of enzymatic activity at varying concentrations of substrate and inhibitors to characterize protein binding [25]. While this measurement allows a much more native measurement of protein binding than SPR and boasts a wide sensitivity range for protein affinities, it is notably most sensitive for proteins that exhibit enzymatic activity for calorimetric measurement and thus does not work well for measuring binding interactions between otherwise inert proteins that act exhibit exclusively modulatory behavior with relatively low enthalpy changes. Further limitations of the technology include (i) necessity of high concentrations for sufficient heat generation, and (ii) relatively long measurement times of 3 hrs/titration, both of which limit potential high throughput applications of this technology [25].

Fluorescence Correlation Spectroscopy (FCS) is a technique that measures the changes in spatial diffusion of biomolecules. As biomolecules bind together and create complexes their diffusion is changed in a measurable way [26]. This technique is the primary focus of chapter 2 and as such will not be discussed in detail here.

Fluorescence Polarization (FP) Assay is similar to FCS, but this technology measures the rotational diffusion of biomolecules labeled with a fluorescent tag. As this labeled molecule binds into complexes with targeted partners, the rotational diffusion slows. This is measurable in the relative change in fluorescence polarization due the orientation of these molecule as they rotate in space. This technique requires

the use of only a single fluorophore labeled species which makes it ideal for a high-throughput screening process, but typically relies on an engineered small peptide tracer that competes with the ligand of interest to bind the same domain upon a sufficiently large receptor to note a change in rotation. Therefore, the technique isn't ideal for measuring binding kinetics, but rather is typically used to find equilibrium binding affinity of a target ligand-receptor pair based upon the fractional binding of tracer in the presence of various concentrations of a competitive substrate [27].

Thermal Shift Assay (TSA) detects the change in thermal stability of a unlabeled target protein under various conditions. One such condition is the inclusion of a suspected binding partner. The binding of two high affinity biomolecules can increase thermal stability, but does so somewhat unpredictably. As such, this methodology is very useful for detection of qualitative changes in thermal stability in the presence and absence of a suspected binding target. However, it is less feasible as a measurement of binding kinetics, and really holds most value as a detector of potential high affinity binding partners and so is a frequent tool used for high throughput screening [28].

1.2.3 Observed Gap in PPI Measurement Technology

All of the surveyed techniques for measurements of PPIs share a failed capability to demonstrate an (i) inexpensive (in both experimental time and actual financial cost), (ii) native, and (iii) high temporal resolution measurement of protein binding kinetics. This gap presents a desirable target of research, particularly given the great leaps and bounds of *in silico* modeling of biodynamic processes in the past decade, due in no small part to the parallel progression of computation power and bioinformatics [29,30]. The development of such a technology would enable a much more efficient, high-throughput characterization of parameters necessary for computational models of PPIs and enable further progress in a rapidly growing field of study.

1.3 Macroscale: Dynamics of the Proteome and Metabolome

1.3.1 Background and Significance

The study of the proteome, the entire library of proteins that exist within an organism, is an entirely different mindset of protein investigation than the microscopic scale of PPIs. Proteomics is a study on the macroscopic scale, focused on the changes in the composition of proteins present within entire tissues or organisms. Whereas PPIs present a bottom-up approach to biological study, elucidating mechanisms of health and disease, proteomics (and it's cousins genomics and transcriptomics) can give a top-down view of all possible contributors to any observed state. Since the advent of genetic sequencing, there has been a powerful push towards these top-down approaches to biomedical study, searching for potential targets of pharmaceutical therapies or for potential actors of future study in classically poorly understood diseases [31]. That being said, it's reasonable to say that proteomics has lagged behind it's peers since the advent of Sanger sequencing and the birth of true 'omic' study. Although newer tools have been hot topics of study in more recent years, for the major part of the past two decades, proteomics has relied solely on the application of a single technology, Mass Spectrometry, albeit in a number of unique and creative ways [31].

Mass spectrometry (MS) is an analytical tool that measures the mass to charge m/z ratio of particles. The core principle of the technology is that the forces exerted on a charged particles in a vacuum is governed by two laws of physics [32]. The first of these is the Lorentz force, or the electromagnetic force, a combination of electrical and magnetical force acting on a point charge due to it's own charge.

$$F_L = q(E + u(B)) \tag{1.10}$$

Where q is the charge of a particle and E and B are the magnitude of the electric and magnetic fields respectively that the particle is passing through with some velocity

u. The second law is, incidentally, Newton's second law, which states that the force is a function of it's mass m and acceleration u'.

$$F_N = m(u') \tag{1.11}$$

Setting these two equations equal to each other yields the governing equation of motion for the particle.

$$m/q = u'^{-}1(E + u(B)) \tag{1.12}$$

This equation can be non-dimensionalized for ions by converting m to an elemental mass number and q to a charge number $z = q/e$ where e is the characteristic charge of an electron. Therefore, the motion of an ion is governed by it's mass ratio, m/z. In MS, typically a sample is ionized into a gas-phase, then passed through a fixed magnetic field that diffracts each ion at an angle which is then measured to determine the m/z ratio of all ions present in the original sample.

For proteomics, samples are typically ionized with a soft ionization, or incomplete fragmentation, using a method such as electrospray ionization (ESI) or matrix-assisted laser desorption (MALDI), that generate mostly whole, multiply charged macromolecules. This allows for distinct, large mass ion profiles that can be easily distinguished to identify and sequence proteins and peptides [33–35]. One of four common types of mass analyzers is then used to quantify these large ions: Quadropole (Q), ion trap (IT), time of flight (TOF), and Fourier-transform cyclotron resonance (FTICR) [35]. Various combinations of these two instruments make up the primary differentiation between MS instruments, and each combination can be used to analyze over different ranges of sensitivity, with different levels of accuracy and therefore different ideal applications.

Regardless of the instrument used, the standard strategy for proteomic MS is to digest all the proteins in a given sample into peptides which are then recorded as

'finger prints' that can be compared to a database to back out the protein composition of the original sample. This protein composition can then be probed to determine what is happening to the proteome under various conditions.

Applications of this technology are broad reaching. Any number of experimental treatments, states of organism development, and states of disease can influence the output proteome. Qualitative variation in the proteome can implicate potential pathways of future study, and thus these tools bear great merit as a top-down model of experimental investigation. This being said, there are limitations of such a shotgun approach. Typically, detection is dependent upon a relatively high mass of protein present in samples, without some way to highlight proteins with lower expression rates (such as a large portion of enzymes) these can be masked behind more dominant players. Similarly, if there is a selected subset of proteins that are of interest, a shotgun approach may not be the most desirable method of targeted study. Full proteome analysis also lacks any sort of temporal specificity, distinct changes in the proteome due to a time resolved stimuli are easily masked by the constituent proteome. To address these issues, and study specific targets within the proteome, a common method is to label the proteome in a way that enables either enrichment of targets or distinct separation of desired peaks.

1.3.2 Survey of Current Metabolic Labeling Techniques

Metabolic labeling refers to any methods by which the endogenous machinery of cells is utilized to generate labeled proteins for targeted analysis. Such labeling can be accomplished through a broad variety of techniques, but at the core are all based on the replacement of a common biomolecule with a labeled analog.

Isotope labeling involves supplementation with a stable isotope of a common biological molecule, typically with ^{15}N. A common practice is to expose an organism to an isotope-variant amino acid, enabling resolution of proteins that incorporate this radio-heavy species. Isotope labeling allows experimental studies to highlight

responses to stimuli, by looking only at protein generated after the introduction of the isotope [36–39]. Such labeling has a few notable limitations. While not radioactive, these isotopes exhibit distinct mass values that can influence biological function and become potentially harmful if present at substantially high concentrations [40]. Furthermore, following their introduction, these isotopes are recycled by naturally processes and are not removed from the system. There is also no simple methodology to enrich proteins labeled in this fashion, and as such low expression is still an issue for experimental detection.

Non-canonical amino acid (ncAA) labeling is a relatively newer technology that involves the introduction of an amino acid analog that can be incorporated into native protein [41]. These amino acids can be designed to exhibit inherent labeling such as photo-reactivity or to include bioorthogonal tags that make them easily modified with some form of label. While this allows for great versatility in possible labeling strategies there are some widespread limitations to presently utilized techniques. Firstly the majority of studies utilizing ncAA labeling have been limited to dietary dosing in complex organisms, with corresponding slow distribution and labeling of the proteome. Additionally, even bioorthogonal ncAAs tend to exert greater influence over biophysical phenomena than radio-heavy isotopes of canonical-amino acids due to greater distinctions in their structure [42]. Current techniques in ncAA labeling therefore tend to demonstrate a limited capacity to span multiple timescales. While there are other considerations, this technique is the primary focus of chapter 4 and as such will not be discussed in further detail here.

1.3.3 Observed Gap in Proteome Labeling Techniques

Amongst the surveyed labeling techniques to enhance proteomic study there is a noted absence of a technique to transiently label newly synthesized proteins *in vivo* for imaging and enrichment. Such a technology would enable study of not only a status post proteome, but additionally would allow a measure of temporal resolution

into the dynamics of the proteome. It could enable study of not just the proteins synthesized under given conditions, but also the relative turnover of those proteins and general metabolome response to stimuli.

2. FLUORESCENCE ROLLING CORRELATION SPECTROSCOPY (FRCS)

Contents of this chapter and the subsequent chapter of preliminary experimental findings will be submitted for publication. This chapter discusses the theory and applications of standard Fluorescence Correlation Spectroscopy (FCS) and the extended applications of single-photon correlation to analyze biomolecular dynamics. Here I discuss the current state of FCS including basic principles, limitations, and methods of data analysis. I then present a new analysis technique and fitting models to measure biomolecule binding behavior with higher temporal sensitivity.

2.1 Introduction

2.1.1 Background and Significance

Fluorescence Correlation Spectroscopy is a powerful tool originally developed by Magde, Elson and Webb in 1972 as a tool to analyze DNA-drug intercalation [43]. At its most basic, the objective of FCS is to take advantage of measurable spontaneous changes in native fluorescent behavior of biomolecules to deduce minute fluctuations in the physical phenomena that modulate fluorescence. When measuring the emitted fluorescence of a sample, many physical properties hold influence over the recorded signal (concentration, position and conformation of particles, chemical and photo reactivity, etc.) [26]. Under most standard conditions many of these properties are constantly fluctuating, generating a typically undesirable noise in the response signal. However, assuming some of these fluctuations are driven by predictable behavior, it stands to reason that this 'noise' would hold information describing the same. Correlating the signal against itself across an interval of time yields frequencies and timescales of significance within the fluctuating signal that can help describe the governing physical processes [43].

At the conception of FCS, this presented a somewhat unique conundrum among fluorescent imaging methods; in most cases, fluorescent studies are designed to observe broad trends in fluorescence over a fluctuating baseline, however for FCS the ideal conditions are reversed, these fluctuations are the signal of interest. Just as it is difficult to understand a single voice among a crowd, so too is it difficult to observe dynamic shifts in fluctuations if many biomolecules are contributing to an averaged signal. With standard fluorescence microscopy, optimizing the signal-to-noise ratio (SNR) for FCS was quite difficult and limited measurable concentrations to far below physiological ranges ($<$ 1nM). The integration of confocal microscopy with FCS allowed an elegant reduction in measurement volume to 1 μm^3 or smaller, and enables concentrations more aligned with physiological ranges [44, 45].

It's worth noting here that while many equilibrium processes can theoretically be measured with FCS, the tool has primarily been used to measure stable Brownian diffusion in aqueous solutions, and subsequently to measure differences in diffusion under distinct, but stable conditions. An common example of such is to detect reduced diffusion of biological molecules as they aggregate due to intramolecular interactions. Traditionally, this phenomenon has often been used to characterize binding affinity between biological molecules. That being said, there have been studies that captured more elusive kinetic parameters, although these often required either (i) relatively slow binding that occurred over the course of minutes (longer than a typical FCS experimental timecourse) or (ii) the application of an advanced FCS technique such as two-photon Fluorescence Cross-Correlation Spectroscopy (FCCS) or applied FRET analysis (FRET-FCS) both of which have unique limitations detailed further below [46, 47]. However, Schwille et al. and others since have shown that at characteristic concentrations for FCS (5-50nM), it is possible to capture rapid kinetics on the order of those observed in PPIs (ka $= 1e4 - 1e6$ $M^{-1} \cdot s^{-1}$) by monitoring a single labeled ligand as it associates with a target, particularly for binding interactions that demonstrate near irreversible binding such as DNA annealing [48, 49]. Notably, FCS

experimental timescales traditionally have limited the ability to observe more rapid and transient binding kinetics, although binding affinity is still very much resolvable.

Here we present an simple data processing algorithm that utilize this singly-labeled technique to improve FCS measurement resolution and capture dynamics of rapid and transient binding kinetics as might be observed in a broad subset of PPIs. We introduce a rolling window correlation algorithm into the standard one-photon FCS workflow. Fluorescence Rolling Correlation Spectroscopy (FRCS) is a novel application of this technology that allows time resolved measurements of biophysical behavior in sub-second time domains, much shorter than the time-course of an FCS experimental measurement. In contrast, traditional FCS and FCCS are capable of generating only information pertaining to the overall average characteristics of a solution in this time domain [50, 51]. The comparison of relative diffusion across various FCS time scales enables a more in depth investigation of reaction progress allowing characterization of not only binding affinity, but of binding kinetics for a much wider range of PPIs.

2.1.2 Principles of Diffusometry

Any form of FCS technique utilized for measurements of binding affinity relies on tenants of microfluidic behavior. As a whole microfluidics is a field which has demonstrated exponential growth as a field in biomedical technology by enabling small sample volumes and high throughput experimental design [52]. Diffusometry, the applied measurement of diffusion, is a subset of microfluidic study that applies the principles of Brownian motion to resolve other biophysical phenomena.

Brownian motion is defined as the random thermal motion of particles suspended in a fluid. This motion was first discovered by and named after botanist Robert Brown, but was mathematically characterized independently by William Sutherland, Albert Einstein, and Marian Smoluchoski in the early 1900's [53–55]. Perhaps the most widely known of these is Einstein's application of Stokes' law to describe diffusion

in a low Reynold's number solution, a solution dominated by viscosity rather than inertia. [54].

$$D_c = \frac{k_b * T}{6\pi\eta R_H} \tag{2.1}$$

Here the diffusion coefficient D_c of a particle is a value with units of $m^2 \cdot s^{-1}$ and describes the average rate at which that particle will traverse an area. Einstein defined this value as a function of Boltzmann's constant (k_b, $m^3 \cdot Pa \cdot K^{-1}$), the temperature ($T$, K), the solution viscosity (η, Pa·s), and the hydrodynamic radius of the particle (R_H, m). Therefore, in a solution with fixed viscosity and temperature, diffusion is modulated solely by the size of a particle within that solution.

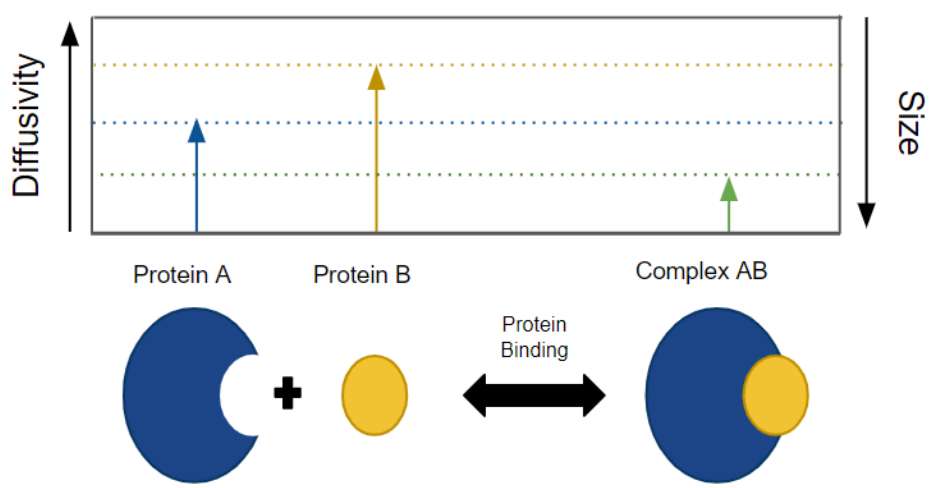


Fig. 2.1. A cartoon showing the reaction of two proteins A and B into complex AB. Each species has a unique size and therefore a unique diffusion.

This principle can be exploited to resolve dynamics of a system. One example of such would be a particle changing in effective size as it aggregates with other particles as might occur in a simple bi-molecular binding interaction between two proteins of differing sizes. As a given, each of these proteins will have a unique hydrodynamic

radius and as such a unique diffusion. As they interact and begin to bind a third species is produced, a complex with a hydrodynamic radius that is larger than either independently (in most cases). As such, in a mixture of these two proteins there would exist three distinct species of diffusing particles, each with a unique D_c, as depicted in figure 2.1. This principle can be exploited to resolve the composition of a solution from diffusion, and is core component of the analysis and algorithms discussed further below in greater detail.

2.1.3 Principles of FCS

FCS Theory

Standard FCS, also known as Fluorescence Auto-Correlation Spectroscopy, requires the use of only single-photon excitation. A laser of appropriate wavelength to excite the chosen fluorophore is focused to a point within a sample solution. This creates a soft-edged volume space known as the confocal volume (CV) within which fluorophores are exposed to excitation (figure 2.2). While the true geometry of the CV is dependent upon the optical system structure, for a small confocal pinhole around 1 Airy Unit (AU) the CV is most often depicted as a Gaussian ellipsoid. The intensity of the excitation signal at a point with radial and axial distance (r, z) from the center of the CV can be modeled as a point spread function (PSF) of laser intensity [44].

$$PSF(r, z) = I_0 e^{-2r^2/\omega_{xy}^2} e^{-2z^2/\omega_z^2} \tag{2.2}$$

Here, I_0 is the peak intensity of the laser, ω_{xy} and ω_z are the radial and axial radii of the CV. fluorophores exposed to this field are excited with a probability proportional to the laser intensity at the spatial coordinates for any given point in time. Excited fluorophores then begin to emit photons which are captured by the objective. Captured photons are filtered to remove wavelengths beyond the emission range of the target fluorophore and the arrival time of each filtered photon is recorded by a

single-photon detector (typically in arbitrary clock units (ACU) that are converted to meaningful units by detection software) [56]. These captured photons hold pseudo-positional data about particles within the CV. The timing and quantity of photons holds information about the number of fluorescent particles present in the space, as well as their location relative to the center of the Intensity PSF.

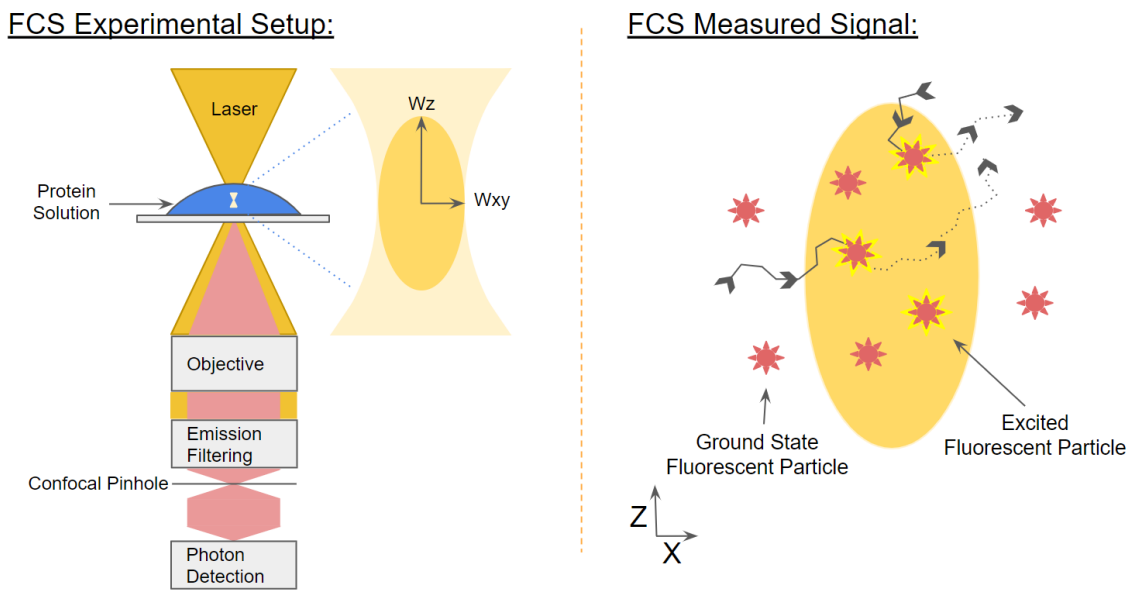


Fig. 2.2. Left: Experimental setup for FCS. Laser (yellow) is passed up through the objective focused to a point within a desired protein sample. Excited fluorophores emit photons captured by the same objective (red), filtered and passed through a confocal pinhole before being captured by the photon counter. Right: freely diffusing fluorescent particles enter the CV and become excited, emitting photons for counting.

By comparing the photon counts at different time points, it is possible to monitor trends as particles diffuse into and out of the CV. To make perform this comparison

over time, the photon count data is auto correlated and normalized by the total photon count.

$$G(\tau) = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I(t)\rangle\langle I(t+\tau)\rangle} \tag{2.3a}$$

$$= \frac{\langle I(t)I(t+\tau)\rangle}{\langle I(t)\rangle^2} \tag{2.3b}$$

Here, angled brackets denote an average over the entire measurement window and τ, the residence time, is the length of time between two points in the intensity function I(t) whose average similarity over the entire measurement window is indicated by G(τ), the unitless correlation magnitude. The numerator in eqs. (2.3) is more commonly known as the time-lagged dot-product function, and the denominator is a normalization factor for either a symmetrical (2.3a) or asymmetrical (2.3b) normalization. This correlation function is typically preformed by a real-time hardware photon correlation module, also known as a time tagged multiple coincidence detector [57].

It is perhaps most intuitive to think of this function as a frequency distribution of particle residency times. Consider the ideal, simple case where I(t) is a function only of the number of particles within the CV at time t, and all particles have an identical diffusion coefficient through the medium, D_c (m^2s^{-1}). A particle might enter the CV at time t_0 and remain within the CV for a very short period (a period of τ seconds). Due to this particle, there is a measurable relationship between the counted photons emitted at time t_0, I(t_0), and those emitted at time $t_0 + \tau$, I($t_0 + \tau$). Similar relationships exist over all possible combinations of I(t) and I(t+τ). The average magnitude of their similarity is analogous to the relative occurrence of particles with a similar residency time over the duration of the entire measurement interval.

It would be expected that low residency times (τ_a in fig. 2.3) would be most likely to occur as particles 'flicker' in and out of the CV at the soft boundary. Conversely, it would be very unlikely for a particle to have a long residency within the CV (τ_b), as this would entail a statistically improbable possibility that the particle 'bounced around' within the CV for an extended period of time. However, somewhere between

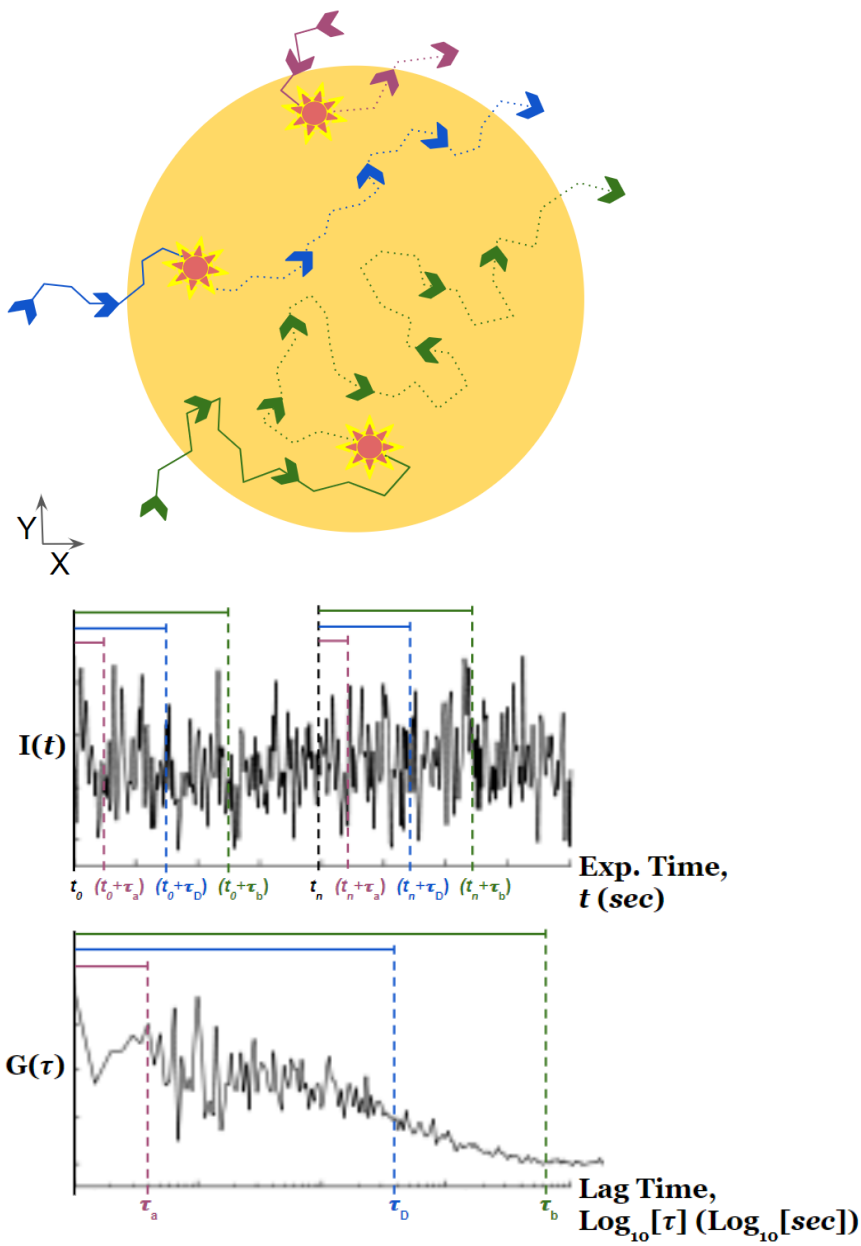


Fig. 2.3. Top: Cartoon showing possible paths of fluorescent particles through an XY cross-section of the CV. For stable diffusion, the particle path determines residency time within the CV. Bottom: example photon count intensity trace, $I(t)$ and autocorrelation, $G(\tau)$. Shown also are possible observed lag times for the color coordinated particles in the top panel (not to scale).

these two extremes exists a residency time (τ_D) that is characteristic of the average time (s) it takes for a particle to diffuse across the cross-section (m^2) of the CV.

This characteristic residency time, τ_D, is the target measurement of a diffusion study, as it is inversely proportional to the D_c of the fluorescent particle. If the dimensions of the CV are known the diffusion coefficient can be calculated from τ_D.

$$D_c = \frac{\omega_{xy}^2}{4\tau_D} \tag{2.4}$$

Here ω_{xy} is radius of the confocal volume perpendicular to the laser and measuring axis, and τ_D is scaled to account for three dimensional geometry of the experimental measurement [44].

Fitting Models

In order to resolve τ_D, the most common approach is to fit the correlation function $G(\tau)$ with a fitting model that accounts for experimental dynamics and the optical geometry of the system. The simplest case of such a model depicts a single particle system dominated by Brownian diffusion.

$$G(\tau) = G(0)M(\tau) + G(\infty) \tag{2.5a}$$

$$M(\tau) = \frac{1}{(1 + (\tau/\tau_D))(1 + \alpha^{-2}(\tau/\tau_D))^{1/2}} \tag{2.5b}$$

Here $G(0)$ is a normalizing factor inversely proportional to the average number of fluorophores present in the CV during measurement, $G(\infty)$ is the fitting offset (typically 0 or 1 depending on the correlation algorithm), and $M(\tau)$ is a measurement function that describes unique dynamics depending on the experimental design. For a system with only one species of diffusing particle, $M(\tau)$ includes only two parameters, the diffusion time and α, the structural parameter of the CV describing the ratio of

ω_z/ω_{xy}. The inclusion of this α parameter adjusts for the three-dimensional geometry of the CV such that it can be used to generate a diffusion time describing two-dimensional motion. The normalizing factor $G(0)$ can also be easily used to describe the geometry of the CV.

$$G(0) = \frac{1}{\langle P \rangle} = \frac{1}{C(V_{eff})} \tag{2.6}$$

$$V_{eff} = \pi^{3/2}\omega_{xy}^2\omega_z \tag{2.7}$$

Here P is the number of particles in the CV, C is the concentration of particles in the sample solution and V_{eff} is the effective volume of the CV. This standard fitting model can also be easily adapted for a poly-disperse solution with distinct species of particles (each with identical fluorescence profiles but unique $D_{c,i}$) by adjusting the $M(\tau)$ function accordingly.

$$M(\tau) = \sum_i \frac{F_i}{(1 + (\tau/\tau_{D,i}))(1 + \alpha^{-2}(\tau/\tau_{D,i}))^{1/2}} \tag{2.8}$$

Here F_i is the fractional composition of particles in the sample with $D_{c,i}$ and thus $\tau_{D,i}$. A similar model can be used to accommodate variation in the relative intensities of fluorescence between species.

$$M(\tau) = \frac{1}{(\sum_i F_i\epsilon_i)^2} \sum_i \frac{F_i\epsilon_i^2}{(1 + (\tau/\tau_{D,i}))(1 + \alpha^{-2}(\tau/\tau_{D,i}))^{1/2}} \tag{2.9}$$

Here ϵ_i is the relative molecular brightness of each species, typically as compared to the molecular brightness of the species of most interest, q_0 ($\epsilon_0 = q_0/q_0$, $\epsilon_1 = q_1/q_0$, etc.). One final adjustment typical for fitting models is descriptor for the system response, sometimes called the triplet state correction. This modification accounts

for biophysical changes that have influence over observed fluctuations. For the majority of organic fluorophores, some fraction of excited particles, T, tend to become further excited into a non-radiating state until they relax back to ground state at a characteristic rate τ_R. There are also a number of other phenomenon, such as rotational diffusion and after-pulsing, that share a similar time domain [43]. Typically on the order of microseconds, these biophysical events usually occur at a much smaller timescale than the desired sensitivity region for measurement, however accommodating this behavior tends to allow much more precise fitting within the range of interest. The triplet correction is often represented independently, but because there are a number of phenomena that occur in this time domain it is could also be bulked into a system correction factor, S, with a characteristic time τ_S.

$$G(\tau) = G(0)\left(1 + \frac{T}{1-T}e^{-\tau/\tau_R}\right)M(\tau) + G(\infty) \tag{2.10a}$$

$$= G(0)\left(1 + Se^{-\tau/\tau_S}\right)M(\tau) + G(\infty) \tag{2.10b}$$

Note that 2.10 demonstrates a triplet/system correction factor adjustment usable with any of the above functions for $M(\tau)$. Further adjustments to the fitting model for single-photon FCS can be utilized to capture a number of other known behaviors depending on the experimental conditions such as defined convective flow and chemical relaxation (as with a FRET system for instance), but further detail on such models can be found elsewhere [43, 58].

FCCS Theory

Two-photon fluorescence microscopy, enables a more complex application of FCS technology. FCCS involves two species of fluorescent particles each with a unique fluorophore and a distinct wavelength of excitation/emission that are present within the same system. Both fluorophores are excited by lasers focused to produce a single

CV, and the resulting emissions are measured independently via separate channels. As differentially labeled particles bind together and interact, the two channels become more similar as they observe the same diffusive behavior in each channel. Unlike in FCS, which utilizes an auto-correlation function, these channels are then cross-correlated against each other to reveal these relationships between the fluctuations present in each channel.

$$G_X(\tau) = \frac{\langle I_1(t) I_2(t+\tau) \rangle}{\langle I_1(t) I_2(t) \rangle} \tag{2.11}$$

This function can be modeled with an only slightly more complex form of the fitting models used for FCS that accounts for some inherent differences in the measurement. Namely, the effective volume is different for each species as each excitation laser will create a slightly different PSF and CV. The effective superimposed observation volume of the two channels can be defined to account for the unique geometry of each channel.

$$V_{eff,12} = \frac{\pi^{3/2}}{2^{3/2}} \left(\omega_{xy,1}^2 + \omega_{xy,2}^2\right) \left(\omega_{z,1}^2 + \omega_{z,2}^2\right)^{1/2} \tag{2.12}$$

Here $V_{eff,12}$ is the effective superimposed volume, and $\omega_{xy,1/2}$ and $\omega_{z,1/2}$ are the radial and axial radius of a confocal ellipsoid for channels 1/2 respectively. This effective volume can be used to inform the concentration of the doubly-labeled species created by binding interactions. In the absence of confounding artefacts, the maximum value of $G_X(\tau)$ is proportional to the concentration of the doubly labeled complex species and the characteristic residency time describes the diffusion of that complex [46].

FCCS is a powerful tool for measurement of binding affinity and enables quick measurements of equilibrium K_D without requiring a concentration titration of reagents, but it has some notable drawbacks. Firstly, both species must be fluorescently labeled and imaged via FCS. This requires engineering and modification of both binding species, which in the case of protein-protein interactions doubles the possibilities for

potential variation from wild type behavior. Similarly, the limitations of FCS demand low concentrations of any fluorescent species, in the range of 5-50 nM. For interactions with K_D of 1 μM or greater it is very difficult to observe binding at these concentrations, so the tool is limited to interactions of high affinity. Like FCS, FCCS also requires a reasonable experimental measurement length ($\sim$ 30 s) to obtain sufficient data for averaging into a reasonable SNR. This characteristic measurement time makes it difficult to observe dynamics faster than the measurement time such as the binding kinetics of proteins with sub second timescales. So while FCCS can characterize binding affinity with ease, binding kinetics such as those used in the Michalis-Menten model of interactions are less obtainable. Lastly, FCCS has known experimental limitations: it requires more expensive setups, contains invasive cross-talk artefacts between channels that skew binding affinity measurements, and increases the rate of photo-deconstruction of fluorophores [46].

FRET-FCS Theory

A slightly newer technique, FRET-FCS is a distinct application of FCS technology to measure binding affinity with differences that address some of the issues presented by FCS. This technology applies FRET technology detailed by Stryer et al. [59]. The general principle relies on the transfer of excitation energy of a donor fluorophore to a nearby acceptor when they are proximal to each other (within a few nanometers). These paired fluorophores then are only able to become excited to a radiative state when localized to a complimentary fluorophore. This localization can be accomplished a variety of ways, but a convenient way is to attach complimentary labels to two particles that when bound together will localize the fluorophores allowing them to excite and emit photons.

FRET-FCS capitalizes on this technology to measure the chemical relaxation of a species fixed within the CV as a diffusive species binds and releases [47].

$$G(\tau) = G(0)e^{-\tau/\tau_B} \tag{2.13}$$

‘

$$G(0) = \frac{1}{\langle N \rangle}\frac{k_a}{k_D} = \frac{1}{\langle N \rangle K_D} \tag{2.14}$$

$$\tau_B = \frac{1}{k_a + k_d} \tag{2.15}$$

‘

While this technology allows direct measurement of binding kinetics. It has it's own limitations, with a notably more complex experimental design requiring the engineering of two labeled species and typically requiring fixation of one of these species in place within the CV to prevent diffusion from affecting the observed measurements of binding. This technique, similar to surface plasmon resonance, therefore is most viable for the study of a 3D freely diffusive species to a 2D surface bound species. Any observed binding affinities under these conditions may be geometrically limited or enhanced by surface interactions.

Correlation

Regardless of the experimental design and fitting model, the actual measurement analyzed remains the same. The raw photon count, as recorded by the counting module is passed into a discrete, time-domain correlation to generate a curve that is then parameterized with non-linear regression. This is typically accomplished with a hardware correlation module, as they are capable of performing a large number of computations in real-time, to rapidly generate a signal output.

Hardware Correlation: Recall as above in eq. (2.3) that the general form of the correlation function is a symmetrically normalized time-lagged dot product.

$$G(\tau) = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I(t)\rangle\langle I(t+\tau)\rangle} \tag{2.3}$$

This form of the equation assumes a continuous function I(t) which is not truly an accurate representation of the experimental measurement. Rather the intensity function is recorded as a stream of discrete photon arrival events typically binned for correlation. The discrete formula for what is known as the bin and multiply (BM) correlation function is very similar to the general form and simply requires redefinition in the discrete domain [60].

$$\hat{G}[k] = \frac{\langle \hat{I}[j]\hat{I}[j+k]\rangle}{\langle \hat{I}[j]\rangle\langle \hat{I}[j+k]\rangle} \tag{2.16}$$

In this instance, j represents the bin or sample index recorded with sampling time Δt ($t = j\Delta t$), $\hat{I}[j]$ is the discrete number of photons received within that sample, and k is a time-lag displacement ($\tau = k\Delta t$). In the discrete form the time-lagged dot product is well named and is defined as the dot product of the data vector with a time shifted copy of itself, with the small addition of a normalization to accommodate fewer data points in each trimmed vector as k increases.

$$\langle \hat{I}[j]\hat{I}[j+k]\rangle = \frac{1}{J-k}\sum_{j=0}^{J-k} \hat{I}[j]\hat{I}[j+k] \tag{2.17}$$

Here J is the length of the original data vector and $J-k$ is thus the length of the two comparable vectors for time shift k. Similarly, the normalization factors for $g(k)$ must account for the difference in vector length associated with k.

$$\langle \hat{I}[j]\rangle = \frac{1}{J-k}\sum_{j=0}^{J-k} \hat{I}[j] \tag{2.18}$$

$$\langle \hat{I}[j+k]\rangle = \frac{1}{J-k}\sum_{j=k}^{J} \hat{I}I[j] \tag{2.19}$$

It is notable that FCS has a very high temporal resolution, but also that the timescales for fluctuations of note in biophysical studies tend to span a wide dynamic

range of several decades which must be captured by the analysis (typical setups measure over a range of approximately (1e−8,1e+0) seconds) [26, 43]. To compute $g(k)$ for every possible value of k at the highest measuring resolution using a linear-tau application of eqs. (2.16)-(2.19) would therefore require $\sum_{k=0}^{J}(J-k)$ computations, an O(n^2) problem. Even with conservative estimates for experimental conditions this is an untenable number of computations (a 10ns temporal resolution and 10 second measurement gives J=1e+9 or total 1e+18 total computations). Typically, these devices employ a multiple-tau correlation algorithm to reduce the number of comparison events to a tenable number. One way to accomplish this is to use overlapping, cascading blocks of linear-tau correlators, with each successive block spanning a wider range of possible τ values [61].

In such a module, the block cascade takes as it's primary input the most recent photon count received during some period Δt as a raw value. This datum is then fed into the first linear-tau correlation block. Within this block there are a given number of linear-shift registers. Each register holds the value of a previous measurement such that the value of a given register of index m = (0,1,2,...,M) represents the photon count over the interval Δt_m at some point in the past given by the lag time $\tau_m = m\Delta t$. When a new datum is recorded each value in the block shifts to the right by a single register, the value in register M is discarded and the new datum is stored in the now vacant first register. To determine the correlation function of this block, the new datum is then multiplied by every value in the block, producing an array of products of length M representing the relationship of the new datum to the data τ_m time prior. This array is then added to a cumulative array tracking the sum-products of this block during the entire course of an experimental measurement. The multiple-tau component comes in as data is passed from the first linear block to the second block. The second block, and indeed each successive block in the cascade receives it's new datum by taking the sum of some number, γ, of elements at the end of the previous block. The photon count stored in each register of any block of index n = (0,1,2,...,N) therefore represents the count over a span $\Delta t_n = \gamma^n \Delta t$. For each

block after the first, the correlation is calculated similarly, with a small additional adjustment of normalizing the products by γ^n to account for the variation in the relative photon count per register in each block with increasing values of Δt_n.

Software Correlation: Software correlation can also be used to generate real-time, simultaneous processing of photon count data into correlation traces, however the possible resolution for real-time processing declines significantly as the analysis slows in the step from dedicated hardware computation structure to a software based data manipulation and calculation [62–64]. That being said, software correlation can be utilized for post-processing of experimental measurements, can often generate better resolution of the experimental data and enables the user to manipulate and trim the raw photon count signal before processing the data into correlated output. Software correlation is discussed in greater detail in section 2.3.2.

2.2 Methods

2.2.1 FRCS Algorithm and GUI Design

FRCS correlation algorithm and custom fitting was evaluated using scripts written in Python and C++ (merged using Ctypes foreign function prebuilt library for python) [65]. All scripts were evaluated on a Linux emulating virtual OS with a Intel i7-6700 CPU @ 3.4 GHz and 8 GB of RAM. For correlation raw data was exported either from custom simulation scripts as detailed in chapter 3 from a Zeiss LSM880 Confocal microscope and converted into PAT lists. Each PAT list was held in temporary memory and experimental time frames were extracted, binned, correlated to generate a rolling correlation over the time-course of the experimental measurement as described in greater detail in section 2.3.2. Fitting was performed using a variety of fitting models and all models were resolved using a combination of Lagrangian Least Square Error and Differential Evolution minimization algorithms using the Lmfit library for Python [66]. A graphical user interface was created using the widget based Tkinter library for Python and the Matplotlib graphical display kit [67].

2.3 Results

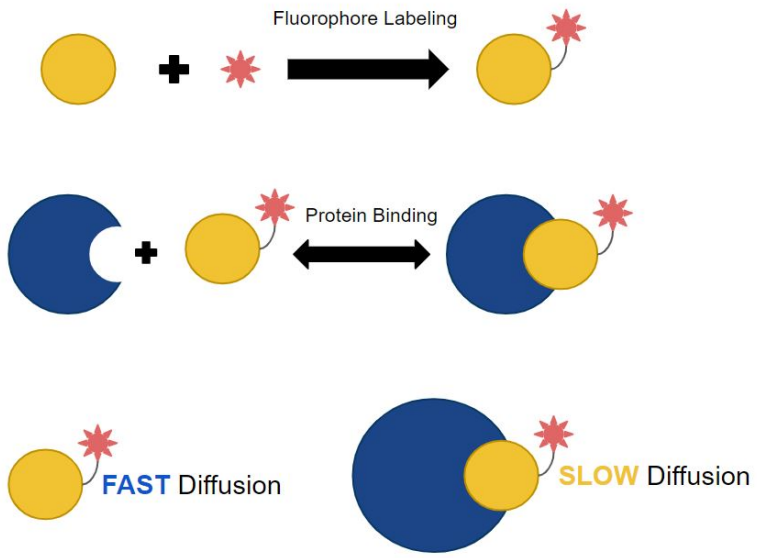
2.3.1 FRCS Theory

Given the limitations associated with FCS, FRET-FCS and FCCS as discussed in section 2.1.3 regarding biomolecular binding interactions, there seems to exist a gap in these technologies. FCS is known to capture data with a time scale resolution on the order of nanoseconds or faster [26]. It should therefore be possible to observe binding kinetics and dynamics of an interaction on a similar scale. Indeed this capability has been verified by the ability to directly measure chemical relaxation through FRET-FCS and by studies that utilize single ligand labeling to demonstrate binding dynamics of biomolecules [46, 47, 49]. However, for both FCCS and traditional FCS that primarily measure diffusion, capturing these fine time resolution dynamics is limited by the need to average measurements over a period of seconds to resolve a valid SNR.

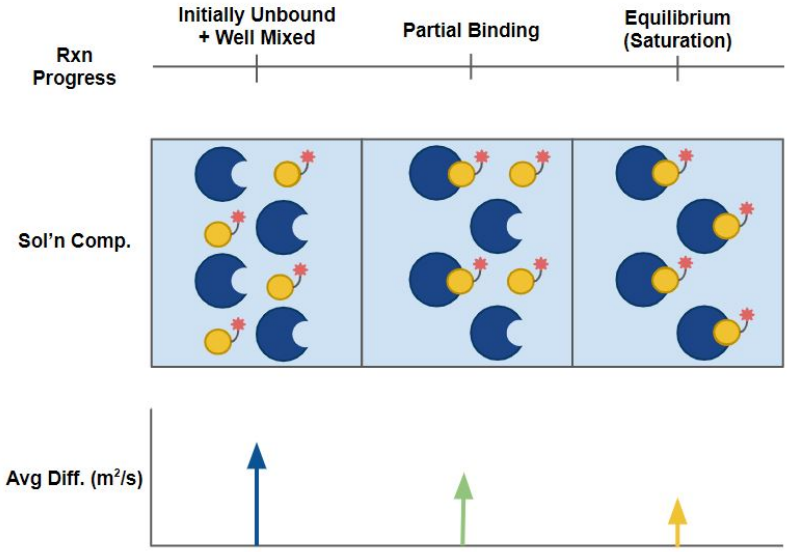
Take a binding reaction of two protein species, one fluorescently labeled and one fluorescently inert. With FCS, the diffusion of the fluorescently labeled species can be measured as it interacts with the larger binding target, increasing in characteristic size and therefore reducing in diffusion.

That being said, the depiction shown in figure 2.4 is a bit of an oversimplification. FCS will not truly measure the *average* diffusion of particles in the solution. Rather a poly-disperse fitting model such as described in equation 2.8 and visualized in figure 2.5 can be used to determine the composition of the solution. In this case there are two possible diffusing species representing the free and bound ligand, each with a unique $\tau_{D,i}$. The diffusion of the free ligand can be calibrated with a relatively simple measurement of the ligand in isolation, therefore requiring parameter fitting only of the fraction of ligand that is bound up, and the diffusion of this bound species.

If the binding interaction proceeds at a sufficiently slow rate, say over the course of a few minutes, subsequent FCS measurements of 15-30 seconds can reasonably detect changes in diffusion for distinct time points in the reaction. Schuler et al.



(a) Cartoon of a two particle binding reaction with one fluorescently labeled ligand binding a larger inert protein.



(b) Cartoon of binding reaction progress, with a corresponding reduction in average diffusivity

Fig. 2.4. Single labeled binding measurable with one-photon FCS

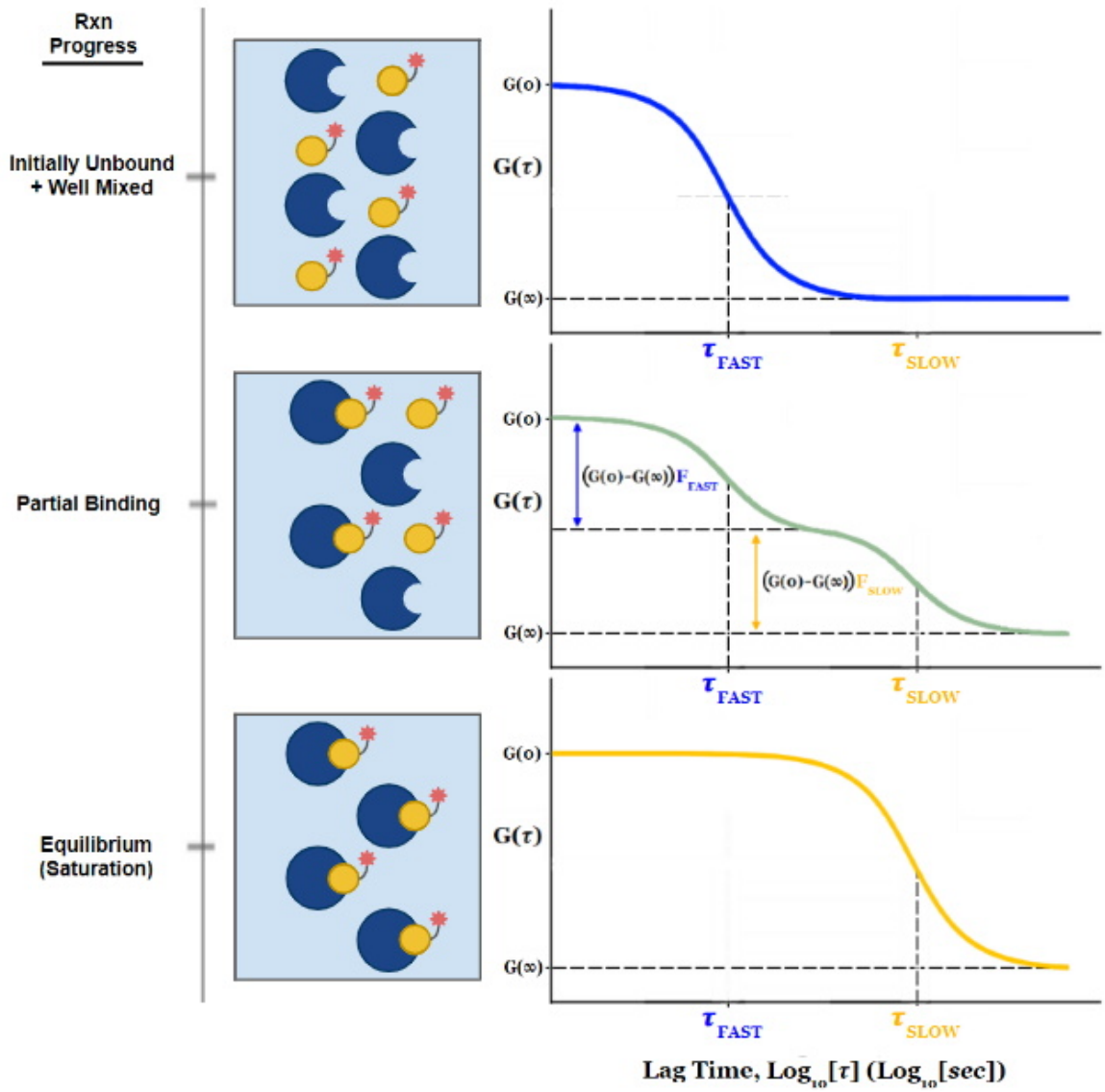


Fig. 2.5. Expected correlation geometry of a two particle fitting model over the timecourse of a binding reaction between a fluorescently labeled ligand and a larger inert protein. Characteristic residency times are τ_{FAST} and τ_{SLOW} for the free and bound ligand respectively, and similarly the fraction of each species present is indicated by F_{FAST} and F_{SLOW}. Note that for a reversible binding interaction full saturation is unlikely to occur with equivalent concentrations of ligand and receptor.

indeed demonstrated this was possible, resolving the binding of a fluorescently labeled protein, transferrin, to a target human transferrin receptor over a 20 minute period

with binding kinetics of $k_a = (1.1 \pm .1)\text{e}4\ \text{M}^{-1}\cdot\text{s}^{-1}$ and affinities of $k_d = (6.0 \pm 4.0)\text{e}{-4}$ s^{-1} (K_D = 15-100 nM) [49].

Having validated this methodology, it is notable that for many PPIs the binding interaction does not necessarily fall into this time domain. Firstly, many high affinity binding interactions occur with faster kinetics on the order of $k_a = 1\text{e}6\ \text{M}^{-1}\cdot\text{s}^{-1}$, and secondly even binding rates below this may result in sub-minute or even sub-second time scale interactions if concentrations are high enough as might be necessary to exhibit binding for lower affinity interactions. In these instances, even with the smallest reasonable FCS measurement length, we may be unable to observe multiple data points within the association phase of binding.

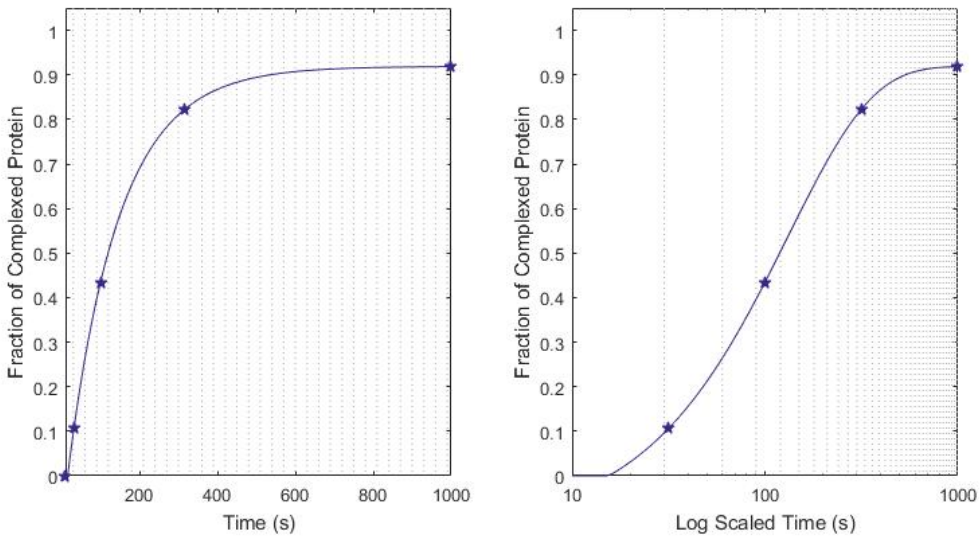


Fig. 2.6. Simulated binding interactions of a fluorescent ligand fA, to a inert receptor B over a time course of minutes. Right panels show the same time course in log scale, grid lines indicate a 30 second period (as subsequent FCS measurements might capture). Plotted is the fraction of complex fAB using kinetics and concentrations as observed by Schuler et al. [49].

We can simulate this type of binding interaction with a simple system of non-linear differential equations to investigate this hypothesis. For a system such as described

above in figures 2.4/2.5 this requires a system of ODES with only 3 mathematical equations as described in chapter 1.

$$fA + B \rightleftharpoons fAB \tag{2.20}$$

$$\frac{d[fA]}{dt} = k_d[fAB] - k_a[fA][B] \tag{2.21}$$

$$\frac{d[B]}{dt} = k_d[fAB] - k_a[fA][B] \tag{2.22}$$

$$\frac{d[fAB]}{dt} = k_a[fA][B] - k_d[fAB] \tag{2.23}$$

This system can be solved with any numerical integration algorithm, such as MATLAB ode45, and the resulting traces used to predict PPI behavior observable with FCS. If we want to convert the results of this ODE simulation to fractional binding as we would observe in FCS this is an even more trivial task.

$$F_A(t) = \frac{[fA](t)}{[fA]_0} \tag{2.24}$$

$$F_{AB}(t) = 1 - F_A \tag{2.25}$$

Shown in figure 2.6, a reaction as observed by Schuler et al. with slow kinetics and high affinity is easy to characterize with subsequent FCS measurements of 30 seconds. However, when we begin to explore the possible parameter space of PPIs ranging over reasonable association rates and binding affinities, it's evident that many PPI kinetics would not be well characterized by FCS measurements in this fashion (figure 2.7). Notably, if we aim for concentrations of binding actors that will attain 50% saturation of the fluorescent ligand (figure 2.7b), such that we can readily observe FCS parameter fitting error both above and below the curve, we note that that under many conditions, the binding curves proceed to equilibrium within the span of only around 2 minutes or so, or the span of ~ 4 FCS measurements.

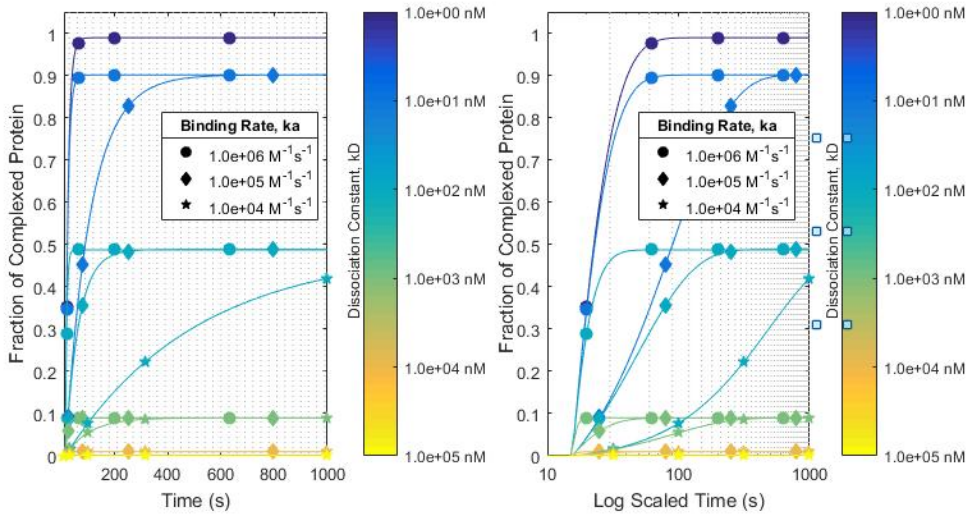
Of these more rapidly equilibrating interactions there are some that occur entirely within the first 30 second FCS measurement, and even those that do not will likely

be poorly described by subsequent measurements due to noisiness in the fractional composition measurement as an inherent product of the experimental process and data fitting required to produce these values. That being said, as aforementioned FCS measurement records data with nanosecond resolution, the problems stem from the need to average this data across a much longer timeframe.

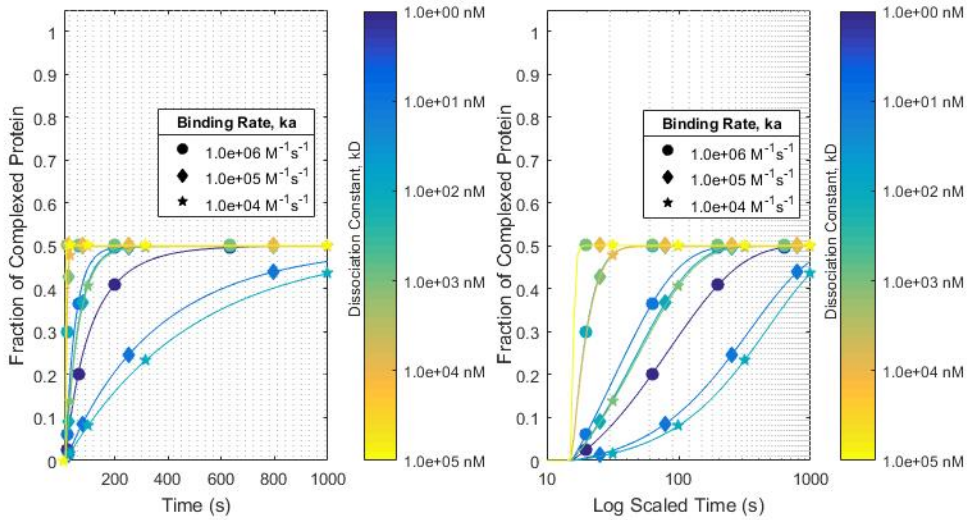
Thus far, we have only considered using FCS data as it's traditionally been applied, with each distinct photon count event being utilized only once and contributing only to the correlation curve of a single measurement. However, there is no reason that the same data points cannot be used to generate more than one correlation. If we take our measured photon count and perform a correlation of overlapping segments we should be able to resolve dynamics on a much finer timescale without compromising the SNR.

By introducing a 'sliding windowed' correlation algorithm we enable a new capability for rapid dynamic characterization to FCS technology. By correlating a small region, a frame, of the raw photon count trace we can characterize the diffusion of particles within that frame of a single experimental measurement. This frame can then be shifted slightly (by a single sample if desired) and a new correlation performed to characterize the change in diffusion between those two frames, and therefore the change in composition of the sample. While computationally expensive, this enables time resolution of a traditionally single measurement. From one experimental measurement of time t_M as many as $(t_M - t_f)/\Delta t$ time resolved points may be generated where t_f is the width of a correlation frame and Δt is the experimental sample rate.

In the next sections of this chapter I detail the software algorithm and graphical user interface that I designed to process data according to this theory which we coin as Fluorescence Rolling Correlation Spectroscopy (FRCS) and present this software as an open source tool for any lab enabled with single-photon FCS.



(a) Fraction of complex fAB over time. Initial concentrations fixed to $fA_0 = 10$ nM, $B_0 = 100$ nM.



(b) Fraction of complex fAB over time. Initial concentrations of species B are varied to obtain half saturation as ideal for FCS, and $fA_0 = 10$ nM.

Fig. 2.7. Simulated binding interactions of a fluorescent ligand fA, to a inert receptor B over a time course of minutes. Right panels show the same time course in log scale, grid lines indicate a 30 second period (as subsequent FCS measurements might capture). Plotted is the fraction of complex species fAB using various kinetics and concentrations reasonable for PPIs

2.3.2 Software Correlation

As discussed with hardware correlation it is unfeasible to compute $G(\tau)$ for every possible value of τ using a software correlation algorithm. The same limitations in number of computational steps still apply. Indeed, as aforementioned software correlators are inherently slower (per computation) than a hardware analog. That said, if real-time acquisition is not demanded, post-processing analysis using software correlation has no limitation beyond user preference as to the computation time for the correlation algorithm. Without the constraints of real-time processing and hardware register storage space, a software correlator is capable of attaining a much finer tau resolution. That being said, a multiple-tau approach is still typically utilized to optimize the ratio of computation time to correlation resolution. Here I utilize a BM algorithm similar to hardware correlation that integrates a discrete time-lagged dot product function developed by Gamari et al. [57].

The photon count raw data is first formatted into a vector of photon arrival times (PAT) with a distinct entry for each photon measured. This vector is then binned into B equivalent bins of width Δt_{n0} equal to the desired lowest tau value in the correlation sample (typically near to the smallest value in the discrete derivative of PAT). Notably, when binned the vector is shifted to arbitrary integer time units bin width Δt_n, to improve memory efficiency. The binned PAT vector, $bPAT_x$, is stored as a paired list of bin left-edges (t_x) containing a non-zero count of photons in the original PAT vector and the associated photon count for each bin (p_x) in the form $bPAT_x = [t_x, p_x]$. The length of this list is B_x, for a dense data set t_x is likely a nearly sequential list and $B_x \approx B$, but for sparse data there may be large jumps between time entries and $B_x << B$. In the latter case, this non-sparse storage may significantly reduce computational expense. For an auto-correlation this list is then duplicated, or in the case of a cross-correlation a second list is produced using an equivalent bin width from the second data channel. This second list is identical in format if not value and can be defined as $bPAT_y$=$[t_y, p_y]$ of length B_y and the two

can be compared against each other using a cross-correlation (equivalent to an auto-correlation if $bPAT_x$=$bPAT_y$) with a discrete approximation of equation (2.11) as described by Gamari et al. [57].

$$\hat{G}[\tau] = \frac{B \sum_i^{B_x} \sum_j^{B_y} p_x[i] p_y[j] \delta[\tau - (t_x[i] - t_y[j])]}{P_x P_y} \tag{2.26}$$

Here, as above, B is the total number of equal width bins (Δt_n) possible given the length of the experimental measurement ($t_M = B\Delta t_n$), P_x and P_y are the total photon counts represented in $bPAT_x$ and $bPAT_y$ respectively, and δ is the Kronecker delta function ($\delta = 1$ only if passed a zero, otherwise $\delta = 0$). Equation (2.26) is essentially equivalent to equation (2.16), although notably is more efficient for a sparse dataset due to comparison only of non-zero bins (accommodated by the inclusion of the Kronecker delta term). As with a hardware correlation however, reduced computation time is accomplished primarily by applying an overlapping, cascading multiple-tau correlation algorithm.

Equation (2.26) is used to compute $\hat{G}[\tau_m]$ with ($\tau_m = m\Delta t_n$, for n = 0: $\Delta t_n = \Delta t_{n0}$) for m = (0, 1, 2,..., M). The two bPAT lists are then re-binned to a bin width of $\Delta t_n = \gamma^n \Delta t$ and $\hat{G}[\tau_m]$ are determined using the newly binned lists for m = ($M/\gamma + 1$,$M/\gamma + 2$,..., M). This process is repeated for blocks n = (0,1,2,...,N) or until a desirable maximum value of τ_m has been achieved. A block diagram of this multiple-tau correlation scheme can be seen in figure 2.8 below.

$\tau_{m,n} = \tau_0(\gamma^n)(m+1)$

n	m													
0	0	1	...	M-1										
1	0 … M/γ+0				M/γ+1	M/γ+2	...	M-1						
...	…								...					
N-1	0 … M/γ+0										M/γ+1	M/γ+2	...	M-1
τ	τ_0	$2(\tau_0)$	...	$M(\tau_0)$	$M\tau_0 + \gamma(\tau_0)$	$M\tau_0 + 2\gamma(\tau_0)$	...	$M\gamma(\tau_0)$	...		$M\gamma^{N-1}(\tau_0) + \gamma^N(\tau_0)$	$M\gamma^{N-1}(\tau_0) + 2\gamma^N(\tau_0)$	...	$M\gamma^N(\tau_0)$

Fig. 2.8. Block Diagram of a multiple-tau scheme that applies the BM Correlation approach.

2.3.3 Python 3.6 Implementation of FRCS

An implementation of this algorithm with reasonable benchmarking times for the generation of many correlation frames, was achieved using custom python scripts that utilize both (i) high order data handling tools from pre-built python packages including Numpy ([68]) and Scipy ([69]) and (ii) integrated functions written and compiled in C++ using the Ctypes library ([65]) to allow more rapid computations than achievable with interpreted Python 3.6. This software and more complete documentation is available open source via the GitHub repository (see appendix A.1). This chapter will discuss these algorithms only at a high level, further detail is available in the documentation.

The base process of this software is a function that takes four inputs describing the intended correlation. These inputs are (i) **srcFiles**: a list of file names for files containing the of raw output of identical FCS measurements (photon arrival times recorded with arbitrary units), (ii) **srcDir**: the hard drive location of the directory containing these files, (iii) **destDir**: the hard drive location of a directory to output correlated traces and (iv) **corrParams**: the correlation parameters. The **corrParams** input is a data structure containing values for the length of experimental measurement (t_M), desired τ resolution (N, M, and γ), and the desired width (t_f)

and number of overlapping correlation frames to generate for each sample. Each file in **srcFiles** is handled separately due to the size of the data vector from a single measurement, which are characteristically around 2 megabytes per million photons counted.

Algorithm 1 funcFRCSAnalysis($srcFiles, srcDir, destDir, corrParams$)

Ensure: All passed variables are of expected format.

```
corr ← corrClass(corrParams)
for file in srcFiles do
    if file.channels > 1 then
        PAT_x ← parseRaw(file, channel1)
        PAT_y ← parseRaw(file, channel2)
    else
        PAT_x ← parseRaw(file, channel1)
        PAT_y ← PAT_x
    end if
    Ĝ ← corr.CCF(PAT_x, PAT_y)
    destDir/file.frcs ← write(Ĝ)
end for
```

Within this base process two simple helper functions are used. A function **parseRaw** detects the format of the input data file (currently supports formats only for Zeiss '*.raw' files and '*.PAT' files from my own simulated FCS data as discussed in chapter 3) and converts the file to a standardized PAT vector with units of seconds. A second function **write** exports correlation traces as comma separated lists to the **destDir** with a custom file identifier of '.frcs'. However, the meat of the computation occurs within an instantiated correlator class (**corrClass**) that takes as input the correlation parameters and contains a sub-method (**CCF**) to perform identical rolling cross correlations of PAT1 and PAT2 for each **file** in **srcFiles**.

Algorithm 2 corrClass.CCF(PAT_x, PAT_y)

Ensure: corrClass is instantiated with appropriate variables including:

corrClass.t_M - The length of experimental measurement

corrClass.$frameEdges$ - Measurement time left edges of desired corr frames

corrClass.t_f - Width of a corr frame (10-30 seconds)

corrClass.$tauResolution$ - Parameters describing the desired τ resolution

corrClass.τ - Vector of all tau values based on **corrClass**.$tauResolution$

$fIndex \leftarrow 0$

for f **in** $frameEdges$ **do**

  $fIndex \leftarrow fIndex + 1$

  $bM \leftarrow 0$

  for n **in** [0, 1, 2, ... **corrClass**.$tauResolution.N$] **do**

    $\Delta t_n \leftarrow$ **corrClass**.$\tau[0] * \gamma^n$

    $B \leftarrow$ **corrClass**.$t_f / \Delta t_n$

    $fL \leftarrow f$

    $fR \leftarrow f+$**corrClass**.t_f

    $bPAT_x, bPAT_y \leftarrow$ **corrClass.binPAT**$(PAT_x, PAT_y, \Delta t_n, (fL, fR))$

    $B_x \leftarrow$ **length**$(bPAT_x)$

    $B_y \leftarrow$ **length**$(bPAT_y)$

Aside from a few aptly named helper functions this correlation is driven primarily by iterative application of two subfunctions. For each frame, **corrClass.CCF** works through subsequent blocks n in the BM algorithm, rebinning the PAT vector as it proceeds from block to block. Within each block, each tau register m is evaluated to compute a normalized time lagged dot product according to equation 2.26. Due to the number of computations involved, performing this function within interpreted Python 3.6 is not ideal. Faster benchmarking is achieved by evaluating this function (**cppFCSLib.tLDP**) from a custom written C++ library that is integrated via the

```
18:       P_x ← sum(bPAT_x)
19:       P_y ← sum(bPAT_y)
20:       for m in [0+bM, 1+bM, 2+bM, ... corrClass.tauResolution.M] do
21:         tauIndex ← m + n * (M − bM)
22:         τ_m ← corrClass.τ[tauIndex]
23:         Ĝ[tauIndex, fIndex] ← cppFCSLib.tLDP(τ_M, B_x, B_y, bPAT_x, bPAT_y)
24:         Ĝ[tauIndex, fIndex] ← Ĝ[tauIndex, fIndex] * B/(P_x * P_y)
25:       end for
26:       bM ← ceil(M/γ)
27:     end for
28:   end for
29:   Return: Ĝ
```

Ctypes Python package. This implementation is based upon the algorithm presented in the same publication that details equation 2.26 [57].

Python 3.6 does however present a convenient set of tools for manipulating large data sets which can be used to efficiently rebin the PAT lists into shorter, frame specific lists in the style of $bPAT_{x/y}$ between each correlation block. This is performed by a relatively straightforward helper function built into the correlator class **corrClass.binPAT** which takes as inputs (i) the PAT vectors, (ii) desired bin width, and (iii) left/right edges of the current correlation frame in experimental time (fL,fR) and exports a binned list in the format of $bPAT_{x/y}$ as described above.

Shown in figure 2.9, this algorithm is capable of rapidly computing many correlation frames for a single measurement, with execution times ranging from 0.4-0.6 s·kHz^{-1} per frame for correlation frames of a length reasonable to generate an appropriate SNR. For a characteristic experiment with correlation frames of 30 seconds and photon count rate in the range of 50kHz, this would indicate an execution time of $\sim$ 20 s/frame.

Algorithm 3 cppFCSLib.tLDP$(\tau_m, B_x, B_y, bPAT_x, bPAT_y)$

Ensure: All passed variables are of C++ **long long** type and that $bPAT_x$ and $bPAT_y$ are contiguous 2D arrays of dimension $2xB_x$ and $2xB_y$ respectively.

```
productSum ← 0
i ← 0
j ← 0
while i < B_x and j < B_y do
    lag ← bPAT_x[1][i] − bPAT_y[1][j] − τ_m
    if lag == 0 then
        productSum ← productSum + bPAT_x[2][i] * bPAT_y[2][j]
        i ← i + 1
        j ← j + 1
    else if lag < 0 then
        i ← i + 1
    else
        j ← j + 1
    end if
end while
Return: productSum
```

Algorithm 4 corrClass.binPAT$(PAT_x, PAT_y, \Delta t_n, (fL, fR))$

Ensure: PAT_x and PAT_y are both photon arrival time lists with arbitrary units (AU, seconds equivalent most easily estimated from the last photon arrival time and experimental measurement length, **corrClass**.t_M).

```
firstBool ← True
for PAT in [PAT_x, PAT_y] do
    unitConversion ← PAT[end]/corrClass.t_M
    B_frame ← (fR − fL)/Δt_n
    fL_bin ← fL/Δt_n
    Δt_n,AU ← unitConversion * Δt_n
    binRange ← np.arange(1, B_frame, 1) + fL_bin
    t_bin, p_bin ← np.unique(np.digitize(PAT, binRange), return_counts=True)
    {Note this function returns two lists, but the first and last values in each list
    contain data for every photon counted on either side of the correlation frame and
    must be removed!}
    if firstBool then
        bPAT_x ← np.stack(t_bin[1 : −1], p_bin[1 : −1])
        firstBool ← False
    else
        bPAT_y ← np.stack(t_bin[1 : −1], p_bin[1 : −1])
    end if
end for
Return: bPAT_x, bPAT_y
```

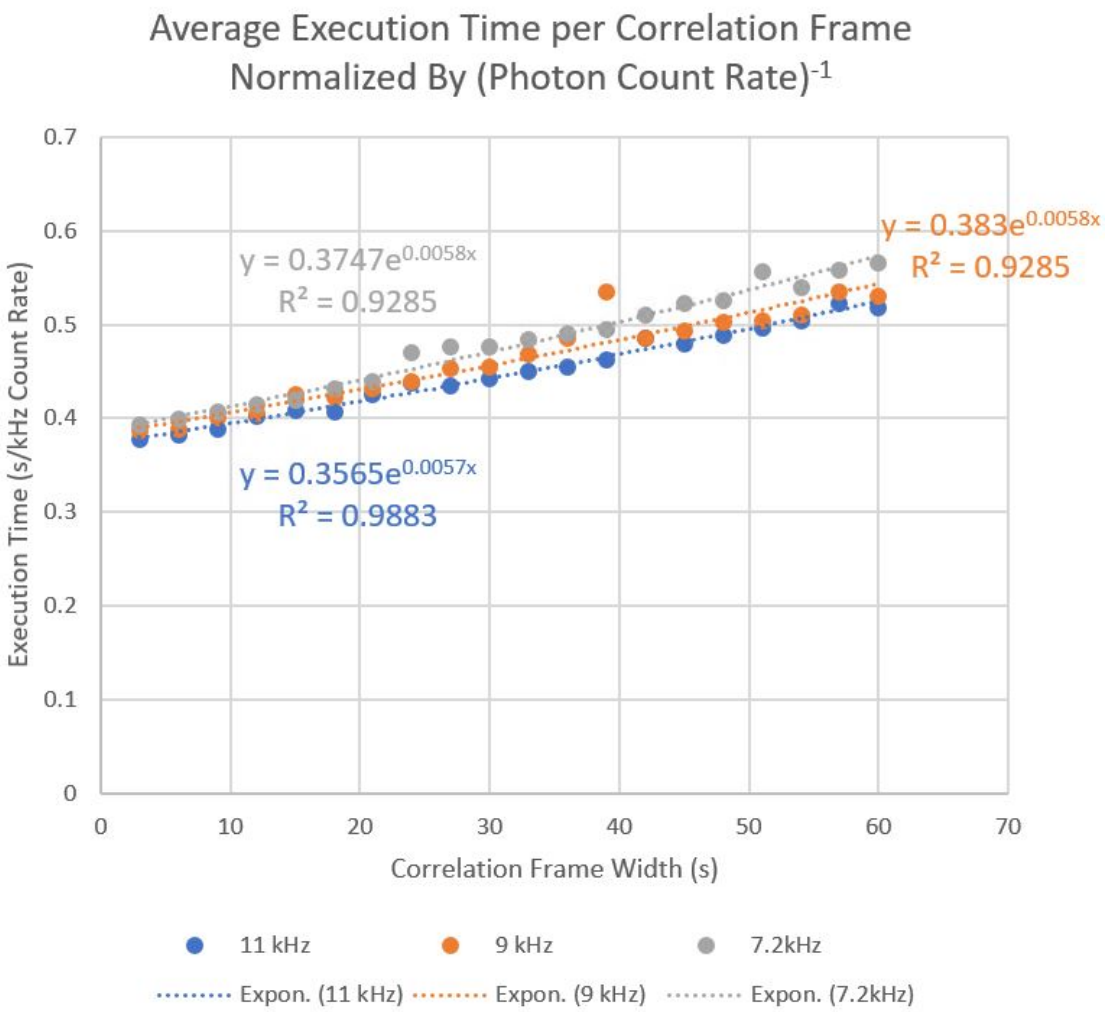


Fig. 2.9. Benchmarking of FRCS algorithm implemenation in Python 3.6. On the x-axis is the width in seconds of the correlation frame, on the y-axis is the average evaluation time for a frame of each width selected randomly from within several seeds of simulated experimental measurements and normalized by average photon count rate for each simulation (n = 10: 2 frames/seed x 5 seeds/photon count rate; 3 simulated photon count rates).

2.3.4 FRCS Data Fitting Algorithm and Workflow

Evaluation of FRCS binding data requires two levels of parameter fitting. The first level is a series of non-linear regression steps that apply an FCS fitting model as described in section 2.1.3 to resolve the fractional composition of free ligand and complexed ligand-receptor in each correlation frame. The second level is a single non-linear regression step that attempts to resolve binding kinetics from this fractional composition, the initial concentrations in the system, and experimental parameters such as fluorescent bleaching which can also be extracted from the fitting models used in the first level. A multilevel fitting approach such as this does introduce a significant potential for fitting error, however we minimize this potential by using a robust fitting workflow that seeks to minimize sources of fitting error.

As described in section 2.3.1 this technique applies a fitting model that assumes a two-species system with a rapidly diffusing free ligand and a slower diffusing complex species. For the purposes of this work, I assume that the complex doesn't introduce a significant amount of fluorescent quenching, under the assumption that the fluorophore label is designed such that it exists a sufficient distance from the binding domain so as not to interfere with binding behavior. In this case, $q_0 = q_1$ or $\epsilon_0 = \epsilon_1$. For experimental measurements, I do include a system correction factor, although for evaluation of simulated FCS data as discussed in greater detail in chapter 3 the correction factor S is railed to 0 (along with a few other parameters as discussed further below) during fitting because this behavior is not enabled in preliminary simulations. The full form of the fitting model used for all measurements is given below.

$$G(\tau) = G(0)(1 + Se^{-\tau/\tau_S}) \left(\sum_{i=1}^{2} \frac{F_i}{(1 + (\tau/\tau_{D,i}))(1 + \alpha^{-2}(\tau/\tau_{D,i}))^{1/2}} \right) + G(\infty) \quad (2.27)$$

For the first level of fitting, each generated FRCS frame is assigned a distinct **Parameters** structure from the Lmfit non-linear regression Python package which is generated and stored into a nested dictionary **frameParams** keyed by first the

file name of the experimental measurement and then by the experimental-time, left-edge of the correlation frame. These Parameters structures essentially contain a set of information describing each parameter in equation 2.27 including initial values, upper/lower bounds, and a boolean to determine whether each parameter is varied during fitting. For all fitting on this level, the same general workflow is used with differing variable parameters of interest: first a rough least squares fitting step is used to appropriately fit the asymptotes ($G(0) and G(\infty)$), then a differential evolution step is used to break any local minima, and a final least squares fitting step is used to tune the less sensitive (but most important) parameters. Precise details on the fitting algorithms can be found in the Lmfit documentation [66].

The model in equation 2.27 is first used to fit calibration measurements of two single particle systems to resolve the (i) structural geometry of the CV and (ii) diffusion coefficient of the free ligand in isolation. For these calibration measurements, F_1 can be railed to 1, F_2 can be railed to 0, $\tau_{D,2}$ can be railed to any non-zero number, and the same FRCS experimental correlation paradigm (tau resolution parameters, experimental measurement length, correlation frame length, and initial concentration of ligand) can be used as for the planned experimental measurements of binding.

In the first calibration, a freely diffusive fluorophore of similar excitation wavelength and with a known concentration and diffusion coefficient in the binding medium is monitored. Applying FRCS allows for more robust fitting of even a simple system like this by allowing a single experimental measurement to generate numerous replicate correlation traces. For this calibration, $G(0)$ and $G(\infty)$ are allowed to vary between frames, but the values of α, τ_D, S, τ_S are fit globally for each measurement file. Depending on the fluorophore and laser intensity, fluorescent bleaching may occur causing $G(0)$ to decrease with successive correlation frames. In this case, the true value for the number of particles in the CV ($P = 1/G(0)$) should be approximated from the output of only the first few frames. From the known initial concentration (C_{cal}), diffusion coefficient ($D_{c,cal}$), and the output values of characteristic residency ($\tau_{D,cal}$) and number of particles in the CV (P_{cal}) the geometry of the CV can be

defined, referring back to definitions in section 2.1.3. Recall that α is the structural parameter describing the confocal arrangement. As a figure of merit, this value should typically be between 1-10 depending on the microscope setup and tends to wildly vary if allowed to fit unbounded. It is recommended to repeat this fitting step until values of α, ω_{xy}, ω_z, and P_{cal} all adhere relatively closely to the expected relationships between them given the known concentration and diffusion coefficient of the calibration fluorophore.

$$\omega_{xy}^2 = 4D_{c,cal}\tau_{D,cal} \tag{2.28}$$

$$V_{eff} = \frac{P_{cal}}{C_{cal}} \tag{2.29}$$

$$= \pi^{3/2}\omega_{xy}^2\omega_z \tag{2.30}$$

$$= \pi^{3/2}\omega_{xy}^3\alpha \tag{2.31}$$

$$\alpha = \frac{P_{cal}}{C_{cal}\pi^{3/2}\omega_{xy}^3} \tag{2.32}$$

$$= \frac{P_{cal}}{C_{cal}\pi^{3/2}(4D_{c,cal}\tau_{D,cal})^{3/2}} \tag{2.33}$$

With determined values for the perpendicular radius (ω_{xy}) and structural parameter (α), the second calibration can then be used to determine the characteristic diffusion coefficient (and concentration) of the free ligand. For this calibration and all following experimental measurements the structural parameter can now be fixed to the value determined in the first calibration, to reduce the flexibility and generate the most accurate residency time for the free ligand species ($\tau_{D,1}$). As before, $G(0)$ and $G(\infty)$ are allowed to vary between frames, but the values of $\tau_{D,1}$, S, τ_S are fit globally for each measurement. From this global fitting, values for $\tau_{D,1}$, S, and τ_S can be determined to use for all of the subsequent experimental fits of actual two-particle binding. In this case also, the fluorescent bleaching, if observed to be significant as in figure 2.10, is a parameter of interest for the second level of fitting for binding kinetics. Assuming such, the declining values of $G(0)$ should be fit separately against an exponential decay model with a horizontal asymptote that represents an equilib-

rium where fluorescent bleaching within the CV is matched by the introduction of new fluorophores from the surrounding solution (which can be considered undepleted over a short time scale).

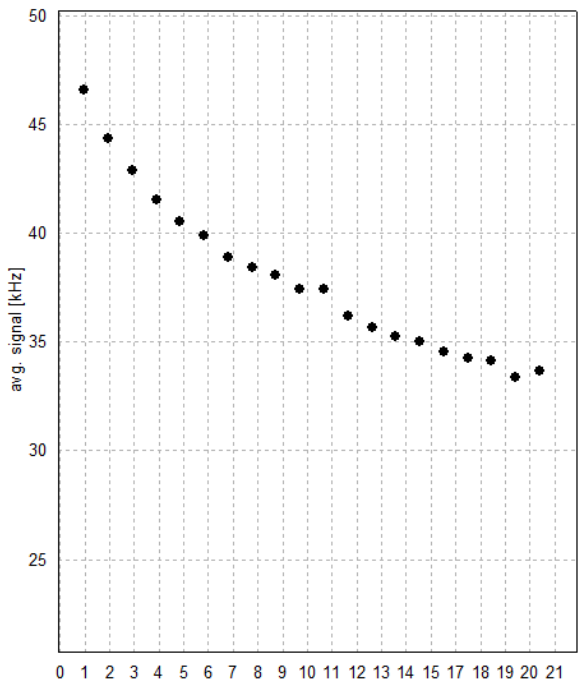


Fig. 2.10. Example of fluorescent bleaching from a preliminary experimental measurement of freely diffusive Alkyne-Alexafluor 647nm. Shown is the photon count rate, declining between linearly spaced correlation frames (n=20) during a 10 minute measurement. Note that Alexafluors are engineered to exhibit robust bleaching resistance but this timescale may be significantly shorter for other fluorophores.

This bleaching fit will be discussed in greater detail as a component of the non-linear regression for binding kinetic determination further below. However, the gen-

eral equation for fitting of this bleaching behavior is an exponential decay with a horizontal asymptote.

$$[A] = [fA] + [bA] \tag{2.34}$$

$$[fA](t) = ([fA]_0 - bM_A)e^{-t(bR_A)} + bM_A \tag{2.35}$$

$$[bA](t) = ([fA]_0 - bM_A)(1 - e^{-t(bR_A)}) \tag{2.36}$$

Here bA is the concentration of bleached free ligand, and bR is the characteristic rate of bleaching, and bM the stable concentration of [fA] at which bleaching is matched by influx of new fluorophores.

With the two calibration fits performed, there are now calibrated parameter values for α, S, τ_S, and $\tau_{D,1}$. If characterizing an interaction between a ligand/receptor pair with a known strong binding affinity that will approach saturation at attainable concentrations of the receptor protein, a third calibration similar to the second can be performed with labeled ligand in the presence of excess receptor to further characterize the diffusion of the complex $\tau_{D,2}$, otherwise this parameter will need to be fit simultaneously with the composition fractions, which may result in skewing of the fraction determinations. In this case, to limit this risk, $\tau_{D,2}$, should be estimated from hydrodynamic radius of the larger protein and bound to a reasonable range according to published findings. Regardless of how $\tau_{D,2}$ is parameterized, the final phase of the first level of parameter fitting is a fit of the fractional composition of free ligand F_A and complex F_AB for each frame of an experimental measurement where ligand and protein are mixed either just before, or during the timecourse of the experiment. A representative example of this fitting step is depicted in figure 2.11 below from a simulated binding reaction between two particles.

With all correlation frames parameterized, these parameters can be passed to the second level of fitting to determine the dynamics of the system across the time course represented by the series of distinct correlation frames each representing a time point in the binding interaction. This is accomplished with a non-linear re-

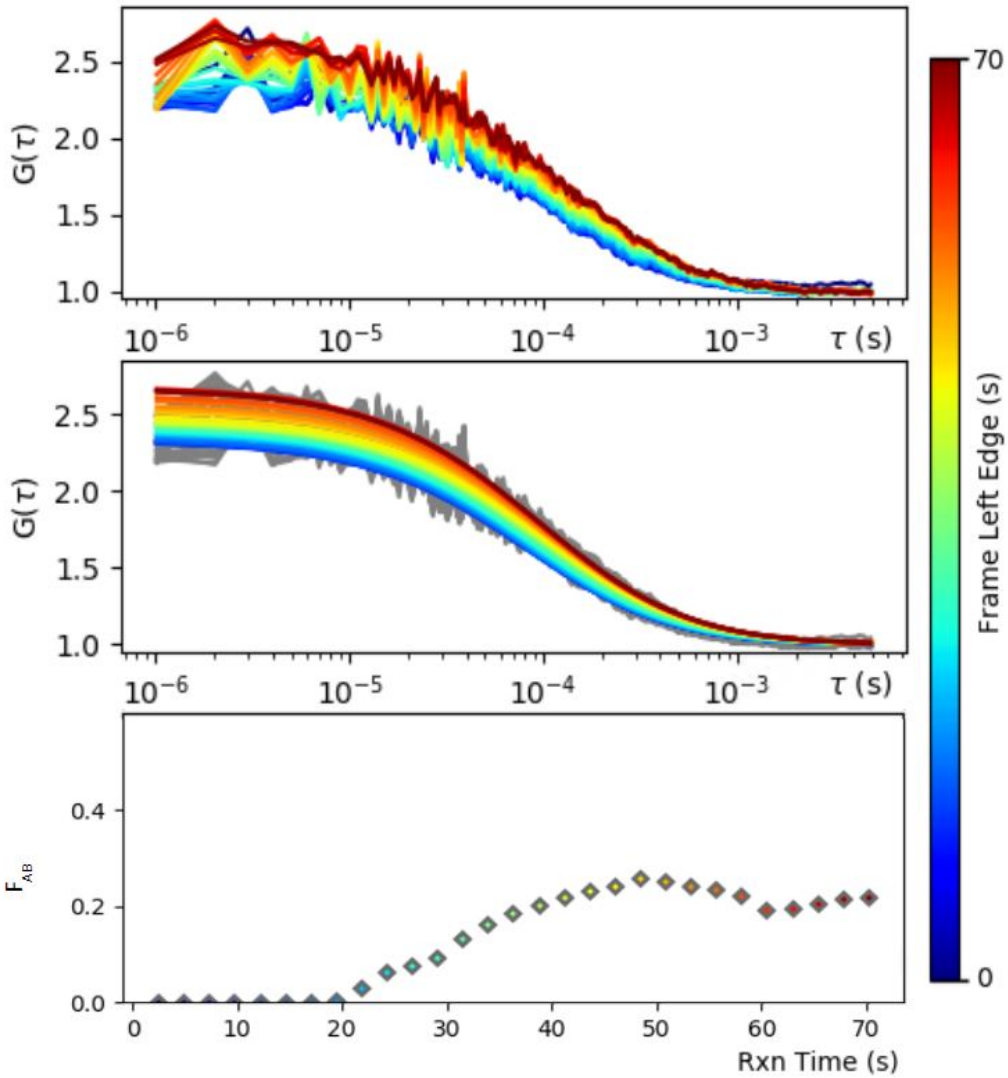


Fig. 2.11. FRCS Evaluation of a simulated binding interaction between two proteins. In this figure reaction time is defined as the left edge of each correlation frame of width 20 s. Free ligand introduced to system at reaction time t = 0, receptor introduced at t = 30 s. Top: All FRCS generated correlation frames. Middle: Fit correlation traces for each frame using fitting model described in equation 2.27. Bottom: Resulting fractional composition of complex as a function of reaction time. Simulated binding data generated as discussed in chapter 3 for a binding interaction with k_a = 2e6 $M^{-1}{\cdot}s^{-1}$, k_d = 5e−2 s^{-1}, $[A]_0 = [B]_0$ = 10 nM, $D_{c,A}$ = 1.5e−10 $m^2{\cdot}s^{-1}$, and $D_{c,B} = D_{c,AB}$ = 0.75e−10 $m^2{\cdot}s^{-1}$

gression according to numerically integrated system of ordinary differential equations describing the binding interaction of ligand to receptor as discussed in section 2.3.1 (equations 2.20-2.25). In the case of observed bleaching this model of binding can be further supplemented with a bleached population of fA and fAB by introducing two more equations to this system. Given that the rate of fluorescent bleaching is assumed to be relatively proportional to the number of absorbed photons per unit time, and occurs at a timescale longer the the typical residency times of any of these particles, it is biophysically realistic to assume that as the complex species becomes more populous, with characteristically longer residency times within the CV, the local fluorophore concentration will bleach at a proportionally faster rate than the free ligand and stabilize at an proportionally lower unbleached concentration (as distal fluorophore enters the local region of the CV). Therefore, if bleaching is observed and parameterized for the free ligand calibration, the parameters for the complex species can be approximated from the determined diffusion of each species.

$$bR_{AB} = bR_A \frac{\tau_{D,AB}}{\tau_{D,A}} \tag{2.37}$$

$$bM_{AB} = bM_A \frac{\tau_{D,A}}{\tau_{D,AB}} \tag{2.38}$$

A final system of ODEs that incorporates this bleaching behavior can be described, and numerically integrated with known initial concentrations of the ligand [A] and receptor [B] over the time scale of the FRCS measurement.

$$[fA](t) = [A](t) - [bA](t) \tag{2.39}$$

$$[fAB](t) = [AB](t) - [bAB](t) \tag{2.40}$$

$$F_A(t) = [fA](t)/([fA](t) + [fAB](t)) \tag{2.41}$$

$$F_{AB}(t) = [fAB](t)/([fA](t) + [fAB](t)) \tag{2.42}$$

$$\frac{d[A](t)}{dt} = k_d[AB](t) - k_a[A](t)[B](t) \tag{2.43}$$

$$\frac{d[B](t)}{dt} = k_d[AB](t) - k_a[A](t)[B](t) \tag{2.44}$$

$$\frac{d[AB](t)}{dt} = k_a[A][B](t) - k_d[AB](t) \tag{2.45}$$

$$\frac{d[bA](t)}{dt} = ([fA](t) - bM_AF_A(t))bR_A \tag{2.46}$$

$$\frac{d[bAB](t)}{dt} = ([fAB](t) - bM_{AB}F_{AB}(t))bR_{AB} \tag{2.47}$$

From this integration, time points can be extracted to fit against the fractional composition of the fluorescent species determined for each correlation frame. The integration can then be iterated in order to determine best fit values for the two unknown binding kinetic parameters k_a and k_d for the system. I note that for a system that starts unbound and drives to equilibrium, k_a tends to be well characterized and k_d tends to be more flexible. If desired, a stable equilibrium measurement can be used to evaluate K_D from determined species concentrations. This value can then be used with the dynamically determined k_a to select a more rigid k_d value ($k_d = K_D k_a$).

An algorithmic representation of this fitting workflow can be found in algorithm 5 below.

Algorithm 5 Fitting Workflow

1: **1st Level of Fitting (FRCS Model Fitting eq. 2.27):**

2: $\omega_{xy}, \omega_z, \alpha \leftarrow$ 1st Cal.: Fluorophore with known $D_{c,cal}$ & C_{cal} (eqs. 2.29-2.33).

3: **if** $G(0)_{start} << G(0)_{end}$ **then** {Bleaching Likely:}

4: $\tau_{D,A}, S, \tau_S, bR_A, bM_A \leftarrow$ 2nd Cal.: Labeled ligand (fA) in isolation using the determined ω_{xy} & α.

5: **else**{Bleaching Insignificant:}

6: $\tau_{D,A}, S, \tau_S \leftarrow$ 2nd Cal.: Labeled ligand (fA) in isolation using the determined ω_{xy} & α.

7: **end if**

8: **if** Expected $K_D << [B]$ **then**

9: $\tau_{D,AB} \leftarrow$ 3rd Cal.: fA mixed with excess of receptor (B) measured after a long incubation ($[fA] \approx 0, [fAB] \approx [fA]_0$).

10: $F_A(t), F_{AB}(t) \leftarrow$ Experimental FRCS: fA mixed with a few different concentrations of B just before or during measurement, frames exhibit fraction of complex fAB formed during time course.

11: **else**

12: $F_A(t), F_{AB}(t), \tau_{D,AB} \leftarrow$ Experimental FRCS: fA mixed with a few different concentrations of B just before or during measurement, frames exhibit fraction of complex fAB formed during time course.

13: **end if**

2nd Level of Fitting (FRCS Model Fitting eq. 2.27):

if $G(0)_{start} << G(0)_{end}$ **then** {Bleaching Likely:}

  $k_a, k_d, K_D \leftarrow$ Numerical integration of eqs. 2.40-2.47. Use $F_A(t), F_{AB}(t), \tau_{D,A}, \tau_{D,AB}, bR_A, bR_AB, bM_A, bM_AB$ as determined from **1st Level of Fitting**.

else{Bleaching Insignificant:}

  $k_a, k_d, K_D \leftarrow$ Numerical integration of eqs. 2.40-2.47. Use $F_A(t), F_{AB}(t)$ as determined from **1st Level of Fitting**.

end if

2.3.5 Graphical User Interface for FRCS Experiments

Due to the relative complexity of the underlying principles, a graphical user interface (GUI) software implementation was created to accelerate data processing from raw measurement files, to correlated FRCS traces, and through both levels of fitting as described in sections 2.3.2-2.3.4. This software is developed to be intuitive and enable FRCS experiments without needing to modify the code or algorithm, making this technique available for use in any lab enabled with single-photon confocal microscopy and reasonable computing resources. This GUI was developed using the Tkinter package in Python 3.6. Full documentation of the tools available in this GUI is available in the the GitHub repository (see appendix A.1), however here we will mention briefly the overall functions available.

The FRCS GUI has two tabs, one for generation of correlation traces, and one for fitting of these traces using the multilevel fitting approach as described in the previous section.

Shown in figure 2.12 the correlation tab allows the user to select a directory containing all calibration and experimental measurements from a single FRCS study. The user can then organize these files as desired, and determine the correlation paradigm, including tau resolution and number of frames per measurement to generate. These correlations can then be performed in place via the GUI, or the GUI can export .bash script files to perform the computations via a job scheduled computational cluster.

The second tab of the GUI as seen in figure 2.13 allows the user to select a directory containing the output correlation files (saved as '.frcs' files) and perform fitting as desired. This tab contains fields for all possible fitting parameters in the model described above, and allows the user to customize the fitting ranges and fixed variables for each correlation being evaluated. The GUI displays both the raw traces and best fit traces, along with a graphical display of the fractional composition over the timecourse of the FRCS measurement, and tabular best fit parameters for each frame.

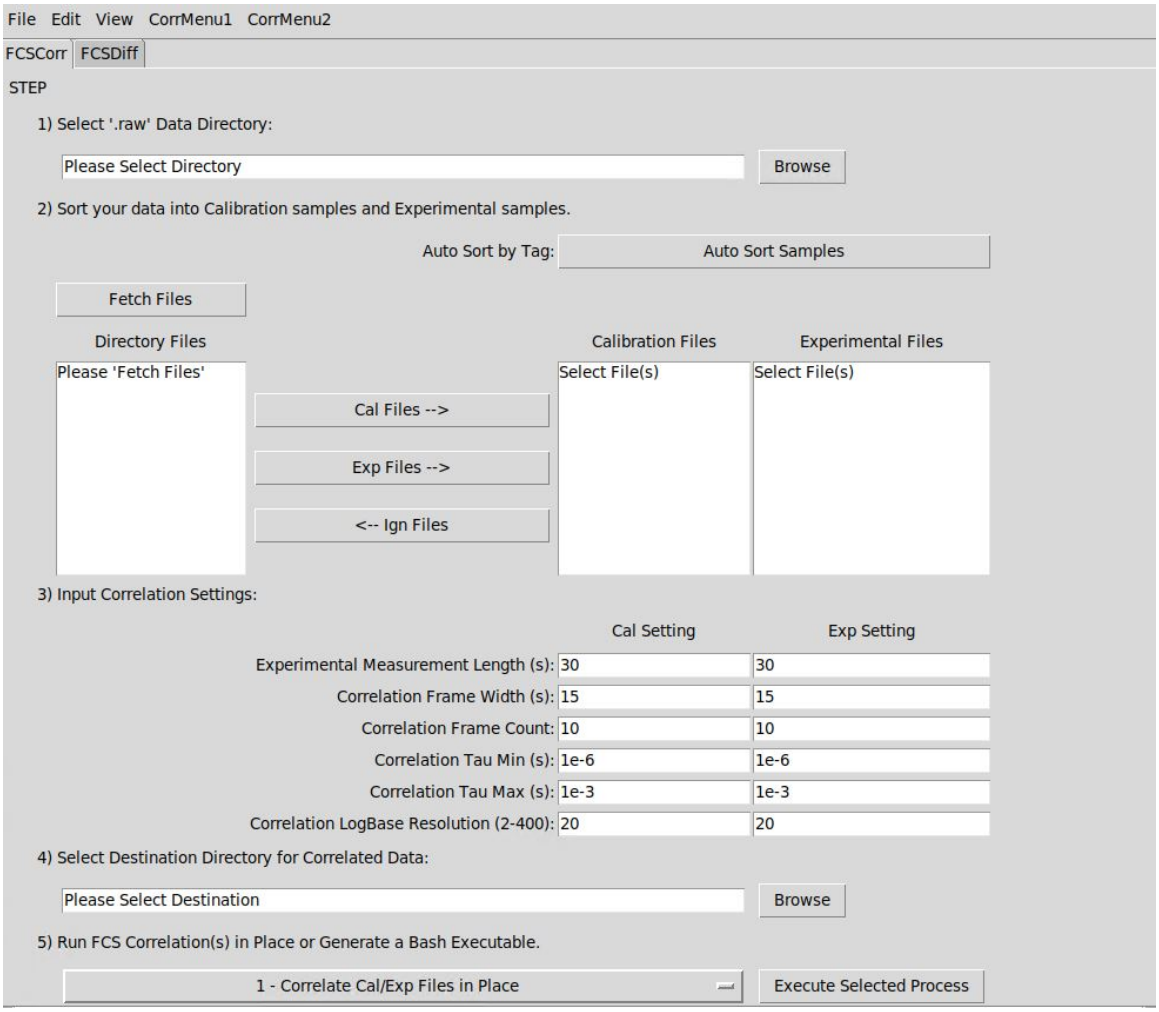
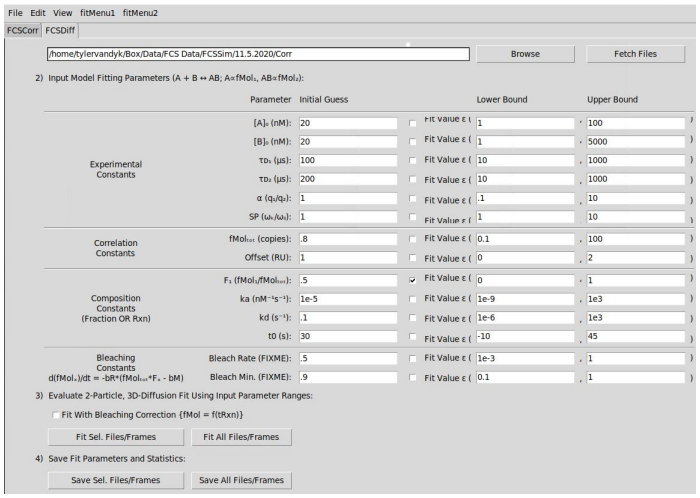


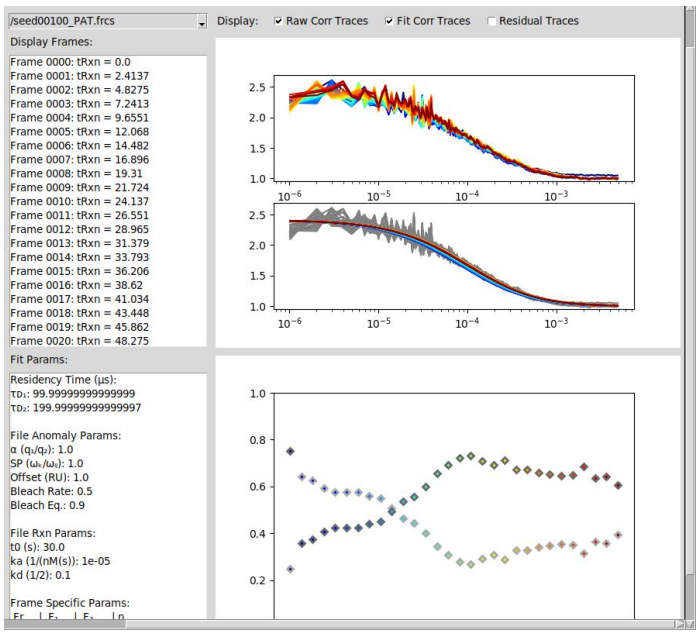
Fig. 2.12. GUI Correlation Tab, contains fields for the source/destination directories and desired correlation parameters as well as organizing functionality for experimental measurement files.

2.4 Conclusions

In this Chapter, we report for the first time the application of a rolling correlation window for characterization of binding reaction progress of biomolecules using FCS. FRCS presents an entirely new tool for the characterization of protein-protein interactions that bears advantages in cost, conditions, and time over previously available gold standard methods. We explain in detail the algorithms for processing of FRCS



(a)



(b)

Fig. 2.13. GUI Fitting Tab, (a) contains fields for correlation file directory as well as all fitting parameters including initial values, fitting constraints and variance boolean check boxes. Also shows (b) graphical and tabular representation of best fit parameters.

data, and present a graphical user interface to make this tool accessible for appropri-

ately enabled biomolecular lab groups. In the next chapter, we demonstrate applied FRCS using spatial stochastic simulations of binding interactions and preliminary experimental measurements of protein binding.

3. PRELIMINARY SIMULATIONS AND EXPERIMENTAL FINDINGS OF FCS/FRCS

This chapter discusses preliminary findings accomplished using the developed techniques and algorithms in the previous chapter. Here I discuss results of both spatial stochastic simulations of bimolecular interactions and preliminary experimental measurements to justify possible future work with this tool.

3.1 Introduction

3.1.1 Background and Significance

Using FRCS, I seek to elucidate the binding affinity and kinetics of subsets of PPIs that our group has previously studied in detail. However, prior to it's application for kinetic determination, the technique first needed to be validated both in theoretical application and biophysical feasibility. To accomplish these validations, FRCS was first applied to simulated reactions of biomolecular behavior using methods similar to work our lab has done in the past [18,19]. By applying *in silico* simulations of this behavior, we are able to rapidly characterize the dynamic range of FRCS, without requiring extensive experimental design with well characterized binding interactions.

Furthermore, I explore here preliminary wetlab experiments to character the behavior of a protein of great interest to our lab: calcium modulated inhibitory protein, calmodulin (CaM). CaM is a highly conserved and ubiquitously expressed protein, particularly dominant in systems of cellular signaling that utilize Ca^{-2} ions as secondary messengers [70]. For studies of binding kinetics, CaM is particularly interesting due it's Ca^{2+} modulated behavior. CaM is known to have 4 Ca^{2+} binding sites, 2 proximal to each terminus, and exhibits unique geometry and binding depending on the saturation of these sites [70, 71]. Furthermore, CaM is well known to have promiscuous modulatory interactions with over 100 binding partners [72–74]. While many of these actors exhibit maximal affinity to fully saturated CaM with 4 bound

Ca^{2+} (CaM_4), it has been proven that unsaturated CaM, apo-CaM (CaM_0) demonstrates higher affinity with some proteins, and that partially saturated CaM species such as CaM with 2 Ca^{2+} bound at the C-terminus (CaM_2C) or Ca^{2+} bound at the N-terminus (CaM_2N) also exhibited modified modulatory behavior [71, 75]. Indeed, while these subspecies of partially saturated CaM are more transient due to high Ca^{2+} affinity, it is likely that each of the 9 total possible configurations of Ca^{2+} saturation demonstrate unique binding that may play a role in the sub-second dynamics of cellular signaling.

Our group has previously noted that CaM promiscuity makes it an ideal target of *in silico* computational studies seeking to elucidate the potential roles of competition in complex biochemical environments, that are difficult to probe with traditional experimental techniques [18, 19]. Conveniently, as a result of it's highly conserved nature, dominant concentrations and high stability, CaM's binding relationships have been extensively characterized and even parameterized to some extent as needed for computational models in the style of Michaelis-Menten [18]. However, parameter spaces for partially saturated CaM states have not been nearly as thoroughly flushed out, as have those for binding of CaM_4 or even CaM_0.

In this chapter, we apply our newly developed FRCS technique to both *in silico* and *in vitro* models of protein binding to provide preliminary characterization of both the technique, and of CaM behavior with a few well characterized binding partners, such as calcineurin (CaN) and CaM-dependent kinase II (CaMKII). This characterization serves as a validation for the efficacy of the novel FRCS technique, but also enables further study of the influence of Ca^{2+} saturation state on CaM binding.

3.1.2 Fluorescent Labeling of Proteins

A notably important component of FCS experimental design is the fluorescent labeling that makes particles observable. Fluorescent labels come with a broad variety of physical properties that should be considered as a component of the experimental

design. For instance, GFP derived labels are all proteins in their own right, relatively large beta barrels around 240AAs (∼30kDa) in length whose fluorescent properties are distinguished by relatively small, site-directed mutagenesis of the original sequence derived from *A. victoria* [76, 77]. Therefore while GFP derived fluorophores are a powerful tool, they are not ideal for a study of diffusion, as simply attaching a GFP label may significantly increase a labeled biomolecule's R_H. Instead, smaller fluorescent dyes such as those in the Cyanine or Tryarylmethane families, which tend to be closer in size to a short peptide (∼1-2kDa) are generally preferable due to their proportionally reduced influence on the diffusion of a labeled particle.

Size is not the only factor of consideration however, fluorophores also differ in wavelength (color), quantum yield (brightness) and extinction rate (bleaching). All of these components can be relevant to experimental design. Wavelength must be tuned to match the available equipment, but also needs to be considered if studying multiple fluorescent labels (as in FCCS) to avoid potential cross-talk. Quantum yield limits the excitation probability of fluorophores and must be accommodated by tuning the laser intensity and concentrations of fluorophore (and therefore the measurement volume) to achieve a desirable SNR. Similarly, given the small sample volume and concentrations of a typical FCS experiment, fluorescent bleaching can quickly reduce the number of visualized fluorophore and require tuning of the laser intensity to reduce this artefact. With all of these additional conditions in mind, new generation fluorophores such as the Alexa Flour family, with the characteristic low molecular weight of biomolecular dyes, but with a broad selection of wavelengths each boasting improved quantum yield and reduced extinction are particularly ideal for FCS.

Attaching these fluorophores to molecules of desired study has been accomplished with a variety of bioengineering methodology. With GFP, proteins can be expressed with a GFP peptide connected to their sequence via a short linker [78]. For linkage of smaller fluorophores, a biomolecule of interest can be generated with some reactive chemical groups such as azides and alkynes, which can be covalently modified to attach a fluorescent tag via a post-translational covalent reaction such as the copper

catalyzed azide-alkyne cycloaddition reaction (a "click chemistry" reaction) proposed by Sharpless et al. [79].

Our group has previously worked extensively with this reaction system, and had success labeling protein in both *in vitro* and *in vivo* systems [80–82]

3.2 Methods

3.2.1 Spatial Stochastic Simulations of FRCS Data

Spatial stochastic simulations of Brownian motion were performed on a supercluster, each simulation was run on a single compute node with dual 12-core Intel Xeon Gold "Sky Lake" CPUs @ 2.60 GHz and 96 GB of memory. MCell 3.5 software, was used to generate positional data. All data were generated using a time step significantly smaller the timescale of diffusion and reaction study (Δt = 1e−6 s). Unless otherwise specified, all simulations were run within a fixed volume sphere of 1 µm radius about the origin, with a smaller 250 nm radius spherical CV. Similarly, unless noted all used identical initial conditions: k_a = 5e6 $M^{-1}s^{-1}$, k_d = 5e−3 s^{-1}, $[fA]_0$ = 10 nM, $[B]_0$ = 10 nM, $[fAB]_0$ = 0 nM, $D_{c,fA}$ = 2e−10 m^2s^{-1}, $D_{c,fAB} = D_{c,B}$ = 1e−10 m^2s^{-1}.

3.2.2 FCS Measurement of Protein Binding Interaction

CaM Expression, Fluorescent Labeling, and Purification

CaM and CaN cloning, expression, tagging and purification was performed based upon our previously published methods [80].

For CaM expression, CaM was engineered to contain a N-myristol transferase (NMT) recognition peptide (hCaNb) allowing it to be myristoylated when coexpressed with NMT. Plasmids containing NMT with kanamycin (KAN) resistance and hCaNb-CaM with ampicillin (AMP) resistance were co-transformed into chemically competent BL21(DE3) *E. coli*. Expression starter cultures were grown overnight at 37°C in

5 mL LB media supplemented with 50 μg·mL^{-1} KAN and 100 μg·mL^{-1} AMP. Primary expression cultures were started using a 1:100 dilution of the starter culture in equivalent media, and were grown at 37°C for 3.5 hr (OD600 = 0.6). Expression was induced with 0.1 mM IPTG and simultaneously supplemented with either 0.5 mM 12-azidodecanoic acid (12-ADA, a "click chemistry" enabled myristic acid analog) to generate azide-labeled CaM (N3-CaM) or with 0.5 mM myristic acid to generate an myristoylated wild type (Myr-CaM). The secondary expression cultures were grown 4 hr at 37°C, then were centrifuged to harvest cells (12,500 x g, 10 min, 4°C). Supernatent was removed and the pellet was resuspended in 5 mL·g^{-1} lysis buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1mM DTT, 1 mg/mL lysozyme, 0.1mM PMSF). Xells were lysed by submersion in a hot water bath at 60°C for 30 minutes then clarified by centrifugation (12,500 x g, 20 min, 4°C).

For N3-CaM fluorescent labeling in lysate, the clarified lysate (CL) was decanted into a new test tube and the semi-pure CaM was flourescently labeled via the copper catalyzed click-chemistry reaction. The CL was reacted for 1 hr at RT supplemented with the following click reagents: 20 mM sodium ascorbate, 50 mM Iodoacetamed, 30 μM alkyne-labeled Alexa Fluor 647 nm Dye (AF647), 2 mM $CuSO_4$, 10 mM tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), and 20 mM aminoguanide.

For CaM purification, CL was supplemented with 3 mM EGTA and 3 mM EDTA then loaded onto a 2 mL column of phenyl-sepharose resin (CL-4B; Sigma-Aldritch) equilibrated with lysis buffer also supplemented with 3 mM EGTA/EDTA. Column was gently agitated by rotating end-over-end at for 30 min at 4°C, then poured into a Bio-Rad gravity column and the flow through (FT1) was collected in the original tube. FT1 was then supplemented with 10 mM $CaCl_2$, and applied to a second phenyl-sepharose column equilibrated with binding buffer (50 mM Tris-HCl pH 7.5, 3 mM $CaCl_2$). Column was again gently agitated by rotating end-over-end for 30 min at 4°C. Column was poured into a clean Bio-Rad gravity column and the flow through (FT2) was discarded. Gravity column was washed once with 4x column

volumes (4xCV) of wash buffer 1 (50 mM Tris-HCl pH 7.5, 1 mM $CaCl_2$), once with 4xCV wash buffer 2 (50 mM Tris-HCl pH 7.5, 1 mM $CaCl_2$, 500 mM NaCl), and again with 4xCV wash buffer 1. CaM was eluted with 4xCV elution buffer (50 mM Tris-HCl pH 7.5, 1.5 mM EGTA) in fractions of 0.5-1 mL. Final concentration was measured via the Lowry Protein Assay, purification and labeling efficiency were evaluated by SDS-PAGE. Samples were aliquoted into samples of 100 μL and supplemented with 10% glycerol, then flash frozen and stored at -80°.

For purified N3-CaM fluorescent labeling, an aliquot of purified N3-CaM was diluted to a concentration of 10 μM and was reacted for 1 hr at RT supplemented with the following click reagents: 20 mM sodium ascorbate, 50 mM Iodoacetamed, 30 μM alkyne-labeled Alexa Fluor 647 nm Dye (AF647), 2 mM $CuSO_4$, 10 mM tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), and 20 mM aminoguanide. Excess flourophore and reaction reagents were removed by desalting with a 7 kDa MWCO desalting column. Column was equilibrated with CaM elution buffer, loaded with 120uL of click reaction mixture, and centrifuged 30 seconds at 1500 x g. Flow through was preserved and flash frozen for future use.

CaN Expression and Purification

For CaN expression, plasmids containing hCaNb-CaN and NMT were transformed into chemically competent BL21(DE3) *E. coli*. Expression starter cultures were grown overnight at 37°C in 5 mL LB media supplemented with 50 $\mu g \cdot mL^{-1}$ KAN and 100 $\mu g \cdot mL^{-1}$ AMP. Primary expression cultures were started using a 1:100 dilution of the starter culture in equivalent media, and were grown at 37°C for 4 hr (OD600 = 0.8). Expression was induced with 0.1 mM IPTG and the cultures were simultaneously supplemented with 0.5 mM myristic acid. The secondary cultures were grown 4 hr at 37°C, then were centrifuged to harvest cells (12,500 x g, 10 min, 4°C). Supernatent was removed and the pellet was resuspended in 5 $mL \cdot g^{-1}$ lysis buffer (50

mM Tris-HCl pH 7.5, 100 mM NaCl, 3 mM $CaCl_2$, 1mM DTT, 1 mg/mL lysozyme, 0.1mM PMSF).

For CaN purification, the resuspended pellet was lysed via sonification and clarified via centrifugation (12,500 x g, 30 min, 4°C). Clarified supernatent was recovered into a fresh tube. CaN was then purified using a CaM-sepharose affinity binding column. Clarified lysate was loaded onto a 1 mL column of CaM-sepharose resin (CL-4B; Sigma-Aldritch) equilibrated with lysis buffer. Column was gently agitated by turning end-over-end 1 hr at 4°C. Column was then pipetted into a Bio-Rad gravity column. Flow through was discarded and column was washed once with 4x column volumes (4xCV) of wash buffer 1 (50 mM Tris-HCl pH 7.5, 1 mM $CaCl_2$), once with 4xCV wash buffer 2 (50 mM Tris-HCl pH 7.5, 1 mM $CaCl_2$, 500 mM NaCl), and again with 4xCV wash buffer 1. CaN was eluted with 4xCV of elution buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 mM EGTA) collected in 0.5 mL fractions. Final concentration was measured via the Lowry Protein Assay and purification efficiency was evaluated by SDS-PAGE. Samples were aliquoted into samples of 200 μL and supplemented with 10%glycerol, then flash frozen and stored at -80°.

A647-CaM FCS Measurements

All FCS measurements were taken using the FCS acquisition module of the Zeiss LSM 780 confocal microscope. Experimental design was guided by previously published literature and Zeiss user documentation [56, 83].

For laser path and focus setup, measurements were performed in an 8-well NUNC chamber with a #1.5 coverglass bottom using the 40x C-Apochromat objective and Zeiss Immersol-W as an immersion medium. To focus into the sample the Zeiss was dropped to it's lowest position, set to measure the reflected light path, and brought up until both the bottom and top of the slide had been imaged as described in their documentation [83]. From the top surface of the slide, the scope was shifted

upward another 50 μm and this position was saved as the *measuring position* within the sample.

For FCS sample preparation, calibration samples were prepared by diluting free alkyne-labeled Alexa Fluor 647 (fA647) to a concentration of 40nM in CaM-CaN binding buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.2 μM EGTA, 1 μM $MgCl_2$). Experimental samples for measurement of A647-CaM were prepared by diluting purified stocks of A647-CaM to a concentration of 80nM in CaM-CaN binding buffer. This working stock was then diluted 1:1 for measurement with binding buffer supplemented with either (i) 2 μM EGTA, (ii) 1 μM $CaCl_2$, or (iii) 1 μM $CaCl_2$ and 500 nM CaN. Measurement samples were gently agitated by flipping end-over-end for 1 hr at 4°C to ensure dynamic binding equilibrium was reached, and gently mixed with slow pipeting prior to measurement.

For FCS calibration measurements, a 20 μL aliquot of calibration sample was added to an empty well with the microscope in loading position. The microscope was then moved to measuring position and a 633 nm Helium Neon laser was turned on. Confocal aperture was set to 1 AU and the Zeiss "auto-adjust" function was used to optimize the pinhole geometry. To ensure appropriate conditions for experimental measurement the photon count rate, correlation, and count-per-molecule were checked to verify expected ranges (1-500 kHz, 1.0, and 1-15 kHz respectively). Laser was then turned off to allow calibration sample to recover from fluorescent bleaching. A calibration measurement was recorded by taking 20 consecutive, 30 second measurements of the calibration sample, and data were exported for analysis.

For FCS experimental measurements, sample measurements were recorded in duplicate by gently mixing the sample mixture by pipeting up and down, then extracting 20 μL aliquots of the sample and placing them in an empty well. The microscope was focused into *measuring position* and measurements were recorded by taking 20 consecutive, 30 second measurements for each replicate, and data were exported for analysis.

For FCS data analysis, hardware correlation traces exported from the Zeiss LSM 780 software were imported into Pycorrfit, a free python based 3rd party FCS analysis software developed by Weidmen et. al [84]. All traces were fit using with a 1x 3D-diffusing species model and triplet correction (3D+T). CV geometry was calibrated based upon the measured residency time and average particle count along with the known aqueous diffusion (3.3E-10 $m^2 \cdot s^{-1}$) of the fA647 fluorophore [85]. Diffusion for the two species of A647-CaM (bound/unbound to CaN) were determined from this CV and the appropriate parameter in the fit model for each system.

3.3 Results

3.3.1 Spatial Stochastic Simulations of Binding Reactions

To validate the feasibility of FRCS, simulations of 3D-interacting proteins under the influence of Brownian motion were performed using MCell 3.5 spatial stochastic simulation software [86–88]. In these simulations each molecule exists as a unique point particle in the simulatory system. These particles are allowed to freely diffuse via a monte-carlo dictated random walk that projects particles along straight line trajectories based off the diffusion coefficient and time step of integration. The position of each molecule is tracked across the entire time course of study. Additionally, MCell utilizes ray tracing to detect if any of these particles will collide during any given time step, and applies collision theory to determine a reaction probability for each collision event from input reaction rates for possible bimolecular reactions [86–89]. For the purposes of this work, these simulations were limited to a very small volume, a sphere of radius 1 μ only slightly larger than the simulated CV to avoid any border influence within the region of interest, and with only three species present in the system. Spatial stochastic simulations of Brownian motion were used to generate simulated photon count data for a reaction of two binding species, one fluorescent species (fA) and one non-fluorescent (B), that form a fluorescent complex (fAB).

$$fA + B \rightleftharpoons fAB \tag{2.20}$$

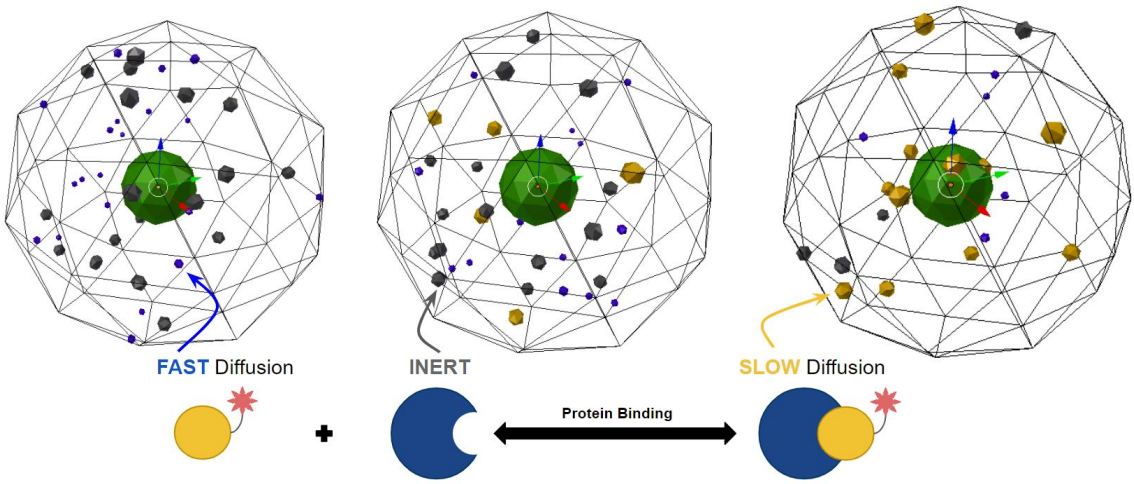


Fig. 3.1. A spatial stochastic simulation of an FCS experiment. Shown are three progressive time points from a simulated binding reaction of a small fluorescent ligand and large inert receptor as in equation 2.20. Small icospheres (blue, grey, yellow) display particle positions, the large green icosphere shows a transparent border defining the CV, and the large wire-mesh show the diffusion boundaries of the simulated solution droplet.

From the simulated positional data of this binding interaction, photon arrival times, as recorded in traditional FCS, were generated. Positional data of particles within a simulated spherical ($\alpha = 1$) CV with radius $\omega_{xy} = \omega_z = 250$ nm were exported for all time steps and a custom python script was used to simulate photon emissions for each simulation time step t, according to a Gaussian excitation probability profile function developed by Dix et al. [90].

$$PSF[t,i] = \frac{-1}{2\omega_{xy}^2}\left((X_i - X\_1)^2 + (Y_i - Y_0)^2 + \frac{(Z_i - Z_0)^2}{\alpha^2}\right) \tag{3.1a}$$

$$P_{ex}[t,i] = P_{abs}e^{K_{pos}[t,i]/\omega_{xy}^2} \tag{3.1b}$$

$$P_{em}[t,i] = P_{ex}[t,i]P_{gs}P_{conv}Q_fQ_{det} \tag{3.1c}$$

$$I[t] = \sum_i H[P_{em}[t,i] - rand(0,1)] \tag{3.1d}$$

Here i denotes the index of each molecule within the defined CV for a given simulation time step t. For each molecule a photon emission probability P_{em} is calculated and compared with a random value from 0-1 for each molecule using the Heaviside step function ($H = 1$ if passed a value greater than or equal to 0). The sum of 'emitted photons' is then tallied for each time point to generate an photon count intensity function, $I[t]$. The emission probability is informed by a PSF similar to eq. (2.2) using positional data for each molecule (X_i, Y_i, Z_i) within the CV with center (X_0, Y_0, Z_0). This location determines the probability of excitation, P_{ex}, given the relevant laser intensity and some photon absorption probability P_{abs} that is related to the molecule orientation, but can be assumed to be an orientation averaged constant over many observed time points. Furthermore, the probability that each particle is in the ground state, P_{gs}, and capable of undergoing photo-conversion, P_{conv}, are assumed to be constant ($P_{gs} = P_{conv} = 1$) given a simulation time step Δt significantly greater than the characteristic relaxation time of a fluorophore and negligible inter-system crossing (such as into a triplet state). Lastly, the quantum yield of fluorescent photons Q_f and quantum yield of photon detection Q_{det} were assumed independent of photon polarization and therefore also represented as constants. The P_{abs}, Q_f, Q_{det} were typically adjusted as a lump constant to generate photon count rates within the range of 5-100 kHz.

This simulated data was then evaluated with FRCS to validate the technique's feasibility and potential range of sensitivity for experimental parameters. Since our simulated data doesn't incorporate system crossing or photo-bleaching, those param-

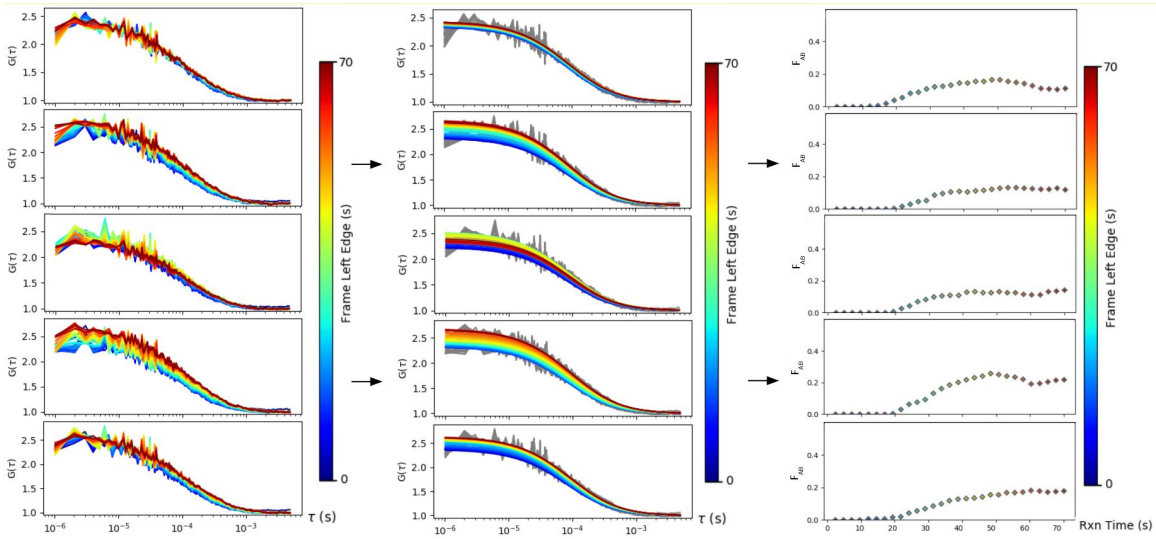


Fig. 3.2. FRCS Evaluation of a simulated binding interaction between two proteins. Shown are data for all seeds from the same simulation (and plotted on the same axes) as displayed in figure 2.11. Free ligand introduced to system at t = 0, receptor introduced at time t = 20 s. Left: FRCS generated correlation frames with frame width of 20 seconds. Middle: Fit correlation traces for each frame using fitting model described in equation 2.27. Right: Resulting fractional composition of complex as a function of reaction time ([fAB](t_{rxn}). Simulated binding data generated for a binding interaction with k_a = 2e6 $M^{-1}{\cdot}s^{-1}$, k_d = 5e−2 s^{-1}, $[A]_0 = [B]_0$ = 10 nM, $D_{c,A}$ = 1.5e−10 $m^2{\cdot}s^{-1}$, and $D_{c,B} = D_{c,AB}$ = 0.75e−10 $m^2{\cdot}s^{-1}$

eters were appropriately railed for fitting. Figure 3.2 displays an example of data generated for a binding simulation with multiple seeds (n = 5) using characteristically rapid kinetic rates as enabled by FRCS.

Frame Time Calibration

One important answer that spatial stochastic simulations can provide is a calibration for the time point of a reaction that a windowed average best represents. It would be naive to assume that a windowed average of diffusing species composition over the course of a 10-30 second measurement frame, such as are generated by FRCS, indicates the true behavior at the beginning, end or even center of that frame. Since in a simulated environment we know the true behavior of the system, this can be used to help characterize the reaction time, t_{rxn} that is best represented by the composition measurement of a frame spanning from $t_{m,f}$ to $t_{m,f} + \Delta t_f$, where $t_{m,f}$ is the experimental measurement time at the beginning of correlation frame with index f. Notably, this characteristic reaction time is likely not constant throughout the measurement. If the FRCS correlation outputs a true average of the composition of diffusing species throughout a frame, then within the linear domain of the initial binding, the composition of particles of an averaged frame likely best represents a time point near the center of the reaction ($t_{rxn,f} = t_{m,f} + 0.5\Delta t_f$). However, as the reaction progresses towards equilibrium, the relative change over the second half of the correlation frame will likely be less than the change over the first half of the frame, and thus the average value would likely represent a time point earlier in that frame $t_{rxn,f} > t_{m,f} + 0.5\Delta t_f$. Similarly, if data can be captured before the binding reaction begins, early time frames at the beginning of the reaction are likely are skewed toward $t_{rxn,f} > t_{m,f} + 0.5\Delta t_f$. The true time point best represented by the frame could therefore be described as a function of some weighting $\lambda_f \epsilon (0, 1)$ that decreases from 1 to 0 as the reaction progresses..

$$t_{rxn,f} = t_{m,f} + \lambda_f \Delta t_f \tag{3.2}$$

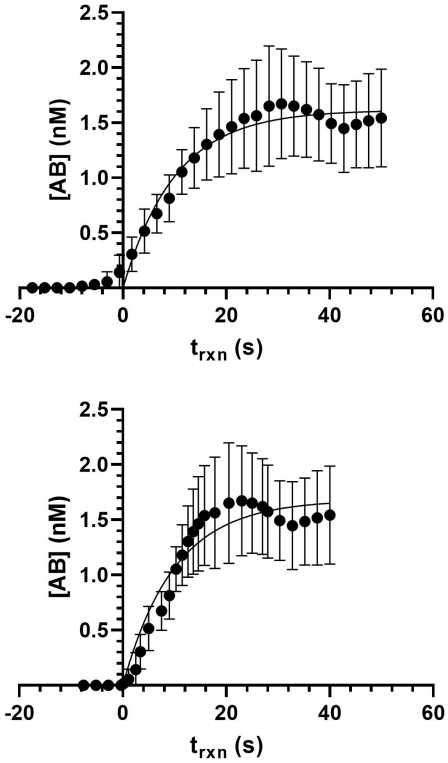


Fig. 3.3. FRCS simulation with and without an adjustment $t_{rxn,f}$. Top: [fAB](t_{rxn}) generated from the data in figure 3.2 and best fit binding kinetic curve (k_a = 1.61e6 $M^{-1}{\cdot}s^{-1}$, k_d = 6.99e−2 s^{-1}) of all seeds (n = 5) without a dynamic $t_{rxn,f}$ ($t_{rxn,f} = t_{m,f} + \Delta t_f$), Bottom: [fAB]($t_{rxn}$) and best fit curve ($k_a$ = 1.70e6 $M^{-1}{\cdot}s^{-1}$, k_d = 7.03e−2 s^{-1}) with a dynamic $t_{rxn,f} = t_{m,f} + \lambda_f \Delta t_f$. Error bars show standard deviation of determined [fAB] across seeds.

If fitting this model with non-linear regression informed by binding kinetics, this behavior can be loosely characterized from the fitting parameters. First the system of ODEs describing binding kinetics is numerically integrated to elucidate system

composition as a function of time using the fitting parameters. From this trace the predicted average composition $F_{AB,f,sim}$ within each each frame f, by numerically integrating the curve over $(t_{m,f}, t_{m,f} + \Delta t_f)$. This value can then be compared to the experimentally measured composition for that frame, $F_{AB,f,exp}$. If desired, the time point of this average composition can be extracted and used to characterize the value for λ_f so that $F_{AB,f,exp}$ can be plotted as a function of the true $t_{rxn,f}$.

Notably, even for relatively rapid kinetics such as simulation shown here with informed by an on rate k_a = 5e6 $\mathrm{M}^{-1}\cdot\mathrm{s}^{-1}$, the best fit does not change significantly between the adjusted and unadjusted model of $t_{rxn,f}$, so inclusion of this modification may not be necessary to resolve binding kinetics.

Realistic Experimental Modeling

Another potential limitation of this technique is that the kinetics being characterized are done so under the Michaelis-Menton assumptions that all species exist at relatively abundant concentrations and are well mixed within the measuring volume [18]. While this is fine for an *in silico* reaction environment where species can be instantiated in a well mixed distribution in a distinct release event at a known point in time, experimental limitations of FCS make this difficult to replicate for *in vitro* measurements.

In order to accomplish a similar goal, the FCS measurement would need to be started on a solution with a known [fA], and somehow supplemented during the course of measurement with a known [B] without reducing the [fA], and rapidly mixed without disturbing the CV. Such an experimental workflow may be possible with a creative and cutting edge microfluidic chip design, however it is much more realistic to simply mix a solution of fA and B and begin the FRCS measurement immediately after mixing. In this case, the first frame of the FRCS correlation would report the dynamics of some midpoint in the binding reaction, rather than capturing the initial binding behavior. Therefore, it is important to justify that the same

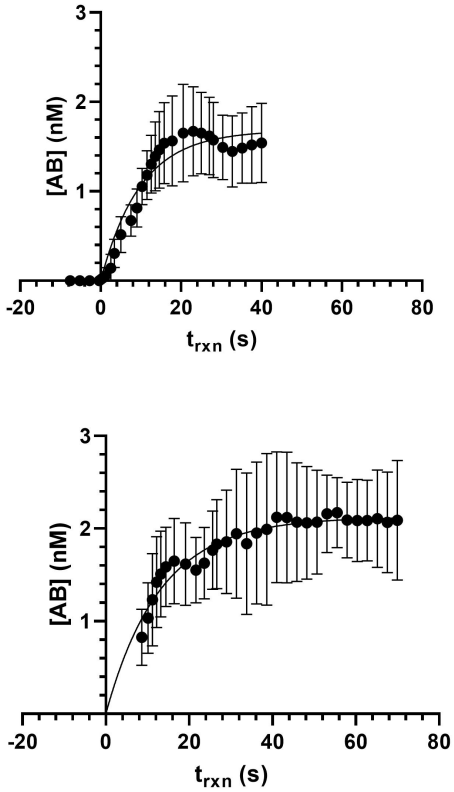


Fig. 3.4. FRCS simulation with and without experimentally limited reaction start time. Top: Determined [fAB](t_{rxn}) and best fit curve (k_a = 1.70e6 $M^{-1}{\cdot}s^{-1}$, k_d = 7.0e−2 s^{-1}) of all seeds (n = 5) with a simulated release event during the experimental time course. Right: Determined [fAB](t_{rxn}) and best fit curve of all seeds (n = 5) with an experimentally mimicked mixing prior to the experimental time course (k_a = 1.64e6 $M^{-1}{\cdot}s^{-1}$, k_d = 4.8e−2 s^{-1}). In both systems, species B was added to the system at $t_{rxn} = 0s$, and conditions were otherwise identical to the simulation data displayed in figures 3.2/3.3. Error bars show standard deviation of determined [fAB] across seeds.

binding kinetics can be observed even given this limitation. Fortunately, at such low concentrations as needed for FCS, binding reactions that occur at rates characteristic for PPIs ($k_a = 1e4 - 1e7$ $M^{-1}{\cdot}s^{-1}$) tend to occur over a time course of several seconds [91].

Indeed, even our preliminary simulations in figure 2.7 demonstrate association behavior in the range of 10-20 seconds that might be captured with a delayed mixing. Therefore, barring rapid association events, we should still be able to capture most of the protein association curve with a delayed first time point. Shown in fig 3.4, we demonstrate nearly equivalent rate kinetic determination in presence and absence of this experimental limitation.

FRCS Parameter Sensitivity Analysis

Another potential value of a spatial stochastic simulation is the opportunity to examine FRCS output under a variety of different experimental conditions. Shown in figure 3.5 are the results of four sets of simulations for which all parameters were held constant with the exception of a single parameter that was swept to explore the dynamic range of this tool. The four swept parameters were selected to elucidate predicted limitations of detection and sensitivity for this technology and a promising capacity for accurate kinetics determinations across a wide range of dynamics. Shown in figure 3.5, modified parameters included (i) the diffusivity of the complex species fAB in comparison to a fixed diffusivity of fA (ii) the initial concentration of species B with a fixed inital concentration of fA, (iii) the binding affinity, K_D, of the reaction by sweeping k_a with a fixed k_d, and (iv) the rate of reaction by holding a constant K_D and sweeping k_a and k_d proportionally. Notably only under two conditions does the fit regression trend differently than the informing experimental reaction.

The first of these is expected, as the diffusivity of the two measured species fA and fAB become closer to each other, as in the top left panel of fig 3.5, the fitting model for a two-particle system with FCS tends to break down. In this case, it seems that at similarly rapid diffusion rates differing by only a factor of 1.3 (proportional to a inverse change in hydrodynamic radius, and a $1.3^3 \approx 2.2$ factor change in protein mass) the bound species is determined (inaccurately) to be slightly more prevalent than predicted. This aligns with previously characterized findings of FCS which

suggest that diffusion can only be differentiated between particles that vary by at least a factor of 1.6 in characteristic diffusion time [26, 92]. Despite this fact, the best fit value of $k_a = 8.33\text{e}6\ \text{M}^{-1}\text{s}^{-1}$, is still well within an order of magnitude of the input value of $k_a = 5.00\text{e}6\ \text{M}^{-1}\text{s}^{-1}$ only differing by a factor of 1.67, which is considered a very tolerable error range for such measurements [18, 19]. Similarly the equilibrium indicates a $K_D = 0.588$ nM, as compared to the input value of $K_D = 1$ nM, again well within desired sensitivity limits.

The second condition under which best fit traces tend to fail to capture realistic kinetics are those with very rapid association rates k_a and relatively high concentrations. As seen in figure 3.5, when the informing reaction proceeds with an extremely rapid association event, the experimentally limited FRCS (first frame recorded after mixing) fails to capture the dynamics occurring in the first few seconds, and thus to accurately capture the initial binding rate k_a. While this is not ideal, it bears mentioning that incredibly rapid binding rates ($k_a > 1\text{e}8\ \text{M}^{-1}\cdot\text{s}^{-1}$), are highly uncommon, and at a certain point reactions become diffusion limited as defined by the Smoluchowski limit which yields binding rates around $1\text{e}9 - 1\text{e}10$ for associations relevant in PPIs [93]. Indeed, most PPIs exhibit binding rates much below this limit, in the range of $1\text{e}4 - 1\text{e}6\ \text{M}^{-1}\cdot\text{s}^{-1}$, which are notably much easier to resolve over the course of an FCS experiment at relevant concentrations. Furthermore, if we remove the experimental limitation (such as by utilizing a microfluidic chip for measurements) this behavior could still be resolved simply by recording a higher density of correlation frames in this region. It's also notable that binding affinity is still captured relatively accurately in this model even in these rapid binding instances, although the approximated K_D tends to indicate a higher binding affinity than the informing model conditions should describe. This is possibly due to an 'simulation overshoot.' At high binding rates, near the beginning of binding the simulated reaction probability is dominated by k_a, which tends to force the reaction past equilibrium, until the k_d becomes equally weighted by a higher complex concentration and drives the binding back down toward equilibrium. Therefore it seems probable that this observation is

an artefact not necessarily of the FRCS methodology, but of the informing spatial stochastic simulations.

Under all other explored conditions, FRCS seems to be very capable of predicting binding kinetics as well as binding behavior. Increasing binding substrate concentrations drives formation of complex higher, while increasing K_D results in lower levels of complex formation as expected. All determined binding constants and the informing simulation parameters can be found in the appendix table D.1.

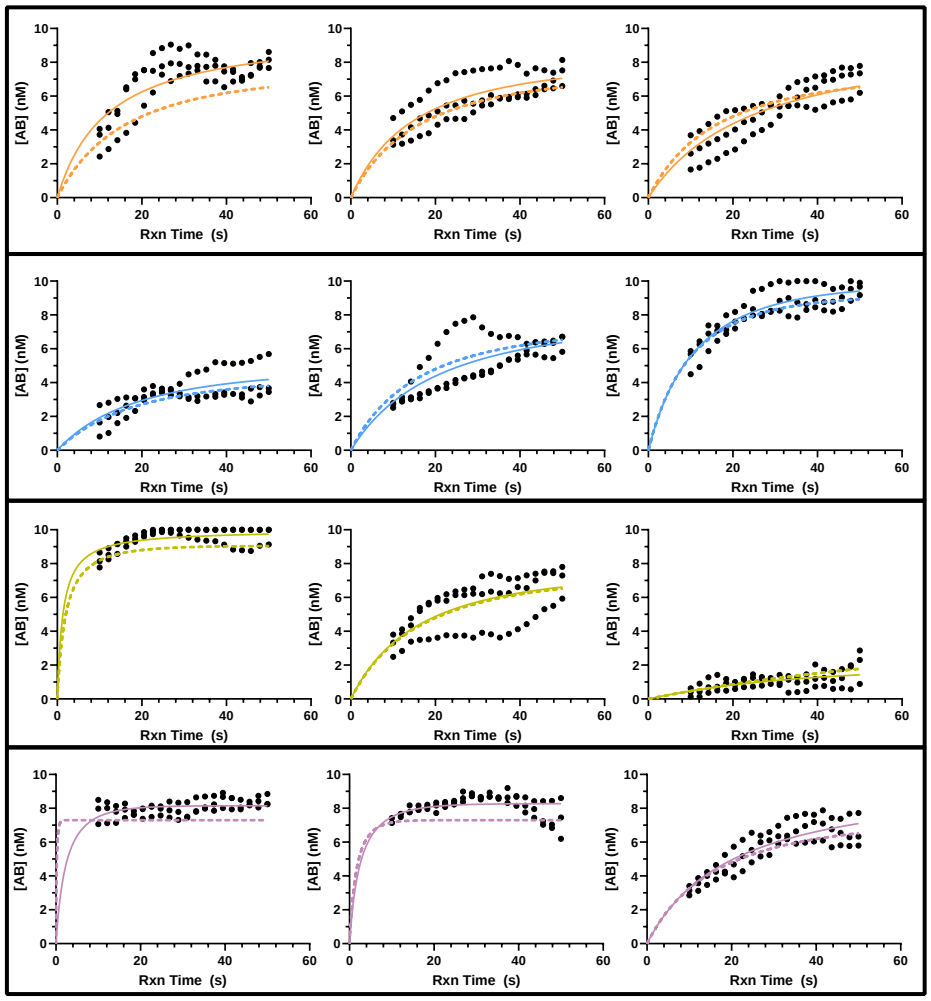


Fig. 3.5. Points indicate a single frame from a simulated measurement (n=3), solid trace indicates the non-linear regression best fit, and the dashed trace indicates the predicted reaction kinetics given input parameters. Parameters studied, top to bottom: (i) the diffusivity of the complex species fAB in comparison to a fixed diffusivity of fA (left-to-right: $D_{c,AB} = 3.0\text{e}{-10}, 2.0\text{e}{-10}, 1.0\text{e}{-10}$ m^2s^{-1}, $D_{c,A} = 4.0\text{e}{-10}$ m^2s^{-1}), (ii) the initial concentration of species B ($[B]_0 = 5, 10, 20$ nM) with a fixed initial concentration of fA ($[fA]_0 = 10$ nM) (iii) the binding affinity, K_D, of the reaction by sweeping k_a ($k_a = 5\text{e}7, 5\text{e}6, 5\text{e}5$ $\text{M}^{-1}\text{s}^{-1}$) with a fixed k_d ($k_d = 5\text{e}{-3}$ s^{-1}), and (iv) the binding rate k_a ($k_a = 5\text{e}8, 5\text{e}7, 5\text{e}6$ $\text{M}^{-1}\text{s}^{-1}$) with a fixed K_D ($k_d = 5\text{e}{-1}, 5\text{e}{-2}, 5\text{e}{-3}$ s^{-1}). Unless otherwise specified, other parameters were held constant to standard values as defined in methods.

3.3.2 FCS Studies of CaM-CaN binding

Preliminary measurements were recorded with purified N3-CaM that was ligated after purification to an Alkyne-labeled Alexa Flour 647 nm via the click reaction. Following the click ligation, click reagents and remaining free fluorophore were removed via a desalting column. These preliminary measurements (30 second measurements, n = 20 per sample) showed distinctly observable changes in the diffusion of the fluorophore when ligated with CaM (figure 3.6). Calibrated against a known diffusion coefficient for unligated Alexa Flour 647 at RT in aqueous solution (3.3e-10 $m^2 \cdot s^{-1}$ [85] the diffusion of A647-CaM was determined to be 1.45e−10 $m^2 \cdot s^{-1}$ and 1.46e−10 $m^2 \cdot s^{-1}$ in the apo-calmodulin (CaM0) and calcium saturated (CaM4) state respectively. This slight reduction in diffusion, and corresponding increase in hydrodynamic radius correlates to a previously published change in the hydrodynamic radius of CaM from 2.5 nm to 3.0 nm when saturated with calcium ions [94]. Furthermore, when 500 nM CaN was introduced to a solution containing 40 nM A647-CaM and 1 μM Ca^{2+}, the diffusion further dropped to 8.42e-11 $m^2 \cdot s^{-1}$.

In a second experiment, CaN was added at various concentrations to a fixed solution of 40nM A647-CaM and 1 μM Ca^{2+}. Seen in figure 3.7 average steady state diffusion dropped consistently with [CaN], demonstrating an average K_{Dapp} of 43.84 nM which is reasonably close to ranges of previously published literature (.028 - 24 nM [18, 95–97]).

Notably, in both experiments the determined diffusion of A647-CaM was faster than the D_c of CaM as predicted by Stokes-Einstein for particles with an R_H of 2.5-3 nm (9.7e−11 and 8.1e−11 $m^2 \cdot s^{-1}$ for CaM0 and CaM4 respectively [94]). This could be due to a smaller experimental hydrodynamic radius than previously characterized, but more likely is due to incomplete removal of free fluorophore after the click reaction by desalting. It also could be an artefact of laser-induced heating of the sample, however this has been shown to generate negligible changes in diffusion [43]. To be certain, I personally characterized the temperature change over the course of a 2

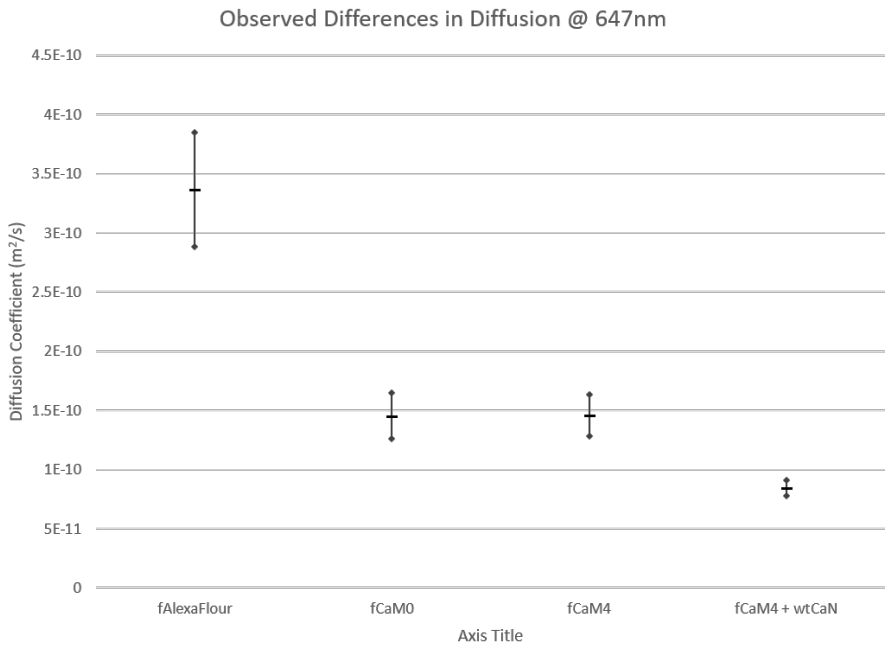


Fig. 3.6. Diffusivity of particles characterized via FCS using a single particle fitting model. Unligated alexafluor was used as a calibration and fluorescently labeled CaM was measured in the presence and absence of Ca^{2+} and CaN

minute measurement at room temperature (n=3, $T \approx 298 \rightarrow 300$ K, see appendix figure C.1), corresponding to a possible change in diffusivity by a factor of $300/298 = 1.006$. Additionally, all measurements are taken under the same conditions, so this change would be reflected in the actual calibration measurement as well.

As seen in figure 3.8, despite Coomassie labeling of highly purified CaM after the click reaction, there seems to exist a large amount of fluorescent species with a smaller mass than the CaM band itself. Without accommodating the presence of some unknown concentration of free Alexa Flour 647, estimates of A647-CaM diffusion are likely to be elevated. Similarly, the determined diffusivity of A647-CaM4 in the presence of excess CaN was higher than the Stokes-Einstein prediction for CaN alone ($R_H = 3.47$ nm, $D_c = 7.2\text{e}{-11}$ [99,100]). While this could be indicative of unsaturated binding of CaM and indeed the [CaN] never appeared to reach a high enough level

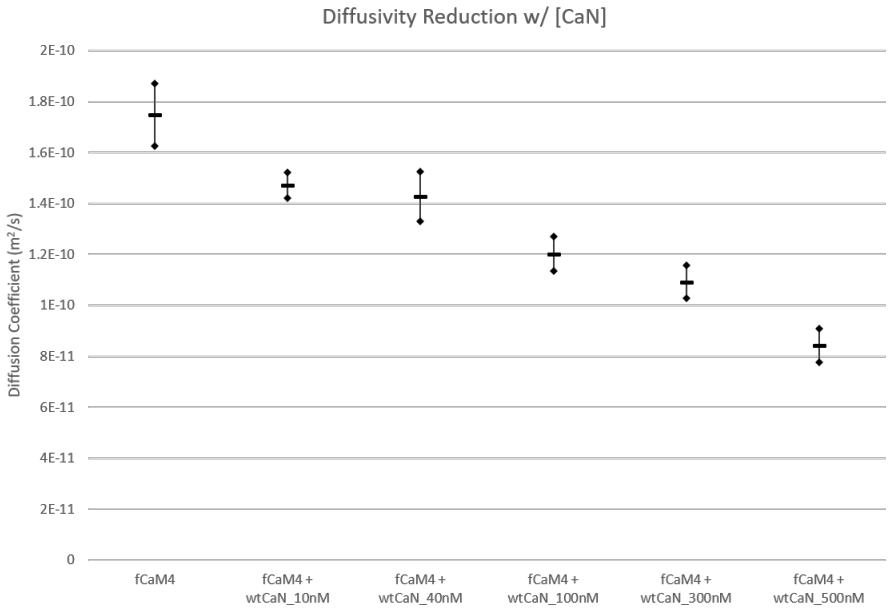
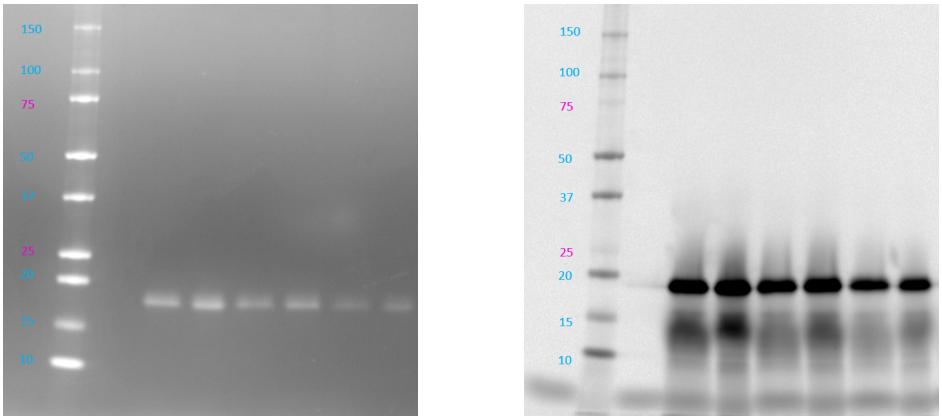


Fig. 3.7. Serial dilution of CaN associating with a fixed 40nM [CaM] measured via FCS. Notably, as the [CaN] increases, the diffusion of CaM appears to slow substantially, indicating generation of complex protein

for saturation, overestimation due to free fluorophore would similarly lead to elevated estimates of diffusion for a complex of A647-CaM and CaN, with a corresponding increase in the determined K_{Dapp}.

Future studies should seek to resolve a higher purity of labeled CaM and to minimize free fluorophore to negligible levels. One way to attain this level of purity is to ligate CaM with Alexa Fluor in lystate, then purify the CaM protein with an affinity binding column as described in the methods.



(a) Coomassie Staining (b) Fluorescent Imaging @ 647 nm

Fig. 3.8. Purified CaM ligated to Alexa Flour 647 and isolated via a 7K MWCO desalting column. Two panels show the same SDS-page gel imaged with fluorescence and coomassie staining. As evident, no significant amounts of protein contaminants are present, however some contaminant fluorescent species with lower MW than CaM (16.8 kDa [98]) become evident when imaged at 647 nm. From left to right, lanes contain a MW ladder, a blank, and 6 lanes with 20 μL dilutions of desalted A647-CaM (2 lanes each 3:1, 2:2, and 1:3 of A647-CaM desalted reaction mixture into CaM-CaN Binding Buffer).

3.3.3 Preliminary FRCS of CaM-mCaMKII

Due to time constraints and limited resources as a result of COVID-19, preliminary attempts to characterize the binding kinetics of CaM and monomeric CaMKII were unable to be resolved with a sufficient SNR to truly discuss or quantify binding dynamics. Results of these preliminary studies can however be seen in the appendices (figure C.2). Future work will be invested in refining the experiment further.

3.4 Conclusions

In this chapter I present an *in silico* model for the generation of FCS data using simulated brownian motion and bimolecular interactions. I use this model to char-

acterize the dynamic range of the FRCS technique and algorithm, and demonstrate it's viability for characterization of PPIs. I further demonstrate preliminary experimental evidence of PPIs measured by FCS technology, and pave the way for future exploration of a broad set of PPI dynamics resolved about CaM, a significant actor in cellular modulation and signalling.

4. KINETICS OF AZIDOHOMOALANINE BIODISTRIBUTION AND IT'S METABOLIC IMPLICATIONS *IN VIVO*

The content of this chapter will be submitted for publication. Dr. Aya Saleh contributed to experimental studies and collection of data that informed the computational model development, parameterization, and validation that I performed and presented in this chapter.

4.1 Introduction

Inspired by the works of Dieterich et al., our group has previously demonstrated the ability to selectively label nascent polypeptides within the murine proteome via a systemic injection of a non-canonical amino acid (ncAA) [81, 101, 102]. In this technique, a methionine (Met) analog, azidohomolaline (Aha) is introduced into the biological system of interest via subcutaneous injection, and immediately begins to become incorporated into newly synthesized polypeptide (NSP) chains using the cell's endogenous translational machinery. This selective labeling of nascent peptides allows for experimentally controlled distinction between the constituent proteome and newly synthesized nascent protein, following Aha injection. To enable this distinction, Aha is enabled with a reactive azide group that can be covalently modified via the azide-alkyne cycloaddition reaction (a click chemistry reaction) [81]. As such, ncAA-labeled NSPs can be selectively conjugated to "clickable" affinity or fluorescent tags for identification or visualization, respectively [41, 81, 102]. Compared to previous studies of ncAA labeling in complex organisms, the Aha injection method allows stable dosing with predictable patterns of Aha distribution [101, 103–105].

Despite the application of ncAA labeling in a number of studies to decipher complex cellular processes in animal models [103–106], understanding the kinetics of the distribution of ncAAs into tissues, especially as it pertains to the rates of protein

incorporation and loss by protein degradation, remains lacking. Our own works in this field have been limited in scope to studies of proteome changes within 24 hours of injection, and without a robust characterization of the dynamics underlying observed changes in NSPs. Determination of the timescale of the ncAA uptake by tissues following its administration and the lag time before maximum protein labeling are critical information to guide the design of robust temporal experiments to study the nascent proteome. Such understanding will also enable optimization of the dosing regimen to attain ideal Aha concentrations that achieves the required degree of protein labeling over the course of the study, depending on the average rate of protein synthesis and turn over in tissues of interest.

In the work presented in this chapter, we aimed to characterize the distribution kinetics of Aha within murine physiology. To study the biodistribution of Aha, we develop here a compartment based model of small molecule biokinetics that characterizes the movement of freely diffusive Aha (fAha) throughout the mouse circulatory system and into distinct tissues including the liver, kidney, brain and skeletal muscles. We further develop a model of incorporation of fAha into proteins within tissues of interest. This incorporation model was used to characterize both the predicted labeling profile for a given experimental treatment of Aha and the relative synthesis and turnover rates of Aha-incorporated proteins. This objective was accomplished with a custom biokinetic compartment model system of ordinary differential equations (ODEs) based, in part, upon a model developed by Kirman et al. describing small molecule transport in mice [107]. This model was parameterized based upon transfer and surface exchange rates experimentally determined through a variety of different techniques [108–110], and then adjusted for Aha by fitting to experimental Aha data. The distribution of fAha in tissues was quantified via liquid chromatography-tandem mass spectrometry (LC-MS/MS) for each relevant tissue over a period of 24 hr following an initial Aha injection. In addition, we used fluorescent western blotting to measure semi-quantitative relative protein labeling in these tissues during the

same period of time. The biokinetic model was then fit within the roughly defined parameter bounds to best represent these experimental data.

We demonstrate here that this fully parameterized biokinetic model can be used to characterize nascent protein synthesis and turnover within distinct tissues. Furthermore, we validate it's capabilities as a predictor of Aha distribution and therefore protein labeling under more complex, multiple injection dosing paradigms. We present this model as a tool to guide future experimental design utilizing Aha labeling, which we have shown to be capable of elucidating metabolome dynamics with previously unattainable selectivity and temporal resolution.

4.2 Methods

4.2.1 Animal Model

Animals used in these studies were derived from female age-matched wild-type C57Bl6 mice (Mus musculus) purchased from The Jackson Laboratory. All experimental protocols were performed in compliance with established guidelines and all methods were approved by Purdue Animal Care and Use Committee (PACUC, protocols 1209000723 and 1801001682). PACUC requires that all animal programs, procedures, and facilities at Purdue University to abide by the policies, recommendations, guidelines, and regulations of the USDA and the United States Public Health Service in accordance with the Animal Welfare Act and Purdue's Animal Welfare Assurance.

4.2.2 Aha Injection, and Plasma and Tissue Collection

The methionine (Met) analog L-azidohomoalanine (Aha; Click Chemistry Tools) was resuspended in 1 x phosphate buffered saline (PBS) to a 10 $mg{\cdot}mL^{-1}$ concentration, pH adjusted to 7.4 with NaOH, sterile filtered and stored at -20°C. All injections were administered to mice subcutaneously at 0.1 $mg{\cdot}g^{-1}$ bodyweight. Mice (n = 3 biological replicates) were euthanized 0.5, 1, 2, 4, 6, 12, 24 hr post injection (hpi).

Blood was collected by cardiac puncture into EDTA-treated tubes and centrifuged at 1,500 x g for 10 min at 4°C. The supernatant (plasma) was transferred into a new tube using a Pasteur pipette, snap frozen in liquid nitrogen and stored at -80°C. Liver, brain, kidney and hindlimb skeletal muscle tissues were dissected at each time point, snap frozen in liquid nitrogen and stored at -80°C. Plasma and tissues were collected as described above from non-injected mice (n = 3 biological replicates) to be used as a control.

4.2.3 Sample Preparation for Aha Analysis

For plasma sample preparation, 50 μL of plasma were mixed with 10 μL of 1 x PBS, pH 7.4 and 5 μL of 100 ng·μL^{-1} L-α-aminobutyric acid (μL of trichloroacetic acid (TCA; Sigma) were added to the mixture to precipitate proteins. The mixture was incubated for 10 min at 4°C and centrifuged for 10 min at 16,000 x g at RT. The supernatant was then mixed with 100% acetonitrile (ACN; Fisher Scientific) at a 1:1 ratio (v/v). The mixture was transferred to an autosampler vial for LC-MS analysis. For calibration curve generation, Aha standards were prepared by mixing 50 μl of non-injected plasma with 10 μl of a known concentration of Aha and 5 μof α-ABA. Proteins were then precipitated with TCA and prepared for LC-MS analysis as described above.

For tissue sample preparation, tissues were rinsed with ice-cold 1 x PBS, pH 7.4 to remove residual blood and homogenized in ice-cold 1 x PBS, pH 7.4 using a TissueRuptor (Qiagen). The final homogenate weight was measured and converted to volume by using a homogenate density of 1 g·mL^{-1}. Samples were then prepared for LC-MS analysis as described for plasma by using 50 μl of the tissue homogenate. The remaining plasma samples and tissue homogenates were snap frozen and stored at -80°C until use for western blot and non-targeted metabolomic analyses as described below.

4.2.4 LC-MS Targeted Analysis of Aha

An Agilent 1260 Rapid Resolution liquid chromatography (LC) system coupled to an Agilent 6470 series QQQ mass spectrometer was used for Aha analysis (Agilent Technologies). An Intrada Amino Acid 2.0 mm x 150 mm, 3.0 μm column (Imtakt Corporatio) was used for LC separation. The buffers were (A) ACN, 0.3 % formic acid (FA; Sigma) and (B) ACN/100 mM ammonium formate (20/80 v/v). The linear LC gradient was as follows: time 0 minutes, 20 % B; time 5 minutes, 20 % B; time 11 minutes, 35 % B; time 20 minutes, 100 % B; time 22 minutes, 100 % B; time 22.5 minutes, 20 % B; time 30 minutes, 20 % B. The flow rate was 0.3 mL·min^{-}1. Multiple reaction monitoring was used for MS analysis. Data were acquired in a positive electrospray ionization (ESI) mode according to Table 1. The jet stream ESI interface had a gas temperature of 325°C, gas flow rate of 9 L·min^{-}1, nebulizer pressure of 35 psi, sheath gas temperature of 250°C, sheath gas flow rate of 7 L·min^{-}1, capillary voltage of 3500 V in a positive mode, and nozzle voltage of 1000 V. The delta electron multiplier voltage was 300 V. Agilent MassHunter Quantitative Analysis software was used for data analysis (version 8.0).

Table 4.1.
Multiple reaction monitoring table for data acquisition.

Compound name	Precursor ion	Product ion	Collision Energy
Aha	145.1	101.3	5
Aha	145.1	71.3	10
Aha	145.1	58.3	40
Ala	90	44	15
Arg	175	116	18
Asn	133	87	12
Asp	134	88	14
Cys	122	76	15
Cys-Cys	241.1	152	15
Gln	147	84	22
Glu	148	130	12
Gly	76	30	15
His	156	110	19
Ile	132	86	15
Leu	132	86	15
Lys	147	84	20
Met	150	104	15
Phe	166	120	15
Pro	116	70	15

4.2.5 Western Blot Analysis of Aha-labeled Tissues

Tissue homogenates were thawed and protein concentration was measured using the Pierce 660 nm Protein Assay (ThermoFisher Scientific). 200 µg of tissue homogenate were alkylated with 40 mM iodoacetamed for 30 min at RT in the dark

with end-over-end rotation. Samples were then reacted for 2 hr at RT with the following click reagents: 50 μM biotin-alkyne (ThermoFisher Scientific), 5 mM tris(3-hydroxypropyltriazolylmethyl)amine (THPTA; Click Chemistry Tools), 2 mM copper sulfate, 20 mM aminoguanidine and 10 mM sodium ascorbate. Following click reaction, proteins were precipitated by adding ice-cold 100% acetone to the samples at a 4:1 ratio (v/v). Samples were incubated overnight at -20°C, centrifuged at 21,100 x g for 20 min at 4°C, supernatant was discarded and protein pellets were vacuum-dried for 15 min at RT using CentriVap (Labconco). Dried pellets were resuspended in (8 M urea in 1 x PBS) and centrifuged at 16,000 x g for 15 min at RT to remove insoluble particles. The supernatants were transferred into new tubes and protein concentration was measured using the Pierce 660 nm Protein Assay (ThermoFisher Scientific). Proteins were resolved on 4 – 20% SDS-PAGE gels (BioRad), transferred to a PVDF membrane (ThermoFisher Scientific) using the Trans-Blot Turbo Transfer System (BioRad) and probed overnight at 4°C with IRDye 680 Streptavidin (LICOR) (1:3000 dilution). Membranes were imaged using an Azure Biosystems c600. Western blot images were analyzed using ImageJ (NIH) to calculate the mean fluorescence intensities of each time point. The intensity of the control sample was used to normalize the intensity of each time point (n = 3 biological replicates/blot). Sample Preparation for Non-targeted Metabolomic Analysis Plasma, and liver and brain homogenates were thawed. The metabolome of non-injected control samples (n = 3 biological replicates) and samples collected 24 hr post Aha injection (n = 3 biological replicates) was extracted by adding methanol: chloroform: water (1:1:1, v/v) to 80 μl of each plasma sample and to 60 μl of each liver and brain sample. Samples were vortexed briefly and centrifuged at 8,000 x g for 5 min at RT. The upper layer was transferred into a new tube and vacuum-dried overnight at RT.

4.2.6 Plasma Sample Preparation for Untargeted Metabolomic Analysis

The plasma metabolome of non-injected control samples (n = 3 biological replicates) and samples collected 24 hr post Aha injection (n = 3 biological replicates) was extracted by adding methanol: chloroform: water (1:1:1 v/v) to 80 μl of each plasma sample. Samples were vortexed briefly and centrifuged at 8,000 x g for 5 min at RT. The upper layer was transferred into a new tube and vacuum-dried overnight at RT. The dried fraction was reconstituted in 75 μl of a diluent composed of 5% ACN and 0.1% FA. Reconstituted samples were sonicated for 5 minutes, centrifuged at 16,000 x g for 8 min at RT, and the supernatants were transferred to HPLC autosampler vials.

4.2.7 LC-MS Untargeted Metabolomic Analysis

Separations were performed on an Agilent 1290 UPLC system (Agilent Technologies). The metabolites were analyzed using a Waters Acquity HSS T3 column (1.8 μm, 2.1 x 100 mm), with a mobile phase flow rate of 0.45 ml·min^{-1}, where the mobile phase A and B were 0.1 FA in double distilled water (ddH2O) and ACN, respectively. Initial conditions were 100:0 A:B, held for 1 min, followed by a linear gradient to 20:80 at 16 min, then 5:95 at 22.5 min. Column re-equilibration was performed by returning to 100:0 A:B at 23.5 minutes and holding until 28.5 minutes.

The mass analysis was obtained using an Agilent 6545 Quadruple Time of Flight (Q-TOF) MS with ESI capillary voltage +3.2 kV, nitrogen gas temperature 325°C, drying gas flow rate 8.0 L·min^{-1}, nebulizer gas pressure 30 psi, fragmentor voltage 130 V, skimmer 45 V, and OCT RF 750 V. MS data scans (m/z 70-1000) were collected using Agilent MassHunter Acquisition software (v.B.06). Mass accuracy was improved by infusing Agilent Reference Mass Correction Solution (G1969-85001). MS/MS was performed in a data-dependent acquisition mode on composite samples.

4.2.8 Metabolomics Data Analysis

Peak deconvolution and integration was performed using Agilent ProFinder (v.10.0). Bioinformatic analyses were performed using Agilent's Mass Profile Professional (v.13.1). Chromatographic peaks were aligned across all samples. Peak areas were normalized by log2-transformation and applying a 75 percentile shift. Metabolites were filtered out if present in only one sample. Only metabolites present in all 3 replicates of at least one group were included in the analysis. Statistical analysis was performed using unpaired student's t-test. Metabolites with $P < 0.05$ and fold change > 2 were considered significant. Peak annotations were performed using the METLIN metabolite database [111], with a mass error of less than 5 ppm. Identifications were aided by MS/MS spectra comparisons. Principal component analysis (PCA), hierarchical clustering analysis (HCA) and metabolic pathway analysis were performed using MetaboAnalyst v.4.0 [112].

4.2.9 Kinetic Modelling of Aha distribution

Simulations were run on a Lenovo Yoga with an Intel Core i7-8550U CPU @ 1.8 GHz and 8 GB RAM. Simulations were preformed using custom modeling scripts written in Python 3.6 (See appendix A). System of ordinary differential equations (See appendix B) were solved using a flexible high order solver (scipy.integrate, [69]) and parameters were fit with a least squares minimization algorithm from 'Lmfit' a prebuilt python library [66]. All best fit values can be found in appendix D tables D.2-D.4.

4.2.10 Model Validation and Sensitivity Analysis

Standard error of fitted parameters were determined for the 19 fitted parameters in the biodistribution model and for all 8 fitted parameters in the protein

incorporation model. Standard error values were determined during the non-linear regression using the built in functionalities of the 'Lmfit' python library [66].

Partial rank correlation coefficients (PRCCs) were determined for all 31 total parameters in the biodistribution model and for all 8 parameters in the protein incorporation model. PRCCs were used to character the influence of each parameter on the sum of square errors (SSE), the optimization metric for non-linear regression. To effectively sample the parameter space, latin hypercube sampling (LHS) was utilized to select unique parameter sets (n = 10000) as detailed by our previous works [18,113]. Simulations used to inform PRCCs were performed on a supercluster, and each was run on a single compute node with dual 12-core Intel Xeon Gold "Sky Lake" CPUs @ 2.60 GHz and 96 GB of memory.

4.3 Results

4.3.1 Kinetic Model of Aha Distribution

Freely diffusive, unbound Aha (fAha) was introduced into the model at the injection site, as in our subcutaneous dosing paradigm. From the injection site it was able to enter and circulate the rodent's blood stream, wherein it underwent surface exchange with distinct tissue compartments. As such the fAha distribution model is easily broken into two stages: transport and exchange (fig 4.1).

Within each compartment, the time rate of change of the fAha plasma concentration ($[fAha_p]$) available for surface exchange with the tissue can be described as a mass balance.

$$\left(\frac{d[fAha_p]_x}{dt}\right)_{transport} = \frac{Q_x}{V_x}\left([fAha_p]_{sysrv} - [fAha_p]_x\right) \tag{4.1}$$

Where Q_x is the blood flow rate between tissue x and a systemic venous reservoir, and V_x is the corresponding volume of plasma relevant to each tissue (Q/V represented as a lumped constant qb in appendix tables D.2-D.4 for parameter details).

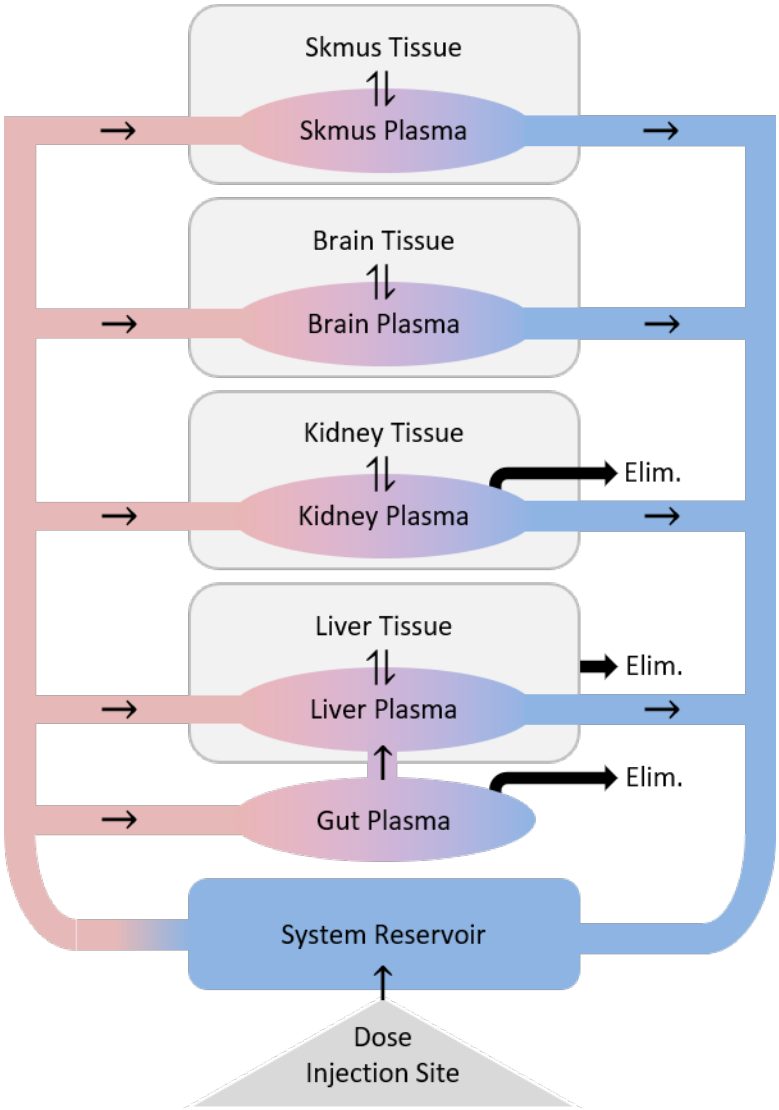


Fig. 4.1. Biodistribution of fAha via transport and exchange. Introduced at a distinct injection site, fAha is allowed to enter the venous circulation (blue). From the venous circulation fAha is driven through the arterial system (red) into distinct tissue compartments where exchange occurs at tissue specific rates. Arrows indicate directional movement of fAha, ellipses indicate distinct compartments of plasma available for surface exchange, and rectangles indicate cumulative intracellular volume for each tissue in the model. Three additional terms exist for elimination of fAha via excretion and metabolization.

The liver tissue, gut and renal plasma compartments all had additional elimination terms accounting for excretion and metabolization of fAha. All kinetic parameters for circulatory transport were normalized by tissue mass to compare relative perfusion rates between tissue compartments of differing size. Once localized to a tissue, fAha in the plasma ($fAha_p$) may also be exchanged across the cell membrane with the trapped intracellular species ($fAha_t$).

$$\left(\frac{d[fAha_p]_x}{dt}\right)_{exchange} = k_{e,x}[fAha_t]_x - k_{i,x}[fAha_p]_x \tag{4.2}$$

$$\left(\frac{d[fAha_t]_x}{dt}\right)_{exchange} = k_{i,x}[fAha_p]_x - k_{e,x}[fAha_t]_x \tag{4.3}$$

Where $k_{i,x}$ and $k_{e,x}$ are the tissue specific import and export rates for fAha across the cell membrane. This trapped species is the concentration experienced by synthesis machinery within the cell. The two stages of distribution were then combined into a single system of ODEs, which was parameterized and bound within reasonable ranges for a model of small molecule pharmacokinetics (appendix tables D.2-D.4) [107–110]. Parameters were then fit to best match time-resolved fAha data obtained from LC-MS.

4.3.2 Kinetic Model of Protein Labeling

Within each tissue compartment $fAha_t$ is expected to become incorporated into a proteinous form (pAha) via protein synthesis, resulting in an experimentally observable fluorescent signal from Aha-labeled protein. As a Met analog, Aha is able to bind methionyl tRNA synthase, albeit at a much slower rate (kcat/Km Aha: 1.42e−3, Met: 5.47e−1 $s^{-1}\cdot\mu M^{-1}$) [114]. Though to our knowledge there is no reliable experimental measurements reported for the amount of Met incorporated into protein in murine tissues (pMet), Met is known to be present at low levels relative to other canonical amino acids. Free Met (fMet) is typically present at concentrations comparable in

magnitude to the peak concentrations of Aha in measured tissues ([fMet] $\approx$ 8, 12 $\mu g \cdot g^{-1}$ in liver and kidney respectively []). Given the difference in the rate constant of association with methionyl tRNA synthase, at similar concentrations, fMet likely dominates translation over fAha. It would be reasonable to assume that concentrations of pAha are therefore lower than pMet by a similar factor of several orders of magnitude or that (i) [pAha] << [pMet] within each tissue. To approximate the [pMet], we can examine total cellular protein synthesis (15 $mg \cdot g^{-1} \cdot day^{-1}$, [116]), and average Met content of the proteome (3-5%, [117]), which yields a maximum pMet synthesis rate of around 750 $\mu g \cdot g^{-1} \cdot day^{-1}$. Since fAha is only in excess for around 4 hours in our distribution model, a reasonable upper bound of [pMet] generated in that time is around 125 $\mu g \cdot g^{-1}$. This is reasonably close to the average [fMet] of our system. Therefore it's reasonable to assume in this time frame that (ii) the generated [pMet]$\approx$[fMet]. With these two assumptions (i/ii), it is reasonable to conclude that [pAha] << [fAha]. Therefore, [pAha] can be assumed dependent upon [fAha], while [fAha] can be assumed relatively independent of depletion via protein synthesis at least in comparison to biotransport on this timescale. To further validate this assumption, we note that despite vast differences in the degree of protein labeling between tissues (for instance liver and skeletal muscles, fig 4.2E-H), there is no discernible difference in their fAha concentration profiles 4.2A-D, supporting the assumption of negligible depletion of fAha by protein synthesis.

$$[fAha_t]_x \gg [pAha_t]_x \tag{4.4}$$

$$\left(\frac{d[fAha_t]_x}{dt}\right)_{synthesis} = -k_{s,x}[fAha_t]_x \approx 0 \tag{4.5}$$

$$\left(\frac{d[pAha_t]_x}{dt}\right)_{synthesis} = k_{s,x}[fAha_t]_x - k_{d,x}[pAha_t]_x \tag{4.6}$$

Where k_s and k_d are the tissue-specific kinetic rates of protein synthesis and degradation with respect to the Aha. Furthermore, if the degree of fluorescent signal relative to the background (rF) from pAha labeled protein is linearly proportional to

the concentration of pAha by some factor, kf, then the generation of rF signal can be predicted from the synthesis kinetics.

$$rF_x = \frac{(signal - background)}{(background)} = k_f[pAha_t]_x \tag{4.7}$$

$$\frac{d(rF_x)}{dt} = k_f \frac{d[pAha_t]_x}{dt} = k_f k_s [fAha_t]_x - k_d(rF_x) \tag{4.8}$$

A distinct third stage was added to the model to predict rF signal from the fAha distribution. This model was fit to fluorescence data, acquired for each tissue via western blot, to (i) estimate relative synthesis and degradation rates of protein in each tissue and (ii) establish a time resolved predictive model of protein labeling given a variety of input dosing paradigms.

4.3.3 Estimation of Relative Protein Synthesis and Turnover Rates

The model was constructed and fit to LC-MS quantifications of [fAha] in all studied tissues over a period of 24 hrs following a single dose subcutaneous injection of fAha (0.1 mg·g^{-1} body weight). Distribution of Aha through the systemic circulation was rapid, with concentration peaking within 1 hr of injection, but the lifetime of Aha under this dosing paradigm was similarly brief (fig. 4.2, panel A-C). Incorporation into protein increased the longevity of Aha, allowing labeled protein to persist even after the fAha concentration was fully depleted (fig. 4.2, panel D-F). Note that despite vast differences in generation of rF, or by proxy the synthesis of florescent protein between tissues (for instance the liver and skeletal muscle), there is no discernible difference in the fAha distribution profiles supporting the assumption of negligible depletion by protein synthesis.

From the model, the half-life of Aha in each tissue was determined (table 4.2). Similarly, the kinetics of protein turnover were also investigated to compare protein synthesis rates and half lives amongst tissues. Notably, protein synthesis rates differed

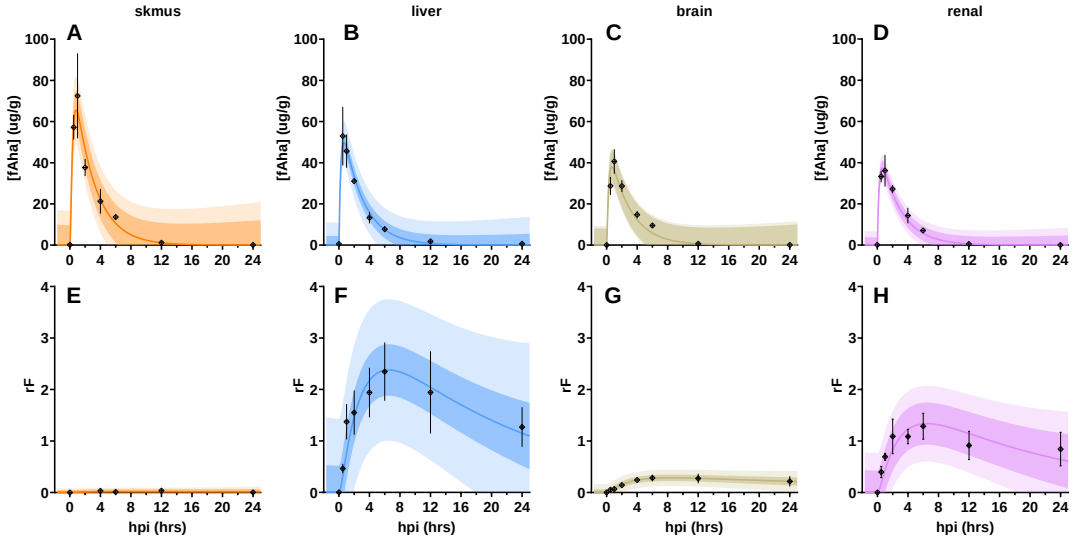


Fig. 4.2. Aha distribution and protein labeling given a dosing paradigm identical to the experimental methods (0.1 mg Aha per g body weight, injected subcutaneously) Top: [fAha] in each tissue over time. Bottom: rF signal in each tissue as a function of the [fAHA]. Filled points indicate mean experimental measurement at each time point, error bars indicate experimental standard deviation. Traces indicate best fit of model to each dataset, with darker and lighter shaded regions showing 95% prediction intervals for residual error based off the mean experimental value and all experimental values, respectively.

Table 4.2.
Tissue relative protein synthesis rates and protein half-lives. *Relative synthesis rate = $k_f k_s$, units include fluoresence response units (RU) and mass pAha/fAha ($\mu g_p/\mu g_f$)

Parameter	**Skmus**	**Liver**	**Brain**	**Renal**
Protein turnover rate, hr^{-1}	2.29e−2	4.91e−2	1.81e−2	4.88e−2
Protein half-life, hr	14.1	14.2	38.3	30.2
Relative syn. rate*, $RU(\mu g_p \mu g_f hr)^{-1}$	4.53e−5	4.19e−3	6.62e−4	5.41e−3

significantly among tissues. In addition, the estimated protein half-lives are consistent with previous studies that utilized isotope labeling to measure tissue protein half-lives using MS [118–120]. In these previous studies, measurements based on absolute quantitation using isotope labeling revealed that liver and kidney have higher turnover rates compared to brain and skeletal muscle. For example, Price et al. showed that the average half-life of brain proteins is 3 times higher than liver (9 and 3 days for brain and liver, respectively) [119]. Notably, our model estimated an average brain protein half-life that is 2.7 higher than the liver (38.3 and 14.1 hours for brain and liver, respectively) (table 4.2). The discrepancies between previously reported values and the half-lives values estimated here can be attributed to the shorter timescale of our experimental setup and the lower accuracy of western blotting measurements compared to MS.

4.3.4 Model Validation

Model validity was examined using a few distinct metrics to investigate parameter stability and goodness of fit.

Firstly, a prediction interval of additional regression values was generated for each tissue studied, for both the biodistribution and protein incorporation models. A distinct 95% prediction interval (PI) was calculated using all experimental replicates as well as a second tighter prediction interval using only the mean for each time point shown in figure 4.2. This prediction interval gives a reasonable expected range for future residuals based upon the observed experimental regression and distance from the informing dataset. For each tissue, the width of a 95% prediction interval (PI) from the average line of best fit ($\hat{y}$) can be approximated using a naively informed forecast interval that assumes a relatively normal distribution of residual error [121].

$$PI = \hat{y} \pm 1.96 \frac{\sum_i (y_i - \hat{y}_i)^2}{n} \sqrt{1 + \frac{1}{n} + \frac{(t - \bar{t})^2}{\sum_i (t_i - \bar{t})^2}} \tag{4.9}$$

Here n is the total number of observations, (t_i, y_i) are the observation time and [Aha] values for each tissue, t is the time point of the predicted residual, $\bar{t}$ is the average time of all experimental observations.

Secondly, the covariance matrix of least squares regression was used to inform a standard error estimate for all 19 fitted parameters in the biodistribution model and for all 8 fitted parameters in the protein incorporation model. This standard error gives an indication of the variability of each parameter, but also inherently reflects upon the definition of the model, parameters that are close to the constraints are less predictable, and so wide error can be indicative of a poorly constrained system. Error for each parameter also increases with the number of fitted parameters, is inversely related to the quantity of data points available for fitting, and is heavily influenced by the metric used for optimization. As such, these errors give a general indication of which parameters are well characterized, but are not an ideal metric for a complex non-linear regression. Notably the widest error ranges were for parameters that were the least well characterized by literature findings specifically the tissue specific import and export of fAha.

As a reinforcement for the standard error ranges determined by fitting covariance, partial rank correlation coefficients (PRCCs) were determined using monte-carlo latin hypercube sampling (LHS) to sweep each parameter over it's bound range and determine the influence of each parameter on the sum of square error metric used for optimization as described in our previous works [18, 113]. Here, PRCC values closer to a value of 1 in magnitude hold more sway over the dynamics of the entire system. It is clear that the most heavily weighted parameters seemed to be those that influence the system removal of fAha. Particularly, elimination rates and import rates into tissues that work actively to remove fAha tend to demonstrate PRCC > 0.1, although none of these parameters is greater than 0.8 which would indicate a poorly defined system.

All standard error and PRCC values can be seen in appendix D tables D.2-D.4.

4.3.5 Predictive Simulations of Dosing Paradigms

Attaining good protein labeling is critical to adequately enrich Aha-labeled proteins with high signal-to-noise ratio for accurate quantitative MS measurements and identification of newly synthesized proteins. Therefore, it is imperative to optimize the labeling strategy for the tissue of interest. Using the developed model, fAha biodistribution and tissue protein labeling could be predicted for alternative dosing regimens. Such predictions would help inform experimental design and predict labeling efficiency depending on the conditions of a study. To that end the model was adapted to accommodate multiple subcutaneous doses of varying magnitude.

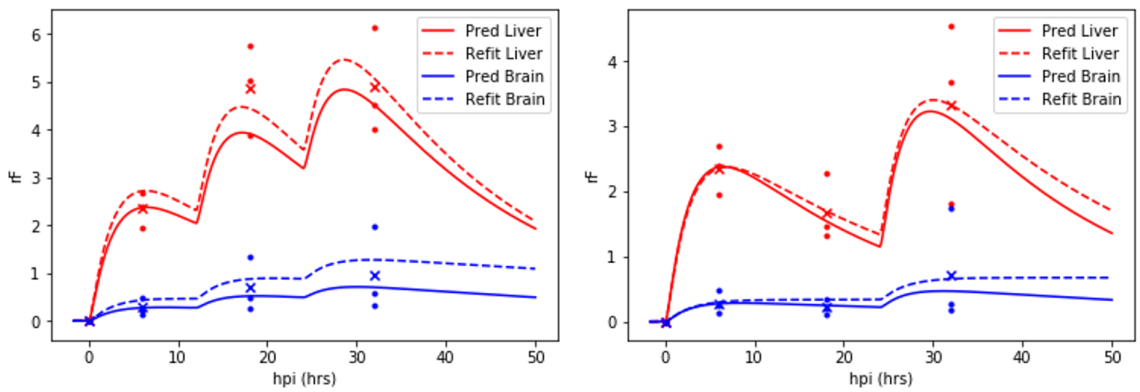


Fig. 4.3. Model accurately and precisely predicts relative labeling in the brain and liver with multiple injection doses. Left: rF experimental data and fit model for 12hrrd. Right: rF experimental data and fit model for 12hrrd. Color indicates liver (red) and brain (blue) tissues, solid line shows best fit of model from original robust dataset, dashed line shows refit model to repeated dose paradigm. Dots indicate individual experimental measurements and crosses represent experimental means for data collected 6, 18, and 32 hpi.

It is notable however, that such an adaptation for extrapolation is naively informed and required further validation. The parameter set for Aha distribution and labeling was characterized by a robust study containing fAha and rF data, recorded with LC-MS and western blot respectively, at 8 distinct time points (n = 3 biological replicates

at 0, 0.5, 1, 2, 4, 6, 12, and 24 hpi). While this study provided sufficient data to develop a well informed model of Aha patterning, it was limited to time points within 24 hr of a single subcutaneous injection of Aha. However, the extrapolated precision and accuracy of an already fully informed model can be reasonably characterized with a much more sparse data set.

Table 4.3.
Parameter and goodness-of-fit statistics related to alternative dosing paradigm predictive models

Tissue	**Parameter**	**Pred. Val. (SE)**	**Refit Val.**	%Δ
Liver	$k_f k_s$, $RU(\mu g_p \mu g_f hr)^{-1}$	3.16e−4 (2.49e−5)	3.65e−4	+15.5%
12 hrrd	k_d, hr^{-1}	8.29e−4 (1.80e−4)	8.67e−4	+4.6%
	SE_{reg}, RU	1.975	1.684	-14.7%
Brain	$k_f k_s$, $RU(\mu g_p \mu g_f hr)^{-1}$	4.30e−5 (2.83e−5)	4.84e−5	+12.5%
12 hrrd	k_d, hr^{-1}	3.54e−4 (1.11e−3)	1.37e−4	-61.3%
	SE_{reg}, RU	1.147	1.082	-5.7%
Liver	$k_f k_s$, $RU(\mu g_p \mu g_f hr)^{-1}$	3.16e−4 (2.49e−5)	3.02e−4	-4.4%
24 hrrd	k_d, hr^{-1}	8.29e−4 (1.80e−4)	6.72e−4	-18.9%
	SE_{reg}, RU	1.558	1.524	-2.2%
Brain	$k_f k_s$, $RU(\mu g_p \mu g_f hr)^{-1}$	4.30e−5 (2.83e−5)	4.28e−5	-0.5%
24 hrrd	k_d, hr^{-1}	3.54e−4 (1.11e−3)	1.37e−9	-99.9%
	SE_{reg}, RU	0.953	0.916	-3.9%

The model was used to predict the rF labeling of brain and liver tissues for two alternative dosing paradigms with (i) 12 hr repeated doses (hrrd) and (ii) 24 hrrd over a 48 hr period. As an internal control for western blot variation, these new experimental values were normalized by a shared time point with the previous study (6 hours post initial injection) for each tissue. The resulting values were used to refit

the relative pAha synthesis rate and degradation for each tissue under each repeated dose paradigm. When compared to these new data there was only a slight reduction in the standard error of regression, a goodness-of-fit metric, between the original and refit parameters in each tissue (12 hrrd: $\Delta SE_{reg} = [-14.7\%, -5.7\%]$; 24 hrrd: $\Delta SE_{reg} = [-2.2\%, -3.9\%]$; for [liver, brain] respectively). Additionally, among all refit parameters, only a single parameter was adjusted beyond a single standard error of fit from the original best fit value (12 hrrd liver $\Delta k_s k_f \approx +1.97SE$), and only the degradation rate in the brain changes by >20%. All changes in parameter values and goodness-of-fit are detailed in table 4.3.

4.4 Conclusions

In this Chapter, we report for the first time the biodistribution kinetics of Aha, the most widely used Met analog, in murine tissues, as well as the associated apparent rates of protein synthesis and turnover. These results show that liver and kidney have faster synthesis and turnover rates compared to brain and skeletal muscle, which is consistent with the results of previous studies that utilized isotope labeling for analyzing protein turnover rates. We also show that the injection technique allows observing maximum protein labeling in a relatively short time (6 h), which enables studying proteins with shorter half-lives, in contrast to the traditional method of introducing the ncAA in diet or using isotope labeled amino acids. Additionally, we report the development of a mathematical framework that describes the distribution kinetics of Aha in murine tissues and its relation to the degree of protein labeling, and computes the apparent relative rates of protein synthesis and turnover. We further validate this framework for predictive modeling of Aha labeling against a distinct dataset including two alternative repeated injection dosing paradigms to demonstrate it's efficacy as a tool for future experimental design.

5. RECOMMENDATIONS

5.1 Implications of FRCS

5.1.1 Applications and Limitations

In this work we have demonstrated that the FRCS algorithm and data processing methodology is capable of capturing dynamics of a system at a sub-measurement-time scale, with a potential temporal resolution equal to the scale at which the experimental photon count rate is non-sparse (∼10-100 ns). At these time scales this technique should be capable of characterizing behavior across a broad dynamic range that covers most of the feasible PPI kinetics that might be captured by any of the presently available technologies such as SPR [$k_a \epsilon(1\text{e}4, 1\text{e}7)$ $\text{M}^{-1}\cdot\text{s}^{-1}$, $k_d \epsilon(1\text{e}{-4}, 1\text{e}1)$ s^{-1}] [22]. Along with the dynamic capabilities FRCS is uniquely simple in experimental design. Due to the singly-labeled nature of the experimental design, FRCS is ideal to characterize the binding behaviors of a target of interest, such as CaM, which can be engineered relatively simply to posses a fluorescent label. This labeled target can then be repeatedly used to characterize binding kinetics of an entire set of other proteins, each which can be expressed as a wild-type without modifications or even purchased in small quantities. Furthermore, unlike current gold-standard methods of binding kinetic determination, FRCS does not require any immobilization or surface binding of proteins and therefore can be used to characterize binding in a more native state. However, it is notable that current experimental constraints do present some additional limitations to this technology that other techniques may not have.

Firstly, it is notable that for single-photon FCS, it has been well characterized to poorly distinguish between particles which characteristic residency times that differ by less than a factor of 1.6 [26, 92]. Given that residency and, by proxy, diffusivity is a function of the hydrophobic radius, and assuming we model proteins as globular spheres with relative uniform density this would demand that the mass of ligand is a factor of $1.6^3 \approx 4$ times less than the mass of ligand-receptor complex. In this case, the

receptor must be at least a factor of 3 heavier than the labeled ligand, to characterize using this single-photon FRCS, assuming that the act of complex formation does not result in a more dense globular protein, which realistically is perhaps a likely. That being said, the same logic used to characterize binding kinetics using single-photon FRCS could be applied to a two-photon system with differentially labeled ligand and receptor. In this case the magnitude of the cross correlation of the two channels would indicate the concentration of complex species. The same overlapping correlation algorithm could easily be used to monitor the time resolved changes in concentration as determined by the cross correlation as could be used to determine the fractional composition in a singly labeled auto-correlation system.

Further complications are present in the actual experimental design of FRCS experiments. Foremost among these is the likely possibility that FRCS experimental measurements will be significantly more noisy than the simulated findings in figure 3.5. While these simulations predict the random noise of diffusion in an otherwise ideal theoretical system in which the diffusivity of both the ligand and complex ligand-receptor species are fixed to a known value. They don't capture true biophysical variation in protein mass, globular density, and orientation specific binding. As such, experimental FRCS will likely require many more replicates than the simulations would imply in order to characterize the true average behavior. Additionally, for the low concentrations of proteins characteristic for FCS, ideal buffer conditions become critically important, slight variations may influence not only the binding kinetics being characterized, but also could render proteins to completely precipitate, or non-specifically bind to each other and the surface of the measurement apparatus. These issues can be minimized by thoughtful design, such as by including inert protein such Bovine Serum Albumin (BSA) in the buffers and by optimizing other conditions such as salt concentration, ion concentration and pH through iterative experimentation or extensive literature searches. Another issue that has been encountered in preliminary studies is the possibility that the fluorophore may interfere with binding if localized too closely with the binding domain of the protein of interest. This could perhaps

be addressed on a case-by-case basis by selectively relocating the fluorophore on the labeled peptide, possibly using some form of flexible linker to isolate the binding domain from potential interference as much as possible.

A more difficult to address limitation lies in defining a mixing protocol for actual experimental FRCS studies. As mentioned previously, the Michaelis-Menten model of protein binding and enzyme kinetics relies on an assumption that the solution is well mixed. This means that each species present in the solution is uniformly distributed at the exact moment that the reaction begins to proceed. As mentioned in 3, this can be accomplished experimentally by mixing the binding species thoroughly just before measurement with gentle pipetting, flicking and inversion, but this inherently means that the earliest binding reaction time point captured by the first FRCS frame will fail to capture the intial binding activity. For a correlation frame width of around 20 seconds, nearly 10 seconds of binding behavior would be lost by such a measurement with this mixing protocol. This time loss is further increased by the length of time that it takes to mix and subsequently begin measuring the solution. As seen in figure 3.5, a loss such as described here is negligible for slower dynamics that occur over the time span of a few minutes. However, for PPIs with rapid association kinetics, particularly those with lower binding affinities that produce a complex that exists only transiently, this loss of time makes it nearly impossible to characterizze the true kinetics. Even in the slower timescales, the length of time between "mixing" and the first correlation frame must be well characterized and consistent. Ideally, it would best to have a running measurement of the free ligand in isolation and without pausing in the measurement, somehow introduce and "instantaneously" mix in the inert receptor without affecting diffusion or concentration of the labeled ligand. Barring this possibility it would at least be good practice to automate the mixing process for consistency between replicates. Proposed methods for mixing protocols are discussed a bit more in the following section.

5.1.2 Future Work

With FRCS algorithm and software developed, and the theory of the technique validated with preliminary simulations and experimental measurements the next steps for this tool are focused on applying the technique to characterizing a diverse set of PPIs. In particular, we are interested in characterizing the sub-saturated binding behavior of CaM with $Ca2^{2+}$ as discussed in chapter 3. To do so, we plan to express mutants of CaM engineered with both fluorescent labels and excised Ca^{2+} binding domains. These mutants will then be used to determine binding affinity with a subset of CaM partners that our lab has worked with including CaMKII (both in monomeric and holoenzyme form), CaN, PSD95, and other proteins. The eventual goal would be to use this technique to begin to developing a shared a database of binding kinetics usable for kinetic models of protein interactions. We would further like to apply this tool to more advanced studies of binding including introduction of competitive inhibitors. In a standard FRCS experiment as described here we anticipate the reduced apparent binding affinity $K_{D,app}$ and even apparent association rates should be able to be captured with this tool. Perhaps an excess of competitive inhibitor could even be added during the time course of FRCS to demonstrate a more rigorous definition of the dissociation rate for a given binding interaction. Among the proteins we hope to study a few notable systems are likely explored by our group in the near future as detailed by table 5.1 below.

In the process of obtaining these measurements, we also would like to begin to develop a more firm experimental design of the mixing protocol of FRCS. The ideal target of this work would be a protocol that enables a 'mid-measurement mixing' experimental design as discussed above and in chapter 3. One possible approach to accomplish this might be with a microfluidic chip that enables simultaneous injection of two solutions into an actively measured reservoir.

In a chip designed for the purpose, such as displayed in the cartoon diagram (figure 5.1), a measurement can be performed that incorporates the desired mixing. For this

Table 5.1.
Binding reaction systems that are good targets of future study with FRCS.

'Ligand' (fA)	**Alternate fA**	**'Receptors' (B)**	**Pred. K_D (nM), Source**
CaM	CaM0	CaN	0.5-50, [18, 95]
	CaM4	mCaMKII	50, [18, 122]
	CaM2N	hCaMKII	< 50, [18, 122]
	CaM2C	mCaMKII + PPP1cα	> 50 , [19]
		hCaMKII + PPP1cα	> 50 , [19]
		PSD95	$(10 - 20)$e3, [123]
		Neurogranin	$(10 - 100)$e3, [18, 124]
Glucosamine	$\pm$ PEG6 Linker	GGBP	200, [125, 126]
Galactoseamine	$\pm$ PEG6 Linker	GGBP	400, [126]

particular design, a central reservoir initially contains a known volume of free ligand (V_0) at a known concentration ($[A]_0$). At some time $t = 0$ an FCS measurement of the reservoir could be started and recorded for the duration of at least one full correlation frame. Then, at some time $t = t_{rxn,0}$ equal volumes of two separate solutions can be quickly pushed into this reservoir, generating a brief turbulance that mixes all three solutions. The first of these injected solutions is a solution of free ligand at twice the concentration already present in the reservoir ($2[A]_0$), such that the final concentration of ligand in the new solution of volume ($3V_0$) is the same as the original concentration. The second of these injected solutions contains the inert receptor at a concentration triple that of the desired FRCS measuring concentration ($3[B]_0$), such that the final mixed concentration is B_0. This mixing step may produce some artefacts in the photon count stream observed during the injection, but when averaged over the entire measurement interval of each frame, this noisy behavior will hopefully be relatively minimal and contribute negligibly to the evaluated binding measurements.

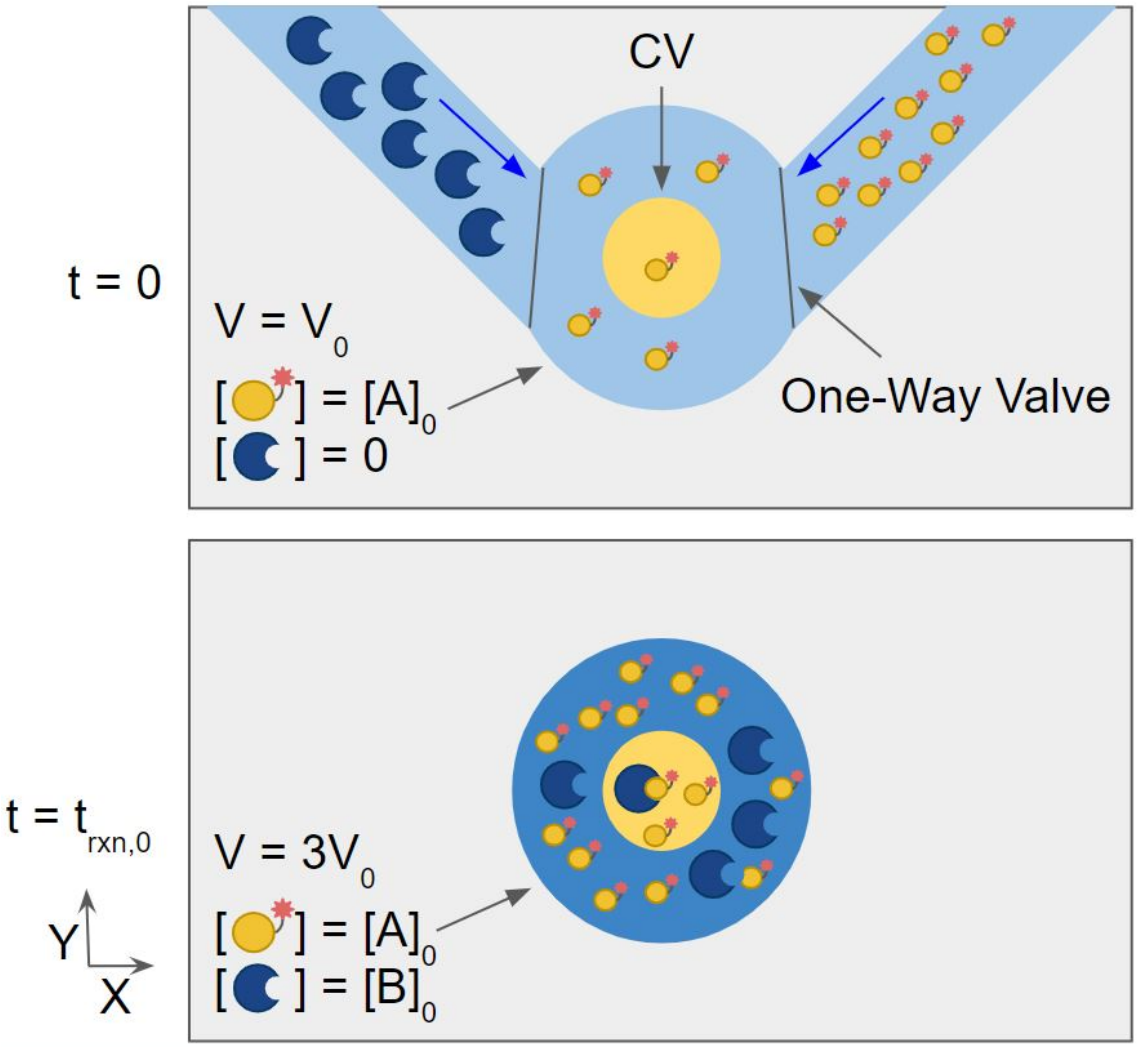


Fig. 5.1. A cartoon diagram of a chip enabling FRCS with a 'mid-measurement' mixing of reagents. This chip allows simultaneous injection and mixing of two solutions at some time $t = t_{rxn,0}$ (injection flow direction described by blue arrows), into a static reservoir during an active measurement started at $t = 0$. Note the CV and particles are not drawn to scale.

Additionally, future efforts will be invested in further advancing the capabilities of both simulations and the GUI for FRCS, in order to capture a broader range of realistic experimental conditions. It is our hope that these tools can be used to

aid experimental design and data processing of even more complex systems than are described here.

5.2 Implications of Aha Labeling

5.2.1 Applications and Limitations

The simulated model and experimental methodology presented in chapter 4 demonstrate predictable labeling of the *in vivo* proteome in a murine model with subcutaneous injections of Aha. The model of ODEs is presented primarily as a tool to guide experimental design of future studies. With rapid distribution to all measured tissues within 1 hpi, this technique of ncAA labeling can be used to probe the proteome response to some applied stimuli with a relatively fine time resolution. Furthermore, with the presented model to predict protein labeling, ideal time points for injections and sacrifice of mice can be selected for optimal labeling of each tissue. While this model is presently useful as a guide for experimental design, there are some apparent limitations that should be considered for a more rigorous application.

Firstly, this model does not have strong quantification of the degree of protein labeling. We have represented proteinous labeling with Aha only by proxy, using the relative fluorescence from western blot analysis of biotinylation. Protein from each tissue labeled with Aha was ligated to biotin, and compared to an uninjected background biotinylated protein concentration from that same tissue. We have not utilized here a universal internal control between tissues to demonstrate truly independent degrees of protein labeling across all tissues. As such, tissues with higher levels of basal biotinylation will result in low predicted relative labeling in our study. Furthermore, because we don't have a strong control between tissues we don't have a good way to characterize the value of k_f, the fluorescence constant, describing the degree of fluorescence expected given some degree of protein incorporated pAha. This is all to say that while the time scale and overall profile of labeling is well charac-

terized by this model of ODES, any further conclusions particularly those regarding predicted pAha synthesis rates between tissues are somewhat naively informed.

Secondly, the timescales of this study are relatively short. Even in our predictive study of multiple doses, we only record data out to 32 hpi. It is possible that at longer timescales, a variety of unpredictable behaviors may begin to occur. A persistance of Aha may for instance result in eventual saturation or breakdown of the biological machinery responsible for removing Aha from the system. Longer Aha lifetimes may also result in labeling of longer lived proteins that persist beyond the expected profiles in figure 4.3. Either of these behaviors would result in correspondingly longer half-lives, and could even result in increased toxicity of Aha in the system.

A third limitation is that this study only captures the dynamics of distribution and labeling following subcutaneous injections. While these injections are ideal for rapid < 1 hr Aha peaks, they do not result in persistence of Aha. All labeling of protein therefore is most likely occur within the short window that Aha is in excess. Alternative dosing mechanisms may be capable of demonstrating more consistent Aha levels that may be favorable for some experimental designs.

5.2.2 Future Work

With the notable limitations described in the previous section, future studies should likely attempt to address these shortcomings. Ideally, another robust study would be performed that characterizes the distribution dynamics of alternate dosing paradigms including injections of variable magnitude (to insure initial absorption from the injection site is unaffected by dose size) as well as oral dosing through feed (slow absorption bolus) or water (relatively continuous dosing). It may be feasible that a combination of these dosing paradigms could be selected that allows for rapid initial distribution of Aha as well as maintained plasma concentrations of Aha after the initial dose, perhaps using an initial Aha injection and subsequent supplementation with dietary Aha. A mechanism of oral dosing is already present in the complete

model of ODEs as described below in B.1, although best fit parameters were not characterized due to absence of supporting data.

Further attempts to truly quantify pAha labeling should also be considered. One possible approach would be to extract protein from labeled tissues via homogenization, desalt the sample to remove free amino acids, then fully digest the labeled protein into independent amino acids with a cocktail of digestive enzymes. The resulting digest could then be measured with LC-MS to quantify total mass of ions with the characteristic mass of Aha and therefore the concentration of pAha in the original homogenate.

Lastly, we hope to pursue application of this technique as a tool for characterization of dynamic response of the proteome not only over the time course of natural function (such as our lab's previous work studying proteome changes during development, [81, 82]), but also in response to unnatural stimuli such as pharmaceuticals. With it's rapid distribution and potential for longer term labeling of the proteome, this technique could be a powerful tool for purposes such as pre-clinical screening of potential future drugs.

5.3 Final Thoughts

It is our hope that each of these tools can be utilized by our lab and others for future experimental studies of protein behavior. FRCS presents an accessibly elegant method of characterizing kinetic parameters of PPIs that should be simple to implement in any confocal microscopy enabled labs. Similarly, we hope that the presented methods and model of Aha biodistribution and labeling will be useful for future planning of experiments for any biochemists with a need for targeted proteome analysis.

All of tools discussed in this work, including models and software for both techniques will be made available and open source (can be accessed via the Kinzer-Ursem

Lab Github repository, appendix A) to any labs interested in their application in the shared pursuit of further depth of knowledge in protein behavior.

REFERENCES

[1] J. D. Watson and F. H. Crick, "Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid," *Nature*, vol. 171, 1953.

[2] G. M. Church and W. Gilbert, "Genomic sequencing," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 81, 1984.

[3] R. Aebersold and M. Mann, "Mass spectrometry-based proteomics," 2003.

[4] W. Widłak, "Protein structure and function," pp. 15–29, 2013.

[5] G. Wu, "Amino acids: Metabolism, functions, and nutrition," pp. 1–17, 5 2009. [Online]. Available: https://link.springer.com/article/10.1007/s00726-009-0269-0

[6] A. J. Link, M. L. Mock, and D. A. Tirrell, "Non-canonical amino acids in protein engineering," pp. 603–609, 12 2003.

[7] A. Stoddart, "Molecular imaging: Seeing the target," pp. 1–1, 7 2016. [Online]. Available: https://www.nature.com/articles/natrevmats:201657

[8] A. F. Altelaar, J. Munoz, and A. J. Heck, "Next-generation proteomics: Towards an integrative view of proteome dynamics," 2013.

[9] P. D. Boyer, "The atp synthase - a splendid molecular machine," 1997.

[10] E. Z. Eisenmesser, O. Millet, W. Labeikovsky, D. M. Korzhnev, M. Wolf-Watz, D. A. Bosco, J. J. Skalicky, L. E. Kay, and D. Kern, "Intrinsic dynamics of an enzyme underlies catalysis," *Nature*, vol. 438, 2005.

[11] B. Yurke, A. J. Turberfield, A. P. Mills, F. C. Simmel, and J. L. Neumann, "A dna-fuelled molecular machine made of dna," *Nature*, vol. 406, 2000.

[12] L. Michaelis, M. L. Menten, R. S. Goody, and K. A. Johnson, "Die kinetik der invertinwirkung/ the kinetics of invertase action," *Biochemistry*, vol. 49, 1913.

[13] J. M. Berg, J. L. Tymoczko, and L. Stryer, "The michaelis-menten model accounts for the kinetic properties of many enzymes," 2002. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK22430/

[14] K. A. Johnson and R. S. Goody, "The original michaelis constant: Translation of the 1913 michaelis-menten paper," *Biochemistry*, vol. 50, 2011.

[15] S. Gianni, A. Engström, M. Larsson, N. Calosci, F. Malatesta, L. Eklund, C. C. Ngang, C. Travaglini-Allocatelli, and P. Jemth, "The kinetics of pdz domain-ligand interactions and implications for the binding mechanism." *The Journal of biological chemistry*, vol. 280, pp. 34 805–12, 10 2005. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16049001

[16] P. G. D. Benedetti and F. Fanelli, "Computational modeling approaches to quantitative structure–binding kinetics relationships in drug discovery," 2018.

[17] R. A. Pearlstein, K. A. MacCannell, G. Erdemli, S. Yeola, G. Helmlinger, Q.-Y. Hu, R. Farid, W. Egan, S. Whitebread, C. Springer, J. Beck, H.-R. Wang, M. Maciejewski, L. Urban, and J. S. Duca, "Implications of dynamic occupancy, binding kinetics, and channel gating kinetics for herg blocker safety assessment and mitigation," *Current Topics in Medicinal Chemistry*, vol. 16, 2016.

[18] D. R. Romano, M. C. Pharris, N. M. Patel, and T. L. Kinzer-Ursem, "Competitive tuning: Competition's role in setting the frequency-dependence of ca2+-dependent proteins," *PLOS Computational Biology*, vol. 13, p. e10058:20, 11 2017. [Online]. Available: https://dx.plos.org/10.1371/journal.pcbi.10058:20

[19] M. C. Pharris, N. M. Patel, T. G. VanDyk, T. M. Bartol, T. J. Sejnowski, M. B. Kennedy, M. I. Stefan, and T. L. Kinzer-Ursem, "A multi-state model of the camkii dodecamer suggests a role for calmodulin in maintenance of autophosphorylation," *PLOS Computational Biology*, vol. 15, p. e1006941, 12 20:19. [Online]. Available: https://dx.plos.org/10.1371/journal.pcbi.1006941

[20] Y. Ou, R. E. Wilson, and S. G. Weber, "Methods of measuring enzyme activity ex vivo and in vivo," pp. 509–533, 6 2018. [Online]. Available: /pmc/articles/PMC6147230/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6147230/

[21] B. Liedberg, C. Nylander, and I. Lunström, "Surface plasmon resonance for gas detection and biosensing," *Sensors and Actuators*, vol. 4, pp. 299–304, 1 1983.

[22] E. Helmerhorst, D. J. Chandler, M. Nussio, and C. D. Mamotte, "Real-time and label-free bio-sensing of molecular interactions by surface plasmon resonance: A laboratory medicine perspective," pp. 161–173, 2012. [Online]. Available: http://www.scirus.com/

[23] P. Schuck and H. Zhao, "The role of mass transport limitation and surface heterogeneity in the biophysical characterization of macromolecular binding processes by spr biosensing." pp. 15–54, 2010.

[24] R. L. Rich and D. G. Myszka, "Survey of the 2009 commercial optical biosensor literature," pp. 892–914, 11 2011. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/jmr.1138 https://onlinelibrary.wiley.com/doi/abs/10.1002/jmr.1138 https://onlinelibrary.wiley.com/doi/10.1002/jmr.1138

[25] S. N. Olsen, "Applications of isothermal titration calorimetry to measure enzyme kinetics and activity in complex solutions," *Thermochimica Acta*, vol. 448, pp. 12–18, 9 2006.

[26] P. Schwille and E. Haustein, "Fluorescence correlation spectroscopy - an introduction to its concepts and applications," *Spectroscopy*, vol. 94(22), 01 2001.

[27] W. A. Lea and A. Simeonov, "Fluorescence polarization assays in small molecule screening," pp. 17–32, 1 2011. [Online]. Available: /pmc/articles/PMC3277431/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3277431/

[28] D. M. Molina, R. Jafari, M. Ignatushchenko, T. Seki, E. A. Larsson, C. Dan, L. Sreekumar, Y. Cao, and P. Nordlund, "Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay," *Science*, vol. 341, pp. 84–87, 7 2013. [Online]. Available: https://science.sciencemag.org/content/341/6141/84 https://science.sciencemag.org/content/341/6141/84.abstract

[29] G. W. Brodland, "How computational models can help unlock biological systems," pp. 62–73, 12 2015.

[30] Z. Ji, K. Yan, W. Li, H. Hu, and X. Zhu, "Mathematical and computational modeling in complex biological systems," 2017.

[31] W. Timp and G. Timp, "Beyond mass spectrometry, the next step in proteomics," *Science Advances*, vol. 6, p. eaax8978, 1 2020. [Online]. Available: http://advances.sciencemag.org/

[32] B. F. Thomas, R. C. Daw, and M. Grabenauer, "General principles of mass spectrometry: Gc-ms, lc-ms, and lc-ms/ms," pp. 1–26, 10 2012. [Online]. Available: http://doi.wiley.com/10.1002/9780470921920.edm091

[33] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse, "Electrospray ionization for mass spectrometry of large biomolecules," pp. 64–71, 10 1989. [Online]. Available: https://science.sciencemag.org/content/246/4926/64 https://science.sciencemag.org/content/246/4926/64.abstract

[34] M. Karas and F. Hillenkamp, "Laser desorption ionization of proteins with molecular masses exceeding 10 000 daltons," pp. 2299–2301, 10 1988. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/3239801/

[35] X. Han, A. Aslanian, and J. R. Yates, "Mass spectrometry for proteomics," pp. 483–490, 10 2008. [Online]. Available: /pmc/articles/PMC2642903/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2642903/

[36] R. E. Cone, "Radiolabeling," pp. 2053–2055, 1 1998.

[37] M. Wexler, D. J. Richardson, and P. L. Bond, "Radiolabelled proteomics to determine differential functioning of accumulibacter during the anaerobic and aerobic phases of a bioreactor operating for enhanced biological phosphorus removal," *Environmental Microbiology*, vol. 11, pp. 3029–3044, 12 2009. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/:19650829/

[38] K. M. Carroll, F. Lanucara, and C. E. Eyers, "Quantification of proteins and their modifications using qconcat technology," pp. 113–131, 1 2011.

[39] A. Chokkathukalam, D. H. Kim, M. P. Barrett, R. Breitling, and D. J. Creek, "Stable isotope-labeling studies in metabolomics: New insights into structure and dynamics of metabolic networks," pp. 511–524, 2 2014. [Online]. Available: /pmc/articles/PMC4048731/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4048731/

[40] P. J. H. Jones and S. T. Leatherdale, "Editorial review stable isotopes in clinical research: safety reaffirmed," pp. 277–280, 1991.

[41] D. Dieterich, A. Link, J. Graumann, D. Tirrell, and E. Schuman, "Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (boncat," in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, pp. 9482–9487,.

[42] A. Trautwein-Schult, S. Maaß, K. Plate, A. Otto, and D. Becher, "A metabolic labeling strategy for relative protein quantification in clostridioides difficile," *Frontiers in Microbiology*, vol. 9, p. 2371, 10 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/fmicb.2018.02371/full

[43] D. Magde, E. L. Elson, and W. W. Webb, "Fluorescence correlation spectroscopy. ii. an experimental realization," *Biopolymers*, vol. 13, pp. 29–61, 1974. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/4818131/

[44] S. T. Hess, S. Huang, A. A. Heikal, and W. W. Webb, "Biological and chemical applications of fluorescence correlation spectroscopy: A review," *Biochemistry*, vol. 41, pp. 697–705, 1 2002. [Online]. Available: https://pubs.acs.org/sharingguidelines

[45] E. Haustein and P. Schwille, "Ultrasensitive investigations of biological systems by fluorescence correlation spectroscopy," *Methods*, vol. 29, pp. 153–166, 2 2003.

[46] P. Schwille, F. Oehlenschläger, and N. G. Walter, "Quantitative hybridization kinetics of dna probes to rna in solution followed by diffusional fluorescence correlation analysis †," 1996. [Online]. Available: https://pubs.acs.org/sharingguidelines

[47] R. Macháň and T. Wohland, "Recent applications of fluorescence correlation spectroscopy in live systems," *FEBS Letters*, vol. 588, pp. 3571–3584, 10 2014.

[48] P. Schwille, F. J. Meyer-Almes, and R. Rigler, "Dual-color fluorescence cross-correlation spectroscopy for multicomponent diffusional analysis in solution," *Biophysical Journal*, vol. 72, pp. 1878–1886, 1997. [Online]. Available: /pmc/articles/PMC1184381/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1184381/

[49] J. Schüler, J. Frank, U. Trier, M. Schäfer-Korting, and W. Saenger, "Interaction kinetics of tetramethylrhodamine transferrin with human transferrin receptor studied by fluorescence correlation spectroscopy," *Biochemistry*, vol. 38, pp. 8402–8408, 6 1999. [Online]. Available: https://pubs.acs.org/sharingguidelines

[50] A. Michelman-Ribeiro, D. Mazza, T. Rosales, T. J. Stasevich, H. Boukari, V. Rishi, C. Vinson, J. R. Knutson, and J. G. McNally, "Direct measurement of association and dissociation rates of dna binding in live cells by fluorescence correlation spectroscopy," *Biophysical Journal*, vol. 97, pp. 337–346, 2009. [Online]. Available: /pmc/articles/PMC2711375/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2711375/

[51] R. Oyama, H. Takashima, M. Yonezawa, N. Doi, E. Miyamoto-Sato, M. Kinjo, and H. Yanagawa, "Protein - protein interaction analysis by c-terminally specific fluorescence labeling and fluorescence cross-correlation spectroscopy," *Nucleic Acids Research*, vol. 34, p. e102, 2006. [Online]. Available: /pmc/articles/PMC:1904107/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC:1904107/

[52] E. K. Sackmann, A. L. Fulton, and D. J. Beebe, "The present and future role of microfluidics in biomedical research," *Nature*, vol. 507, pp. 181–189, 2014. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/24622:198/

[53] W. Sutherland, "Lxxv. a dynamical theory of diffusion for non-electrolytes and the molecular mass of albumin," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 9, pp. 781–785, 6 1905. [Online]. Available: https://zenodo.org/record/1430774

[54] A. Einstein, "Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen," *Annalen der Physik*, vol. 322, pp. 549–560, 1905. [Online]. Available: http://sedici.unlp.edu.ar/handle/10915/2785

[55] M. von Smoluchowski, "Zur kinetischen theorie der brownschen molekularbewegung und der suspensionen," *Annalen der Physik*, vol. 326, pp. 756–780, 1906. [Online]. Available: https://zenodo.org/record/1424073

[56] A. Z. Politi, Y. Cai, N. Walther, M. J. Hossain, B. Koch, M. Wachsmuth, and J. Ellenberg, "Quantitative mapping of fluorescently tagged cellular proteins using fcs-calibrated four-dimensional imaging," *Nature Protocols*, vol. 13, pp. 1445–1464, 6 2018. [Online]. Available: /pmc/articles/PMC6609853/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6609853/

[57] B. D. Gamari, D. Zhang, R. E. Buckman, P. Milas, J. S. Denker, H. Chen, H. Li, and L. S. Goldner, "Inexpensive electronics and software for photon statistics and correlation spectroscopy," *American Journal of Physics*, vol. 82, pp. 712–722, 7 2014. [Online]. Available: /pmc/articles/PMC4768833/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4768833/

[58] D. Magde, W. W. Webb, and E. L. Elson, "Fluorescence correlation spectroscopy. iii. uniform translation and laminar flow," *Biopolymers*, vol. 17, pp. 361–376, 2 1978. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/bip.1978.360170208 https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.1978.360170208 https://onlinelibrary.wiley.com/doi/10.1002/bip.1978.360170208

[59] L. Stryer, "Fluorescence energy transfer as a spectroscopic ruler," *Annual Review of Biochemistry*, vol. 47, pp. 8:19–846, 6 1978. [Online]. Available: http://www.annualreviews.org/doi/10.1146/annurev.bi.47.070178.004131

[60] K. Schätzel, "Noise on photon correlation data. i. autocorrelation functions," *Quantum Optics: Journal of the European Optical Society Part B*, vol. 2, pp. 287–305, 8 1990. [Online]. Available: https://iopscience.iop.org/article/10.1088/0954-8998/2/4/002 https://iopscience.iop.org/article/10.1088/0954-8998/2/4/002/meta

[61] S. Gong, I. Labanca, I. Rech, and M. Ghioni, "A 32-channel photon counting module with embedded auto/cross-correlators for real-time parallel fluorescence correlation spectroscopy," *Review of Scientific Instruments*, vol. 85, p. 103101, 10 2014. [Online]. Available: http://aip.scitation.org/doi/10.1063/1.4896695

[62] L. L. Yang, H. Y. Lee, M. K. Wang, X. Y. Lin, K. H. Hsu, Y. R. Chang, W. Fann, and J. D. White, "Real-time data acquisition incorporating high-speed software correlator for single-molecule spectroscopy," *Journal of Microscopy*, vol. 234, pp. 302–310, 6 2009. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/:19493109/

[63] E. Schaub, "F2cor: fast 2-stage correlation algorithm for fcs and dls," *Optics Express*, vol. 20, p. 2184, 1 2012.

[64] S. M. Trimberger, "Three ages of fpgas: A retrospective on the first thirty years of fpga technology: This paper reflects on how moore's law has driven the design of fpgas through three epochs: The age of invention, the age of expansion, and the age of accumulation," *IEEE Solid-State Circuits Magazine*, vol. 10, pp. 16–29, 3 2018.

[65] G. Kloss, "Automatic c library wrapping ctypes from the trenches," *The Python Papers*, vol. 3, no. 3.

[66] M. Newville, T. Stensitzki, D. B. Allen, and A. Ingargiola, "Lmfit: Non-linear least-square minimization and curve-fitting for python," 9 2014. [Online]. Available: https://zenodo.org/record/11813

[67] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science and Engineering*, vol. 9, p. 90–95, publisher link.

[68] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. Fernández del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, p. 357–362, 2020.

[69] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[70] Y. S. Babu, C. E. Bugg, and W. J. Cook, "Structure of calmodulin refined at 2.2 Å resolution," *Journal of Molecular Biology*, vol. 204, pp. 191–204, 11 1988. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/3145979/

[71] J. M. Shifman, M. H. Choi, S. Mihalas, S. L. Mayo, and M. B. Kennedy, "Ca2+/calmodulin-dependent protein kinase ii (camkii) is activated by calmodulin with two bound calciums," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, pp. 13 968–13 973, 9 2006. [Online]. Available: www.pnas.orgcgidoi10.1073pnas.0606433103

[72] K. Ando, Y. Kudo, K. Aoyagi, R. Ishikawa, M. Igarashi, and M. Takahashi, "Calmodulin-dependent regulation of neurotransmitter release differs in subsets of neuronal cells," *Brain Research*, vol. 1535, pp. 1–13, 10 2013.

[73] H. Tabuchi, H. Yamamoto, K. Matsumoto, K. Ebihara, Y. Takeuchi, K. Fukunaga, H. Hiraoka, Y. Sasaki, M. Shichiri, and E. Miyamoto, "Regulation of insulin secretion by overexpression of ca2+/calmodulin-dependent protein kinase ii in insulinoma min6 cells," *Endocrinology*, vol. 141, pp. 2350–2360, 2000.

[74] K. Török and D. R. Trentham, "Mechanism of 2-chloro-(ϵ-amino-lys75)-[6-[4-(n,n-diethylamino)phenyl]-1,3,5-triazin-4-yl]calmodulin interactions with smooth muscle myosin light chain kinase and derived peptides," *Biochemistry*, vol. 33, pp. 12 807–12 820, 11 1994. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/7947686/

[75] P. Zhang, S. Tripathi, H. Trinh, and M. S. Cheung, "Opposing intermolecular tuning of ca2+ affinity for calmodulin by neurogranin and camkii peptides," *Biophysical Journal*, vol. 112, pp. 1105–11:19, 3 2017. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28355539/

[76] R. Y. Tsien, "The green fluorescent protein," pp. 509–544, 1 1998. [Online]. Available: https://api.semanticscholar.org/CorpusID:8138960

[77] R. N. Day and M. W. Davidson, "The fluorescent protein palette: Tools for cellular imaging," *Chemical Society Reviews*, vol. 38, pp. 2887–2921, 9 2009. [Online]. Available: /pmc/articles/PMC2910338/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2910338/

[78] M. Grabner, R. T. Dirksen, and K. G. Beam, "Tagging with green fluorescent protein reveals a distinct subcellular distribution of l-type and non-l-type ca2+ channels expressed in dysgenic myotubes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, pp. 1903–1908, 2 1998. [Online]. Available: www.pnas.org.

[79] H. C. Kolb, M. G. Finn, and K. B. Sharpless, "Click chemistry: Diverse chemical function from a few good reactions," pp. 2004–2021, 6 2001.

[80] J. G. Fraseur and T. L. Kinzer-Ursem, "Next generation calmodulin affinity purification: Clickable calmodulin facilitates improved protein purification," *PLoS ONE*, vol. 13, 6 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29864125/

[81] A. Saleh, K. Wilding, S. Calve, B. Bundy, and T. Kinzer-Ursem, "Non-canonical amino acid labeling in proteomics and biotechnology," *Journal of Biological Engineering*, vol. 13, pp. 43,.

[82] A. Saleh, K. Jacobson, T. Kinzer-Ursem, and S. Calve, "Dynamics of non-canonical amino acid-labeled intra- and extracellular proteins in the developing mouse," *Cell Mol Bioeng*, vol. 12, no. 5, pp. 495–509,.

[83] "Guide to basic fcs experiments on zeiss lsm 780/confocor3/big systems running zen 2010." [Online]. Available: http://www.invitrogen.com/site/us/en/home/References/Molecular-Probes-The-Handbook/Technical-

[84] P. Müller, P. Schwille, and T. Weidemann, "Pycorrfit - generic data evaluation for fluorescence correlation spectroscopy," *Bioinformatics*, vol. 30, no. 17, p. 2532–2533.

[85] P. Kapusta, "Absolute diffusion coefficients: Compilation of reference data for fcs calibration," 2010. [Online]. Available: http://dx.doi.org/10.1002/cphc.200600638

[86] J. R. Stiles and T. M. Bartol, *Monte Carlo Methods for Simulating Realistic Synaptic Microphysiology using MCELL*, 2001.

[87] J. R. Stiles, D. V. Helden, T. M. Bartol, E. E. Salpeter, and M. M. Salpeter, "Miniature endplate current rise times ¡ 100 µs from improved dual recordings can be modeled with passive acetylcholine diffusion from a synaptic vesicle," *Biophysics*, vol. 93, pp. 5747–5752, 1996.

[88] R. A. Kerr, T. M. Bartol, B. Kaminsky, M. Dittrich, J.-C. J. Chang, S. B. Baden, T. J. Sejnowski, and J. R. Stiles, "Fast monte carlo simulation methods for biological reaction-diffusion systems in solution and on surfaces." *SIAM journal on scientific computing : a publication of the Society for Industrial and Applied Mathematics*, vol. 30, p. 3126, 10 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20151023 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2819163

[89] S. Gupta, J. Czech, R. Kuczewski, T. M. Bartol, T. J. Sejnowski, R. E. C. Lee, and J. R. Faeder, "Spatial stochastic modeling with mcell and cellblender," 2018.

[90] J. A. Dix, E. F. Hom, and A. S. Verkman, "Fluorescence correlation spectroscopy simulations of photophysical phenomena and molecular interactions: A molecular dynamics/monte carlo approach," *Journal of Physical Chemistry B*, vol. 110, pp. 1896–1906, 2 2006. [Online]. Available: /pmc/articles/PMC3129548/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3129548/

[91] S. H. Northrup and H. P. Erickson, "Kinetics of protein-protein association explained by brownian dynamics computer simulation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, pp. 3338–3342, 1992. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/1565624/ https://pubmed.ncbi.nlm.nih.gov/1565624/?dopt=Abstract

[92] U. Meseth, T. Wohland, R. Rigler, and H. Vogel, "Resolution of fluorescence correlation measurements," *Biophysical Journal*, vol. 76, pp. 16:19–1631, 3 1999.

[93] M. Schlosshauer and D. Baker, "Realistic protein-protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness," *Protein Science*, vol. 13, pp. 1660–1669, 6 2004. [Online]. Available: /pmc/articles/PMC2279981/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2279981/

[94] A. L. Papish, L. W. Tari, and H. J. Vogel, "Dynamic light scattering study of calmodulin-target peptide complexes," *Biophysical Journal*, vol. 83, pp. 1455–1464, 9 2002.

[95] A. R. Quintana, D. Wang, J. E. Forbes, and M. N. Waxham, "Kinetics of calmodulin binding to calcineurin," *Biochemical and Biophysical Research Communications*, vol. 334, pp. 674–680, 8 2005. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/16009337/

[96] M. J. Hubbard and C. B. Klees, "Calmodulin binding by calcineurin ligand-induced renaturation of protein immobilized on nitrocellulose*," *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, vol. 262, pp. 15 062–15 070, 1987.

[97] E. Takano, M. Hatanaka, and M. Maki, "Real-time analysis of the calcium-dependent interaction between calmodulin and a synthetic oligopeptide of calcineurin by a surface plasmon resonance biosensor," *FEBS Letters*, vol. 352, pp. 247–250, 9 1994. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/7925982/

[98] J. B. Sperry, R. Y. Huang, M. M. Zhu, D. L. Rempel, and M. L. Gross, "Hydrophobic peptides affect binding of calmodulin and ca2+ as explored by h/d amide exchange and mass spectrometry," *International Journal of Mass Spectrometry*, vol. 302, pp. 85–92, 4 2011. [Online]. Available: /pmc/articles/PMC3134286/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3134286/

[99] C. B. Klee, T. H. Crouch, and M. H. Krinks, "Calcineurin: A calcium- and calmodulin-binding protein of the nervous system," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 76, pp. 6270–6273, 1979. [Online]. Available: /pmc/articles/PMC411845/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC411845/

[100] "Hydrodynamic radius calculator." [Online]. Available: https://www.fluidic.com/resources/Toolkit/hydrodynamic-radius-Converter/

[101] S. Calve, A. Witten, A. Ocken, and T. Kinzer-Ursem, "Incorporation of non-canonical amino acids into the developing murine proteome," *Scientific Reports*, vol. 6, pp. 1–7,.

[102] D. Dieterich, "In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons," *Nat Neurosci*, vol. 13, no. 7, pp. 897–905,.

[103] D. McClatchy, "Pulsed azidohomoalanine labeling in mammals (palm) detects changes in liver-specific lkb1 knockout mice," *J Proteome Res*, vol. 14, no. 11, pp. 4815–22,.

[104] B. Alvarez-Castelao, "Cell-type-specific metabolic labeling of nascent proteomes in vivo," *Nature Biotechnology*, vol. 35, pp. 1196–1201,.

[105] D. McClatchy, Y. Ma, D. Liem, D. Ng, P. Ping, and J. Yates, "Quantitative temporal analysis of protein dynamics in cardiac remodeling," *Journal of Molecular and Cellular Cardiology*, vol. 121, pp. 163–172,.

[106] H. Evans, L. Bodea, and J. Gotz, "Cell-specific non-canonical amino acid labelling identifies changes in the de novo proteome during memory formation," *Elife*, vol. 9.

[107] C. R. Kirman, S. M. Hays, L. L. Aylward, M. Suh, M. A. Harris, C. M. Thompson, L. C. Haws, and D. M. Proctor, "Physiologically based pharmacokinetic model for rats and mice orally exposed to chromium," *Chemico-Biological Interactions*, vol. 200, pp. 45–64, 10 2012.

[108] R. P. Brown, M. D. Delp, S. L. Lindstedt, L. R. Rhomberg, and R. P. Beliles, "Physiological parameter values for physiologically based pharmacokinetic models," *Toxicology and Industrial Health*, vol. 13, pp. 407–484, 7 1997. [Online]. Available: http://journals.sagepub.com/doi/10.1177/074823379701300401

[109] C. A. Boswell, E. E. Mundo, S. Ulufatu, D. Bumbaca, H. S. Cahaya, N. Majidy, M. V. Hoy, M. G. Schweiger, P. J. Fielder, S. Prabhu, and L. A. Khawli, "Comparative physiology of mice and rats: Radiometric measurement of vascular parameters in rodent tissues," *Molecular Pharmaceutics*, vol. 11, 2014.

[110] J. U. Streif, K.-H. Hiller, C. Waller, M. Nahrendorf, F. Wiesmann, W. R. Bauer, E. Rommel, and A. Haase, "In vivo assessment of absolute perfusion in the murine skeletal muscle with spin labeling mri," *Journal of Magnetic Resonance Imaging*, vol. 17, pp. 147–152, 1 2003. [Online]. Available: http://doi.wiley.com/10.1002/jmri.10229

[111] C. Smith, "Metlin: a metabolite mass spectral database," *Ther Drug Monit*, vol. 27, no. 6, pp. 747–51,.

[112] J. Chong, D. Wishart, and J. Xia, "Using metaboanalyst 4.0 for comprehensive and integrative metabolomics data analysis," *Curr Protoc Bioinformatics*, vol. 68, no. 1, pp. 86,.

[113] T. L. Kinzer-Ursem and J. J. Linderman, "Both ligand- and cell-specific parameters control ligand agonism in a kinetic model of g protein–coupled receptor signaling," *PLoS Computational Biology*, vol. 3, p. e6, 1 2007. [Online]. Available: https://dx.plos.org/10.1371/journal.pcbi.0030006

[114] K. Kiick, E. Saxon, D. Tirrell, and C. Bertozzi, "Incorporation of azides into recombinant proteins for chemoselective modification by the staudinger ligation," in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 19–24,.

[115] J. R. Wiśniewski, M. Y. Hein, J. Cox, and M. Mann, "A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards," *Molecular and Cellular Proteomics*, vol. 13, pp. 3497–3506, 12 2014. [Online]. Available: /pmc/articles/PMC4256500/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4256500/

[116] M. Piques, W. X. Schulze, M. Höhne, B. Usadel, Y. Gibon, J. Rohwer, and M. Stitt, "Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in arabidopsis," *Molecular Systems Biology*, vol. 5, 1 2009. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/:19888:209/ https://pubmed.ncbi.nlm.nih.gov/:19888:209/?dopt=Abstract

[117] A. Moura, M. A. Savageau, and R. Alves, "Relative amino acid composition signatures of organisms and environments," *PLoS ONE*, vol. 8, 10 2013. [Online]. Available: /pmc/articles/PMC3808408/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3808408/

[118] D. McClatchy, M. Dong, C. Wu, J. Venable, and J. Yates, "15n metabolic labeling of mammalian tissue with slow protein turnover," *Journal of Proteome Research*, vol. 6, pp. 2005–2010,.

[119] J. Price, S. Guan, A. Burlingame, S. Prusiner, and S. Ghaemmaghami, "Analysis of proteome dynamics in the mouse brain," in *Proceedings of the National Academy of Sciences*, vol. 107, pp. 14 508–14 513,.

[120] D. Hammond, "Proteome dynamics: Tissue variation in the kinetics of proteostasis in intact animals," *Mol Cell Proteomics*, vol. 15, no. 4, pp. 1204–19,.

[121] R. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed. OTexts: Melbourne, Australia., 2018. [Online]. Available: https://otexts.com/fpp2/

[122] S. Pepke, T. Kinzer-Ursem, S. Mihalas, and M. B. Kennedy, "A dynamic model of interactions of ca2+, calmodulin, and catalytic subunits of ca2+/calmodulin-dependent protein kinase ii," *PLoS Computational Biology*, vol. 6, 2 2010. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/20168991/

[123] D. Chowdhury, M. Turner, T. Patriarchi, A. C. Hergarden, D. Anderson, Y. Zhang, J. Sun, C. Chen, J. B. Ames, and J. W. Hell, "Ca 2+ /calmodulin binding to psd -95 mediates homeostatic synaptic scaling down," *The EMBO Journal*, vol. 37, pp. 122–138, 1 2018. [Online]. Available: /pmc/articles/PMC5753037/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5753037/

[124] Y. Kubota, J. A. Putkey, and M. N. Waxham, "Neurogranin controls the spatiotemporal pattern of postsynaptic ca 2+/cam signaling," *Biophysical Journal*, vol. 93, pp. 3848–3859, 12 2007. [Online]. Available: /pmc/articles/PMC:2084249/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC:2084249/

[125] T. J. Amiss, D. B. Sherman, C. M. Nycz, S. A. Andaluz, and J. B. Pitner, "Engineering and rapid selection of a low-affinity glucose/galactose-binding protein for a glucose biosensor," *Protein Science*, vol. 16, pp. 2350–2359, 11 2007. [Online]. Available: /pmc/articles/PMC2211708/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2211708/

[126] M. N. Vyas, N. K. Vyas, and F. A. Quiocho, "Crystallographic analysis of the epimeric and anomeric specificity of the periplasmic transport/chemosensory protein receptor for d-glucose and," pp. 4762–4768, 1994. [Online]. Available: https://pubs.acs.org/sharingguidelines

A. APPENDIX: CODE

A.1 FRCS Scripts and GUI

https://github.itap.purdue.edu/TamaraKinzerursemGroup/FRCS

A.2 Aha Models and Scripts

https://github.itap.purdue.edu/TamaraKinzerursemGroup/ncAABiokinetics

B. APPENDIX: EQUATIONS

B.1 System of ODEs for Aha Model

The following equations (eqs. B.1-B.30) are separated into functional components and should be evaluated in order to appropriately determine the rate of change of Aha in each compartment for each time step.

Dose Forcing Functions:

$$(fAha)_{subcu} = (fAha)_{subcu} + F(t)_{subcu} \tag{B.1}$$

$$(fAha)_{gasin} = (fAha)_{gasin} + F(t)_{gasin} \tag{B.2}$$

F(t) is typically either a cumulative step function (for injections or dietary bolus) or a continuous accumulation function (IV or watering). The model tracks the depleted pool from each dosing site ($(fAha)_{subcu}$ for injections, $(fAha)_{gasin}$ for dietary) and uses this depletion and the cumulative forcing function to determine the actual Aha mass present in each site (without changing the nomenclature) for the given time step.

Blood Flow:

$$\frac{d(fAha_p)_{skmus}}{dt} = qb_{skmus}\left(\frac{(fAha_p)_{sysrv}}{mb_{sysrv}} - \frac{(fAha_p)_{skmus}}{mb_{skmus}}\right) \tag{B.3}$$

$$\frac{d(fAha_p)_{liver}}{dt} = qb_{hepar}\left(\frac{(fAha_p)_{sysrv}}{mb_{sysrv}} - \frac{(fAha_p)_{liver}}{mb_{liver}}\right) \tag{B.4}$$

$$\frac{d(fAha_p)_{brain}}{dt} = qb_{brain}\left(\frac{(fAha_p)_{sysrv}}{mb_{sysrv}} - \frac{(fAha_p)_{brain}}{mb_{brain}}\right) \tag{B.5}$$

$$\frac{d(fAha_p)_{renal}}{dt} = qb_{renal}\left(\frac{(fAha_p)_{sysrv}}{mb_{sysrv}} - \frac{(fAha_p)_{renal}}{mb_{renal}}\right) \tag{B.6}$$

$$\frac{d(fAha_p)_{smint}}{dt} = qb_{smint}\left(\frac{(fAha_p)_{sysrv}}{mb_{sysrv}} - \frac{(fAha_p)_{smint}}{mb_{smint}}\right) \tag{B.7}$$

$$\frac{d(fAha_p)_{sysrv}}{dt} = qb_{smint}\left(\frac{(fAha_p)_{liver}}{mb_{liver}} - \frac{(fAha_p)_{sysrv}}{mb_{sysrv}}\right) - \sum_x \frac{d(fAha_p)_x}{dt} \tag{B.8}$$

$$\frac{d(fAha_p)_{liver}}{dt} \mathrel{+}= qb_{smint}\left(\frac{(fAha_p)_{smint}}{mb_{smint}} - \frac{(fAha_p)_{liver}}{mb_{liver}}\right) \tag{B.9}$$

Subcutaneous Dose Absorption:

$$\frac{d(fAha_p)_{sysrv}}{dt} += ka_{subcu}(fAha)_{subcu} \tag{B.10}$$

$$\frac{d(fAha)_{subcu}}{dt} = -ka_{subcu}(fAha)_{subcu} \tag{B.11}$$

Oral/Dietary Gastrointestinal Dose Absorption:

$$\frac{d(fAha_p)_{smint}}{dt} += ke_{gasin}(fAha)_{gasin} - ki_{gasin}(fAha_p)_{smint} \tag{B.12}$$

$$\frac{d(fAha)_{gasin}}{dt} = ki_{gasin}(fAha_p)_{smint} - ke_{gasin}(fAha)_{gasin} \tag{B.13}$$

Even if the function $F(t)_{gasin}$ is zero for all time values, this set of equations enables excretion and is included in the model.

Tissue Exchange:

$$\frac{d(fAha_t)_{skmus}}{dt} = ki_{skmus}(fAha_p)_{skmus} - ke_{skmus}(fAha_t)_{skmus} \tag{B.14}$$

$$\frac{d(fAha_t)_{liver}}{dt} = ki_{liver}(fAha_p)_{liver} - ke_{liver}(fAha_t)_{liver} \tag{B.15}$$

$$\frac{d(fAha_t)_{brain}}{dt} = ki_{brain}(fAha_p)_{brain} - ke_{brain}(fAha_t)_{brain} \tag{B.16}$$

$$\frac{d(fAha_t)_{renal}}{dt} = ki_{renal}(fAha_p)_{renal} - ke_{renal}(fAha_t)_{renal} \tag{B.17}$$

$$\frac{d(fAha_p)_{skmus}}{dt} -= \frac{d(fAha_t)_{skmus}}{dt} \tag{B.18}$$

$$\frac{d(fAha_p)_{liver}}{dt} -= \frac{d(fAha_t)_{liver}}{dt} \tag{B.19}$$

$$\frac{d(fAha_p)_{brain}}{dt} -= \frac{d(fAha_t)_{brain}}{dt} \tag{B.20}$$

$$\frac{d(fAha_p)_{renal}}{dt} -= \frac{d(fAha_t)_{renal}}{dt} \tag{B.21}$$

Proteome Incorporation:

$$\frac{d(rF)_{skmus}}{dt} = kf * ks_{skmus}\frac{(fAha_t)_{skmus}}{mt_{skmus}} - kd_{skmus}(rF)_{skmus} \tag{B.22}$$

$$\frac{d(rF)_{liver}}{dt} = kf * ks_{liver}\frac{(fAha_t)_{liver}}{mt_{liver}} - kd_{liver}(rF)_{liver} \tag{B.23}$$

$$\frac{d(rF)_{brain}}{dt} = kf * ks_{brain}\frac{(fAha_t)_{brain}}{mt_{brain}} - kd_{brain}(rF)_{brain} \tag{B.24}$$

$$\frac{d(rF)_{renal}}{dt} = kf * ks_{renal}\frac{(fAha_t)_{renal}}{mt_{renal}} - kd_{renal}(rF)_{renal} \tag{B.25}$$

$$\tag{B.26}$$

Aha Removal:

$$\frac{d(fAha_t)_{liver}}{dt} -= kr_{liver}(fAha_t)_{liver} \tag{B.27}$$

$$\frac{d(fAha_p)_{renal}}{dt} -= kr_{renal}(fAha_p)_{renal} \tag{B.28}$$

$$\frac{d(fAha)_{gasin}}{dt} -= kr_{colon}(fAha)_{gasin} \tag{B.29}$$

$$\frac{d(fAha)_{cumrm}}{dt} = kr_{liver}(fAha_t)_{liver} + kr_{renal}(fAha_p)_{renal} + kr_{colon}(fAha)_{gasin} \tag{B.30}$$

B.2 All Listed Equations

List of equations

$$S + E \rightleftharpoons ES \rightarrow EP \rightleftharpoons E + P \tag{1.1}$$

$$S + E \rightleftharpoons ES \rightarrow E + P \tag{1.2}$$

$$\frac{d[S]}{dt} = k_d[ES] - k_a[E][S] \tag{1.3a}$$
$$\frac{d[E]}{dt} = (k_d + k_{cat})[ES] - k_a[E][S] \tag{1.3b}$$
$$\frac{d[ES]}{dt} = k_a[E][S] - (k_d + k_{cat})[ES] \tag{1.3c}$$
$$\frac{d[P]}{dt} = k_{cat}[ES] \tag{1.3d}$$

$$\frac{d[P]}{dt} = V_{max}\frac{[S]}{K_M + [S]} = k_{cat}[E]_0\frac{[S]}{K_M + [S]} \tag{1.4}$$

$$K_D = \frac{[E]_{eq}[S]_{eq}}{[ES]_{eq}} = \frac{1}{K_{eq}} = \frac{k_d}{k_a} \tag{1.5a}$$
$$K_M = \frac{k_d + k_{cat}}{k_a} \tag{1.5b}$$

$$S + E \rightleftharpoons ES \rightarrow E + P \tag{1.2}$$

$$I + E \rightleftharpoons EI \tag{1.6}$$

$$\frac{d[S]}{dt} = k_d[ES] - k_a[E][S] \tag{1.7a}$$
$$\frac{d[I]}{dt} = k_{d,I}[EI] - k_{a,I}[E][I] \tag{1.7b}$$
$$\frac{d[EI]}{dt} = k_{a,I}[E][I] - k_{d,I}[EI] \tag{1.7c}$$
$$\frac{d[E]}{dt} = (k_d + k_{cat})[ES] + k_{d,I}[EI] - k_a[E]([S] + [I]) \tag{1.7d}$$
$$\frac{d[ES]}{dt} = k_a[E][S] - (k_d + k_{cat})[ES] \tag{1.7e}$$
$$\frac{d[P]}{dt} = k_{cat}[ES] \tag{1.7f}$$

$$\frac{d[P]}{dt} = k_{cat}[E]_0\frac{[S]}{K_{M,I} + [S]} \tag{1.8}$$
$$K_{M,I} = K_M\left(\frac{1 + [I]}{k_{d,I}}\right) \tag{1.9}$$

$$F_L = q(E + u(B)) \tag{1.10}$$

$$F_N = m(u') \tag{1.11}$$

$$m/q = u'^{-}1(E + u(B)) \tag{1.12}$$

$$D_c = \frac{k_b * T}{6\pi\eta R_H} \tag{2.1}$$

$$PSF(r, z) = I_0 e^{-2r^2/\omega_{xy}^2} e^{-2z^2/\omega_z^2} \tag{2.2}$$

$$G(\tau) = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I(t)\rangle\langle I(t+\tau)\rangle} \tag{2.3a}$$

$$= \frac{\langle I(t)I(t+\tau)\rangle}{\langle I(t)\rangle^2} \tag{2.3b}$$

$$D_c = \frac{\omega_{xy}^2}{4\tau_D} \tag{2.4}$$

$$G(\tau) = G(0)M(\tau) + G(\infty) \tag{2.5a}$$

$$M(\tau) = \frac{1}{(1 + (\tau/\tau_D))(1 + \alpha^{-2}(\tau/\tau_D))^{1/2}} \tag{2.5b}$$

$$G(0) = \frac{1}{\langle P\rangle} = \frac{1}{C(V_{eff})} \tag{2.6}$$

$$V_{eff} = \pi^{3/2}\omega_{xy}^2\omega_z \tag{2.7}$$

$$M(\tau) = \sum_i \frac{F_i}{(1 + (\tau/\tau_{D,i}))(1 + \alpha^{-2}(\tau/\tau_{D,i}))^{1/2}} \tag{2.8}$$

$$M(\tau) = \frac{1}{(\sum_i F_i\epsilon_i)^2} \sum_i \frac{F_i\epsilon_i^2}{(1 + (\tau/\tau_{D,i}))(1 + \alpha^{-2}(\tau/\tau_{D,i}))^{1/2}} \tag{2.9}$$

$$G(\tau) = G(0)\left(1 + \frac{T}{1-T}e^{-\tau/\tau_R}\right)M(\tau) + G(\infty) \tag{2.10a}$$

$$= G(0)\left(1 + Se^{-\tau/\tau_S}\right)M(\tau) + G(\infty) \tag{2.10b}$$

$$G_X(\tau) = \frac{\langle I_1(t) I_2(t+\tau)\rangle}{\langle I_1(t) I_2(t)\rangle} \tag{2.11}$$

$$V_{eff,12} = \frac{\pi^{3/2}}{2^{3/2}} \left(\omega_{xy,1}^2 + \omega_{xy,2}^2\right) \left(\omega_{z,1}^2 + \omega_{z,2}^2\right)^{1/2} \tag{2.12}$$

$$G(\tau) = G(0) e^{-\tau/\tau_B} \tag{2.13}$$

$$G(0) = \frac{1}{\langle N\rangle} \frac{k_a}{k_D} = \frac{1}{\langle N\rangle K_D} \tag{2.14}$$

$$\tau_B = \frac{1}{k_a + k_d} \tag{2.15}$$

$$G(\tau) = \frac{\langle I(t) I(t+\tau)\rangle}{\langle I(t)\rangle \langle I(t+\tau)\rangle} \tag{2.3}$$

$$\hat{G}[k] = \frac{\langle \hat{I}[j] \hat{I}[j+k]\rangle}{\langle \hat{I}[j]\rangle \langle \hat{I}[j+k]\rangle} \tag{2.16}$$

$$\langle \hat{I}[j] \hat{I}[j+k]\rangle = \frac{1}{J-k} \sum_{j=0}^{J-k} \hat{I}[j] \hat{I}[j+k] \tag{2.17}$$

$$\langle \hat{I}[j]\rangle = \frac{1}{J-k} \sum_{j=0}^{J-k} \hat{I}[j] \tag{2.18}$$

$$\langle \hat{I}[j+k]\rangle = \frac{1}{J-k} \sum_{j=k}^{J} \hat{I} I[j] \tag{2.19}$$

$$fA + B \rightleftharpoons fAB \tag{2.20}$$

$$\frac{d[fA]}{dt} = k_d[fAB] - k_a[fA][B] \tag{2.21}$$

$$\frac{d[B]}{dt} = k_d[fAB] - k_a[fA][B] \tag{2.22}$$

$$\frac{d[fAB]}{dt} = k_a[fA][B] - k_d[fAB] \tag{2.23}$$

$$F_A(t) = \frac{[fA](t)}{[fA]_0} \tag{2.24}$$

$$F_{AB}(t) = 1 - F_A \tag{2.25}$$

$$\hat{G}[\tau] = \frac{B \sum_i^{B_x} \sum_j^{B_y} p_x[i] p_y[j] \delta[\tau - (t_x[i] - t_y[j])]}{P_x P_y} \tag{2.26}$$

$$G(\tau) = G(0)(1 + Se^{-\tau/\tau_S}) \left(\sum_{i=1}^{2} \frac{F_i}{(1 + (\tau/\tau_{D,i}))(1 + \alpha^{-2}(\tau/\tau_{D,i}))^{1/2}} \right) + G(\infty) \tag{2.27}$$

$$\omega_{xy}^2 = 4D_{c,cal}\tau_{D,cal} \tag{2.28}$$

$$V_{eff} = \frac{P_{cal}}{C_{cal}} \tag{2.29}$$

$$= \pi^{3/2}\omega_{xy}^2\omega_z \tag{2.30}$$

$$= \pi^{3/2}\omega_{xy}^3\alpha \tag{2.31}$$

$$\alpha = \frac{P_{cal}}{C_{cal}\pi^{3/2}\omega_{xy}^3} \tag{2.32}$$

$$= \frac{P_{cal}}{C_{cal}\pi^{3/2}(4D_{c,cal}\tau_{D,cal})^{3/2}} \tag{2.33}$$

$$[A] = [fA] + [bA] \tag{2.34}$$

$$[fA](t) = ([fA]_0 - bM_A)e^{-t(bR_A)} + bM_A \tag{2.35}$$

$$[bA](t) = ([fA]_0 - bM_A)(1 - e^{-t(bR_A)}) \tag{2.36}$$

$$bR_{AB} = bR_A \frac{\tau_{D,AB}}{\tau_{D,A}} \tag{2.37}$$

$$bM_{AB} = bM_A \frac{\tau_{D,A}}{\tau_{D,AB}} \tag{2.38}$$

$$[fA](t) = [A](t) - [bA](t) \tag{2.39}$$

$$[fAB](t) = [AB](t) - [bAB](t) \tag{2.40}$$

$$F_A(t) = [fA](t)/([fA](t) + [fAB](t)) \tag{2.41}$$

$$F_{AB}(t) = [fAB](t)/([fA](t) + [fAB](t)) \tag{2.42}$$

$$\frac{d[A](t)}{dt} = k_d[AB](t) - k_a[A](t)[B](t) \tag{2.43}$$

$$\frac{d[B](t)}{dt} = k_d[AB](t) - k_a[A](t)[B](t) \tag{2.44}$$

$$\frac{d[AB](t)}{dt} = k_a[A][B](t) - k_d[AB](t) \tag{2.45}$$

$$\frac{d[bA](t)}{dt} = ([fA](t) - bM_A F_A(t))bR_A \tag{2.46}$$

$$\frac{d[bAB](t)}{dt} = ([fAB](t) - bM_{AB} F_{AB}(t))bR_{AB} \tag{2.47}$$

$$fA + B \rightleftharpoons fAB \tag{2.20}$$

$$PSF[t,i] = \frac{-1}{2\omega_{xy}^2}\left((X_i - X\_1)^2 + (Y_i - Y_0)^2 + \frac{(Z_i - Z_0)^2}{\alpha^2}\right) \tag{3.1a}$$

$$P_{ex}[t,i] = P_{abs} e^{K_{pos}[t,i]/\omega_{xy}^2} \tag{3.1b}$$

$$P_{em}[t,i] = P_{ex}[t,i] P_{gs} P_{conv} Q_f Q_{det} \tag{3.1c}$$

$$I[t] = \sum_i H[P_{em}[t,i] - rand(0,1)] \tag{3.1d}$$

$$t_{rxn,f} = t_{m,f} + \lambda_f \Delta t_f \tag{3.2}$$

$$\left(\frac{d[fAha_p]_x}{dt}\right)_{transport} = \frac{Q_x}{V_x}\left([fAha_p]_{sysrv} - [fAha_p]_x\right) \tag{4.1}$$

$$\left(\frac{d[fAha_p]_x}{dt}\right)_{exchange} = k_{e,x}[fAha_t]_x - k_{i,x}[fAha_p]_x \tag{4.2}$$

$$\left(\frac{d[fAha_t]_x}{dt}\right)_{exchange} = k_{i,x}[fAha_p]_x - k_{e,x}[fAha_t]_x \tag{4.3}$$

$$[fAha_t]_x \gg [pAha_t]_x \tag{4.4}$$

$$\left(\frac{d[fAha_t]_x}{dt}\right)_{synthesis} = -k_{s,x}[fAha_t]_x \approx 0 \tag{4.5}$$

$$\left(\frac{d[pAha_t]_x}{dt}\right)_{synthesis} = k_{s,x}[fAha_t]_x - k_{d,x}[pAha_t]_x \tag{4.6}$$

$$rF_x = \frac{(signal - background)}{(background)} = k_f[pAha_t]_x \tag{4.7}$$

$$\frac{d(rF_x)}{dt} = k_f\frac{d[pAha_t]_x}{dt} = k_f k_s[fAha_t]_x - k_d(rF_x) \tag{4.8}$$

$$PI = \hat{y} \pm 1.96\frac{\sum_i(y_i - \hat{y}_i)^2}{n}\sqrt{1 + \frac{1}{n} + \frac{(t-\bar{t})^2}{\sum_i(t_i - \bar{t})^2}} \tag{4.9}$$

$$(fAha)_{subcu} = (fAha)_{subcu} + F(t)_{subcu} \tag{B.1}$$

$$(fAha)_{gasin} = (fAha)_{gasin} + F(t)_{gasin} \tag{B.2}$$

$$\frac{d(fAha_p)_{skmus}}{dt} = qb_{skmus}\left(\frac{(fAha_p)_{sysrv}}{mb_{sysrv}} - \frac{(fAha_p)_{skmus}}{mb_{skmus}}\right) \tag{B.3}$$

$$\frac{d(fAha_p)_{liver}}{dt} = qb_{hepar}\left(\frac{(fAha_p)_{sysrv}}{mb_{sysrv}} - \frac{(fAha_p)_{liver}}{mb_{liver}}\right) \tag{B.4}$$

$$\frac{d(fAha_p)_{brain}}{dt} = qb_{brain}\left(\frac{(fAha_p)_{sysrv}}{mb_{sysrv}} - \frac{(fAha_p)_{brain}}{mb_{brain}}\right) \tag{B.5}$$

$$\frac{d(fAha_p)_{renal}}{dt} = qb_{renal}\left(\frac{(fAha_p)_{sysrv}}{mb_{sysrv}} - \frac{(fAha_p)_{renal}}{mb_{renal}}\right) \tag{B.6}$$

$$\frac{d(fAha_p)_{smint}}{dt} = qb_{smint}\left(\frac{(fAha_p)_{sysrv}}{mb_{sysrv}} - \frac{(fAha_p)_{smint}}{mb_{smint}}\right) \tag{B.7}$$

$$\frac{d(fAha_p)_{sysrv}}{dt} = qb_{smint}\left(\frac{(fAha_p)_{liver}}{mb_{liver}} - \frac{(fAha_p)_{sysrv}}{mb_{sysrv}}\right) - \sum_x \frac{d(fAha_p)_x}{dt} \tag{B.8}$$

$$\frac{d(fAha_p)_{liver}}{dt} \mathrel{+}= qb_{smint}\left(\frac{(fAha_p)_{smint}}{mb_{smint}} - \frac{(fAha_p)_{liver}}{mb_{liver}}\right) \tag{B.9}$$

$$\frac{d(fAha_p)_{sysrv}}{dt} \mathrel{+}= ka_{subcu}(fAha)_{subcu} \tag{B.10}$$

$$\frac{d(fAha)_{subcu}}{dt} = -ka_{subcu}(fAha)_{subcu} \tag{B.11}$$

$$\frac{d(fAha_p)_{smint}}{dt} += ke_{gasin}(fAha)_{gasin} - ki_{gasin}(fAha_p)_{smint} \tag{B.12}$$

$$\frac{d(fAha)_{gasin}}{dt} = ki_{gasin}(fAha_p)_{smint} - ke_{gasin}(fAha)_{gasin} \tag{B.13}$$

$$\frac{d(fAha_t)_{skmus}}{dt} = ki_{skmus}(fAha_p)_{skmus} - ke_{skmus}(fAha_t)_{skmus} \tag{B.14}$$

$$\frac{d(fAha_t)_{liver}}{dt} = ki_{liver}(fAha_p)_{liver} - ke_{liver}(fAha_t)_{liver} \tag{B.15}$$

$$\frac{d(fAha_t)_{brain}}{dt} = ki_{brain}(fAha_p)_{brain} - ke_{brain}(fAha_t)_{brain} \tag{B.16}$$

$$\frac{d(fAha_t)_{renal}}{dt} = ki_{renal}(fAha_p)_{renal} - ke_{renal}(fAha_t)_{renal} \tag{B.17}$$

$$\frac{d(fAha_p)_{skmus}}{dt} -= \frac{d(fAha_t)_{skmus}}{dt} \tag{B.18}$$

$$\frac{d(fAha_p)_{liver}}{dt} -= \frac{d(fAha_t)_{liver}}{dt} \tag{B.19}$$

$$\frac{d(fAha_p)_{brain}}{dt} -= \frac{d(fAha_t)_{brain}}{dt} \tag{B.20}$$

$$\frac{d(fAha_p)_{renal}}{dt} -= \frac{d(fAha_t)_{renal}}{dt} \tag{B.21}$$

$$\frac{d(rF)_{skmus}}{dt} = kf * ks_{skmus}\frac{(fAha_t)_{skmus}}{mt_{skmus}} - kd_{skmus}(rF)_{skmus} \tag{B.22}$$

$$\frac{d(rF)_{liver}}{dt} = kf * ks_{liver}\frac{(fAha_t)_{liver}}{mt_{liver}} - kd_{liver}(rF)_{liver} \tag{B.23}$$

$$\frac{d(rF)_{brain}}{dt} = kf * ks_{brain}\frac{(fAha_t)_{brain}}{mt_{brain}} - kd_{brain}(rF)_{brain} \tag{B.24}$$

$$\frac{d(rF)_{renal}}{dt} = kf * ks_{renal}\frac{(fAha_t)_{renal}}{mt_{renal}} - kd_{renal}(rF)_{renal} \tag{B.25}$$

$$\tag{B.26}$$

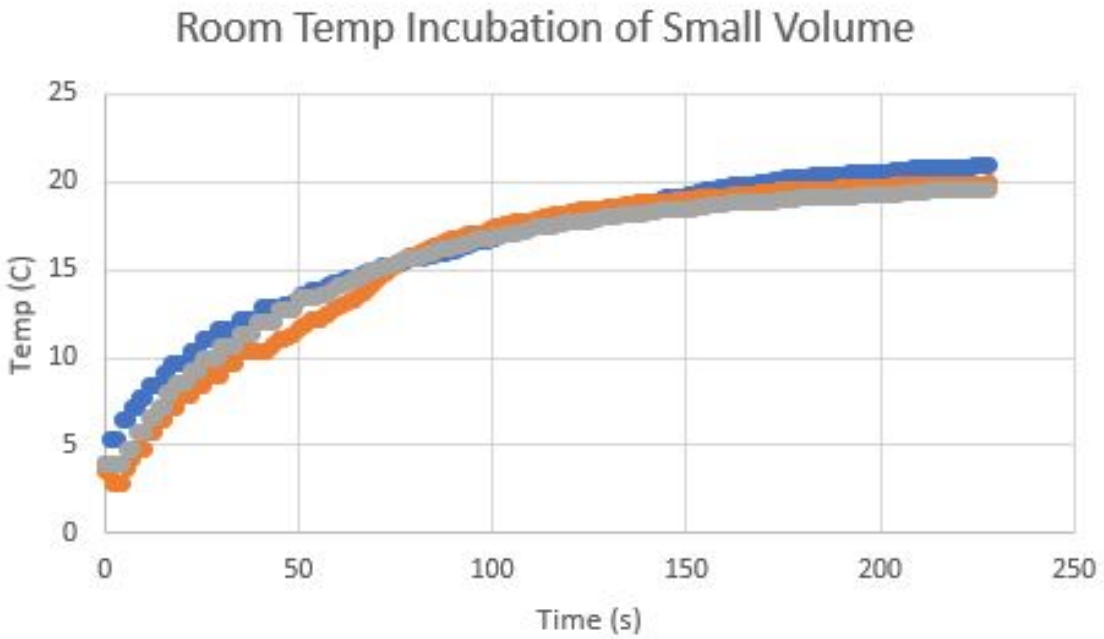
$$\frac{d(fAha_t)_{liver}}{dt} -= kr_{liver}(fAha_t)_{liver} \tag{B.27}$$

$$\frac{d(fAha_p)_{renal}}{dt} -= kr_{renal}(fAha_p)_{renal} \tag{B.28}$$

$$\frac{d(fAha)_{gasin}}{dt} -= kr_{colon}(fAha)_{gasin} \tag{B.29}$$

$$\frac{d(fAha)_{cumrm}}{dt} = kr_{liver}(fAha_t)_{liver} + kr_{renal}(fAha_p)_{renal} + kr_{colon}(fAha)_{gasin} \tag{B.30}$$

C. APPENDIX FIGURES



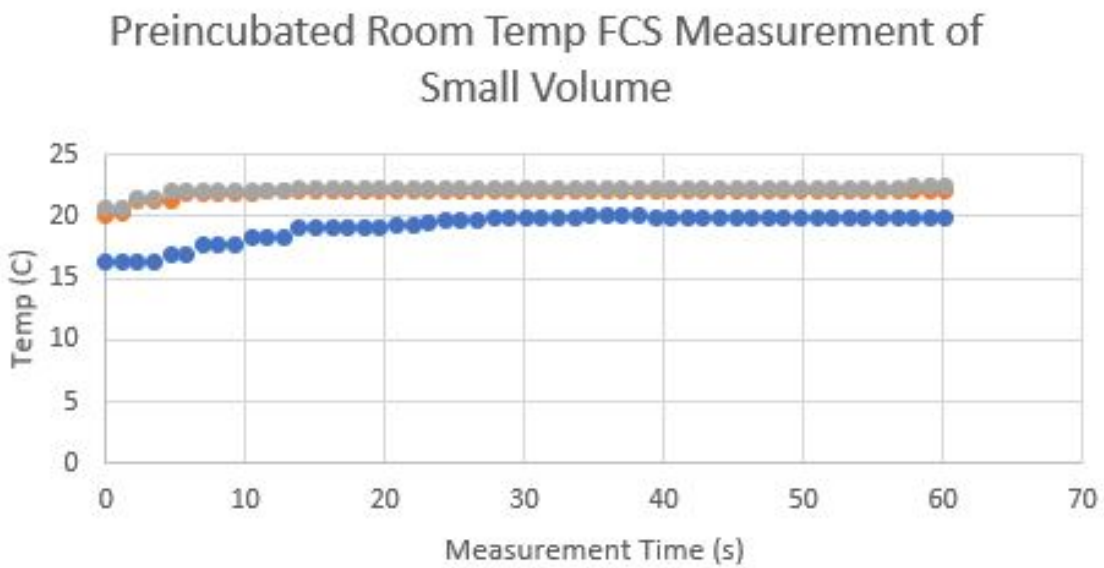


Fig. C.1. Temperature change of a small 15uL sample as it is moved first from an Ice Bath (4C) to RT, then measured with FCS under a 633 He-Ne laser at RT for 60 s. Temperature recorded using a thermocouple that samples at 1Hz.

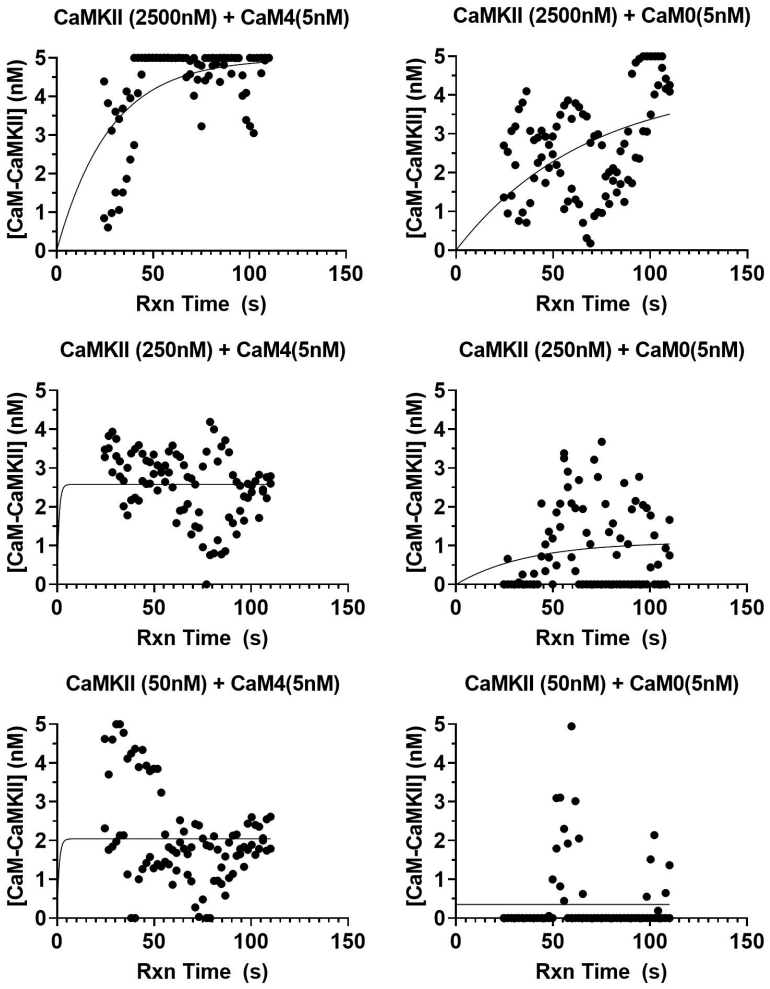


Fig. C.2. Binding of N3CaM-647 to various concentrations of mCaMKII in the presence and absence of Ca^{2+}. Binding reactions were measured in duplicate for 120 s and correlated into 50 frames of width 25 s, spaced linearly over the timecourse of measurement. Correlated frames were fit with according to calibrated measurements of the diffusion time of CaM0/4 (382/437 µs) and CaM-CaMKII (463µs). Notably, these calibrated diffusion times do not differ by a factor of least 1.6 as desired for FCS fitting analysis [26]. This may explain very noisy measurements that are not altogether firmly aligned with the expected fitting model behavior. All binding reactions performed in the same buffer (20 mM Hepes (pH 7.5), 100 mM NaCl, 1 mg/mL BSA, 1mM EGTA, 1mM EDTA, $< 2\%$ Glycerol) supplemented with either 12 mM Ca^{2+} or an additional 2 mM EGTA for CaM4 and CaM0 measurements respectively.

D. APPENDIX: TABLES

Table D.1.
FRCS simulation parameters, determined kinetics from FRCS fitting, and goodness of fit statistics for all studies discussed in chapter 3. Seeds 1-39 are represented in the plots shown in figure 3.5 and results from seed 100-110 are displayed in figures 3.2-3.4. $S_{F,x}$ is a standard deviation of the residuals of complex composition fraction (F_{AB}) from the best fit kinetic binding curve and is shown because R^2 is poor fitting statistic for non-linear regression.

	Simulation Conditions		**Binding Constants (Input → Best Fit Value) {%Δ}**		**Goodness of Fit**	
Seed(s)	**$[A]_0$ / $[B]_0$ (nM)**	**$D_{c,A}$ / $D_{c,B/AB}$ (m^2s^{-1})**	**k_a ($M^{-1}s^{-1}$)**	**K_D (nM)**	**$S_{F,t}$**	**R^2**
1-3	10 / 10	4.0e-10 / 3.0e-10	(5.0 → 8.33)e6 {+66.6%}	1.0 → 0.59 {-41.0%}	0.103	.58
4-6	10 / 10	4.0e-10 / 2.0e-10	(5.0 → 5.82)e6 {+16.4%}	1.0 → 1.29 {+29.0%}	0.090	.56
7-9	10 / 10	4.0e-10 / 1.0e-10	(5.0 → 3.87)e6 {-22.6%}	1.0 → 1.69 {+69.0%}	0.086	.73
11-13	10 / 5	4.0e-10 / 2.0e-10	(5.0 → 5.50)e6 {+10.0%}	1.0 → 1.10 {+10.0%}	0.073	.49
14-16	10 / 10	4.0e-10 / 2.0e-10	(5.0 → 4.07)e6 {-18.6%}	1.0 → 2.09 {+109.0%}	0.099	.55
17-19	10 / 20	4.0e-10 / 2.0e-10	(5.0 → 4.89)e6 {-2.2%}	1.0 → 0.68 {-32.0%}	0.066	.76
21-23	10 / 10	4.0e-10 / 2.0e-10	(5.0 → 7.53)e7 {+50.3%}	0.10 → 0.064 {-36.0%}	0.045	.40
24-26	10 / 10	4.0e-10 / 2.0e-10	(5.0 → 5.13)e6 {+2.6%}	1.0 → 1.75 {+75.0%}	0.116	.41
27-29	10 / 10	4.0e-10 / 2.0e-10	(5.0 → 5.46)e5 {+9.2%}	10.0 → 52.8 {+428.0%}	0.045	.27
31-33	10 / 10	4.0e-10 / 2.0e-10	(5.0 → 0.38)e8 {-92.4%}	1.0 → 0.42 {-58.0%}	0.042	.09
34-36	10 / 10	4.0e-10 / 2.0e-10	(5.0 → 3.63)e7 {-27.4%}	1.0 → 0.37 {-63.0%}	0.056	.16
37-39	10 / 10	4.0e-10 / 2.0e-10	(5.0 → 4.89)e6 {-2.2%}	1.0 → 1.14 {+14.0%}	0.069	.75
101-105	10 / 10	1.5e-10 / 0.75e-10	(2.0 → 1.71)e6 {-14.5%}	25 → 40.9 {+63.8%}	.035	.73
106-110	10 / 0 → 10 @ t=30	1.5e-10 / 0.75e-10	(2.0 → 1.94)e6 {-3.0%}	25 → 29.17 {+16.7%}	.050	.31

Table D.2.
Distribution model fitting parameters including best fit values, fitting ranges, standard error of fit, PRCC values and informing literature. Parameters without published literature values are sourced from this study (TS) either from an experimental measurement or best estimate initial value that produced a single time-step Δ[Aha] that was one order of magnitude lower than the maximum recorded experimental value and swept upward and downward 1.5 orders of magnitude for fitting.

Name	Best Fit Value	Units	Fitting Range	Source(s)	Stderr of Fit	PRCC
kiBrain	2.55E+1	min^{-1}	(5.0e-1, 5.0e+2)	TS	1.37E+2	1.06E-2
kiLiver	2.34E+1	min^{-1}	(5.0e-1, 5.0e+2)	TS	7.13E+1	-2.15E-1
kiRenal	6.94E+0	min^{-1}	(5.0e-1, 5.0e+2)	TS	1.66E+1	-7.27E-2
kiSkmus	6.64E+1	min^{-1}	(5.0e-1, 5.0e+2)	TS	1.57E+2	1.92E-2
keBrain	4.08E-1	min^{-1}	(1.0e-2, 1.0e+1)	TS	2.12E+0	-2.43E-1
keLiver	8.60E-1	min^{-1}	(1.0e-2, 1.0e+1)	TS	1.53E+0	-1.19E-1
keRenal	1.23E+0	min^{-1}	(1.0e-2, 1.0e+1)	TS	3.18E+0	-2.29E-1
keSkmus	3.32E-1	min^{-1}	(1.0e-2, 1.0e+1)	TS	7.65E-1	-9.64E-2
krColon	2.88E-1	min^{-1}	(3.3e-3, 3.3e-1)	[107]	5.96E+0	4.94E-2
krLiver	2.29E-1	min^{-1}	(1.0e-2,1.0e+0)	TS	1.61E-1	-2.14E-1
krRenal	1.38E-1	min^{-1}	(1.0e-2, 1.0e+0)	[107]	2.71E+0	-1.63E-1
kiGasin	5.67E-2	min^{-1}	(1.0e-3, 1.0e-1)	[107]	4.49E-2	-1.08E-1
kaSubcu	6.18E-3	min^{-1}	(1.0e-3, 1.0e-1)	TS, [108]	9.18E-3	2.16E-2
keGasin	3.08E-2	min^{-1}	(5.0e-3, 5.0e-1)	[107]	1.88E-1	-1.08E-1
qbBrain	3.69E-1	g/min	(8.8E-3,8.8E-1)	[108,109]	2.02E+0	8.03E-2
qbHepar	1.69E-1	g/min	(6.8E-3,6.8E-1)	[108,109]	1.99E+0	9.34E-2
qbRenal	4.32E-1	g/min	(9.2E-3,9.2E-1)	[108,109]	1.68E-1	3.38E-4
qbSkmus	4.44E-2	g/min	(6.8E-3,6.8E-1)	[108,110]	7.40E-1	-6.69E-2
qbSmint	8.79E-1	g/min	(1.0E-2,1.0E+0)	[108,109]	1.34E+1	1.20E-2

Table D.3.

Distribution model static parameters including model value, references utilized for scaling and PRCC values. All mass parameters are sourced from literature and experimental measurements from this study (TS) and were held constant to generate an "average" behavior according to our murine animal model. Compartment plasma masses are parameterized from average mass of plasma per unit mass tissue and appropriately scaled to match tissue mass utilized in this model for measurements of Aha.

Name	Model Value	Units	Physiological Range	Source(s)	PRCC
mbBrain	4.52E-3	g	(9.54e-3, 1.17e-2) g/g_{tissue}	TS, [108,109]	6.50E-3
mbLiver	3.96E-2	g	(3.39e-2, 6.15e-2) g/g_{tissue}	TS, [108,109]	-3.63E-3
mbRenal	3.07E-2	g	(9.65e-2, 1.37e-1) g/g_{tissue}	TS, [108,109]	4.17E-3
mbSkmus	1.04E-3	g	(4.65e-3, 6.98e-3) g/g_{tissue}	TS, [108,109]	3.16E-3
mbSmint	5.18E-3	g	(2.53e-4, 2.65e-4) g/g_{body}	TS, [108,109]	5.92E-3
mbSysrv	1.09E+0	g	(4.45e-2, 6.45e-2) g/g_{body}	TS, [108,109]	6.00E-3
mbSysto	1.17E+0	g	(4.85e-3, 6.85e-3) g/g_{body}	TS, [108,109]	-9.99E-4
mtBody	2.00E+1	g	(1.45e+1, 2.55e+1) g	TS, [108,109]	6.03E-2
mtBrain	4.26E-1	g	(2.77e-1, 5.75e-1) g	TS, [108,109]	-5.66E-3
mtLiver	8.30E-1	g	(5.36e-1, 1.12e+0) g	TS, [108,109]	-2.51E-2
mtRenal	2.63E-1	g	(1.91e-1, 3.35e-1) g	TS, [108,109]	1.20E-3
mtSkmus	1.79E-1	g	(1.25e-1, 2.33e-1) g	TS, [108,109]	-1.34E-2
To minimize the number of variable parameters for fitting, plasma (mb) and tissue (mt) mass parameters from each model were held constant to match the average recorded experimental mass for each tissue. Therefore, all model values represent kinetics for an experimentally average mouse from our animal model.					

Table D.4.
Incorporation model fitting parameters including best fit values, fitting ranges, standard error of fit, PRCC values and informing literature. Parameters without published literature values are sourced from this study (TS) either from an experimental measurement or best estimate initial value that produced a single time-step Δ[Aha] that was one order of magnitude lower than the maximum recorded experimental value and swept upward and downward 1.5 orders of magnitude for fitting.

Name	Best Fit Value	Units	Fitting Range	Source(s)	Stderr of Fit	PRCC
kdBrain	3.54E-4	min^{-1}	(5.0e-1, 5.0e+2)	TS	1.37E+2	1.06E-2
kdLiver	8.29E-4	min^{-1}	(5.0e-1, 5.0e+2)	TS	7.13E+1	-2.15E-1
kdRenal	8.44E-4	min^{-1}	(5.0e-1, 5.0e+2)	TS	1.66E+1	-7.27E-2
kdSkmus	3.72E-4	min^{-1}	(5.0e-1, 5.0e+2)	TS	1.57E+2	1.92E-2
ksBrain	4.30E-5	min^{-1}	(1.0e-2, 1.0e+1)	TS	2.12E+0	-2.43E-1
ksLiver	3.16E-4	min^{-1}	(1.0e-2, 1.0e+1)	TS	1.53E+0	-1.19E-1
ksRenal	2.38E-4	min^{-1}	(1.0e-2, 1.0e+1)	TS	3.18E+0	-2.29E-1
ksSkmus	1.60E-6	min^{-1}	(1.0e-2, 1.0e+1)	TS	7.65E-1	-9.64E-2
kf	1E+0	$RU(g_t)(\mu g_p)^{-1}$	kf is included only as a scalar 1 to show [pAha] $\propto$ rF			