5-2018

High Precision Fret to Study Bimolecular Dynamics with High Temporal Resolution

Junyan Ma
Clemson University, junyanm@g.clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations

HIGH PRECISION FRET TO STUDY BIMOLECULAR DYNAMICS WITH HIGH TEMPORAL RESOLUTION

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Chemistry

by
Junyan Ma
May 2018

Accepted by:
Hugo Sanabria, Committee Chair
Andrew Tennyson
Feng Ding
Leah Casabianca

ABSTRACT

Förster Resonance Energy Transfer (FRET) is a promising methodology which is used by researchers in determining structural information of biomolecules. Other methodologies such as X-ray diffraction, nuclear magnetic resonance (NMR) and electron microscopy (EM), and X-ray scattering (SAXS) and circular dichroism (CD) can only give snapshots of the functional conformers. In the applications of most cases, FRET can only be used to give a qualitative result such as "yes" or "no" pattern. In our lab, we are able to expand this ability by applying FRET to quantitatively determine the structure and dynamic information in the biological system with high accuracy.

We first apply this methodology in measuring our standard double strand DNA in order to identify its ability to obtain the interdye distance with high accuracy. Then, we are able to identify the fast dynamic which happens in mRNA followed by a kinetic pathway of its folding pattern. Finally, we used FRET as well as discrete molecular dynamics (DMD) simulation in determining dynamic behavior of PDZ domains.

DEDICATION

I dedicate this dissertation to my parents for their love and support.

ACKNOWLEDGMENTS

I would like to thank lots of people who have helped me along the way to my Ph.D. journey.

First of all, I give my deep thank to all my committee members. My advisor Dr. Hugo Sanabria act as a mentor with magnanimous, honorable and knowledgeable characteristics. He is always ready to help and guide me through my way. This dissertation would not be done without his endless support and help. Dr. Tennyson is a good friend. He can always provide me with useful suggestions and directions. Dr. Ding is very generous and always with lots of inspirations and ideas. Dr. Casabianca is an honorable expert and has a fantastic knowledge in research.

I would like to thank Inna Yanez-Orozco as she provides me with lots of help and support in biophysics laboratory.

I would like to thank all the faculty and staff working in chemistry and physics department.

I thank my parents for their unconditional support and love to me.

TABLE OF CONTENTS

Table of Contents (Continued)

LIST OF TABLES

LIST OF FIGURES

List of figures (continued)

CHAPTER ONE

BIOLOGY

1.1) Central dogma of molecular biology

The central dogma of molecular biology is a description of the genetic

information flow which happens in a biological system. Previously, this is often stated as

"DNA makes RNA, and RNA makes protein," even though this is an over-simplification.

This description was first stated by Dr. Francis Crick [1] in the middle of 1960's.

Figure 1. The modern description of central dogma of molecular biology. This includes
flow from DNA to DNA (DNA self-replication); flow from DNA to RNA, and then flow
to the protein (transcription and translation); from RNA to RNA (RNA self-replication);
flow from RNA to DNA (reverse transcription.

The central dogma of molecular biology (Fig. 1) is one of the most important and

basic laws in modern biology which has played an enormous role in exploring the

essence and general laws of life phenomena. It has greatly promoted the development of

modern biology and is the theoretical foundation of modern biology. Also, it leads the

direction for the unification of the basic theory of biology, occupying an important

position in the development of biological sciences. In general, genetic material can be

DNA or RNA. Cell genetic material is DNA, and only some of the virus's genetic

material is RNA. This RNA-based virus is called a retrovirus [2], and during this virus's

1

infection cycle, a single-stranded RNA molecule can be reverse transcribed by a reverse transcriptase into single-stranded DNA, and then forms double-stranded DNA via a single-stranded DNA template. Double stranded DNA can become part of the host cell's genome and can be delivered to the daughter cells along with the host cell's genome. With the help of reverse transcriptase-catalyzed, RNA molecules produce its complementary DNA [3] sequence, the DNA sequence is called complementary DNA, abbreviated as cDNA. This process is a reversible transcription.

Thus, the genetic information does not necessarily flow from the DNA unidirectionally to the RNA, and the genetic information carried by the RNA can also flow to the DNA. However, the genetic information contained in DNA and RNA flows only unidirectionally to the protein. However, so far no genetic information of protein has been found to be able to flow back to the nucleic acid. The flow of this genetic information is the genetic significance of the central dogma that Francis Crick summarizes.

Any kind of assumptions are subject to the test of scientific fact. The discovery of reverse transcriptase allows the central rule to modify the one-way flow of genetic information from DNA into RNA. Flow can happen between DNA and RNA. So, when accounting for the flow of genetic information between DNA and RNA and protein molecules is only one-way flow of nucleic acid to protein molecules, or the information of protein molecules can also flow back to the nucleic acid, the central law still affirms the former. However, the behavior of the prion [4] proposes a serious challenge to the traditional central principle.

1.2) Nucleotides

Nucleotide, a class of organic molecules comprise purine base or pyrimidine base, ribose or deoxyribose and phosphate compounds. Also known as nucleotide acid. [5] Nucleosides are synthesized from pentose and organic bases, and nucleotides are synthesized with phosphoric acid. Nucleic acids are composed of four nucleotides. Nucleotides are mainly involved in the formation of nucleic acids, many of which also have a variety of important biological functions, such as energy metabolism related to adenosine triphosphate (ATP), dehydrogenase [6].

Some nucleotide can interfere with nucleotide metabolism, thus can be used as anti-cancer drugs. Depending on the type of five carbon sugar, the nucleotides have two types: ribonucleotides and deoxynucleotides [7]. Depending on the base (Fig. 2), there are adenine nucleotide (AMP), guanine nucleotide (GMP), cytosine nucleotide (CMP), uracil nucleus (Uridine acid, UMP), thymidine (thymidylate, TMP) and inosine (IMP). The number of phosphoric acid in nucleotide can be one molecule, two molecules and three molecules which is available in several forms (Fig. 2). In addition, nucleotide molecules can also be dehydrated and condensed into cyclic nucleotides.
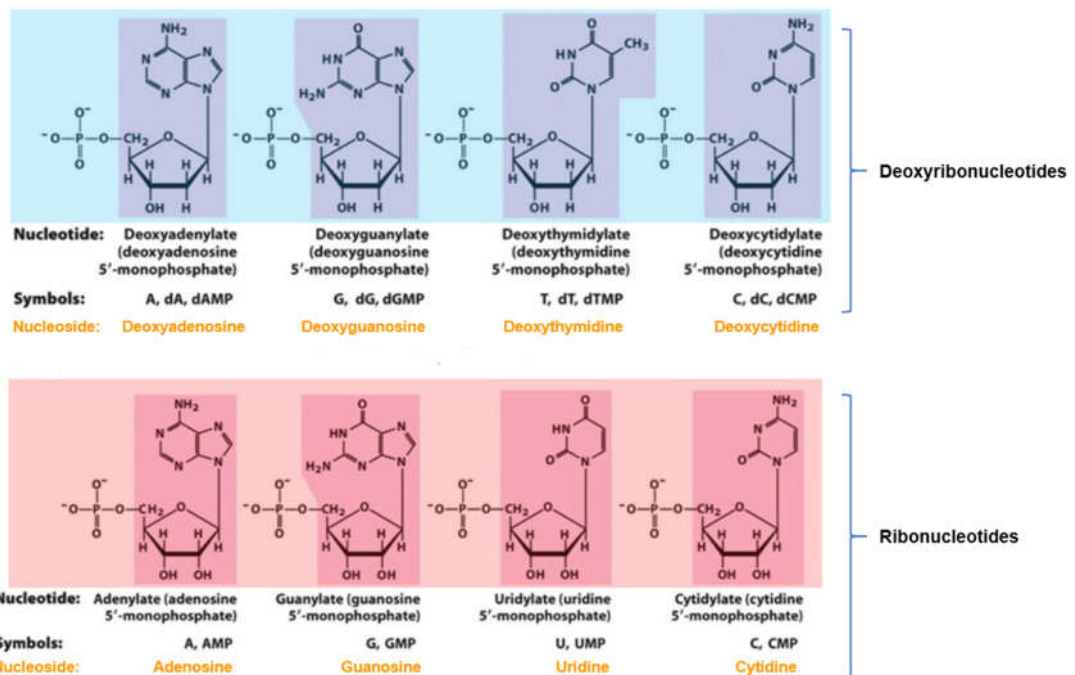
Figure 2. Eight types of nucleotides and their structures. Depending on the sugar, the nucleotides have two types of ribonucleotides and deoxynucleotides. Depending on the base, there are adenine nucleotide (AMP), guanine nucleotide (GMP), cytosine nucleotide (CMP), uracil nucleotide (UMP), thymine (TMP).

Nucleotides are the basic components of RNA and DNA as well as some precursors to synthesis of nucleic acids in vivo. Nucleotides are widely distributed in the nucleus and cytoplasm of various organs, tissues and cells along with the nucleic acid, and are also involved in the basic life activities of the organism such as genetics, development and growth as the components of the nucleic acid. There are also a considerable number of nucleotides that exist in free form in living organisms. For example, adenosine triphosphate plays a major role in cellular energy metabolism[8]. The energy release and absorption is mainly reflected in the production and consumption of adenosine triphosphate. In addition, uridine triphosphate, cytidine triphosphate and

4

guanosine triphosphate are also sources of energy in the synthesis. Adenosine is also the

component of some coenzyme [9], such as coenzyme I, II and coenzyme A.

1.3) DNA and RNA
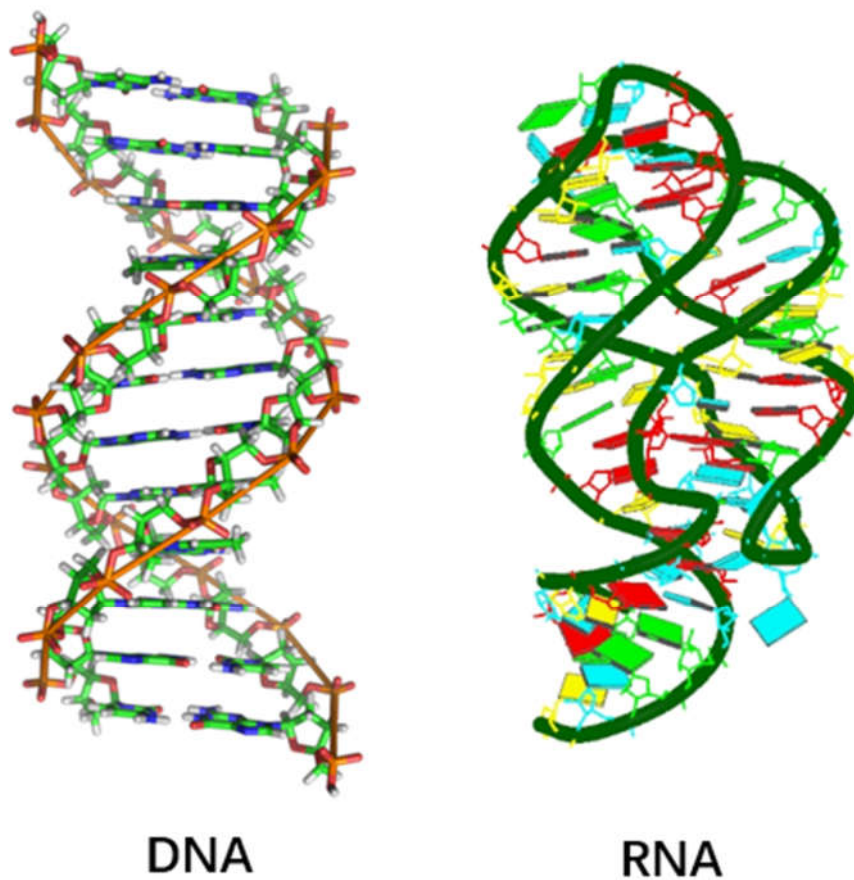
Figure 3. General structure of DNA and RNA. DNA is a long chain of a number of deoxynucleotides that are connected with each other by 3 ', 5'-phosphodiester bonds in a certain base sequence, and most DNA contains two such long chains. RNA are typically single-stranded molecules, which do not form a double helix structure. However, many RNA also need to form a secondary structure and even the tertiary structure to perform biological functions based on the principle of pairing. (DNA structure obtained from wiki; RNA structure obtained from PDB bank: (1Y27)

DNA is a biological macromolecule [10] that contains genetic information which could lead a way in biological functions. The major function of DNA is to store genetic information, which can be refer as "blue-print". The genetic instructions contained within DNA are needed to guide other intracellular compounds such as various of proteins and RNAs. Also, DNA fragments which is encoded by protein are referred as genes.

DNA is a long chain with many deoxynucleotides that are connected with each other by 3 ', 5'-phosphodiester bonds in a certain sequence (Fig. 3). Most of double strand DNA contains two chains, some of which are single-stranded, such as E. coli phage φX174, G4, M13 and so on. In some types of DNA, 5-methylcytosine can be substituted for cytosine in some situations. In some phages, cytosine is replaced by 5-hydroxymethylcytosine. In the late 1940s, Dr. E.Chargaff [11] discovered that the DNA composition various in different kinds of species. However, the number of adenines is same as the number of thymines (A = T), and the number of guanines is same as the number of cytosines (G = C). Thus, the sum of purines equals the sum of pyrimidines, which generally characterizes the structure of DNA in several levels.

The genetic matter of a prokaryote is a long DNA molecule, but prokaryotes have no real nucleus. Eukaryotic nuclei [12] have more than one chromosome, each chromosome contains only one DNA molecule. However, they are generally larger than the DNA molecules in prokaryotes. The function of a DNA molecule is: i) to store the entire genetic information that determines all the proteins and RNA structures of a species; ii) to plan the time and space for the organisms to synthesize the cellular and tissue components in an orderly pattern; iii) to determine the biological activity throughout the

life cycle and to determine the biological properties. Except for chromosomal DNA, a few DNAs with different structures also exist in the mitochondria and chloroplasts of eukaryotic cells.

Ribonucleic acid (RNA) is a genetic information carrier present in biological cells as well as some viruses. RNA chain is formed by phosphodiester bond condensation (Fig. 3). A ribonucleotide molecule consists of phosphoric acid, ribose and bases. There are four types of RNA bases, namely A adenine, G guanine, C cytosine, U uracil.

RNA species	Biological Functions
mRNA	carries information about a protein sequence to the ribosomes
tRNA	act as the catalytic component of the ribosomes
rRNA	transfers a specific amino acid to a growing polypeptide chain

Table 1. Three main types of RNA present in cells

Unlike DNA, RNA is typically a single-stranded long molecule that does not form a double helix structure, but many RNAs also perform the biological function by having secondary and tertiary structures through base pairing. The base sequence of RNA is basically the same as that of DNA. In cells, there are three main types of RNAs[13], tRNAs (transport RNAs), rRNAs (ribosomal RNAs), and mRNAs (messenger RNAs) depending on their structural function (Table. 1). mRNA is a template for the synthesis of proteins, the content of which is transcript by DNA in the nucleus; tRNA is the identifier of the

base sequence (ie, the genetic code) on the mRNA and the transporter of amino acids; rRNA is a component of ribosomes and is a protein synthesis region.

1.4) Amino acids

Amino acids were first discovered by Louis-Nicolas Vauquelin and Pierre Jean Robiquet [14] in the early 19th century. Aspartic acid was isolated from the asparagus [15] by French chemists and later several amino acids were found separately. Amino acids were finally named by chemists in the laboratory around 1900s after the hydrolysis of different proteins. A variety of different amino acids, have a structure consists of a carboxyl group and a side chain of substances. In 1820, the glycine was found in the hydrolyzate of proteins. By 1940, around 20 kinds of amino acids (Fig. 4) were found which are available in nature.

Amino acids are a group of organic compounds containing amino and carboxyl groups collectively, the basic constituent unit of a biological protein. Amino acids can be synthesized in plant, animal tissues and hydrolyzed by proteins. They play an important role in the metabolism, growth, maintenance and repair of tissues. Amino acids are defined as organic compounds that contain one basic amino group and one acidic carboxyl group, but the general definition of amino acids are the structural units that build up the protein. Proteins are important active molecules in the body, including enzymes that catalyze metabolism. As an important nutrient, amino acids can be obtained from food, the main sources are: milk, meat, eggs, soy products.

Most of the amino acids in the body can be decarboxylated to produce the corresponding primary amine. Amino acid decarboxylase is very specific, each amino

8

acid has a decarboxylase, and their coenzyme is pyridoxal phosphate. Amino acid decarboxylation reaction is widely found in animals, plants and microorganisms, some products have important physiological functions, such as decarboxylation of L-Glu in brain which produce r-aminobutyric acid [16], is an important neurotransmitter. His decarboxylation to produce histamine (also known as histamine), has a function in lowering blood pressure. Tyramine [17] which comes from Tyr decarboxylation, has the function of raising blood pressure. However, most amines are toxic to animals, and amine oxidase can oxidizes the amine into aldehydes and ammonia.
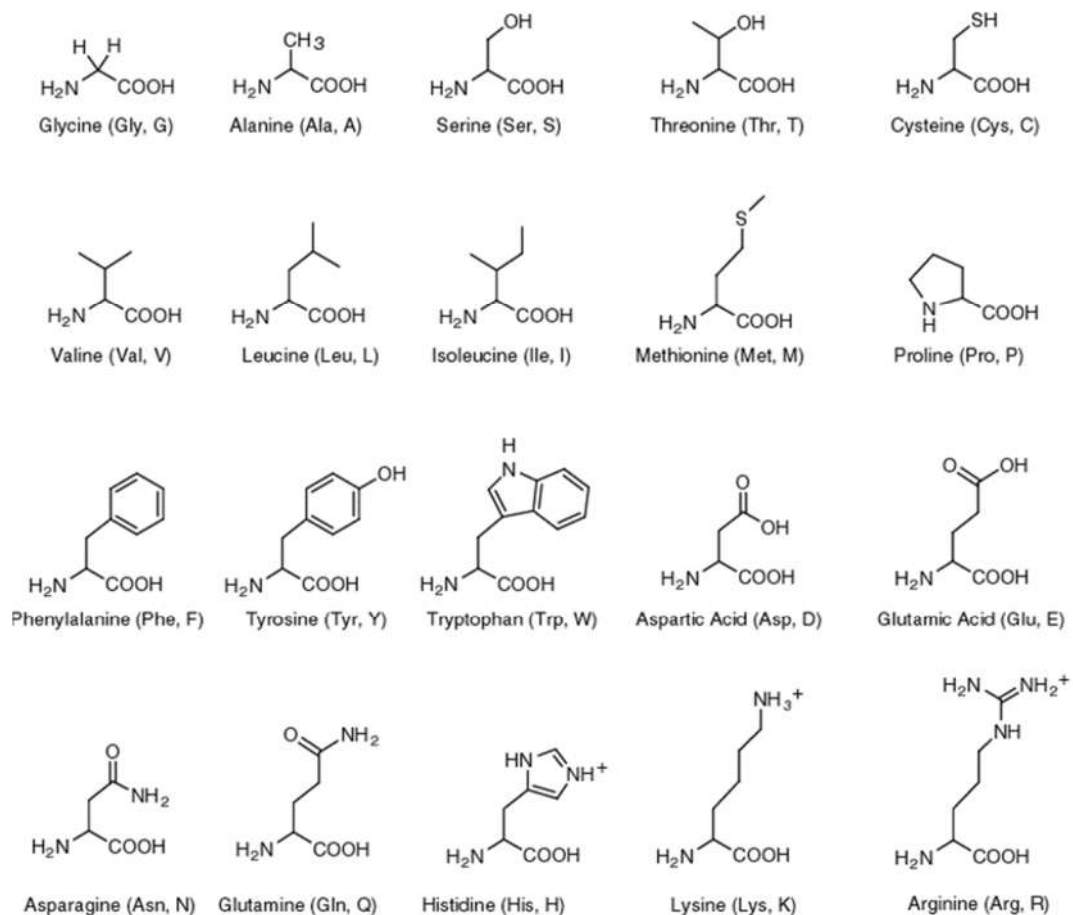
Figure 4. 20 types of essential amino acids widely present in proteins. Amino acids are the basic substances that build the protein which is needed for animal nutrition. It is an organic compound containing a basic amino group and an acidic carboxyl group. The amino group attaching to the α-carbon is called an α-amino acid. Most of the amino acids that make up proteins are alpha-amino acids.

1.5) Protein

Protein is the basic component of life. It is a biomolecule which is the basic organic material that builds the cell. In other words, there will be no life without protein present. Within proteins, amino acids are the basic components of proteins. Each cell in human body and almost all components have proteins involved in. There are many types

of proteins in the human body which have different properties and functions. They all

consist around 20 kinds of amino acids arrange in different proportions. Also, proteins

are constantly metabolized and renewed in the body.

Amino acid	Codons	Amino acid	Codons
Ala / A	GCT, GCC, GCA, GCG	Leu / L	TTA, TTG, CTT, CTC, CTA, CTG
Arg / R	CGT, CGC, CGA, CGG, AGA, AGG	Lys / K	AAA, AAG
Asn / N	AAT, AAC	Met / M	ATG
Asp / D	GAT, GAC	Phe / F	TTT, TTC
Cys / C	TGT, TGC	Pro / P	CCT, CCC, CCA, CCG
Gln / Q	CAA, CAG	Ser / S	TCT, TCC, TCA, TCG, AGT, AGC
Glu / E	GAA, GAG	Thr / T	ACT, ACC, ACA, ACG
Gly / G	GGT, GGC, GGA, GGG	Trp / W	TGG
His / H	CAT, CAC	Tyr / Y	TAT, TAC
Ile / I	ATT, ATC, ATA	Val / V	GTT, GTC, GTA, GTG
START	ATG	STOP	TAA, TGA, TAG

Table 2. Commonly used codon table. mRNA can determine the type and order of amino acids in protein molecules. The four nucleotide (base) sequences in a messenger RNA molecule determine the sequence of 20 amino acids in a protein molecule. Three bases on the messenger RNA molecule determine one amino acid. (Table obtained from wiki and modified)

        Amino acids are the basic units that build up proteins, and amino acids form

peptide chains by dehydration condensation. Proteins are biological macromolecules

which are composed of one or more polypeptide chains. Each chain contains 20 to

several hundred of amino acids. A number of amino acid residues are arranged into a

certain order. The amino acid sequence of a protein is encoded by the corresponding gene.

In addition to the 20 essential amino acids[18] encoded by the genetic code (Table 2),

certain amino acid residues in a protein can also be post-translationally modified to

undergo chemical changes that could activate or regulate the protein[19]. Multiple proteins

11

can collaborate together, often by binding together to form stable protein complexes which will have a specific structure. These complexes will have certain biological functions. The organelles for synthesizing peptides are ribosomes on the matted endoplasmic reticulum in the cytoplasm. Proteins differ in their amino acid species, number, arrangement order, and spatial structure of the peptide chain (Fig. 5).

COOH · · · · · · · · · · · · · · · · NH₂

Primary Structure

α helix

β sheet

Secondary Structure

Tertiary Structure

Quaternary Structure
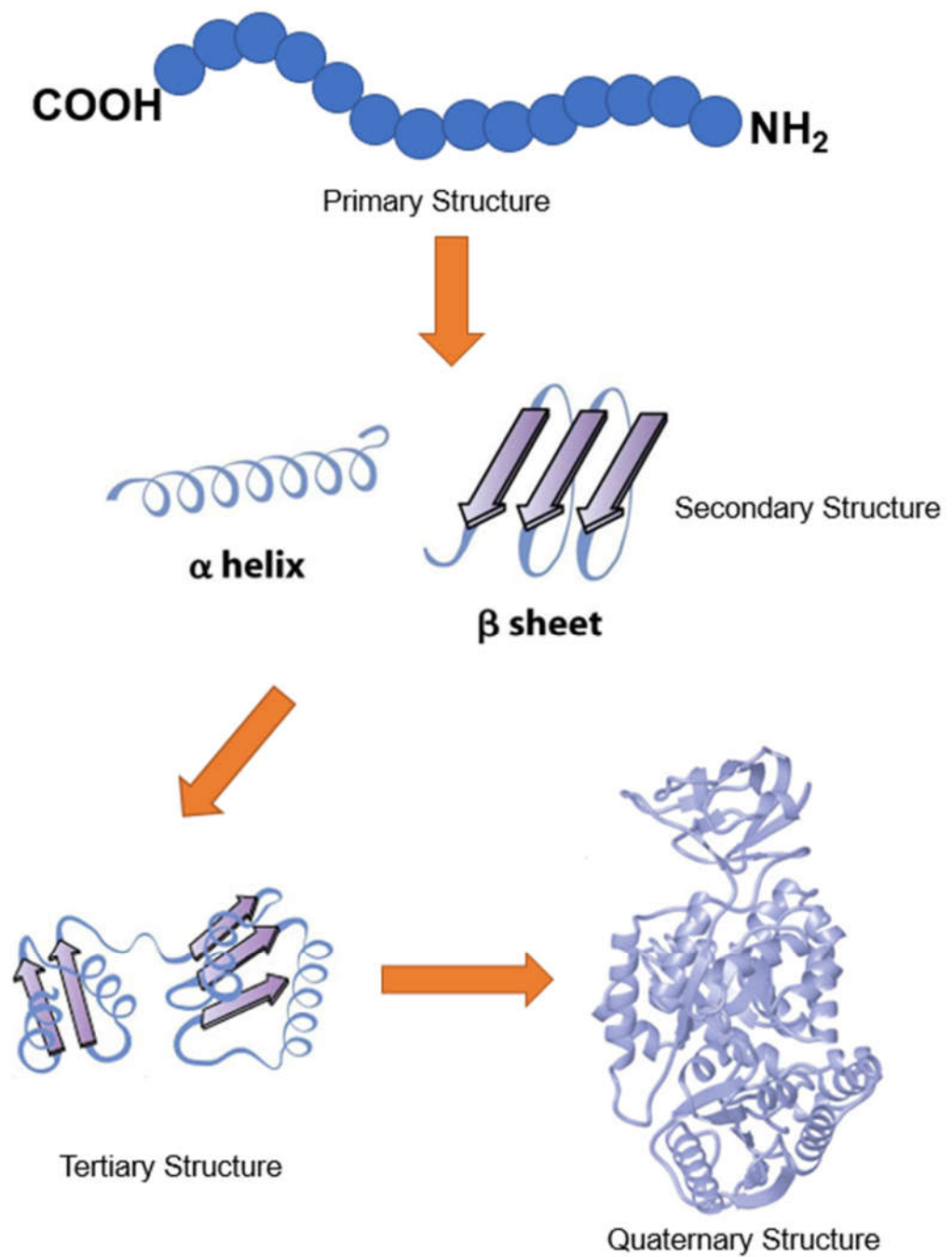
Figure 5. Structure of proteins. Proteins are biopolymers composed of amino acids. The amino acid sequence of the protein molecule and the three-dimensional structure determine the diversity of the protein structure. Proteins have primary, secondary, tertiary, and quaternary structures, and the structure of protein molecules determines its function.

Primary structure: The sequence of amino acid residues in the peptide chain of a protein is called the primary structure of the protein. Each protein has a unique amino acid sequence.

Secondary structure: The protein molecules in the peptide chain is not linear, but folded in a certain pattern such as α-helix or β-sheet [20] to form a specific spatial structure, which is called protein secondary structure. The secondary structure of a protein depends mainly on the hydrogen bond formed between the hydrogen atom on the amino acid residue (-NH-) in the peptide chain and the oxygen atom on the carbonyl group.

Tertiary structure (tertiary structure): The peptide chain forms a more complicated tertiary structure in a certain pattern on basis of the secondary structure.

Quaternary structure: An aggregate structure formed by arranging polypeptide chains with tertiary structure in a certain spatial arrangement is called the quaternary structure of protein. For example, hemoglobin is composed of four polypeptide chains with tertiary structure. Two of them are α-chains, the other two are β-chains, and the quaternary structure is approximately ellipsoidal.

Different protein has its unique function and this is quite important to biological system. For example, fibrous protein provides mechanical strength to individual cells or entire organisms and thus playing a protective or structural role, chaperone can prevent incorrect folding intermediates from forming and incorrect, myoglobin is responsible to transport oxygen from the lungs to peripheral tissues.

1.6) Biomolecular Dynamics

Historically, structure and dynamics belong to two different realms of biophysics [21]. An important goal in biological structural studies is to obtain information of the relation between structure and bio-function. Since the 1950s, our view on biomolecules (e.g., proteins and nucleic acids) has been given, by X-ray crystallography [22,23], which provides high-resolution static structural information determined by the constraints in the crystal. Thus, the immobility of X-ray models puts less emphasis on the dynamic nature of biomolecules, even though most biological systems are not static. With the idea that a rigid biological structure would not be functional. Then, the need of protein dynamics is called for. Later, with the introduction of nuclear magnetic resonance (NMR) [24], it was possible to unravel the intrinsic dynamics of biomolecules and conformational ensembles, thus bridging structure, dynamics and function. However, NMR [25] [26] [27] [28] is limited by sample size, the large amounts of sample required, and, often, sample solubility at those conditions. Nevertheless, nowadays it is possible to integrate various sources of information to provide a hybrid model generating biomolecular structural and dynamic models. Integrating various experimental and computational approaches remains a challenge for the scientific community as we seek to understand the intrinsic relationship between structure and dynamics.
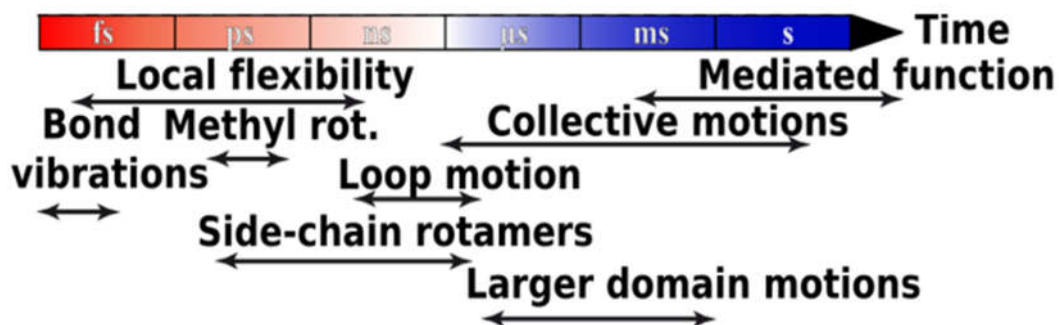
Figure 6. Biologically relevant motions and timescales. Biomolecular dynamics cover a wide range of timescales from local flexibility occurring on the fs to ns time scales, to longer range motions which occur on much slower timescales (sub-ms to seconds)

Biomolecular dynamics timescales range[29-31] from local motions (with the time scale from femtosecond to nanosecond), to long range motions (macromolecular conformational changes) which occurs on a much slower timescale (Fig. 6). Thus, the main challenge here is to get temporal as well as spatial information in a high resolution.

1.7) Methods to monitor dynamics

Many methodologies can only provide snapshots of functional conformers of biomolecules. However, each of them has its limitations. Small angle X-ray scattering [23,32,33 34] (SAXS) loses the atomistic details but can provide more information on the dynamics. Electron microscopy/tomography (EM/ET) and cryo-EM are applied to map large structural assemblies together with their structural rearrangements. However, the achievable resolution is limited in particular in asymmetric structures. Circular dichroism (CD) provides information only about the secondary structure of biomolecules but can be used in kinetic experiments. Biophysical or biochemical methods, such as mass spectrometry or locally specific chemical reactivity (e.g. foot printing, crosslinking,

16

hydroxyl radical probing), can also provide dynamic structure information (>
milliseconds) which needs a good structural model to interpret the observations. In
contrast, other spectroscopic approaches (such as UV-Vis, IR, fluorescence, EPR) can
provide very detailed temporal and spatial information, which is limited to the local
environment of the spectroscopic probe. Moreover, labeling biomolecules specifically
with two unique spectroscopic probes, pulsed EPR [32] and FRET[35-38] can be used to
provide precise long-range distance information on structure and dynamics.

The combination[39-41] of FRET with computational methods and other biophysical
tools such as hydrogen exchange experiments (HDX), NMR[25,28], small-angle X-ray
scattering (SAXS) or electron paramagnetic resonance (EPR) could then be used to
generate an atomistic-like view of biomolecular structures.

1.8) Summary

FRET[42-53] has not been taken as an important role in studying biological structure.
However, several recent developments have considered FRET as a promising structural
biology method. People have shown that FRET has many advantages in determining
structures and dynamics in biological system: (i) high sensitivity due to fluorescence,
smFRET experiments[54] can identify heterogeneous ensembles through counting and
simultaneously characterizing the molecule structures at the same time; (ii) Reaction
pathways could be directly depicted in smFRET studies; (iii) FRET can access a wide
range of domains which range over 10 decades in time covering varieties of biologically
conformation dynamics; (iv) FRET experiments can be conducted in any solution
conditions, such as in vitro as well as in vivo; (v) Finally, FRET can be used to

characteristic the systems with almost any size: from polyproline oligomers to larger

organelles.

In this thesis, we start from an introduction of FRET (chapter 2), followed by how

to measure double stranded DNA sample (chapter 3) in order to verify the high accuracy

of using FRET as a tool in identifying biological structure. Then, we apply this powerful

tool in determining the dynamic behavior of TPP riboswitch (chapter 4). Finally, we

combined FRET experiments with DMD simulation to obtain information about structure

and dynamics of PDZ tandem (chapter 5).

CHAPTER TWO

FÖRSTER RESONANCE ENERGY TRANSFER

2.1) Fluorescence

Fluorescence phenomena was first observed by Bernardino de Sahagun in 1560. Matlaline, which is product of the flavonoids oxidation in the tree species is the reason of the fluorescence. Later, George Gabriel Stokes described more about this phenomenon and named this fluorescence in his paper. By absorbing photon, a molecule can be excited from its ground state to an excited state with higher energy. When this molecule relaxes back to its ground state, it can give out photon which is in the form of luminescence. Typically, luminescence contains two classes: fluorescence and phosphorescence. In the condition of fluorescence, the photon is emitted from the singlet state to the ground state directly. However, in terms of phosphorescence, the system first goes from singlet state to the triplet state through intersystem crossing. Then, the photon is emitted from the triplet state to the ground state (Fig. 7). In most cases, phosphorescence is longer and can take several minutes. The processes of luminescence are described in the Jablonski diagram.
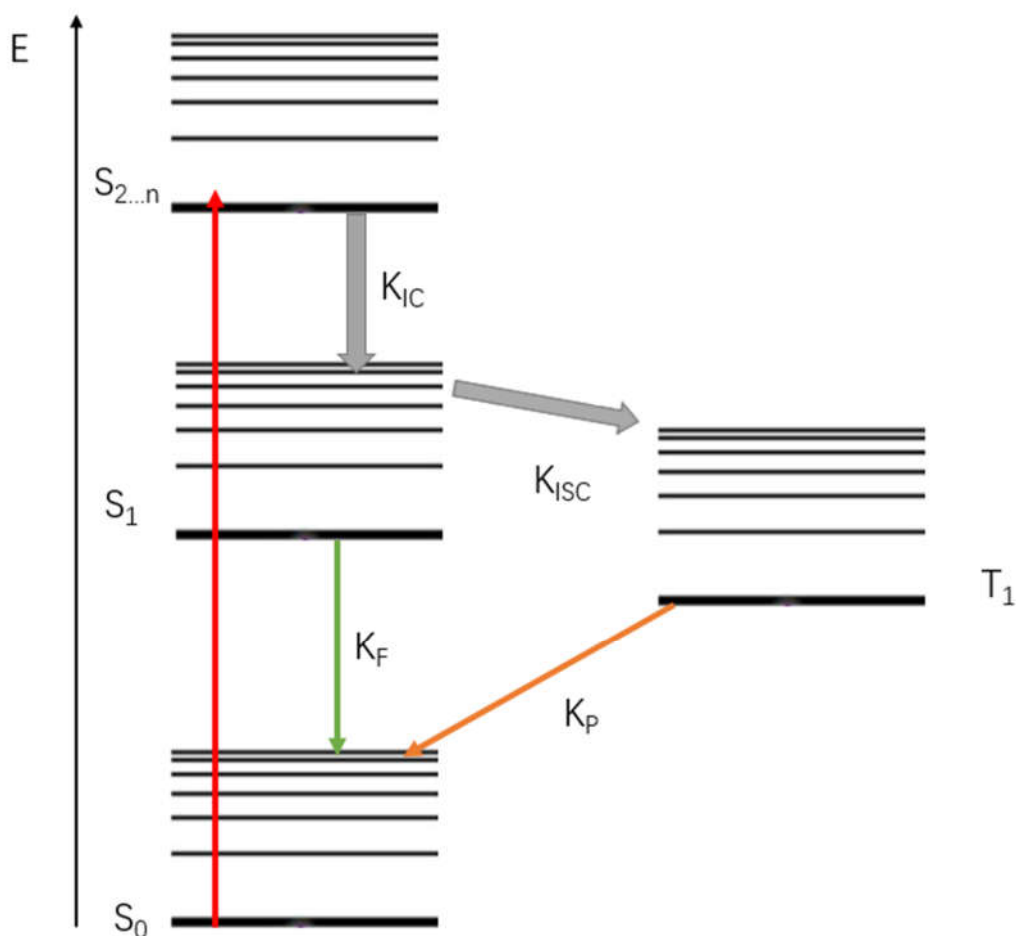
Figure 7. Jablonski Diagram. After absorption of a photon in a fluorphore, serveral relaxation processes could happen: Fluorescence (K_F), internal conversion (K_{IC}), intersystem crossing (K_{ISC}) and phosphorescence (K_P).

The quantum yield of the fluorophore[55] is typically defined by the ratio of photons emitted divided by the photon absorbed, which is show in this equation.

$$\Phi_F = \frac{\#\text{ of emitted fluorescence photons}}{\#\text{ of absorbed photons}} = \frac{k_F}{k_F + k_{IC} + k_{ISC}} \leq 1$$

(Eq. 2.1)

Where the k_F is the fluorescence rate, k_{IC} is the internal conversion rate and k_{ISC} is the intersystem crossing rate.

Fluorescence lifetime is defined by the average time needed by the electrons from excited to the ground state, and this process follows the equation:

$$\frac{dN}{dt} = -k_f N$$

(Eq. 2.2)

N is the total number of molecules in excited state, and k_f is the fluorescence emission rate constant.

Fluorescence lifetime is originally defined by the time needed of a fluorophore where the fluorescence intensity dropped to 1/e of its initial intensity. This process can be described as an exponential decay equation. And the inverse of the decay constant k_f is known as the lifetime (τ) of the corresponding fluorophore.

$$F(t) = F(0)\exp(t \cdot k_f) = F(0)\exp(t/\tau)$$

(Eq. 2.3)

Light is polarized, and anisotropy gives information of the characteristics property of a sample based on the polarization of light. When using polarized light for excitation, the sample can absorb photon which is parallel to the dipole moment of itself. Also, the emission light can be in forms of parallel or perpendicular. Thus, the anisotropy of fluorescence can be defined by the equation:

$$r = \frac{F_\parallel - g \cdot F_\perp}{F_\parallel + 2 \cdot g \cdot F_\perp}$$

(Eq. 2.4)

$F_{\parallel}$ is the emission intensity parallel to the incident bean while $F_{\perp}$ is

perpendicular emission intensity, g is the device factor which is corrected based on

instruments non idealities. The fundamental anisotropy is defined by the equation:

$$r_0 = 0.4 \frac{3\cos^2(\beta)-1}{5}$$

(Eq. 2.5)

Where β is the angle of the excitation and emission dipole which can be

determined. Also, anisotropy of a molecule which is in freely diffused state can be

determined by its lifetime and rotational correlation time, and it's given by equation:

$$r = \frac{r_0}{1+\frac{\tau}{\rho}}$$

(Eq. 2.6)

Where, ρ is the rotational correlation time which is determined by temperature,

size of the molecule and viscosity of solution. r0 is the maximum anisotropy value. τ is

the lifetime of the fluorophore.

2.2) FRET theory

Energy transfer from a donor (D) to an acceptor (A) dye may occur via radiative

or non-radiative processes, which have different underlying mechanisms. Förster's theory

deals with the case in which only coulombic interactions are considered. This is called

the "weak coupling regime" since the coulombic coupling does not perturb the energy

levels of D and A. Within this regime the energy transfer can be conceptualized with the

help of the classical picture of two oscillators being in resonance. The overlap integral J

which is shown in the spectral overlap of donor emission and the acceptor absorption region is required for FRET[56].

$$J(\lambda) = \int_0^\infty F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda$$

(Eq. 2.7)

Here, λ is the wavelength, $F_D(\lambda)$ is the donor emission intensity, $\varepsilon_A(\lambda)$ is the coefficient wavelength of acceptor. The rate of the energy transfer can be calculated by the following equation:

$$k_{FRET} = \frac{1}{\tau_D}\left(\frac{R_0{}^6}{R_{DA}{}^6}\right)$$

(Eq. 2.8)

Here, R_{DA} is the distance between donor and acceptor, and R0 is referred as the Forster radius which is defined by the distance between Donor and Acceptor where 50% of the energy transfer happens. And this is defined by equation:

$$R_o^6 = 8.79 * 10^{-5} \cdot J \cdot \kappa^2 \cdot \Phi_D \cdot n^{-4}\Phi_D$$

(Eq. 2.9)

Where Φ_D is the fluorescence quantum yield of donor, n is the refractive index of the medium between the dyes, k^2 is the orientation factor, and J is the spectral overlap integral.

Fluorescence lifetime, intensity and distance of dye pair will affect the efficiency of FRET:

$$E = \frac{n_A}{n_D + n_A} = \frac{F_A / \Phi_{F,A}^{(D,A)}}{F_D / \Phi_{F,D}^{(D,0)} + F_A / \Phi_{F,A}^{(0,A)}} = \frac{k_{FRET}}{k_0 + k_{FRET}} = 1 - \frac{\tau_D^{(D,A)}}{\tau_D^{(D,0)}} = 1 - \frac{F_D^{(D,A)}}{F_D^{(D,0)}}$$

(Eq. 2.10)

Where n_A, n_D are the number of photons emitted by the acceptor and donor, $k_0 = 1/\tau_D^{(D,0)}$, is the relaxation rate constant of the excited donor in absence of acceptor, $\tau_D^{(D,A)}$ is the lifetime of donor in presence of A, $F_D^{(D,0)}$ and $F_D^{(D,A)}$ are the corrected fluorescence intensities of donor in absence and presence of acceptor, F_A refers as the corrected fluorescence intensity of acceptor, $\Phi_{F,D}^{(D,0)}$ and $\Phi_{F,A}^{(0,A)}$ are the fluorescence quantum yields of the unperturbed donor and acceptor. FRET efficiency also can be defined by the dye distance in the following equation:

$$E = \frac{1}{1 + \left(\dfrac{R_{DA}}{R_0}\right)^6}$$

(Eq. 2.11)

2.3) Sample labeling

Labelling fluorophore (Fig. 8) into biomolecules[57] needed to be quantitatively. For DNA or RNA samples, fluorophore can be labelled through chemical synthesis by using fluorescently labeled nucleotides. In the case of protein labeling, the most commonly used method is to introduce two cysteines at the interested position. Another method is to perform a chemical reaction which involves a conjugation of amine reactive probes with the w-amine group of lysines.
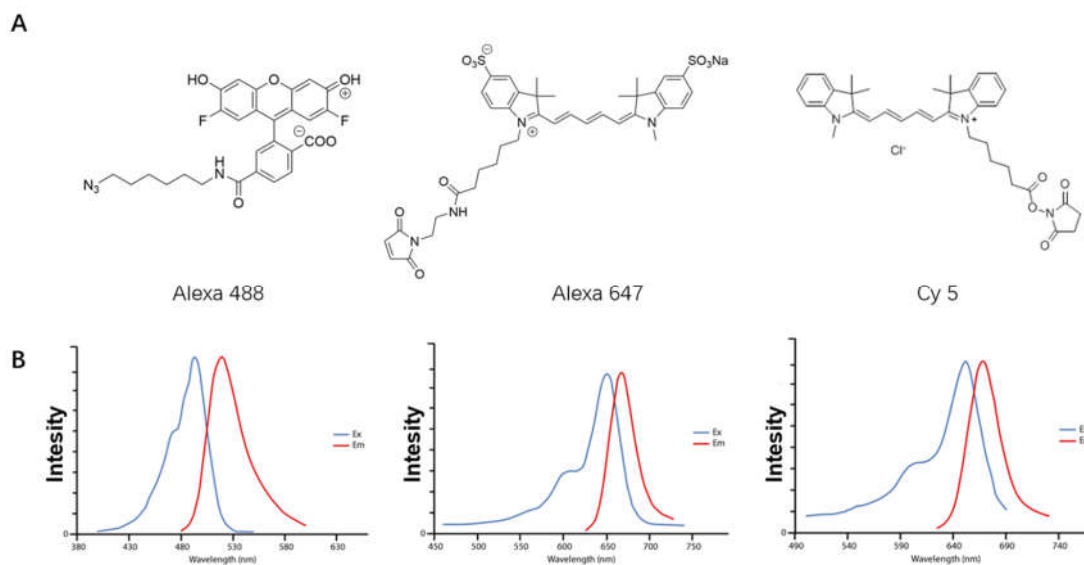
Figure 8. Three commonly used dyes structure and their spectra. Alexa 488, Alexa 647 and Cy 5 are three the widely used fluorophores. (Ex: excitation; Em: emission)

Dye species	Absorption (nm)	Emission (nm)	Quantum Yield
Alexa 488	495	519	0.92
Alexa 647	650	665	0.33
Cy 5	650	670	0.27

Table 3. Different dyes fluorescence properties

2.4) MFD setup

MFD refers to the experiment (Fig. 9A) in which several dimensions such as anisotropy, lifetime, spectral properties can be determined in a confocal experiment. When combined with FRET, we can obtain dynamic information as well as structural information within biomolecules.

25
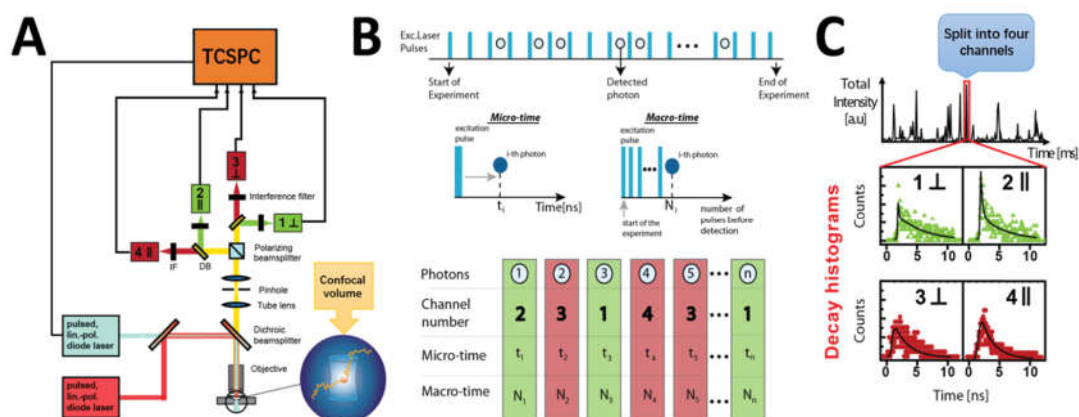
Figure 9. Multiparameter fluorescence detection setup and data collection for smFRET experiment. (Modify from Clue Seidel's group)

For MFD experiments[58,59], the biomolecule samples are diluted into extremely low picomolar concentration and placed on a glass slide facing a home-build confocal microscope. A sub-nanosecond laser pulse excites dye-labeled biomolecules freely diffused[60] through a detection volume. The typical confocal volume for the measurement is below 4 femtoliters in order to make sure only single molecule is individually detected in this situation (Fig. 9B). The emitted fluorescence signal from the dye-labeled biomolecules is collected via the objective and then are selected by a pinhole with ~100 µm. Then, the signal photos are separated into parallel and perpendicular parts (Fig. 9C) and separated by four detectors (two for donor and two for acceptor). Each photon detector channel is then coupled with time correlated single photon counting(TCSPC) electronics for later data analysis.

2.5) Fluorescence correlation spectroscopy (FCS)

Fluorescence correlation spectroscopy [61-64] (FCS) is one of the most widely used tools to get fluctuational information of any signal. Combined with FRET[65], it allows us to study fluctuations of biomolecule system over a large time range from ns to second. Fluctuations happens within structures can be obtained by the correlation function, which in turn could provide restraints on the number of dynamic states. The auto or cross-correlation of two correlation channels S_A and S_B is shown by:

$$G_{A,B}(t_c) = 1 + \frac{\langle \delta S_A(t) \cdot \delta S_B(t+t_c) \rangle}{\langle S_A(t) \rangle \cdot \langle S_B(t) \rangle}$$

(Eq. 2.12)

To increase contrast between correlation channels, fFCS[66] was developed by Felekyan by using MFD to optimally separate molecular species by their specific fluorescence lifetime-, polarization- and spectral-resolved fluorescence decays. Typically, fFCS is quite useful to separate two species within a system. fFCS takes advantage of the ability to generate optimal filters (f_j) for the fluorescence species and correlates the fluorescence signal ($S_j(t)$) trace weighted by the corresponding filters:

$$G^{(i,m)}(t_c) = \frac{\langle F^{(i)}(t) \cdot F^{(m)}(t+t_c) \rangle}{\langle F^{(i)}(t) \rangle \cdot \langle F^{(m)}(t+t_c) \rangle} = \frac{\left\langle \left(\sum_{j=1}^{2 \cdot L} f_j^{(i)} \cdot S_j(t) \right) \cdot \left(\sum_{j=1}^{2 \cdot L} f_j^{(m)} \cdot S_j(t+t_c) \right) \right\rangle}{\left\langle \sum_{j=1}^{2 \cdot L} f_j^{(i)} \cdot S_j(t) \right\rangle \cdot \left\langle \sum_{j=1}^{2 \cdot L} f_j^{(m)} \cdot S_j(t+t_c) \right\rangle}$$

(Eq. 2.13)

represents the Species Auto- (sAC) when $i = m$, and Cross-Correlation (sCC) function when $i \neq m$. Filter f_j is created by a bimodal decay histogram H_j. The conditional probabilities p_j is the weighted species specific histogram decay and satisfy equation:

$$H_j = \sum_{i=1}^{n(=2)} w^{(i)} p_j^{(i)} \quad , \text{with} \quad \sum_{j=1}^{d \cdot L} p_j^{(i)} = 1 \tag{Eq. 2.14}$$

Thus, filter can be generated which will minimize the photon count per species

($w^{(i)}$) and the weighted histogram $f_j^{(i)} \cdot H_j$:

$$\left\langle \left(\sum_{j=1}^{d \cdot L} f_j^{(i)} \cdot H_j - w^{(i)} \right)^2 \right\rangle \to \min \tag{Eq. 2.15}$$

Here, filter for each species are optimal, and it contains information such as

fluorescence lifetime decay and time resolved anisotropy decay.

CHAPTER THREE

HIGH PRECISE FRET MEASUREMENT

3.1) Introduction

In this chapter, we present how single molecule FRET spectroscopy [42] [66] [50] [35] can be used to derive distances for distance-restrained structural modeling with high accuracy. MFD [53] is advantageous because it considers eight dimensions of fluorescence information such as excitation and emission spectrum, anisotropy, fluorescence lifetime, quantum yield, macroscopic time, and the fluorescence intensities influenced by the stoichiometry and distance between fluorophores. Pulsed interleaved excitation (PIE) or alternating-laser excitation (ALEX) in combination with FRET measurements make the last criterion possible. To extract the most plausible information regarding biomolecular dynamics using fluorescence methods, we combine ensemble time correlated single-photon counting [67] (eTCSPC) with our custom-built MFD microscope using pulsed excitation and time-tagging acquisition. In MFD, each single-molecule burst is analyzed by FRET indicators such as the FRET efficiency, fluorescence lifetime ($\langle \tau_{D(A)} \rangle_f$), or fluorescence of donor over acceptor ratio (F_D/F_A). To analyze FRET distances, we use photon distribution analysis and apply theoretical analysis known as "FRET lines" to identify fast or slow interconversion between states.

The combination of FRET with computational methods and other biophysical tools such as hydrogen exchange experiments (HDX), NMR, small-angle X-ray scattering (SAXS) or electron paramagnetic resonance (EPR) could then be used to generate an atomistic-like view of biomolecular structures [40] Results

Accessible volume simulation (AVsim) which was first developed by Kalinin[68], can give a reasonable model in less than 1 s of computational time, providing a great advantage over MD simulations. In addition, AVs are a good solution for screening the large set of biomolecular models. In the AV method, fluorophores are typically simplified by a sphere with a defined radius R_{dye}, which is obtained from the molecules physical dimensions (Fig. 10). The connecting linker between them is modeled as a flexible cylinder with length L_{link} and width W_{link}. In case of nucleic acids, fluorophores are linked *in silico* to the C5 or C2 atom of the nucleic acid base, while in proteins the linkers are coupled *in silico* to C_β atoms of amino-acids.
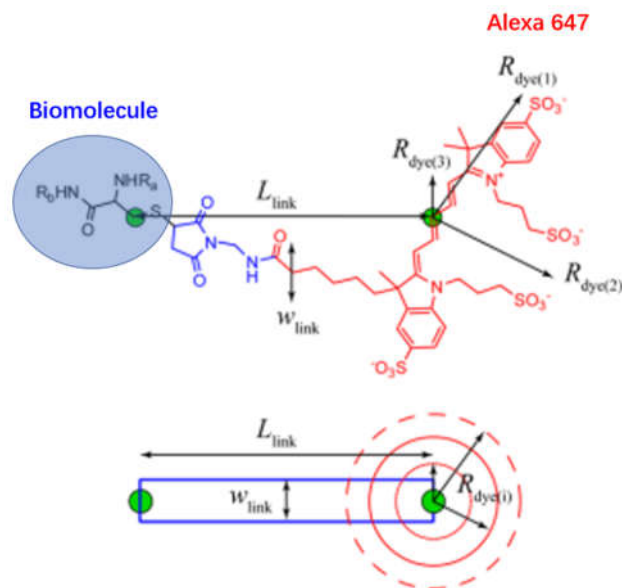
Figure 10. Schematic representation of a dye approximation Fluorophores (Alexa Fluor 647 C2 maleimide) are approximated by a sphere with a defined radius $R_{dye(i)}$, a connecting linker with length L_{link} and width w_{link}

The distance between mean dye positions Rmp is not accessible in FRET experiments. Rmp provides a graphical description of inter-dye distances and it is defined as

$$R_{mp} = \left| \left\langle \overrightarrow{R_D^{(i)}} \right\rangle - \left\langle \overrightarrow{R_A^{(j)}} \right\rangle \right| = \left| \frac{1}{n} \sum_{i=1}^{n} \overrightarrow{R}_D^{(i)} - \frac{1}{m} \sum_{j=1}^{m} \overrightarrow{R}_A^{(j)} \right|$$

(Eq. 3.1)

where $\overrightarrow{R_D^{(i)}}$ and $\overrightarrow{R_A^{(j)}}$ are all the possible positions (modeled by AV or MD simulations) that the donor and the acceptor fluorophores can take. In contrast to R_{mp}, two FRET observables are experimentally accessible.

In time-resolved fluorescence experiments the distribution of the donor-acceptor distances $p(R_{DA})$ is accessible. However, in real experiments often only the the mean and the width of the distribution $p(R_{DA})$ can be determined. Given an arrangement of flurophores in space the mean donor-acceptor distance $\langle R_{DA} \rangle$ is defined as:

$$\langle R_{DA} \rangle = \left| \left\langle \overrightarrow{R_D^{(i)}} - \overrightarrow{R_A^{(j)}} \right\rangle \right| = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \left| \overrightarrow{R_D^{(i)}} - \overrightarrow{R_A^{(j)}} \right|$$

(Eq. 3.2)

In contrast to time-resolved methods with ns-time resolution, intensity based measurements have a limited time-resolution. Therefore, information on the distribution function $p(R_{DA})$ of the fluorophores cause by dye linker motion is completely lost. Still, a quantity with a unit of distance can be determined:

$$\langle R_{DA} \rangle_E = R_0 \left(\langle E \rangle^{-1} - 1 \right)^{1/6}$$

(Eq. 3.3)

$$\langle E \rangle = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\frac{R_0}{R_0 + \left| \overrightarrow{R_D^{(i)}} - \overrightarrow{R_A^{(j)}} \right|^6} \right)$$

where the average efficiency is defined as

.
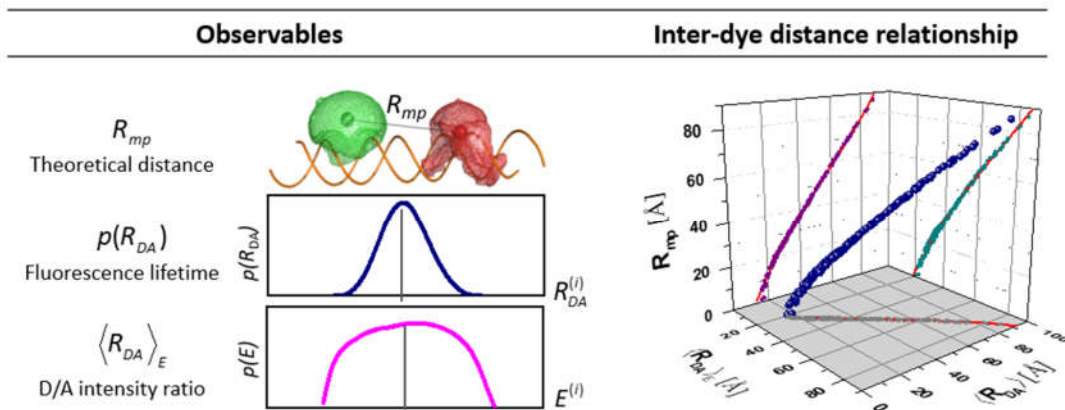
Figure 11. The various inter-dye distance definitions (R_{mp}, $\langle R_{DA} \rangle$ and $\langle R_{DA} \rangle_E$) are related to different phenomenons or fluorescence-observables. All inter-dye distances can be interrelated with empirical polynomials of 3rd order, shown as a 3D plot. Inter-dye distance definitions.

The R_{mp}, $\langle R_{DA} \rangle$ and $\langle R_{DA} \rangle_E$ average distances (Fig. 11) show differences depending on the size, shape and the mutual orientation of the dye distributions. The difference is referred, particularly for the distances below R_0 and can be of up to 10 Å (or ~30%). If not considered properly, this error contributes to large uncertainties in determined distances. R_{mp} can be approximated from experimentally obtained $\langle R_{DA} \rangle$ or $\langle R_{DA} \rangle_E$ using a conversion function based on the set of calculated dye distributions on known structures.

3.2) Results
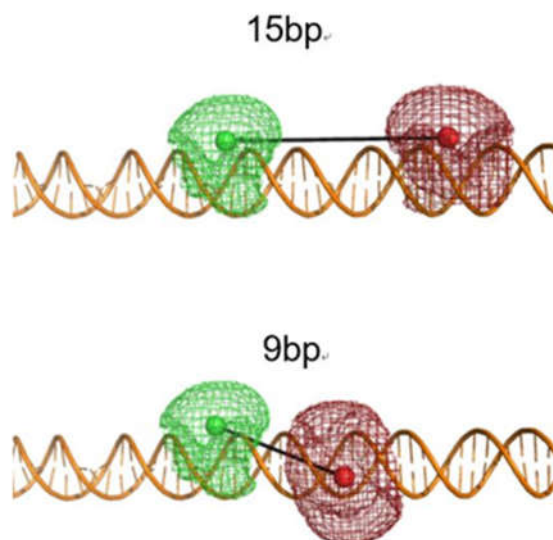
Figure 12. Two dsDNA with different interdye distance as standards. The dsDNA molecules are labeled with the dye pair (Alexa488–Cy5) at two distinct labeling sites using accessible volume simulations (9 and 15 bp separation)

We designed a set of complementary oligonucleotides, then we site specifically assign the location of the donor and acceptor fluorophores. Three unique dsDNA samples are designed (Fig 12). 1) No FRET, only donor strand with a complementary sequence with no dye; 2) A Low FRET (LFRET); and 3) A High FRET (HFRET). Below are the sequences of them:

1. Donor only dsDNA sequence

      NO label complimentary strand: 5′ -d(CGG CCT ATT TCG GAG TTG TAA ACA GAG ATC GCC TTA AAC GTT CGC CTA GAC TAG TCC AAG TAT TGC)

      Donor Strand: 5′ -d(GCA ATA CTT GGA CTA GTC TAG GCG AAC GTT TAA GGC GAT CTC TGT TT(Alexa488)A CAA CTC CGA AAT AGG CCG)

2. Low FRET dsDNA sequence

Donor Strand: 5´ -d(GCA ATA CTT GGA CTA GTC TAG GCG AAC GTT TAA GGC GAT CTC TGT TT(Alexa488)A CAA CTC CGA AAT AGG CCG)

LRET Acceptor Strand: 5´-d(CGG CCT ATT TCG GAG TTG TAA ACA GAG ATC GCC TT(Cy5)A AAC GTT CGC CTA GAC TAG TCC AAG TAT TGC)

3. High FRET dsDNA sequence

Donor strand: 5´ -d(GCA ATA CTT GGA CTA GTC TAG GCG AAC GTT TAA GGC GAT CTC TGT TT(Alexa488)A CAA CTC CGA AAT AGG CCG)

HFRET Acceptor Strand: 5´-d(CGG CCT ATT TCG GAG TTG TAA ACA GAG AT(Cy5)C GCC TTA AAC GTT CGC CTA GAC TAG TCC AAG TAT TGC)

DNA hybridization Process: stay at 95 ℃ for 10 min, 90 ℃ for 10 min, 80 ℃ for 10 min, 70 ℃ for 10 min, 60 ℃ for 10 min, 50 ℃ for 10 min, 40 ℃ for 10 min, 30 ℃ for 10 min, 20 ℃ for 10 min, 10 ℃ for 10 min and then remain at 5 ℃. All the above hybridization is done by a thermal mixer.

First, turn on the 485 nm and 640 nm laser power and all detectors of the single molecule measurement equipment which is built in our lab. Open the software (Sepia II and HydraHarp, PicoQuant GmbH) in order to control the measurement and make sure repetition rate of laser is 40 Mhz. Set the 485 nm pulsed laser power at 60 μW at image plane of the x60 water immersion objective and 640 nm pulsed laser power at 23 μW. Add objective immersion liquid (drop of double distilled water) between the objective lense and a glass slide (purchased from Fisher Scientific, 24X60-1.5). Then, move forward the focus knob one and a half cycle after finding two bright focuses. Add 0.1 μL 100 nM Rhodamine 110 into 50 μL distilled water to the center of the glass cover. Make

sure the solution is in the center of the objective lense as well. Adjust the pinhole (size 70 μm) positions (x and y direction one at a time) to maximize the number photons detected. Add 0.5 μL 100 nM Rhodamine 110 (purchased from BioChemika) into 50 μL distilled water and mix well. Place this mix at the center of the objective lens. Collect 120s of count rate in HT3 format. Calculate offline Fluorescence Correlation Spectroscopy (FCS) to determine the value of diffusion time(tdiff), number of molecules in the confocal volume, and molecular brightness46. These parameters need to be compared to theoretical and standard values. Add 10 μL 100 nM Rhodamine 101 (purchased from BioChemika) into 50 μL dd water and mix them well. Place this mixture at the center of the objective lens. Perform FCS measurement for 120s.

Add 1 μL 100 nM Alexa 647 (purchased from BioReagent) into 50 μL distilled water and mix well. Place this mixture at the center of the objective lens. Perform FCS measurement for 120s. Place 50 μL distilled water at the center of the objective lens. Perform FCS measurement for 300s. Place 50 μL PBS buffer at the center of the objective lens. Perform FCS measurement for 300s. Take 1 μL of the mix from above step and mix with 50 μL PBS buffer. Place this mix at the center of the objective lens. First, measure this for 10s. Then, check the burst from software Paris (BIFL software, MFD suite). If the burst is around 35 per 10 second, it is appropriate for single molecule measurement. Then, continue collecting count rate in HT3 mode for 1 hour to treat as a standard single molecule measurement sample.

Get the background factors of green (BG), red (BR) and yellow (BY) using the water sample got from above step. Determine the G-factor of green and "yellow" channels using sub ensemble Time Correlated single photon counting of Rhod 110 and Alexa 647 respectively. Due to the large file size, we split raw file into smaller pieces to load and processes simultaneously using a multi-core approach, reducing the analysis time. Load the files got from above step, select appropriate threshold and fit the related photo parameters. Then, get photo information files. Finally, use the photo information files got from above step. Check the stoichiometry. If the stoichiometry is at 1.0, it means the setup is fine ready for following measurements.

All fluorescence parameters information is gained in our measurement for each burst. in the case of mono-exponential donor decay and no conformational dynamics, this relation is defined by:

$$E_{static} = \left[1 + \gamma' \cdot \frac{F_{D|D}}{F_{A|D}}\right]^{-1} = 1 - \frac{\tau_D}{\tau_D^{(D,0)}}$$

$$\left(\frac{F_{D|D}}{F_{A|D}}\right)_{static} = \left[\gamma' \cdot \left(\frac{\tau_D}{\tau_D^{(D,0)}} - 1\right)\right]^{-1}$$

(Eq. 3.4)

Where $F_{D|D}$ is the intensity of the donor, $F_{A|D}$ is the intensity of the acceptor in presence of donor, and τ_D is the lifetime of the donor.

FRET lines are the theoretical relation between intensity based FRET-indications (e.g. F_D/F_A or E) and the fluorescence weighted average lifetime. For example, a system having interconverting FRET-levels (1) and (2) with corresponding lifetimes $\tau_D^{(1)}$ and $\tau_D^{(2)}$,

36

a dynamic FRET line which has been derived to account for the averaging of E nd $F_{D|D}/F_{A|D}$, this is given by:

$$E_{dyn} = 1 - \frac{\tau_D^{(1)} \cdot \tau_D^{(2)}}{\tau_D^{(D,0)}[\tau_D^{(1)} + \tau_D^{(2)} - \langle \tau_D \rangle_f]}$$

$$\left(\frac{F_{D|D}}{F_{A|D}}\right)_{dyn} = \frac{\langle \tau_D \rangle_x}{\gamma' \cdot \left(\tau_D^{(D,0)} - \langle \tau_D \rangle_x\right)} = \frac{1}{\gamma' \cdot \tau_D^{(D,0)}} \cdot \frac{\tau_D^{(1)} \cdot \tau_D^{(2)}}{\left(\tau_D^{(1)} + \tau_D^{(2)} - \langle \tau_D \rangle_f\right) - \frac{\tau_D^{(1)} \cdot \tau_D^{(2)}}{\tau_D^{(D,0)}}}$$

(Eq. 3.5)

The PIE stoichiometry is defined by

$$S_{PIE} = \frac{F_{R|D} - \beta F_{R|A} - \alpha F_{G|D} + \gamma F_{G|D}}{F_{R|D} - \beta F_{R|A} - \alpha F_{G|D} + \gamma F_{G|D} + F_{R|A}}$$

(Eq. 3.6)

$F_{G|D}$, $F_{R|D}$, and $F_{R|A}$ are fluorescence intensities [67] with background corrected,

for example $F_{G|D}$ =I - B_G, I is the intensity of donor, B_G is the background parameter from

the green channel. α is the correction factor for crosstalk happens from donor

fluorescence to the acceptor channel, β is the correction factor for acceptor excitation by

the donor excitation source, and $\gamma = \frac{\Phi_{F,A}^{(A,0)}}{\Phi_{F,D}^{(D,0)}} \cdot \frac{g_R}{g_G}$ is a function of the quantum yields Φ_F

and detection efficiency g.

Diffusion time sets the time-scale of the experiment and the FRET-levels

determine the donor lifetimes. After fitting the shape of expected distribution to the

experimental data obtained, the distance of donor and acceptor can be revealed. PDA[69] is

accomplished by calculating the probability $P(S_G, S_R)$ to observe a certain combination

of photons collected in the "green" (G) and "red" (R) detection channel given a certain time-window and given by equation:

$$P(S_{G},S_{R}) = \sum_{F_{G}+B_{G}=S_{G};F_{R}+B_{R}=S_{R}} P(F)P(F_{G},F_{R}|F)P(B_{G})P(B_{R})$$

(Eq. 3.7)

Here, $P(F)$ referred as the intensity distribution of fluorescence which is got from the total signal intensity distribution $P(S)$ with the assumption that the background signals of B_{G} and B_{R} are distributed following Poisson distributions, $P(B_{G})$ and $P(B_{R})$, with known mean intensities B_{G} and B_{R}. $P(F_{G}, F_{R}|F)$ is the conditional probability of observing a specific mixture of green and red fluorescence photons(F_{G} and F_{R}).
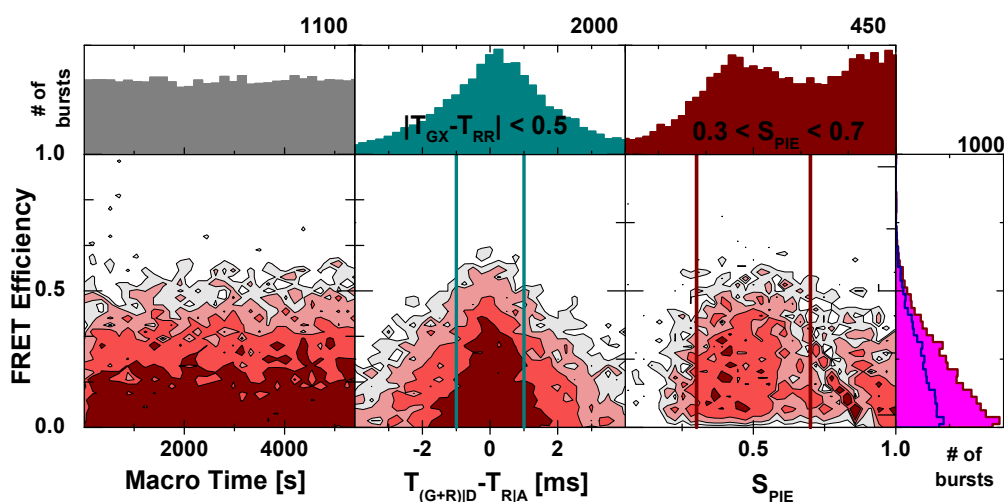
Figure 13. Mechanism of burst analysis via various parameters within fluorescence. FRET Efficiency vs. Macrotime, $T_{(G+R)|D} - T_{R|A}$ and S_{PIE} for the LFRET or 15bp dsDNA.

PIE-MFD[70] is applied to analyze variety of parameters of a molecule, such as mean mircro or macro time of photo, fluorescence lifetime, burst integrated anisotropy,

FD/FA (Fig. 14) as well as our calibration dsDNA standards. For example, we used the

LFRET DNA and the stoichiometry parameters (S_{PIE}) for the donor only, acceptor and

LFRET samples to calibrate the proper instrumental factors parameters in order to satisfy

S_{PIE}=1 for donor only sample, S_{PIE}=0 for acceptor only sample and around S_{PIE}=0.5 for

the most of FRET samples (Fig. 13).
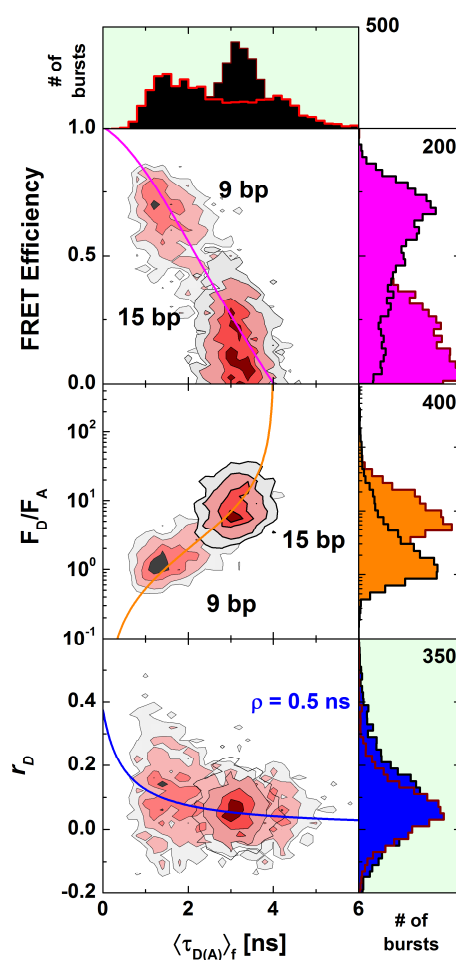
Figure 14. Histograms of intensity based parameters (E, FD/FA), donor anisotropy, and the donor lifetime. FD/FA is the ratio of the donor fluorescence over the acceptor fluorescence after correction for background counts, and cross-talk. In this panel the histograms share a common lifetime axis. Bursts selected are overlaid in red.
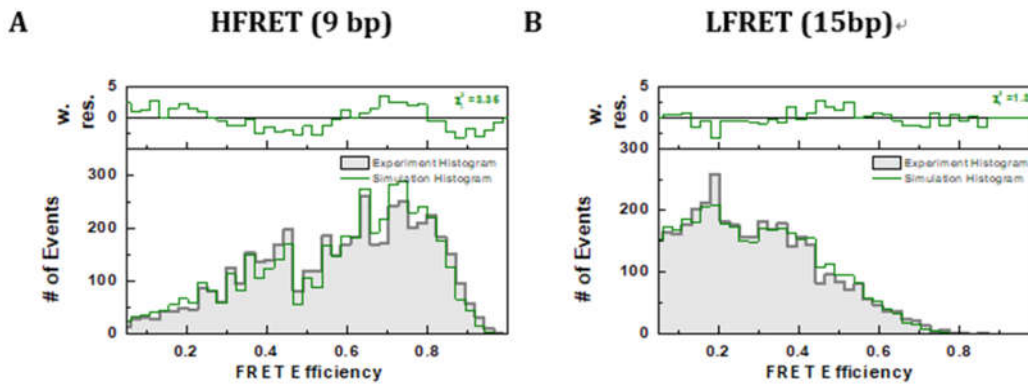
Figure 15. PDA comparisons of HFRET and LFRET double strand DNA. Time window was set at two ms with the hw$_{DA}$=6% of the mean FRET efficiency distance. We considered that each distance follows Gaussian distribution with 6 % of the $\langle R_{DA} \rangle_E$ as width (hw$_{DA}$). **(A)** For the HFRET DNA sample, the interdye distance is $\langle R_{DA} \rangle_E^{(HFRET)}$ = 45.7 Å, **(B)** LFRET that the distance $\langle R_{DA} \rangle_E^{(LFRET)}$ = 59.7 Å.

Sample	$\langle R_{DA} \rangle_E$ [Å]	$\langle R_{DA} \rangle_{E,AV}$ [Å]
	Experiment	AV simulation
HFRET	45.7 ± 2.3	44.7
LFFRET	59.7 ± 2.9	59.1

Table 4. Interdye distances obtained by smFRET and AV simulation

MFD reduces photo-physical and experimental errors artifacts and let one to be able to measure interdye distance with high accuracy up to ~ 1 Å in rigid biomolecules (Table. 4). The S_{PIE} parameter (dye stochiometry) is particularly informative because it serves to properly calibrate the acceptor quantum yield and select doubly labeled biomolecules. To calibrate the set-up, we use double stranded DNA (dsDNA) standards, where the donor and acceptor dyes are separated by 9 or 15 base pairs (bp). For the 9 bp

40

dsDNA, the measured interdye distance $\langle R_{DA} \rangle_E = 45.3$ Å is comparable to the expected

mean interdye distance $\langle R_{DA} \rangle_{E,sim} = 44.7$ Å by using the accessible volume (AV)

simulations. In the AV simulations, we coarse-grained the fluorescent dye and used

Monte Carlo simulations to model the space distribution of the dye. For the 15 bp dsDNA,

the measured distance is $\langle R_{DA} \rangle_E = 59.5$ Å and the expected distance is $\langle R_{DA} \rangle_{E,sim} = 59.1$ Å.

If the wrong quantum yield is used, or if the same quantum yield is used for both samples,

the corresponding distance of the 15bp dsDNA is $\langle R_{DA} \rangle_E = 63.7$ Å, with an error of 8%

compared to an error of 0.7% using the proper quantum yield. Recently, we participated

in a worldwide blind study directed by Prof. Thorsten Hugel (Freiburg University,

Germany) addressing a similar issue. In this study, participating laboratories around the

world were requested to determine the interdye distance of "unknown"-labeled dsDNA.

Our results deviated by only 1.3% of the expected values. When applying this

methodology (Fig. 15) to studies of more dynamic biomolecules, we found an agreement

of less than 3 Å.

One of the fundamental goals in structural biology studies is to obtain the

information of relation between structure and function of biomolecules. Since the 1950s,

our view on biomolecules (e.g., proteins and nucleic acids) has been formed, to a

significant extent, by X-ray crystallography, which provides high-resolution static

structural information determined by the constraints in the crystal. Therefore, the

immobility of X-ray structural models puts less emphasis on the dynamic nature of

biomolecules, even though most biological systems are not static. With the introduction

of nuclear magnetic resonance (NMR), it has been possible to recover the intrinsic

dynamic nature of biomolecules and conformational ensembles, thus bridging structure, dynamics and function. However, NMR is limited by sample size, the large amounts of sample required, and, often, sample solubility at those conditions. Nevertheless, nowadays it is possible to integrate various sources of information to provide a hybrid model generating biomolecular structural and dynamic models. Integrating various experimental and computational approaches remains a challenge for the scientific community as we seek to understand the intrinsic relationship between structure and dynamics.

3.3) Conclusion

In this project, we present how to align, calibrate, and measure with high precision interdye distances using PIE-MFD single molecule Förster Resonance Energy Transfer(smFRET) experiments. The accuracy can be increased on the measured distances by carefully calibrating all instrumental parameters [71] [52] and it can reach angstrom accuracy. In order to achieve this, a variety of multi-dimensional histograms are needed to identify populations for further analysis. Mean Macro time is used to verify stability of the measured biomolecule samples; donor and acceptor photobleaching are corrected and FRET populations is selected via the stoichiometry parameter. In addition, it shows that photophysical properties of the acceptor is changing depending on the labelling location, so we use SPIE distribution in order to correct for the acceptor quantum yield. Gamma factor (ɣ), as well as other correction factors can be used to improve the accuracy of the interdye distance measurement. We verified this method by

using two designed dsDNA standard samples [72] and found an accuracy of ~ 1 Å when compared with simulation values.

In summary, smFRET experiments in MFD [53] mode allows us to obtain interdye distance within the range of 25 Å ~70 Å. With network of distances applied, making it possible for us to use these as restrains in structural modeling, especially in cases when it is not easy to characterize with standard analytical methods in structural determination [73] [34 74 75 76].

CHAPTER FOUR

TPP RIBOSWITCH PROJECT

4.1) Introduction

Riboswitches, first reported by Henkin and later named by Breaker [77] [78] and co-worker, are mRNA elements that by binding directly to a small-molecule ligand control the transcription machinery [79] in a cis-fashion [80] [81]. Typically, two distinct functional domains compose riboswitches (Fig. 16A). The effector molecule or ligand is recognized by an aptamer domain, which adopts different three-dimensional conformations to allow for ligand binding. Target recognition with high selectivity for the effector is needed to elicit the appropriate regulatory [82] [78] [77] response. The region of the expression platform contains a secondary structural switch that interfaces with the transcriptional or translational machinery (e.g. RNA polymerase). Regulation is achieved at the overlap region between these two domains, known as the switching sequence. The interactions of this regime determine the conformation that the RNA adopts. Typically, two mutually exclusive structures correspond to the on and off states of the Riboswitch (Fig. 16B). Thiamine pyrophosphate (Fig. 16C) is the active form of vitamin B_1, an essential conenzyme commonly present in bacteria, fungi and plants. TPP riboswitch [83] [84] (Fig. 16D), also known as Thi-box riboswitch, is one of the widely known riboswitch which is quite sensitive towards thiamine pyrophosphate (TPP). It controls genes which would affect the transportation and synthesis of thiamine and its phosphorylated derivatives [80] [85].
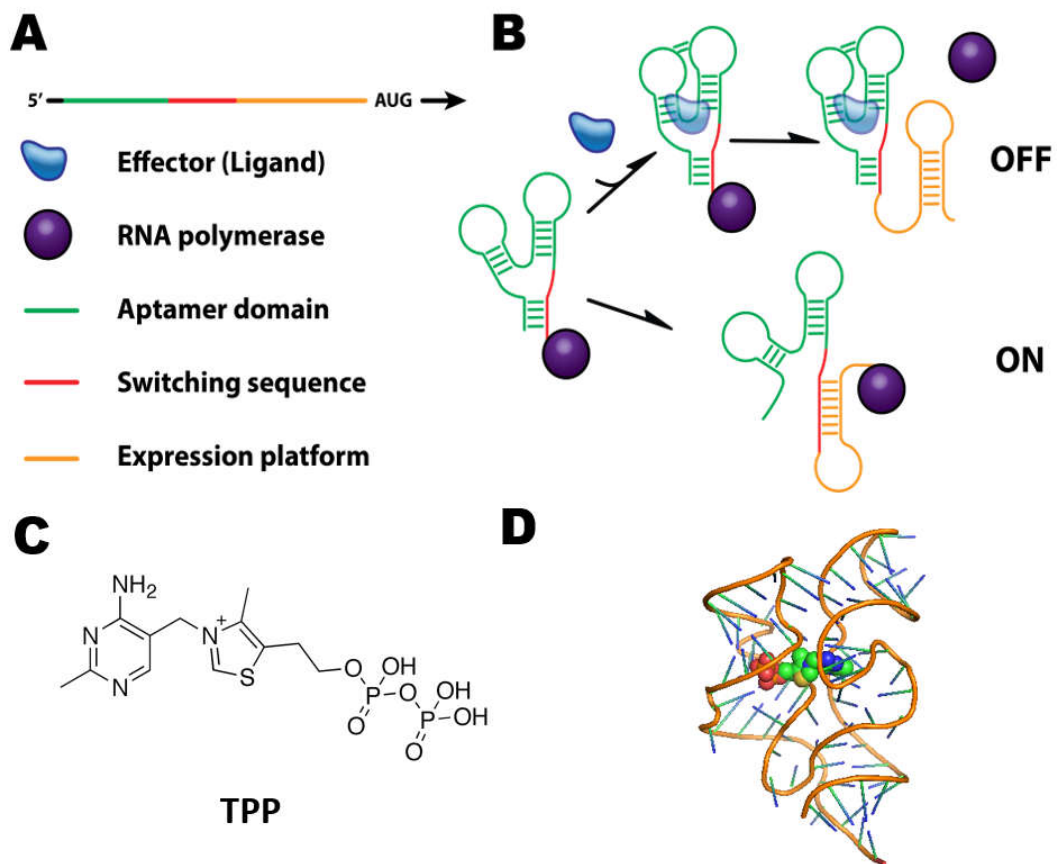
Figure 16. Typical organization and mechanism of riboswitch. A) Primary sequence of the mRNA: aptamer domain (green), switching sequence (red) and the expression platform (orange). B) The regulatory outcome during transcription depends on the potential binding of the effector molecule (blue). C) Chemical structure of thiamine pyrophosphate D) Cartoon-rendered structure of TPP riboswitch.
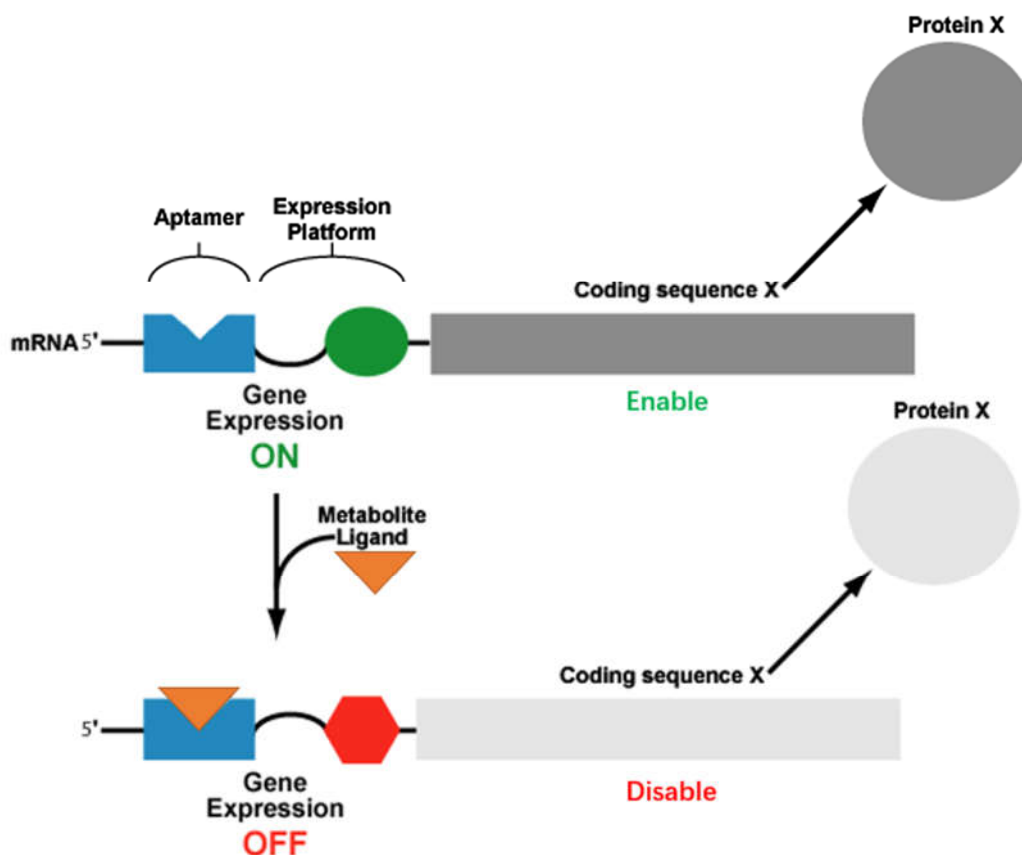
Figure 17. Mechanism of riboswitch when binding with small molecule. In normal state, aptamer is not bind with ligand and coding sequence would be able to perform the translation function. After metabolite ligand bind with aptamer, expression platform will change its conformation and leads coding sequence disabled. (Figure obtained and modified from Palmer slideplayer)

Previous researches on TPP riboswitch have been focused on the TPP ligand recognition process (Fig. 17) through several biophysical methods. Investigations of these include small angle X-ray scattering indicating multiple conformational states, isothermal titration calorimetry [86] [87,88] (ITC) showing that binding is highly dependent on Mg concentration, optical tweezer [89] reveals molecular extension and single molecule optical trapping methods to obtain information of TPP riboswitch folding and unfolding energy

landscape [90] [91] [91]. These researches have given us a sight into the mechanism of how TPP riboswitchs function. First, P2/P3 and P4/P5 secondary structure ligand sensor arms form with Mg in present. Second, the two arms will not have any interaction without TPP small molecule. Moreover, it is still not clear how the small molecule such as TPP affect the formation of binding pocket [92] [90].

More than 200 vital proteins exist in bacteria; however, only three antibiotic targets [93] [94] are commonly used: the cell wall synthesis machinery, the DNA gyrase, and the ribosomal RNA (rRNA). Many antibiotics selectively target ribosomal RNAs (rRNA), but with the increase of antibiotic resistant bacterial strains, it is evident that new targets are needed. In 2013, 2 million people suffered infection with an antibiotic resistant strain of bacteria; over 23,000 people died as a direct result of resistant bacteria according to the 2013 CDC report. Recently, it was proposed to use riboswitches as an innovative target avoiding the antibiotic resistance generated by changes in the rRNA. Riboswitches are attractive antibiotic targets[95].

Although promising, this structure-guided rational design depends significantly on the knowledge of the riboswitch three dimensional structures, which remains the main challenge on this field. Moreover, it would be ideal to identify all the conformational space of these molecules, such that all possible binding pockets that have regulatory functions are identified. Even though many studies have been made in decades, only in few cases the ligand-free or apo form of riboswitches have been determined using X-ray crystallography. One reason might be due to the intrinsic flexibility of the apo form. This is a critical aspect of the structure-guided rational design. Other methods, which are more

amenable to flexible biomolecules like Nuclear Magnetic Resonance, small-angle X-ray scattering and cryo-EM, are useful in these scenarios. However, these tools also face difficulties when determining structures of RNA.

Current structural biology methods face significant challenges for characterizing highly dynamic molecules[96,97] such as mRNAs. However, Förster Resonance Energy Transfer (FRET) could bring solutions to many of the potential challenges. With the basic knowledge of FRET [36], we will perform Ensemble Time Correlated Single Photon Counting (eTCSPC), Multiparameter fluorescence detection (MFD) and Filtered fluorescence correlation spectroscopy (fFCS) experiments. eTCSPC allows us to determine those states that are stable in the nanosecond timescales, and single-molecule experiments are useful in determining the sample heterogeneity and the dynamic exchange of the molecule. MFD allows for the selection of a burst (a single molecule event) with specific properties, such as the use of macrotime in identifying slow kinetics. FCS is commonly used to identify dynamics of a system occurring faster than the diffusion time of the molecule or even to the nanosecond range quantitively. Particularly, fFCS [66] can be used to identify more than two species based on species-specific MFD patterns. Therefore, based on our extensive expertise and the potential of use of this approach, a combination of these methods will make it achievable for us to derive the conformational landscape of the selected riboswitches.

Arabidopsis thaliana riboswitch is selected for our research. We used a coarse-grained approach to model the linker and the dye as a cylinder and spheres of dimensions obtained from their chemical structure. We then use Monte Carlo dynamic Simulations to

probe the accessible volume (AV) of the dyes. We used these simulations to determine

the donor-acceptor distance of all possible pair combinations *in silico*. Based on all

possible interdye distances, we selected position 25 and 56. After selection, we require

site-specific labeling of the riboswitch with orthogonal chemistry, we present the

sequence of the synthesized FRET-labeled-TPP riboswitch in SI (Y= 2'-dG-(N2-C6-

aminoCy5); X= rU(2'-O-propargyl Al 488)). The next step is to measure this sample.

4.2) Results

      To obtain dynamic information in submillisecond between states, we measured

smFRET with MFD/PIE in four different conditions. MFD 2D histogram include

fluorescence lifetime and intensity which is the integral of the overall time when the

labeled biomolecule passing through the confocal volume. Then conformation change

can be revealed by identifying the change in the liftetime and FD/FA ratio.
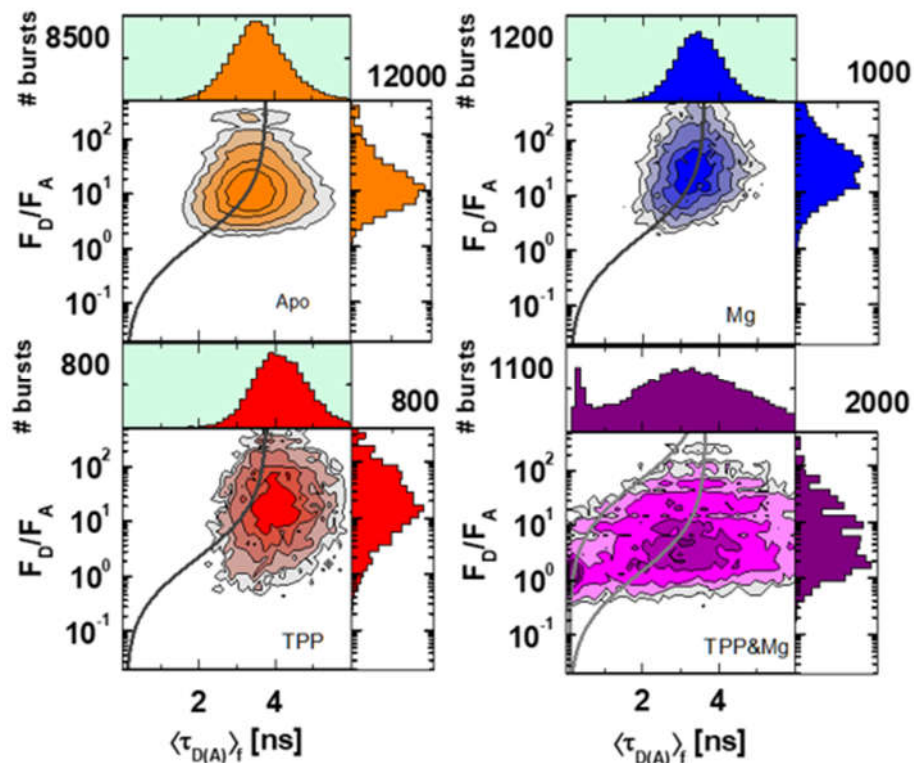
Figure 18. smFRET 2D histogram for riboswitch of the average lifetime per burst vs. ratio of donor over acceptor fluorescence of the RNA in variant conditions (Apo, Mg, TPP, Mg&TPP).

MFD is a kind of experiment using single molecule burst analysis. From the MFD plots (Fig. 18) that we got, we found some interesting results. In the apo condition, it shows a broad unimodal distribution center at low FRET region. When riboswitch is in Mg^{2+} condition, it still gives us a broad unimodal distribution slightly shifted toward lower FRET region with a tail pointing toward higher FRET region. For the TPP condition, it gives a much broader distribution with a significant tail pointing towards the high FRET region. The most interesting part is in the condition when both Mg and TPP are existing. The plots shows a bimodal broad distribution with an accumulation of burst

at higher FRET region and a broad distribution showing all possible states centers at slightly higher than mid FRET. As MFD is a useful tool for detecting dynamic information mainly at ms timescales, this dynamic equilibrium happens faster than ms. With only Mg^{2+} present, it shows a tailing going through an open state. The TPP condition shows that it goes through a higher FRET intermediate state. However, with both Mg and TPP, it shows two distinct FRET efficiency profiles with a stable X-tal like state.

Mg^{2+} ion is known to have an ability in binding with the aptamer domain of the riboswtich. In order to address the role in the conformation change of the riboswitch, we measured smFRET under with a range of Mg^{2+} ion (0-500mM). We calculated the FRET average efficiency for each condition. We depicted the plot which shows the relationship when adding Mg^{2+} ion into the buffer (Fig. 3A). It can be clearly see that the efficiency almost has no change just by changing the concentration of Mg^{2+} ion, which indicates riboswitch is under a fast dynamic conformation change. By measuring the FRET efficiency with a range of TPP (0-4.8μM), we can see there is a significant increase with higher concentration of TPP (Fig. 3B). This can be explained by riboswitch is going through a intermediate state which require a larger conformation change. The last measurement is done with fixing TPP concentration at 4.8 μM while changing the concentration of Mg ion. Again, it shows almost no significant change in the FRET efficiency (Fig. 3C), meaning TPP has an important role in larger conformation change while Mg ion only induce this process.
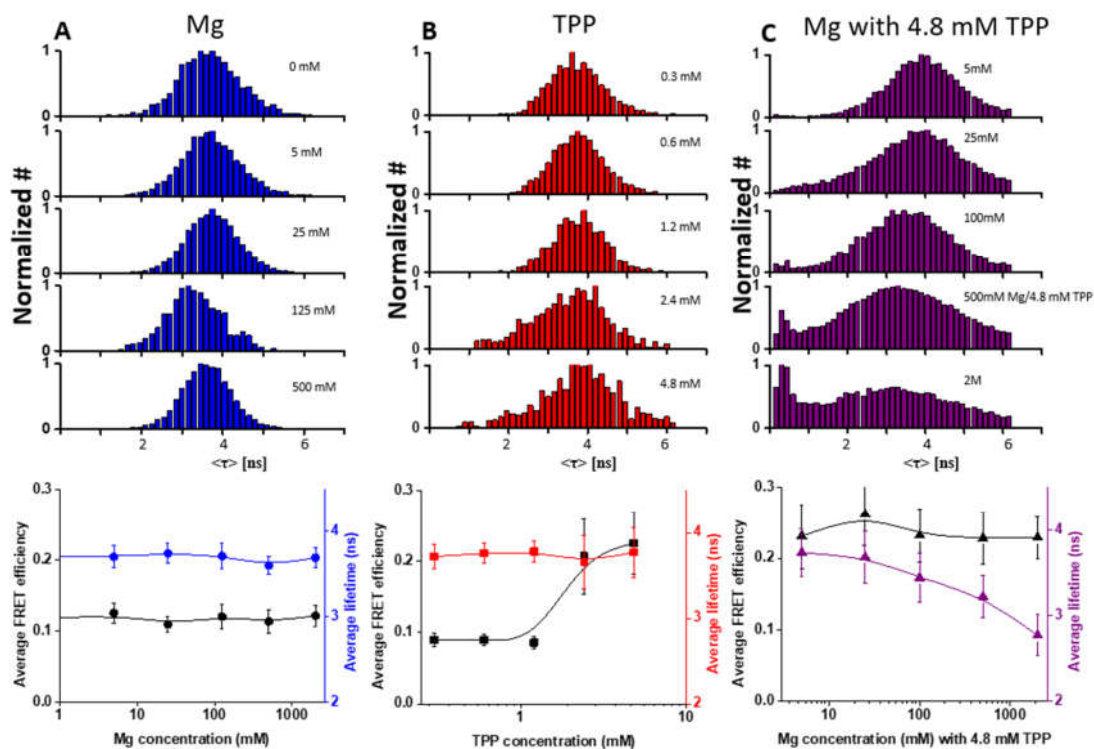
Figure 19. 1D histograms of smFRET lifetime in different buffer solutions (upper) and average FRET efficiency (lower). (A) with different concentration of MgCl2 starting from 5mM to 0.5M (B) with different concentration of TPP start from 0.15 µM to 4.8 µM. (C) only change concentration of MgCl2 with TPP concentration fixed at 4.8 µM.

In Mg^{2+} only buffer (Fig. 19A), the lifetime almost has no change with an increase in the concentration of $MgCl_2$. This indicates that riboswitch is in a fast dynamic motion. With an increase of the TPP concentration, the lifetime has a broader distribution meaning TPP has changed the conformation of riboswitch and this dynamic change is not fast (Fig. 19B). With both Mg and TPP present, the center of lifetime population shift to lower region and shows acceptor quenching (Fig. 19C).

Using ensemble time correlated single photon counting (eTCSPC), we measured the fluorescence lifetimes for double labeled and DOnly riboswitch (Fig. 20) in four

different buffers in order to identify the number of limiting states or conformational states. eTCSPC use the ps pulsed laser and can reveal structural states which are stable on ns timescale. To obtain the limiting states, a multi exponential decay model was applied to the fluorescence decay of the DOnly sample and the lifetime parameters are then used to fit the DA sample. A multiple Gaussian distribution model was then applied to the DA sample with each peak depicting the mean interdye distance for a limiting state or conformation.

In order to ensure the precision of the distribution fitting to the DA decay, we first apply the Gaussian model with three states and then reduced it one by one. To achieve a relatively low χ^2, we ended up with a two Gaussian distribution model which can be used to describe the conformational information best. Here, we present the plot with two limiting states in four different conditions.
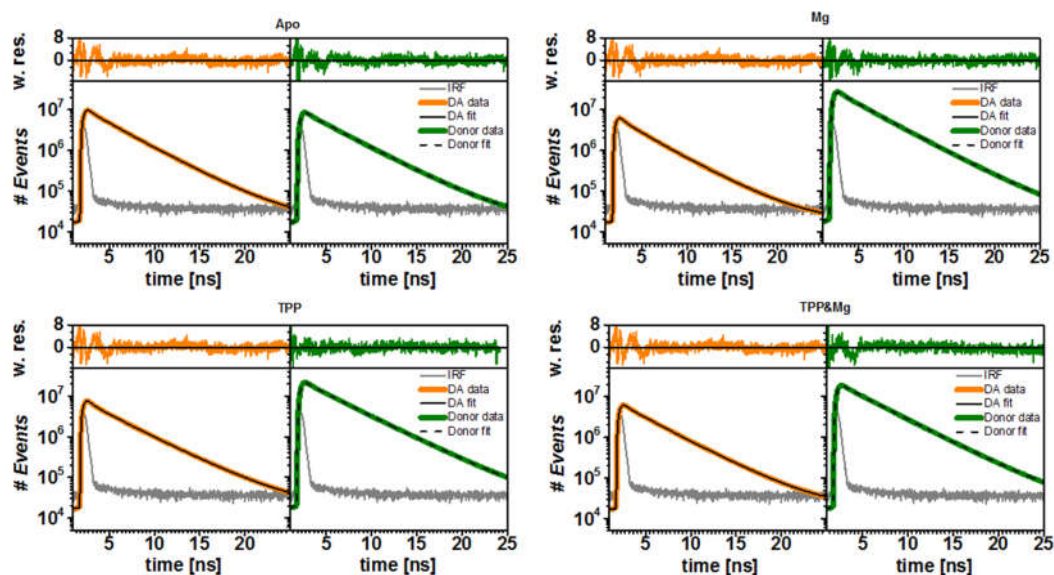
Figure 20. TPP riboswitch eTCSPC data in different conditions

53

Species	T1 (ns)	x1	T2 (ns)	x2	T3 (ns)	x3	T4 (ns)	x4	$\langle\tau\rangle_x$ (ns)	$\langle\tau\rangle_f$ (ns)	χ_r^2
Apo	3.95	0.59	2.42	0.14	0.79	0.11	0.09	0.16	2.771	3.618	2.01
TPP	3.85	0.54	2.29	0.15	0.80	0.12	0.09	0.19	2.536	3.492	2.23
Mg	4.18	0.53	2.69	0.17	0.82	0.11	0.10	0.19	2.781	3.822	2.75
TPP&Mg	4.04	0.57	2.77	0.17	0.88	0.09	0.10	0.17	2.867	3.724	2.21

Table 5. Lifetime fitting for donor only sample with TCSPC data.
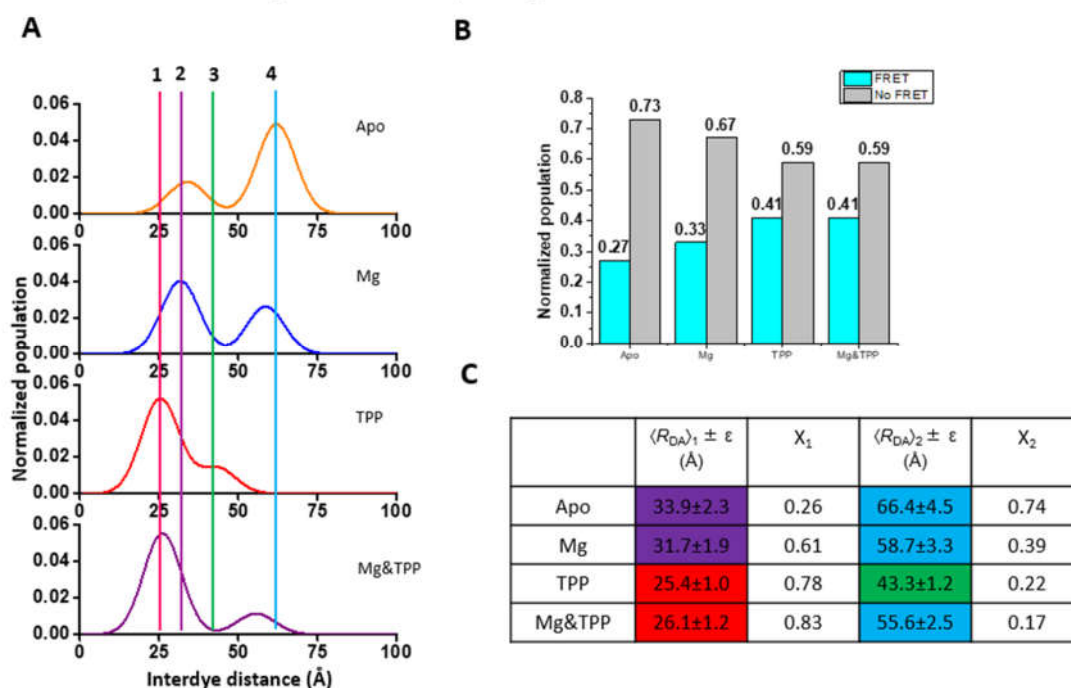
Figure 21. Determine limiting states by TCSPC. A) Interdye distance distribution with two Gassian distribution fitting model. Each color indicates different buffer.1. close state; 2. Intermediate state II; 3. Intermediate state I; 4. Semi-open/open state. B) Columns showing population of FRET and No-FRET.

Time resolved fluorescence decays (F(t)) were modeled using a multi-exponential

model which is shown as

$$F(t) = \sum_i x^{(i)} \exp(-t / \tau^{(i)})$$ (Eq. 4.1)

54

where $x^{(i)}$ is the corresponding population fraction and $\tau^{(i)}$ is the fluorescence lifetime. Here, we used two-exponential model to fit TCSPC data in order to get a relatively low χ^2.

At apo conditions, there is an equilibrium between a very open state and a close state (Fig. 21). This result in a semi-extended state with interdye distance at 66Å. With Mg present, there is an equilibrium between a semi-extended and semi-closed state with an average interdye distance at 58Å. With only TPP, the interdye distance stays at 43 Å. And with both Mg and TPP exist, the interdye distance is the shortest. The x-tal like state is highly enriched and the x-tal conformation has five Mg ion and one TPP bound. So, there are at least two intermediate states, one x-tal like configuration and a set of unfolded configurations from the data we get.
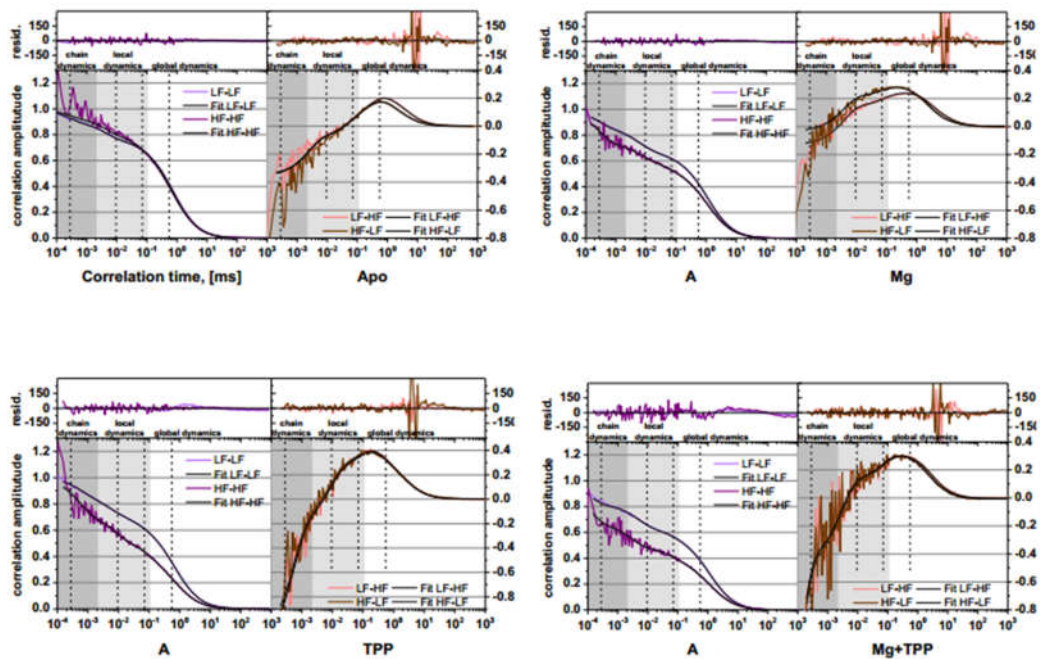
Figure 22. fFCS fitting for different conditions
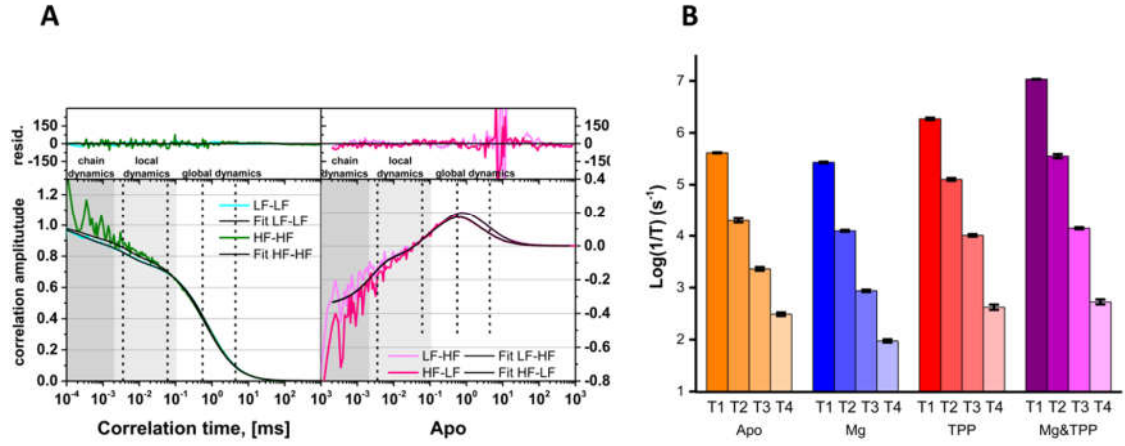
Figure 23. Four relaxation time obtained from fFCS. A) Filtered FCS species auto and cross-correlation (sACF and sCCF) function of FRET experiments for riboswitch in variant conditions. B) Four rexalation time populations in variant conditions. (From T1 to T4)

Here, we used a set of equations to fit the species into auto/cross correlation which are shown as,

$$G_{i,i}(t_c)=1+\frac{1}{N_{Br}}\cdot G_{diff}^{(i)}(t_c)\cdot\left[1-T^{(i)}+T^{(i)}\cdot \mathbf{exp}\left(-t_c/t_T^i\right)+\sum_{R=1}^{4}AC_{i,i}^{(R)}\cdot\left(\mathbf{exp}\left(-\frac{t_c}{t_R}\right)-1\right)\right]\cdot G_B^{(i)}(t_c)$$

$$G_{m,m}(t_c)=1+\frac{1}{N_{Br}}\cdot G_{diff}^{(m)}(t_c)\cdot\left[1-T^{(m)}+T^{(m)}\cdot \mathbf{exp}\left(-t_c/t_T^m\right)+\sum_{R=1}^{4}AC_{m,m}^{(R)}\cdot\left(\mathbf{exp}\left(-\frac{t_c}{t_R}\right)-1\right)\right]\cdot G_B^{(m)}(t_c)$$

$$G_{i,m}(t_c)=1+\frac{1}{N_{CC}}\cdot G_{diff}^{(i,m)}(t_c)\cdot\left[1-CC_{i,m}\cdot\sum_{R=1}^{4}X_{i,m}^{(R)}\cdot \mathbf{exp}\left(-\frac{t_c}{t_R}\right)\right]\cdot\left(1-B_{i,m}\cdot \mathbf{exp}\left(-\frac{t_c}{t_B}\right)\right)$$

(Eq. 4.2)

Where t_R are the relaxation times which are referred as the exchange times between selected species with corresponding absolute amplitudes of the sACF$^{(R)}$ and the relative normalized amplitudes of the sCCF*T$^{(x)}$ is the triplet amplitude. Here, we used four relaxation time for both auto and cross correlation fitting in order to get a relatively good fit to simply kinetic model.

56

Species	T1(ms)	T2(ms)	T3(ms)	T4(ms)
Apo	0.0024241	0.0502355	0.434764	3.24609
Mg	0.0037027	0.0800246	1.15034	10.5978
TPP	0.000537459	0.00790622	0.0983016	2.3812
Mg&TPP	0.00009284	0.00281428	0.0715076	1.88312

Fractions	x1	x2	x3	x4
Apo	0.194	0.089	0.202	0.515
Mg	0.235	0.084	0.219	0.462
TPP	0.45	0.205	0.121	0.224
Mg&TPP	0.628	0.148	0.051	0.173

Table 6. Relaxation times and its amplitude for fFCS fitting.

We performed fFCS experiments (Fig. 22) in order to track the relaxation times under ms timescale. For all the conditions, we find four relaxation times faster than the diffusion time which is consistent with MFD experiments. These relaxation times range from 100ns to ms. It gives a chevron like plot (Figure. 23) behavior if the order is from folded to unfolded state. From the data we get, five different conformational states are needed in order to satisfy the fast dynamics model that we apply. We observe chain dynamics which is around 100 ns. This can be explained by the situation that riboswitch has an "unfolded state" sampling very fast energy level which are very similar. Then, we observe large conformation changes from 100 microsecond to ms which reveals the changes between open and close state.

4.3) Conclusion

Here, we present the kinetic scheme and energy landscape for TPP riboswitch.
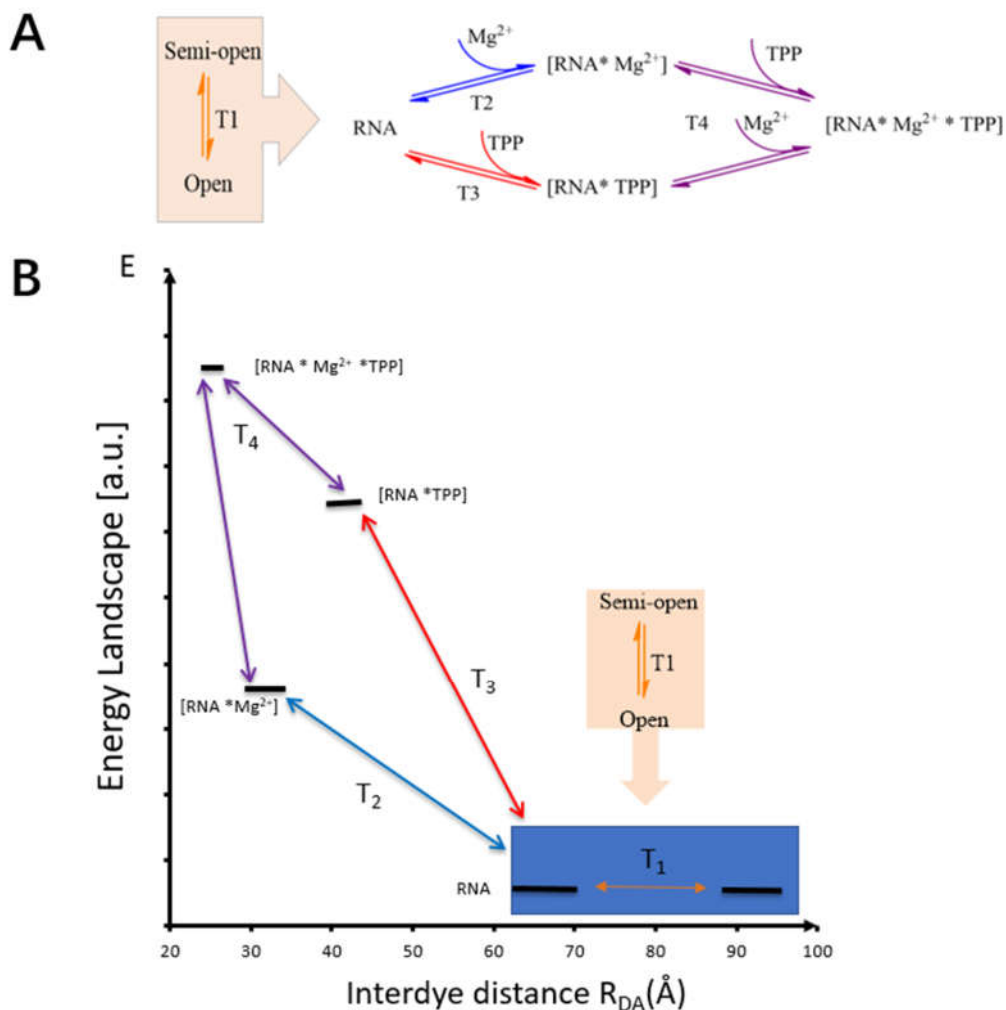
Figure 24. Predicted kinetic pathway and energy landscape. A) Two pathways from open/semi-open state to the close state. B) Energy landscape showing the mechanism of how TPP binding with riboswitch.

RNA(semi-open) and RNA(open) are two exit pathway (Fig. 24) which go from the freely diffused state to the final x-tal state. They share similar energy levels and exchange very fast. These two are extended (very low FRET or no FRET). RNA(Mg^{2+}) corresponds to an intermediate state which has a fast dynamic motion. RNA(TPP) corresponds to another intermediate state which has a larger conformation change.

58

In apo state, relaxation time T1 is fastest which means energy barrier is low between RNA(semi-open) and RNA(open). Also, it has the largest none-FRET population. Relaxation time T2 stands for the process of binding with metal ions which leads to a semi-closed state. We can see from figure that this deceases None-FRET population as well as the semi-open state. Moreover, this induces an intermediate state I2 with Mg^{2+} ion present. Relaxation time T3 is slower than T2 which means a larger conformation change occurs in this state. It is very obviously to see that semi-open state is completely gone with a significant increase in the FRET population. Also, both close state and an intermediate state I1 present with TPP present in buffer. T4 is the slowest relaxation time which shows the largest population of close state. Here, the None-FRET population decease combined with an increase in the energy barrier with both Mg^{2+} ion and TPP exist in buffer.

CHAPTER FIVE

PDZ PROJECT

5.1) Introduction

In this chapter, we utilize FRET along with MD simulation to study the dynamic and structure information of a type protein domains called PDZ. Multidomain proteins, whose folded domains act as the primary sequence and whose domains are thought to interact with one another, gave birth to the term supertertiary structures[98]; particularly when each stable structure has specific functional implications [99] [100]. Therefore, these interdomain interactions, which must be driven by long-range interactions, could find various stable configurations. Here, we identify two supramodule configurations [101] of a tandem of two 95 kDa/Disc-large/ZO-1 (PDZ) domains, which dynamically exchange at timescales faster than milliseconds. Our conclusions, arises from integrating, ensemble and single molecule Förster Resonance Energy methods with a temporal resolution that spans from ps to seconds, and compare them with replica exchange molecular dynamic simulations that sample a large enough configuration space[102]. The supramodule configuration with shorter interdomain distance is driven by hydrophobic interactions, while the supramodlue configuration with longer interdomain distance is mostly driven by charged amino acids found at the interface. However, the low energy barrier between these configurations is indicative of mainly weak interactions. Therefore, the tandem can sample a large configuration space at multiple timescales.

Polypeptides typically fold into a three-dimensional structure which is driven by the intramolecular interactions among its native sequences. High affinity interactions as well as other intramolecular interactions can result in a relatively static biomolecular conformation. However, relatively weak interactions within molecules can permit a dynamic ensemble of alternate conformations, which can be hard to predict as well as a challenge to identify experimentally. Our current understanding of protein folding has advanced to the point where the prediction of tertiary structure is quite accurate for small protein domains. However, many proteins composed of a series of independently-folded domain[103]s that subsequently assemble into a supertertiary structure. For such, multidomain proteins, the folded subdomains act as the primary sequence and their intramolecular interactions give rise to the supertertiary structure.

The same forces govern the folding of individual domains and multidomain proteins[104,105]; however, as surface of folded domains are generally polar, the affinity of intramolecular interactions can be much lower. This can result in a dynamic supertertiary structure capable of different spatial arrangements[106] that interconvert between similar free energy states on a range of timescales. Despite their prevalence, there are few methods to predict the supertertiary structure of multidomain proteins even when the structure of individual subdomains is known. Because of the long length and timescales sampled by such dynamic assemblies, there is no single experimental technique that can fully capture supertertiary protein dynamics and such proteins are a challenge for simulations.

The postsynaptic density protein of 95 (PSD-95, Fig. 25A) is a prototypical multidomain scaffolding protein found in the postsynapse of excitatory neurons. PSD-95 contains five protein-interaction domains connected by intrinsically-disordered linkers. The structures all five domains are known to high resolution but there is little knowledge of how to assemble these pieces into a model for the full-length protein. Single molecule fluorescence measurements confirmed that the supertertiary structure was dynamic, but the time resolution was insufficient to identify and limiting conformational states.

There is great interest in the two N-terminal PDZ (PSD-95/Dlg/ZO-1) domains [107], which are the target of pharmaceutical compounds in clinical trials as a treatment for ischemic stroke. As such, there has been intense interest in the supertertiary structure of the PDZ tandem, which has been studied by NMR, x-ray crystallography and single molecule fluorescence.

The relative orientation of tandem PDZ domains has been suggested to influence their binding specificity and interactions [108]. Studies of tandem PDZ domains have revealed a range of interdomain affinities resulting in interdomain orientation that could be dynamic or static. Some PDZ tandems have minimal interdomain interactions despite a short linker. For example, the Mint2 PDZ tandem [109] was sufficiently dynamic that only the first domain of the tandem was seen as ordered within a protein crystal. In contrast, intramolecular interactions within Mint1 lock the PDZ tandem into a single conformation resulting in Mint1 autoinhibition. Thus, modulation supertertiary dynamics can be an important regulatory mechanism for binding affinity in PDZ tandems.

The N-terminal PDZ domains from PSD-95, which are separated by a 10-residue linker, have been suggested to function as a "supramodule" capable of binding dimeric ligands [110] [111]. NMR studies suggested that the PDZ domainshad limited rotation with respect to one another. Ligand binding by the PDZ domains unlocked the conformation suggesting that interdomain interactions may be autoinhibitory. Unfortunately, NMR wasn't able to identify any limiting states [112], as a fact of lacking of identifiable interdomain contacts (Fig. 25B). That the PDZ tandem could be crystalized argues for limited dynamics. However, the crystal revealed two conformations that both lacked interdomain contacts (Fig. 25B). Single molecule Förster/fluorescence resonance energy transfer (smFRET) measurements [46] [49] were compatible with a Subsequent single molecule fluorescence resonance energy transfer (smFRET) measurements, performed with camera detection at 10Hz, suggested a "mean occupied" conformation for the tandem (still without interdomain contacts. Fig. 25B) but also detected conformational dynamics
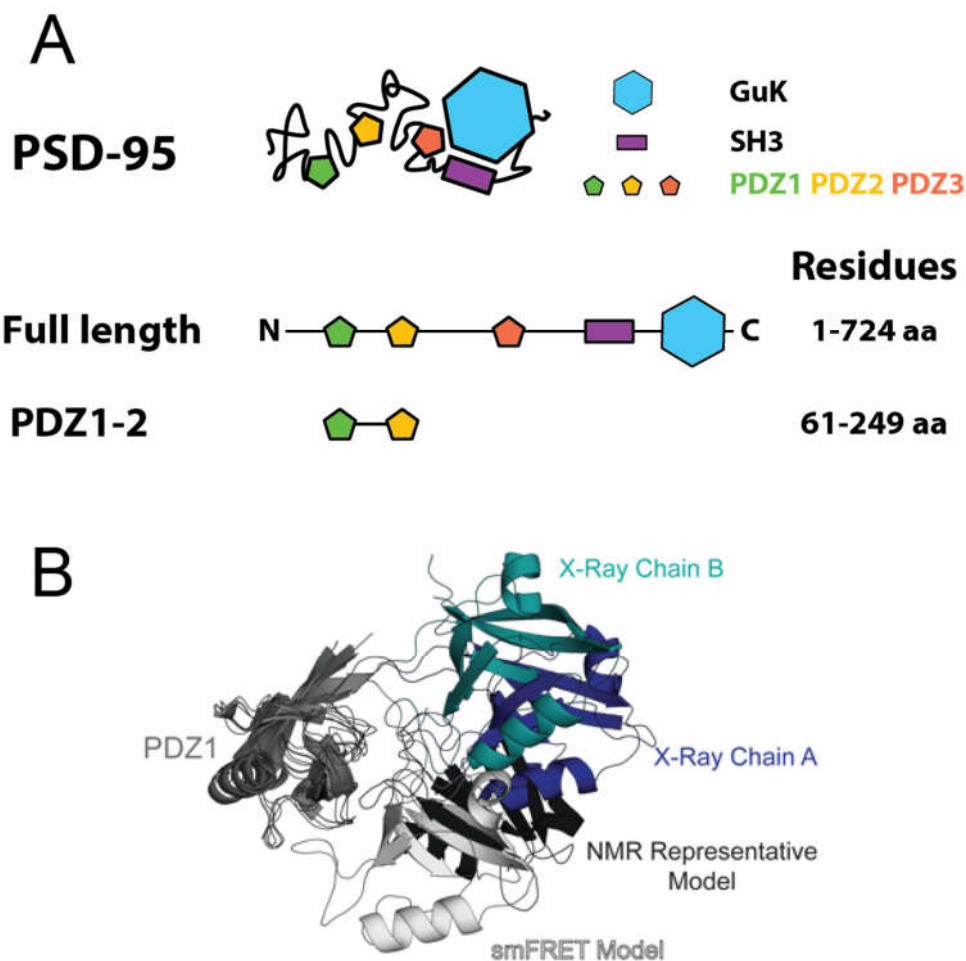
Figure 25. The PDZ1-2 tandem from PSD-95. **A**) Domain organization in PSD-95, which contains five protein interaction domains connected by linkers of varying length. The N-terminal PDZ1-2 tandem contains the first two PDZ domains corresponding to residues 61-249. **B**) Cartoon representation of structural models of the PDZ-1-2 tandem as derived by different methods. PDZ1 (grey) from each model was aligned to highlight conformational differences. PDZ2 is colored according the method used to resolve the structure. Models from the crystal structure (blue and cyan) [PDB ID: 3GSL].

To better understand supertertiary PDZ domain interactions[113-115], we revisited smFRET experiments[116] on the PSD-95 PDZ tandem with a orders of magnitude faster time resolution, at both ensemble and single molecule conditions, to identify any limiting states. Additionally, we used Discrete Molecular Dynamics (DMD) simulations to map

the energy landscape and identify any interdomain interactions[117-119]. The combination of these methods provided us a unique sampling spanning ~12 orders of magnitude in time with spatial distance precision in the Å regime. Both experiments and simulations suggested two limiting conformations for the PDZ tandem: a closed conformation driven by hydrophobic interactions and a more open conformation with charged amino acids at the interface. However, the energy barrier between these configurations is shallow, which permits the sampling of a large conformational space.

5.2) Results

In order to probe the conformation of the PDZ tandem, we used a set of ten protein variants[120-122] each containing two unique cysteines, one in each PDZ domain (Fig. 26A). The labeling positions were chosen based on surface accessibility of the selected residues and visually maximizing the spatial distribution of labeling sites across the PDZ domains (Fig. 26B). FRET depends of the distance between fluorescent dyes while our goal is to use this information to model the protein structure. To define the position of the fluorophore relative to the protein backbone, the fluorescent dyes were modeled to estimate the accessible volume (AV). The AV simulations were coarse grained in that the fluorescent dye was represented by an ellipsoid while the flexible linker was represented by a cylinder.

For experiments, proteins were randomly labeled the FRET donor (Alexa Fluor 488) and the FRET acceptor (Alexa Fluor 647), which have a Förster radius (R_0) of 52 Å.

First, the protein was sub-stoichiometrically labeled with donor. An aliquot of this reaction was then quenched as used as a paired donor-only sample before adding an access of the acceptor to the remainder. The final labeling ratio was approxiamtely 1:4 donor to acceptor. This strategy minimizes the creation of protein doubly-labeled with the donor, which complicates fluorescence measurements[123,124]. Despite this, the samples still showed the presence of donor-only protein presumably due to labeling with inactive acceptor.
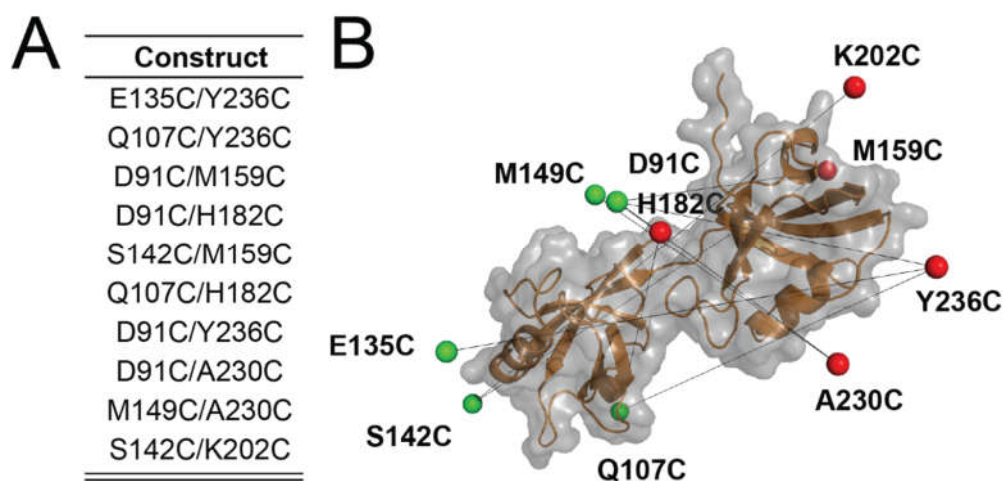
Figure 26. Network of FRET Restraints in the PDZ Tandem. A) Position of cysteine residues in the PDZ tandem (Residues 60-250 of PSD-95) for the ten FRET variants used in this study. The first residue is in PDZ1 while the second is in PDZ2. B) Mean position of dyes as taken from modeling of the accessible volume. (PDBID: 3GSL)

We first measured ensemble time correlated single photon counting (eTCSPC) for all 10 FRET variants to identify the number of limiting states and assign distances based on the rate of energy transfer (Fig. 27). By using a pulsed laser, eTCSPC can identify structural states that are stable on the timescale of fluorescence emission (ns), which should be faster than the rigid body dynamics of entire protein domains (μs- ms). We first

66

analyzed the time resolved fluorescence of the donor with a multi exponential decay model rather than fitting the decay with simple exponentials, which gives incorrect information. The information about the fluorescence decay for the donor-only sample is then used to fit the fluorescence decay of the DA sample under the assumption that FRET only changes with rate of energy transfer without affecting the distribution of energy states. Rather than extract the fluorescence lifetime, we directly fit the fluorescence decay using an interdye distance distribution modeled as two Gaussian distributed states, where the width of the Gaussian distribution accounts for dye mobility within the AV. Benchmark studies have suggested that a standard deviation of 6 Å reasonably represents dye mobility as well as distributions obtained by simulations of the dye. Thus, the eTCSPC results can be represented by the mean dye separation for each state with an associated width representing dye mobility (Fig. 27).

We systematically compared several models to fit the time resolved fluorescence decays. We fit the decays with one or two singular FRET states and compared this to fits with one or two Gaussian distributed states. We also examined the addition of a no FRET or donor-only state. Models were compared by assessing the figure of merit χ^2. A comparison of the $\Delta\chi^2$ for the different models revealed that a single Gaussian distributed state was insufficient to describe the fluorescence decay. This suggested the presence of two distinct lifetimes for each of the ten variants in addition to a fraction of donor-only.
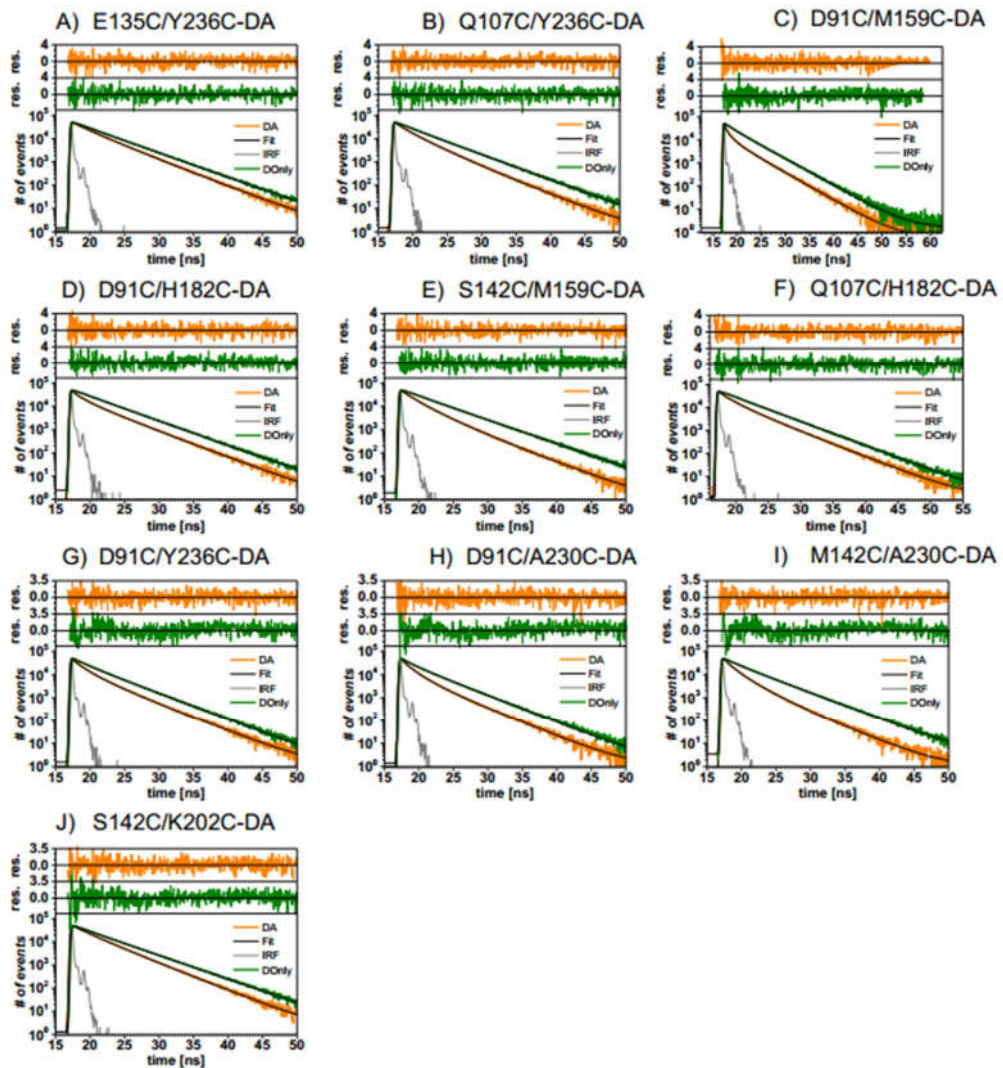
Figure 27. Time Resolved Fluorescence of labeled PDZ tandems. Time Resolved Fluorescence of donor only (DOnly) and donor-acceptor (DA) labeled PDZ1-2 tandems. DOnly decays shown in green, and corresponding black overlay uses the Eq. 1. DA decays are shown in orange and model fit function uses Eq. 3 to describe the Gaussian distribution of states. Instrument response function (IRF) is shown in gray. A) E135C/Y236C-DA B) Q107C/Y236C-DA C) D91C/M159C-DA D) D91C/H182C-DA, E) S142C/M159C-DA, F) Q107C/H182C-DA, G) D91C/Y236C-DA, H) D91C/A230C-DA, I) M142C/A230C-DA, J) S142C/K202C-DA . Table 1A 1B summarizes the fit results.

Based on Förster theory, we calculated the expected correlation between intensity

and lifetime for a static molecule: the 'static' FRET line (red lines in Fig. 28) as well as

the expected correlation for molecules making transitions between each of the 3 states

identified by eTCSPC (green lines in Fig. 28). The mean interdye distances R_{DA} for the

major and minor populations are marked by purple blue lines, respectively. The dynamic

FRET lines illustrate the interconnecting pathways between these states. Thus, any FRET

species in a fixed conformation will fall along the red, static-FRET line. If the molecule

deviates from this line, it is an indication of dynamic averaging. Dynamic averaging

shifts the population to larger $<\tau_{D(A)}>_f$ values when the dynamics change the FRET

efficiency. In contrast, dynamics due to photophysical artifacts shift $<\tau D(A)>_f$ to smaller

values [48]. Shot noise provides statistical fluctuations resulting on the width of the

distribution in both axes.

For ease of interpretation, we display the fluorescence intensity with two FRET

indicators: FD/FA and FRET Efficiency, which are plotted against the average

fluorescence lifetime per single molecule $<\tau D(A)>_f$. FD/FA is only corrected for

background counts and the detection efficiency ratio. FRET efficiency includes an

additional correction for the ratio of donor and acceptor quantum yields. For all 10

variants, most of the single molecule events lie within the area enclosed by the dynamic

FRET lines, which few events falling along the static FRET lines (Fig. 28). The

frequency histograms, atop and aside the main panel, display a single distribution for all

variants, which in several cases is asymmetric. The existence of a single peak indicates

dynamic averaging on the sub-millisecond timescale as the limiting states are not visible.

69

In some cases, such as the Q107C/H182C variant (Fig. 26A) one would be tempted to consider this a single static FRET state; however, the peak of the distribution lies to the right of the red line near the limiting state from eTCSPC. This indicates the interconversion of states and is clear evidence of dynamic averaging.

As reported for TIRF mode, the MFD histograms show FRET values across a broad range (Figs. 28). The Q107C/H182C restraint is close to mid FRET while D91C/M159C is centered at high FRET, and S142C/K202C-DA shows low FRET (Fig. 4, lower panel). In these cases the equilibrium lays closer to the major state identified from eTCSPC (Fig. 28, purple lines). For all constructs, the peak in the 2D plots falls off the static FRET line indicating dynamics (Fig. 28, red line). Additionally, the lifetime distributions are narrow for most measurements, which indicates dynamics faster than the measurement timescale. The individual peaks fall at various locations with respect to the dynamic FRET lines that connect the limiting states (Fig. 28, green lines). This indicates that not all sites are responding equally to the underlying transitions.
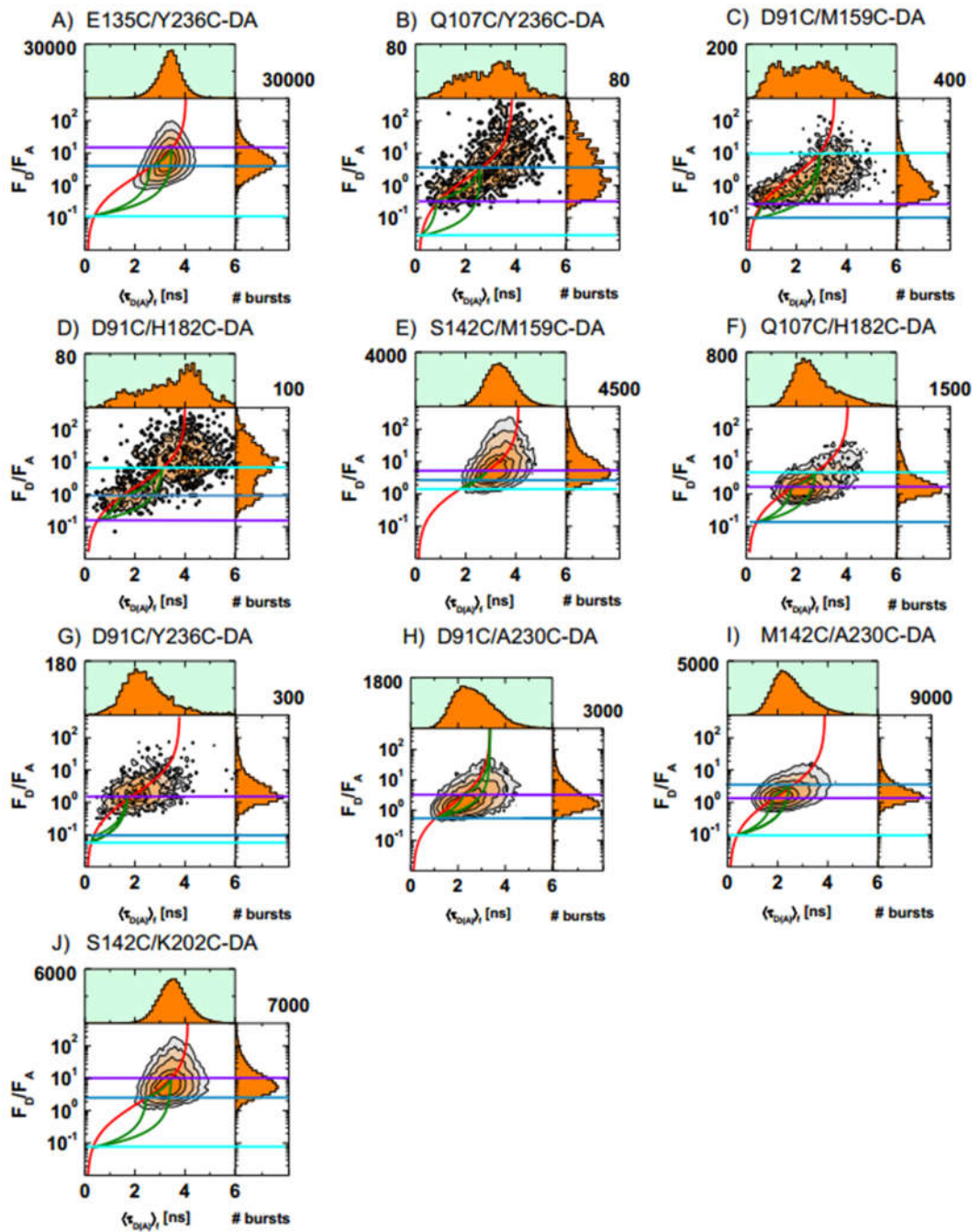
Figure 28. Multiparameter Fluorescence Detection histograms. A) E135C/Y236C-DA B) Q107C/Y236C-DA C) D91C/M159C-DA D) D91C/H182C-DA, E) S142C/M159C-DA, F) Q107C/H182C-DA, G) D91C/Y236C-DA, H) D91C/A230C-DA, I) M149C/A230C-DA, J) S142C/K202C-DA. Table 1A 1B summarizes the fit results. Horizontal guidelines

71

for the major and minor states 1 and 2 are shown in purple and blue, respectively. Red and green lines correspond to the static and dynamic FRET lines.

To get an unbiased estimate of the conformational dynamic[125]s of the PDZ tandem, we performed molecular dynamics (MD) simulations, which could be compared with the experimental results. Given the large molecular size and long timescale associated with the inter-domain dynamics, we used the atomistic discrete molecular dynamics (DMD), a rapid MD algorithm which has been benchmarked with *ab inito* folding of several fast-folding proteins and successfully applied to engineer *de novo* allosterically-regulated proteins by capturing their coupled conformational dynamics. Replica exchange simulations were used to ensure sufficient sampling of the conformational dynamics of the PDZ tandem. We used the smFRET model as the starting conformation. To model the FRET restraints [126], we added non-interactive dyes at all the labeling locations. We imposed a structure-based potential constrain for each PDZ subdomain to prevent unfolding. Eight replica DMD simulations[127], each lasting 100 ns, were performed with an accumulative simulation time of ~0.8 µs. Based on the replica exchange trajectories, we applied the weighted histogram analysis method (WHAM) to analyze the interdomain conformational dynamics and estimate the potential mean force (PMF) at the room temperature, ~300 K.

Using the distance between centers of mass of the two PDZ domains as a reaction coordinate, we observed two distinct minima in the distribution domain separation, in agreement with two limiting states observed by eTCSPC (Fig. 31). In experiments, we only observed the interdye distance so we also extracted the interdye distance from the

conjugated fluorescent dyes present in the simulation. To represent the energy landscape with respect to the interdye distance, we calculated the distance distribution ($p(R_{DA})_{DMD}$) by computing where E_{PMF} is the integrated free energy over interdomain distance as the function of the interdye distance, R_{DA}; and k_B is Boltzmann constant. The resulting two-dimensional PMF plots represent the free energy landscape (Figs. 30). The computed free energy landscape (Fig. 29) indicates that the two states have a low free energy barrier that allows for fast exchange with each other, while their separation from the rest is larger. While the two states are reasonably well separated in terms of interdomain distance, the projected dye distributions are substantially overlapped for most constructs. Thus, only a handful of the FRET restraints would be capable of resolving these two states.

Table 7. Time Resolved fit results
A) Donor Acceptor Distances and fractions as model with a three Gaussian distributed and the addition of a No FRET state (Eq. S4). There is a major population representing the 43.5%. The second most populated accounts for 32.4%, and the least populated corresponds to 24.1%.

Sample	D_A [Å]	$\langle R_{DA}\rangle_1 \pm \varepsilon$ [Å]	DA [Å]	$\langle R_{DA}\rangle_2 \pm \varepsilon$ [Å]	DA [Å]	$\langle R_{DA}\rangle_3 \pm \varepsilon$ [Å]	No FRET	χ_r^2
E135C/Y236C-DA	2.6	74.8 ± 1.8	8.7	57.7 ± 1.7	5.8	32.6 ± 0.8	0.28	1.233
Q107C/Y236C-DA	1.3	37.8 ± 0.9	8.4	57.8 ± 1.4	0.7	15.7 ± 1.1	0.15	1.095
D91C/M159C-DA	9.1	37.0 ± 1.2	5.0	31.9 ±1.0	3.0	85.2 ± 2.6	0.09	1.104
D91C/H182C-DA	8.0	34.5 ± 1.1	6.1	44.9 ± 1.4	8.2	65.7 ± 2.1	0.21	1.206
S142C/M159C-DA	0.5	60.3 ± 1.1	3.4	53.1 ± 0.9	13.9	47.8 ± 0.8	0.09	1.078

Sample							
Q107C/H182C-DA	4.4	50.8 ± 0.9	4.5	33.0 ± 0.6	4.8	81.7 ± 1.4	0.12	1.183
D91C/Y236C-DA	3.9	54.3 ± 2.0	15.5	31.8 ± 1.1	2.8	29.7 ± 1.1	0.22	1.083
D91C/A230C-DA	10.6	38.0 ± 0.9	3.7	53.1 ± 1.3	2.7	105.7 ± 2.7	0.11	1.183
M149C/A230C-DA	3.3	49.6 ± 0.5	2.3	59.2 ± 0.6	11.1	31.8 ± 0.3	0.07	1.078
S142C/K202C-DA	5.7	75.4 ± 2.3	4.5	56.1 ± 1.7	1.9	31.2 ± 1.0	0.00	1.290

B) Donor only Lifetime Decay $\Phi_{FD(0)}$ determination is described in methods section.

Sample	x_1	τ_1 [ns]	x_2	τ_2 [ns]	$\langle \tau \rangle_x$ [ns]	$\langle \tau \rangle_f$ [ns]	$\Phi_{D(0)}$	χ_r^2
E135C/Y236C-DA	0.92	4.13	0.08	1.24	3.90	4.06	0.78	1.172
Q107C/Y236C-DA	0.87	3.99	0.13	1.46	3.65	3.85	0.73	1.288
D91C/M159C-DA	0.77	3.76	0.23	1.17	3.16	3.54	0.63	1.386
D91C/H182C-DA	0.92	4.10	0.08	1.10	3.86	4.03	0.77	1.361
S142C/M159C-DA	0.90	4.24	0.10	1.54	3.98	4.14	0.80	1.343
Q107C/H182C-DA	0.94	4.14	0.06	1.15	2.85	4.09	0.79	1.295
D91C/Y236C-DA	0.74	3.29	0.26	1.03	2.71	3.07	0.54	1.556
D91C/A230C-DA	0.66	3.72	0.34	1.15	2.85	3.37	0.57	1.473
M149C/A230C-DA	0.76	3.85	0.24	1.13	3.20	3.62	0.64	1.711
S142C/K202C-DA	0.89	4.22	0.11	1.05	3.86	4.12	0.77	1.674

C) Acceptor Only Lifetime Decay Φ_{FA} determination is described in methods section

Sample	x_1	τ_1	x_2	τ_2	$\langle \tau \rangle_x$ [ns]	$\langle \tau \rangle_f$ [ns]	Φ_{FA}	χ_r^2
E135C/Y236C-DA	0.68	1.54	0.32	0.77	1.30	1.40	0.35	1.174
Q107C/Y236C-DA	0.48	1.71	0.52	1.19	1.44	1.49	0.39	1.748
D91C/M159C-DA	0.60	1.65	0.40	0.72	1.28	1.44	0.35	1.849
D91C/H182C-DA	0.66	1.73	0.34	0.72	1.38	1.55	0.38	1.815
S142C/M159C-DA	0.58	1.53	0.42	0.86	1.25	1.34	0.34	1.447
Q107C/H182C-DA	0.59	1.57	0.41	0.67	1.20	1.36	0.33	0.885
D91C/Y236C-DA	0.65	1.69	0.35	0.86	1.40	1.51	0.38	1.643
D91C/A230C-DA	0.62	1.83	0.38	0.93	1.49	1.62	0.41	1.423
M149C/A230C-DA	0.66	1.79	0.34	0.90	1.49	1.61	0.41	1.413
S142C/K202C-DA	0.60	1.91	0.40	1.00	1.54	1.67	0.42	1.423

Table 8. Time Resolved Anisotropy

A) Donor Only anisotropy.

Sample	b_1	τ_1 [ns]	b_2	τ_2 [ns]	b_3	τ_3 [ns]	$\chi^2_{r,sum}$	$\chi^2_{r,diff}$	r_{ss} +/- std
E135C/Y236C-DA	0.12	0.33	0.09	1.99	0.06	12.24	2.86	1.03	0.1 ±0.01
Q107C/Y236C-DA	0.16	0.31	0.11	2.32	0.12	37.00	2.90	1.14	0.2 ±0.00
D91C/M159C-DA	0.10	0.22	0.09	1.27	0.17	8.81	1.69	1.10	0.2±0.00
D91C/H182C-DA	0.12	0.35	0.08	1.91	0.15	10.97	2.59	0.95	0.2 ±0.00
S142C/M159C-DA	0.15	0.29	0.10	2.06	0.08	14.00	4.40	1.51	0.1 ±0.02
Q107C/H182C-DA	0.25	0.22	0.10	2.63	0.03	13.49	4.17	1.51	0.1 ±0.00
D91C/Y236C-DA	0.14	0.62			0.16	7.39	2.23	1.16	0.1 ±0.01
D91C/A230C-DA	0.11	0.35	0.12	2.89	0.12	15.47	3.42	1.28	0.2 ±0.01
M149C/A230C-DA	0.17	0.36	0.07	2.59	0.10	14.94	3.51	1.32	0.1 ±0.02
S142C/K202C-DA	0.14	0.27	0.09	1.45	0.10	7.91	3.68	1.33	0.1 ±0.01

B) Direct excitation of the Acceptor anisotropy

Sample	b_1	τ_1 [ns]	b_2	τ_2 [ns]	b_3	τ_3 [ns]			r_{ss} +/- std
E135C/Y236C-DA	0.19	0.58	0.15	6.94			9.28	1.74	0.2 ±0.01
Q107C/Y236C-DA	0.17	0.48	0.12	6.04	0.05	5691	6.55	1.36	0.2 ±0.06
D91C/M159C-DA	0.14	0.64	0.19	9.83			6.14	1.33	0.2 ±0.02
D91C/H182C-DA	0.15	0.56	0.22	13.79			31.58	4.52	0.2 ±0.02
S142C/M159C-DA	0.15	0.44	0.17	8.81			2.78	1.30	0.2 ±0.03
Q107C/H182C-DA	0.26	0.43	0.08	18.24			3.17	1.18	0.1 ±0.01
D91C/Y236C-DA	0.12	0.41	0.17	7.46			3.16	1.29	0.2 ±0.03
D91C/A230C-DA	0.11	0.39	0.19	8.94			3.52	1.11	0.2 ±0.02
M149C/A230C-DA	0.17	0.19	0.19	8.28			4.35	1.37	0.2 ±0.05
S142C/K202C-DA	0.14	0.43	0.16	6.97			5.03	1.82	0.2 ±0.04

C) Sensitized by FRET acceptor anisotropy

Sample	b_1	τ_1 [ns]	b_2	τ_2 [ns]	b_∞	τ_∞ [ns]	$\chi^2_{r,sum}$	$\chi^2_{r,diff}$	r_{ss} +/- std
E135C/Y236C-DA	0.22	0.16	0.10	1.28	-0.01	285	3.31	1.10	0.0 ±0.03
Q107C/Y236C-DA	0.10	0.32	0.04	1.78	-0.03	61887	2.44	1.11	0.0 ±0.08
D91C/M159C-DA	0.003	0.31			0.005	15	1.79	1.08	0.0 ±0.03
D91C/H182C-DA	0.02	0.24			0.007	154633	1.84	1.07	0.0 ±0.02
S142C/M159C-DA	0.07	0.40			0.03	18	2.78	0.98	0.0 ±0.11

75

Q107C/H182C-DA	0.09	0.34	0.01	5	2.50	0.98	0.0 ±0.03
D91C/Y236C-DA	0.04	0.59	-0.03	59417	2.43	1.05	-0.0±0.07
D91C/A230C-DA	0.04	0.28	0.02	3	2.28	0.96	0.0 ±0.07
M149C/A230C-DA	0.06	0.51	0.08	45	1.98	1.01	0.1 ±0.14
S142C/K202C-DA	0.10	0.82	-0.03	164115	3.34	1.19	-0.0±0.08
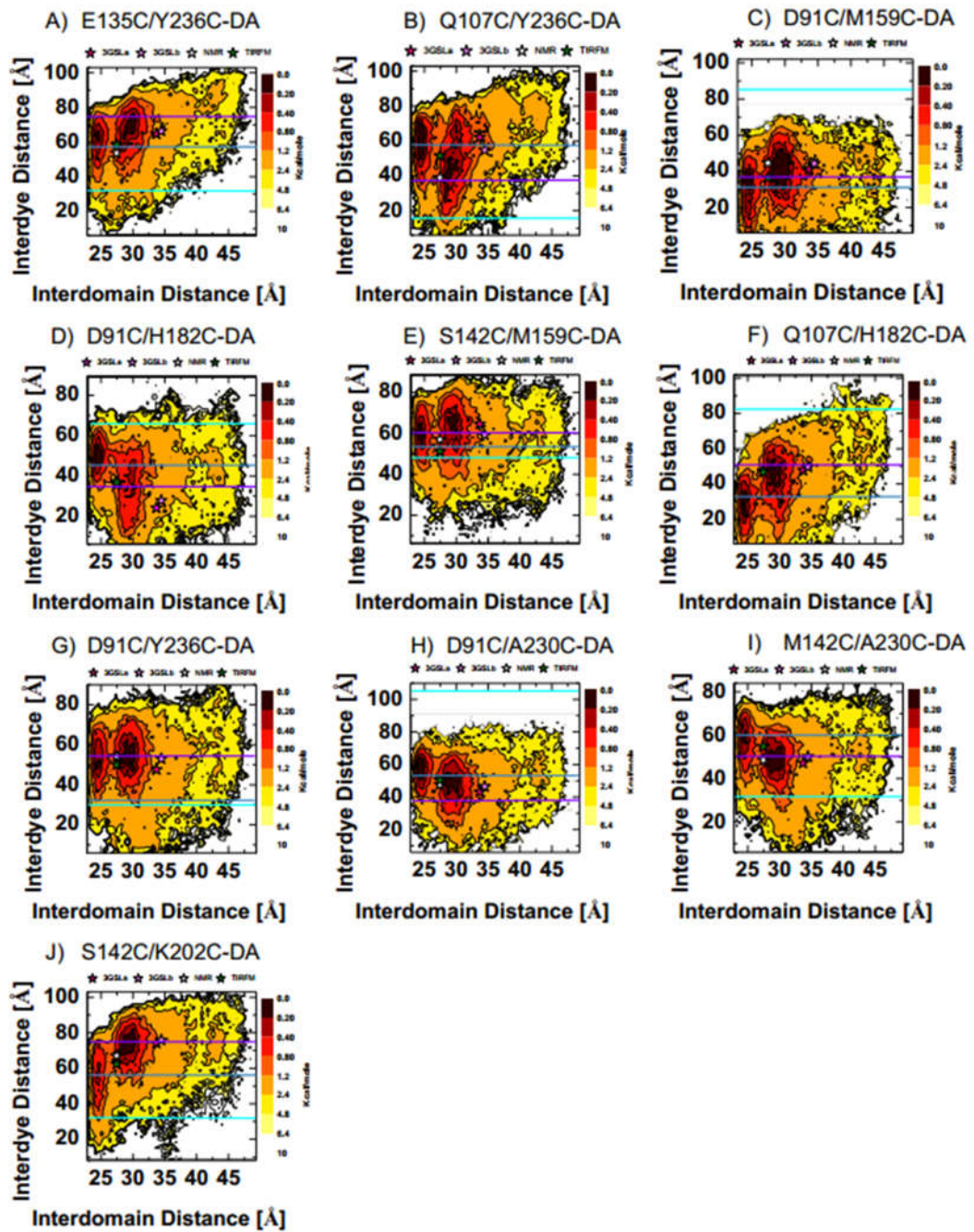
Figure 29. Energy landscape of PDZ1-2 tandem. A) E135C/Y236C-DA B) Q107C/Y236C-DA C) D91C/M159C-DA D) D91C/H182C-DA, E) S142C/M159C-DA, F) Q107C/H182C-DA, G) D91C/Y236C-DA, H) D91C/A230C-DA, I) M149C/A230C-DA, J) S142C/K202C-DA. Table 1A 1B summarizes the fit results. Horizontal guidelines

for the mean interdye states determined by eTCSPC are shown in purple and blue, respectively. Symbols represent the X-ray, NMR and TIRFM structural models.

Figure 30. Comparison of Interdye distribution as derived from DMD simulation.

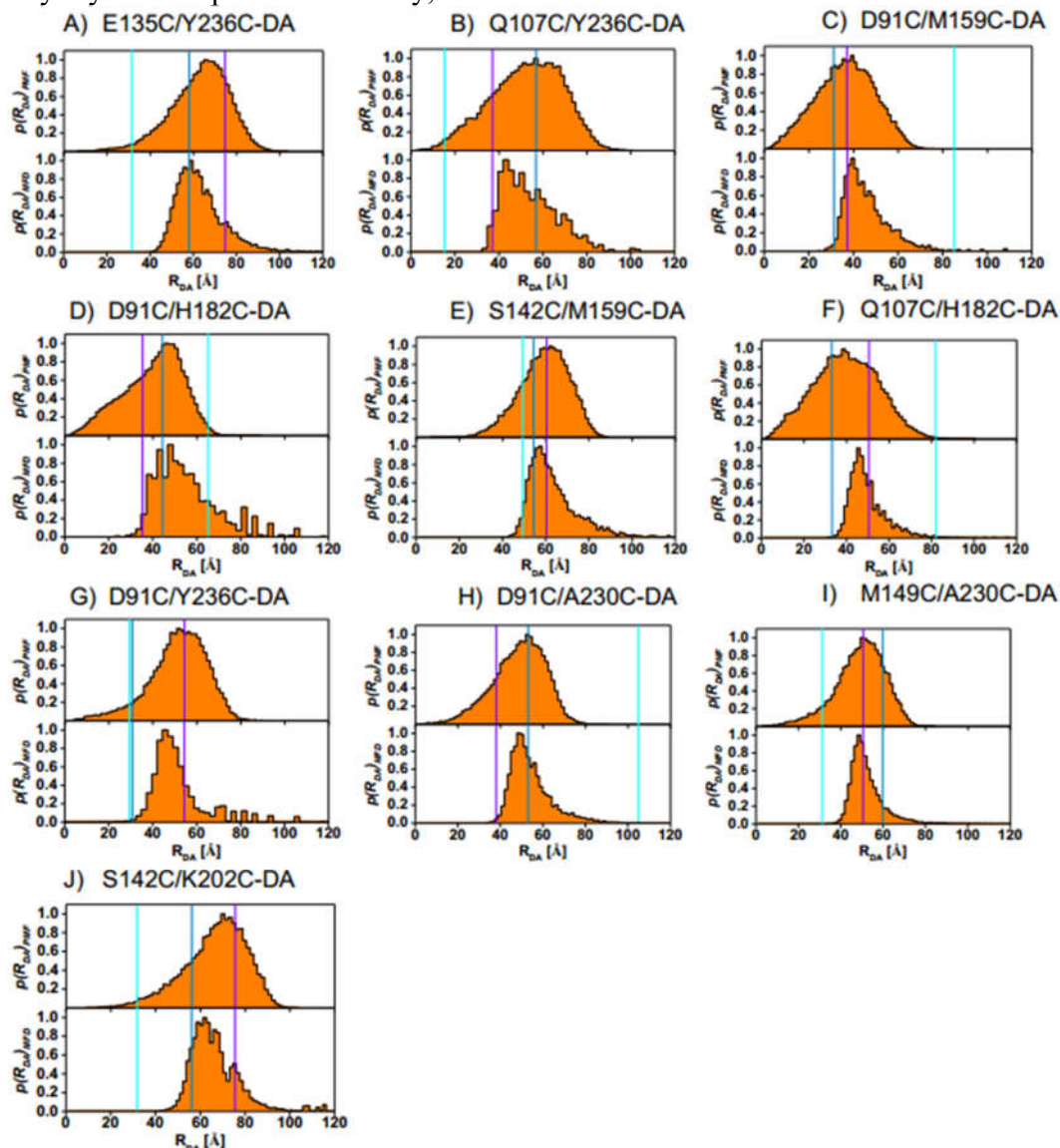$$p(R_{DA})_{PMF} = exp\left(-\frac{E_{PMF}}{0.6}\right)$$ ()nd MFD. A) E135C/Y236C-DA B) Q107C/Y236C-DA C) D91C/M159C-DA D) D91C/H182C-DA, E) S142C/M159C-DA, F) Q107C/H182C-DA, G) D91C/Y236C-DA, H) D91C/A230C-DA, I) M149C/A230C-DA, J) S142C/K202C-DA.

We extracted the simulated interdye distances $p(R_{DA})_{DMD}$ and fit the distributions with two Gaussian distributed states (Fig 31D-F, black and gray lines). As with the MFD, the distances showed varying proportions of two states; however the states show varying degrees of agreement with the observed eTCSPC distances (Fig. 31 purple lines). For Q107C/H182C the experimentally derived distances were $\langle R_{DA} \rangle_1$=54.7 Å and $\langle R_{DA} \rangle_2$=36.1 Å, while the mean of the two Gaussian distributions that reproduce $p(R_{DA})_{DMD}$ are $\langle R_{DA} \rangle_{1,DMD}$=54.8 Å, and $\langle R_{DA} \rangle_{2,DMD}$=38.8 Å (Fig. 31D). Although, it is clear that the amplitudes do not agree. The experimental population fractions were x_1 = 0.55 and x_2 =0.45, respectively. For the $p(R_{DA})_{DMD}$ the fractional amplitudes are $x_{1,DMD}$ = 0.11 and $x_{2,DMD}$ =0.89. For D91C/M159C (Fig. 31E), DMD and eTCSPC obtained a similar distance for the major population with $\langle R_{DA} \rangle_1$=34.7 Å (x_1 = 0.90) and $\langle R_{DA} \rangle_{2,DMD}$=39.2 Å x_2 =0.10. In this case only one distance agreed very well with the $p(R_{DA})_{DMD}$ ($x_{2,DMD}$ =0.81). However, the minor population was at a much shorter distance than observed experimentally. For S142C/K202C, DMD reports shorter distances for both populations and a much more even population fraction than experimentally determined. Similar variations were observed for the other variants. In some cases, experimental results are within 5 Å of the DMD distance while the in worst case the difference approaches 20 Å. As in experimental results, the simulated interdye distance distributions showed a lack of agreement regarding the underlying state distribution with the estimates ranging from near even to close to 90% occupancy of a single state
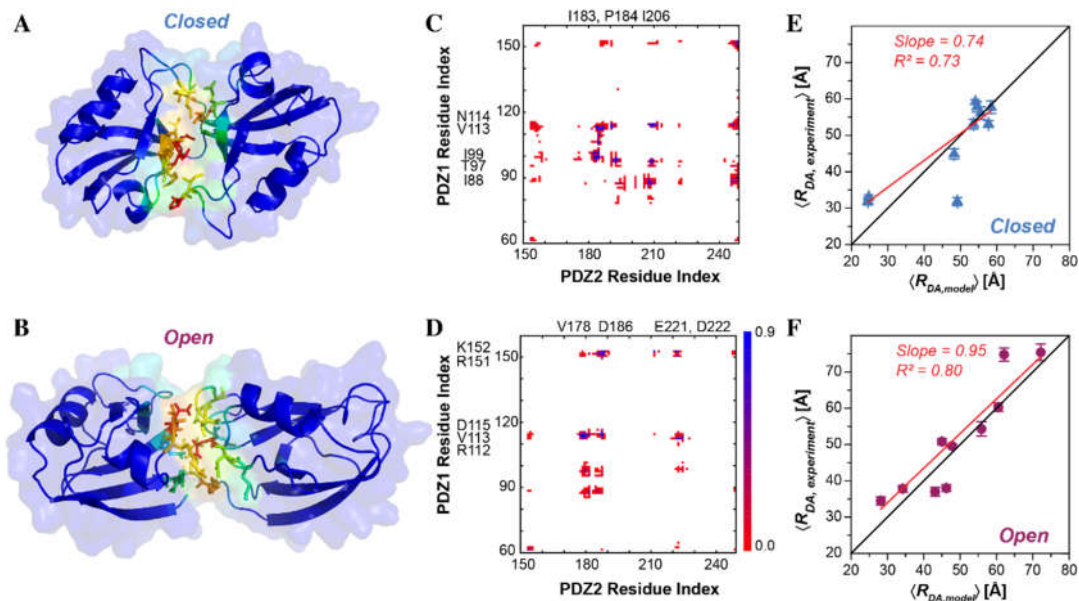
Figure 31. Structural states of PDZ1-2 tandem. The representative structures of two conformational states, (A) the closed state with a shorter inter-domain distance and larger contact interface and (B) the open state with a longer distance and smaller contact interface, were derived from DMD simulations. The two PDZ domains are shown in cartoon representation with transparent surface, and residues are colored according to their frequencies of making contact with the other PDZ domain in rainbow from blue (low frequency) to red (high frequency). Based on the conformational ensembles of both (C) closed and (D) open states, the residue-wise inter-domain contact frequency map are also computed. Residues with high inter-domain contact frequencies are highlighted along the axis for each PDZ domain. To compare simulation with experiment, we plotted the interdye distances from experiment (y-axis) against those taken from simulations (x-axis) for the **E)** Closed State and **F)** Open State.

5.3) Conclusion

According to DMD simulations [128] at least two basins are identified in the conformational landscape (PMF) of the PDZ1-2 tandem of PSD95. These basins correspond to possible supramoldule configurations, two of which show interdomain distances that predict contacts amongst them. Note that none of the previous X-ray crystallography structures revealed no interdomain contacts (Fig. 25), and the NMR

80

structure showed minimal interdomain contacts. Based on our eTCSPC analysis, we identify a major and a minor populations and the equilibrium agrees very well with the observation in MFD FRET [129]. MFD FRET is ideal to identify dynamic averaging at timescales faster than ms. Even in the cases where the distributions are heavily weighted towards the major population, and where the two identified configurations show similar interdye distances, the minor distance is capturing the asymmetry observed in the energy landscape. Such is the case of E135C/Y236C-DA, where the minor population is only 20%, and MFD captures the molecules that transition to this apparent high FRET state. However, if we focus on the minima of the basins of the energy landscape according to the DMD simulations, none of the identified configurations show a mean interdye distance that is that short. But there is a tail effect that is observed on the $p(R_{DA})_{DMD}$. This would represent molecules that go over that pathway in order to transition between states. Therefore, in order to differentiate between configuration states, the intedye distances have to be well separated and be almost equally distributed, such as the case of Q107C/H182C-DA. But even in cases, where this happens such as D91C/Y236C-DA, and where there is an agreement between the $p(R_{DA})_{PMF}$ and the interdye distance distribution derived by eTCSPC, it happens that DMD does not differentiate between the two configuration states. Therefore, in order to capture the two states, there are two factors that need to occur: *i*) the populations are equally distributed as observed experimentally; and *ii*) the expected interdye distance between two states have to be significantly different.

One observation of the $p(R_{DA})_{DMD}$ distributions, is that the modeled Gaussians have a much larger width than those we considered in the eTCSPC models. Our assumption was to represent the mobility of the dyes, where benchmark studies have suggested that a $R_{DA} = 6$ Å. However, that assumption considers that the protein or biomolecule is a rigid object. In this case, we can assign the additional broadening to the intrinsic dynamics of the molecule within each basin in the rugged energy landscape. Thus, if we consider the distanced determined by fitting the $p(R_{DA})_{DMD}$ distributions with two Gaussian distributed states, and we used the determined distances, it is possible to fit the eTCSCP by allowing the population contributions and their corresponding widths.

To compare against other structural studies of the PDZ1-2 tandem, we map in Fig. 31, the corresponding interdomain distances of the crystallographic (PDB ID 3GSL) and the NMR solved structures (Atomic coordinates were provided by Dr. Zhang since the structure was not deposited in the PDB). We add the corresponding expected interdye distances to have the coordinate system for each pair on the PMF. The expected interdye distance was calculated considering these states as rigid biomolecules and approximating the fluorescence dyes by the accessible volume. For the variant Q107C/H182C-DA, we find that the interdomain distances for the two-unit cells of the crystallographic structure (PDB ID 3GSLa/b) are 33.7 Å and 34.5 Å, respectively, while their expected mean interdye distances are 47.8 Å and 49.7 Å. The two structures cannot be differentiated based on FRET alone, and their interdomain distances are larger than those corresponding to the two free energy minima in DMD simulations. Similar behavior is

82

observed by comparing other variants. Therefore, it is possible that the crystal contacts add to the stabilization of this open state.

When comparing the NMR structure, the interdomain distance falls in the middle of the two minima, closer to the one basin with longer interdomain distance. The expected interdye distances either agree to one or the other state, or to what it seems to be an average configuration. For example, in Fig. 26A, for the variant Q107C/H182C-DA the expected measured intedyes distance for both states are 54.7 Å and 36.1 Å for the major and minor states, respectively. Considering the average of the two states, the fluorescence dyes is ~46.3 Å. On the other hand, the calculated interdye distance using the AV model on the NMR structure, we obtained a similar value of ~47.7 Å. Because dipolar couplings provide an orientation effect, and not distance restrains; it is possible that the overall interdomain orientation is skewed by the dynamic averaging identified in our FRET experiments, thus complicating the derivation of a possible structural model of the PDZ1-2 tandem. *Summary*

We combined FRET experiments with different temporal resolution and DMD simulations to study the possible supramodule configurations of the PDZ1-2 tandem. A 10 inter-domain FRET distance network was engineered to probe the structure and dynamics of the tandem. With the wide temporal resolution of our FRET approaches from subnanoseconds to tenths of seconds, we showed that the PDZ1-2 tandem is highly dynamic exchanging between at least two supramodule configurations and displays a heterogeneous free energy landscape. Two minima were observed in the energy landscape derived from DMD simulations. The upper limit of the transition between

83

different states is set by the dynamic averaging observed in MFD experiments, with an average diffusion time of < 1 ms. Structural analysis of different conformational states indicated that closer interdomain interactions are driven by hydrophobic interactions, while longer range interactions are determined by salt bridges. An experimentally observed fast inter conversion (< 1ms) was consistent with a low free energy barrier separating the two states in simulations, suggesting weak interdomain interactions. Hence, our results explain the high degeneracy observed and why previous structural studies of the same tandem protein had eluted characterizations because such states are only transiently populated.

APPENDICES

Appendix A

Materials and instruments

Name of Material/ Equipment	Company	Catalog Number
charcoal	Merck KGaA	K42964486 320
syringe filter	Fisherbrand	09-719C
chambered coverglass	LAB-TEK	LOT 071614-8-0
glass slide	Fisher Scientific	LOT# 063014-9
sterilized water	Fisher Scientific	148859
tween-20	Thermo Scientific	28320
DNA strand with acceptor (High FRET)	Integrated DNA Technologies	178124895
DNA strand with acceptor (Low FRET)	Integrated DNA Technologies	177956424
DNA strand with donor	Integrated DNA Technologies	177951437
thermol cycler	Eppendorf	E6331000025
485nm laser	PicoQuant	LDH-D-C-485
640nm laser	PicoQuant	LDH-D-C-640
microscope cover glass	Fisher Scientific	063014-9
Alexa 647	Life Technologies	A20347
Rhodamine 110	Sigma-Aldrich	83695-250MG
Rhodamine 101	Sigma-Aldrich	83694-500MG
photon detector	PicoQuant	PMA hybrid 40
computer	Dell	optiplex 7010
Detection software	PicoQuant GmbH	Heinrich Heine Unviersity
IPTG	Fisher Scientific	BP1755
HiTrap Chelating HP	GE Life Sciences	17-0409-01
Imidazole	Sigma-Aldrich	56749
Ni-NTA	Agarose Qiagen	30210
PD-10 Column	GE Life Sciences	17-0851-01
AktaPurifier	GE Life Sciences	28406264
Dialysis tubing	Spectrum labs	132562
Dichroics	Semrock	FF500/646
50/50 Beam splitter polarizer	Qioptiq Linos	G33 5743 000
Green pass filer	Chroma	ET525/50m
Red pass filter	Chroma	ET720/150m

Power Meter	ThorLabd	PM200
UV-Vis	Varian	Cary300Bio
Fluorolog 3 fluorometer	Horiba	FL3-22-R3
TCSPC controller	Horiba	Fluorohub-B
NanoLed 485L	Horiba	485L
NanoLed 635L	Horiba	635L
Olympus IX73	Olympus	IX73P2F
PMA 40 Hybrid Detector	PicoQuant GmbH	932200, PMA 40
PMA 50 Hybrid Detector	PicoQuant GmbH	932201, PMA 50
485 nm laser	PicoQuant GmbH	LDH-D-C-485
640 nm laser	PicoQuant GmbH	LDH-D-C-640

Appendix B

Methods

Multiparameter Fluorescence Detection (MFD) for smFRET experiments

MFD for confocal single-molecule Förster Resonance Energy Transfer (smFRET) experiments of single-molecules was performed using Pulsed Interleaved Excitation (PIE) with two diode lasers (Model LDH-D-C- 485 at 485 nm and laser LDH-D-C-640 at 640 nm; PicoQuant, Germany) operating at 40 MHz with 25 ns interleaved time. The power at objective was set for 60 µW for the 485 nm laser line and 23 µW for the 640 nm excitation. Labeled riboswitch molecules are excited when they pass though the focal point of a 60X Olympus objective in freely diffused condition. The emitted photo signal was also collected from the same objective and then filtered by a 70 µm pinhole. The signal was separated into parallel and perpendicular parts with two different spectral windows which are commonly defined as "green" and "red". Moreover, four photon-detectors are used: two for green (PMA Hybrid model 40 PicoQuant, Germany) and two for red channels (PMA Hybrid model 50, PicoQuant, Germany). A time-correlated single photon counting (TCSPC) module (HydraHarp 400, PicoQuant, Germany) with Time-Tagged Time-Resolved (TTTR) mode and 4 synchronized input channels were used for data registration.

For smMFD measurements, samples were diluted to pM in four different buffers (Apo buffer: 20mM Tris 200mM NaCl pH 7.0; Mg buffer: 20mM Tris 200mM NaCl 1M $MgCl_2$ pH 7.0; TPP buffer: 20mM Tris 200mM NaCl 2mg/mL TPP pH 7.0; Mg &TPP buffer: 20mM Tris 200mM NaCl 1M $MgCl_2$ 2mg/mL TPP pH 7.0), which had

been charcoal filtered to remove residual impurities. At pM concentration, we assure

that we observe ~2 molecules per second. To prevent adsorption artifacts, NUNC

chambers (Lab-Tek, Thermo Scientific, Germany) were pre-treated with a solution of

0.1% Tween 20 (Thermo Scientific) in water for 30 min and then rinsed with ddH2O.

Collection time varied from several minutes up to 10 hours. In long time

measurements, an oil immersion liquid was used to prevent drying out of the

immersion water. Standard controls consisted of measuring distilled water to identify

the instrument response function (IRF), buffers for background subtraction and the

nM concentration of green and red standard dyes (Rhodamine 110, Rhodamine 101,

and Alexa 647) in water solutions for calibration of green and red channels.

Discrete Molecular Dynamic (DMD) simulations

DMD is a special type of molecular dynamics (MD) algorithm where continuous

potentials are approximated by step-wise potential functions. A detailed description of

the DMD algorithm can be found elsewhere. Briefly, all molecules were represented

by the united-atom model will all heavy atoms and polar hydrogen atoms explicitly

modelled, and an implicit solvent model was used. We adopt an implicit solvent

approach, appropriate for studying long timescale dynamics of biomolecules due to

the separation from the fast water dynamics. The step-wise potential functions were

obtained by mimicking the continuous inter-atomic interaction potentials in the

molecular mechanics-based force field, the Medusa force field. Following the same

physical laws, the dynamics in DMD is equivalent to continuous potential MD at

timescales longer than *picoseconds* with differences mainly at short timescales within

the sub-*picosecond range* (i.e., the average time step between two consecutive inter-atomic collisions where a potential energy step is encountered). The bonded interactions include covalent bonds, bond angles, and dihedrals. The interatomic interactions include van der Waals, solvation, hydrogen bond, and electrostatics. The Medusa force field has been shown to be effective in accurate prediction of protein stability changes upon mutations for a large set of experimental data. The solvation energy is model by the Lazaridis-Karplus implicit solvent model. Screened electrostatic interactions are computed by the Debye-Hückel approximation. A Debye length of 1 nm is used by assuming a water dielectric constant of 80 and a monovalent electrolyte concentration of 0.1 M. The Anderson's thermostat is used to maintain constant temperature and a periodic boundary condition is applied.

The model dyes are composed of a linear tail and the aromatic head. In our simulations, the atoms in the tails are effectively non-interacting by setting a very small hardcore distance with all other atoms (0.001 Å). The atoms in the head have a hard sphere interaction with other atoms with the hardcore distance of 3 Å, while themselves are effectively non-interactive with a small hardcore distance of 0.001 Å. Each dye is constrained to the corresponding residues by mimicking a disulfide bond as in experiments.

Protein Expression and Purification

The PDZ1-2 tandem from *Rattus norvegicus* PSD-95 (residues 61-249) was expressed in *E. coli* and purified by a combination of Ni-affinity, ion exchange and size exclusion chromatography as previously described. For smTIRF, proteins were

90

labeled with 5:1 molar ratio (dye:protein) of Alexa Fluor 555 C_2 maleimide and Alexa Fluor 647 C_2 maleimide at an equimolar ratio. For MFD, proteins were first labeled with a 1:2 ratio of Alexa 488 C_5 maleimide for 1 hour at 4°C followed by addition of a 5:1 molar ratio of Alexa Fluor 647 C_2 maleimide, which was reacted overnight at 4°C. Unconjugated dye was removed by desalting with Sephadex G50 (GE Healthcare) followed by dialysis.

Ensemble Time correlated single photon counting (eTCSPC)

Ensemble time-correlated single-photon-counting (eTCSPC) measurements are performed using a Fluorolog3 spectrofluorometer in T shape with a PDX detector and Horiba Yvon photon (Horiba Yvon, Country) system. The excitation source was a NanoLED 485 nm diode laser (Model Number, Horiba, Country) for Rhodamine excitation. The corresponding monochromator slits were set to 2 nm (excitation path) and 16 nm (emission path).

References

1   Crick, F. Central dogma of molecular biology. *Nature.* **227** (5258), 561 (1970).

2   Poiesz, B. J. *et al.* Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proceedings of the National Academy of Sciences.* **77** (12), 7415-7419 (1980).

3   Denhardt, D. T. A membrane-filter technique for the detection of complementary DNA. *Biochemical and biophysical research communications.* **23** (5), 641-646 (1966).

4   Bussard, A. E. A scientific revolution?: The prion anomaly may challenge the central dogma of molecular biology. *EMBO reports.* **6** (8), 691-694 (2005).

5   Hurlbert, R. B., Schmitz, H., Brumm, A. F. & Potter, V. R. Nucleotide metabolism II. Chromatographic separation of acid-soluble nucleotides. *Journal of Biological Chemistry.* **209** (1), 23-39 (1954).

6   Wróblewski, F. & Ladue, J. S. Lactic dehydrogenase activity in blood. *Proceedings of the Society for Experimental Biology and Medicine.* **90** (1), 210-213 (1955).

7   Downing, M., Adams, A. & Hellenga, L. Enzymatic conversion of ribonucleotides to deoxynucleotides. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis.* **108** (2), 233-242 (1965).

8   Wojtczak, L., Bogucka, K., Sarzala, M. & Zaluska, H. Effect of fatty acids on energy metabolism and the transport of adenine nucleotides in mitochondria and other cellular structures. *Mitochondria, Structure and Function (Ernster, L. and Drahota, Z., eds.).* 79-92 (1969).

9   Jones, M. E., Black, S., Flynn, R. M. & Lipmann, F. Acetyl coenzyme a synthesis through pyrophosphoryl split of adenosine triphosphate. *Biochimica et biophysica acta.* **12** (1-2), 141-149 (1953).

10   Watson, J. D. & Crick, F. H. A structure for deoxyribose nucleic acid. *Nature.* **171** (4356), 737-738 (1953).

11   Chargaff, E., Lipshitz, R., Green, C. & Hodes, M. The composition of the desoxyribonucleic acid of salmon sperm. *Journal of Biological Chemistry.* **192** (1), 223-230 (1951).

12   Holliday, R. in *Organization and Contror in Prokaryotic and Eukaryotic Cells, 20th Symp. Soc. gen. Microbiol.(HP Charles and BCJG Knight, eds.).* 359-380.

13   Johnson, L., Abelson, H., Green, H. & Penman, S. Changes in RNA in relation to growth of the fibroblast. I. Amounts of mRNA, rRNA, and tRNA in resting and growing cells. *Cell.* **1** (2), 95-100 (1974).

14   Vauquelin, L. N. Chapter I Discovery of the Amino Acids. *Outline of the amino acids and proteins.* 13 (1944).

15   Tower, D. B. in *The Neurochemistry of Nucleotides and amino acids*   173-204 (Wiley New York, 1960).

16 TAKAHASHI, H., TIBA, M., IINO, M. & TAKAYASU, T. THE EFFECT OF γ-AMINOBUIC ACID ON BLOOD PRESSURE. *The Japanese journal of physiology.* **5** 334-341 (1955).

17 Udenfriend, S. & Cooper, J. R. The chemical estimation of tyrosine and tyramine. *J. biol. Chem.* **196** (1), 227-233 (1952).

18 Halver, J. E., DeLong, D. C. & Mertz, E. T. Nutrition of salmonoid fishes: V. Classification of essential amino acids for chinook salmon. *The Journal of nutrition.* **63** (1), 95-105 (1957).

19 Kendrew, J. C. Architecture of a protein molecule. *Nature.* **182** (4638), 764-767 (1958).

20 Pauling, L., Corey, R. B. & Branson, H. R. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences.* **37** (4), 205-211 (1951).

21 Von Bertalanffy, L. The History and Status of General Sytems Theory.   (1950).

22 Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature.* **181** (4610), 662-666 (1958).

23 Sheldrick, B. Structure determination by X‐ray crystallography: By MFC Ladd and RA Palmer. pp 502. Plenum Press, New York.£ 39.50. ISBN 0‐30641878‐9. *Biochemistry and Molecular Biology Education.* **14** (2), 93-93 (1986).

24 Sandström, J. *Dynamic NMR spectroscopy.* (Academic Press, 1982).

25 Havel, T. F. & Wuthrich, K. An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution. *J Mol Biol.* **182** (2), 281-294 (1985).

26 Kline, A. D. & Wuthrich, K. Secondary structure of the alpha-amylase polypeptide inhibitor tendamistat from Streptomyces tendae determined in solution by 1H nuclear magnetic resonance. *J Mol Biol.* **183** (3), 503-507 (1985).

27 Williamson, M. P., Havel, T. F. & Wuthrich, K. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. *J Mol Biol.* **182** (2), 295-315 (1985).

28 Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nature protocols.* **2** (11), 2728-2733 (2007).

29 Henzler-Wildman, K. A. *et al.* A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature.* **450** (7171), 913-U927 (2007).

30 Henzler-Wildman, K. A. *et al.* Intrinsic motions along an enzymatic reaction trajectory. *Nature.* **450** (7171), 838-U813 (2007).

31 Dolino, D. M., Adariani, S. R., Shaikh, S. A., Jayaraman, V. & Sanabria, H. Conformational selection and submillisecond dynamics of the ligand-binding domain of the N-methyl-D-Aspartate receptor. *Journal of Biological Chemistry.* jbc. M116. 721274 (2016).

32 Boura, E. *et al.* Solution structure of the ESCRT-I complex by small-angle X-ray scattering, EPR, and FRET spectroscopy. *Proc Natl Acad Sci U S A.* **108** (23), 9437-9442, doi:10.1073/pnas.1101763108, (2011).

33   Zheng, W. & Tekpinar, M. Accurate flexible fitting of high-resolution protein structures to small-angle x-ray scattering data using a coarse-grained model with implicit hydration shell. *Biophysical journal.* **101** (12), 2981-2991 (2011).

34   McCann, J. J., Zheng, L., Chiantia, S. & Bowen, M. E. Domain orientation in the N-Terminal PDZ tandem from PSD-95 is maintained in the full-length protein. *Structure.* **19** (6), 810-820, doi:10.1016/j.str.2011.02.017, (2011).

35   Best, R. B. *et al.* Effect of flexibility and cis residues in single-molecule FRET studies of polyproline. *Proc Natl Acad Sci U S A.* **104** (48), 18964-18969, doi:10.1073/pnas.0709567104, (2007).

36   Choi, U. B. *et al.* Single-molecule FRET-derived model of the synaptotagmin 1-SNARE fusion complex. *Nat Struct Mol Biol.* **17** (3), 318-324, doi:10.1038/nsmb.1763, (2010).

37   Dolino, D. *et al.* Intra and Interdomain Motions of the NMDA Receptor using Single Molecule FRET. *Biophysical Journal.* **110** 288a (2016).

38   Brunger, A. T., Strop, P., Vrljic, M., Chu, S. & Weninger, K. R. Three-dimensional molecular modeling with single molecule FRET. *J Struct Biol.* **173** (3), 497-505, doi:10.1016/j.jsb.2010.09.004, (2011).

39   Steven, A. C. & Baumeister, W. The future is hybrid. *J Struct Biol.* **163** (3), 186-195, doi:10.1016/j.jsb.2008.06.002, (2008).

40   Cowieson, N. P., Kobe, B. & Martin, J. L. United we stand: combining structural methods. *Curr Opin Struct Biol.* **18** (5), 617-622, doi:10.1016/j.sbi.2008.07.004, (2008).

41   Orrit, M. & Bernard, J. Single Pentacene Molecules Detected by Fluorescence Excitation in a p-Terphenyl Crystal. *Physical Review Letters.* **65** (21), 2716-2719 (1990).

42   Hoefling, M. *et al.* Structural heterogeneity and quantitative FRET efficiency distributions of polyprolines through a hybrid atomistic simulation and Monte Carlo approach. *PLoS One.* **6** (5), e19791, doi:10.1371/journal.pone.0019791, (2011).

43   Stryer, L. Fluorescence energy transfer as a spectroscopic ruler. *Annual Review of Biochemistry.* **47** 819-846 (1978).

44   Ma, J. *et al.* High Precision FRET at Single-molecule Level for Biomolecule Structure Determination. *J Vis Exp.* (123), e55623, doi:10.3791/55623, (2017).

45   Sisamakis, E., Valeri, A., Kalinin, S., Rothwell, P. J. & Seidel, C. A. Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods in enzymology.* **475** 455-514 (2010).

46   Kalinin, S., Valeri, A., Antonik, M., Felekyan, S. & Seidel, C. A. M. Detection of structural dynamics by FRET: a photon distribution and fluorescence lifetime analysis of systems with multiple states. *J Phys Chem B.* **114** (23), 7983-7995, doi:10.1021/jp102156t, (2010).

47   Schuler, B., Lipman, E. A., Steinbach, P. J., Kumke, M. & Eaton, W. A. Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence *Proceedings of the National Academy of Sciences of the United States of America.* **102** (8), 2754-2759 (2005).

48   Kalinin, S., Sisamakis, E., Magennis, S. W., Felekyan, S. & Seidel, C. A. On the origin of broadening of single-molecule FRET efficiency distributions beyond shot noise limits. *The journal of physical chemistry B.* **114** (18), 6197-6206 (2010).

49   Kalinin, S. P., T.; Sindbert, S.; Rothwell, P. J.; Berger, S.; Restle, T.; Goody, R. S.; Gohlke, H.; Seidel, C. A. M. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nature Methods.* (2012).

50   Jares-Erijman, E. A. & Jovin, T. M. FRET imaging. *Nature Biotechnology.* **21** (11), 1387-1395 (2003).

51   Margittai, M. *et al.* Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1. *Proceedings of the National Academy of Sciences.* **100** (26), 15516-15521 (2003).

52   Renner, A. *et al.* High Precision FRET Reveals Dynamic Structures in the Drosophila Scaffold Protein Complex Stardust-DPATJ-DLin-7 Mediated by L27 Domains. *Biophysical Journal.* **106** (2), 256a (2014).

53   Kudryavtsev, V. *et al.* Combining MFD and PIE for Accurate Single‐Pair Förster Resonance Energy Transfer Measurements. *ChemPhysChem.* **13** (4), 1060-1078 (2012).

54   Muschielok, A. & Michaelis, J. Application of the Nano-Positioning System to the Analysis of Fluorescence Resonance Energy Transfer Networks. *Journal of Physical Chemistry B.* **115** (41), 11927-11937, doi:Doi 10.1021/Jp2060377, (2011).

55   Iyer, V., Rossow, M. J. & Waxham, M. N. Peak two-photon molecular brightness of fluorophores is a robust measure of quantum efficiency and photostability. *JOSA B.* **23** (7), 1420-1433 (2006).

56   Förster, T. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen der Physik.* **2** 55-75 (1948).

57   DeRocco, V., Anderson, T., Piehler, J., Erie, D. A. & Weninger, K. Four-color single-molecule fluorescence with noncovalent dye labeling to monitor dynamic multimolecular complexes. *Biotechniques.* **49** (5), 807-816, doi:10.2144/000113551, (2010).

58   Widengren, J. *et al.* Single-molecule detection and identification of multiple species by multiparameter fluorescence detection. *Analytical chemistry.* **78** (6), 2039-2050 (2006).

59   Kühnemuth, R. & Seidel, C. A. Principles of single molecule multiparameter fluorescence spectroscopy. *Single Molecules.* **2** (4), 251-254 (2001).

60   Dale, R. E., Eisinger, J. & Blumberg, W. E. The orientational freedom of molecular probes. The orientation factor in intramolecular energy transfer. *Biophys J.* **26** (2), 161-193, doi:10.1016/S0006-3495(79)85243-1, (1979).

61   Elson, E. L. & Magde, D. Fluorescence correlation spectroscopy. I. Conceptual basis and theory. *Biopolymers.* **13** (1), 1-27 (1974).

62 Felekyan, S. *et al.* Full correlation from picoseconds to seconds by time-resolved and time-correlated single photon detection. *Review of scientific instruments.* **76** (8), 083104 (2005).

63 Magde, D., Elson, E. L. & Webb, W. W. Fluorescence correlation spectroscopy. II. An experimental realization. *Biopolymers.* **13** (1), 29-61 (1974).

64 Antonik, M., Felekyan, S., Gaiduk, A. & Seidel, C. A. Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *The Journal of Physical Chemistry B.* **110** (13), 6970-6978 (2006).

65 Gaiduk, A., Kühnemuth, R., Antonik, M. & Seidel, C. A. Optical Characteristics of Atomic Force Microscopy Tips for Single-Molecule Fluorescence Applications. *ChemPhysChem.* **6** (5), 976-983 (2005).

66 Felekyan, S., Kalinin, S., Sanabria, H., Valeri, A. & Seidel, C. A. Filtered FCS: Species Auto-and Cross-Correlation Functions Highlight Binding and Dynamics in Biomolecules. *ChemPhysChem.* **13** (4), 1036-1053 (2012).

67 Böhmer, M., Wahl, M., Rahn, H.-J., Erdmann, R. & Enderlein, J. Time-resolved fluorescence correlation spectroscopy. *Chemical Physics Letters.* **353** (5), 439-445 (2002).

68 Sindbert, S. *et al.* Accurate Distance Determination of Nucleic Acids via Forster Resonance Energy Transfer: Implications of Dye Linker Length and Rigidity. *Journal of the American Chemical Society.* **133** (8), 2463-2480, doi:10.1021/ja105725e, (2011).

69 Kalinin, S., Felekyan, S., Antonik, M. & Seidel, C. A. Probability distribution analysis of single-molecule fluorescence anisotropy and resonance energy transfer. *The Journal of Physical Chemistry B.* **111** (34), 10253-10262 (2007).

70 Kask, P. *et al.* Two-dimensional fluorescence intensity distribution analysis: theory and applications. *Biophysical Journal.* **78** (4), 1703-1713 (2000).

71 Kalinin, S. *et al.* A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nature methods.* **9** (12), 1218-1225 (2012).

72 Fessl, T. *et al.* Towards characterization of DNA structure under physiological conditions in vivo at the single-molecule level using single-pair FRET. *Nucleic Acids Res.* **40** (16), e121, doi:10.1093/nar/gks333, (2012).

73 Michalet, X., Weiss, S. & Jager, M. Single-molecule fluorescence studies of protein folding and conformational dynamics. *Chem Rev.* **106** (5), 1785-1813, doi:10.1021/cr0404343, (2006).

74 Furukawa, H. & Gouaux, E. Mechanisms of activation, inhibition and specificity: crystal structures of the NMDA receptor NR1 ligand-binding core. *EMBO J.* **22** (12), 2873-2885, doi:10.1093/emboj/cdg303, (2003).

75 Furukawa, H., Singh, S. K., Mancusso, R. & Gouaux, E. Subunit arrangement and function in NMDA receptors. *Nature.* **438** (7065), 185-192, doi:10.1038/nature04089, (2005).

76 Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature.* **450** (7172), 964-972 (2007).

77   Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nature Reviews Molecular Cell Biology.* **5** (6), 451-463 (2004).

78   Winkler, W. C. & Breaker, R. R. Genetic control by metabolite-binding riboswitches. *Chembiochem.* **4** (10), 1024-1032 (2003).

79   Nudler, E. & Mironov, A. S. The riboswitch control of bacterial metabolism. *Trends in biochemical sciences.* **29** (1), 11-17 (2004).

80   Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature.* **419** (6910), 952-956 (2002).

81   Sudarsan, N., Barrick, J. E. & BREAKER, R. R. Metabolite-binding RNA domains are present in the genes of eukaryotes. *Rna.* **9** (6), 644-647 (2003).

82   Sudarsan, N. *et al.* Tandem riboswitch architectures exhibit complex gene control functions. *Science.* **314** (5797), 300-304 (2006).

83   Yamauchi, T. *et al.* Roles of Mg2+ in TPP-dependent riboswitch. *FEBS letters.* **579** (12), 2583-2588 (2005).

84   Kubodera, T. *et al.* Thiamine-regulated gene expression of Aspergillus oryzae thiA requires splicing of the intron containing a riboswitch-like domain in the 5′-UTR. *FEBS letters.* **555** (3), 516-520 (2003).

85   Serganov, A., Polonskaia, A., Phan, A. T., Breaker, R. R. & Patel, D. J. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature.* **441** (7097), 1167-1171 (2006).

86   Baird, N. J., Kulshina, N. & Ferré D'Amaré, A. R. Riboswitch function: flipping the switch or tuning the dimmer? *RNA biology.* **7** (3), 328-332 (2010).

87   Kulshina, N., Edwards, T. E. & Ferré-D'Amaré, A. R. Thermodynamic analysis of ligand binding and ligand binding-induced tertiary structure formation by the thiamine pyrophosphate riboswitch. *RNA.* **16** (1), 186-196 (2010).

88   Burnouf, D. *et al.* kinITC: a new method for obtaining joint thermodynamic and kinetic data by isothermal titration calorimetry. *Journal of the American Chemical Society.* **134** (1), 559-565 (2011).

89   Ritchie, D. B. & Woodside, M. T. Probing the structural dynamics of proteins and nucleic acids with optical tweezers. *Current opinion in structural biology.* **34** 43-51 (2015).

90   Anthony, P. C., Perez, C. F., García-García, C. & Block, S. M. Folding energy landscape of the thiamine pyrophosphate riboswitch aptamer. *Proceedings of the National Academy of Sciences.* **109** (5), 1485-1489 (2012).

91   Baird, N. J. & Ferré-D'Amaré, A. R. Idiosyncratically tuned switching behavior of riboswitch aptamer domains revealed by comparative small-angle X-ray scattering analysis. *Rna.* **16** (3), 598-609 (2010).

92   Ali, M., Lipfert, J., Seifert, S., Herschlag, D. & Doniach, S. The ligand-free state of the TPP riboswitch: a partially folded RNA structure. *Journal of molecular biology.* **396** (1), 153-165 (2010).

93   Blount, K. F. & Breaker, R. R. Riboswitches as antibacterial drug targets. *Nature biotechnology.* **24** (12), 1558-1564 (2006).

94   Hamad, B. The antibiotics market. *Nature Reviews Drug Discovery.* **9** (9), 675-676 (2010).

95   Lewis, K. Platforms for antibiotic discovery. *Nature reviews Drug discovery.* **12** (5), 371-387 (2013).

96   Ding, F. *et al.* Ab initio RNA folding by discrete molecular dynamics: From structure prediction to folding mechanisms. *RNA.* **14** (6), 1164-1173 (2008).

97   Muschielok, A. *et al.* A nano-positioning system for macromolecular structural analysis. *Nat Methods.* **5** (11), 965-971, doi:10.1038/nmeth.1259, (2008).

98   Yin, S., Ding, F. & Dokholyan, N. V. Modeling backbone flexibility improves protein stability estimation. *Structure.* **15** (12), 1567-1576 (2007).

99   Hertig, S., Goddard T , D., Johnson G , T. & Ferrin T , E. Multidomain Assembler (MDA) Generates Models of Large Multidomain Proteins. *Biophys J.* **108** (9), 2097-2102, doi:10.1016/j.bpj.2015.03.051, (2015).

100   Xu, D., Jaroszewski, L., Li, Z. & Godzik, A. AIDA: ab initio domain assembly for automated multi-domain protein structure prediction and domain-domain interaction prediction. *Bioinformatics.* **31** (13), 2098-2105, doi:10.1093/bioinformatics/btv092, (2015).

101   Tompa, P. On the supertertiary structure of proteins. *Nat Chem Biol.* **8** (7), 597-600 (2012).

102   Sakon, J. J. & Weninger, K. R. Detecting the conformation of individual proteins in live cells. *Nat Methods.* **7** (3), 203-205, doi:10.1038/nmeth.1421, (2010).

103   Ye, F. & Zhang, M. Structures and target recognition modes of PDZ domains: recurring themes and emerging pictures. *Biochem J.* **455** (1), 1-14, doi:10.1042/bj20130783, (2013).

104   Wang, W., Weng, J., Zhang, X., Liu, M. & Zhang, M. Creating conformational entropy by increasing interdomain mobility in ligand binding regulation: a revisit to N-terminal tandem PDZ domains of PSD-95. *J Am Chem Soc.* **131** (2), 787-796, doi:10.1021/ja807602210.1021/ja8076022 [pii], (2009).

105   Andrecka, J. *et al.* Nano positioning system reveals the course of upstream and nontemplate DNA within the RNA polymerase II elongation complex. *Nucleic Acids Research.* **37** (17), 5803-5809, doi:Doi 10.1093/Nar/Gkp601, (2009).

106   Karginov, A. V., Ding, F., Kota, P., Dokholyan, N. V. & Hahn, K. M. Engineered allosteric activation of kinases in living cells. *Nature biotechnology.* **28** (7), 743-747 (2010).

107   Gianni, S. *et al.* Sequence-specific long range networks in PSD-95/discs large/ZO-1 (PDZ) domains tune their binding selectivity. *J Biol Chem.* **286** (31), 27167-27175, doi:10.1074/jbc.M111.239541, (2011).

108   Bach, A. *et al.* A high-affinity, dimeric inhibitor of PSD-95 bivalently interacts with PDZ1-2 and protects against ischemic brain damage. *Proc Natl Acad Sci U S A.* **109** (9), 3317-3322, doi:10.1073/pnas.1113761109, (2012).

109   Eildal, J. N. N. *et al.* Rigidified Clicked Dimeric Ligands for Studying the Dynamics of the PDZ1‑2 Supramodule of PSD‑95. *Chembiochem.* **16** (1), 64-69 (2015).

110 Hung, A. Y. & Sheng, M. PDZ domains: structural modules for protein complex assembly. *J Biol Chem.* **277** (8), 5699-5702, doi:10.1074/jbc.R100065200, (2002).

111 Xie, X. *et al.* Open–closed motion of Mint2 regulates APP metabolism. *Journal of Molecular Cell Biology.* **5** (1), 48-56, doi:10.1093/jmcb/mjs033, (2013).

112 Petit, C. M., Zhang, J., Sapienza, P. J., Fuentes, E. J. & Lee, A. L. Hidden dynamic allostery in a PDZ domain. *Proc Natl Acad Sci U S A.* **106** (43), 18249-18254, doi:10.1073/pnas.0904492106, (2009).

113 Dominguez, C., Boelens, R. & Bonvin, A. M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc.* **125** (7), 1731-1737, doi:10.1021/ja026939x, (2003).

114 Best, R. B., Li, B., Steward, A. & Clarke, J. Can Non-Mechanical Proteins Withstand Force? Stretching Barnase by Atomic Force Microscopy and Molecular Dynamics Simulation. *Biophysical Journal.* **81** (4), 2344-2356 (2001).

115 Long, J.-F. *et al.* Supramodular structure and synergistic target binding of the N-terminal tandem PDZ domains of PSD-95. *Journal of Molecular Biology.* **327** (1), 203-214 (2003).

116 Knight, J. L., Mekler, V., Mukhopadhyay, J., Ebright, R. H. & Levy, R. M. Distance-Restrained Docking of Rifampicin and Rifamycin SV to RNA Polymerase Using Systematic FRET Measurements: Developing Benchmarks of Model Quality and Reliability. *Biophysical Journal.* **88** (2), 925-938 (2005).

117 Long, J.-F. *et al.* Autoinhibition of X11/Mint scaffold proteins revealed by the closed conformation of the PDZ tandem. *Nat Struct Mol Biol.* **12** (8), 722-728 (2005).

118 Hickerson, R., Majumdar, Z. K., Baucom, A., Clegg, R. M. & Noller, H. F. Measurement of internal movements within the 30 S ribosomal subunit using Forster resonance energy transfer. *J Mol Biol.* **354** (2), 459-472, doi:10.1016/j.jmb.2005.09.010, (2005).

119 Weninger, K., Bowen, M. E., Chu, S. & Brunger, A. T. Single-molecule studies of SNARE complex assembly reveal parallel and antiparallel configurations. *Proceedings of the National Academy of Sciences of the United States of America.* **100** (25), 14800-14805 (2003).

120 Fanning, A. S. & Anderson, J. M. Protein-protein interactions: PDZ domain networks. *Curr Biol.* **6** (11), 1385-1388 (1996).

121 Ha, T. *et al.* Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc.Natl.Acad.Sci.USA.* **93** 6264-6268 (1996).

122 McCann, J. J. *et al.* Supertertiary structure of the synaptic MAGuK scaffold proteins is conserved. *Proc Natl Acad Sci U S A.* **109** (39), 15775-15780, doi:120025410910.1073/pnas.1200254109, (2012).

123 Weiss, S. Fluorescence spectroscopy of single biomolecules. *Science.* **283** (5408), 1676-1683 (1999).

124 Eggeling, C., Fries, J., Brand, L., Günther, R. & Seidel, C. Monitoring conformational dynamics of a single molecule by selective fluorescence

spectroscopy. *Proceedings of the National Academy of Sciences.* **95** (4), 1556-1561 (1998).

125 Kumawat, A. & Chakrabarty, S. Hidden electrostatic basis of dynamic allostery in a PDZ domain. *Proc Natl Acad Sci U S A.* **114** (29), E5825-E5834, doi:10.1073/pnas.1705311114, (2017).

126 Ha, T. & Tinnefeld, P. Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging. *Annu Rev Phys Chem.* **63** 595-617, doi:10.1146/annurev-physchem-032210-103340, (2012).

127 Ding, F. & Dokholyan, N. V. Emergence of Protein Fold Families through Rational Design. *PLoS Comput Biol.* **2** (7), e85, doi:10.1371/journal.pcbi.0020085, (2006).

128 Shirvanyants, D., Ding, F., Tsao, D., Ramachandran, S. & Dokholyan, N. V. Discrete Molecular Dynamics: An Efficient And Versatile Simulation Method For Fine Protein Characterization. *The Journal of Physical Chemistry B.* **116** (29), 8375-8382, doi:10.1021/jp2114576, (2012).

129 Kudryavtsev, V. *et al.* Combining MFD and PIE for Accurate Single-Pair Förster Resonance Energy Transfer Measurements. *ChemPhysChem.* **13** (4), 1060-1078, doi:10.1002/cphc.201100822, (2012).