ABSTRACT

Title of Document: **SINGLE MOLECULE FLUORESCENCE INVESTIGATION OF PROTEIN FOLDING**

Jianwei Liu, Ph.D., 2008

Directed By: Associate Professor, Victor Muñoz, Department of Chemistry and Biochemistry

In addition to the well known two-state folding scenario, the energy landscape theory of protein folding predicts the possibility of downhill folding under native conditions. This intriguing prediction was extended by Victor Muñoz and coworkers to include global downhill folding. i.e. a barrierless free energy surface and unimodal conformational distributions at all degrees of unfolding stress. A small protein, BBL, has been shown to follow this behavior as evidenced from experiments and simulations. However, the identification of BBL as a global downhill folder has raised a significant amount of controversy with some groups claiming that it still folds in a two-state fashion.

The objective of this thesis is to characterize the conformational distribution of BBL using single molecule Förster resonance energy transfer (SM-FRET) to obtain direct evidence for the downhill folding in BBL. We carried out SM-FRET measurements at 279 K to slow down the protein dynamics to 150 μs thus enabling the use of a 50 μs binning time (the short binning time being a first in SM

measurements). By optimizing the microscope system setup and employing a novel Trolox-cysteamine fluorophore protection system, we obtained sufficient signal to construct reliable 50 μs SM-FRET histograms. The data show clear unimodal conformational distributions at varying denaturant concentrations thus demonstrating the downhill folding nature of BBL.

Further SM-FRET measurements on a two-state folder, α-spectrin SH3 produced bimodal histograms indicating that our experimental setup works well and that the unimodal distributions of BBL are not due to instrumental errors. The comparison of ensemble FRET measurements on labeled proteins (both BBL and α-spectrin SH3) with CD measurements on the corresponding unlabeled proteins shows that the fluorophores do not affect the protein stability. We also simulated the expected histograms if BBL were a two-state folder using Szabo's photon statistics theory of SM-FRET. The two-state simulation results are inconsistent with the experimental histograms even under very conservative assumptions about BBL's relaxation time. Therefore, all the control experiments and simulations exclude any possible artifacts, which shows our results are quite robust. Additionally, we estimated the relaxation time of BBL from the histogram width analysis to be consistent with independent kinetic measurements.

SINGLE MOLECULE FLUORESCENCE INVESTIGATION OF PROTEIN FOLDING

By

Jianwei Liu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:
Associate Professor Victor Muñoz, Chair
Professor Dorothy Beckett
Assistant Professor Arthur La Porta
Associate Professor Wolfgang Losert,
Professor Devarajan Thirumalai

Dedication

To my sister Jin-Mei

Acknowledgements

I thank my wife Yan for accompanying me and inspiring me to work hard, for her self-sacrificing devotion to our daughters Wen-Xi (Evelyn) and Ruo-Xi (Irene). I am greatly indebted to my parents, my sisters and my brother. They shaped me into the person I am now. I also thank my in-laws for their support and encouragement.

I would like to thank my supervisor Dr. Victor Muñoz for providing me with the opportunity to work on such an intriguing research project and his original, insightful guidance. I am greatly inspired by his enthusiasm for the pursuit of knowledge.

The Muñoz group has been a joyous environment. I have enjoyed working with the present and past members: Mourad, Urmi, Rani, Athi, Tanay, Christina, Peng and Raquel. Particularly, I learned a lot about molecular biology from the three year unforgettable collaboration with Luis. I am also deeply grateful to Athi for helping me with my thesis writing and the many stimulating discussions about protein folding.

Many thanks to our collaborators Prof. Douglas English and Dr. Xiang Wang for their great efforts in the single molecule experiments. Other members of English group: Sara, Cathy and Amy have also been very instrumental to my work. In addition, I appreciate the help from Daoning and Sarah.

At the end, I should acknowledge the Chemical Physics program for their considerations about their students. I am happy to have joined this program. I am also grateful to my committee members for their comments and suggestions.

Table of Contents

Chapter 1 Introduction and Research Plan

1.1 Energy Landscape Theory of Protein Folding

1.1.1 General Description of Protein Folding

Following the pathway DNA→ mRNA→ protein, the genetic information stored in DNA is used to encode the amino acid sequence of proteins that perform most of the functions of life. However, such 1-dimensional arrays of amino acids do not perform any function *per se* until they fold into a specific 3-dimensional structure, i.e. the native structure. Hence, structure conveys function and if a protein loses its native structure it no longer functions.

This in turn leads to the question: how does the sequence of amino acids determine the final 3-D structure of a protein? This is typically phrased as the 'protein folding problem'. Folding is controlled by the interactions between amino acids themselves and those between amino acids and the medium (solvent). Such interactions include hydrogen bonds, salt-bridges, van der Waals forces and hydrophobic effect. Among these interactions, the hydrophobic force plays a dominant role in protein folding. To date we still do not understand the mechanisms of protein folding to the extent that we can reliably predict the folding dynamics or the final structure of a protein based on these interactions alone. However, we have learned many important aspects of protein folding.

It is well known that a protein will denature with increasing temperature or denaturants such as urea and guanidinium hydrochloride (GuHCl). In 1973, Anfinsen

found that chemically denatured RNase is able to regain its native structure upon dilution of denaturants *in vitro* [1]. This proves that the process is reversible and that all the information required is encoded in the amino acid sequence of a protein. Later, it was recognized that the folding of many proteins can be described as a first-order phase transition [2, 3]. One criterion of first-order phase transition is that the slope of the transition as a function of any physical parameter must be proportional to the number of units (i.e. amino acids) [3]. Privalov and Ptitsyn tested and found evidence for this criterion in their calorimetric experiments of protein denaturation [2, 3]. First-order phase transitions are equivalent to an all-or-none process. It means all the units of the system engage simultaneously in the change in conformation upon unfolding. The more units the system has, the closer to an all or none process and therefore the higher its cooperativity (which refers to the property that all residues of a protein change simultaneously during the transition). Although a protein molecule may have hundreds of thousands of degrees of freedom that can produce astronomical numbers of partially folded structures, the probability of observing those intermediates is almost reduced to zero (especially for single domain proteins) as a result of an all-or-none folding transition.

Folding Rate and Levinthal Paradox. It is known that the average difference in conformational entropy per residue between unfolded and native states is: $\Delta S \sim 4.3 cal/mol \bullet K$ [4]. Based on the Boltzmann definition of entropy, we have $\Delta S = kN_A \ln(\gamma / \gamma_f)$, where N_A is Avogadro's number, k is Boltzmann's constant and γ_f is the average number of orientations of each monomer in native state. γ_f is usually taken as unity because the native state is seen as possessing a fixed structure.

From these numbers we can estimate the average number of conformations for each monomer in the unfolded state: $\gamma \sim 8.7$. The exact value of γ is not critical, but since it is larger than unity, a protein composed of N monomers will populate an enormous number of conformations: γ^N [5, 6]. If the protein randomly searches all of the γ^N configurations in order to reach the native configuration, the folding time will be much longer than the experimentally observed time of less than 1 minute. This apparent paradox was first noted by Levinthal who proposed that proteins fold following specific pathways [7]. This idea has therefore been widely used to rationalize protein folding.

1.1.2 Statistical Energy Landscape Theory of Protein Folding

In contrast to the above "chemical" view, the energy landscape approach [8, 9] proposes a statistical view to describe protein folding. Here, the energy of the protein is a function of protein configuration, resulting in an energy landscape. Such energy has three main contributions: the energy of residues themselves $\varepsilon_i(\alpha_i)$, interactions between neighboring residues $J_{i,i+1}(\alpha_i, \alpha_{i+1})$ and interactions between non-neighboring residues $K_{i,j}(\alpha_i, \alpha_j, r_i, r_j)$ [8]:

$$E = -\sum_i \varepsilon_i(\alpha_i) - \sum_i J_{i,i+1}(\alpha_i, \alpha_{i+1}) - \sum_{i,j} K_{i,j}(\alpha_i, \alpha_j, r_i, r_j) \tag{1.1}$$

where α_i and r_i refer to the state and position of the i-th amino acid residue. From the configuration density $N(E)$ we can get the entropy:

$$S(E) = k \ln(N(E)) \tag{1.2}$$

Each minimum of the free energy

$$F = E - T \cdot S(E) \tag{1.3}$$

would be a possible state for the protein.

The above considerations would lead to a hyper-dimensional surface with a certain degree of roughness (local minima separated by small barriers). Protein folding can be described as the movement among these many local minima of the surface in search for a global minimum. The motions on such a rough free energy surface will be driven by thermal fluctuations ($\sim k_B T$).

The free energy surface rises and falls with configuration coordinates thereby producing surfaces with various degrees of roughness. The roughness will affect the folding rate. Rough and smooth energy landscapes have distinct thermodynamic and kinetic behaviors [9]. In a rough landscape the competition between sets of interactions is often high, which is called frustration [10]. A system with frustration (i.e., spin glass) requires a very long time to reach its global minimum on the energy landscape. A random compact heteropolymer is expected to have a very rough energy landscape. Obviously, most proteins are unlikely to have such rough/frustrated energy landscapes. Otherwise folding into the native structure would require much greater time than 1 minute. On the other hand, the landscape of proteins cannot be completely smooth, although protein folding shows some characteristics of smooth energy landscapes such as cooperativity. Now it is believed that the energy landscapes of natural proteins have intermediate degrees of roughness [9].

1.1.3 Two-State and Downhill Folding Scenarios

The conformational space of a protein is multi-dimensional. Any calculation in such a high dimensional space is impractical. So it is useful to begin with simple

representations obtained by projecting the free energy surface onto a few or even one order parameter (or reaction coordinate for kinetics). Using the reaction coordinate approach, and once we have a surface of low dimensionality, we can employ well known reaction rate theories such as transition state theory [11] and Kramer's theory [12] to describe the protein folding process. These treatments facilitate comparison with experiments.

Depending of the shape of the projected free energy surface two extreme scenarios arise. In the first scenario the surface displays two major minima separated by a high barrier. Thus, each molecule of the ensemble must be in one of those two minima. Such a two-state scenario can be described with a simple chemical two-state model:

$$N \leftrightarrow U \tag{1.4}$$

where N and U refer to native and unfolded (i.e. denatured) states respectively. The equilibrium constant $K_{eq} = [U]/[N]$ is determined by the free energy difference between the two states ΔG: $K_{eq} = \exp\left(-\frac{\Delta G}{RT}\right)$.

Many proteins have been identified to fold in a two-state manner. This has led to the proposal that two-state folding is a biological requirement of natural single domain proteins [13]. Practically speaking, a two-state scenario does not mean that there are no partially folded structures along the pathway, but that the probability of such structures is very small so that they are hard to observe and, thus, negligible.

The other folding scenario arises when the projected surface presents a small (< *3 kT*) barrier or a purely downhill shape. Here the projected free energy surface has two shallow minima or even a single well, as shown in Figure 1.1A. This scenario

is called downhill folding. It was originally proposed by Wolynes and collaborators [9], and first identified experimentally by Muñoz and coworkers [14]. In the one-state downhill folding case the ensemble of protein molecules will shift from higher order parameter values (i.e. folded state) to lower order parameter values (i.e. unfolded state) gradually as the temperature or concentration of chaotropic agents such as urea and GuHCl increases (see Figure 1.1B). Since the free energy surface of a downhill folding protein has only one minimum at all conditions, all intermediate stages during folding could be potentially populated in equilibrium. Such information about partially folded structures cannot be obtained directly for a protein that follows a two-state transition [14].

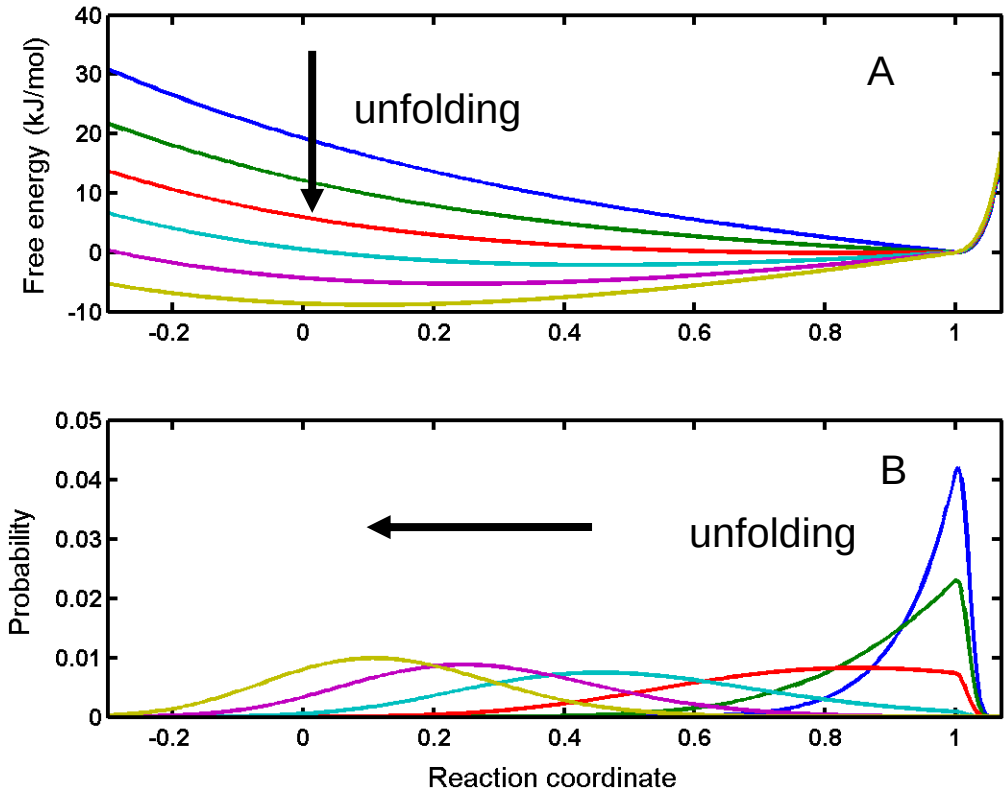

Figure 1.1 Illusion of the free energy surface and conformational distributions of downhill folding protein upon denaturation.

1.2 Brief Review of Downhill Folding

1.2.1 Identification of BBL as a Downhill Folder

Muñoz and co-workers used several techniques including differential scanning calorimetry (DSC), far UV circular dichroism spectroscopy (CD), FRET and Dansyl fluorescence to probe the thermal denaturation of protein BBL. The DSC profile was found to be very broad spanning 60 K and without a clear pre-transition baseline. The apparent transition temperatures (T_m) obtained from different techniques ranged from 295 to 335 K. Moreover, a wavelength-dependent unfolding transition was clearly evident from far-UV CD experiments. Also a basis-spectra analysis of far-UV CD spectra showed gradual unwinding of α-helices. Double perturbation experiments using temperature and urea also revealed steep pre-transition slopes and gradual unraveling of secondary structure. These findings do not agree with the fundamental expectations of a two-state folding protein [14].

A simple statistical-mechanical model (ME model) was used to globally fit all the thermal denaturation experiments with just six thermodynamic parameters. The resulting free-energy profiles were barrierless at all temperatures, indicative of global downhill folding behavior in BBL [14]. Later, they proposed the Variable-Barrier model that can estimate the free energy barrier height from the calorimetry data based on the Landau theory of phase transitions [15]. Using this model Muñoz and co-workers again obtained barrierless free energy surfaces. Thioredoxin, a two-state protein, gave a significant barrier height (~39.5 kJ mol^{-1}) thus validating the model. Therefore, all the experimental and theoretical work from the Muñoz group indicate the absence of barrier in BBL.

1.2.2 Controversy about Downhill Folding in BBL

However, the initial proposal of global downhill folding in BBL has raised a significant amount of controversy. Fersht and coworkers studied the unfolding of a slightly longer version of BBL (QNND-BBL) under higher ionic strength conditions [16]. They claimed that the T_m obtained from far-UV CD, DSC and NMR was similar and that the difference (which is more than 5 K) was due to experimental or fitting error. In effect, they claimed that BBL follows a barrier-limited transition. They suggested that extrinsic fluorophores present in Muñoz's version of BBL (Naf-BBL, with a Naphthyl-alanine at its N-terminus) [14] perturbed the equilibrium and caused protein aggregation. Possibly together with the shorter protein boundaries, the aggregation made the transition less cooperative.

Later, Naganathan et al. experimentally excluded the possibility of protein aggregation and perturbation caused by the fluorophore. They also showed that another BBL version with its ends protected (Ac-Naf-BBL-NH2) still folded in a downhill manner which meant that protein boundaries do not affect the folding behavior [17]. They explained the small differences in T_m obtained in Fersht's far-UV CD measurement to be a result of a wrong choice of the wavelengths. Their analysis also indicated that QNND-BBL folds in a downhill fashion.

NMR experiments on BBL. Both the Fersht and Muñoz groups employed NMR to characterize the unfolding of BBL, but their findings were still not consistent with each other. Fersht and co-workers followed the thermal unfolding of QNND-BBL by ^{13}C NMR under high ionic strength conditions and at pH 7.0 [18]. They obtained a T_m range of 320-329 K from individual two-state fitting of 15 probes (9

residues) and a global T_m of 324 K. The Muñoz's group, on the other hand, working on Naf-BBL at pH 5.3 (the pH at which the structure was originally reported) monitored the proton chemical shifts of 158 atoms [19]. Interestingly, and in contrast to the results of Fersht and colleagues, they obtained a T_m range spanning more than 60 K. Some probes even showed clear three-state transitions. They introduced a novel concept – the mean thermodynamic coupling index (MTCI) - to estimate the degree of similarity between the individual atomic unfolding curves and hence calculate the coupling between distant residues. The MTCI calculation thus provided a first ever quantitative estimate of the folding cooperativity in a protein.

1.2.3 Simulations of BBL Unfolding

In the same period, some theoretical and simulation work have also been carried out to test the experimental observations in BBL. Wang and coworkers employed an off-site Go-model to simulate the unfolding of BBL and other proteins. They found that BBL had the lowest cooperativity in their database with no obvious barrier at different temperatures, which is consistent with the experiment and modeling works of the Muñoz group [20]. By employing Cα native centric Go-model, Knott and Chan also obtained similar barrierless free energy surface and unimodal distribution of conformations for BBL [21].

Later, Wolynes and coworkers performed purely additive native structure-based simulations of QNND-BBL and Naf-BBL, obtaining a significant (~4 RT_B) and very low barrier (<0.5 RT_B), respectively. They attributed the discrepancy between those two constructs of BBL to modest tertiary structural differences. Some critical long-range contacts absent in Naf-BBL were claimed to result in the very low

estimated barrier. Their simulation also showed that ionic strength of the medium did not affect the free energy barrier [22].

As an exception, Chang and coworkers obtained a marginal barrier (~2 RT_m) for PDB entry 1BBL (37 residues) and a higher barrier for other BBL versions using the extended Eaton-Muñoz model with multiscale-heterogeneous pairwise interactions [23]. However, the authors failed to parameterize this purely phenomenological model with the experimental data thus raising questions as to the validity of the conclusions. Replica-exchange molecular dynamics simulations from two independent groups also reveal very marginal barriers for BBL under different reaction coordinate assumptions [24, 25].

1.2.4 Summary of the Controversy about Downhill Folding in BBL

In summary, it is now clear that the downhill folding in BBL is not an artifact caused by the extrinsic fluorophore [17] and ionic strength conditions have also been shown to have very modest effects on the barrier height [22]. Most theoretical simulations further support barrierless free energy surface in BBL under different degrees of denaturational stress.

One of the equilibrium signatures of downhill folding is the large spread in apparent T_m when different probes are employed. Here the precondition is to choose probes that monitor as many different structural features as possible, as done in the first experiments on BBL. If similar probes are chosen it is bound to give identical melting temperatures even if the protein folds downhill as the same unfolding process is monitored. This is a possible reason for the similarity in T_m observed by Fersht and co-workers in far-UV CD experiments on QNND-BBL[16]. In addition, the NMR

experiments on BBL by Munoz and co-workers reveal a normal distribution of melting temperatures centered around the global T_m. The corollary is that the smaller the number of probes the more probable that they cluster around the global T_m. So to give convincing evidence to one- or two-state transition, it is necessary to follow as many probes as possible, as done by the Munoz group. On the other hand, the Fersht group monitored just 15 probes from 9 residues in contrast to the 158 probes followed by Munoz and co-workers.

Moreover, the determination of T_m is based on a simple two-state fit that mostly involves 6 free parameters: 2 each for the folded and unfolded baselines, T_m and ΔH_m (the enthalpy of unfolding at the T_m). The Fersht group succeeded in globally fitting the chemical shifts of 15 probes whose individual two-state fittings produced a spread in T_m of ~ 9 K [18]. They have apparently been successful in globally fitting even the NMR unfolding curves of 158 protons with spread in individual T_m of ~ 60 K [26]. The main reason that unfolding curves with very different T_ms can be globally fitted to an average T_m is due to the free-floating nature of the baselines. The baselines are supposed to represent the signals of native and unfolded states as a function of the denaturational stress. When the native baseline crosses the unfolded baseline, the two-state fitting becomes meaningless even though it produces very small fitting error. Therefore extreme care should be taken in interpreting the results of global fits with respect to the physical meaning of resulting baselines which unfortunately was not the case with Fersht and coworkers.

1.3 Research Proposal: Single Molecule Förster Resonance Energy Transfer (SM-FRET) to Measure Protein Conformational Distributions

As discussed above, the proposal of global downhill folding in BBL has raised a significant amount of controversy with some groups arguing that BBL is a two-state folder. The principal difference between two-state folding and global downhill folding is the presence or absence of a free energy barrier at the apparent midpoint. Consequently a global downhill folding protein should have unimodal probability distribution at all conditions. In contrast, two subpopulations corresponding to the folded and unfolded states co-exist in the middle of transition for a two-state folder. However, conventional methods are not able to detect the probability distribution directly. Models are needed to analyze the experiments quantitatively, which is the root of different opinions about the folding mechanism of BBL.

In the last two decades, optical single molecule spectroscopy (SMS) has obtained great progress enabling the detection of single molecules possible. As one type of SMS techniques, SM-FRET measures the FRET efficiency between two fluorophores labeled mostly at the ends of a macromolecule e.g. protein. Because FRET efficiency depends on the inter-dye distance which in turn depends on the macromolecular conformation, SM-FRET measurement produces the conformational distribution of the molecule.

For the downhill folder BBL, we would expect a unimodal conformational distribution from the SM-FRET measurement. That would provide a direct evidence to demonstrate downhill folding in BBL.

1.4 Introduction to Single Molecule Förster Resonance Energy Transfer (SM-FRET)

1.4.1 Single molecule Spectroscopy (SMS)

Optically based single molecule spectroscopy (SMS) uses a laser beam to electronically excite and then detect the emission of ONE fluorescent molecule. This molecule can be immobilized at an interface or it can be freely diffusing in a liquid. Moerner and co-workers [27] were the first to measure the absorption of a single molecule directly in an experiment that had the very strict requirement of cryogenic temperatures. A year later Orrit and co-workers [28] measured the absorption of a single molecule indirectly by detecting its fluorescence emission. Emission detection is used in most current single molecule spectroscopy experiments because fluorescence is much more sensitive and more convenient to measure than absorption. SMS removes the ensemble averaging completely, thus providing information on the stochastic behavior of molecules and enables the monitoring of equilibrium fluctuations.

Typical setup of SMS. For immobilized molecules or very big particles such as cells, the wide field method and scanning method are two successful setups to conduct single molecule measurements. In wide field methods, the field of view (10 to 100 microns) of a microscope is imaged onto a 2-dimensional detector, which is usually a charge coupled device (CCD). It can monitor several chromophores simultaneously and track their positions at frame rates faster than the video scan rate. The spatial resolution is set by the diffraction limit: $\frac{1.22\lambda}{N.A.}$ where λ is the wavelength of beam and N.A is the numerical aperture of the objective. With a sufficient signal-

to-noise ratio (SNR), the location of the chromophores can be determined to better than the diffraction limit. The scanning method, on the other hand, probes very small illumination volumes each time and finally builds up a full image by scanning the view field. The confocal microscope SMS is one of the widely used scanning methods.

Total internal reflection SMS is another widely used setup working for surface immobilized samples. Because the refractive index of the glass slip is larger than that of the water underneath, the pump beam will be totally reflected when the incidence angle goes beyond the critical angle. This generates an exponentially decaying electromagnetic field (evanescent field) perpendicular to the surface of the glass cover slip. The density of an evanescent beam is given by $I(z) = I(0)\exp(-z/d)$ where *I(0)* is the initial density, *z* is the depth into solution from the surface and the decay distance *d* is ~150 nm [29]. Therefore, only those molecules that are very close to the cover slip will be excited and the background contribution from molecules in the bulk solution is minimized.

If the molecule is relatively small (e.g. a single domain protein) and not immobilized, it will be able to diffuse fast. In this case, it is difficult to get a clear image of the diffusing molecules by either wide field or scanning method. A practical way is to fix the probe volume and measure the signal intensity when a molecule diffuses through it. Such single molecule spectroscopy for free diffusing molecules [30-33] was employed in our measurement which will be described in Chapter 2.

Critical Issues of SMS for Freely Diffusing Molecules. The prerequisite of SMS is to make sure only one molecule can be excited in the probe volume that can

be as small as 0.1 fL. For a solution at the typical operating concentration of 100 pM, the average number of molecules in the probe volume is 0.005 or less. This ensures the presence of just one molecule in that probe volume for most of the time. Good signal-to-background-ratio is the other key point of SMS. The majority of background is from the scattering of excitation beam which includes both Rayleigh scattering (i.e. elastic scattering) and Raman scattering (i.e. inelastic scattering) [29]. So using proper filters to remove the excitation scattering is very important. Another kind of background is from the out-of-focus stray light. This can be reduced by a spatial filter such as pinhole aperture.

As one type of SMS method, single molecule Förster resonance energy transfer (SM-FRET) measures the energy transfer efficiency between the donor and acceptor fluorophores in a doubly-fluorophore-labeled molecule. This particular technique has been exploited by our group to characterize the denaturation of BBL and a two-state-like protein as a control. This technique is explained in detail as follows.

1.4.2 Förster Resonance Energy Transfer (FRET)

Förster resonance energy transfer (FRET) has been widely applied in the ensemble biophysical measurement for several decades [34]. It is the transfer of energy through a dipole-dipole interaction between the excited state of the donor and the ground state of the acceptor. The efficiency of the transfer process depends on the overlap between the donor emission spectrum and the acceptor absorption spectrum. The energy transfer rate is given by:

$$k_T(r) = \frac{QY_D\kappa^2}{\tau_D \cdot r^6}\left(\frac{9000\ln(10)}{128\pi^5 N_A \cdot n^4}\right)J \tag{1.5}$$

QY_D is the donor quantum yield; τ_D is the donor lifetime; κ^2 is the orientation factor decided by the dipoles of the donor and acceptor; r is the distance between the donor and acceptor; N_A is Avogadro's number; n is the refractive index of the medium; J is the overlap integral of the donor emission spectrum and the acceptor absorption spectrum:

$$J = \frac{\int F_D(\lambda)\varepsilon(\lambda)\lambda^4 d\lambda}{\int F_D(\lambda)d\lambda} \tag{1.6}$$

where $F_D(\lambda)$ is the donor fluorescence spectrum and $\varepsilon(\lambda)$ is the acceptor extinction spectrum in molar extinction units ($cm^{-1}M^{-1}$).

The orientation factor κ^2 is given by:

$$\kappa^2 = (\cos\theta_T - 3\cos\theta_A \cos\theta_D)^2 \tag{1.7}$$

As shown in Figure 1.2, θ_D and θ_A are the angles between the dipoles and the vector joining the donor and acceptor. θ_T is the angle between the two dipoles. If it is assumed that the donor and acceptor rapidly change orientations on the time scale of the fluorophore's lifetime we are able to use an average $\kappa^2 = \frac{2}{3}$. This assumption may not work in certain cases. But in the case of labeled proteins in aqueous solution, using steady state polarization measurements, several groups have measured anisotropies of about 0.05~0.1 [33, 35]. This indicates that the orientation change is fast enough compared to the fluorophore lifetime. Our steady state measurements also gave similar results. Moreover, in our analyses we do not intend to extract inter-dye

distances from the FRET efficiency. So the validity of the orientational averaging will not affect our results.

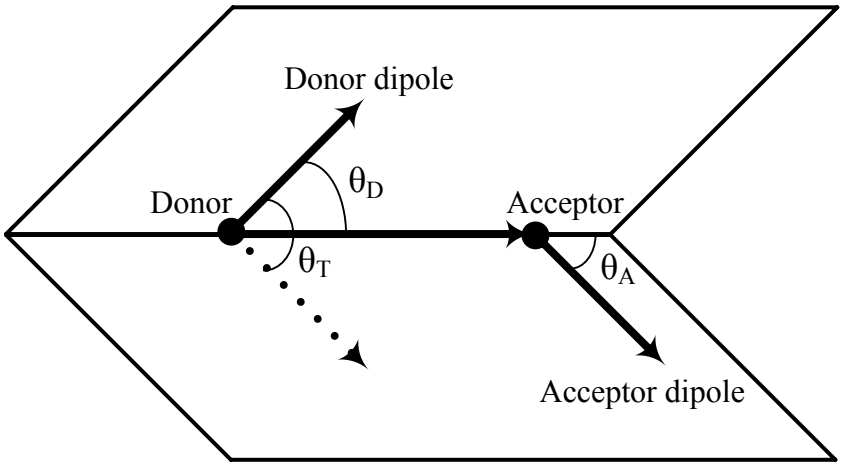

Figure 1.2 Illustration of orientation factor κ^2: θ_D and θ_A are the angles between dipoles and the vector joining the donor and acceptor, θ_T is the angle between two dipoles.

From Eq. (1.5) we can find the transfer rate is $\propto 1/r^6$. So it is often written as:

$$k_T(r) = \frac{1}{\tau_D}\left(\frac{R_0}{r}\right)^6 \tag{1.8}$$

Förster distance R_0 is defined as the distance between the donor and acceptor when $k_T(r) \approx {\tau_D}^{-1}$. It can be calculated by:

$$R_0^6 = QY_D\kappa^2\left(\frac{9000\ln(10)}{128\pi^5 N_A \cdot n^4}\right)J \tag{1.9}$$

Then the energy transfer efficiency is given by:

$$E = \frac{k_T}{\tau_D^{-1} + k_T} = \frac{1}{1+(r/R_0)^6} \tag{1.10}$$

R_0 is usually in the range from 2 to 9 nm which makes FRET very useful in the studies of biological macromolecules including proteins. Since the energy transfer is decided by many factors and sensitive to the immediate chemical environment of the fluorophore, it is more reliable to compare the relative inter-dye distance rather than calculate the absolute distance.

It is challenging to experimentally measure the energy transfer rate k_T and donor life time τ_D and then use Eq. (1.10) to calculate the FRET efficiency. Alternatively, the fluorescence intensities I_A and I_D of the donor and acceptor in a certain acquisition window is used to calculate the FRET efficiency:

$$ET = \frac{I_A}{I_A + I_D} \tag{1.11}$$

Since the measured signals are only parts of I_A and I_D due to the effect of QY, transmissivity and detector efficiency, proper corrections of the measured signals are needed in order to get the accurate I_A and I_D which will be discussed in Chapter 2.

1.4.3 SM-FRET

The combination of single molecule spectroscopy and FRET provides us a very powerful technique. The SM-FRET measurement on freely diffusing molecules produces a FRET efficiency distribution that can be used to extract the conformational distribution of the molecule. When the molecule is immobilized, SM-FRET detects the FRET trajectory of an individual molecule which in turn provides dynamical information at the level of single molecule. When different species are labeled with donor and acceptor, SM-FRET can track the reaction progress between

those species. Several groups have applied SM-FRET to studies of protein folding and obtained important information on protein dynamics [32, 33, 36, 37].

1.5 Brief Review of SM-FRET Applications in Protein Folding

SM-FRET was first applied to protein folding by the Weiss group [32] and Hochstrasser group [31], respectively. They observed the bimodal FRET efficiency distribution close to the midpoint of the transition in protein CI2 and GCN4, respectively, which is the landmark of two-state folding. Then the Eaton group estimated the free energy barrier of protein Csp *Tm* using SM-FRET [33].

Measuring the end-to-end distance of proteins is one of the most direct applications of SM-FRET that have been attempted by couple of groups [38, 39]. Schuler et. al. however obtained inconsistent results when comparing the inter-dye distance calculated from FRET efficiency and the expected length of rigid poly-proline peptides [38]. Later, Best et. al found that these inconsistencies are due to the non-rigid nature of poly-proline peptides as a result of cis-trans peptide bond transitions [40]. In fact, one of the major challenges in applying SM-FRET to measure protein end-to-end distance is the lack of rigid and well characterized spacer molecules. The other challenge is that the calculation of accurate FRET efficiency needs complicated corrections due to instrumentations, thus adding significant uncertainty to the result.

Observing the heterogeneity of folding pathways is another advantage of SM-FRET over ensemble measurements. The Haran group mapped the change of FRET efficiency between two consecutive acquisition time windows for protein adenylate kinase (AK) [35]. Since different FRET efficiencies represent different conformations,

the diverse distribution of the FRET efficiency changes indicated disperse conformation changes, i.e folding / unfolding pathways. Ideally, we should be able to extract information about the free energy surface and dynamics from these measurements, but so far there have been no quantitative estimates. The key points are to reduce the acquisition time and the effects of photoblinking and photobleaching which can introduce artificial transitions.

The Weiss group [41, 42] introduced the alternative excitation technique. They excited the fluroephore by several beams (usually two beams) with different wavelengths alternatively. Each beam was switched to the other in a certain adjustable frequency. This technique enables illuminating the donor and acceptor of labeled proteins almost simultaneously. That facilitates the recognition of donor-only proteins which results the zero-peak in the SM-FRET efficiency histogram.

Another technical progress in SM-FRET measurement of protein folding is the use of pulse laser and time correlated single photon counting (TCSPC) technique [37, 39, 40, 43-45]. It not only measures the photon number but also the timing information of each photon. With such techniques, Schuler and et.al obtained the reconfiguration time ~ 50 ns for the unfolded state of protein Csp and found the collapse of the peptide affected the unfolded state dynamics [37]. The photon timing information can also be used to calculate the fluorescence correlation function (FCS). The Haran group calculated the hydrodynamic radius of protein L from the FCS and compared it with the radius of gyration determined from the mean of SM-FRET efficiency [39].

1.6 Research Plan and Chapter Summary

My objective is to measure SM-FRET efficiency histograms of BBL and obtain direct information about protein conformational distributions in order to demonstrate global downhill folding in BBL. As a control experiment, the SM-FRET measurement of a two-state folding protein − α-spectrin SH3 has also been carried out. We expect to obtain the unimodal and bimodal FRET efficiency distributions for BBL and α-spectrin SH3, respectively. Finally, we will try to get some other information about protein dynamics from the analysis of SM-FRET efficiency histogram width. This thesis is organized as follows:

Chapter 2: Methods and Materials. Here, the freely diffusing SM-FRET confocal microscope system that was used to perform our single molecule measurement is described in detail. Other ensemble measurement instruments including the UV-Vis spectrophotometer and spectrofluorimeter are also introduced. The formula to calculate accurate FRET efficiency from both single molecule and ensemble measurements is developed and the parameters needed to calculate the FRET efficiency are determined. The design of our home made sample temperature control system is presented. At the end, the purification and labeling procedures of BBL and α-spectrin SH3 are detailed.

Chaper 3: Fluorophore Protection. To characterize protein conformational distributions, the binning time of SM-FRET has to be shorter than the relaxation time of a protein. Otherwise the obtained FRET efficiency will be the average of several possible conformational states. From dynamical studies, Li, et.al measured the relaxation time of BBL to be about 20 μs at room temperature and 120 μs at 279 K [46].

So the SM measurements have been carried out at 279 K with a binning time of 50 μs or shorter. As a result, the pump power had to be increased to 160 μW or higher to increase photon counts. Such a high excitation power will cause very frequent formations of triplet and other dark states in fluorophores. To avoid this problem we found the effective Trolox-cysteamine system to protect the fluorophore. In chapter 3, the method to protect fluorophores is presented.

Chapter 4: BBL Experimental Results and Two-State Simulation. We performed both ensemble and single molecule FRET measurements of the denaturation of BBL. With careful correction, they give consistent results. The denaturation profiles from both urea and GuHCl show clear unimodal profiles in SM-FRET histograms. We also simulated the SM-FRET histogram of BBL based on a two-state folding assumption that resulted in a bimodal distribution. The discrepancy between the experimental results and simulation clearly demonstrates global downhill folding in BBL.

Chapter 5: Characterization of α-spectrin SH3. To compare with BBL, we characterized the denaturation of α-spectrin SH3 in urea and GuHCl using ensemble and single molecule FRET measurements. The clear bimodal SM-FRET efficiency histogram of this protein confirmed that our instruments detected sufficient signals from single molecules and the observation of downhill folding in BBL is not an artifact of the instrumental setup or experimental conditions.

Chapter 6: Analysis of SM-FRET Efficiency Histogram Width. The width of SM-FRET efficiency histogram is mostly due to shot noise. However, it has been noted that the unfolded peak has some extra width which can be caused by protein

dynamics or other effects [33]. If the extra width is due to dynamics, it should decrease with larger binning time. In this chapter, the analysis of the extra width with varying binning times is presented. The single molecule measurements of several model systems are analyzed, which include a fluorescent protein, BBL and α-spectrin SH3. Possible explanations for the extra width are also discussed.

Chapter 2 Methods and Materials

2.1 Single Molecule Fluorescence Microscope System

2.1.1 Brief Description of the System

We collaborated with Dr. English's group to perform our single molecule fluorescence measurements on their home made single molecule fluorescence microscope system [47, 48]. The fundamental component of this system is an inverted microscope (Axiovert 200, Carl Zeiss MicroImaging). As shown in **Error! Reference source not found.**B, the 488 nm excitation laser beam from an Argon Ion laser (532-AP-A01, Melles Griot) is reflected by a 515 nm dichroic mirror (515DCLP, Chroma Technology) into an objective (N.A 1.3 ×100, oil immersion FLUAR, Carl Zeiss MicroImaging). Then the objective focuses the beam into the sample solution at a depth of 20 μm.

The fluorescence emission is collimated by the same objective becoming a parallel beam. After passing through the 515 nm dichroic mirror and a 488 nm holographic notch filter (HNF-488.0-1.0, Kaiser Optical Systems), the beam is focused by the tube lens into a beam spot which was the image of focus (probe volume) in principle.

2.1.2 New Design of the Transform Lens Working without Pinhole Aperture

In a typical single molecule confocal microscopy setup, a pinhole is placed in the image plane of objective focus as shown in **Error! Reference source not found.**A to block out-of-focus light [29, 49, 50]. Then a pair of transform lenses is used to

collimate and refocus the beam which passes through the pinhole onto the detector window in the same way as an objective and tube lens [51]. However, we do not use the pinhole in our system with the transform lens modification.

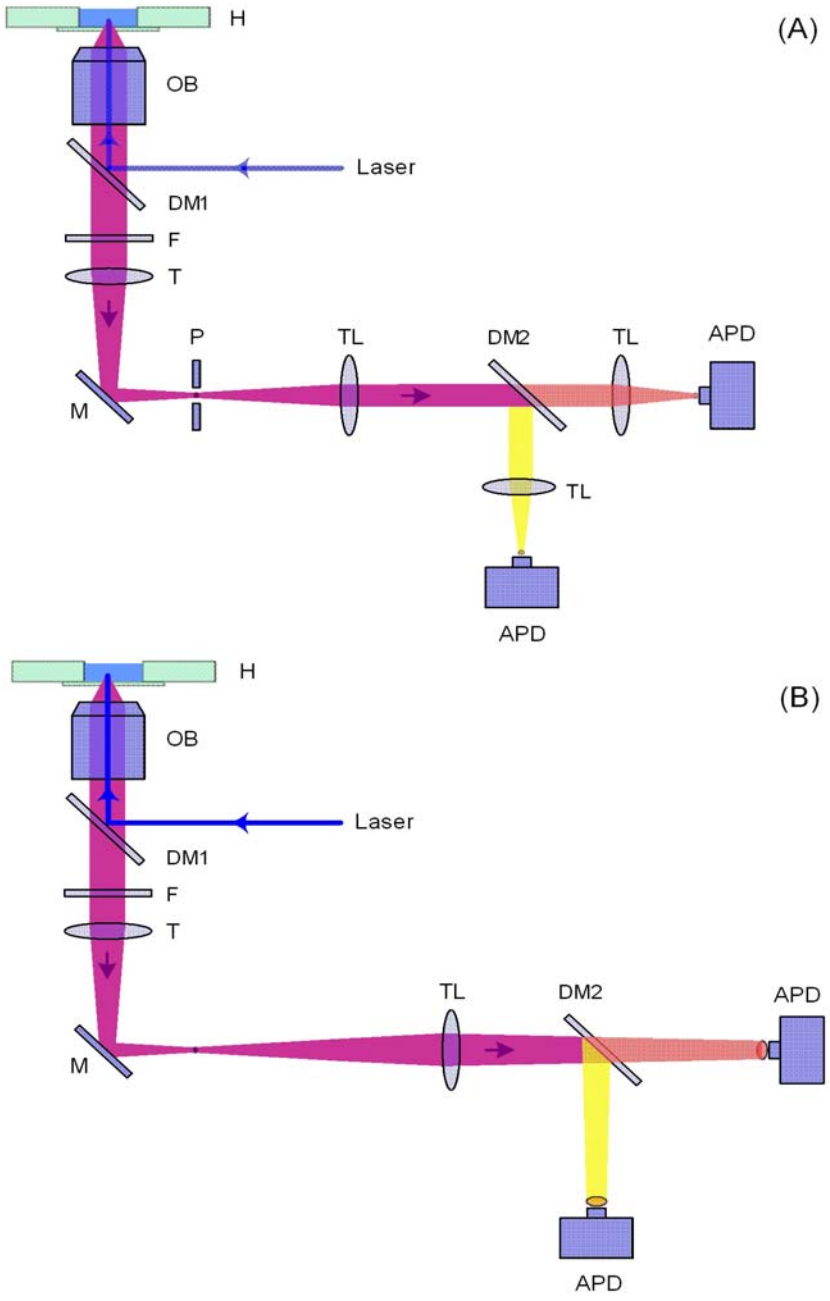

Figure 2.1 Experimental single molecule confocal microscope setup. (A), the typical setup using pinhole aperture and (B), our system without pinhole aperture. H is the sample holder, OB is the objective, DM1 is the 515 nm dichroic mirror, F is the laser

filter, T is the tube lens, M is a mirror, P is the pinhole aperture, TL is the transform lens, DM2 is the 585 nm dichroic mirror, APD is the avalanche photon detector.

We know that a fluorescence beam has very broad spectrum in terms of wavelength. Ideally the superposition of beams with different wavelengths should be focused in the same spot because the objective and tube lens are designed to be achromatic optics. The pinhole will then be placed in that spot where the beams are focused.

However, in the beam path there are a few components such as glass cover slip, objective oil, aqueous solution and so on which are not achromatic. As a result, the focus image of different wavelength lights is not overlapped but dispersed along the beam transmission direction. When we put a pinhole aperture in the beam path it is inevitable that it will block some fluorescence light. This loss of photons and the excitation power limit (large excitation power will cause triplet and other dark states thus affecting the photon emission, which is discussed in Chapter 3) result in the typical binning time of freely diffusing SM-FRET measurements on the order of several hundred μs. However, the relaxation time of downhill protein BBL was characterized as short as 20 μs at room temperature [46]. Even at 279 K, it is still 120 μs [46] at which the binning time of several hundred μs is not sufficient. So we propose a new scheme that does not use a pinhole in order to reduce the loss of fluorescence photons as shown in **Error! Reference source not found.**B.

Initially, people used photomultiplier tube (PMT) detectors in confocal microscope. Since the PMT has a large detection window, the pinhole had to be employed to block out-of-focus light. However, the window size of avalanche photon

detector (APD) used in our system is only ~180 μm [52]. It is twice that of the 100 μm diameter of typical pinhole aperture used in single molecule microscope. So if we set up the transformation lens properly and magnify the focus image twice, the APD window can itself work as a pinhole aperture. Also, we use only one transformation lens to transform the focus image to the detector instead of two transformation lenses as shown in **Error! Reference source not found.**B.

To determine the proper position of the transformation lens, we can use the image equation $\frac{1}{u}+\frac{1}{v}=\frac{1}{f}$ and the magnification equation $M=\frac{|v|}{|u|}$ where u, v, f and M are object distance, image distance, focus length and optical magnification amplitude, respectively. Given the geometry of the instrument, we chose a 6 cm convex lens as the transform lens. The object distance (from the focus image which is the object here to the transformation lens) and image distance (from the transformation lens to the detector) were set as 9 and 18 cm, respectively. Theoretically, the focus image will be amplified twice with this setup. However, the identification of focus image position is with large error. To check the beam amplification, we estimated the beam spot size on the detector window by monitoring the change of photon counts when the APD detector was moved laterally. We obtained the FWHM size of beam spot on the detector window to be ~130 μm. Given that the FWHM of the focus was ~0.6 μm and the focus was amplified 100 times by the objective, the real amplification amplitude of the transformation lens was close to 2 as we designed.

2.1.3 Photon Collection

After passing through the transform lens, the fluorescence beam is divided by the 585 nm dichroic mirror (DCXR585, Chroma Technology) into donor and acceptor beams that are detected by two APD (SPCM-AQR-14, PerkinElmer), respectively. Photon counts from two detectors are recorded by a counter/timer board (PCI 6602, National Instruments) with binning times from 20 μs to 1 ms. Then the data is transferred from the PCI board to a personal computer. The data acquisition process is controlled by a Lab windows (National Instruments) program.

2.2 Ensemble Measurement Instrumentation

2.2.1 Absorbance Measurement

According to the quantum mechanics, molecules have many electronic, vibrational and rotational states available, but usually electrons of the molecule are in the ground electronic and vibrational states. When the excitation light has the photon energy equal to the energy difference between certain higher energy state (excited state) and the ground state of the molecule, some excitation photons will be absorbed and the electron undergoes a transition to the excited state. The absorption of visible light happens when electrons transfer between different electronic states.

We used a UV-Vis spectrophotometer (Cary 100, Varian) to measure the optical absorbance spectrum. The temperature of the sample block was always set at 298 K. The integration time was chosen between 0.1~0.5 s/nm according to the sample's concentration. 1 cm pathlength cuvette was used.

2.2.2 Fluorescence Measurement

The excited electrons usually go back to the original ground state by dissipating the energy as heat into the medium. However, the excited electrons will sometimes emit photons as fluorescence or phosphorescence when they revert to the ground states. The fluorescence is emitted from the electronic singlet state which has a shorter lifetime in the order of ns while phosphorescence is from triplet which has a longer lifetime of about μs.

A spectrofluorimeter (Flurolog-3, Jobin Yovin, Inc.) was used to measure the ensemble fluorescence spectrum. It has a peltier system to control the temperature of sample holder. The emission intensity S and excitation intensity R are measured simultaneously. The final fluorescence signal is given by S/R that corrects for the effect of lamp fluctuations. The quartz cuvette has a pathlength of 1 cm. For the ensemble FRET measurement, the sample concentration was 20 nM and the spectrofluoremeter parameters were: excitation and emission slits 3 nm, integration time 0.5 s, excitation wavelength 488 nm. For the Quantum yield measurements, the sample concentration varied and the instrument setup was: excitation and emission slits 2 nm, integration time 0.5 s, excitation wavelengths 488, 590 nm for donor and acceptor, respectively. For other measurements, the experimental condition will be noted specifically.

2.2.3 Circular Dichroism (CD) Measurement

Any light beam can be regarded as the superposition of left-handed circularly (L-) polarized beam and right-handed circularly (R-) polarized beam. CD is the phenomenon that L-polarized and R-polarized beams have different absorbances

when a beam passes through a medium. The modern CD instrument usually use ellipticity to characterize CD signal which is defined as $\theta(\lambda) = 32.98(\varepsilon_L(\lambda) - \varepsilon_R(\lambda)) \cdot l \cdot C$. $\varepsilon_L(\lambda)$, $\varepsilon_R(\lambda)$ are molar absorptivities of L- and R-polarized beams at wavelength λ in units of $M^{-1}cm^{-1}$; l is the pathlength of quartz cuvette in unit of cm; C is the protein molar concentration.

Biomolecules are usually asymmetric and thus have nonzero CD values (optically active). The amide group (peptide bond) of proteins produces very strong CD signals in the range of 190 ~ 250 nm. Importantly, different protein secondary structures show signature features in their CD spectra making CD a powerful tool to monitor protein folding. For example, α-helix spectrum has a negative band at 222 nm, a negative couplet at 208 nm and a positive couplet at 190 nm.

The CD measurements on BBL (shown in Chapter 4) was carried out in a Jasco CD spectrometer (J-815, JASCO, UK). The protein concentration was about 50 mM. The pathlength of the quartz cuvette was 1 mm. The instrument parameters were set as: continuous scanning mode, response time 16 s, bandwidth 2 nm, scanning rate 10 $nm \cdot min^{-1}$ and repetition 2. The sample temperature was maintained at 279 K.

2.3 Formula to Calculate Accurate FRET Efficiency

2.3.1 Basics of FRET Efficiency Calculation

Experimentally FRET efficiency (ET) is calculated as the fraction of donor absorbed photons that are transferred to the acceptor. If we refer the transferred

photons as I_A and untransfered photons as I_D (sum of radiatively and nonradiatively emitted photons from donor) ET will be:

$$ET = \frac{I_A}{I_A + I_D} \tag{2.1}$$

Because of nonradiative processes such as excited state quenching and the photon-loss during transmission, corrections are needed to calculate I_D and I_A from the measured signals of the donor and acceptor channels: S_D and S_A. All the variables and parameters used to calculate accurate ET are listed in Table 2.1.

Table 2.1 Description of variables and parameters for accurate ET calculation

Variable /parameter	Description	Value or ratio
I_D	Untransferred photons (sum of radiatively and nonradiatively emitted photons from the donor)	
I_A	Transferred photons (sum of radiatively and nonradiatively emitted photons from the acceptor)	
S_D, S_A	Measured signals (photon counts) from the donor and acceptor channels	
QY_D, QY_A	Quantum yields of donor and acceptor	$\Gamma_{QY} = \frac{QY_A}{QY_D}$
t_D, t_A	Transmission efficiencies for the 515 nm dichroic mirror for donor and acceptor beams	$\Gamma_{trans} = \frac{t_A}{t_D} = 1.26 \pm 0.03$
Φ_D, Φ_A	APD detection efficiencies of donor and acceptor channels	$\Gamma_{dtector} = \frac{\Phi_A}{\Phi_D} = 1.08 \pm 0.03$
Γ	$\Gamma = \Gamma_{dtector}\Gamma_{trans}\Gamma_{QY}$, photon collection efficiency ratio between acceptor and donor channels	
C_{DA}	Leak through from donor to acceptor channels (leaked photons divided by total photons)	0.08 ± 0.01

C_{AD}	Cross talk from acceptor to donor channels (leaked photons divided by total photons)	0.04 ± 0.01
r_{abs}	$r_{abs} = \varepsilon_A^{488} / \varepsilon_D^{488}$, extinction coefficient ratio between acceptor and donor s at 488 nm	0.05 ± 0.01
C_{label}	Acceptor labeling efficiency, fraction of donor-labeled protein that also has acceptor labeled	

First we consider only donor's absorption at 488 nm that produces most of the photons. The contribution from acceptor's absorption at 488 nm will be added into the formula later. In a doubly-labeled protein (protein labeled with one donor and one acceptor) the donor absorbed photons can be split into two parts: transferred photons I_A and radiatively /nonradiatively emitted photons I_D from the donor. The effect of nonradiative emission is represented by quantum yield (QY): the radiative emission is $QY_D I_D$. Those photons which are transferred to the acceptor are also being emitted finally and the part of radiative emission is $QY_A I_A$. Only the radiative emission is detected by one of the two channels partially based on the photon collection efficiency of the channel while the nonradiative emission is dissipated and is not observed.

After passing through the 515 nm dichroic mirror, both the donor and acceptor beams lose some intensities. The beam intensities immediately after this mirror are $t_D QY_D I_D$ and $t_A QY_A I_A$. Then in the 585 nm dichroic mirror, the cross talk between two channels happens. A small portion of donor beam leaks to the acceptor channel: $\Phi_A C_{DA} t_D QY_D I_D$ while some acceptor beam enters the donor channel: $\Phi_A C_{AD} t_A QY_A I_A$. Here we should use donor channel detector efficiency Φ_D, but to simplify the

derivation we use Φ_A. Based on the above analysis we get the formula for the measured signals:

$$\begin{aligned} S_D &= \Phi_D(1-C_{DA})t_D QY_D I_D + \Phi_A C_{AD} t_A QY_A I_A \\ S_A &= \Phi_A C_{DA} t_D QY_D I_D + \Phi_A(1-C_{AD})t_A QY_A I_A \end{aligned} \tag{2.2}$$

Direct Excitation of Acceptor. At 488 nm, acceptor's extinction coefficient is about 5% of donor's, which can introduce some error in ET calculation and hence needs consideration. Let the ratio of acceptor and donor's extinction coefficient at 488 nm be:

$$r_{abs} = \frac{\varepsilon_A^{488}}{\varepsilon_D^{488}} \tag{2.3}$$

Then acceptor's absorption can be written as:

$$I_A^{abs} = I_D^{abs} r_{abs} = r_{abs}(I_A + I_D) \tag{2.4}$$

Since I_A^{abs} is just several percent of the whole signal $I_A + I_D$ and since the cross talk between two channels can modify I_A^{abs} by only a small amount, we neglect the effect of cross talk on I_A^{abs}. In this way, all those photons arising from acceptor direct excitation will be added to the acceptor channel signal, which is $\Phi_A t_A QY_A r_{abs}(I_A + I_D)$. Then Eq. (2.2) becomes:

$$\begin{aligned} S_D &= \Phi_D(1-C_{DA})t_D QY_D I_D + \Phi_A C_{AD} t_A QY_A I_A \\ S_A &= \Phi_A(C_{DA} t_D QY_D + t_A QY_A r_{abs})I_D + (1-C_{AD}+r_{abs})\Phi_A t_A QY_A I_A \end{aligned} \tag{2.5}$$

Simplification of QY_A, t_A and Φ_A. Since only the ratios of quantum yield, transmission efficiency and detector efficiency can affect the calculation of ET instead of the absolute values, we can assume QY_A, t_A and Φ_A to be units and regard them as references: $QY_A = 1$, $t_A = 1$ and $\Phi_A = 1$. Then for the donor we have:

$$QY_D = \frac{QY_A}{\Gamma_{QY}} = \frac{1}{\Gamma_{QY}}$$

$$t_D = \frac{t_A}{\Gamma_{trans}} = \frac{1}{\Gamma_{trans}}$$

$$\Phi_D = \frac{\Phi_A}{\Gamma_{dtector}} = \frac{1}{\Gamma_{dtector}} \tag{2.6}$$

So the detected intensity of the signals (Eq. (2.5)) can be simplified as:

$$\begin{aligned} S_D &= (1 - C_{DA}) t_D QY_D \Phi_D I_D + C_{AD} I_A \\ S_A &= (C_{DA} t_D QY_D + r_{abs}) I_D + (1 - C_{AD} + r_{abs}) I_A \end{aligned} \tag{2.7}$$

where QY_D, t_D and Φ_D will be replaced by Γ_{QY}, Γ_{trans} and $\Gamma_{dtector}$ later. Now, we obtain the relation between the detected signals and the original transferred and untransferred photon numbers.

2.3.2 Formula of Accurate SM-FRET Efficiency

Actually Eq. (2.7) can be modified into:

$$\frac{(1 - C_{AD} + r_{abs})}{C_{AD}} S_D = \frac{(1 - C_{AD} + r_{abs})}{C_{AD}} (1 - C_{DA}) t_D QY_D \Phi_D I_D + (1 - C_{AD} + r_{abs}) I_A \tag{2.8}$$

and

$$(1 - C_{AD} + r_{abs}) I_A = S_A - (C_{DA} t_D QY_D + r_{abs}) I_D \tag{2.9}$$

Inserting Eq. (2.8) into Eq. (2.9) we have:

$$\frac{(1 - C_{AD} + r_{abs})}{C_{AD}} S_D = \frac{(1 - C_{AD} + r_{abs})}{C_{AD}} (1 - C_{DA}) t_D QY_D \Phi_D I_D + S_A - (C_{DA} t_D QY_D + r_{abs}) I_D \tag{2.10}$$

which is equivalent to :

$$\frac{(1 - C_{AD} + r_{abs})}{C_{AD}} S_D - S_A = \frac{(1 - C_{AD} + r_{abs}) S_D - C_{AD} S_A}{[1 - C_{AD} + r_{abs} - C_{DA} + C_{DA}(C_{AD} - r_{abs})] t_D QY_D \Phi_D - (C_{AD} C_{DA} t_D QY_D + C_{AD} r_{abs})} \tag{2.11}$$

First-order Approximation. Since C_{DA} , C_{AD} and r_{abs} are all small percentages, we neglect their corresponding second order terms and the products between them. This first-order approximation results in a simplified but accurate enough form of Eq. (2.11):

$$I_D = \frac{(1 - C_{AD} + r_{abs})S_D - C_{AD}S_A}{(1 - C_{AD} + r_{abs} - C_{DA})QY_D\Phi_D} \tag{2.12}$$

Because $r_{abs} - C_{AD}$ is almost zero, the formula can be simplified further:

$$I_D = \frac{S_D - C_{AD}S_A}{(1 - C_{DA})t_DQY_D\Phi_D} \tag{2.13}$$

By inserting Eq. (2.13) into Eq. (2.7) we get I_A:

$$I_A = \frac{S_A - \frac{(C_{DA}t_DQY_D + r_{abs})S_D - (C_{DA}t_DQY_D + r_{abs})C_{AD}S_A}{(1 - C_{DA})t_DQY_D\Phi_D}}{(1 - C_{AD} + r_{abs})} \tag{2.14}$$

As discussed above, all the second-order terms on C_{DA}, C_{AD} and r_{abs} can be neglected. So Eq. (2.14) is simplified into:

$$I_A = \frac{1}{(1 - C_{AD} + r_{abs})}\left[S_A - \frac{(C_{DA}t_DQY_D + r_{abs})}{(1 - C_{DA})t_DQY_D\Phi_D}S_D\right] \tag{2.15}$$

Given that C_{AD} and r_{abs} are similar, we get a more simplified formula:

$$I_A = S_A - \frac{(C_{DA}t_DQY_D + r_{abs})}{(1 - C_{DA})t_DQY_D\Phi_D}S_D \tag{2.16}$$

Now we replace QY_A, t_A and Φ_A in Eq. (2.13) and Eq. (2.16) with their corresponding ratios shown in Eq. (2.6):

$$I_D = \frac{S_D - C_{AD}S_A}{(1 - C_{DA})}\Gamma \tag{2.17A}$$

$$I_A = S_A - \frac{\left(C_{DA}\Gamma_{dtector} + r_{abs}\Gamma\right)}{\left(1 - C_{DA}\right)} S_D \qquad (2.17B)$$

This formula can be directly used to calculate accurate SM-FRET efficiency.

2.3.3 Calculation of Accurate Ensemble FRET Efficiency

We can employ the accurate SM-FRET efficiency formula of Eq. (2.17) to analysis ensemble FRET data except that some parameters have to be modified in the ensemble measurement.

In ensemble measurements the beam does not pass through a 515 nm dichroic mirror to remove the excitation beam. So the transmission efficiencies of 515 nm dichroic mirror are: $t_A = t_D = 1$ and the ratio of transmission efficiency is $\Gamma_{trans} = \frac{t_A}{t_D} = 1$.

Secondly, we collect the whole spectra including both donor and acceptor signals and then split it at 585 nm into two portions. The short and long wavelength portions are referred to donor and acceptor signals, respectively. There is no 585 nm dichroic mirror used. Because of the splitting at 585 nm, the cross talks between two channels are still present. Ensemble C_{DA} is only slightly different from the value used in single molecule measurements. So we use 8%. However, C_{AD} is only around 1% and so we neglect it in the calculation.

Moreover, the detection efficiency of the detector in the spectrofluoremeter has been corrected for all wavelengths. The detector efficiency for both donor and acceptor signals are identical thus $\Gamma_{dtector} = \frac{\Phi_A}{\Phi_D} = 1$. Incorporating the above changes,

we get the proper formula for ensemble FRET efficiency calculation based on Eq. (2.17):

$$I_D = \frac{S_D}{(1 - C_{DA})} \Gamma_{QY} \tag{2.18A}$$

and

$$I_A = S_A - \frac{(C_{DA} + r_{abs}\Gamma_{QY})}{(1 - C_{DA})} S_D \tag{2.18B}$$

Effect of Donor-Only Proteins. There is another difference between ensemble and single molecule measurements. Experimentally the sample always has some donor-only proteins (also called acceptor-inactive proteins). In the single molecule measurement, molecules are detected individually, and thus the presence of donor-only proteins does not affect the signal of other doubly-labeled proteins. However, in case of ensemble measurements, the signal from donor-only proteins will be mixed with those from doubly-labeled proteins. This means the untransfered photon number calculated by Eq. (2.18A) has the contribution from donor-only proteins.

If there is small amount of donor-only proteins, it is very reasonable to assume all the acceptor channel signal I_A is from doubly-labeled proteins. In other words, we can neglect the leak through of donor-only protein signal to acceptor channel. So for the donor channel signal, we have:

$$I_D = I_D^0 + I_D^{'} \tag{2.19}$$

where I_D^0 and $I_D^{'}$ refer to donor channel signals emitted by doubly-labeled proteins and donor-only proteins, respectively. Introducing the acceptor labeling efficiency

C_{label} (ratio between doubly-labeled proteins and all proteins with donor), we have the relation:

$$\frac{I_D^0 + I_A}{I_D^{'}} = \frac{C_{label}}{1 - C_{label}} \tag{2.20}$$

From Eq. (2.19) and (2.20) we get:

$$I_D^0 = C_{label} I_D - (1 - C_{label}) I_A \tag{2.21}$$

which is the donor signal produced by doubly-labeled proteins. Given the effect of donor-only protein, we get the formula to calculate ensemble FRET efficiency:

$$ET = \frac{1}{C_{label}} \frac{I_A}{I_D + I_A} \tag{2.22}$$

We can see the effect of donor-only protein on the ensemble FRET efficiency can be easily corrected. However, it is difficult to quanlify the percentage of doubly-labeled proteins C_{label} in the sample.

2.3.4 Background Subtraction

Background subtraction is a complicated issue for single molecule ET calculation. Schuler and coworkers used the photon count of solution without labeled proteins as the background [33]. But this does not sufficiently represent the background noise since the majority of background is from the scattering of out-of-focus molecules instead of the excitation beam scattering after the beam passes through the 488 nm laser filter and 515 nm dichroic mirror. So a better way is to select a piece of trajectory which has few photon counts and can be safely assumed to be no molecule present, and then use its average photon count as the background. Obviously such

method is very arbitrary. Since the background intensity is very low compared to the signal, we did not apply any background correction to our calculations.

2.4 Change of Förster distance R_0

If Förster distance R_0 is constant under different conditions, the accurate ET calculated using the above formula will be a good indicator of protein end-to-end distance. However, as discussed in Section 1.4.2, R_0 is dependent on medium refractive index (n) which changes with concentration of urea and GuHCl, and donor QY (QY_D) which also changes with denaturant concentration in some cases. This means the ET may vary with denaturant concentrations when the inter-dye distance does not change. If we want ET to present the protein end-to-end distance more accurately, we should correct for the changes in R_0 which is shown below.

Eq. (1.9) can be rewritten as:

$$R_0^6 = \frac{QY_D}{n^4}\kappa^2\left(\frac{9000\ln(10)}{128\pi^5 N_A}\right)J = \frac{QY_D}{n^4}\alpha \tag{2.23}$$

where $\alpha = \kappa^2\left(\frac{9000\ln(10)}{128\pi^5 N_A}\right)J$ is a constant under the assumption that fluorophore freely rotates. Assuming the inter-dye distance r is fixed, The ET in water and in a solution with denaturant concentration c is given by:

$$ET_0 = \frac{1}{1+(r/R_0)^6} = \frac{1}{1+r^6\Big/\frac{QY_{D0}}{n_0^4}\alpha} \tag{2.24}$$

and

$$ET(c) = \frac{1}{1+(r/R_0(c))^6} = \frac{1}{1+r^6\Big/\frac{QY_D(c)}{n^4(c)}\alpha} \tag{2.25}$$

From the above equations we have:

$$\frac{r^6 \Big/ \frac{QY_D(c)}{n^4(c)}\alpha}{r^6 \Big/ \frac{QY_{D0}}{n_0^4}\alpha} = \frac{\frac{1}{ET(c)} - 1}{\frac{1}{ET_0} - 1} \tag{2.26}$$

which is equivalent to:

$$ET_0 = \frac{ET(c)}{ET(c) + (1 - ET(c))\left(\frac{QY_D(c)}{QY_{D0}} \Big/ \left(\frac{n(c)}{n_0}\right)^4\right)} \tag{2.27}$$

If we set ratios $\gamma_{QY_D} = \frac{QY_D(c)}{QY_{D0}}$ and $\gamma_n = \frac{n(c)}{n_0}$, Eq. (2.27) will become:

$$ET_0 = \frac{ET(c)}{ET(c) + (1 - ET(c))\left(\gamma_{QY_D} / \gamma_n{}^4\right)} \tag{2.28}$$

Or

$$ET_0 = \frac{1}{1 + (\frac{1}{ET(c)} - 1)\left(\gamma_{QY_D} / \gamma_n{}^4\right)} \tag{2.29}$$

The accurate ET calculated in Section 2.3 is equivalent to *ET(c)*.With this equation we will be able to determine the corresponding ET_0.

2.5 QY Measurements

2.5.1 QY Measurement in Aqueous Solution

To calculate accurate ET, we have to know the QY ratio between donor and acceptor. QY of a fluorophore is the fraction of absorbed photons that are emitted. The typical way to measure a sample's QY involves the use of a well characterized

QY standard [34, 53]. By comparing the fluorescence and absorbance ratio between the unknown sample and standard we can obtain the QY of the unknown sample. To improve the accuracy, people usually measure the sample and standard with varying concentrations and then the QY is calculated as [34]:

$$QY = QY_{ref} \frac{Slope}{Slope_{ref}} \left(\frac{n}{n_{ref}} \right)^2 \tag{2.30}$$

The subscript $_{ref}$ refers to the QY standard, n is the refractive index of solution, $Slope$ is the slope of linear fit between total fluorescence intensity and absorbance at the excitation wavelength. This equation is based on the presumption that QY is independent of fluorophore concentration, which requires the absorbance of the fluorophore to be less than 0.1. Our measurement had the fluorophore absorbance below 0.04 as suggested by Lacowicz [34].

Correction Factor. For accurate QY calculations we need to consider a correction for the spectrum cut. In ensemble fluorescence measurements we usually take the spectrum from a wavelength which is couple of nms longer than the excitation wavelength to avoid including the excitation beam. Therefore, we might lose some emission whose wavelength is very close to the excitation wavelength. This means we do not collect the whole emission spectrum. So we use a correction factor f^{cor} to deal with this issue. f^{cor} is the ratio between the integration of whole emission spectrum and the measured spectrum. If we multiply the fitting slope by the correction factor f^{cor} we will get a more accurate slope.

QY Standards. We chose Rhodamine 6G (R6G) and Cresyl violet perchlorate (Cresyl violet) as the standards for the QY measurement of donor and

acceptor, respectively, because they are widely used standards and have similar absorption and emission spectra as our labels. Another standard Fluorescein was also measured in order to check the measurement accuracy of R6G and Cresyl violet. The integrated fluorescence intensity of the emission spectrum vs. absorbance at the excitation wavelength was fitted linearly as shown in Figure 2.2. The fitting slope was used to determine QY of other samples which is listed in Table 2.2. Using R6G as the reference, the measured QY of fluorescein and Cresyl violet are consistent with the literature values with 98% agreement. This indicates that our QY measurement is very reliable.

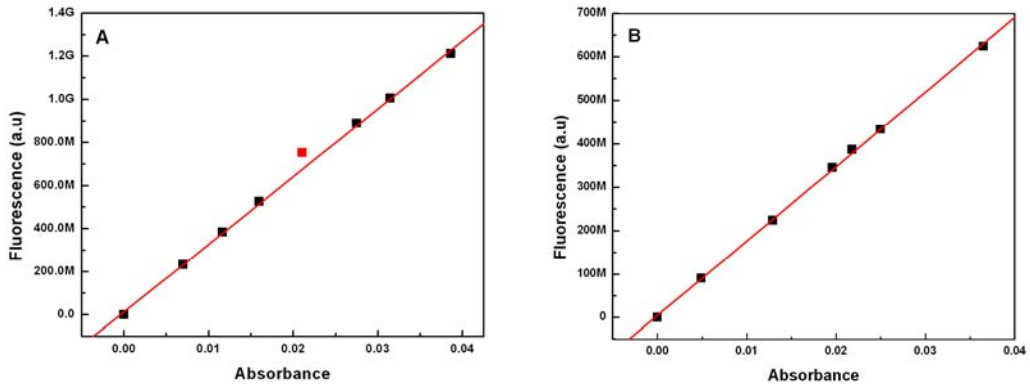

Figure 2.2 Measurement results of QY standards R6G (A) and Cresyl violet (B). The fluorescence intensity is plotted vs. the absorbance at the excitation wavelength (488 nm for R6G and 590 nm for Cresyl violet). The slope of the linear fit (red line) is used to calculate QY of other samples. The marked point (red square) is not used in the fitting.

Labeled proteins in aqueous solution. Because a protein may protect the fluorophore from quenching by the solvent [34], the labeled protein usually has different QY compared to free label. Also, different proteins could have different effects on the label. For both BBL and α-spectrin SH3, we measured the QY of donor-only and acceptor-only proteins. The sample preparation is described in Section 2.7.3. The

fluorescence intensity vs. absorbance at the excitation wavelength was fitted linearly as we did with the QY standards. The fitting slope and obtained QY are listed in Table 2.2.

Table 2.2 QY measurement results

Sample or standard	R6G	D-BBL	D-SH3	Cresyl violet	A-BBL	A-SH3
Refractive Index	1.359	1.335	1.333	1.329	1.335	1.333
Fitted Slope ($\times 10^{10}$)	3.14±0.04	2.62±0.06	2.02±0.05	1.71±0.02	2.07±0.04	1.94±0.02
Correction factor	1.00	1.01	1.01	1.03	1.04	1.04
Corrected Slope ($\times 10^{10}$)	3.14±0.04	2.65±0.05	2.06±0.05	1.76±0.02	2.15±0.03	2.02±0.02
QY	0.95	0.77±0.03	0.60±0.04	0.53	0.65±0.01	0.61±0.01

2.5.2 QY in Urea and GuHCl Solution

The other issue with QY is that it may change with chemical denaturant. So in order to accurately calculate ET during protein denaturation we measured the QY of single labeled proteins (BBL and α-spectrin SH3) at different concentrations of urea and GuHCl.

In this measurement, we used a mixed sample of donor-only and acceptor-only proteins that produced similar absorbance at their respective maximum absorption wavelengths. Using such a mixed sample we took the absorption spectrum from 700 nm to 400 nm that included both acceptor and donor's absorption range. Then the sample was excited at 488 nm and 590 nm to take the fluorescence spectrum of donor and acceptor, respectively.

This method of using a mixture of proteins meant that the measurements were taken in exactly identical conditions for both proteins, thus avoiding some errors. In the measurement, a small percentage of acceptor signal was introduced into the donor measurement, but the effect was negligible. The acceptor's absorption at 488 nm is around 5% of its maximum absorption at 590 nm which is pretty small. Moreover, not only was the acceptor's absorption at 488 nm included in donor's absorption but also the resulting acceptor emission was treated as donor emission. Since QY is decided by the ratio between fluorescence and absorbance, the induced error was much less than 5% and hence negligible.

For each measurement, we made two sets of samples. One had the maximum absorbance around 0.02 and the other one around 0.04. The buffer conditions were identical to the ensemble and single molecule FRET measurements of a given protein (Section 2.8). All the measurements were made at 298 K.

The QY data was quite noisy, but it is not used to calculate ET. It is only used to analyze the effect of Förster distance R_0 changes on ET. We did a second-order fit to donor QY dependence on the denaturants. The QY ratio $\Gamma_{QY} = \frac{QY_A}{QY_D}$ did not fluctuate much and can be adequately fitted to a linear function. The results of QY dependence on denaturant concentration and their fitting are given below.

QY in Urea Solution. The results for labeled BBL and α-spectrin SH3 in urea are shown in Figure 2.3. For BBL, donor QY increases with urea concentration. The second order fit gave: $QY_D = (0.766 \pm 0.012) + (0.025 \pm 0.007)[Urea] - (0.0014 \pm 0.0007)[Urea]^2$. From the data and fit, we can find that the donor QY of BBL increases with low urea

concentrations and then becomes relatively flat at high urea concentrations (larger than 6 M). This shows that the addition of urea in solution enhances donor QY but the enhancement gets saturated at 6 M urea.

The QY ratios $\Gamma_{QY} = \frac{QY_A}{QY_D}$ of most data points are very close to 0.84 (the value of labeled BBL in buffer) with the discrepancy less than 0.02. Since QY measurement usually has the error of several percent it is very reasonable to assume $\Gamma_{QY} = 0.84 \pm 0.04$ works for all urea solutions.

For α-spectrin SH3, the donor QY increases steeply with urea concentration. The second order fit gave $QY = (0.603 \pm 0.019) + (0.053 \pm 0.010)[Urea] - (0.0022 \pm 0.0011)[Urea]^2$. Shown in Figure 2.3D, Γ_{QY} decreases almost linearly with urea concentration. So the linear function $\Gamma_{QY} = (0.98 \pm 0.01) - (0.020 \pm 0.001)[Urea]$ fits the data very well.

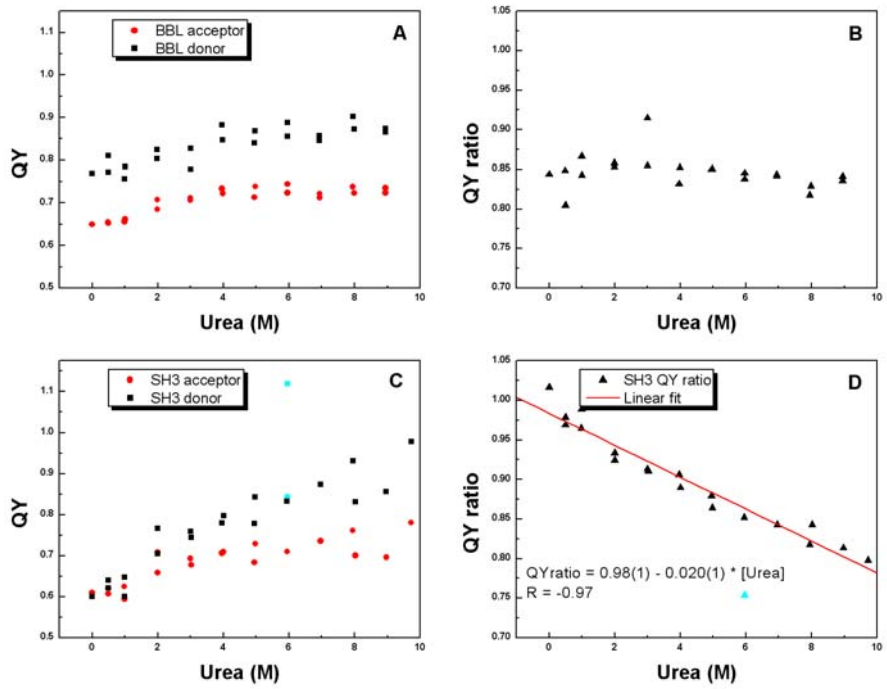

Figure 2.3 QY (black square for acceptor-labeled protein and red circle for donor-labeled protein) and QY ratio (black square) of singly-labeled proteins in urea solution. (A) and (B) are QY and QY ratio of BBL. (C) and (D) are QY and QY ratio of α-spectrin SH3. The QY ratio of α-spectrin SH3 is fitted linearly (red line) without the cyan point.

QY in GuHCl. The QY measurements in GuHCl solution were carried out in the same way as that in urea. As shown in Figure 2.4A, BBL donor QY was fitted as: $QY = 0.781(25) + 0.023(17)[\text{GuHCl}] - 0.0011(23)[\text{GuHCl}]^2$. Shown below in Figure 2.4B, BBL QY ratio fluctuates significantly. The absorption measurement of low concentration samples may have large errors due to the baseline change. However, the QY ratio vs. GuHCl concentration only shows a small decreasing trend. So a linear fit is good enough to describe the change: $\Gamma_{QY} = (0.83 \pm 0.02) - (0.003 \pm 0.004)[GuHCl]$.

For α-spectrin SH3, the donor QY was fitted by $QY = 0.589(17) + 0.095(12)[\text{GuHCl}] - 0.0081(17)[\text{GuHCl}]^2$. The QY ratio drops dramatically between 0 M GuHCl and the rest of samples. With GuHCl concentration being larger than 1 M, the QY ratio changes slightly, so we can just use a linear function to fit the QYratio: $\Gamma_{QY} = (0.85 \pm 0.01) - (0.004 \pm 0.002)[GuHCl]$. For 0 M GuHCl, we chose $\Gamma_{QY} = 1.02$.

In the fitting equations of Γ_{QY}, the slope has very large relative error which is > 100 % for BBL and 50 % for α-spectrin SH3. However, because the slope is small, even with 7 M GuHCl, the absolute QY error caused by the slope fitting error is < 0.03. Provided that typical QY measurement error is around 0.02~ 0.05 [34], the above fit of QY-ratio is acceptable.

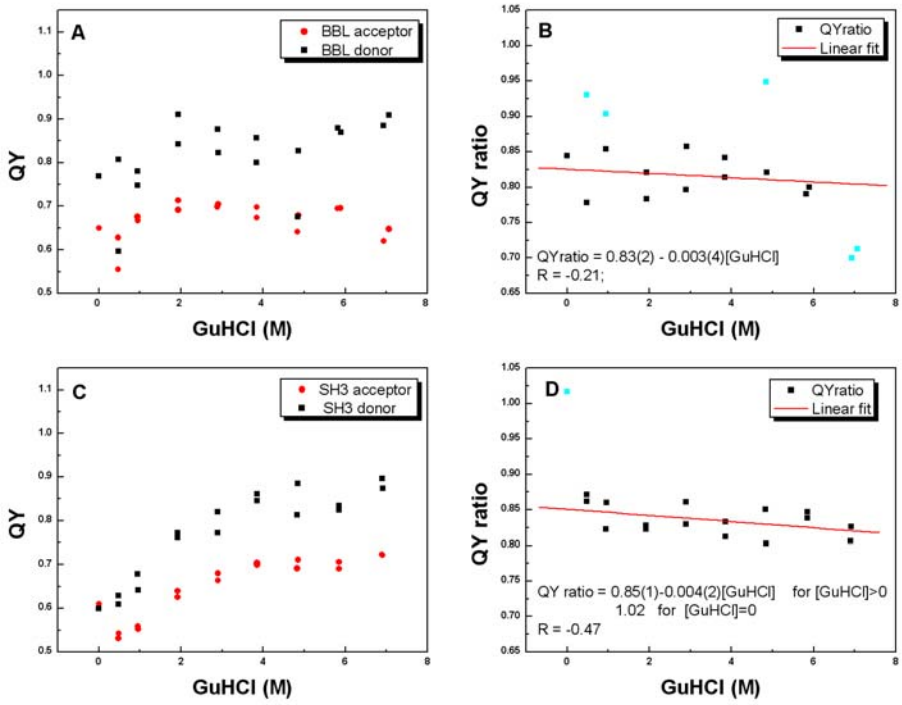

Figure 2.4 QY and QY ratio in GuHCl. The QY of BBL (A) and α-spectrin SH3 (C) is plotted vs. GuHCl concentration where the black square refers to acceptor-labeled

protein and red dot is for donor-labeled protein. (B), The QY ratio of BBL is fitted by a linear function (red line) where cyan points are not used because of poor quality. (D), The QY ratio of α-spectrin SH3 drops dramatically from 0 M GuHCl to the rest samples. The data points with nonzero GuHCl concentrations are fitted by a linear function (red line).

2.6 Measurement of Other Parameters for FRET Efficiency Calculation

2.6.1 Effect of Dichroic Mirrors on Fluorescence

There are two dichroic mirrors in the single molecule measurement setup. Fluorescence first passes through the 515 nm dichroic mirror. When we multiply the spectrum of incident donor beam by the transmittance spectrum of the mirror (shown in Figure2.5A) we get the transmission spectrum of donor (see Figure 2.5C). The deduction of the transmission spectrum from the original spectrum of incoming beam gives the reflection spectrum (also shown in Figure2.5C). The ratio between integrated transmission spectrum and the original spectrum is donor's transmission efficiency: $t_D = 0.744$. In the same way, we get the acceptor's transmission efficiency $t_D = 0.940$ (shown in Figure 2.5E). Thus the transmission ratio for the 515 nm dichroic mirror is $\Gamma_{trans} = \frac{t_A}{t_D} = 1.26$.

The fluorescence then passes through the 585 nm dichroic mirror whose transmittance spectrum is shown in Figure 2.5B. As above, we multiply the incident beam spectrum by the transmittance spectrum of 585 nm dichroic mirror and obtain the transmission spectrum (Figure2.5D, F). About 8.2% donor signal can pass through the dichroic mirror and enter the acceptor channel which is referred to as leak through from donor to acceptor channel C_{DA}. Similarly, 4.0% acceptor signal is

reflected by the 585 nm mirror and enters donor channel, which is the cross talk from acceptor channel to donor channel C_{AD} .

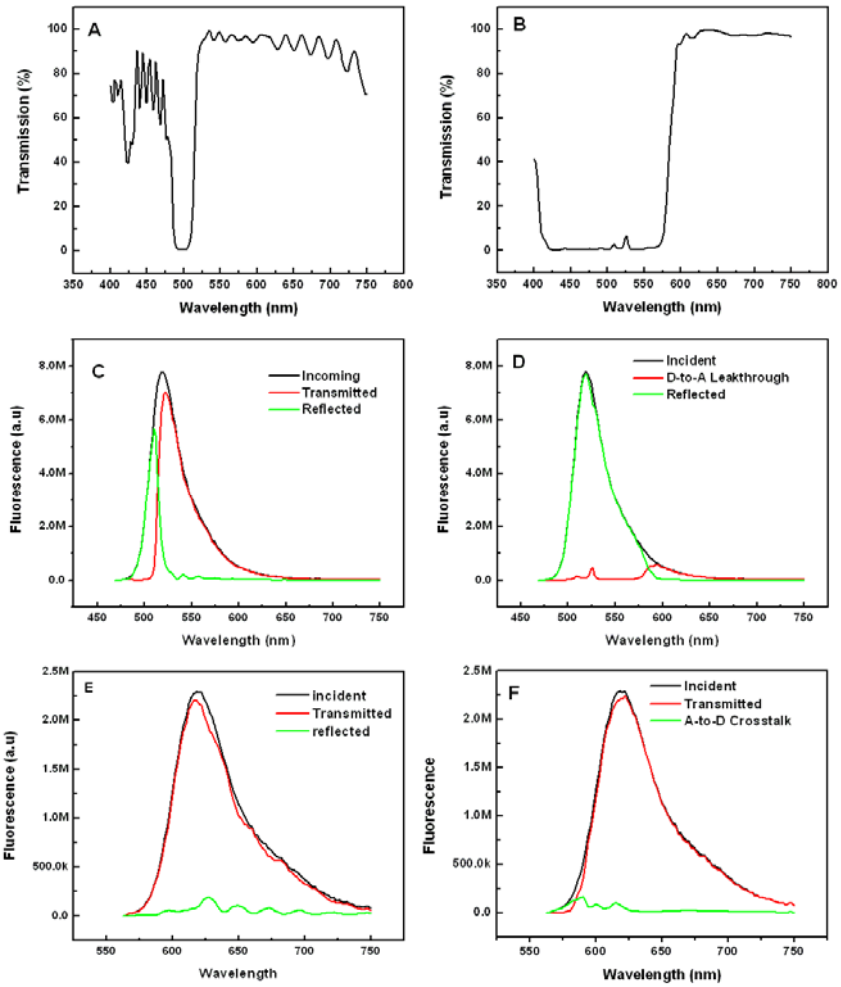

Figure 2.5 Transmittance spectra of dichroic mirrors and the transmission / reflection spectra of donor and acceptor beams. (A) and (B), transmittance spectra of 515 and 585 nm dichroic mirrors. (C) and (E), the original (black), transmission (red) and reflection (green) spectra of donor (C) and acceptor (E) beams after passing through 515 nm dichroic mirror. (D) and (F), the original (black), transmission (red) and reflection (green) spectra of donor (D) and acceptor (F) beams after passing through 585 nm dichroic mirror.

2.6.2 Detector Efficiency

The APD detector has a slightly better response at longer wavelengths. This results in the unequal detector efficiencies of acceptor and donor channels. Because it

is difficult to measure the detector efficiencies Φ_D, Φ_A directly, we measured the ratio $\Gamma_{dtector} = \frac{\Phi_A}{\Phi_D}$ by employing the fluorescent protein R-phycoerythrin (R-PE).

Present in cyanobacteria and some algae, the phycobiliprotein is made of complex between protein and covalently bound phycobilin. It absorbs light and then transfers energy to Chlorophyll. Phycobilin proteins include phycoerythrin, allophycocyanin and other species which have different absorption and emission properties. RP-E is one type of phycoerythrin which is called R-phycoerythrin.

RP-E is a very good fluorophore to calibrate the detector efficiency ratio for our system. The extremely high absorption at 488 nm (~ $10^6 cm^{-1}M^{-1}$) and a large QY (~ 0.82) results in a strong fluorescence emission. Because of its size its diffusion is slow and the residence time in the probe volume is much longer than other small fluorophores, which also benefits the burst intensity. The most important thing is its very large stokes shift. With excitation at 488 nm, the emission peak is around 580 nm which is so close to the cutting edge of the 585 nm dichroic mirror (See Figure 2.6). So the emission contributes to two channels pretty equally and thus RP-E mimics a donor-acceptor FRET system very well.

The fluorescence measurement of PR-E was carried out by using the spectrofluorimeter with sample concentration 1 nM. The setup was excitation / emission slits 2 nm, integration time 0.5 s and excitation wavelength 488 nm.

Based on RP-E emission spectrum and the transmittance spectrum of 585 nm dichroic mirror we expect the equivalent ET should be 0.46 ± 0.01. From the SM-FRET measurement we obtained the ET histogram centered at 0.49±0.01. Since the emission is far from 515 nm, the 515nm dichroic mirror does not affect the SM-FRET

efficiency. Also there is no effect of QY. The discrepancy between the measured and expected FRET efficiencies is uniquely due to the inequality of detector efficiencies. Thus we obtained $\Gamma_{dtector} = \frac{\Phi_A}{\Phi_D} = 1.08$.

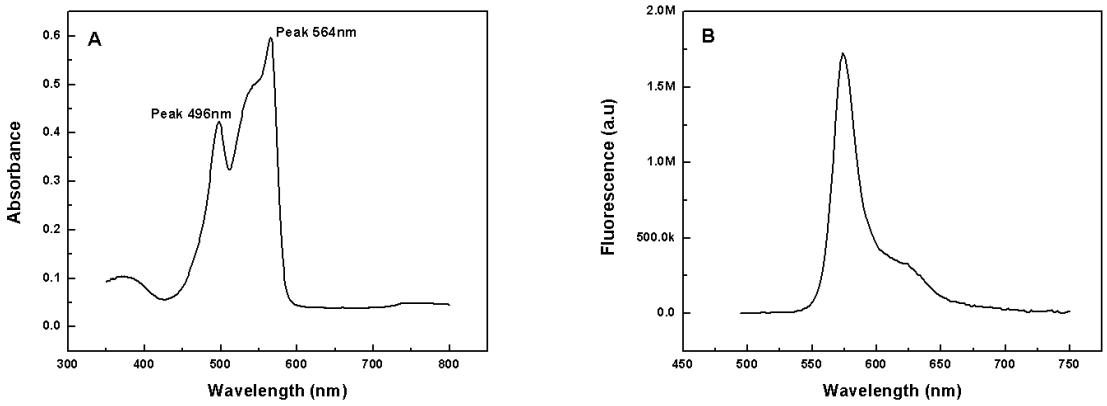

Figure 2.6 Absorption (A) and emission spectra (B) of RP-E. The absorption spectrum has two maximums at 496 nm and 564 nm, respectively.

2.7 Sample Temperature Control System

In order to slow down the dynamics of BBL we did the SM-FRET measurement at 279 K by using a sample temperature control system. In the microscope there is a sample plate where a nano-positioning stage is usually placed to position the sample holder. To control the sample temperature we designed a cold plate to replace that nano-positioning stage. Whenever we performed a single molecule measurement at 279 K, we placed the cold plate on the sample plate and then put sample holder on the cold plate.

Shown in Figure 2.7, the dimension of cold plate is 3.25 x 6 x 1 inches. There is a through hole with a diameter 0.75 inch at the center to enable the movement of the objective. On the top of the cold plate two heat sinks are used to cool down the

hot sides of the peltier (TE-71-1.4-1.5P, TE Technology) which is used to cool down the cold plate. The peltier is between the cold plate and the heat sink. Cold tap water flows through the copper tube inside the heat sink to take away heat generated by the peltier.

The peltier is controlled by an automatic controller (TC-24-12, TE Technology). It monitors the cold plate temperature by obtaining the voltage cross the thermistor (MP-2379, TE Technology) that is inserted inside a small hole of the cold plate.

The glass sample holder is 0.5 inch thick and has a through hole with a diameter of 0.5 inch in the cener. Every time before the experiment, we glued a standard cover slip on the bottom of the hole and made a chamber with volume around 0.8 ml.

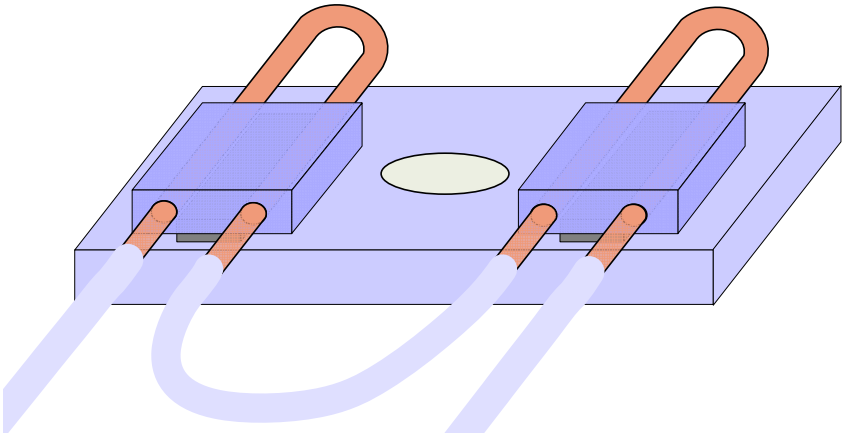

Figure 2.7 Diagram of the cold plate. The cold plate (light blue) has a through hole in the center. Between the cold plate and heat sink (dark blue) there is peltier (black). Tap water flows in the copper tubes (red) which are connected with plastic tubes (grey) to take away the heat generated by peltiers.

2.8 Two-State Analysis of Equilibrium Chemical Denaturation

In the two-state model, the chemical denaturation of protein is regarded as the transition between native and unfolded states. The equilibrium constant $K_{eq} = [U]/[N]$ is determined by the free energy of the unfolded state (native state is the reference) ΔG:

$$K_{eq} = \exp\left(-\frac{\Delta G}{RT}\right) \tag{2.31}$$

Then we can obtain the probabilities of unfolded and native states:

$$P_u = \frac{[U]}{[N]+[U]} = \frac{\exp\left(-\frac{\Delta G}{RT}\right)}{1+\exp\left(-\frac{\Delta G}{RT}\right)} \tag{2.32A}$$

and

$$P_n = 1 - P_u = \frac{1}{1+\exp\left(-\frac{\Delta G}{RT}\right)} \tag{2.25B}$$

The free energy ΔG is usually assumed to be a linear function of the denaturant concentration $[D]$:

$$\Delta G = \Delta G_{H_2O} - m[D] \tag{2.33}$$

where ΔG_{H_2O} is the unfolded state free energy in water, m is the slope.

During the chemical denaturation, the signal S of any ensemble measurement is the weighted mean of signals from native and unfolded state proteins:

$$S = P_n S_n + P_u S_u \tag{2.34}$$

The signals of native and unfolded states which are also referred as baselines are usually assumed to be free-floating linear lines vs. [D]. So each baseline is decided by

two parameters. Including ΔG_{H_2O} and m, six parameters are necessary to characterize the equilibrium chemical denaturation using a two-state model.

The fitting was made by a Matlab (MathWorks) program. The program iteratively changes the fitting parameters of the above two-state model, calculates the observed output data and then compares it with the input data. When the difference between the output data and input data falls below a certain criteria, the iteration stops and we obtain the final fitting parameters of the two-state analysis.

2.9 Protein Purification

2.9.1 Expression and Purification of BBL

The design of BBL sequence is discussed in Chapter 4. The cloned gene was inserted in the plasmid vector (pBAT-4) and expressed in competent cell strain (B121 DE3 gold, Stratagene). We first incubated approximately 10 ml LB solution with transformed cells and 100 μg/ml Ampicilin (Fisher Biotech) at 37 ºC overnight. Next morning, We added the 10 ml culture solution into 4 L LB solution which had 100 μg/ml Ampicilin and left it at 37 ºC to grow until the optical density at 600 nm increased to 1.2~1.4. Then We added isopropyl-beta-D-thiogalactopyranoside (IPTG, Invitrogen) to induce the gene expression and changed the incubation temperature to 298 K. After overnight incubation, the cell was harvested by centrifuging the culture at 8000 rpm for 30 minutes.

The collected cells were dissolved in pH 7.0 20 mM Phosphate buffer with phenylmethylsulphonyl fluoride (PMSF, Invitrogen) and 2mM tris(2-carboxyethyl) phosphine hydrochloride (TCEP, Soltec Ventures, Inc.). Then the solution was

sonicated using the ultrasonication instrument (240/450 Sonifier, Branson Ultrasonics) to disrupt the cell. After sonication, the crude extract was ultracentrifuged at 25000 rpm for one hour. Then we added ammonium sulfate to 50% of its solubility in the supernatant. The addition was very slow and with sufficient stirring of the solution to avoid its immediate accumulation of ammonium sulfate. After overnight precipitation the solution was ultracentrifuged again and many contaminant proteins were separated from the solution which was ready to pass through HPLC columns.

We purified BBL using size-exclusion column (HiLoad 26/60 Superdex 75 prep grade, Amersham Biosciences) first and then the C-18 reverse phase column (Higgins Analytical, Inc.). For size-exclusion column, the buffer was 20 mM pH 7.0 Phosphate with 0.15 M NaCl and 2 mM TCEP. The purification fractions were checked by gel electrophoresis and those pure fractions were pooled together and concentrated by centrifuge filter (Milipore) to reduce the volume. For the reverse phase column, the buffer A was 5% Acetonitrile (ACN) and 0.1% Trifluoroacetic acid (TFA); buffer B was 95% ACN and 0.1% TFA. The gradient started from 20% B and ended at 50% B after 60 minutes. During the purification, the solution was monitored by taking the absorbance at 205 nm. By running the gel electrophoresis, the pure fractions were pooled and lyophilized. The obtained protein power was kept in - 80 ºC for conservation.

2.9.2 Expression and Purification of *α*-spectrin SH3

α-spectrin SH3 was purified almost in the same way as BBL. The small modification was made in the ammonia sulfate cut. For *α*-spectrin SH3, we first

added ammonia sulfate to 25% of the solubility in order to precipitate some large proteins. Then we added ammonia sulfate to 75% of the solubility in order to precipitate SH3. After centrifuging, we got the SH3 in the precipitant and dissolved it again with Phosphate buffer. Other procedures were similar to BBL purification.

2.10 Fluorophore Labeling Reaction and Purification

2.10.1 Fluorophore Labels

Maleimide Alexa 488 / 594 (Invitrogen Molecular Probes) were used as donor and acceptor in our project, respectively. The maleimide probe is very thiol-selective. It does not react with histidine or methionine and only reacts with cysteine, which makes the labeling site-specific. Also, the generated covalent bond is strong making the labeled product very stable.

Alexa 488 / 594 have high extinction coefficients and quantum yields which are very critical for single molecule measurement. Their pertinent features are shown in Table 2.3. The absorption and fluorescence spectra of Alexa 488 and 594 are shown in Figure 2.8. It shows that the emission spectrum of Alexa 488 and absorption spectrum of Alexa 594 have significant overlap which results in strong FRET between two fluorophores.

Table 2.3 Basic features of donor and acceptor

	Molecular Weight	Absorption Peak (nm)	Extinction coefficient ($M^{-1}cm^{-1}$)	Emission Peak (nm)	Buffer	Lifetime (ns)
Alexa 488	720.66	493	72,000	516	pH 7	4.1
Alexa 594	908.97	588	96,000	612	pH 7	3.9

Note: the lifetime is measured at pH 7.4 buffer and 20 ^{o}C. (from http://probes.invitrogen.com/handbook)

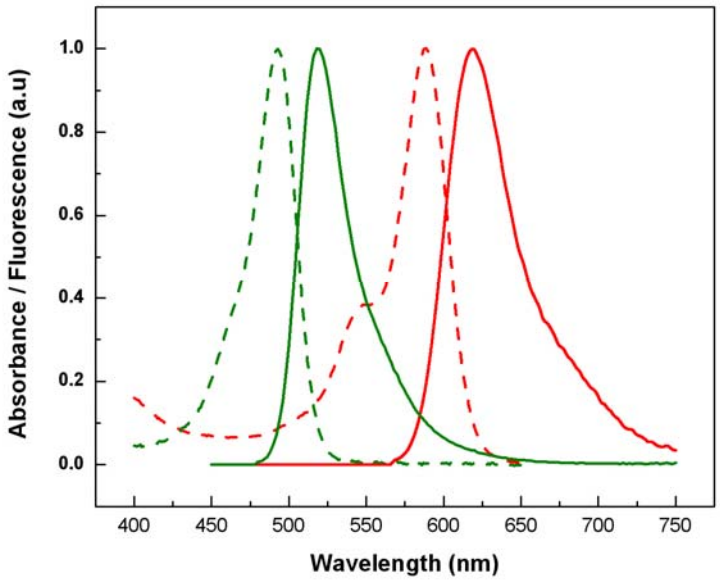

Figure 2.8 Absorbance (dash line) and emission spectra (solid line) of Alexa 488 (olive) and 594 (red). The donor emission spectrum (solid olive) has significant overlap with the acceptor absorption spectrum (dash red) which generates the FRET.

2.10.2 Labeling Reaction and Purification

For most samples of BBL and α-spectrin SH3 we first labeled the protein with donor and purified it using reverse phase column. Then the selected one-donor protein was reacted with acceptor and purified again. Only the labeled BBL sample that was used for the GuHCl denaturation measurement was labeled with acceptor first.

All the labeling reactions were made in 3 ml of pH 7.0 20 mM Phosphate buffer. 2 mM TCEP was always used in order to avoid protein aggregation. In BBL acceptor labeling and all α-spectrin SH3 labeling reactions, the reaction mixture had 6 M urea in order to speed up reactions. The free label and lyophilized unlabeled or

single-labeled proteins were dissolved separately. Then we added drops of the label solution into the protein solution slowly with sufficient mixing. The reaction mixture was left in 4 ºC for 18~24 hours and later the reaction was stopped by adding 1 µl β-Mercaptoethanol (BME).

According to our experience, the optimized reactant ratio of unlabeled protein and donor was 8~10 mg protein vs. 1mg label which gives good yield of one-donor protein. However, the reaction progress can be different even with exactly same reactant ratios. So sometimes the reactant ratio was slightly adjusted based on the yield of reactions we had obtained before. For the second labeling reaction, the purified one-donor protein was only about 2~3 mg, but we stilled used 1 mg acceptor which was in excess. For that BBL labeling reaction which was labeled acceptor first and then donor, the similar protocol was used.

The purification was made by the same reverse phase column and conditions as we used to purify the unlabeled protein. Then purified fractions were checked using MALDI mass spectrometer (AXIMA-CFR, SHIMADU, Japan). In Figure 2.9 there are some typical mass spectra of doubly-labeled BBL. We can see that one-donor, two-donor and doubly-labeled proteins were separated pretty well from each other. Finally, those fractions of doubly-labeled protein were pooled and filtered with centrifugal filter (Centricon, YM-3, Millipore) to remove free label. The filtered protein solution was then stored in -80 ºC.

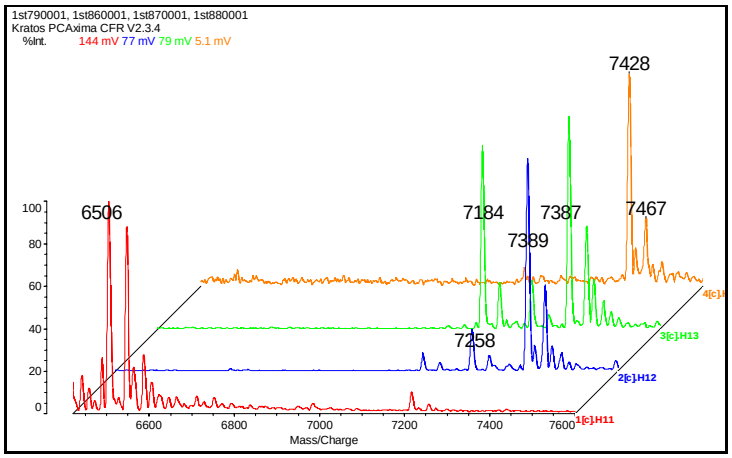

Figure 2.9 Mass spectra of the purification fractions of BBL labeling reaction. One-donor, two-donor and doubly-labeled BBL proteins have the molecular weight of 6511, 7209 and 7397 Da, respectively. Because of the error of calibration and measurement, the measured peak values slightly deviate from the expected values. We can see the separation of different species was good. Fraction 88 (orange) was very pure and only had the doubly-labeled BBL**.**

2.10.3 Preparation of Two-Donor and Two-Acceptor Protein Samples

We also prepared some two-donor or two-acceptor proteins (both BBL and α-spectrin SH3) to carry out the QY measurement. In such labeling reactions, 2 mg protein and 0.5 mg label were mixed in 3 ml solution which was same to the reaction solution of acceptor labeling described in Section 2.7.2. The purification and filtration were also same as the preparation of doubly-labeled proteins.

2.11 Chemicals and Buffers

The fluorescent protein R-PE was purchased from Molecular probes (Cat. No. P801). QY standard R6G, Cresyl Violet and fluorescein were obtained from Sigma. The sample with glucose catalyse/oxydase oxygen scavenger system was made of 1.2% w/w glucose, 2170 U/ml catalyse (Roche), 165 U/ml oxidase (Sigma-Aldrich).

In the Oxidase stock, 2mM $MgCl_2$. was present. We used the combination of Trolox and cysteamine to protect fluorophore from photobleaching and photoblinking. The chira select compound (S)-Trolox methyl ether (Trolox, Sigma-Aldrich) is a soluble derivative of vitamin E. Cysteamine has the formula $HSCH_2CH_2NH_2$ (Sigma-Aldrich).

The measurements of BBL were carried out in pH 6.0 20 mM acetate buffer. NaCl was added to adjust the ionic strength to 36 mM. It is very similar to that of pH 7.0 20 mM phosphate buffer (43 mM) used in the original experiments of Naf-BBL thus excluding the possible effect of ionic strength on protein stability. The measurements of α-spectrin SH3 were performed in pH 3.5 20 mM citric buffer or pH 5.0 20 mM acetate buffer.

Chapter 3 Enhancement of SM-FRET Temporal Resolution Using Effective Fluorophore Protector

3.1 Introduction

As introduced in Chapter 1, we intend to measure the SM-FRET efficiency histogram of BBL upon denaturation. However, to detect the true conformational distribution, the acquisition time of SM-FRET measurements needs to be short enough. If the acquisition time is longer than the dynamic relaxation time of protein, the measured FRET efficiency of this process will become an average of different conformations. In this case, SM-FRET will not be able to provide the distribution of protein conformations. As indicated in the simulation performed by Szabo [54], a 2-state protein will show a unimodal distribution in the SM-FRET efficiency histogram given a long enough acquisition time.

In freely diffusing SM-FRET measurements of protein folding, researchers usually choose a laser excitation power of approximately 40 μW and acquisition time of the order of several hundred microseconds [33, 37, 44]. Using these parameters it is then possible to collect enough photon bursts above thresholds of 25 counts or higher to build up a SM-FRET histogram in reasonable measurement duration. However, the relaxation time of BBL is 20 μs and 120 μs at room temperature and 279 K, respectively. With the typical excitation power and acquisition time SM-FRET is therefore no longer effective to detect the dynamics of BBL.

The first approach to solve this problem is to shorten the acquisition time in order to obtain reliable dynamic information. However, this decrease reduces the amount of photons detected per burst. Therefore the SM-FRET statistics will be of poor quality because of the stochastic process of photon emission. Increasing the excitation power can increase photon emission, but too much excitation power will damage the fluorophore very easily. Also, when the inter-photon time is reduced, the triplet state or other dark states may interfere with the photon emission process. The solution is to use proper agents to protect the fluorophore from photobleaching and photoblinking processes and thus increase the excitation power in order to reduce the acquisition time.

In photobleaching and photoblinking processes, the oxygen in solution plays a complicated but very important role. As a very effective triplet quencher, triplet oxygen can suppress the triplet state of the fluorophore. However, the singlet oxygen and other oxygen radicals in solution generated by radiation can affect the emission by damaging the fluorophore. The proper way to conserve the signal is to remove singlet oxygen and oxygen radicals leaving normal triplet oxygen in solution.

In surface immobilized single molecule fluorescence measurements, several groups have used enzymatic oxygen scavenger systems to protect the fluorophores from photobleaching and photoblinking [55-57], which have some effect on the oxygen in solution. For example, Ha and collaborators use a combination of enzymatic oxygen scavenger system and Trolox, a derivative of vitamin E with antioxidant properties, which works very efficiently to eliminate the photoblinking of surface immobilized Cys3 and Cys5 [55].

Nevertheless, there are some fundamental drawbacks to implement these methods using enzymatic oxygen scavenger into freely diffusing SM-FRET measurement of protein folding. Firstly, the glucose oxidase/catalase used in the enzymatic oxygen scavenger system usually makes the solution cloudy that significantly decreases the detected signal. Secondly, in protein folding experiments, it is very common to use chemical denaturants such as urea and GuHCl in the solution to unfold the protein. Obviously such chemical denaturants affect the activity of the enzymes to a large extent. And finally, there is always the chance of enzymes interacting with the protein of interest. So finding some non-enzymatic agent to prevent photobleaching and photoblinking will be very advantageous in SM-FRET study of protein folding.

Enlightened by previous studies [55, 57, 58], we have found that the combination of Trolox and cysteamine can protect the fluorophore very efficiently and increase the signal several folds. This particular mixture has not been used before for these purposes and works better than each compound alone.

3.2 Experimental Methods

The single molecule measurement was carried out in the old scheme confocal microscope system whose transform lens setup was not optimized and thus the photon collection was lower than the new scheme which I described in Chapter 2. So the binning time was set as 100 μs. The sample concentration was 75 pM except that the experiment showed in Figure 1 had the concentration of 500 pM. A threshold of 25 photon count was used to select bursts.

For the ensemble fluorescence measurement, sample concentration was about 0.7 μM. The spectrofluoremeter was set as: excitation / emission slits 1nm, integration time 0.25 μs. The excitation wavelengths were 495 nm and 590 nm for donor and acceptor, respectively.

3.3 Results

3.3.1 Trolox Alone is Better than Glucose Oxidase/Catalase Oxygen Scavenger plus Trolox

When we increase the excitation power without any protection of the fluorophore, the effect of photobleaching and photoblinking become serious producing a larger population of molecules with donor-only signal (figure 3.1A). The use of glucose oxidase/catalase oxygen scavenger system plus Trolox reduces the percentage of molecules in the acceptor-inactive peak (known as zero-peak), but it also decreases the signal many fold (Figure 3.1B). Trolox along gives much more bursts than the use of enzymatic oxygen scavenger system plus Trolox while the percentage of zero-peak is not very high (Figure 3.1C). Although Trolox alone produces more bursts than the combination of enzymatic Oxygen scavenger and Trolox, the number of bursts using Trolox alone is still not large enough provided that the protein concentration in this measurement was as high as 500 pM and the typical protein concentration of SM-FRET measurement is 75 pM. The excitation power was 180 μW.

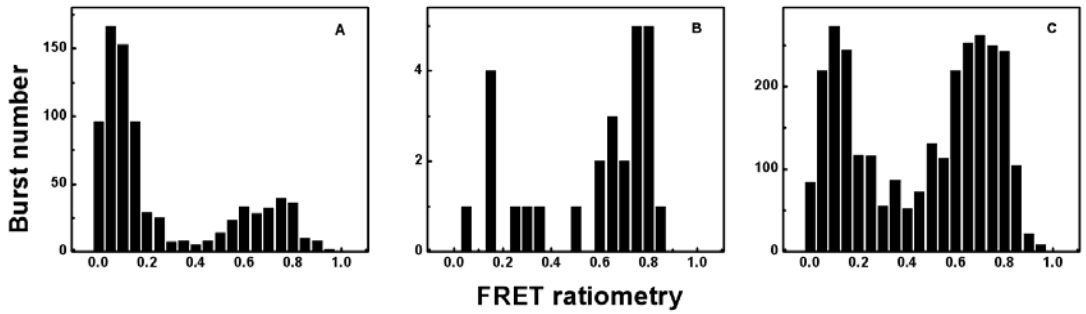

Figure 3.1. Comparison of SM-FRET efficiency histograms obtained with a sample being protected by: oxygen scavenger plus Trolox (b), Trolox (c) or either of them (a). The excitation power was 180 μW and the protein sample concentration was 500 pM.

3.3.2 Combination of Trolox-Cysteamine Works Efficiently

As shown in Figure 3.1, the only use of Trolox drastically increases the amount of bursts above threshold of 25. However, the protein concentration used in this experiment was 500 pM. When it is reduced to 75 pM, the typical concentration used in freely diffusing protein SM-FRET experiments, the number of bursts drops and hence it is not possible to build a reliable histogram (Figure 3.2A). To solve this problem we have found that the addition of cysteamine in solution increases the signal dramatically (Figure 3.2B). Under the same excitation power and threshold to select burst, the use of 10 mM cysteamine and 1 mM Trolox produces almost ten fold bursts more than the Trolox alone.

While the solubility of Trolox is ~2 mM and thus 1 mM Trolox was used in our measurements, the concentration of cysteamine is not limited by the solubility. To investigate the effect of cysteamine on SM-FRET signal systematically we have done the measurement with different cysteamine concentrations and excitation powers. For a certain excitation power, the number of bursts was maximal at 10 mM cysteamine. When we increased the cysteamine concentration beyond 10 mM the number of

bursts dropped. At 50 mM cysteamine even the singlet fluorophore was quenched by cysteamine and there were almost no bursts above the threshold of 25 (figure 3.3A). With respect to the percentage of zero-peak, it decreased monotonically with the concentration of cysteamine (figure 3.3B). This is because the higher the concentration of cysteamine, the better the photobleaching and photoblinking are suppressed. However, the price of minimizing photobleaching and photoblinking (represented by the zero-peak) is the quenching of the fluorophore singlet state. Since the zero-peak is not critical if it is not too high, we suggest that the optimal concentration of cysteamine is 10 mM. Of course, changes in fluorophore, protein, buffer or other experimental conditions may change this optimal concentration.

With 1 mM Trolox and 5~25 mM cysteamine, the number of bursts increased linearly with the excitation power. It did not show any saturation until 210 μW (Figure 3.3A). We actually did some measurements with an excitation power 240 μW under the protection of Trolox and cysteamine and it still produced a reasonable SM-FRET histogram. However, the protein peak of the histogram shifted marginally to smaller ET, indicating that the Trolox-cysteamine system started to be ineffective (see Chapter 4).

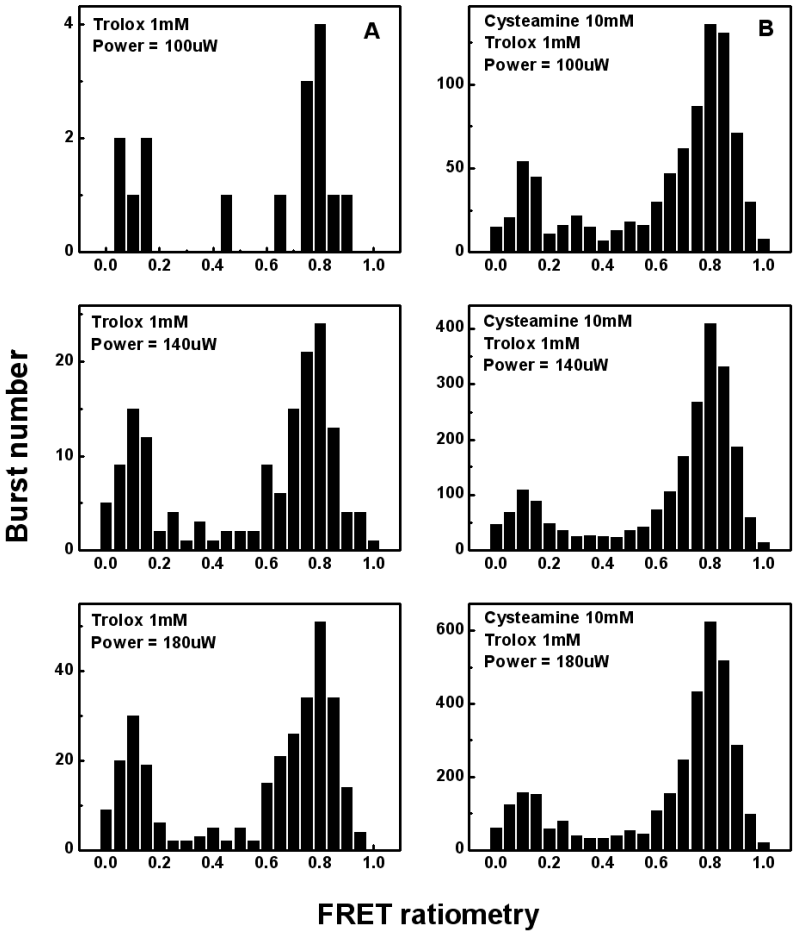

Figure 3.2. Histograms obtained with 1 mM Trolox (A) and 10 mM Cysteamine / 1mM Trolox samples (B) at varying excitation powers. The protein concentration was 75 pM.

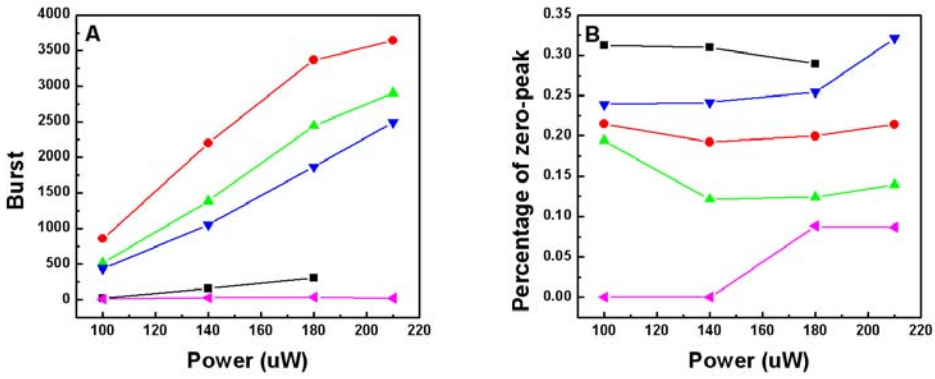

Figure 3.3. Bursts (A) and zero-peak percentages (B) of samples of 1 mM Trolox without cysteamine (black), with 5 mM (blue), 10 mM (red), 15 mM (green) and 50 mM cysteamine (pink). We can see the best condition is 10mM cysteamine and 1mM Trolox.

Cysteamine Alone is not as Good as Trolox-Cysteamine. Since the combination of Trolox and cysteamine works so efficiently, and Trolox alone is not enough, the next step is to test the effect of cysteamine alone. Maybe just the optimal concentration is different. However, our results show that cysteamine alone cannot reproduce the effect of the Trolox-cysteamine combination. With or without Trolox, the optimal concentration of cysteamine is still approximately 10 mM, but cysteamine combined with Trolox can increase bursts two fold as shown in Figure 3.4A where the excitation power was 140 μW. On the other hand, the combination of cysteamine and Trolox seems to reduce the percentage of zero-peak more efficiently than cysteamine alone, although experimental noise makes it difficult to quantify the effect (Figure 3.4B).

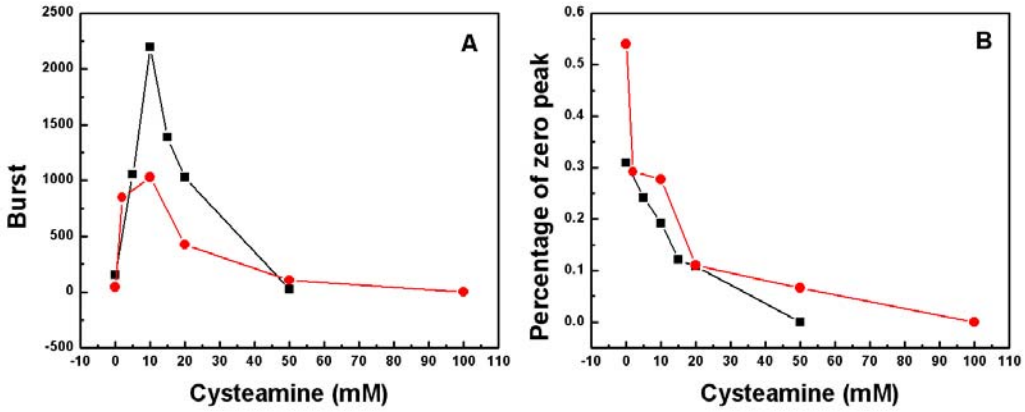

Figure 3.4. Effect on the burst number (A) and zero-peak percentage (B) of cysteamine in samples with (black square) and without Trolox (red circle). The protein concentration was 75 pM and excitation power was 140 μW.

3.3.3 Quenching Effect of Cysteamine on Fluorophore Singlet

From Figure 3.3A, it is clear that cysteamine has a quenching effect on the fluorophere singlet. To quantify this effect, we measured the ensemble fluorescence of free fluorophores and labeled proteins with varying concentrations of cysteamine.

Compared to the power intensity of the focused excitation beam in the probe volume of SM-FRET measurement, the excitation power intensity of ensemble measurement is very low. Thus photobleaching and photoblinking occur at a much lower frequency than it in the single molecule measurement. Therefore, the main contribution to the signal change is the singlet quenching produced by cysteamine.

The ensemble fluorescence signal of labeled proteins (donor-labeled and acceptor-labeled BBL) tend to decrease remarkably with increasing cysteamine concentration (Figure 3.5B, D). In contrast, for the free fluorophores (both Alexa 488 and Alexa 594), there is an increase of fluorescence intensity between 0 ~ 10 mM cysteamine. Also, the maximum fluorescence intensity of the free fluorophore with 10 mM cysteamine is very close to the maximium value of the labeled protein without cysteamine.

To summarize, it seems that a small amount of cysteamine can protect the singlet of free fluorophore from being quenched by singlet oxygen or other species while the labeled protein avoids such effects by some protein-specific processes and presence of cysteamine cannot improve it any more.

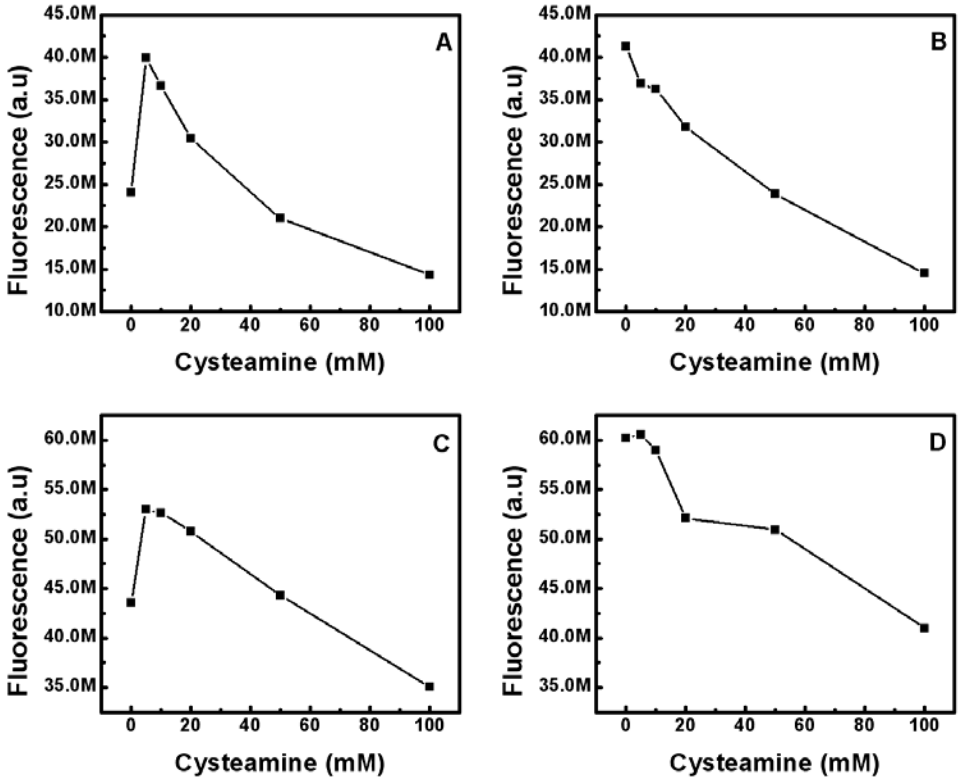

Figure 3.5 The ensemble fluorescence intensity of samples of free donor (A), donor-labeled BBL (B), free acceptor (C) and acceptor-labeled BBL (D) with varying concentrations of cysteamine. The fluorescence signal was normalized by the maximum absorbance of respective samples to avoid any error caused by concentration measurement.

3.4 Conclusions

Both Trolox and cysteamine have been used previously in SM-FRET experiments with enzymatic oxygen scavenger system but not together [55, 57, 58]. We demonstrated that without the enzymatic oxygen scavenger system the combination of both compounds works more efficiently. Also the combination of Trolox and cysteamine works better than each one alone. According to our results, both compounds affect to the percentage of zero-peak, related to the photobleaching and photoblinking of acceptor, and to the total amount of bursts, related with the

photobleaching and photoblinking of donor. The solubility limit for Trolox is 2 mM and although it is a powerful protector, Trolox is not sufficient to avoid photobleaching and photoblinking with the high excitation power that has to be used in order to collect enough photons within short acquisition times. Cysteamine is very soluble so the only limitation for the concentration is the quenching effect on fluorophore emission. Combining both, the protection is effective with excitation power of 160 μW or even higher. The optimal concentrations for our system are 1 mM Trolox and 10 mM cysteamine, but the optimal concentration may depend on the fluorophore, protein and experimental conditions.

The possibility to use SM-FRET method for very fast folding processes appears with the use of efficient fluorophore protectors. Here the binning time is 100 μs due to the inefficient design of the transform lens. With the optimized transform lens setup that is described in Chapter 2 and the fluorophore protector Trolox-cysteamine system, we measured reliable SM-FRET efficiency histograms of BBL and α-spectrin SH3 under 50 μs binning time. The results will be discussed in Chapter 4 and 5, respectively.

Chapter 4 Ensemble and Single Molecule FRET Measurements on BBL

4.1 Introduction

The most fundamental characteristics of downhill folding are the barrierless free energy surface and the corresponding unimodal conformational distributions. However, with traditional ensemble equilibrium or kinetics methods neither of them is accessible. Because of the ability to detect FRET efficiency (ET) of single molecules, SM-FRET is able to measure the conformational distribution of protein ensembles. We carried out the SM-FRET measurements on BBL under varying conditions to obtain the conformational distribution as discussed below.

As the good reference of single molecule measurement, we performed some ensemble measurements. The two-state analysis of ensemble FRET curves provided us the FRET baselines of native and unfolded states which were used in the single molecule simulation. We compared the denaturation profiles of BBL by urea and GuHCl and found GuHCl to be 2.75 times stronger than urea.

Because of the higher aggregation propensity of doubly-labeled BBL in typical far-UV CD concentrations (30-50 μM) we compared the ensemble FRET unfolding profiles of labeled proteins with the far-UV CD profile of unlabeled proteins. The good superposition and the similar thermal parameters obtained from the two-state analysis of FRET and CD curves indicate that labeling has no effect on the protein stability.

All ensemble and single molecule FRET measurements were carried out at 279 K in 20 mM acetate buffer at pH 6.0 except when noted otherwise. BBL folds

very fast with a relaxation time of ~20 μs at room temperature [46]. However, it slows down drastically to ~120 μs at 279 K enabling the use of a practical acquisition time of 50 μs in single molecule measurements. Trolox-cysteamine system was used in the single molecule measurement to protect fluorophores from photobleaching and photoblinking. Details about the experimental methods are described in Chapter 2.

We also applied Szabo' photon statistics theory about SM-FRET to simulate the expected SM-FRET efficiency histogram by assuming BBL folds in a two-state manner under our experimental conditions. By comparing the simulation and experimental results we provide direct evidence to the downhill folding nature of BBL.

4.2 Structure and Sequence of BBL

4.2.1 Background

BBL is the E3-binding domain of the dihydrolipoamide succinyltransferase core from the 2-oxoglutarate dehydrogenase multienzyme complex of Escherichia coli [59]. The wild-type BBL has 40 residues and its sequence is shown below. The wild type BBL has two parallel α-helices and a 3_{10} helix turn connected by long loops as shown in Figure 4.1 that is from the PDB website.

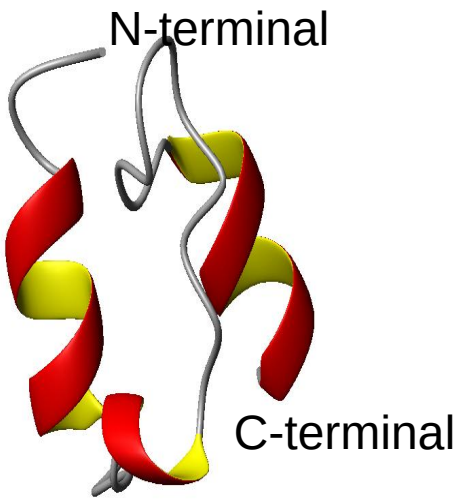

Figure 4.1 Structure of wild-type BBL. It contains two parallel α-helices and a 3_{10} helix turn connected by long loops (from PDB website).

4.2.2 Sequence Design

The variant that was used in our studies has longer tails than the wild-type BBL in both sides:

Wild-type: ALSPAIRRLL AEHNLDASAI KGTGVGGRLT REDVEKHLAK

Variant: MDCKKNND ALSPAIRRLL AEHNLDASAI KGTGVGGRLT REDVEKHLAK APAKKC

QNND-BBL: QNND ALSPAIRRLL AEHNLDASAI KGTGVGGRLT REDVEKHLAK A

In the sequence of variant BBL, the central part (NNDAL..... KAPAK) is extracted from the larger multidomain protein and is identical to the wild-type BBL and QNND-BBL. On both N- and C-termini additional residues MDCKK and KC are engineered into the sequence. The design of the variant sequence is based on several considerations.

As discussed in Chapter 1, one of the claims of Fersht group is that wild-type BBL folds less cooperatively due to shorter protein boundaries [16]. They base this claim, among other reasons, on the lower stability of the wild-type BBL sequence.

They go on to propose that the addition of the short tail QNND to the wild-type BBL enhances the barrier significantly thus making it fold in a two-state manner. The simulation of Wolynes et.al apparently supports this idea. However, earlier work by Naganathan et al. indicates that the tail of QNND-BBL does not affect the free energy barrier significantly [17]. To further clarify this controversy, we designed our new variant BBL to include almost the entire sequence of QNND-BBL except the first residue glutamine.

The residues MDCKK and KC were designed to speed up the labeling reaction since the positively charged lysine attracts the negatively charged fluorophores. The two cysteines are the anchoring groups for the fluorescent label.

The long tails also introduce an additional advantage. FRET efficiency changes most effectively with respect to variation of the end-to-end distance when it is near the Förster distance (R_0). To get a significant change in FRET efficiency during unfolding process, we need a donor-acceptor pair with a Förster distance that is approximately half-way between the end-to-end distances of the native and unfolded proteins. The Alexa488 / Alexa594 pair has a Förster distance of about 5.4 nm [38] that is too large for the wild-type BBL. With the flexible tails, we expect the variant BBL to have a larger dynamic range of end-to-end distances centered at about 5.4 nm upon denaturation. In the experiments we obtained a FRET efficiency of ~ 0.8 for the native state and ~0.4 for the unfolded state thus validating the need for long tails.

4.3 Ensemble FRET Measurement

4.3.1 Ensemble FRET Measurements at Varying pH and Temperatures

In order to obtain a sufficient dynamic FRET efficiency range upon denaturation, we initially planned to use pH 6.0 acetate buffer which is expected to destabilize the protein slightly compared to pH 7.0. As shown in Figure 4.2A, at 298 K the FRET efficiency ratiometry (donor signal divided by sum of two channel signals) in pH 6.0 is lower than in pH 7.0, indicating that pH 6.0 destabilizes BBL compared to pH 7.0. However, the unfolding curves at pH 6.0 and 7.0 almost overlap with each other at 279 K. This means at 279 K the stability of BBL at pH 6.0 matches that of pH 7.0 BBL. The crossover of pH 6.0 and 7.0 curves may be due to quantum yield (QY) difference since the FRET efficiency ratiometry was without any correction.

However, we found pH 6.0 had another advantage: it enhances the fluorescence signal. As shown in Figure 4.2B, for the same concentration of labeled BBL, the sample produces more signals at pH 6.0 than pH 7.0. This enhancement was also confirmed in single molecule measurements. It is probably because of the pH-effect on the fluorophore QY or the protein adsorption to the sample holder.

In addition, the total fluorescence signal of both channels increased dramatically between 0 and 2 M urea while remaining constant at higher concentrations. It is very possibly due to the adsorption of protein molecules on the quartz cuvette. The presence of urea made the solution more hydrophobic and then less protein adhered to the cuvette.

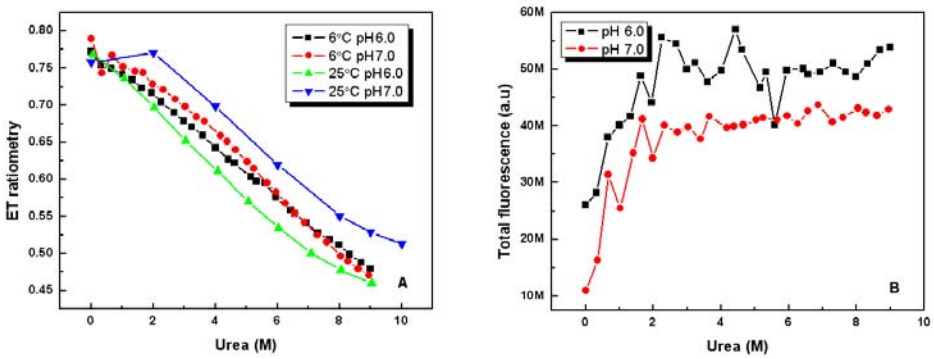

Figure 4.2 Comparison of ensemble FRET measurements: pH 6.0 (black square) and pH 7.0 (red circle) at 279 K, pH 6.0 (green upward triangle) and pH 7.0 (blue downward triangle) at 298 K. (A), the ET ratiometry. (B), the total fluorescence intensity.

4.3.2 Ensemble FRET Measurements of BBL Denaturation

Based on the above comparison of pH conditions, we decided to perform the ensemble and single molecule measurements of BBL in pH 6.0. Also, the protein was unfolded by both urea and GuHCl.

The accurate ET change upon BBL denaturation is shown in Figure 4.3A where the cross talk between the two channels and the acceptor direct excitation were corrected using those equations introduced in Chapter 2. The denaturant effect on R_0 was not corrected for this curve. The unfolding curve is broad, spanning almost 5 M GuHCl with long post-transition and little pre-transition baselines.

In the GuHCl unfolding measurement, the FRET efficiency of BBL sample without GuHCl dropped significantly when the temperature was lowered from 298 K to 279 K. We also found the total fluorescence intensity of acceptor dropped by 36 % and that of donor only dropped by 8 % upon the temperature change. In other words, the amount of acceptor in solution dropped faster than that of donor during cooling.

A possible reason for these effects is the relatively higher hydrophobic nature of the acceptor making it more prone to surface adsorption or aggregation than donor. Thus the adsorption of doubly-labeled proteins to the glass cuvette is more likely than donor-only proteins during cooling. Therefore, the relative amount of doubly-labeled proteins among all proteins dissolved in solution decreases, which in turn causes the decrease in ET during cooling. However, for the urea denaturation measurement, we did not find such large ET change during cooling. This is probably because protein samples purified and labeled apart have different characteristics.

The urea unfolding curve is broader than its GuHCl counterpart with no evident pre- or post-transition baselines (Figure 4.3B). To compare with the urea denaturation profile, I multiplied the original ET of GuHCl unfolding curve by 1.11, which made the ET of 0 M GuHCl sample at 279 K equal to the ET of the same sample at 298 K and the ET of 0 M urea sample in urea denaturation measurement. The modified GuHCl ET profile will be used as the ensemble ET profile of BBL GuHCl denaturation as shown below.

The difference in the apparent broadness of the two unfolding profiles is expected as GuHCl is a stronger denaturant than urea, the strength of which is determined by the *m*-value (see Chapter 2). However, it is challenging to estimate the *m*-value from the urea-unfolding curve as it has little information on the baselines. This problem was overcome by matching the abscissa to superimpose the two curves. The two unfolding profiles exactly superimposed on one another at a scaling factor (defined as m_{GuHCl}/m_{urea}) of 2.75, as shown in Figure 4.3C. This unfolding strength

ratio is very similar to that for SH3 which is 2.85 (see Chapter 5). This value also lies within the range expected for small single-domain proteins.

The GuHCl and urea ET profiles were globally fitted to a two-state model that is described in Chapter 2. The fitting baselines and fitting curve are shown in Figure 4.3C. The baselines are $ET_n = (0.81 \pm 0.03) - (0.03 \pm 0.05)[GuHCl]$ for native state and $ET_u = (0.45 \pm 0.04) - (0.011 \pm 0.006)[GuHCl]$ for unfolded state. The values of ΔG_{H2O} and m are 5 ± 2 kJ·mol^{-1} and 2.7 ± 0.5 kJ·mol^{-1}·M^{-1} (for GuHCl concentration), respectively.

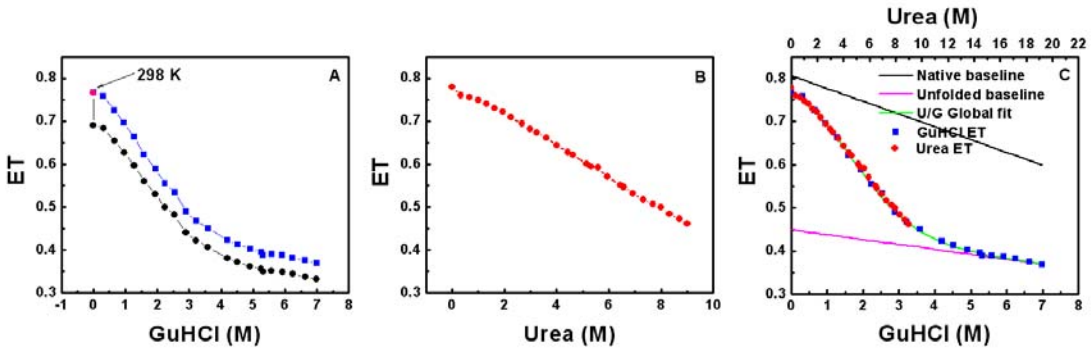

Figure 4.3 Ensemble FRET measurement results of the denaturation of BBL at 279 K. (A), the original ET of GuHCl denaturation (black circle) and the modified curve (blue square). The point of aqueous sample measured at 298 K is marked (pink circle). (B), the ET of urea denaturation (red circle). (C), the superposition of urea and modified GuHCl ET curves. The native baseline (black line), unfolded baseline (magenta line) and fitting curve (green line) from the global two-state fit are plotted as well.

4.3.3 R_0 Corrected Ensemble ET

The Förster distance R_0 usually changes during the denaturation because the donor QY (QY_D) and the refractive index (n) of the solution are dependent on the denaturant concentration. That means the ET could be different though the inter-dye distances do not change. The formula to correct for R_0 change is presented in Chapter

2. The ET curve corrected for R_0 change is plotted in Figure 4.4. As can be seen, the correction has very little effect on urea denaturation profile while it becomes apparent only at high concentrations for the GuHCl unfolding curve.

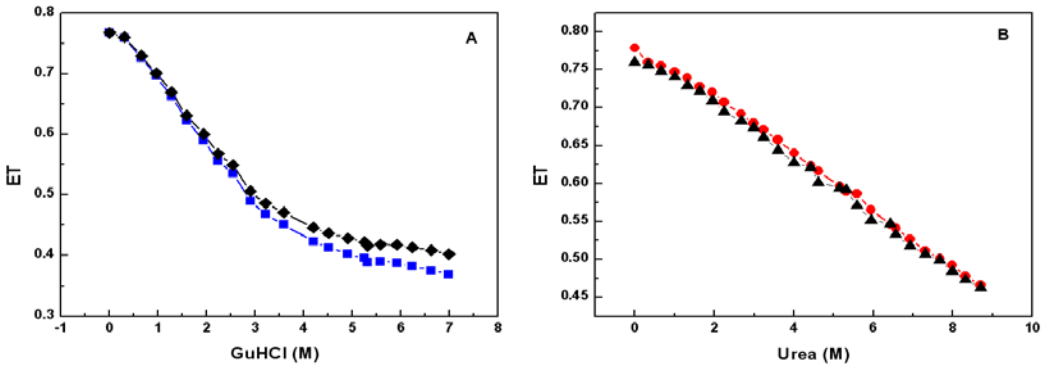

Figure 4.4 ET unfolding curves without and with R_0 correction. (A), the ET (blue square) of BBL GuHCl denaturation is slightly lower than the R_0 corrected ET (black diamond). (B), for urea denaturation of BBL, the ET without (red filled circle) and with R_0 correction (black triangle) are almost identical.

4.4 Comparing Ensemble CD and FRET Measurements: Effect of Fluorophores on Protein Stability

To check for the effect of fluorophores on protein stability we followed the GuHCl denaturaion of unlabeled BBL using far-UV circular dichroim spectroscopy (CD). Because we did not have enough labeled BBL sample, we used the unlabeled BBL to perform the CD measurement. Moreover, the doubly-labeled protein is more prone to aggregation related problems at the far-UV CD concentrations.

Figure 4.5 compares the ellipticity at 222 nm from the CD measurement on the unlabelled protein with the ET of the GuHCl and urea denaturations on the labeled protein. We can see that the CD and ensemble FRET curves almost overlay on one another. The two-state fitting of the CD profile produced a ΔG_{H2O} and m of 5.9 ± 0.7 kJ·mol^{-1} and 3.0 ± 0.2 kJ·mol^{-1}·M^{-1} (for GuHCl concentration), respectively.

They are identical within error to the values we obtained from the global fitting of urea and GuHCl ET curves. The similarity between the unfolding curves and the resulting parameters provides strong evidence that labeling does not affect the behavior of the protein. Moreover, the Alexa 488 / Alexa 594 dye pair has been widely used with no reported effect on the folding behavior [33, 43, 51].

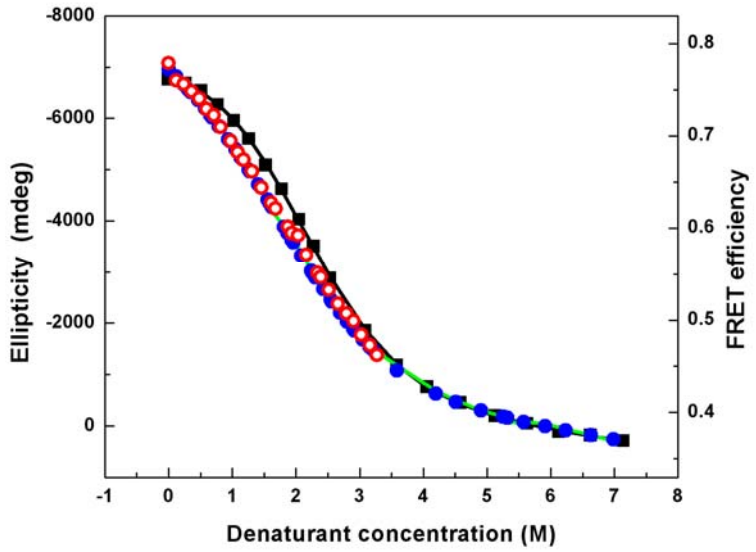

Figure 4.5 Comparison of CD and ensemble ET unfolding profiles of BBL. All of the measurements were conducted at 279 K. The ellipticity at 222 nm (Black square) vs. GuHCl concentration was fitted by the two-state model (black line). The ensemble ET of urea denaturation (red open circle) is plotted together with the ensemble GuHCl ET curve (blue filled circle) where the urea concentration is divided by 2.75. The two-state global fit to the two ensemble ET curves is also plotted (green line).

4.5 SM-FRET Measurements of BBL Denaturation

4.5.1 Urea and GuHCl Denaturation of BBL at pH 6.0

We carried out the SM-FRET measurements of urea and GuHCl chemical unfolding of BBL under the same condition as mentioned before: temperature of 279 K, binning time of 50 μs and excitation power of 160 μW. Thresholds of 40 and 30

were used to select bursts for urea and GuHCl data, respectively. The experiment was carried out at pH 6.0 that gave a higher signal as we discussed previously. Moreover, the aggregation propensity of the protein solution at pH 6.0 was lower compared to pH 7.0.

The measured SM-FRET efficiency histograms are shown in Figure 4.6 without correction for R_0 changes. The zero-peak centered on ET ~ 0 is due to the donor-only protein (also called acceptor-inactive protein). The ET distribution under each condition is fitted by the sum of a Gaussian function (for protein peak) and a lognormal function (for zero-peak) to determine the means and variances of histogram peaks. Figure 4.6 also shows the expected ET values of native and unfolded states under the same concentration of denaturants assuming a two-state like transition (see Section 4.3.2).

From the histogram, we see the clear unimodal ET distribution under all degrees of unfolding stress. The ET peak shifts from the higher value to the lower value gradually when the concentration of chemical denaturant increases. No bimodal distribution is observed in the middle of transition that is a hallmark of two-state behavior. These results provide the clearest evidence as yet to the downhill folding manner in BBL.

At low denaturant concentrations, i.e. 0~3 M urea and 0~1 M GuHCl, the magnitude of the zero-peak is very small but increases with increasing concentrations. This indicates that the Trolox-cysteamine system works less efficiently at high denaturant concentrations. Another possible reason for this observation is that the amount dissolved Oxygen is reduced at higher concentrations of denaturant. Since

Oxygen is a very efficient triplet quencher, the fluorophores in the high denaturant concentration sample have more triplets thus resulting in more inactive acceptors and hence a larger zero-peak. Moreover, the high viscosity of urea and GuHCl solution progressively slows down the diffusion of molecules that can aid in triplet-quenching thereby increasing the magnitude of zero-peak.

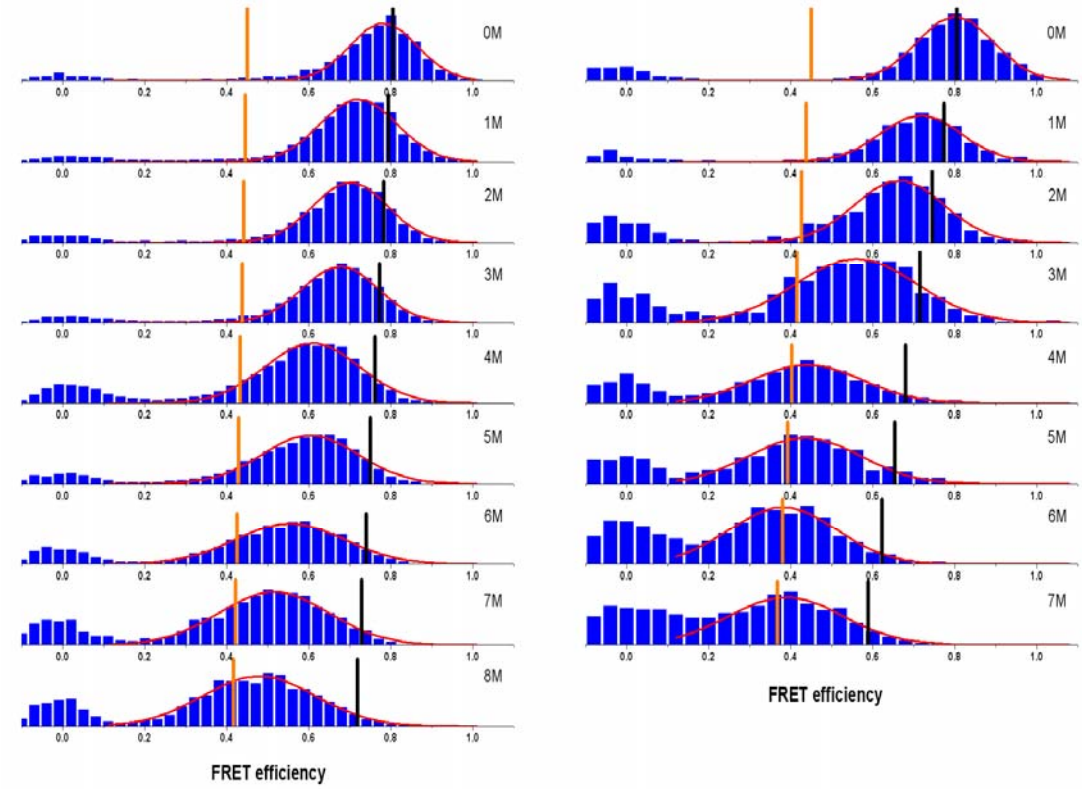

Figure 4.6 SM-FRET efficiency histograms of BBL urea (left column) and GuHCl (right column) chemical denaturations at pH 6.0. The measurement was carried out at 279 K and with a binning time of 50 μs. The histogram was fitted by the sum of a Gaussian function for protein peak (red line), and a lognormal function for zero-peak (not shown). The native (black bar) and unfolded (orange bar) state ET obtained from ensemble ET global fitting are also plotted.

Comparison of SM and Ensemble FRET Measurements. In Figure 4.7, the means of protein peaks obtained from the SM-FRET histogram fitting are plotted along with the ensemble ET measurements. The single molecule ET and ensemble ET are almost identical for the urea denaturation, which validates the ET corrections as

well as indicates there were only small amount of donor-only proteins in this sample. For the GuHCl denaturation, the single molecule ET deviates from the corresponding ensemble ET when GuHCl concentration is 2~3 M. This difference is possibly due to the correction factor of 1.1 that was used to deal with cooling-related changes in ET.

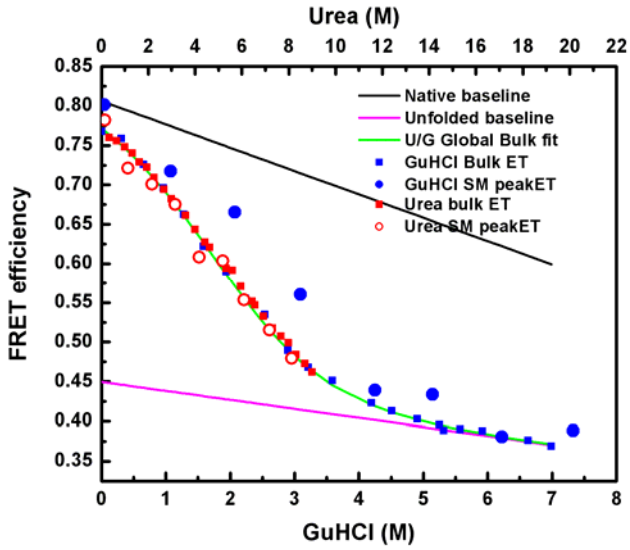

Figure 4.7 Comparison of ensemble and single molecule FRET measurements of BBL unfolding. The ensemble ET of urea (red square) and GuHCl (blue square) unfolding are plotted together with a scale 2.75 between urea and GuHCl concentration. The ensemble urea and GuHCl unfolding profiles are globally fitted to a two-state model. The fitting curve (green line), native (black line) and unfolded (magenta line) baselines are shown. The mean ET from single molecule measurements are plotted for urea (red open circle) and GuHCl (blue filled circle), respectively.

4.5.2 Urea denaturation of BBL at pH 7.0

Although we know that modest changes in pH do not induce any free energy barrier in BBL [19], we still characterized the BBL denaturation with SM-FRET at pH 7.0 in order to exclude the possible effect of pH on the barrier. All the experimental parameters were identical to the pH 6.0 measurements. The informational bursts were selected by a threshold of 40.

As shown in Figure 4.8A, the unimodal distribution of ET is retained at pH 7.0 as expected. The shift in protein peak between 0 and 1 M urea is quite large compared to the ET shift in other measurements. Probably it is because the 1 M sample had poor quality and some aggregation. The zero-peak is not centered on zero for several samples. It is very possibly due to the ET correction — the correction parameters were measured at pH 6.0 and maybe had different values for pH 7.0.

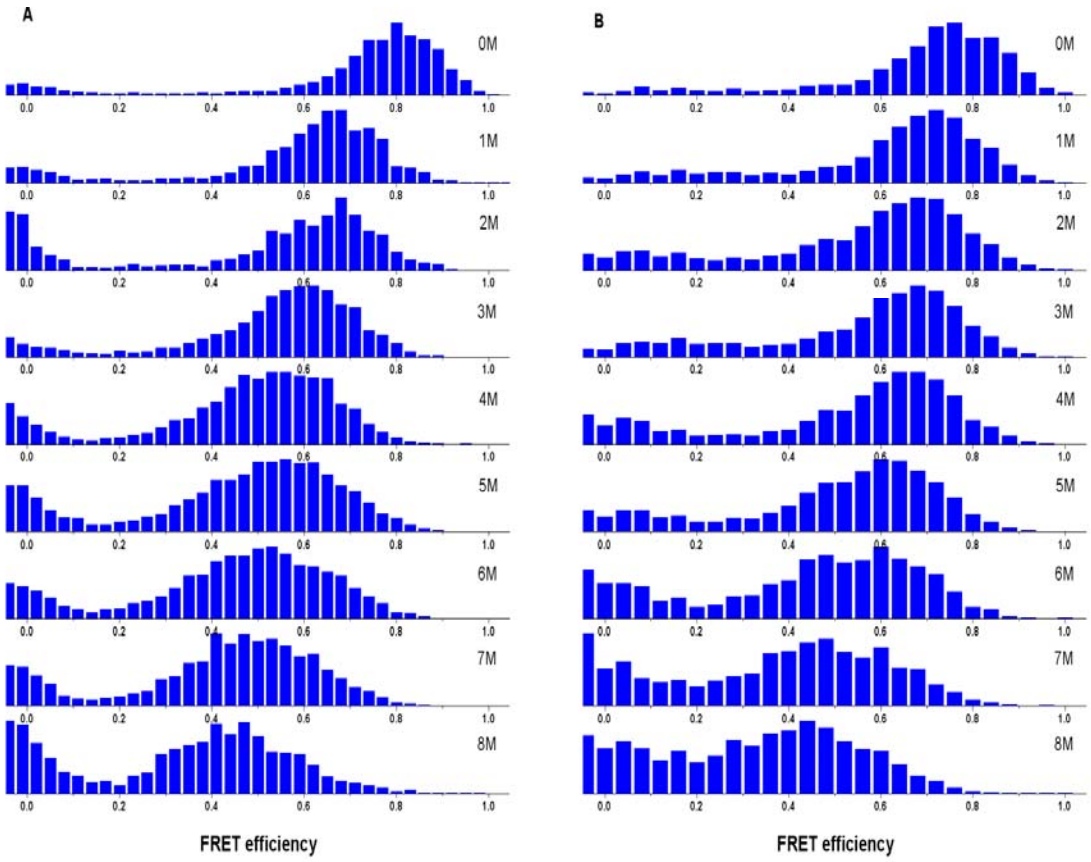

Figure 4.8 SM-FRET histograms of the BBL urea denaturation at 279 K. (A), the measurement was made at pH 7.0 with a binning time of 50 μs. The histogram step size is 0.03. (B), the measurement was made at pH 6.0 with binning time 20 μs. The histogram step size is 0.04.

4.5.3 20 μs SM-FRET Measurement of BBL Urea denaturation at pH 6.0

As discussed before, at 279 K the relaxation time of BBL is 120 μs. All of the experiments above have been performed with an acquisition time of 50 μs that is just

about 2 times smaller than the folding relaxation time. This then raises the question of averaging over conformations. Ideally, the shorter the acquisition time compared to the relaxation time the more reliable is the measured histogram and the dynamic analyses. However, if the acquisition time is too short to collect sufficient photons, the signal will be overwhelmed by shot noise and background noise. We therefore decided to carry out the measurement of BBL at 20 μs as a control. In order to collect enough photons within 20 μs, we increased the excitation power from 160 μW which was used in the 50 μs measurement to 240 μW. Assuming the photon emission increases linearly with the excitation power, the average burst intensity would be 60% of that in the 50 μs measurement, which means that we can use a threshold of 25 or higher to select the burst.

The resulting ET histograms are shown in Figure 4.8B. Compared with the 50 μs binning time histograms, the 20 μs histograms have broader protein-peak and smaller mean (maximum of the peak). The relative population of zero-peak is also larger and the valley between protein-peak and zero-peak has more events. The smaller average burst intensity is one reason for this observation. Moreover, with such a high excitation, our Trolox-cysteamine protector is no longer sufficient to avoid the transitions to triplet and other dark states.

Although the quality of data is not as good as that of the 50 μs binning time measurements, the SM-FRET histogram still shows that BBL has the unimodal conformal distributions under all conditions even with the 20 μs binning time.

4.6 Photon Counting Histogram and Protein Aggregation

Photon counting histogram (PCH) is the probability distribution of photon counts measured within a certain binning time. The analysis of PCH can extract very useful information from the single molecule trajectory, i.e., the average molecule number in the probe volume and the photon count rate per molecule [60-62]. Here I am going to point out how PCH can be used to check for protein aggregation without getting into the complicated analysis of PCH.

The PCH calculated from two pieces of SM-FRET trajectory are compared in Figure 4.9. Both pieces were 5 minutes long and measured using the same sample in 3 M urea and pH 7.0 with an interval of about 10 minutes. The one measured later shows a long tail in its PCH (Figure 4.9B) while the one measured earlier has a normal PCH profile (Figure 4.9A) which is similar to other experimentally and theoretically obtained PCH results [61, 62].

When many labeled protein molecules aggregate together, they may emit several photons at the same time. This would result in some extraordinarily big bursts constituting the tail of the PCH as shown in Figure 4.9B. For the above trajectory the SM-FRET histogram is also modified as seen in Figure 4.9D. However, sometimes the SM-FRET histogram still looks fine even with severe protein aggregation. In freely diffusing single molecule measurements, it is very hard to avoid molecular aggregation entirely. The practical way is to identify any possible aggregation from the measured trajectory and discard it. PCH provides a fast way to identify the aggregation.

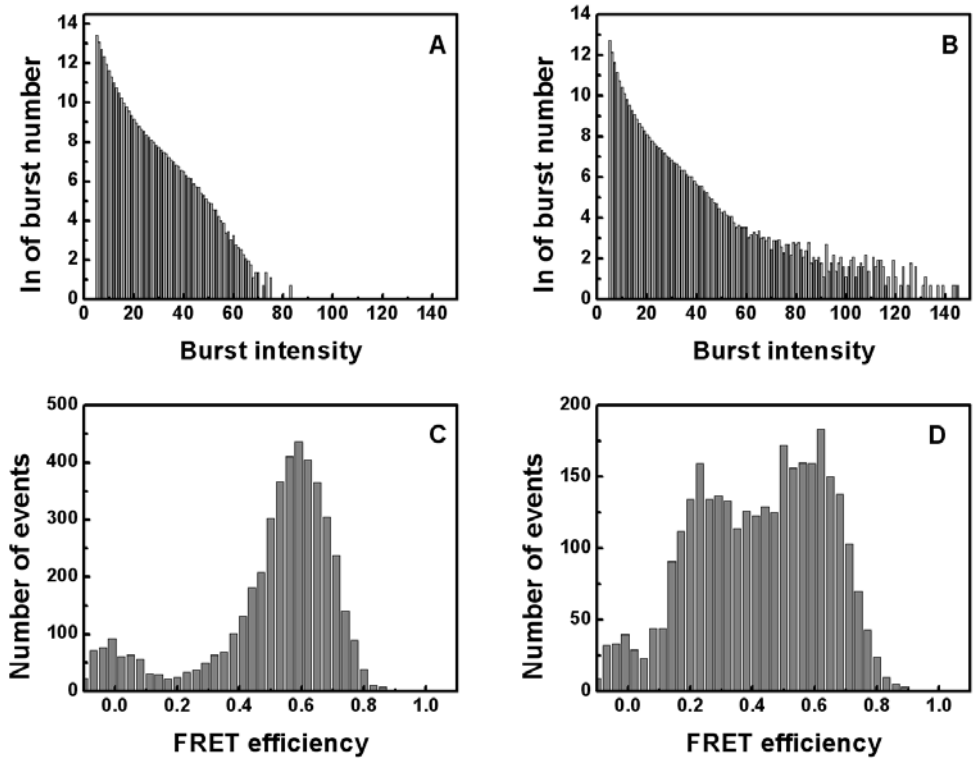

Figure 4.9 The PCH and SM-FRET histogram of two single molecule trajectory pieces which were 5 minutes long and measured using the same sample in the condition of 3 M urea and pH 7.0 with an interval of about 10 minutes. (A) and (B), the PCH of the trajectory measured earlier and the one measured later, respectively. (C) and (D), the corresponding SM-FRET histograms.

4.7 Two-State Simulation of SM-FRET Histograms

4.7.1 Theory and Method

In the single molecule FRET measurements of BBL which were carried out at 279 K and with the time resolution of 50 μs, we observed unimodal ET distributions in both urea and GuHCl unfolding of BBL. This strongly supports the proposal of global downhill folding in BBL. We know that if the ET of native and unfolded states are too close or the binning time is too long compared to the protein relaxation time, even a two-state protein may have unimodal distributions of ET during the unfolding transition [44, 54]. To exclude the possible effect of the ET baselines and make a more

convincing argument for downhill folding in BBL, we simulated the expected SM-FRET histogram if BBL were a two-state folder.

Szabo and Gopich have developed a comprehensive theory of single molecule FRET under varying conditions including the freely diffusing measurement [54, 62-68]. We did the simulation by applying their theory and using the parameters obtained from our measurements.

Introduction to SM-FRET Photon Statistics Theory. The photon emission of a doubly-labeled molecule is usually regarded as a stochastic process [66, 69, 70] and thus the photon statistics of SM-FRET can be adequately described by the Poisson distribution. For an immobilized molecule, the probability distribution of donor photons I_D and acceptor photons I_A is given by [54, 66]:

$$P(I_A, I_D \mid T_{bin}) = \frac{(n_A T_{bin})^{I_A}}{I_A!} \frac{(n_D T_{bin})^{I_D}}{I_D!} e^{-(n_A+n_D)T_{bin}} \tag{4.1}$$

where n_D and n_A are the mean photon count rates of donor and acceptor, respectively. T_{bin} is the binning time. If any dynamic process with a shorter relaxation time than the inter-photon time (interval time between two consecutive photons) are present, it does not affect the photon statistics and Eq. (4.1) still works [54].

If the molecule is not only able to freely diffuse through the probe volume but also involves dynamics with longer time scale than the inter-photon time, the probability distribution of acceptor photon I_A and donor photon I_D would become very complicated. Szabo and Gopich have given a general formula under these conditions:

$$P(I_A, I_D \mid T_{bin}) = \left\langle \frac{\left[\int_0^{T_{bin}} \sum_i n_{Ai}(t)dt\right]^{I_A}}{I_A!} \frac{\left[\int_0^{T_{bin}} \sum_i n_{Di}(t)dt\right]^{I_D}}{I_D!} e^{-\int_0^{T_{bin}} \sum_i (n_{Ai}(t)+n_{Di}(t))dt} \right\rangle_{R,r} \quad (4.2)$$

where $n_{Ai}(t)$ and $n_{Di}(t)$ refer to the acceptor and donor photon count rates of i-th molecule at time t. $\langle\ \rangle_{R,r}$ denotes to the average over all possible trajectories of all molecules' diffusional and conformational changes [54]. Eq. (4.2) is almost impossible to be solved analytically.

Simplified Formula for Freely Diffusing Molecules with Two-state Dynamics. However, if the following two conditions can be satisfied to a good approximation as well as only one molecule is detected in the probe volume, we may simplify the analysis of Eq. (4.2) [54].

Condition I: the FRET efficiency is independent of where the molecule stays in the probe volume. This means the ET only depends on the molecular conformation:

$$E(R,r) = \frac{n_A(R,r)}{n_A(R,r) + n_D(R,r)} = E(r) \quad (4.3)$$

where R and r refer to spatial coordinates and conformational coordinates.

Condition II: the total photon count rate of the two channels does not depend on the molecular conformation:

$$n_{AD}(R,r) = n_A(R,r) + n_D(R,r) = n_{AD}(R) \quad (4.4)$$

where $n_{AD}(R)$ is the photon count rate of the sum of two channels. If for different locations of the probe volume the confocal microscope system has different photon collection efficiency ratios, this condition will not be satisfied. However, usually researchers think the variation of photon collection efficiency ratio is negligible.

For a freely diffusing molecule having conformational dynamics, Szabo and Gopich suggest that if the molecule is quasi-immobilized in the time scale of binning time, the distribution of I_D, I_A (Eq.(4.2)) can be approximately factored into a product of two parts [54]:

$$P(I_A, I_D \mid T_{bin}) = P(I_{AD} \mid T_{bin}) B_c(I_A, I_D \mid T_{bin}) \tag{4.5}$$

The first term $P(I_{AD} \mid T_{bin})$ denotes the photon count distribution of the sum of two channels which is independent of the conformation. Actually it equals the PCH of the single molecule measurement.

The second term $B_c(I_A, I_D \mid T_{bin})$ denotes the photon distribution that is solely due to the conformational dynamics. For two-state conformational dynamics, $B_c(I_A, I_D \mid T_{bin})$ has a clear analytical expression [54]:

$$\begin{aligned} B_c(I_A, I_D \mid T) = \frac{I_{AD}!}{I_A! I_D!} \Big(& E_1^{I_A} (1-E_1)^{I_D} p_1 e^{-k_1 T_{bin}} + E_2^{I_A} (1-E_2)^{I_D} p_2 e^{-k_2 T_{bin}} \\ & + \frac{2kT_{bin} p_1 p_2}{E_2 - E_1} \int \varepsilon^{I_A} (1-\varepsilon)^{I_D} \left[I_0(y) + kT_{bin}(1-z) I_1(y)/y \right] e^{-k_2 T_{bin}} d\varepsilon \Big) \end{aligned} \tag{4.6}$$

where E_1 and E_2 ($E_2 > E_1$) are FRET efficiencies, p_1 and p_2 are relative populations, k_1 and k_2 are transition rates, $k = k_1 + k_2$, $p_1 = 1 - p_2 = k_2 / k$, $y = 2kT_{bin}(p_1 p_2 (E_2 - \varepsilon)(\varepsilon - E_1) / (E_2 - E_1)^2)^{1/2}$, $z = (p_1(\varepsilon - E_1) + p_2(E_2 - \varepsilon)) / (E_2 - E_1)$, and $I_n(y)$ are modified Bessel functions of the first kind.

Once we determine the joint probability distribution $P(I_A, I_D \mid T_{bin})$, it will be very easy to calculate the ET distribution, i.e., SM-FRET histogram. We just need to add up all the probability $P(I_A, I_D \mid T_{bin})$ that produces FRET efficiency $E = I_A / (I_A + I_D)$ falling into the step interval of the histogram, i.e. between $E - h/2$ and $E + h/2$ (h is the histogram step size).

Photon Discarding. As I discussed before, we can use the PCH of the single molecule trajectory as the two-channel-sum photon count distribution function: $P(I_{AD} \mid T_{bin})$. However, in the real measurement, because of unequal QY, transmission efficiencies and detector efficiencies between the two channels, the total photon count rate actually varies when the molecule lies in different conformations with different ET. To deal with this issue, Gopich and Szabo introduced the method of photon discarding [54, 66]. Instead of multiplying photon collection efficiency ratio Γ to the donor channel signal, they suggested to discard some photons from the acceptor channel (usually acceptor channel has higher photon collection efficiency, otherwise we should discard the donor photons). The discarded amount of photons is decided by a Poisson random number and the expectation of this random number is determined by Γ. After using this random photon discarding process, the trajectory qualifies the presumption that the photon count rate of two channel sum is independent of conformations.

Then we can construct PCH based on the processed trajectory and choose part of the PCH by using a threshold in the same way as we analyze the experimental SM-FRET data.

Simulation of BBL. Our single molecule measurements were conducted with a binning time of 50 μs which is much shorter than the typical diffusion time of molecules in the probe volume (~ 1 ms). Therefore, the "Quasi-immobilized" condition is fulfilled. The FRET efficiencies and relative populations of native and unfolded states E_2, E_1, p_2 and p_1 were determined from the global two-state fits to the ensemble urea and GuHCl FRET profiles.

4.7.2 Simulation Results

Simulation of BBL Urea Denaturation with a Relaxation Time 120 μs. The binning time is set to 50 μs. Li et al. determined the relaxation time of BBL in 279 K water to be around 120 μs (equivalent relaxation rate of 8.3 ms^{-1}) by using laser T-jump spectroscopy [46]. It is well known that relaxtion rates of proteins slow down upon chemical denaturation until the denaturation midpoint beyond which it increases [71]. So using a relaxation time of 120 μs for all denaturant concentrations was a very conservative assumption. A threshold of 40 was used to truncate the PCH.

The simulated SM-FRET histogram for a hypothetical BBL urea denaturation is shown in Figure 4.10. The clear bi-modal distribution during the transition is in contrast to the unimodal distribution obtained from the experimental single molecule measurements. Similar results were also obtained for the simulation of GuHCl denaturation (not shown). We also calculated the Gaussian distributions for native and unfolded states using the means of E_2, E_1 and the variances determined by the shot noise. The Gaussian distributions are plotted together with the histograms in Figure 4.10. The native peak of the histogram deviates slightly to lower ET values when compared to the calculated Gaussian distribution. This is mainly due to the conformational changes. Since the binning time of 50 μs and relaxation time of 120 μs are quite close, the protein may transfer from the native state to the unfolded state within the binning time thus producing an measured ET lower than the mean ET of native state E_2.

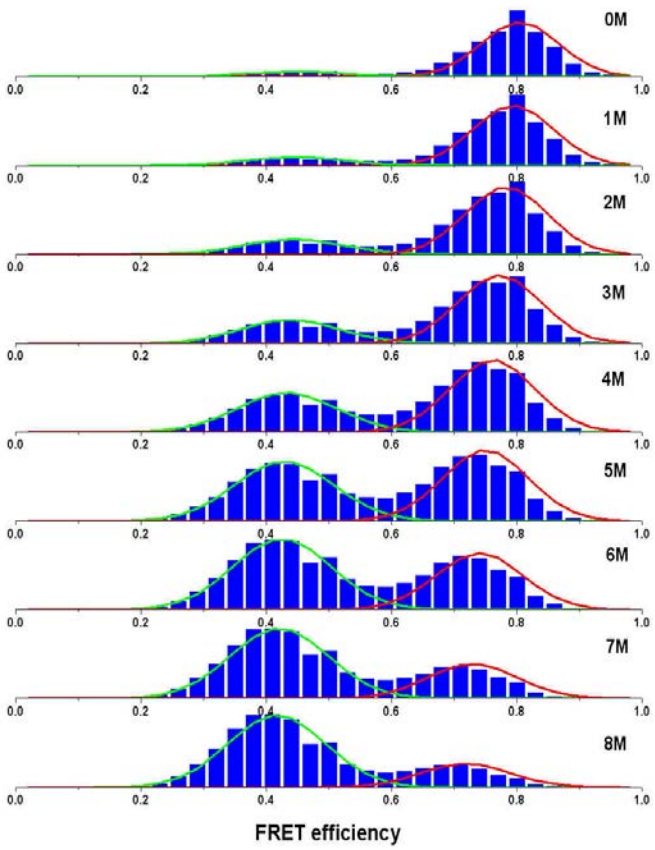

Figure 4.10 SM-FRET histogram simulation of BBL urea denaturation at 279 K with a binning time of 50 μs. The protein relaxation time was assumed to be 120 μs. The FRET efficiencies and relative populations of native and unfolded states E_2, E_1, p_2 and p_1 were determined by the two-state global fitting of the ensemble FRET profiles. The Gaussian profiles were calculated using the means of E_2, E_1, and the variances calculated from shot noise (red and green lines for native and unfolded state, respectively).

Simulation of the Transition Middle Point with Varying Relaxation Times. The relative ratio between the binning time and protein relaxation time is crucial to determine if the SM-FRET is able to extract information about conformational distributions. Although laser T-jump measurements point to a relaxation time of ~ 120 μs at 279 K [46], it is of interest to know how the ET distribution looks under different relaxation times.

With relaxation times between 20 ~ 150 μs, we simulated the SM-FRET histograms at the transition midpoint of 5.1 M. The ET were E_2 = 0.75 and E_1 = 0.43 for native and unfolded states, respectively. The binning time was 50 μs and a threshold of 40 was used to truncate the PCH.

As shown in Figure 4.11A, if the relaxation time is 40 μs or shorter, we will not be able to see a clear unimodal distribution. Considering the peak broadening due to scattering noise in the real measurement, it is quite probable that the valley between two peaks will not be recognized even with a relaxation time of 80 μs. However, even if the native and unfolded peaks merge into one peak, the simulated histogram is still significantly different from the one constructed from the measurement because the variance of the very broad peak is much larger than what it supposed to be solely due to the shot noise. In contrast, the experimental histogram is just a little bit wider than the width determined by shot noise. The histogram width will be analyzed carefully in the following simulation with a relaxation time 20 μs.

Simulation with Relaxation Time 20 μs. At 279 K, the relaxation time of BBL must be much larger than 20 μs according to the Laser T-jump measurements of BBL. So setting the relaxation time as 20 μs will be a rigorous test to the single molecule measurements on BBL. The simulation result of BBL urea denaturation with 20 μs relaxation time and 50 μs binning time is shown in Figure 4.11B.

As expected, the native and unfolded peaks merge into one peak when they coexist using the above parameters. However, the profile of the merged peak is significantly asymmetrical. In Figure 4.11B, Gaussian profiles do not fit the

histograms cleanly except at low denaturant concentrations. In contrast, all the experimental histograms can be perfectly fitted to Gaussian profiles.

We also compared the width of the simulated histogram and the experimental histogram. The variance of single molecule ET distribution is mostly due to the shot-noise [54, 66]:

$$\sigma_{shot}^2 = \langle ET \rangle (1 - \langle ET \rangle) \langle N^{-1} \rangle \tag{4.7}$$

where $\langle ET \rangle$ is the mean of the ET distribution and $\langle N^{-1} \rangle$ is mean of the burst intensity reciprocals which is very close to the reciprocal of threshold. The non-shot-noise variances is given by

$$\sigma_{non-shot}^2 = \sigma_{fit}^2 - \sigma_{shot}^2 \tag{4.8}$$

where σ_{fit}^2 is the variance determined from the Gaussian fit of the histogram. The $\sigma_{non-shot}^2$ of the simulated and experimental histograms are compared in Figure 4.12. We can see the $\sigma_{non-shot}^2$ of the simulated histogram reaches a maximum at the transition middle point and then decreases. This is because two states are equally distributed in the transition midpoint and thus the superposition of two peaks is the broadest. However, the histogram $\sigma_{non-shot}^2$ from the measurement monotonically increases.

The other more important issue is that the noise in measurements will broaden the histogram further. This means the experimental $\sigma_{non-shot}^2$ must be larger than the simulation $\sigma_{non-shot}^2$. However, as far as we can see from Figure 4.12, the experimental $\sigma_{non-shot}^2$ is lower than the $\sigma_{non-shot}^2$ obtained from simulation between 0 ~ 6 M urea.

Based on the above analysis, we can conclude that BBL is still not two-state folder even if the relaxation time of BBL were 20 μs at 279 K.

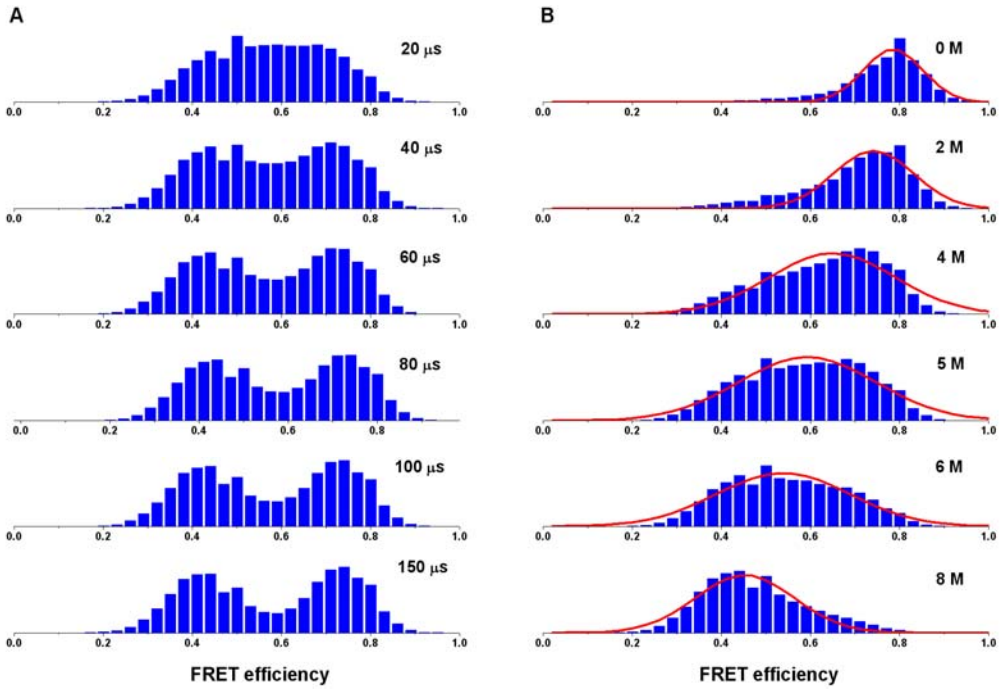

Figure 4.11 50 μs binning time histogram simulation of BBL urea denaturation. (A), the transition midpoint of 5.1 M urea and E_2 = 0.75, E_1 = 0.43 were obtained from global fitting of ensemble FRET curves; the relaxation time was assumed to be 20, 40, 60, 80, 100 and 150 μs. (B), urea concentrations were 0, 2, 4, 5, 6, 8 M, respectively; E_2 and E_1 were obtained from global fitting of ensemble FRET curves; the relaxation time was set as 20 μs.

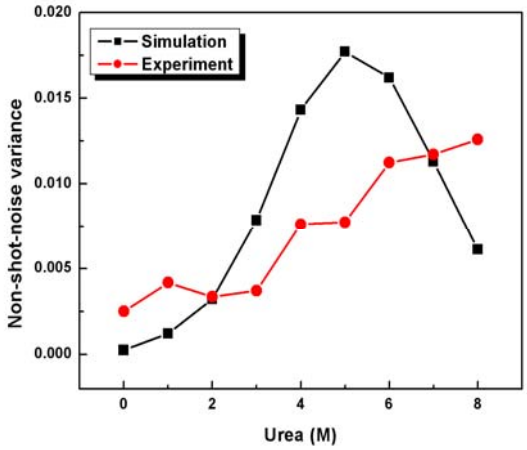

Figure 4.12 Non-shot-noise variances of the simulated histogram (black square) and experimental histogram (red circle).

4.8 Conclusions

We successfully measured the SM-FRET efficiency histogram of BBL chemical denaturation at 279 K and pH 6.0 using a binning time of 50 μs. The histograms show unimodal profiles at all degrees of unfolding stress. We also tried a binning time of 20 μs with increasing excitation power by 50 %. The data was not as good, but it still showed unimodal ET distributions.

As a reference, the ensemble FRET measurements of BBL unfolding by urea and GuHCl were performed in the same condition as the single molecule measurement, i.e. 279 K and pH 6.0. The urea and GuHCl unfolding curves superimposed on each other very well with a scaling factor of 2.75 between the urea and GuHCl concentrations. The means of the single molecule ET distributions were pretty consistent with the ensemble FRET efficiencies. Only two single molecule data points (2 and 3 M GuHCl) deviated from the ensemble curve which is probably related to the variation of ensemble ET during cooling.

The ensemble ET curves of labeled BBL were superimposable on the far-UV CD curve of unlabelled BBL. Moreover, two-state fits to the far-UV CD and ensemble ET curves produced identical ΔG_{H2O} and m-values within the fitting error. So we can safely conclude that the labeling does not significantly affect the folding dynamics or the stability of BBL.

To exclude the possible effect of the modest pH change on the conformational distributions, the same ensemble and single molecule measurements of BBL were also carried out in pH 7.0. We obtained similar results in pH 7.0 measurements as

well. However, we found the doubly-labeled BBL had a higher propensity to aggregate and adsorb at pH 7.0 compared to pH 6.0.

Employing Szabo's photon statistics theory of SM-FRET, we simulated the possible SM-FRET efficiency histograms if BBL were a two-state folder. The simulation showed clear bimodal ET distributions under the real conditions. Assuming a very conservative relaxation time of 20 μs at 279 K, the ET distribution showed the unimodal profiles as the result of the merging of native and unfolded peaks. However, such unimodal profiles were asymmetrical and deviated from a Gaussian profile at the midpoint. Furthermore, if our measured unimodal SM-FRET histogram was a result of the merging of native and unfolded peaks, we should have obtained a significantly larger non-shot-noise variance of the histogram than the one experimentally measured.

While the control experiments confirmed that the unimodal ET distribution was not an artifact caused by modest pH change or labeling, the SM-FRET measurement and simulation provide strong evidence to the downhill folding nature in BBL.

Chapter 5 FRET Characterization of Two-State Folding Protein α-spectrin SH3

5.1 Introduction

5.1.1 Research Objective

SM-FRET measurements are inevitably affected by scattering from out-of-focus molecules which is generally referred to as background noise. If the background noise is too high, even a two-state protein may produce a unimodal SM-FRET efficiency histogram [72]. So it is critical to ensure that the signal from only ONE molecule is detected in the probe volume with a good signal-to-noise ratio. If we are able to observe the coexistence of two subpopulations (i.e. folded and unfolded states) from the SM-FRET histogram of a previously well characterized two-state protein, i.e. α-spectrin SH3, the instrument could be guaranteed to have a good signal-to-noise ratio to detect the single molecule signal. This would also mean that the unimodal SM-FRET efficiency histogram of BBL was not an artifact due to any instrumental error. Moreover, it serves as a very good control experiment for many experimental variables including the fluorophore protector system and extrinsic fluorescent labels.

As with BBL, we followed the unfolding of α-spectrin SH3 using both urea and GuHCl. The effects of those two denaturants on α-spectrin SH3 are compared. A two-state analysis of ensemble FRET measurement shows that the labeling does not affect the protein stabilities. The comparison of ensemble and single molecule results provides some information about native and unfolded states' conformations of α-

spectrin SH3 as a function of the chemical denaturant. All the ensemble and single molecule measurements were carried out at room temperature.

α-spectrin SH3 and BBL are good model systems for two-state and downhill folding proteins, respectively. A comparison of their ensemble experimental behavior provides deep insights into protein folding mechanism as discussed below.

5.1.2 Background of *α*-spectrin SH3 Domain

The SH3 (Src homology region 3) domain is present in many proteins [73, 74]. It was first identified in *src* tyrosine kinases [75], then in cytoskeletal protein *α*-spectrin [73] and later in other proteins [76]. The biological role of SH3 domain is not very clear, but people have found that SH3 domains bind proline-rich ligands having the sequence xP-x-xP, where "x" represents any amino acid [77, 78].

The SH3 domain of *α*-spectrin was previously characterized as a two-state folding protein [79, 80]. The sequences of wild type and variant *α*-spectrin SH3 are shown below. In the variant, two cysteines are introduced at the termini that act as labeling sites. The additional lysines beside the cysteines are designed to attract fluorophores and speed up the labeling reaction.

Wild type M DETGKELVLA LYDYQEKSPR EVTMKKGDIL TLLNSTNKDW WKVEVNDRQG FVPAAYVKKL D
Variant MACKK DETGKELVLA LYDYQEKSPR EVTMKKGDIL TLLNSTNKDW WKVEVNDRQG FVPAAYVKKL DGKKC

5.1.3 Different Buffer Conditions

α-spectrin SH3 domain is a very stable protein and thus it is not possible to observe a complete urea unfolding curve under the normal conditions (i.e. at room temperature and pH 7.0). Therefore, most of the reported experiments have been carried out at lower pH to destabilize the protein. In our measurement we initially compared the denaturation in pH 5.0 acetate and pH 3.5 citric buffers. The protein

was well destabilized at pH 3.5 resulting a long post-transition regime. However, at pH 3.5 the protein showed a high propensity to aggregate or adsorb onto surfaces thus precluding a clean single molecule measurement. At pH 5.0 the protein solution was relatively well-behaved. So the single molecule measurement of α-spectrin SH3 was performed in 20 mM acetate buffer at pH 5.0. Both urea and GuHCl ensemble unfolding measurements were obtained providing clear pre- and post-transition baselines.

5.2 Ensemble FRET Measurement

5.2.1 Urea Denaturation of SH3 at pH 3.5 and pH 5.0

We initially compared the ensemble FRET measurements of urea denaturation at pH 3.5 and pH 5.0. The results are shown in Figure 5.1. The FRET efficiency ratiometry (ET_{ratio}) is defined as the donor intensity divided by total fluorescence intensity while the accurate FRET efficiency (ET) is calculated by using the formula explained in Chapter 2. The corrections for R_0 change are not included in this figure. We can see that ET is marginally lower than ET_{ratio} and the ET corrections do not change it too much. Obviously, pH 3.5 produces a longer post-transition regime, but the single molecule measurement under this condition was plagued by protein aggregation or adhesion.

Using a two-state model, we obtained ΔG_{H_2O} and m-value of 11.0 ± 0.6 kJ mol^{-1} and 2.69 ± 0.12 kJ mol^{-1} M^{-1}, respectively for the pH 3.5 data. At pH 5.0, the corresponding values are 13.5 ± 0.6 kJ mol^{-1} and 2.50 ± 0.12 kJ mol^{-1} M^{-1},

respectively. As shown in Table 5.1 the fitting results are consistent with those values reported for the wild type α-spectrin SH3 [80]. That means the fluorophore labeling does not affect the protein stability and overall behavior.

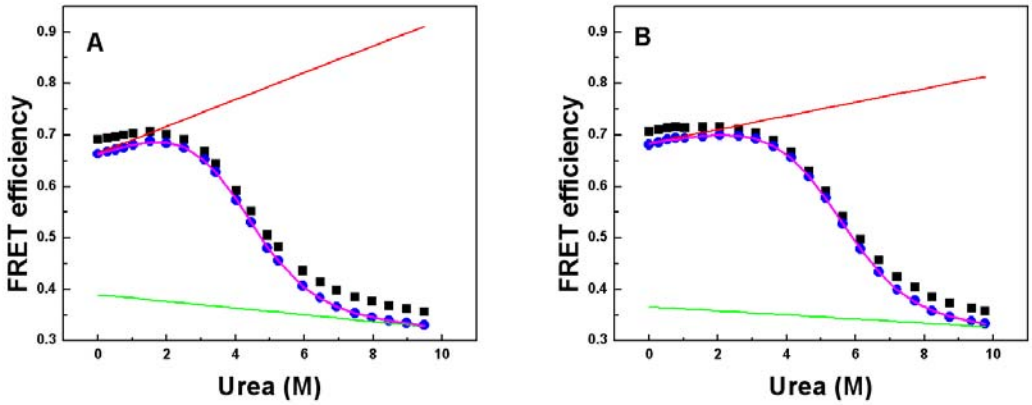

Figure 5.1 The urea denaturation profiles of α-spectrin SH3 at pH 3.5 (A) and pH 5.0 (B). The accurate FRET efficiency (blue dot) was fitted to a two-state model. The native and unfolded state baselines (red and green lines, respectively) and the fitting curve (magenta line) are plotted. The ratiometry of FRET efficiency (black square) is also displayed.

5.2.2 Urea and GuHCl Denaturations of α-spectrin SH3 at pH 5.0

Similar to the BBL denaturation data, the urea and GuHCl unfolding curves of a-spectrin SH3 can be superimposed using a scaling factor of 2.85 as shown in Figure 5.2. Because GuHCl is charged and urea is not, they bind to the protein by different interactions [81-83]. For some relatively big proteins, the denaturations by GuHCl and urea can result in different intermediates [84]. Also, at low concentration, GuHCl may stabilize the protein in some cases [82, 85]. However, for most small proteins, GuHCl and urea globally unfold the protein in the similar way. The denaturation of α-spectrin SH3 and BBL (see Chapter 4) are such cases. Greene and Pace compared the denaturation of several proteins by urea and GuHCl [81]. They found that each protein

had the similar urea and GuHCl denaturation profiles, but the ratio of the unfolding strength between GuHCl and urea varied from 2.8 for ribonuclease to 1.7 for lysozyme. We can see the scaling factors of α-spectrin SH3 and BBL (it is 2.75) lies in the high range.

With free floating baselines, the global fitting of the urea and GuHCl ET unfolding curves gives the values of ΔG_{H_2O} and m-value of 12.9 ± 1.4 kJ mol^{-1} and 6.9 ± 0.6 kJ mol^{-1} M^{-1} (for GuHCl), respectively. The equivalent m-value is 2.4 kJ mol^{-1} M^{-1} for urea. The fitting parameters are consistent with previous individual fitting results (Section 5.2.1) of urea denaturation curve within the fitting error.

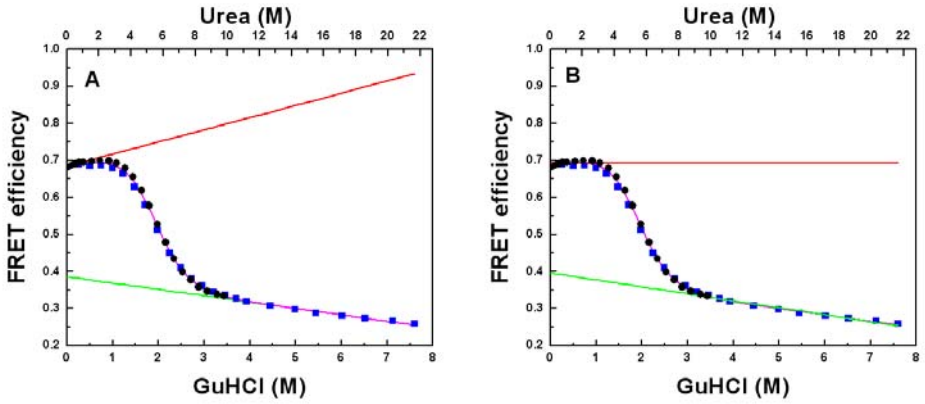

Figure 5.2 Superposition of urea (black dot) and GuHCl (blue square) denaturation profiles of α-spectrin SH3 with a scaling factor 2.85 between urea and GuHCl concentrations. The FRET efficiency is calculated with necessary corrections, but the R_0 changes are not considered. The global fitting curve (magenta line), native (red line) and unfolded (green line) state baselines are also plotted. (A), free floating native baseline. (B), zero-slope native baseline.

However, as shown in Figure 5.2A, the native baseline increases significantly with denaturant thus not characterizing the ET of native state well. So I also tried the global fit with the native baseline slope fixed to be zero as shown in Figure 5.2B. The obtained values of ΔG_{H_2O} and m-value are 15.7 ± 2.2 kJ mol^{-1} and 8.0 ± 1.0 kJ mol^{-1}

M^{-1} (for GuHCl), respectively. These two fitting parameters are both larger than the corresponding values obtained from the above free floating baseline fit. But the difference seems not large given the fitting error.

Table 5.1 Two-state fitting parameters of α-spectrin SH3 denaturation

Method	pH	Denaturant	ΔG_{H_2O} (kJ mol-1)	m (kJ mol-1 M-1)
FRET*	3.5	urea	11.0 ± 0.6	2.69 ± 0.12
FRET*	5.0	urea	13.5 ± 0.6	2.50 ± 0.12
FRET*	5.0	urea and GuHCl‡	12.9 ± 1.4	6.9 ± 0.6‡‡
FRET*	5.0	urea and GuHCl§	15.7 ± 2.2	8.0 ± 1.0‡‡
Fluorescence†	3.5	urea	12 ± 0.2	3.1 ± 0.03
CD†	3.5	urea	12 ± 0.1	3.2 ± 0.06
Fluorescence†	5~7⊥	GuHCl	15.5 ± 1.3	7.9 ± 0.5

*These measurements were carried out on the labeled α-spectrin SH3 domain by us. †These measurements were performed on the wild type α-spectrin SH3 by Viguera and co-workers [80]. ‡Urea and GuHCl denaturation curves are globally fitted to the two-state model with free floating baselines. §The slope of native baseline is fixed to be zero in the global fitting of Urea and GuHCl denaturation curves. ‡‡These m-values are for GuHCl and their equivalent values for urea should be divided by 2.85. ⊥Between pH 5 and 7 ΔG_{H_2O} and m-value are obtained with relatively large error by Serrano and co-workers [80].

5.3.3 R_0 Corrected ET

As discussed in Chapter 2, the Förster distance R_0 actually changes upon denaturation because it is dependent of the donor QY and refractive index of the medium. As a result, even if the inter-fluorophore distance does not change, the measured ET may change upon denaturation. We can correct for this effect and calculate the R_0 corrected ET using the formula developed in Chapter 2. The original ET and R_0 corrected ET curves are plotted in Figure 5.3.

In both urea and GuHCl denaturations, the ET without R_0 correction shows a positive slope at low denaturant denaturation, which is more pronounced in the urea denaturation curve. After the correction for R_0 change, the pre-transition baseline becomes relatively flat and hence more reasonable. Similar changes are evident in the post-transition regime of the GuHCl denaturation: the R_0 corrected ET almost does not change after 6 M GuHCl while the uncorrected curve shows a negative slope. However, the R_0 correction only affects the baseline of a two-state fit with little effect on ΔG_{H_2O} and m-value.

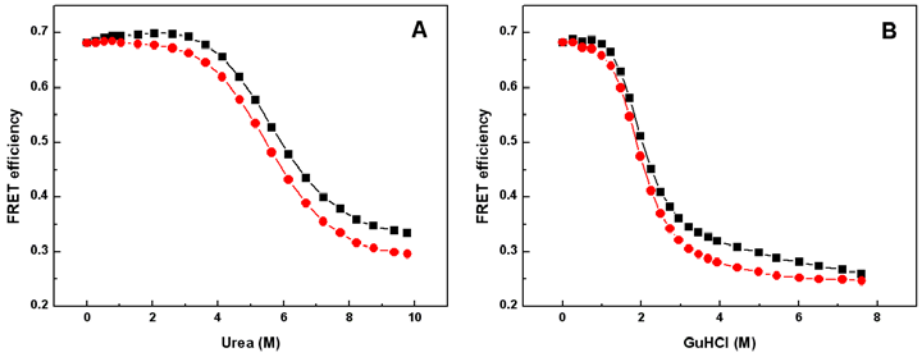

Figure 5.3 Denaturation profiles of original ET (black) and R_0 corrected ET (red). Panel (A) and (B) are for urea and GuHCl, respectively.

5.3 SM-FRET Measurement

5.3.1 SM-FRET Histograms of Urea and GuHCl Denaturations

The SM-FRET histograms are shown in Figure 5.4. The peak centered around 0.85 corresponds to the native state and the other peak with a lower ET is the unfolded state. As the concentration of the denaturant is increased the intensity of the folded peak drops while that of the unfolded state simultaneously increases. The third peak with an ET around zero is due to donor-only (or acceptor-inactive) molecules.

The ET of zero-peak maximum is not exactly zero, indicating that the ET correction is not exact. Since we do not intend to calculate the inter-dye distances, the uncertainty in ET does not affect our conclusions.

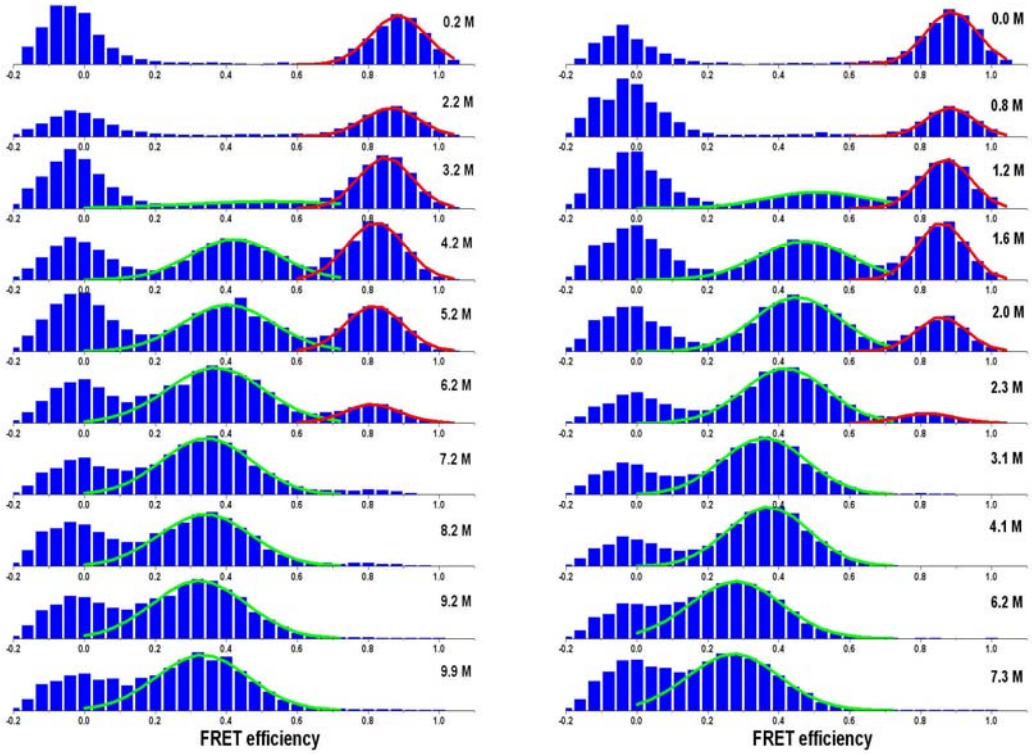

Figure 5.4 SM-FRET histograms of α-spectrin SH3 denaturation. Left column is the urea denaturation and right column is the GuHCl denaturation. The histogram was fitted to a sum of one lognormal function for the zero peak and two Gaussian functions for the native (red line) and unfolded peaks (green line).

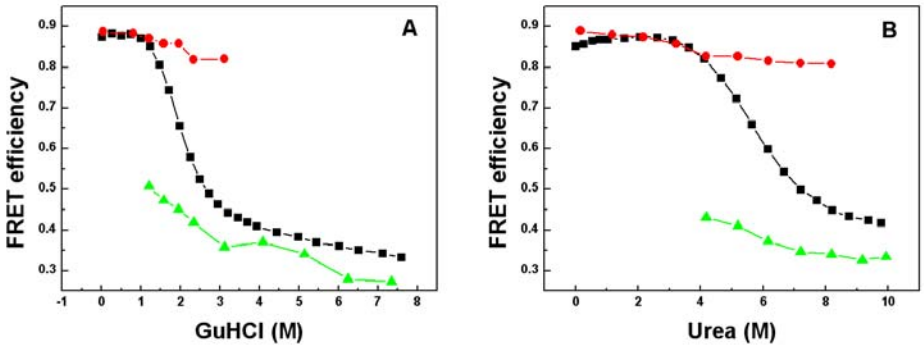

Figure 5.5 The fitted mean values from SM-FRET histogram's naive (red circle) and unfolded (green triangle) peaks. The ensemble ET (black square) is multiplied by 1.28 to correct the effect of donor-only molecules in the sample. Panel (A) and (B) are for GuHCl and urea denaturations, respectively.

5.3.2 Analysis of SM-FRET Histograms

To get the mean and variance of native and unfolded peaks, we fit the SM-FRET histograms to a sum of a lognormal function for the zero-peak and two Gaussian functions corresponding to the native and unfolded peaks as shown in Figure 5.4. The histogram variance will be analyzed in Chapter 6. The fitted mean values of native and unfolded peaks: $\overline{ET_n}$ and $\overline{ET_u}$, are plotted in Figure 5.5 along with the ensemble ET curves. The ensemble ET has to be multiplied by 1.28 to correct the effect of donor-only molecules in the sample for comparison with the SM-FRET means, as discussed in Chapter 2. The number 1.28 was determined by overlapping the ensemble ET and single molecule $\overline{ET_n}$ since it is challenging to estimate the acceptor labeling efficiency.

$\overline{ET_n}$ shows a slight dependence on urea and GuHCl concentrations while $\overline{ET_u}$ displays a larger dependence, with both decreasing. In the above analysis of the ensemble FRET denaturation curve, we have found that the pre-transition and post-transition regimes were very flat once corrected for the R_0 changes. So the decrease in both $\overline{ET_n}$ and $\overline{ET_u}$ is unlikely due to changes in inter-dye distance. Instead, it is mostly due to background scattering and changes in R_0.

When the denaturant concentration is increased and the unfolded state becomes more populated, the signal of the native protein is mixed with increased scattering from the unfolded molecules. So $\overline{ET_n}$ becomes smaller for higher denaturant concentration samples. The scattering from donor-only protein also

decreases the native state ET, but it has identical effect on different denaturant concentration samples, and hence the effect is not explicit.

The mean of the unfolded peak in GuHCl denaturation experiments $\overline{ET_u}$ decreases rapidly between 1 and 3 M. A major reason for this observation is the reduced scattering from native proteins. After the completion of the transition at about 3 M GuHCl, $\overline{ET_u}$ decreases very slowly. This is mostly caused by the variation in R_0 as there is no any scattering from the native protein under these conditions. However, it is hard to explain why $\overline{ET_u}$ drops significantly between 5 and 6 M GuHCl. This is probably due to stronger scattering at large GuHCl concentrations that in turn increases the background scattering from donor-only protein.

The $\overline{ET_u}$ from urea denaturation studies also decreases rapidly around the transition region and then decreases slowly beyond this region due to changes in R_0 as observed in GuHCl experiments. It is noticeable that the $\overline{ET_u}$ is significantly lower than the ensemble ET in the case of urea denaturation. Previously, we noted that 1 M GuHCl is as strong as 2.85 M urea which means that10 M urea is equivalent to only 3.5 M GuHCl. So at 10 M urea the transition is still not complete resulting in a tiny population of native-like proteins. This in turn makes the ensemble FRET higher than $\overline{ET_u}$. However, because of background scattering, the small portion of native protein is not clearly shown in the histogram of 9.9 M urea sample (see left column of Figure 4.5).

Actually, it is very unreliable to relate the populations of native and unfolded states from the SM-FRET histogram to the equilibrium constant between the two

states. As discussed above, the background scattering will induce the mixing of signals between proteins of different conformations. Moreover, because of the different photon collection efficiencies of the two channels, the fraction of emitted photons that are detected is different for proteins with different ET (in other words, proteins in different conformations). So when we use a certain threshold to select informational bursts, bursts from different conformations may have different probabilities of being selected. This would result in their inaccurate representation in a SM-FRET histogram complicating the interpretation.

5.4 Conclusions

The clear observation of a bi-modal distribution in the SM-FRET histogram of α-spectrin SH3 demonstrates that our experimental setup has a good "ensemble resolution" to detect sufficient signals from single molecules. The SM-FRET measurement of α-spectrin SH3 was carried out in the identical conditions as the BBL experiment except that it was conducted at room temperature instead of 279 K: the incident excitation beam power was 160 μW and the binning time was 50 μs; Alexa 488 / 594 dye pair was protected by 1 mM Trolox and 10 mM cysteamine. So we can definitely exclude the possibility that the unimodal ET distribution of BBL was caused by the background scattering from out-of-focus molecules.

In most SM-FRET measurements, the unfolded peak shifts to lower ET values with increasing denaturant concentrations [33, 39]. The origin of the shift is still not very clear. It can be either due to the expansion of the protein or measurement issues. In the case of α-spectrin SH3 domain, we see similar shifts in the unfolded peak.

However, in the ensemble FRET measurement of α-spectrin SH3 denaturaiton by GuHCl, the post-transition regime becomes very flat after the correction for R_0 changes. That indicates the inter-dye distance does not increase once the protein is completely unfolded. Therefore, we propose that the observed changes in SM-FRET histogram $\overline{ET_u}$ value are mostly due to background scattering. Especially it seems that very high denaturant concentrations can enhance the background scattering significantly. Thus the unfolded peak of samples with high denaturant concentrations shifts notably if there are some donor-only proteins in solution.

Chapter 6 Dynamic Analysis of SM-FRET histograms

6.1 Introduction

One fundamental character of a SM-FRET histogram is that its broad distribution is mostly due to the stochastic emission of photons. Such stochastic processes cause fluctuations in photon burst intensity when measured within a certain binning time and consequently result in the fluctuation of FRET efficiency (ET), which is called the shot noise [66]. Usually the photon burst intensity can be described by a Poisson distribution and the shot noise variance of single molecule ET is given by:

$$\sigma_{shot}^2 = \langle ET \rangle (1 - \langle ET \rangle) \left\langle \frac{1}{N} \right\rangle \tag{6.1}$$

in the freely diffusing experiment [66]. $\langle ET \rangle$ is the mean single molecule ET and $\left\langle \frac{1}{N} \right\rangle$ is the mean of the reciprocal of burst intensity which is very close to the threshold reciprocal.

However, people have found that the unfolded peak in a SM-FRET histogram has larger variance than σ_{shot}^2 while the native peak variance is very close to σ_{shot}^2. The extra variance is referred as non-shot-noise variance as discussed in Chapter 4:

$$\sigma_{non-shot}^2 = \sigma^2 - \sigma_{shot}^2 \tag{6.2}$$

The origin of $\sigma_{non-shot}^2$ is not very clear. One possible origin is the protein conformational dynamics. If the timescale of protein dynamics is similar to or longer than the binning time, different bins (acquisition interval time window) may catch

different protein conformations and thus the measured ET varies. If this is the case, $\sigma^2_{non-shot}$ should decrease when we increase the binning time.

I change the binning time by merging adjacent bins into one bin and check for the effect on $\sigma^2_{non-shot}$. From the change I try to explain the possible origin of $\sigma^2_{non-shot}$. SM-FRET histograms of BBL and α-spectrin SH3 at varying conditions are analyzed as well as the histogram of the fluorescence protein RP-E.

6.2 Analysis Methods

6.2.1 How to Select Informational Bins

Informational bins refer to those bins having a good signal-to-noise ratio and thus able to provide sufficient information about the detected molecules. The typical method to select informational bins is by using a threshold and choosing all those bursts above the threshold. When we change the original trajectory into trajectories with larger binning times, the simplest way to select bins is to apply the same threshold to trajectories with different binning times. However, in this way the calculated $\sigma^2_{non-shot}$ does not decrease with increasing binning times. Actually it increases as shown below in Figure 6.1A.

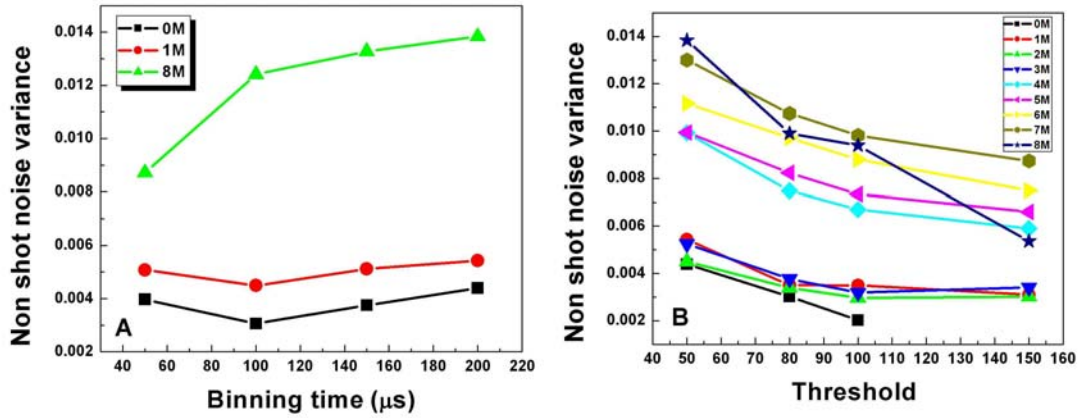

Figure 6.1 SM-FRET histogram $\sigma^2_{non-shot}$ of BBL urea denaturation at 279 K. (A), the informational bins were selected by a constant threshold of 30 in different binning time trajectories. $\sigma^2_{non-shot}$ increases with binning time. (B), the informational bins of 200 μs trajectory were selected by different thresholds. $\sigma^2_{non-shot}$ decreases with increasing thresholds.

It seems more reasonable to use different thresholds for different binning times instead of the same threshold. Here a question arises: does the threshold affect the calculation of $\sigma^2_{non-shot}$? The answer is yes as shown in Figure 1B. This means the threshold is very critical for the analysis. However, it is impossible to derive a reasonable formula to calculate thresholds for varying binning times.

Burst Marking Method to Select Informational Bins. To avoid the problem to decide thresholds for varying binning time trajectories, I introduce another method — burst marking method: select only the informational bins in the original trajectory (50 μs binning time); for larger binning time trajectories, instead of selecting the bins by thresholds, we choose those bins that include at least one bin which is already selected in the original trajectory.

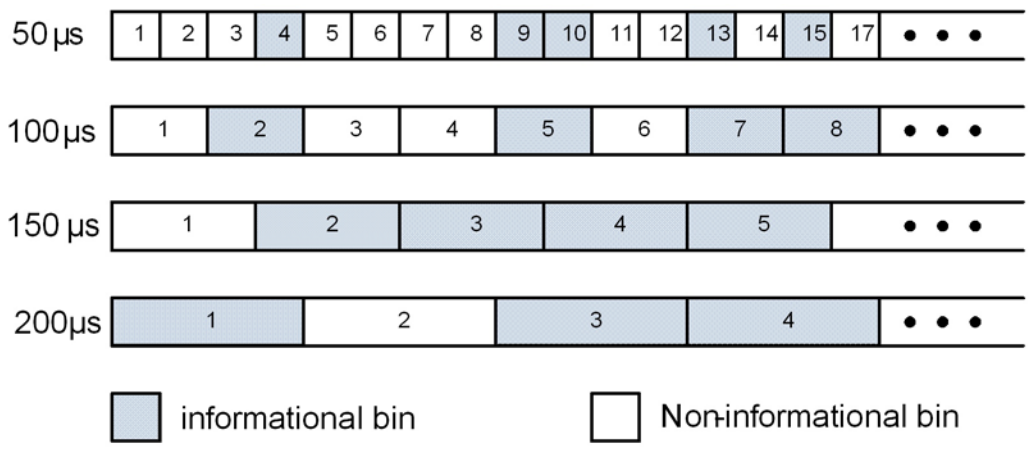

Figure 6.2 Illustration of the burst marking method to select informational bins. The light blue box represents the informational bin and the empty box represents the non-informational bin.

This method is illustrated in Figure 6.2. In the original 50 µs trajectory we decide on the informational bins 4, 9, 10, 13, 15 by using a certain threshold. Then in 100 µs trajectory, bin-1 which is made of two non-informational 50 µs bins is still regarded as a non-informational bin. Bin-2 which is made of one informational 50 µs bin and one non-informational 50 µs bin is recognized as an informational bin. For any longer binning time trajectories, bins are identified as informational or non-informational bins similarly based on if they include at least one informational bin of the original trajectory.

Using this method we can avoid choosing thresholds for different binning time trajectories. Also, since we only select the bursts just once, the informational bursts with different binning times can be regarded as coming from the same "detected molecular ensemble", which makes the comparison among different binning time trajectories more practical.

The threshold that is used to select bursts in the 50 μs trajectory can not be too large. Otherwise the histogram does not have enough bursts and the fitting results of the histogram will fluctuate greatly.

6.2.2 Issues of Photon Discarding

As discussed in Chapter 4, Szabo's theory has a basic assumption: the photon count rate from the summation of the two channels should be a constant at all conformations of the molecule [54, 66]. Due to the inequality of photon collection efficiencies in the two channels, the photon count rate varies for different conformations having different ET. So it is necessary to use the photon discarding method to delete some photons in order to fulfill the above assumption, which is explained in Section 4.7.1.

To obtain accurate ET we also need to correct for the effects of cross talks between two channels and acceptor direct excitation as discussed in Chapter 2. However, with photon discarding method, such corrections become complicated. Also those effects are not supposed to affect the analysis of the histogram width ($\sigma^2_{non-shot}$ changes very little when I made the above corrections in one set of data). So in the following analysis we do not make such corrections in the ET calculation.

6.2.3 Fitting of SM-FRET Efficiency Histograms

After selecting the informational bins and discarding some photons, we can build up the SM-FRET efficiency histograms of varying binning times as illustrated in Figure 6.3. The original 50 μs trajectory was obtained from the 279 K measurement of BBL in 4 M urea. The informational 50 μs bins were selected using a

threshold of 30. Because we did not make any correction for the cross talks and the acceptor direct excitation the zero-peak was not centered at zero, but it did not affect the analysis of $\sigma^2_{non-shot}$.

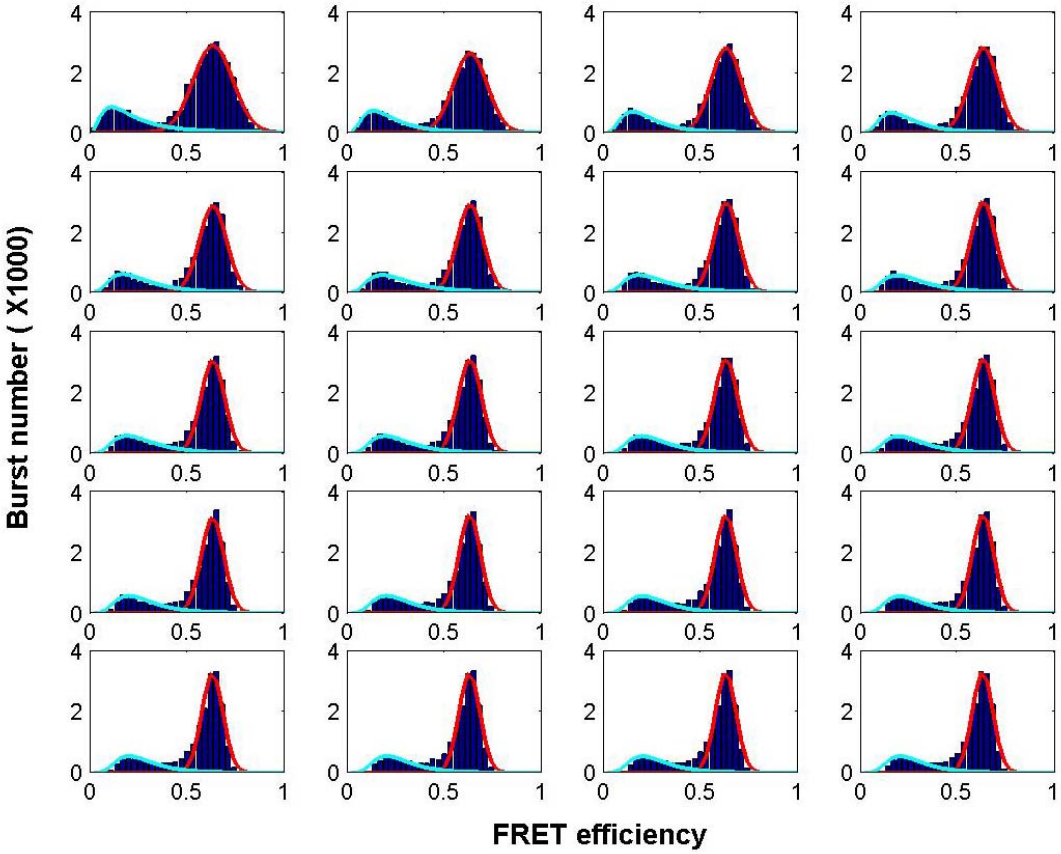

Figure 6.3 SM-FRET histograms of BBL in 4M urea solution and at 279 K. From left to right and then from top to bottom, the binning time changes from 50 μs to 1ms with the increasing step of 50 μs. The histograms were fitted to the sum of a lognormal function for the zero-peak (blue line) and a Gaussian function for the protein peak (red line).

In order to obtain the histogram variance σ^2 a sum of lognormal and Gaussian functions was used to fit the histogram as shown in Figure 6.3. The lognormal function fits the zero-peak better than Gaussian function because the zero-peak is not symmetrical. However, we also tried a sum of two Gaussian functions to

fit the histogram and found that the fitting of zero-peak almost did not affect the fitted variance σ^2 of the protein peak.

6.3 Decay of BBL $\sigma^2_{non-shot}$ vs. Binning Times

Applying the methods of burst selection, photon discarding and histogram fitting we introduced above, we analyzed the $\sigma^2_{non-shot}$ of BBL urea denaturation at 279 K as shown in Figure 6.4A. Except for 0 M sample other samples show clear decays in $\sigma^2_{non-shot}$ vs. binning time plots. All the $\sigma^2_{non-shot}$ plots approach asymptotically to some baselines. The baseline is higher for samples with higher denaturant concentrations. If such decay is caused by protein conformational dynamics, the time-constant of the decay will be able to provide information on the protein relaxation time and other dynamic insights.

Exponential Fit of $\sigma^2_{non-shot}$ Decay. The most straight forward way to fit the decay is to use a simple exponential function plus a term for the baseline:

$$\sigma^2_{non-shot} = \sigma^2_{other} + \sigma^2_c e^{-kT_{bin}} \tag{6.3}$$

where σ^2_{other} is the baseline which corresponds to other unknown contribution to the histogram width. σ^2_c is the changing amplitude of $\sigma^2_{non-shot}$ which is supposed to come from the protein conformational relaxation. k describes the decay rate of the conformational variance and equals the protein relaxation rate. T_{bin} is the binning time.

The fitting to the exponential function is good as shown in Figure 6.4A. The fitting of 0 M sample is meaningless because its $\sigma^2_{non-shot}$ does not change. The fitting parameters are plotted in Figure 6.4B, C and D. We can see the decay rate *k* decreases with urea concentrations. This is qualitatively consistent with what we expect from BBL kinetics measurements. The corresponding relaxation times of 1~3 M urea samples are of the order of 100 μs which is close to the 120 μs relaxation time obtained from T-jump measurement of BBL at 279 K water [46]. The decay times of 4~8 M samples are slightly higher than what we expect. Since the increasing step of binning time is 50 μs, the fitting of $\sigma^2_{non-shot}$ decay could not be very accurate.

The conformational variance σ^2_c increases monotonically with urea concentrations. This is reasonable indicating that the protein becomes more loose, flexible and unstructured upon denaturation. The baseline σ^2_{other} also increases with urea. It is hard to explain why samples with high concentration of urea have very high baselines. One possible reason is that σ^2_{other} is caused by background scattering which becomes stronger at higher urea concentrations. The assumption of σ^2_{other} to be constant may not be the case as well, but it is difficult to estimate its dependence on the binning time.

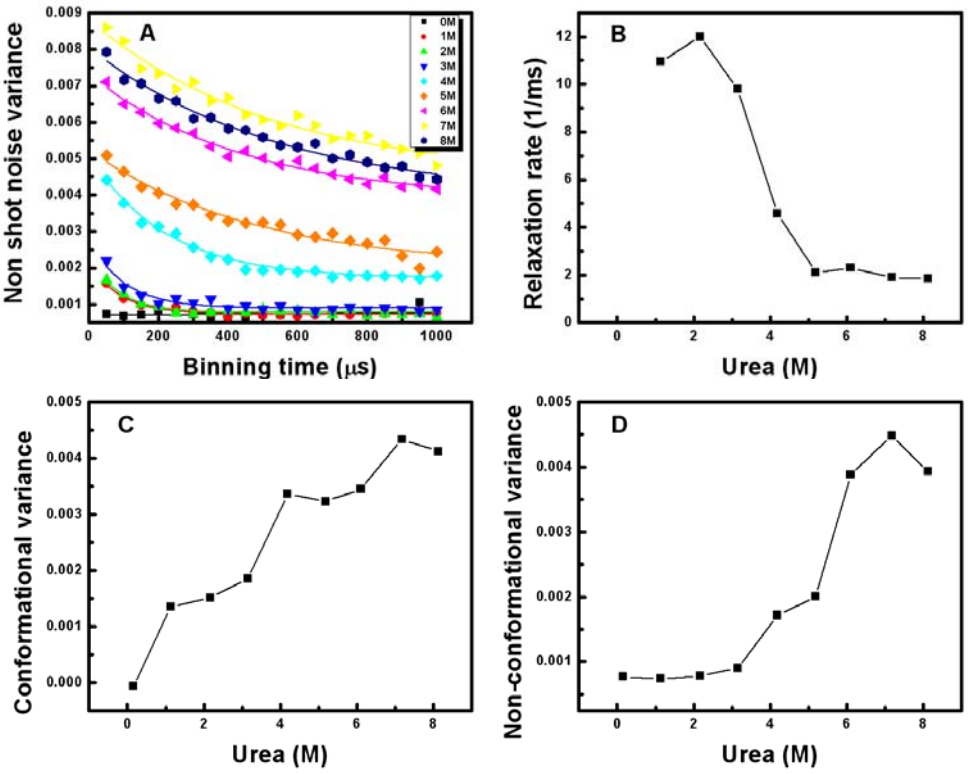

Figure 6.4 $\sigma^2_{non-shot}$ analysis of BBL urea denaturation at 279 K. (A), $\sigma^2_{non-shot}$ decay (the scattering plot) and the exponential fit (solid line). (B), decay rate obtained from the fitting. (C), Conformational variance σ^2_c from the fitting. (D), fitting baseline σ^2_{other}.

Fit of $\sigma^2_{non-shot}$ Decay Using Gopich Formula. Gopich and Szabo derived an equation to describe the dependence of conformational variance on binning time under the Gaussian well approximation of free energy surface [86]:

$$\sigma^2_{conformation}(T_{bin}) = \frac{2\tau\left(\langle E^2\rangle - \langle E\rangle^2\right)}{T_{bin}}\left[1 - \frac{\tau}{T_{bin}}\left(1 - e^{-T_{bin}/\tau}\right)\right] \tag{6.4}$$

where $\left(\langle E^2\rangle - \langle E\rangle^2\right)$ is the variance due to conformational distribution in the free energy well which is equivalent to σ^2_c in the simple exponential fitting function of Eq. (6.3), τ is the relaxation time and T_{bin} is the binning time. In principle this function is

still based on the exponential decay assumption about conformational dynamics, but it considers the conformational distribution and the average of all conformations' dynamics. We add a linear baseline (the baseline slope is assumed to non-negative) term and use it to fit $\sigma^2_{non-shot}$:

$$\sigma^2_{non-shot} = a + bT_{bin} + \frac{2\tau\sigma_c^2}{T_{bin}}\left[1 - \frac{\tau}{T_{bin}}\left(1 - e^{-T_{bin}/\tau}\right)\right] \tag{6.5}$$

Since this function has one more fitting parameter, it fits $\sigma^2_{non-shot}$ decay very well as shown in Figure 6.5.

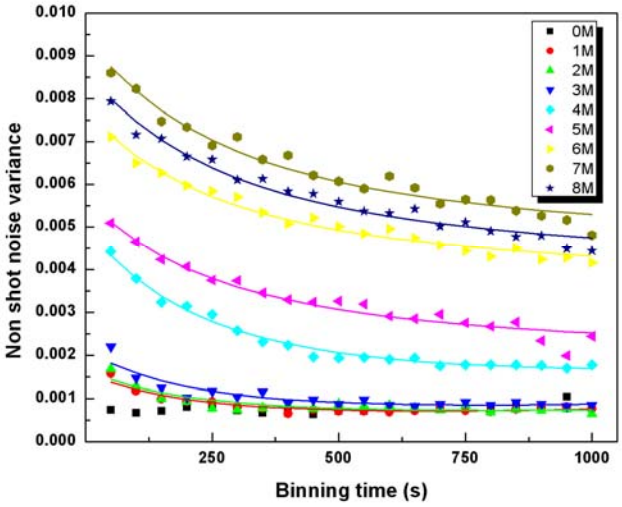

Figure 6.5 $\sigma^2_{non-shot}$ of BBL urea denaturation at 279 K (scattering plots) and the corresponding fit using the Gopich formula (solid lines). The data of 0 M sample is not fitted as no decay is evident.

The dependence of the fitting parameters on urea concentration is similar to what we obtained from the exponential fit as shown in Figure 6.6. The relaxation time is 50~80 μs for 1~3 M samples and it increases to ~150 μs gradually at high urea concentrations. The relaxation time is smaller than the T-jump results, but still

reasonable. The conformational variance increases with urea concentration except that 4 M urea point deviates.

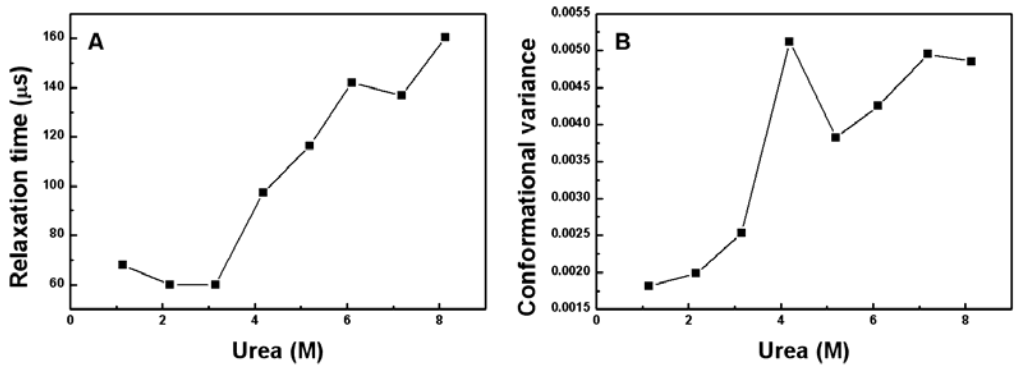

Figure 6.6 $\sigma^2_{non-shot}$ decay fitting results of BBL urea denaturation at 279 K by using Gopich formula: (A) relaxation time and (B) conformational variance.

We also tried several variations to fit the $\sigma^2_{non-shot}$ decay including the exponential function with free-floating baseline and Gopich formula with a constant baseline. All of these fitting functions give similar results with only marginal differences. The basic results are: the relaxation time is around 50~100 μs for 1~3 M samples and increases with urea; σ_c^2 also increases monotonically with urea concentration.

6.4 $\sigma^2_{non-shot}$ Decay Analysis of Other Proteins

Although BBL $\sigma^2_{non-shot}$ decays with binning time in an exponential manner and the decay rate is consistent with the relaxation rate of BBL, it is still not straightforward to claim that $\sigma^2_{non-shot}$ decay reflects protein dynamics. Scattering from the out-of-focus molecules and interstate transitions of acceptor to triplet or

other dark states can also cause $\sigma^2_{non-shot}$ changes. We need to compare the results from more proteins to make a solid argument. Recognizing this, here we analyze the $\sigma^2_{non-shot}$ decay of the fluorescent protein RP-E, α-spectrin SH3 and also BBL at room temperature.

6.4.1 $\sigma^2_{non-shot}$ of RP-E

As introduced in Chapter 2, the fluorescent protein RP-E has intense fluorescence with the maximum emission at 576 nm and can be used as a donor-acceptor FRET system. To be consistent with BBL measurements, the SM-FRET trajectory of RP-E was recorded with a 50 μs binning time at a concentration of 75 pM, but with an excitation power of just 30 μW. The advantages of using RP-E as a control are: (1). No donor-only molecule or zero-peak so the histogram width is not affected by them. Even if we get signal from multiple molecules, they do not add extra variance to the histogram since all molecules are identical. (2). No dynamic processes so $\sigma^2_{non-shot}$ should be zero ideally.

I used thresholds from 30 to 70 to select informational bins from the original 50 μs trajectory. The resulting $\sigma^2_{non-shot}$ is plotted in Figure 6.7. We can find that the magnitude of $\sigma^2_{non-shot}$ is very small. Also $\sigma^2_{non-shot}$ becomes smaller with higher thresholds. The decay amplitude of $\sigma^2_{non-shot}$ between 50 μs and 1ms is very small (<0.0006). With a threshold of 30, $\sigma^2_{non-shot}$ shows a decay trend, but with higher thresholds, $\sigma^2_{non-shot}$ almost does not show any decay. Only threshold 50 has the first point slightly higher than the baseline.

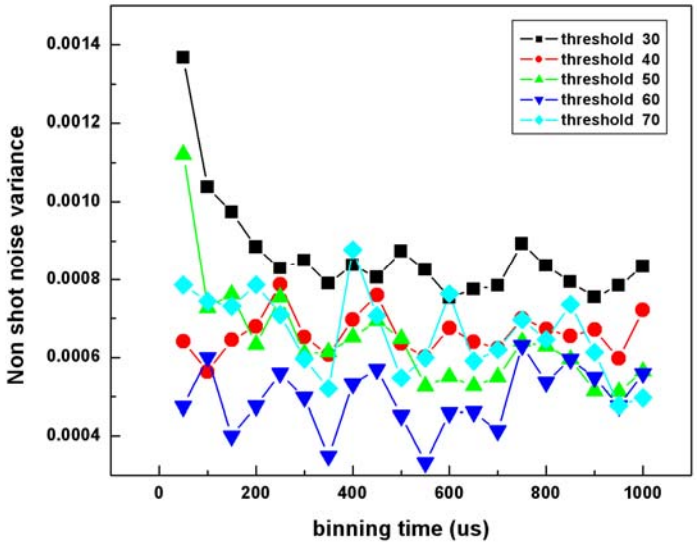

Figure 6.7 RP-E $\sigma^2_{non-shot}$ with different thresholds. The excitation power was 30 μW and the protein concentration was 75 pM.

Origin of RP-E $\sigma^2_{non-shot}$. Based on the above observations, a possible origin for $\sigma^2_{non-shot}$ in RP-E is the excitation beam scattering. With a small threshold the burst intensity is low on average, so the effect of scattering is relatively big. With higher thresholds the burst intensity becomes larger and thus the relative effect of scattering is reduced. Another possible reason is that the formula to calculate shot noise is not adequate. The derivation of FRET efficiency distribution formula for the freely diffusing single molecule measurement is very complicated and therefore some approximations have to be used. So it is not surprising if the shot noise variance formula has some error which might become larger with smaller burst intensity.

From the above analysis, we can conclude that the theory about σ^2_{shot} and $\sigma^2_{non-shot}$ is effective. Any $\sigma^2_{non-shot}$ decay with small amplitude (<0.001) is possibly

due to the excitation beam scattering or the intrinsic error in σ^2_{shot} derivation, and cannot be used as an indicator of protein relaxation.

6.4.2 $\sigma^2_{non-shot}$ of SH3

Since α-spectrin SH3 is a two-state protein, analyses of $\sigma^2_{non-shot}$ decay of both folded and unfolded states will possibly provide insights into the origin of $\sigma^2_{non-shot}$ decay. The $\sigma^2_{non-shot}$ of urea denaturation is presented in Figure 6.8. Informational bins in the original 50 μs trajectory were selected using a threshold of 30.

Native State. As plotted in Figure 6.8A the native state $\sigma^2_{non-shot}$ does not show any decay with binning time. This means that there is no dynamic process in the native state. Also the $\sigma^2_{non-shot}$ is very small for 0~3 M urea samples, but keeps increasing with higher concentrations of urea.

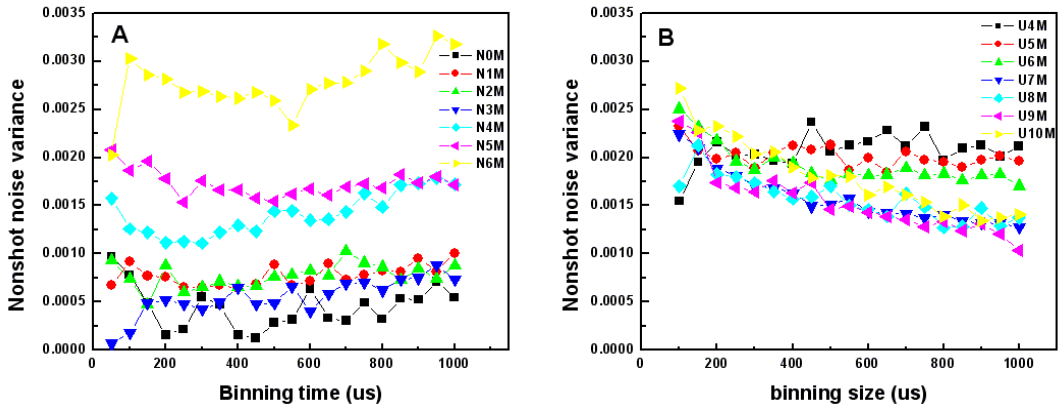

Figure 6.8 $\sigma^2_{non-shot}$ of α-spectrin SH3 denaturation by urea at room temperature. (A), native peak $\sigma^2_{non-shot}$ which is independent of binning time. (B), unfolded peak $\sigma^2_{non-shot}$ which decays with binning time. The 50 μs point is not plotted because it is much higher than the rest of data points.

Unfolded State. The unfolded state $\sigma^2_{non-shot}$ shows a decay trend similar to BBL. However, the magnitude of $\sigma^2_{non-shot}$ for the 50 μs binning time are very high (the data is not shown here), but becomes pretty small from 100 μs on as shown in Figure 6.8B. What could be origin of this phase or is it a dynamic process at all? Since the selected informational bins from the 50 μs trajectory had an average photon count of about 30 (which is not high), the mixing of background noise from donor-only molecules (α-spectrin SH3 had more donor-only proteins than BBL) may have significant effect on the width of the unfolded peak. For longer binning times, the average photon count became higher and the effect of background noise from donor-only molecules was not so serious. Thus the unfolded peak was not broadened as badly as that of the 50 μs trajectory. This is possibly the reason why $\sigma^2_{non-shot}$ drops drastically between 50 and 100 μs binning times.

Origin of α-spectrin SH3 $\sigma^2_{non-shot}$. In α-spectrin SH3, the native state does not show any decay while the unfolded state has a decay albeit with a small amplitude. Their baselines cluster around 0.0005~0.0035. Interestingly, when the protein becomes more unfolded (higher urea concentration), the baseline of native state $\sigma^2_{non-shot}$ becomes higher while the baseline of unfolded state $\sigma^2_{non-shot}$ becomes lower. In other words, when either the native state or unfolded state gets more populated, its $\sigma^2_{non-shot}$ baseline becomes smaller. This strongly supports the hypothesis that out-of-focus scattering from unfolded and native proteins are respobsible for the $\sigma^2_{non-shot}$ baselines of the native and unfolded peaks, respectively.

The inter-state transition of acceptor into triplet and other dark states is the other possible origin of $\sigma^2_{non-shot}$. Although the Trolox-cysteamine protector works efficiently, it is impossible to avoid the occurrence of triplet and other dark states completely. The typical triplet lifetime of a fluorophore is on the order of μs [87] and dark states can have even longer lifetimes [88, 89]. The mean inter-photon time of the selected burst from the original 50 μs binning time trajectory is around 1 μs. So acceptor interstate transitions are possibly to interfere with the photon collection process and hence broaden the histogram (donor inter-state transitions do not affect ET as they just reduce burst intensity). When we merge adjacent bins in order to obtain longer binning time trajectories, the photon counts of the informational bins do not increase proportional to the binning time. So the mean inter-photon time will increase. Actually, the mean inter-photon time of 1 ms trajectory is about thrice that of the 50 μs trajectory. With longer inter-photon time, acceptor inter-state transitions that induce histogram broadening are probably diminished. This in turn can cause the decay of $\sigma^2_{non-shot}$.

However, it is difficult to explain why inter-state transitions do not cause $\sigma^2_{non-shot}$ decay in the native state. Since the native protein is more compact it possibly protects the fluorophore from inter-state transitions better than the unfolded protein.

6.5 Conclusions

Ideally, SM-FRET can be used to probe the dynamics of protein folding or other dynamic process. One way to achieve this is to analyze $\sigma^2_{non-shot}$ decay as a

function of the binning time. If there is no other contribution to $\sigma^2_{non-shot}$ and if the experimental binning time is shorter than the protein relaxation time, we would expect a non-zero $\sigma^2_{non-shot}$ that decays with increasing binning times. By properly fitting the decay curve, protein relaxation time can be determined. From the analysis of BBL 279 K urea denaturation we do find a $\sigma^2_{non-shot}$ decay and the fitted decay time is consistent with BBL relaxation measurement from independent studies.

However, in real measurements, the scattering of out-of-focus molecules and acceptor inter-state transitions are inevitable. The analysis of SH3 $\sigma^2_{non-shot}$ strongly suggests that the scattering of out-of-focus molecules contributes to the non-zero baseline of $\sigma^2_{non-shot}$. Acceptor inter-state transition is the other source of $\sigma^2_{non-shot}$ and probably the source of $\sigma^2_{non-shot}$ decay as well.

In the SM-FRET measurement of BBL urea denaturation conducted at room temperature (the data is not shown here), the 6 M sample shows a clear and slow decay curve of $\sigma^2_{non-shot}$ with binning time. This is not expected from the room temperature measurements because BBL conformational relaxation time is as fast as 20 μs at this temperature. However, since the data from other samples in these measurements were poor the calculated $\sigma^2_{non-shot}$ is very noisy. Therefore, we do not have enough evidence to conclude that at room temperature BBL $\sigma^2_{non-shot}$ still decays and thus $\sigma^2_{non-shot}$ is not caused by protein conformational relaxation.

From the above analysis, we can safely conclude that the decay of $\sigma^2_{non-shot}$ may have information about protein relaxation. But scattering of out-of-focus

molecules and acceptor inter-state transitions make it challenging to convincingly extract quantitative dynamical information from $\sigma^2_{non-shot}$. Further investigation about the background scattering and fluorophore inter-state transitions is needed in order to obtain protein dynamical information from the SM-FRET measurements.

Chapter 7 Summary and Prospective

We proposed to measure the SM-FRET efficiency distribution of BBL to obtain direct evidence about global downhill folding. Unimodal SM-FRET efficiency histogram is expected from the barrierless free energy surface and unimodal conformational distribution of downhill folding. To clarify the effect of protein tails on free energy barrier we designed a new BBL sequence to include the tail residues of QNND-BBL that has been claimed to fold in a two-state fashion by some groups [16, 22].

However, the short relaxation time of BBL posed a big challenge to SM-FRET measurements. As pointed by Szabo and Gopich, if the acquisition time of single molecule measurements is longer than the protein relaxation time, even a two-state protein can produce unimodal SM-FRET efficiency histograms [54]. Laser T-jump spectroscopy measurements estimated the relaxation time of BBL in water to be 20 μs at 298 K and 150 μs at 279 K, respectively [46]. So we carried out the single molecule experiments at 279 K using a binning time of 50 μs that is one-third of the folding relaxation time under these conditions.

Improvement on SM-FRET Method. In order to maintain the sample at 279 K we built a peltier-cooled cold plate with an electronic board to control the temperature. The typical binning times in SM-FRET studies are several hundred μs. The need to use a 50 μs binning time in our measurements required an increase in photon count of almost ten times than that of the usual SM-FRET measurement. We first increased the photon count by discarding the pinhole aperture in the confocal

microscope system and using the detector window as a spatial filter. A transform lens was used to double the beam diameter and then the majority of out-of-focus scattering was blocked by the detector window. We were also able to increase the photon count by ramping up the excitation power.

However, a large excitation power can result in bad photoblinking and photobleaching of fluorophores which ruins the single molecule measurement. We found that the combination of Trolox and cysteamine can protect the fluorophore efficiently for excitation powers less than 210 μW. We determined the optimal concentrations to be 1 mM Trolox and 10 mM cysteamine.

We further improved signals by performing the experiments of BBL at pH 6.0 acetate buffer which yielded more photons with reduced protein aggregation compared to pH 7.0. Control experiment showed that such modest pH change did not affect the ET distribution.

SM-FRET Measurement on BBL and α-spectrin SH3. With our improved confocal microscope system and the fluorophore protector we performed the SM-FRET measurements on BBL at 279 K. In the unfolding of BBL. We observed clear unimodal SM-FRET efficiency histograms on BBL upon urea and GuHCl denaturations. The protein peak shifted gradually from higher to lower ET values with increasing denaturant concentrations as expected of global downhill folders. We also carried out the measurements using a binning time of 20 μs and still obtained unimodal SM-FRET efficiency histograms although the data quality was poor due to the photophysics processes of fluorophores.

As suggested by Gopich and Szabo, even a two-state protein can result in a unimodal ET distribution due to the multiple molecule effects if a detector collects signals from multiple molecules when protein concentration is too high or the out-of-focus scattering is too strong [72]. So we performed SM-FRET measurements on a previously characterized two-state protein α-spectrin SH3 under the same conditions except that it was at room temperature. The bimodal ET distribution of α-spectrin SH3 demonstrated that our instrument worked fine and that the unimodal conformational distribution of BBL was not an artifact due to instrumental anomalies.

Ensemble Control Experiments. Apart from the single molecule measurements on α-spectrin SH3, ensemble measurements were performed to exclude the possibility of any labeling-related artifact. In the case of BBL GuHCl unfolding, the ensemble ET profile of doubly-labeled BBL and CD profile of unlabeled BBL can be superimposed very well. Using a two-state model, the thermodynamic parameters from the global fitting of GuHCl and urea ET unfolding curves are identical within errors to the corresponding values obtained from GuHCl CD unfolding curve alone.

The global fitting of α-spectrin SH3 urea and GuHCl ET unfolding curves also produced similar thermodynamic parameters as those obtained from the wild-type α-spectrin SH3. All these experiments on BBL and α-spectrin SH3 therefore confirmed that fluorophores do not affect protein stability. The relative strength of GuHCl to urea denaturation was also similar for BBL and α-spectrin SH3.

Simulation of SM-FRET Efficiency Histogram. We simulated the SM-FRET efficiency histogram of BBL by assuming it were a two-state folder using

experimental parameters and a shorter relaxation time of 120 μs. The simulated histogram had clear bimodal distribution around the denaturation midpoint. Another simulation with a strict and unrealistic relaxation time assumption (20 μs at 279 K) showed a unimodal distribution. However, the simulation predicted very large $\sigma^2_{non-shot}$ for conditions around the denaturation midpoint while the $\sigma^2_{non-shot}$ of the experimental ET distribution was significantly smaller. Taking into account the histogram broadening due to background scattering and unfolded state dynamics (which were not considered in the simulation), the experimental $\sigma^2_{non-shot}$ must be larger than the simulated $\sigma^2_{non-shot}$ if the two-state assumption about BBL is correct. So we conclude that BBL is still a downhill folder even if its relaxation time was 20 μs at 279 K.

Dynamic Analysis of SM-FRET Efficiency Histogram. From the analysis of the SM-FRET efficiency histogram width we tried to extract further information about protein dynamics. We found that the $\sigma^2_{non-shot}$ of BBL exponentially decayed with binning time at non-zero urea concentrations. The decay time from this analysis was calculated to be around 100 μs that is consistent with the relaxation time of BBL from laser T-jump studies. Interestingly, the BBL sample without urea and the native state of α-spectrin SH3 did not show any decay while the unfolded state of α-spectrin SH3 showed a similar decay. Such decays might be protein relaxation processes within individual free energy well, but can also be caused by fluorophore photophysics or background scattering. Therefore the origin of $\sigma^2_{non-shot}$ decay is not clear.

Prospective for Future Investigation. Apart what we have achieved from SM-FRET measurements on BBL and α-spectrin SH3, there is clear room for improvement in both the technique and analysis to extract more information in the near future.

Higher Time Resolution. It is critical to reduce the binning time. If we can approach a time resolution in the order of μs, SM-FRET will be able to follow the dynamic processes of even the fastest folding proteins. A shorter binning time is also helpful in histogram width analysis. A typical single photon detector's dead time is around 50 ns that makes it possible to collect about 25 photons in several μs.

In SM-FRET experiments, most of the emitted photons from fluorophore are not detected. So there is a large room to improve the photon collection efficiency. In our measurement the focus was 20 μm deep into the solution. Actually, we can get a higher signal with the focus being closer to the surface. But this increases the background noise. One possible direction is to find ways to reduce the background noise using a better optical setup thus enabling a focus closer to the surface.

Moreover, more efficient fluorophore protectors are needed so as to increase the excitation power. To start with, the compound propyl gallate can be tried. It is probably more efficient than Trolox because it is more soluble in water [57].

Characterization of Fluorophore Photophysics. The typical lifetime of triplet is in the order of μs and other dark states have even longer lifetimes. In order to minimize the influence of fluorophore inter-state transitions on the single molecule measurement, the photophysics of fluorophore should be understood more thoroughly.

SM-FRET Efficiency Histogram Width Analysis. Extract relaxation times from the analysis of SM-FRET efficiency histogram widths under varying binning times (Chapter 6) is a promising development. Ideally, for a two-state protein, the measured relaxation rates as a function of denaturant should have a "V-shape" (Chevron plot). An important issue here is that we need to exclude the effect of fluorophore inter-state transitions. A good control experiment is to perform SM-FRET measurements on some very rigid spacer molecules labeled with the same dye-pair. If this rigid molecule does not show a $\sigma^2_{non-shot}$ decay it is a clear evidence that $\sigma^2_{non-shot}$ decay represents protein dynamics. Also, the binning time should be shorter than 50 μs. In this way it is possible to obtain even chevron plots from single molecule experiments.

Application of Rigid Spacer Molecules. A rigid spacer molecule has very broad applications in the field of SM-FRET spectroscopy. It can also used to compare the measured ET and theoretical value calculated from Förster formula, thus checking the validity of Förster formula and the instrument calibration.

Protein Immobilization and Tracking. The freely diffusing SM-FRET technique provides many advantages, the obvious one being that it does not affect the protein. However, the diffusion of proteins complicates the measurement and analysis. Firstly, the selection of informational bins from the single molecule trajectory accounts for a lot of error. It will be easier and more accurate to identify immobilized molecules from background noise. Moreover, the free diffusion makes photon statistics theory of SM-FRET very complicated. It is much easier to simulate

the immobilized molecules. The typical surface immobilization may affect the protein dynamics. The encapsulation of proteins in vesicle probably works better [35].

The ideal way to overcome the limitation imposed by the molecular diffusion is to track the molecule when it diffuses in solution. To do that we must use an image detector such as charge coupled device (CCD) rather than the photon detector. Currently, advanced CCD cameras can capture images at a rate of 1 frame per μs which is fast enough to track protein diffusion. The challenge is to determine protein movement and then control instruments to follow the protein. If it is realized, we would be able to easily identify ONE molecule from the background noise and make measurements on it for an extended period. Obviously we can get more information about protein folding using the molecular tracking technique than the current SM-FRET method.

Bibliography

[1] C. B. Anfinsen, Science **181**, 223 (1973).

[2] P. L. Privalov, Advances in Protein Chemistry **33**, 167 (1979).

[3] O. B. Ptitsyn and V. N. Uversky, Febs Lett **341**, 15 (1994).

[4] A. Akmal and V. Munoz, Proteins **57**, 142 (2004).

[5] V. S. Pande, A. Y. Grosberg, and T. Tanaka, Rev. Mod. Phys. **72**, 259 (2000).

[6] A. Y. Grosberg and A. R. Khokhlov, *Statistical physics of macromolecules* (American Institute of Physics, New York, 1994).

[7] C. Levinthal, in *Proceedings of a meeting held at Allerton House* (University of Illinois Press, Allerton House, Monticello, Illinois, 1969), p. 22.

[8] J. D. Bryngelson and P. G. Wolynes, P Natl Acad Sci USA **84**, 7524 (1987).

[9] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Proteins-Structure Function and Genetics **21**, 167 (1995).

[10] P. W. Anderson, Journal of the Less Common Metals **62** 291 (1978).

[11] P. Hanggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990).

[12] H. A. Kramers, Physica (Utrecht) **7**, 284 (1940).

[13] S. Jackson, I. M. Leigh, and A. Storey, J Invest Dermatol **110**, 529 (1998).

[14] M. M. Garcia-Mira, M. Sadqi, N. Fischer, J. M. Sanchez-Ruiz, and V. Munoz, Science **298**, 2191 (2002).

[15] V. Munoz and J. M. Sanchez-Ruiz, Proc. Natl. Acad. Sci. U. S. A. **101**, 17646 (2004).

[16] N. Ferguson, P. J. Schartau, T. D. Sharpe, S. Sato, and A. R. Fersht, J. Mol. Biol. **344**, 295 (2004).

[17] A. N. Naganathan, R. Perez-Jimenez, J. M. Sanchez-Ruiz, and V. Munoz, Biochemistry **44**, 7435 (2005).

[18] N. Ferguson, T. D. Sharpe, P. J. Schartau, S. Sato, M. D. Allen, C. M. Johnson, T. J. Rutherford, and A. R. Fersht, J. Mol. Biol. **353**, 427 (2005).

[19] M. Sadqi, D. Fushman, and V. Munoz, Nature **442**, 317 (2006).

[20] G. H. Zuo, J. Wang, and W. Wang, Proteins **63**, 165 (2006).

[21] M. Knott and H. S. Chan, Proteins **65**, 373 (2006).

[22] S. S. Cho, P. Weinkam, and P. G. Wolynes, Proc. Natl. Acad. Sci. U. S. A. **105**, 118 (2008).

[23] W. Yu, K. Chung, M. Cheon, M. Heo, K. H. Han, S. Ham, and I. Chang, Proc. Natl. Acad. Sci. U. S. A. **105**, 2397 (2008).

[24] J. Zhang, W. Li, J. Wang, M. Qin, and W. Wang, Proteins **72**, 1038 (2008).

[25] J. W. Pitera, W. C. Swope, and F. F. Abraham, Biophys. J. **94**, 4837 (2008).

[26] N. Ferguson, T. D. Sharpe, C. M. Johnson, P. J. Schartau, and A. R. Fersht, Nature **445**, E14 (2007).

[27] W. E. Moerner and L. Kador, Phys Rev Lett **62**, 2535 (1989).

[28] M. Orrit and J. Bernard, Phys Rev Lett **65**, 2716 (1990).

[29] W. E. Moerner and D. P. Fromm, Rev. Sci. Instrum. **74**, 3597 (2003).

[30] T. J. Ha, A. Y. Ting, J. Liang, W. B. Caldwell, A. A. Deniz, D. S. Chemla, P. G. Schultz, and S. Weiss, P Natl Acad Sci USA **96**, 893 (1999).

[31] Y. W. Jia, D. S. Talaga, W. L. Lau, H. S. M. Lu, W. F. DeGrado, and R. M. Hochstrasser, Chem. Phys. **247**, 69 (1999).

[32] A. A. Deniz, T. A. Laurence, G. S. Beligere, M. Dahan, A. B. Martin, D. S. Chemla, P. E. Dawson, P. G. Schultz, and S. Weiss, P Natl Acad Sci USA **97**, 5179 (2000).

[33] B. Schuler, E. A. Lipman, and W. A. Eaton, Nature **419**, 743 (2002).

[34] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy* (Plenum Pub Corp, 1999).

[35] E. Rhoades, E. Gussakovsky, and G. Haran, P Natl Acad Sci USA **100**, 3197 (2003).

[36] Y. Jia, D. S. Talaga, L. Wei, H. S. M. Lu, W. F. DeGrado, and R. M. Hochstrasser, Biophys J **76**, A10 (1999).

[37] D. Nettels, I. V. Gopich, A. Hoffmann, and B. Schuler, P Natl Acad Sci USA **104**, 2655 (2007).

[38] B. Schuler, E. A. Lipman, P. J. Steinbach, M. Kumke, and W. A. Eaton, P Natl Acad Sci USA **102**, 2754 (2005).

[39] E. Sherman and G. Haran, P Natl Acad Sci USA **103**, 11539 (2006).

[40] R. B. Best, K. A. Merchant, I. V. Gopich, B. Schuler, A. Bax, and W. A. Eaton, P Natl Acad Sci USA **104**, 18964 (2007).

[41] A. N. Kapanidis, N. K. Lee, T. A. Laurence, S. Doose, E. Margeat, and S. Weiss, P Natl Acad Sci USA **101**, 8936 (2004).

[42] X. Michalet, S. Weiss, and M. Jager, Chem Rev **106**, 1785 (2006).

[43] A. Hoffmann, et al., P Natl Acad Sci USA **104**, 105 (2007).

[44] K. A. Merchant, R. B. Best, J. M. Louis, I. V. Gopich, and W. A. Eaton, P Natl Acad Sci USA **104**, 1528 (2007).

[45] E. Sherman, A. Itkin, Y. Y. Kuttner, E. Rhoades, D. Amir, E. Haas, and G. Haran, Biophys J **94**, 4819 (2008).

[46] P. Li, F. Y. Oliva, and V. Munoz, Submitted (2008).

[47] M. A. Morgan, K. Okamoto, D. S. English, and J. D. Kahn, Biophys J **88**, 382A (2005).

[48] M. A. Morgan, K. Okamoto, J. D. Kahn, and D. S. English, Biophys J **89**, 2588 (2005).

[49] T. Wilson and A. R. Carlini, Opt. Lett. **12**, 227 (1987).

[50] D. R. Sandison, D. W. Piston, R. M. Williams, and W. W. Webb, Appl. Optics **34**, 3576 (1995).

[51] B. Schuler and W. A. Eaton, Curr Opin Struc Biol **18**, 16 (2008).

[52] PerkinElmer.

[53] A. T. R. Williams, S. A. Winfield, and J. N. Miller, Analyst **108**, 1067 (1983).

[54] I. V. Gopich and A. Szabo, J Phys Chem B **111**, 12925 (2007).

[55] I. Rasnik, S. A. McKinney, and T. Ha, Nat. Methods **3**, 891 (2006).

[56] X. W. Zhuang, H. Kim, M. J. B. Pereira, H. P. Babcock, N. G. Walter, and S. Chu, Science **296**, 1473 (2002).

[57] J. R. Grunwell, J. L. Glass, T. D. Lacoste, A. A. Deniz, D. S. Chemla, and P. G. Schultz, J. Am. Chem. Soc. **123**, 4295 (2001).

[58] M. Heilemann, E. Margeat, R. Kasper, M. Sauer, and P. Tinnefeld, J. Am. Chem. Soc. **127**, 3801 (2005).

[59] L. C. Packman and R. N. Perham, FEBS Lett. **206**, 193 (1986).

[60] J. D. Muller, Y. Chen, Q. Q. Ruan, W. W. Mantulin, and E. Gratton, Biophys J **76**, A359 (1999).

[61] Y. Chen, J. D. Muller, P. T. C. So, and E. Gratton, Biophys J **77**, 553 (1999).

[62] I. V. Gopich and A. Szabo, J Phys Chem B **109**, 17683 (2005).

[63] I. Gopich, G. Hummer, and A. Szabo, Abstr Pap Am Chem S **226**, U287 (2003).

[64] I. Gopich and A. Szabo, Abstr Pap Am Chem S **227**, U317 (2004).

[65] I. Gopich and A. Szabo, J Phys Chem B **109**, 6845 (2005).

[66] I. Gopich and A. Szabo, J Chem Phys **122** (2005).

[67] I. V. Gopich and A. Szabo, J Chem Phys **118**, 454 (2003).

[68] I. V. Gopich and A. Szabo, J Chem Phys **124** (2006).

[69] J. W. Goodman, IEEE J. Quantum Electron. **QE 1**, 180 (1965).

[70] G. D. Harp and B. J. Berne, Physical Review A **2**, 975 (1970).

[71] H. S. Chan and K. A. Dill, Proteins **30**, 2 (1998).

[72] I. V. Gopich, J Phys Chem B **112**, 6214 (2008).

[73] A. Musacchio, M. Noble, R. Pauptit, R. Wierenga, and M. Saraste, Nature **359**, 851 (1992).

[74] A. Musacchio, T. Gibson, V. P. Lehto, and M. Saraste, Febs Lett **307**, 55 (1992).

[75] B. J. Mayer, M. Hamaguchi, and H. Hanafusa, Nature **332**, 272 (1988).

[76] C. J. Morton and I. D. Campbell, Curr. Biol. **4**, 615 (1994).

[77] S. B. Feng, J. K. Chen, H. T. Yu, J. A. Simon, and S. L. Schreiber, Science **266**, 1241 (1994).

[78] H. T. Yu, J. K. Chen, S. B. Feng, D. C. Dalgarno, A. W. Brauer, and S. L. Schreiber, Cell **76**, 933 (1994).

[79] J. C. Martinez, A. R. Viguera, R. Berisio, M. Wilmanns, P. L. Mateo, V. V. Filimonov, and L. Serrano, Biochemistry-Us **38**, 549 (1999).

[80] A. R. Viguera, J. C. Martinez, V. V. Filimonov, P. L. Mateo, and L. Serrano, Biochemistry-Us **33**, 2142 (1994).

[81] R. F. Greene and C. N. Pace, J. Biol. Chem. **249**, 5388 (1974).

[82] O. D. Monera, C. M. Kay, and R. S. Hodges, Protein Sci. **3**, 1984 (1994).

[83] E. P. O'Brien, R. I. Dima, B. Brooks, and D. Thirumalai, J Am Chem Soc **129**, 7346 (2007).

[84] E. Deu and J. F. Kirsch, Biochemistry-Us **46**, 5810 (2007).

[85] L. M. Mayr and F. X. Schmid, Biochemistry-Us **32**, 7994 (1993).

[86] I. V. Gopich and A. Szabo, J Phys Chem B **107**, 5058 (2003).

[87] J. Bernard, L. Fleury, H. Talon, and M. Orrit, J Chem Phys **98**, 850 (1993).

[88] R. Zondervan, Universiteit Leiden, 2006.

[89] R. Zondervan, F. Kulzer, S. B. Orlinskii, and M. Orrit, J. Phys. Chem. A **107**, 6770 (2003).