Springer Theses
Recognizing Outstanding Ph.D. Research

Sharona Cohen

Single-Molecule Fluorescence Spectroscopy of the Folding of a Repeat Protein

Springer

Springer Theses

Recognizing Outstanding Ph.D. Research

Aims and Scope

The series "Springer Theses" brings together a selection of the very best Ph.D. theses from around the world and across the physical sciences. Nominated and endorsed by two recognized specialists, each published volume has been selected for its scientific excellence and the high impact of its contents for the pertinent field of research. For greater accessibility to non-specialists, the published versions include an extended introduction, as well as a foreword by the student's supervisor explaining the special relevance of the work for the field. As a whole, the series will provide a valuable resource both for newcomers to the research fields described, and for other scientists seeking detailed background information on special questions. Finally, it provides an accredited documentation of the valuable contributions made by today's younger generation of scientists.

Theses are accepted into the series by invited nomination only and must fulfill all of the following criteria

- They must be written in good English.
- The topic should fall within the confines of Chemistry, Physics, Earth Sciences, Engineering and related interdisciplinary fields such as Materials, Nanoscience, Chemical Engineering, Complex Systems and Biophysics.
- The work reported in the thesis must represent a significant scientific advance.
- If the thesis includes previously published material, permission to reproduce this must be gained from the respective copyright holder.
- They must have been examined and passed during the 12 months prior to nomination.
- Each thesis should include a foreword by the supervisor outlining the significance of its content.
- The theses should have a clearly defined structure including an introduction accessible to scientists not expert in that particular field.

More information about this series at http://www.springer.com/series/8790

Sharona Cohen

Single-Molecule Fluorescence Spectroscopy of the Folding of a Repeat Protein

Doctoral Thesis accepted by
Weizmann Institute of Science,
Rehovot, Israel

Springer

Author
Dr. Sharona Cohen
Department of Chemical Physics
Weizmann Institute of Science
Rehovot
Israel

Supervisor
Prof. Gilad Haran
Department of Chemical Physics
Weizmann Institute of Science
Rehovot
Israel

ISSN 2190-5053 ISSN 2190-5061 (electronic)
ISBN 978-3-319-09557-8 ISBN 978-3-319-09558-5 (eBook)
DOI 10.1007/978-3-319-09558-5

Library of Congress Control Number: 2014946189

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Parts of this thesis have been published in the following documents:

Sharona S. Cohen, Inbal Riven, Aitziber L. Cortajarena, Lucia De Rosa, Luca D. D'Andrea, Lynne Regan, Gilad Haran, *Probing the molecular origin of native-state flexibility in repeat proteins*. J Am Chem Soc, Volume 137, 2015, Pages 10367–10373.

Supervisor's Foreword

The protein folding field has undergone enormous transformation in recent years, led mainly by the introduction of new techniques that facilitate the study of folding on the single-molecule level, based on either sensitive fluorescence methodology or force spectroscopy. The progress in the field has also been driven by the realization that, in addition to the more familiar globular proteins, there are novel groups of proteins whose folding behavior requires the development of new paradigms. Among these proteins are two important groups. The first is the group of intrinsically unfolded proteins, many of which fold only upon interaction with their binding partners. The second group includes the repeat proteins, a diverse set of proteins made of short structural units that repeat themselves and form nonglobular, elongated structures resembling solenoids.

The Ph.D. thesis of Dr. Sharona Cohen marries the two themes discussed above. Dr. Cohen employed single-molecule fluorescence spectroscopy to probe the folding and dynamics of tetratricopeptide repeat proteins, whose repeat units consist of 34 amino acids. In particular, she studied a designed protein made of three repeats, with fluorescent probes inserted in several positions. Fluorescence resonance energy transfer (FRET) was used to learn how the conformation of the protein changes in the presence of chemical denaturants.

The big surprise in Dr. Cohen's results is that in addition to the unfolding transition of the protein, clearly observed in single-molecule FRET histograms, there is interesting behavior related to changes in the structure of the folded state. As it turns out, the folded protein becomes more and more loose as the concentration of denaturant is increased, like a solenoid whose spring constant is weakened gradually. This finding immediately suggests that the source of the spring behavior of this protein is to be found in the interfaces between repeat units, which are mostly hydrophobic in nature, and can be loosened by the denaturant. This explanation is commensurate with molecular dynamics simulations conducted by others on similar proteins.

Multiple questions are raised by this study such as: What are the timescales of the dynamics related to the spring behavior of these repeat proteins? What is

the functional role of this behavior? Single-molecule techniques, in combination with computer simulations, will no doubt answer some of these questions in the not-so-far future. The beautiful set of experiments documented in this thesis lays the foundation for these future studies.

Rehovot, June 2014 Prof. Gilad Haran

Acknowledgments

I express sincere gratitude to my mentor, Prof. Gilad Haran, for his support and generous assistance throughout my research program. I also thank my fellow lab members, Menahem Pirchi, Inbal Riven, Haim Aviram, Mila Goomanovsky, Nir Zohar, Yunmin Jung, and Rita August for their help. I thank Ellen Merlo for financial support. I am indebted to the Weizmann Institute of Science for giving me the opportunity and for providing me a superb environment to research my field of interest. For their never ending support and dedication, I thank my dear parents, Mehrdad Sedghani and Diana Soleimani. Last but not least, I am grateful to my husband Shlomo Cohen for love, patience, understanding, and encouragement.

Contents

Abbreviations

ADP	Avalanche Photo Diodes
ANK	Ankyrin
CD	Circular Dichroism
CTPR	Consensus TPR
CTPR3	CTPR with three repeats
DIC	Dichroic mirror
FCS	Fluorescent correlation spectroscopy
FRET	Förster Resonance Energy Transfer
GuHCl	Guanidine hydrochloride
Heat	Huntington, Elongation factor 3, protein phosphatase 2A, and Yeast Kinase TOR1
HSQC	Heteronuclear Single Quantum Coherence
HX	Hydrogen Exchange
MS	Motorized stage
NMR	Nuclear magnetic resonance
PDB	Protein Data Bank
PH	Pinhole
PMT	Photomultiplier tube
SM-FRET	Single-Molecule FRET
TPR	Tetratrico Peptide Repeat

Chapter 1
Introduction

Repeat proteins account for more than 5 % of the proteins in multi-cellular organisms. In contrast to globular proteins that have extensive long-range contacts, the structures of repeat proteins show mainly intra- and inter-repeat local contacts. These simple topologies facilitate modeling of the folding mechanism. Tetratricopeptide repeat (TPR) is a 34-amino acid helix-turn-helix motif found in tandem arrays in many natural proteins [1].

The design and characterization of a series of consensus TPR (CTPR) proteins, built as arrays of multiple tandem copies of a consensus sequence, enable further simplification of the folding model. In a series of experiments we used Single Molecule Förster Resonance Energy Transfer (SM-FRET) to study the folding of a CTPR protein with three repeats.

In our experiment, single molecules of double-labeled CTPR were allowed to diffuse through the focus of a confocal microscope. These measurements, which were carried out for two FRET pairs, map the folding behavior of two and three repeats. Measured photon bursts were used to obtain FRET efficiency histograms in different chemical denaturant concentrations. For both FRET pairs, a continuous shift of FRET efficiency histograms is found at mild denaturant concentrations, followed by an all-or-none transition to the unfolded state. Global fitting of histograms showed that the continuous shift is not a result of an interchange between two stable populations with similar FRET efficiencies. Statistical analysis of single-molecule trajectories obtained from measurements of immobilized molecules prove the absence of a slow transition between possible FRET populations. Correlation analysis was used to observe the dynamics of FRET efficiency. These measurements indicate that the conformational dynamics is faster than a microsecond timescale. We conclude that the initial continuous shift of FRET efficiency is caused by a continuous expansion of the structure.

Combining the result of nuclear magnetic resonance (NMR) and SM-FRET measurements we determine that during global expansion the secondary structure is preserved and the tertiary structure slightly expanded. This expansion is followed by a barrier crossing event which involves unfolding of the secondary structure. Modification of a previously proposed Ising model for the folding of CTPR proteins is suggested.

S. Cohen, *Single-Molecule Fluorescence Spectroscopy of the Folding of a Repeat Protein*, Springer Theses, DOI 10.1007/978-3-319-09558-5_1

Our model is in agreement with previous studies of the flexibility of tertiary structure in repeat proteins, suggesting that this flexibility plays a role in binding to other proteins or ligands [2].

1.1 Repeat Proteins

Repeat proteins, whether as independent folds or as segments within larger proteins, are composed of tandem arrays of a small structural unit. The non-globular folds of repeat proteins are formed from stacks of repeated motifs of 20–40 amino acids, often producing extended structures. Repeat proteins are widely distributed in various biological species [3], and are involved in a myriad of essential biological processes. They account for more than 5 % of the proteins in the multi-cellular organism category of the Swiss-Prot database. Ankyrin (ANK) repeats, tetratricopeptide (TPR), leucine-rich repeats, Heat (Huntington, Elongation factor 3, protein phosphatase 2A, and Yeast Kinase TOR1), beta helical repeats, and armadillo repeats are common examples of repeat proteins.

In contrast to globular proteins with complicated and irregular tertiary structures, the tertiary structures of repeat proteins are simple and modular. The degree of the simplicity of this topology is usually determined by the contact order, which is the average sequence distance between all pairs of contacting residues normalized by the total sequence length. Typical contact order values for repeat proteins range from 0.05 to 0.07 (equal to sequence separations of about 6–7 residues), whereas the contact order values of globular proteins range from 0.1 to 0.2 (equal to sequence separations of about 10–30 residues) [4]. Owing to the modularity and simplicity of their structure, repeat proteins have received increasing attention by scientists in various fields such as protein engineering, protein design, protein folding, and molecular dynamics simulations.

The non-globular, elongated shapes of repeat proteins result in larger surfaces for molecular recognition and interactions. Recognition of different binding proteins, based on their sequence and structure, is the most common function of repeat proteins.

In repeat proteins, amino acid residues that are important for binding are distinct and different from those that are important for maintaining the stability of the fold; therefore, consensus designed proteins with fixed framework positions and variable binding positions can be designed to construct novel binding molecules. Additionally, the linearity in the structure of repeat proteins suggests that repeat protein modules, based on the consensus sequence, may be added, allowing extended structures without altering the overall topology [5]. Design of repeats in the laboratory, based on the consensus sequences of individual repeats [6], is therefore possible. Consensus TPR (CTPR) [7], ANK [8, 9], leucine-rich [10], beta-propeller [11], and armadillo repeats [12] with the same structure as the natural repeat have been successfully designed.

TPRs [1, 13] are one of the common structural motifs, and are found in tandem arrays of 2–16 motifs of 34 amino acid residues. The three-dimensional structure

of a single TPR is a helix–turn–helix. Adjacent TPR units are stacked in parallel and form a right-handed superhelix. More than 5,000 TPR-containing proteins have been identified in different organisms. For more than 100 of them, their structures have been solved and deposited in the Protein Data Bank (PDB). Similar to other repeat proteins, the large surface area presented by their super helix elongated structures [1] makes them particularly well suited to mediate interactions with other proteins [14–16].

TPRs participate in a diverse set of processes: they are involved in peroxisomal targeting [17], in co-chaperones interacting with heat shock proteins 70 and 90 [18], and in synaptic vesicle fusion [19]. In addition, TPRs interact with transcriptional repressor proteins [20] and are associated with bacterial pathogenesis [21, 22], protein sorting, and translocation to the outer membrane of chloroplasts and mitochondria [23, 24], and more. They are also found in magnetosome-associated proteins in magnetotactic bacteria [25] and in the outer membrane assembly of alphaproteobacteria [26]. Mutations in TPR-containing proteins have been associated with a variety of human diseases [13].

The engineered structure of CTPRs [27, 28], with high stability and desired binding specificity, make them unique and useful for diverse applications. Recent studies have used CTPRs for generating anti-cancer agents [29] and functional biofilms [30].

1.2 Folding Stability of Repeat Proteins

The effect of insertion of repeat modules on the stability and folding kinetics for two series of repeat proteins, namely, the CTPR and ANK series [31, 32], was previously studied. In both series the insertion of repeat modules stabilizes the protein and reduces the unfolding rate.

In general, individual repeats are intrinsically unstable; the addition of a repeat with favorable interface interactions between repeats stabilizes the fold. Since there is no long-range interaction in repeat proteins, assuming that nonadjacent repeats are independent of one another, to a good approximation, the stabilities of a construct can be described as the sum of energy terms associated with each repeat [5]. One-dimensional Ising models have been used with great success to explain the unfolding behavior of repeat proteins at equilibrium.

The Ising model is a statistical model, conceived in 1920, to describe the magnetization of a series of spins with only nearest neighbor interactions between them. In this model each spin is an independent unit that can be either up or down. The free energy of the system is a linear summation of energy associated with each unit [33]. The one-dimensional Ising model has been used for more than half a century to describe the helix-coil transition in the formation of α helices. In this model each residue is considered to be ordered in a helical state, or be in a coil state [34]. Local contact between adjacent residues results in cooperative folding behavior.

An extended version of the Ising model has been successfully applied to repeat proteins. Assuming that each repeat module is an independent unit and can be folded or unfolded, the one-dimensional Ising model can be used to explain the free energy of the ANK domain of Notch [5], and consensus designed ANK and TPR proteins [31, 35]. Consensus-designed identical interactions between different units reduce the parameters of the Ising model and only a few parameters, describing the stability of each unit and the interaction between adjacent units, were used to globally explain the thermodynamic behavior of a series with a different number of repeats.

Using this model, it was shown that the global stability of a CTPR protein can be changed by modifying the stability of a unit, without affecting the coupling interaction between units, in a predictable fashion [36]. Modulating the interactions between repeats, however, reduces the population of intermediate states for CTPR [37] and induces a two-state like folding mechanism.

In the case of ANK proteins folding was shown to be a two-state process [38, 39]. Nevertheless stability of ANK proteins can be explained by using a one-dimensional Ising model [5]. A molecular dynamics simulation, based on native topology, was used to explain the difference in folding between the TPR and ANK proteins [40, 41]. These simulations demonstrate the sensitivity of the folding landscape to the topology and demonstrate the inter-domain interaction. Modifying the energy strength of the interface was shown to affect the protein stability, folding rate, and the folding pathways [41]. Therefore, it was proposed that the difference in the interaction between units is the cause of the dissimilar folding mechanisms of TPR and ANK.

The Ising model predicts the existence of partially folded stable species with one or more units unfolded. Owing to their coexistence with fully folded and fully unfolded species, in ensemble experiments, these species are quite difficult to identify. Single-molecule microscopy provides a means by which we can analyze such intermediates. Recent single-molecule measurements of ANK unfolding under force by atomic force microscopy show that the folding of ANK proteins is more complicated than was originally thought. Heterogeneous unfolding steps were observed during the unfolding pathways. These steps are not necessarily equal to one or even half of the repeat unfolding, as was suggested by the Ising model. Folding occurs in multiple distinct pathways and with several intermediates [42]. In contrast to ANK, the folding of TPR proteins has not been studied on the single molecule level. Here we used single-molecule fluorescence spectroscopy to study the folding of CTPR.

1.3 Protein Folding

Historically, two main simple mechanisms were suggested to explain protein folding: (1) the diffusion-collision model, in which microdomains are folded by a random event initiating the formation of the secondary structure, and these secondary

structures fluctuate until they collide and pack tightly together to form the native structure; and (2) the nucleation-condensation model, in which the secondary and tertiary structures of the protein are formed at the same time [43]. Recent studies in the field show that there are many proteins that exhibit the characteristics of both the diffusion-collision and nucleation-condensation models [44]. The Ising model is one model that describes such a behavior. According to this model, folding can be initiated from any of the helices, a la the diffusion collision model, but propagation of folding occurs simultaneously for secondary and tertiary structures, as in the nucleation condensation model.

Modern theories and recent experimental results propose that the folding process can be explained by a funnel-shaped energy landscape [45, 46]. According to this view, which is opposed to a historical view, various parallel pathways are involved in the folding process. These pathways encompass numerous structures, although all lead to the final state, which is the most stable thermodynamic state under native conditions. At any stage during its folding, the protein exists as an ensemble of conformations and can be trapped transiently in many local energy minima. Detailed structural characterization of those conformations that are sampled by the protein during folding is not yet technically possible. However, new single-molecule techniques, combined with statistical analysis [47–50], enable us to obtain some insight about the folding pathways. Owing to the simplicity of the fold in repeat proteins, structural characterization of intermediate states and folding pathways is feasible.

Studying the thermodynamics of folding, in particular, is a powerful method to get insight into the mechanism involved in the process. This method allows us to map the local minima on the energy landscape and calculate the propensity for the protein to be trapped in these minima. The rate of transition between any two local minima and the order of transitions during the process can also be obtained.

The thermodynamic behavior of small proteins is typically characterized by a two-state model [51–55]. A one-dimensional free energy surface is used to explain the folding of such proteins, and the crossing of the energy barrier between folded and unfolded conformations is the rate-limiting step [56]. According to this model, proteins rarely exist in the transition state region, and the transition over the barrier is an all-or-none event.

During the folding of certain proteins, stable intermediate states have been reported to exist under mild denaturation conditions [57–59]. Circular dichroism and NMR measurements indicate that considerable amounts of secondary structure exist in such intermediate states. However, as indicated by near-UV circular dichroism, as well as by the NMR and fluorescent spectroscopic behavior of aromatic residues [60–62], significant mobility seems to exist in the hydrophobic core, which is not present in the native state of these proteins. This kind of intermediate state was termed the molten globule, which was defined as a structure with loosened tertiary structure that conserves its native secondary structure.

Single-molecule techniques provide new insights into the protein folding field. Using these techniques, partially unfolded stable states have been found in large proteins [48, 63, 64]. In these reports the diversity of the folding pathways

of proteins was observed experimentally. In a few other proteins the continuous changes in an order parameter during folding were observed, suggesting that even for small proteins, folding may be more complicated than was previously thought [65–69]. Clear structural characterization of such processes, however, remains a challenge.

1.4 Single-Molecule Fluorescence Spectroscopy

Single-molecule spectroscopy provides a fundamental advantage in resolving and quantifying the properties of individual molecules or subpopulations by direct examination of the molecular processes. This information is usually inaccessible in classical ensemble experiments due to the averaging. Generally, two different single-molecule methods are used to investigate biomolecules at a single-molecule level [49]. The first method is force spectroscopy, where either the atomic force microscope [70, 71] or the optical trap [72] is used to obtain a detailed structural understanding of molecules under mechanical force. The second method is fluorescence spectroscopy [50, 73–76], which in combination with Förster Resonance Energy Transfer (FRET), is widely used as an accurate technique for distance measurements, and which was termed "a spectroscopic ruler". FRET was discovered in 1948 by Forster [77], who demonstrated that dipole-dipole interactions between donor to acceptor fluorophores can lead to efficient energy transfer between them. This energy transfer strongly depends on the distance between the two probes. Single-Molecule FRET (SM-FRET) has been successfully and extensively used to investigate a variety of processes such as protein folding [50, 63, 78] and misfolding [79], the conformational dynamics of proteins [80], protein-protein interactions [81, 82], intrinsically unstructured proteins [83–85], protein function in vitro [86] and in vivo [87]. It has also been used to study structures involving DNA [88] and RNA [89].

In particular, in the field of protein folding, the modern view proposing the existence of a diversity of folding pathways has attracted much attention to single-molecule spectroscopy techniques. Single-molecule fluorescence spectroscopy was used to obtain detailed information about the folding pathways of single-molecule proteins in equilibrium for fast [64] and for slow [63] folding processes. This was achieved by visualizing the change in FRET efficiency as a function of time at the single-molecule level in different denaturant concentrations and by using statistical analysis to analyze them. SM-FRET was also used to study the protein folding transition path time [90]. In order to investigate the behavior of single molecules under conditions far from equilibrium, time-resolved measurements of FRET efficiency were carried out by abrupt change of the denaturant concentration in micro-fluidic devices [91–93]. SM-FRET was also used to obtain information about unfolded states. Expansion of the unfolded state by increasing the denaturant concentration was observed by single-molecule FRET equilibrium and non-equilibrium measurements [94, 95]. FRET, in combination with fluorescent

correlation spectroscopy, was used to quantify the internal conformational friction of the unfolded states and the intrinsic denatured states [96].

Single-molecule measurements show that the protein folding process for some proteins is very complicated. Diversity in the folding of different proteins was experimentally observed [63, 66, 69].

1.5 The Main Findings of This Work

A CTPR protein with three repeats (CTPR3) was labeled at different positions with two pairs of fluorophores suitable for FRET. Single-molecule measurements were carried out on freely diffusing molecules. Histograms of FRET efficiency were measured in different denaturant concentrations. These histograms revealed two populations: one at a low FRET value, and another at a high FRET value. The population with a low FRET value was found to represent the unfolded state. The population with a high FRET value was found to represent the folded state under denaturant-free conditions. This population was observed to shift continuously to lower FRET values with increasing denaturant concentration.

We demonstrated that continuous loosening of the tertiary structure is responsible for this shift and that crossing the barrier to the unfolded state occurs when the secondary structure melts. This folding model is in opposition to the nucleation condensation model, which proposes coincident propagation of folding of secondary and tertiary structures. The loosening of the tertiary structure is believed to be important in TPR functionality because of its direct relationship to its affinity for binding to other proteins.

References

1. Blatch GL, Lassle M (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. Bioessays 21(11):932–939
2. Forwood JK et al (2010) Quantitative structural analysis of importin-beta flexibility: paradigm for solenoid protein structures. Structure 18(9):1171–1183
3. Andrade MA, Perez-Iratxeta C, Ponting CP (2001) Protein repeats: structures, functions, and evolution. J Struct Biol 134(2–3):117–131
4. Javadi Y, Itzhaki LS (2013) Tandem-repeat proteins: regularity plus modularity equals design-ability. Curr Opin Struct Biol 23(4):622–631
5. Mello CC, Barrick D (2004) An experimentally determined protein folding energy landscape. Proc Natl Acad Sci USA 101(39):14102–14107
6. Forrer P et al (2004) Consensus design of repeat proteins. ChemBioChem 5(2):183–189
7. Main ER et al (2003) Design of stable alpha-helical arrays from an idealized TPR motif. Structure 11(5):497–508
8. Mosavi LK, Minor DL Jr, Peng ZY (2002) Consensus-derived structural determinants of the ankyrin repeat motif. Proc Natl Acad Sci USA 99(25):16029–16034

9. Binz HK et al (2003) Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. J Mol Biol 332(2):489–503
10. Stumpp MT et al (2003) Designing repeat proteins: modular leucine-rich repeat protein libraries based on the mammalian ribonuclease inhibitor family. J Mol Biol 332(2):471–487
11. Nikkhah M et al (2006) Engineering of beta-propeller protein scaffolds by multiple gene duplication and fusion of an idealized WD repeat. Biomol Eng 23(4):185–194
12. Parmeggiani F et al (2008) Designed armadillo repeat proteins as general peptide-binding scaffolds: Consensus design and computational optimization of the hydrophobic core. J Mol Biol 376(5):1282–1304
13. D'Andrea LD, Regan L (2003) TPR proteins: the versatile helix. Trends Biochem Sci 28(12):655–662
14. Tang KS et al (1999) Stability and folding of the tumour suppressor protein p16. J Mol Biol 285(4):1869–1886
15. Das AK, Cohen PW, Barford D (1998) The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. EMBO J 17(5):1192–1199
16. Zeytuni N, Zarivach R (2012) Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. Structure 20(3):397–405
17. Fransen M et al (2008) Comparison of the PTS1- and Rab8b-binding properties of Pex5p and Pex5Rp/TRIP8b. Biochim Et Biophys Acta-Mol Cell Res 1783(5):864–873
18. Ramsey AJ, Russell LC, Chinkers M (2009) C-terminal sequences of hsp70 and hsp90 as non-specific anchors for tetratricopeptide repeat (TPR) proteins. Biochem J 423:411–419
19. Young JC, Barral JM, Hartl FU (2003) More than folding: localized functions of cytosolic chaperones. Trends Biochem Sci 28(10):541–547
20. Smith RL, Redd MJ, Johnson AD (1995) The tetratricopeptide repeats of ssn6 interact with the homeo domain of alpha-2. Genes Dev 9(23):2903–2910
21. Edqvist PJ et al (2006) Tetratricopeptide repeats in the type III secretion chaperone, LcrH: their role in substrate binding and secretion. Mol Microbiol 59(1):31–44
22. Tiwari D et al (2009) Key residues in mycobacterium tuberculosis protein kinase G play a role in regulating kinase activity and survival in the host. J Biol Chem 284(40):27467–27479
23. Mirus O et al (2009) Evolutionarily evolved discriminators in the 3-TPR domain of the Toc64 family involved in protein translocation at the outer membrane of chloroplasts and mitochondria. J Mol Model 15(8):971–982
24. Baker MJ et al (2007) Mitochondrial protein-import machinery: correlating structure with function. Trends Cell Biol 17(9):456–464
25. Zeytuni N et al (2011) Self-recognition mechanism of MamA, a magnetosome-associated TPR-containing protein, promotes complex assembly. Proc Natl Acad Sci USA 108(33):E480–E487
26. Gatsos X et al (2008) Protein secretion and outer membrane assembly in Alphaproteobacteria. FEMS Microbiol Rev 32(6):995–1009
27. Cortajarena AL et al (2010) Designed proteins to modulate cellular networks. ACS Chem Biol 5(6):545–552
28. Cortajarena AL et al (2004) Protein design to understand peptide ligand recognition by tetratricopeptide repeat proteins. Protein Eng Des Sel 17(4):399–409
29. Cortajarena AL, Yi F, Regan L (2008) Designed TPR modules as novel anticancer agents. ACS Chem Biol 3(3):161–166
30. Grove TZ, Regan L, Cortajarena AL (2013) Nanostructured functional films from engineered repeat proteins. J Roy Soc Interface 10(83):20130051
31. Wetzel SK et al (2008) Folding and unfolding mechanism of highly stable full-consensus ankyrin repeat proteins. J Mol Biol 376(1):241–257
32. Javadi Y, Main ER (2009) Exploring the folding energy landscape of a series of designed consensus tetratricopeptide repeat proteins. Proc Natl Acad Sci USA 106(41):17383–17388

33. Gallavotti G (1999) Statistical mechanics : a short treatise. Texts and monographs in physics, Springer, Berlin, xiv, 339 p
34. Scott RA 3rd (1967) Statistical mechanical studies of polypeptides. 1. Theory of the helix-coil transition including right- and left-handed helical states. Biopolymers 5(10):931–951
35. Kajander T et al (2005) A new folding paradigm for repeat proteins. J Am Chem Soc 127(29):10188–10190
36. Cortajarena AL, Mochrie SGJ, Regan L (2011) Modulating repeat protein stability: the effect of individual helix stability on the collective behavior of the ensemble. Protein Sci 20(6):1042–1047
37. Phillips JJ et al (2012) Modulation of the multistate folding of designed TPR proteins through intrinsic and extrinsic factors. Protein Sci 21(3):327–338
38. Bradley CM, Barrick D (2002) Limits of cooperativity in a structurally modular protein: response of the Notch ankyrin domain to analogous alanine substitutions in each repeat. J Mol Biol 324(2):373–386
39. Zweifel ME, Barrick D (2001) Studies of the ankyrin repeats of the Drosophila melanogaster Notch receptor. 1. Solution conformational and hydrodynamic properties. Biochemistry 40(48):14344–14356
40. Ferreiro DU et al (2008) The energy landscapes of repeat-containing proteins: topology, cooperativity, and the folding funnels of one-dimensional architectures. PLoS Comput Biol 4(5):e1000070
41. Hagai T et al (2012) Modulation of folding kinetics of repeat proteins: interplay between intra- and interdomain interactions. Biophys J 103(7):1555–1565
42. Serquera D et al (2010) Mechanical unfolding of an ankyrin repeat protein. Biophys J 98(7):1294–1301
43. Karplus M, Weaver DL (1976) Protein-folding dynamics. Nature 260(5550):404–406
44. Travaglini-Allocatelli C et al (2009) Folding and stability of globular proteins and implications for function. Curr Opin Struct Biol 19(1):3–7
45. Sali A, Shakhnovich E, Karplus M (1994) How does a protein fold? Nature 369(6477):248–251
46. Karplus M (1997) The Levinthal paradox: yesterday and today. Fold Des 2(4):S69–S75
47. Schuler B (2005) Single-molecule fluorescence spectroscopy of protein folding. ChemPhysChem 6(7):1206–1220
48. Stigler J et al (2011) The complex folding network of single calmodulin molecules. Science 334(6055):512–516
49. Borgia A, Williams PM, Clarke J (2008) Single-molecule studies of protein folding. Annu Rev Biochem 77:101–125
50. Haran G (2003) Single-molecule fluorescence spectroscopy of biomolecular folding. J Phys Condens Matter 15:R1291–R1317
51. Privalov PL (1979) Stability of proteins: small globular proteins. Adv Protein Chem 33:167–241
52. Weikl TR, Palassini M, Dill KA (2004) Cooperativity in two-state protein folding kinetics. Protein Sci 13(3):822–829
53. Segawa SI, Sugihara M (1984) Characterization of the transition-state of Lysozyme unfolding. 1. effect of protein solvent interactions on the transition-state. Biopolymers 23(11):2473–2488
54. Jackson SE, Fersht AR (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. Biochemistry 30(43):10428–10435
55. Baker D (2000) A surprising simplicity to protein folding. Nature 405(6782):39–42
56. Kubelka J, Hofrichter J, Eaton WA (2004) The protein folding 'speed limit'. Curr Opin Struct Biol 14(1):76–88
57. Ptitsyn OB et al (1990) Evidence for a molten globule state as a general intermediate in protein folding. FEBS Lett 262(1):20–24
58. Christensen H, Pain RH (1991) Molten globule intermediates and protein folding. Eur Biophys J 19(5):221–229

59. Ohgushi M, Wada A (1983) Molten-globule state—a compact form of globular-proteins with mobile side-chains. FEBS Lett 164(1):21–24
60. Carrey EA, Pain RH (1978) Conformation of a stable intermediate on the folding pathway of Staphylococcus aureus penicillinase. Biochim Biophys Acta 533(1):12–22
61. Baum J et al (1989) Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig alpha-lactalbumin. Biochemistry 28(1):7–13
62. Roder H, Elove GA, Englander SW (1988) Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. Nature 335(6192):700–704
63. Pirchi M et al (2011) Single-molecule fluorescence spectroscopy maps the folding landscape of a large protein. Nat Commun 2:493
64. Chung HS et al (2011) Extracting rate coefficients from single-molecule photon trajectories and fret efficiency histograms for a fast-folding protein. J Phys Chem A 115(16):3642–3656
65. Knott M, Chan HS (2006) Criteria for downhill protein folding: calorimetry, chevron plot, kinetic relaxation, and single-molecule radius of gyration in chain models with subdued degrees of cooperativity. Proteins 65(2):373–391
66. Liu J et al (2012) Exploring one-state downhill protein folding in single molecules. Proc Natl Acad Sci USA 109(1):179–184
67. Favrin G et al (2003) Two-state folding over a weak free-energy barrier. Biophys J 85(3):1457–1465
68. Naganathan AN, Orozco M (2011) The native ensemble and folding of a protein molten-globule: functional consequence of downhill folding. J Am Chem Soc 133(31):12154–12161
69. Lindhoud S (2012) Visualization and charachterisation of apoflavodoxin folding, wageningen university. p 140
70. Engel A, Gaub HE, Muller DJ (1999) Atomic force microscopy: a forceful way with single molecules. Curr Biol 9(4):R133–R136
71. Rief M et al (1997) Reversible unfolding of individual titin immunoglobulin domains by AFM. Science 276(5315):1109–1112
72. Kellermayer MS et al (1997) Folding-unfolding transitions in single titin molecules characterized with laser tweezers. Science 276(5315):1112–1116
73. Moerner WE (2002) A dozen years of single-molecule spectroscopy in physics, chemistry, and biophysics. J Phys Chem B 106(5):910–927
74. Haustein E, Schwille P (2004) Single-molecule spectroscopic methods. Curr Opin Struct Biol 14(5):531–540
75. Michalet X et al (2003) The power and prospects of fluorescence microscopies and spectroscopies. Annu Rev Biophys Biomol Struct 32:161–182
76. Roy R, Hohng S, Ha T (2008) A practical guide to single-molecule FRET. Nat Methods 5(6):507–516
77. Forster T (1948) Zwischenmolecular engergiewanderung and floureszenz. Ann Phys Ny 2:55
78. Schuler B, Lipman EA, Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. Nature 419(6908):743–747
79. Borgia MB et al (2011) Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. Nature 474(7353):662–665
80. Hanson JA et al (2007) Illuminating the mechanistic roles of enzyme conformational dynamics. Proc Natl Acad Sci USA 104(46):18055–18060
81. Bae W, Choi M-G, Hyeon C, Shin Y-K, Yoon T-Y (2013) Real-time observation of multiple-protein complex formation with single-molecule FRET. J Am Chem Soc 135:10254–10257
82. Truong K, Ikura M (2001) The use of FRET imaging microscopy to detect protein-protein interactions and protein conformational changes in vivo. Curr Opin Struct Biol 11(5):573–578
83. Ohashi T et al (2007) An experimental study of GFP-based FRET, with application to intrinsically unstructured proteins. Protein Sci 16(7):1429–1438
84. Schuler B et al (2012) Application of confocal single-molecule FRET to intrinsically disordered proteins. Methods Mol Biol 896:21–45

85. Nath A et al (2012) The conformational ensembles of alpha-synuclein and tau: combining single-molecule FRET and simulations. Biophys J 103(9):1940–1949
86. Blanchard SC (2009) Single-molecule observations of ribosome function. Curr Opin Struct Biol 19(1):103–109
87. Adjobo-Hermans MJW et al (2011) Real-time visualization of heterotrimeric G protein Gq activation in living cells. Bmc Biol 9:32
88. Weiss S (1999) Fluorescence spectroscopy of single biomolecules. Science 283(5408):1676–1683
89. Zhuang XW (2005) Single-molecule RNA science. Annu Rev Biophys Biomol Struct 34:399–414
90. Chung HS et al (2012) Single-molecule fluorescence experiments determine protein folding transition path times. Science 335(6071):981–984
91. Lipman EA et al (2003) Single-molecule measurement of protein folding kinetics. Science 301(5637):1233–1235
92. Wunderlich B et al (2013) Microfluidic mixer designed for performing single-molecule kinetics with confocal detection on timescales from milliseconds to minutes. Nat Protoc 8(8):1459–1474
93. Kinoshita M et al (2007) Development of a technique for the investigation of folding dynamics of single proteins for extended time periods. Proc Natl Acad Sci USA 104(25):10453–10458
94. Schuler B, Eaton WA (2008) Protein folding studied by single-molecule FRET. Curr Opin Struct Biol 18(1):16–26
95. Sherman E, Haran G (2006) Coil-globule transition in the denatured state of a small protein. Proc Natl Acad Sci USA 103(31):11539–11543
96. Soranno A et al (2012) Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. Proc Natl Acad Sci USA 109(44):17800–17806

Chapter 2
Methods

2.1 Protein Purification and Labeling

This protein purification and labeling was done in collaboration with Luca Domenico D'Andrea, Istituto di Biostrutture e Bioimmagini, CNR, Napoli, Italy.

The protein cloning, expression and purification was carried out according to the protocol published previously [1].

The amino acid sequence of CTPR3 is as follows: GNSAEAWYNLGNAYYKQGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDY DEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIE YYQKALELDPNNAEAKQNLGNAKQKQG. The crystal structure of CTPR3 with the same color coding is shown in Fig. 2.1.

A novel strategy was developed for site specific labeling of CTPR3 [2]. This method is based on the use of the expressed protein ligation. Schematic representation of this process is shown in Fig. 2.2. This technique requires the splitting of the protein of interest into two fragments, each bearing a single cysteine. The N-terminal fragment is expressed in a bacterial host as a fusion protein with the *Mycobacterium xenopi* gyrase A intein (Mxe GyrA intein). The expressed chimeric protein is also endowed with a C-terminal His6-tag, allowing its purification by affinity chromatography. After purification, the fusion protein is incubated with a thiol, which induces the splicing of Mxe GyrA intein and the release of the protein fragment as a C-terminal thioester. The thioester protein is then isolated from the intein and selectively labeled on its single Cys residue upon reaction with a probe containing a maleimide group as sulfhydryl-reactive moiety. The C-terminal fragment, carrying an N-terminal Cys residue, can be chemically synthesized or expressed in bacteria depending on the fragment size. The native chemical ligation reaction between the mono-labeled thioester fragment and the C-terminal fragment affords the full-length mono-labeled protein.

S. Cohen, *Single-Molecule Fluorescence Spectroscopy of the Folding of a Repeat Protein*, Springer Theses, DOI 10.1007/978-3-319-09558-5_2

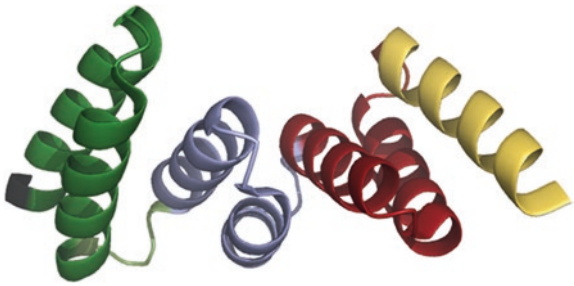

Fig. 2.1 Crystal structure of CTPR3, PDB ID:1NA0

The Cys residue involved in NCL reaction is then exploited to introduce the second probe into the protein (Fig. 2.2, strategy A). A simpler approach is used when the second probe is located at the C-terminus. In this case, the full-length protein containing a single Cys at the required position is expressed as an intein fusion protein. After purification and first labeling, the full-length thioester protein is reacted with L-Cys in a NCL reaction. The newly introduced Cys residue is then labeled with the second probe (Fig. 2.2, strategy B). A final re-folding step is required before their utilization.

The details of the labeling process was reported for CTPR with Atto dyes [2]. This method was also used for the labeling of CTPR3 with Alexa-maleimide dyes (Life technologies Invitrogen). Amino acid sequences of the CTPR3 variants and a schematic representation of double labeled CTPR are shown in Fig. 2.3.

CTPR31C Atto is the CTPR3 double labeled with Atto 488 on site 121, and Atto 647 N on site 36.
CTPR32C Atto is the CTPR3 double labeled with Atto 488 on site 121, and Atto 647 N on site 70.
CTPR3NC Atto is the CTPR3 double labeled with Atto 488 on site 121, and Atto 647 N on site −1.
CTPR313 Atto is the CTPR3 double labeled with Atto 488 on site 105, and Atto 647 N on site 12.
CTPR31C Alexa is the CTPR3 double labeled with Alexa 488 on site 121, and Alexa 594 on site 36.
CTPR3NC Alexa is the CTPR3 double labeled with Alexa 488 on site 121, and Alexa 594 on site −1.

Finally, for control measurements, CTPR3 is labeled only with Alexa 488 in position 121(CTPR3SL-Alexa488).

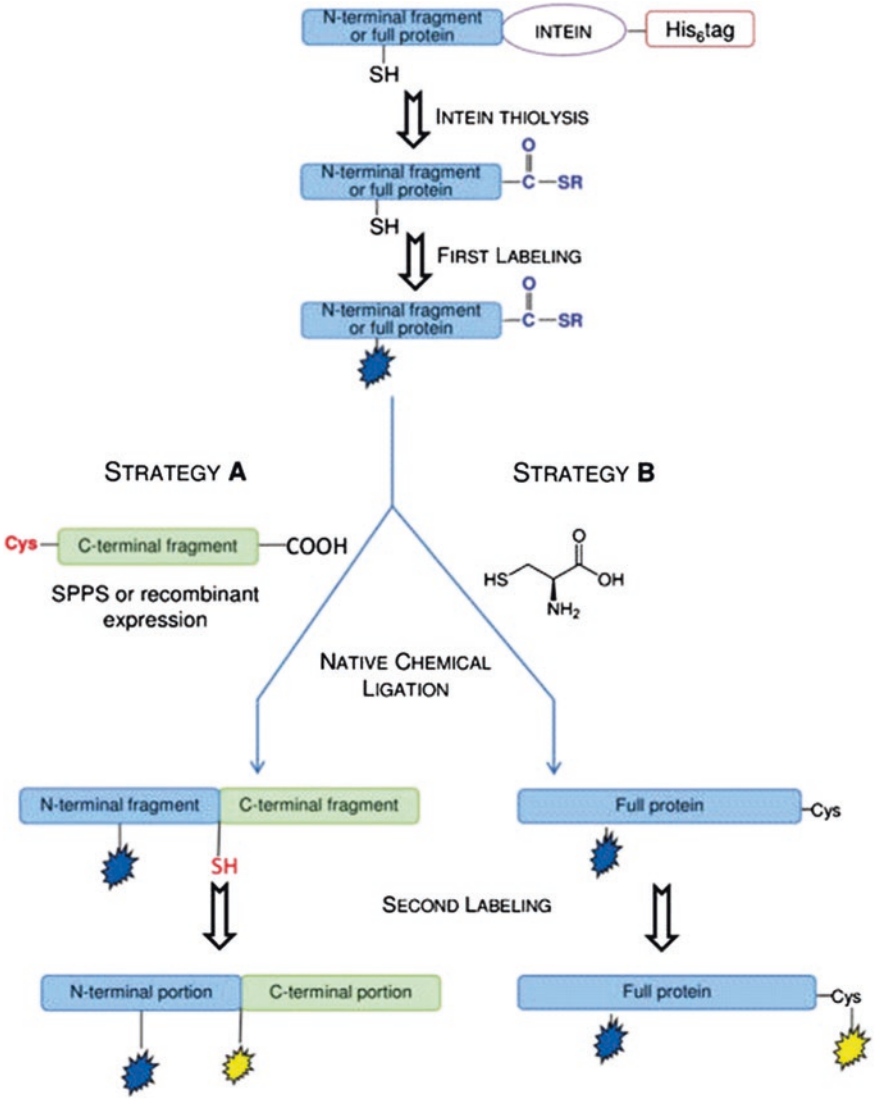

Fig. 2.2 A schematic representation of the semi-synthetic strategy adopted for the preparation of doubly labeled proteins. The N-terminal portion of the protein was prepared as a C-terminal thioester through expression with the Mxe GyrA intein as fusion partner. After splicing, the N-terminal fragment of the protein, carrying a C-terminal thioester group, was labeled with the first probe and then reacted by native chemical ligation with the C-terminal remaining portion of the protein containing an N-terminal Cys residue (strategy A) or a L-Cys (strategy B). Finally, the full protein is labeled with the second probe on the Cys reactive in the native chemical ligation reaction

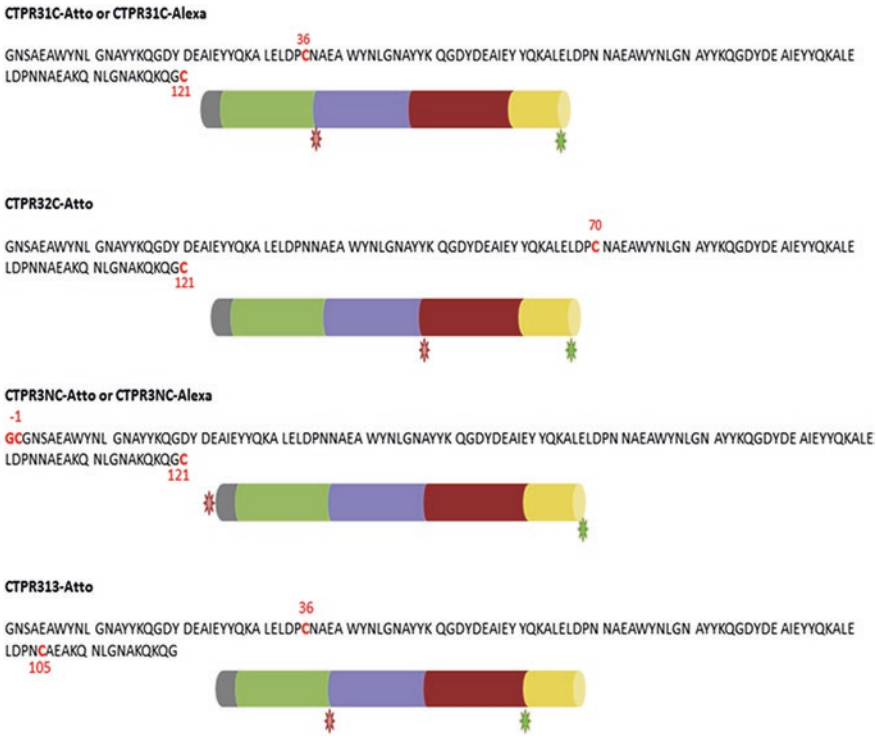

Fig. 2.3 Amino acid sequences and schematic representation of the CTPR3 variants and double labeled proteins. Red colored amino acids represent insertion and/or mutation. When double labeled with Atto dyes, green stars represent Atto 488 and red stars represent Atto 647 N. When double labeled with Alexa dyes, green stars represent Alexa 488 and red stars represent Alexa 594

2.2 General Experimental Methods

All measurements were performed in 50 mM phosphate buffer, 150 mM NaCl, at pH 6.8. Index of refraction of stock solution was determined using a Fisher Abbe refractometer (Fisher Scientific Co.). Concentration of guanidine hydrochloride (GuHCl) was obtained using the equation below

$$\text{Concentration of GuHCl} = 57.147\Delta n + 38.68\Delta n^2 - 91.6\Delta n^3 \qquad (2.1)$$

where, Δn is the difference between index of refraction of buffer with and without GuHCl.

2.3 Circular Dichroism Spectroscopy

Circular Dichroism (CD) spectroscopy was performed on a AVIV Model 215 CD spectrophotometer (AVIV Instruments, Lakewood, NJ). CD experiments were performed at protein concentration of 6 μM, 25 °C, in 222 nm wavelength.

2.4 Ensemble/Bulk Steady-State Measurements

Ensemble/bulk fluorescence studies were conducted with a Fluorolog spectrofluorometer (Jobin-Yvon) equipped with Glan-Thompson polarizers. CTPR3 labeled with Alexa 488 and the double-labeled CTPR31C were used for control experiments, including evaluation of fluorescence quantum yield and anisotropy values as a function of GuHCl concentration (see Sect. 3.12). CTPR31C was also used for measuring the ensemble/bulk denaturation curve.

2.5 Single-Molecule Setup and Measurement Methodology

Single-molecule setup

A custom-made single-molecule microscope for automated data collection was constructed, as described in Fig. 2.4. The beam of an argon ion laser (Spectra-Physics model 163, 488 nm) was passed through a laser-line cleanup filter (LP, Chroma Z488/10x) and expanded five times (BX) to fill the back aperture of a 100x oil-immersion objective (OBJ, Zeiss FLUAR 100x/1.3NA). A sample cell (SAM) was mounted on a custom-designed sample holder on top of a capacitance-feedback 100 × 100 × 20 μm piezo stage (PIEZO, PI P-517) controlled by a dedicated DSP (PI E-710). The sample was excited by the focused

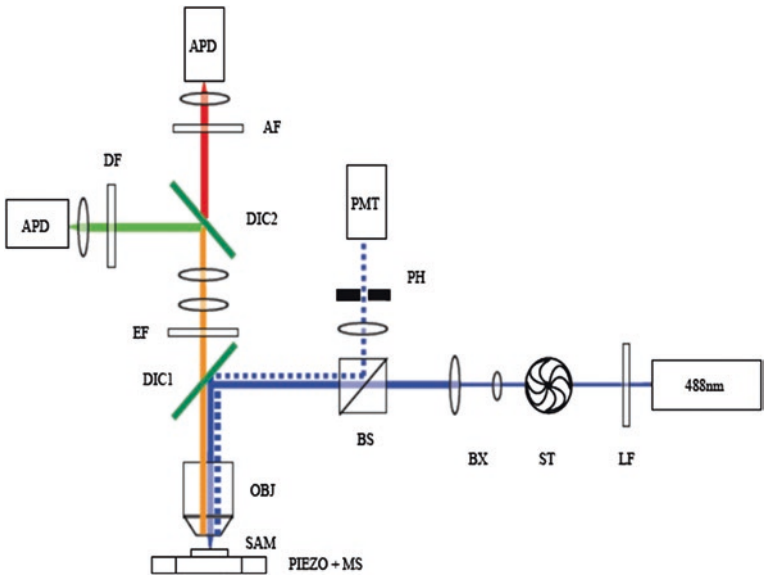

Fig. 2.4 Schematic representation of the confocal set up

laser beam and the fluorescent emission from the sample was collected by the objective, passed through the first dichroic mirror (DIC1, Chroma 505DCLP), filtered from remaining excitation light (EF, Semrock LP02-488RU), and split into two detection channels by a second dichroic mirror (DIC2, Chroma 565DCLP). The donor dye emission and the acceptor dye emission were filtered by respective emission filters (DF, Semrock FF01-536/40 and AF, Semrock FF01-624/40) and focused onto two single-photon avalanche photodiodes (APD, Perkin-Elmer SPCM-AQR-15). Each APD delivered outputs a TTL pulse output for each incoming photon. The arrival times of each of these photons was recorded by a counter/timer card (National Instruments NI6602) with time resolution of 12.5 ns. A standalone TCSPC module (PicoHarp 300, PicoQuant) was used instead of the counter/timer card in free diffusion experiments. In order to generate an auto-focus mechanism, back-reflected laser light from the sample surface was focused on a pinhole (PH) and measured by a photomultiplier tube (PMT, Hamamatsu H8249). When exactly focused, the PMT position was adjusted to register the maximum photon flux. The piezo stage was mounted on top of a coarse motorized stage (MS) to allow larger-scale motion of the sample. When measurements were not in progress, to prevent excess illumination, a mechanical shutter (ST, Uniblitz LS6T2) blocked the laser.

The encapsulation of proteins

Liposomes made of egg phosphatidylcholine and a fraction of 1:500 of 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine-N-(cap biotinyl) (both from Avanti Lipids) were prepared by extrusion [3] in the appropriate buffer, using a disposable 0.1 μm syringe connected to Anopore filter (Whatman Anotop-10). To prepare protein-loaded liposomes, we added labeled protein to a final concentration of ~0.2 μM, chosen so that one out of ~10 liposomes would contain a single molecule [4]. Liposomes were separated from non-encapsulated proteins by size exclusion column (GE Healthcare MicroSpin S-400 HR).

Flow cells

Sample cells were made of two glass #1.5 coverslips (Thomas Scientific), glued together with two Teflon strips. After incubation, to prevent evaporation, the cell was sealed with silicon grease.

2.5.1 Measurements on Freely-Diffusing Molecules

Sample preparation

Flow cells filled with a dilute solution of protein molecules (20–45 pM), and 0.01 % Tween to minimize protein adsorption, at different denaturant concentrations were prepared.

Data acquisition and analysis

The laser beam was focused 20 μm deep into the solution, and data were collected continuously for 1 h for each sample. Several such samples were collected at each denaturant concentration to get enough statistics.

In free diffusion experiments, a standalone TCSPC module (PicoHarp 300, PicoQuant) with 4 ps resolution was used instead of the counter/timer card to record the arrival time lags of photons at the detectors. Data analysis was performed using similar methods to those [5] reported by Seidel and coworkers. A running-average of 15 photons was first used to smooth the photon trajectory. A cut-off time (50 μs) was then determined from the histogram of the time lags, and used to effectively separate background photons from those related to fluorescent bursts during the passage of molecules in the beam. Fluorescent bursts were identified using this cut-off time, and only bursts with a total of 35 photons or more were selected for further analysis. In some part of analysis, higher thresholds were used to choose the bursts as mentioned in the corresponding section. FRET efficiency was then calculated for each burst as the number of photons arriving from the acceptor channel divided by total number of photons. FRET efficiency histograms were then constructed.

2.5.2 Measurements on Diffusing Encapsulated Molecules

Sample preparation

A supported lipid bilayer was formed on glass surfaces by incubating a solution containing empty liposomes in flow cells to prevent interaction with the surface. For preparation of protein-loaded liposomes, we added labeled proteins to a final concentration of ~0.05 μM, so that one out of ~60 liposomes would contain a single molecule [6]. This ensures low probability of producing liposomes containing more than one molecule. Protein-loaded liposomes were later diluted 5–10 times to reduce the background after they were separated from non-encapsulated proteins and were then introduced into the cell.

Data analysis

A similar analytical method was used to find the bursts as previously explained in Sect. 2.5.1. Bursts with more than 100 photons were selected to obtain FRET histograms, which were calculated for bins of 100 photons. A small fraction of bursts, whose FRET efficiency from the first to the last bin changed by more than 15 %, were discarded.

Temperature control of the sample

A coil of cold water circulation around the objective was used to lower the temperature of the objective. To monitor the temperature of the sample, readings were taken constantly using a thermocouple (Minco). To avoid water condensation on

the lenses of the objective, the objective and the sample were kept in an argon filled atmosphere, and a low flow of nitrogen gas was injected continuously close to the back aperture to avoid air entrance.

2.5.3 Measurements on Immobilized Molecules

Sample preparation

A supported lipid bilayer was formed on glass surface by incubating a cell with a solution containing empty liposomes in flow cells. Incubation with a solution of 1 mg/ml streptavidin (Sigma) was followed by introduction of protein-loaded liposomes into the cell. For preparation of protein-loaded liposomes, the same protocol as Sect. 2.5.2. Was used, with the difference that a 0.2 M protein solution was used instead of a 0.05 M solution to obtain higher concentration of protein-loaded liposomes.

Data acquisition

Data acquisition was fully automated using dedicated software programmed in LabWindows/CVI and MATLAB. A 5×5 μm region of the sample was scanned, and the position of vesicles loaded with molecules was identified with subpixel resolution. A piezo stage was used to position each of molecules in the focus of the laser, in order to obtain a fluorescence time trace (trajectory). For each molecule, either donor and acceptor signals were recorded for 5 s, or until background level was reached. After acquiring trajectories of all molecules in that field, the piezo stage was moved to a new region, and the acquisition cycle was repeated. The arrival time of each of these photons were recorded by a counter/timer card (National Instruments NI6602) with time resolution of 12.5 ns. The auto-focus device ensured that the laser beam was focused on the surface of the sample throughout the data collection.

Data analysis

First fluorescence trajectories (accumulated as photon arrival times on the two detectors) were binned in time bins. To ensure that only trajectories generated by individual molecules were included in the analysis, and to prevent the occurrence of various artifacts, we used a series of computational filters (see additional details below). Trajectories passing the filtration stage were corrected for background and leakage of photons from donor to acceptor channel. These trajectories amounted to about 9 % of the total data set. Using the equation below, the FRET efficiency was calculated for each bin:

$$FRET\ efficiency = \frac{I_A(t)}{I_A(t) + \gamma I_D(t)} \tag{2.2}$$

γ is the correction factor for differences in donor and acceptor quantum yields as well as differences in the detection efficiencies at the respective wavelengths (see Sect. 3.8).

To identify transitions between FRET efficiency levels in a model-independent manner, single-molecule trajectories were subjected to change-point analysis [7]. Bootstrapped trajectories were used to estimate the statistical significance of the identified transitions.

Additional details on the data analysis procedures

The initial treatment of the measured trajectories.

In order to ensure the quality the results, and to avoid artifacts, the following steps were taken in the treatment of single-molecule trajectories before analysis. These steps were implemented automatically, and resulted in selection of 7–10 % of the trajectories (depending on the data set), for further analysis.

First, we binned the two data sets in each trajectory (donor and acceptor) in time-windows to get $I_{D,raw}(t)$ and $I_{A,raw}(t)$. The bin windows where chosen according to the laser power to get a total count of about 50 counts per bin.

We used change-point analysis to identify trajectories that did not show a photobleaching step, and removed them from the set.

The lowest intensity region in each single-molecule trajectory was identified as the background level for that molecule. Trajectories with individual channel background levels (b_D, b_A) larger than threshold were removed. The threshold value was chosen depending on the laser intensity, and was between 10–14 counts per bin for each channel.

Any trajectory with a total intensity larger than the chosen threshold was also removed, as it might have arisen from two molecules in the same or nearby vesicles.

The intensity of the remaining trajectories was then corrected for background and leakage of photons from donor to acceptor channel using:

$$\begin{aligned} I_D(t) &= \big(I_{D,raw}(t) - b_D\big)(1+\lambda) \\ I_A(t) &= \big(I_{A,raw}(t) - b_A\big) - \lambda\big(I_{D,raw}(t) - b_D\big) \end{aligned} \tag{2.3}$$

Using a solution of free Alexa-488 molecules, the leakage factor $\lambda = 0.088$ was measured.

In the trajectories, to ensure that essentially only anti-correlated intensity changes in donor and acceptor occur (as expected for FRET transitions), we cut the trajectories whenever the total intensity ($= I_D(t) + \gamma\, I_A(t)$) changed by more than 25 %. In most cases, this point occurred when the donor photobleached, although in some trajectories we identified additional intensity changes.

We also made certain that the acceptor data set of each trajectory included a photobleaching step (which could occur either simultaneously with the donor photobleaching step or independently of it). Trajectories that did not show such a photobleaching step were discarded.

2.6 FRET-Fluorescence Correlation Spectroscopy

In these measurements, the single molecule set up was used. The dichroic mirror was replaced by a polarizer beam splitter. The emitted light was separated using a polarizer beam splitter into two channels, and focused on the avalanche photo diodes. The cross correlation between the two channels was calculated by a photon-counting digital hardware correlator (Flex02-12D). The donor and acceptor auto-correlation functions were separately measured by locating the proper emission filter before the polarizer, and by measuring the cross correlation between the two channels. The cross correlation function was calculated after placing emission filters for donor and acceptor fluorophores in donor and acceptor channels, respectively.

References

1. Main ER et al (2003) Design of stable alpha-helical arrays from an idealized TPR motif. Structure 11(5):497–508
2. De Rosa L et al (2012) Site-specific protein double labeling by expressed protein ligation: applications to repeat proteins. Org Biomol Chem 10(2):273–280
3. Macdonald RC et al (1991) Small-volume extrusion apparatus for preparation of large. Unilamellar Vesicles Biochimica Et Biophysica Acta 1061(2):297–303
4. Boukobza E, Sonnenfeld A, Haran G (2001) Immobilization in surface-tethered lipid vesicles as a new tool for single biomolecule spectroscopy. J Phys Chem B 105:12165–12170
5. Fries JR et al (1998) Quantitative identification of different single molecules by selective time-resolved confocal fluorescence spectroscopy. J Phys Chem A 102(33):6601–6613
6. Boukobza E, Sonnenfeld A, Haran G (2001) *Immobilization in surface-tethered lipid vesicles as a new tool for single biomolecule spectroscopy*. J Phys Chem B 105(48):12165–12170
7. Taylor WA (2000) *Change-point analysis: a powerful new tool for detecting changes*. Available from: http://www.variation.com/cpa/tech/changepoint.html

Chapter 3
Results

3.1 Measurements on Free Diffusing CTPR31C-Alexa

To follow the folding process of the protein, single-molecule burst experiments [1] on CTPR31C-Alexa molecules were performed. Protein molecules were set to diffuse through the focus of the confocal microscope. Bursts of emitting photons from donor and acceptor fluorophores were registered by two avalanche photo diodes (APDs). Signal detection was performed using photon counting instrumentation (PicoHarp, PicoQuant). A running-average of 15 photons was first used to smooth the photon trajectory. Photons arriving at any of the detectors with a time difference smaller than 50 μs were considered as belonging to the same fluorescent burst. Bursts with more than 35 photons per burst were analyzed. The experiment was repeated for a series of GuHCl concentrations, and histograms of FRET values were constructed. The histograms suggested three populations: (1) A population with zero FRET, representing molecules with only donor fluorophores activated. (2) A population at a low FRET value of 26 %, representing the unfolded state. (3) A population at a high FRET value of 76 %, which shifted to a lower FRET value of 63 % by increasing GuHCl concentration (Fig. 3.1). There are two interpretations for this shift. (1) A continuous change of the FRET. (2) The existence of stable intermediate state.

A continuous change of the FRET efficiency

To characterize the continuous shift of the folded state, a three-Gaussian model was used. In this model, Gaussian distribution functions (Gs) were used to model molecules with only donor fluorophores, unfolded molecules and, the shifting population.

$$\mathrm{Gs}(A_i, w_i, FRET_{C_i}) = \left(\frac{A_i}{w_i * sqrt\left(\frac{PI}{2}\right)}\right) * exp\left(-2 * \left(\frac{FRET - FRET_{C_i}}{w_i}\right)^2\right) \tag{3.1}$$

$$Hist = \sum_{i=0}^{i=2} Gs(A_i, w_i, FRET_{C_i}) \tag{3.2}$$

S. Cohen, *Single-Molecule Fluorescence Spectroscopy of the Folding of a Repeat Protein*, Springer Theses, DOI 10.1007/978-3-319-09558-5_3

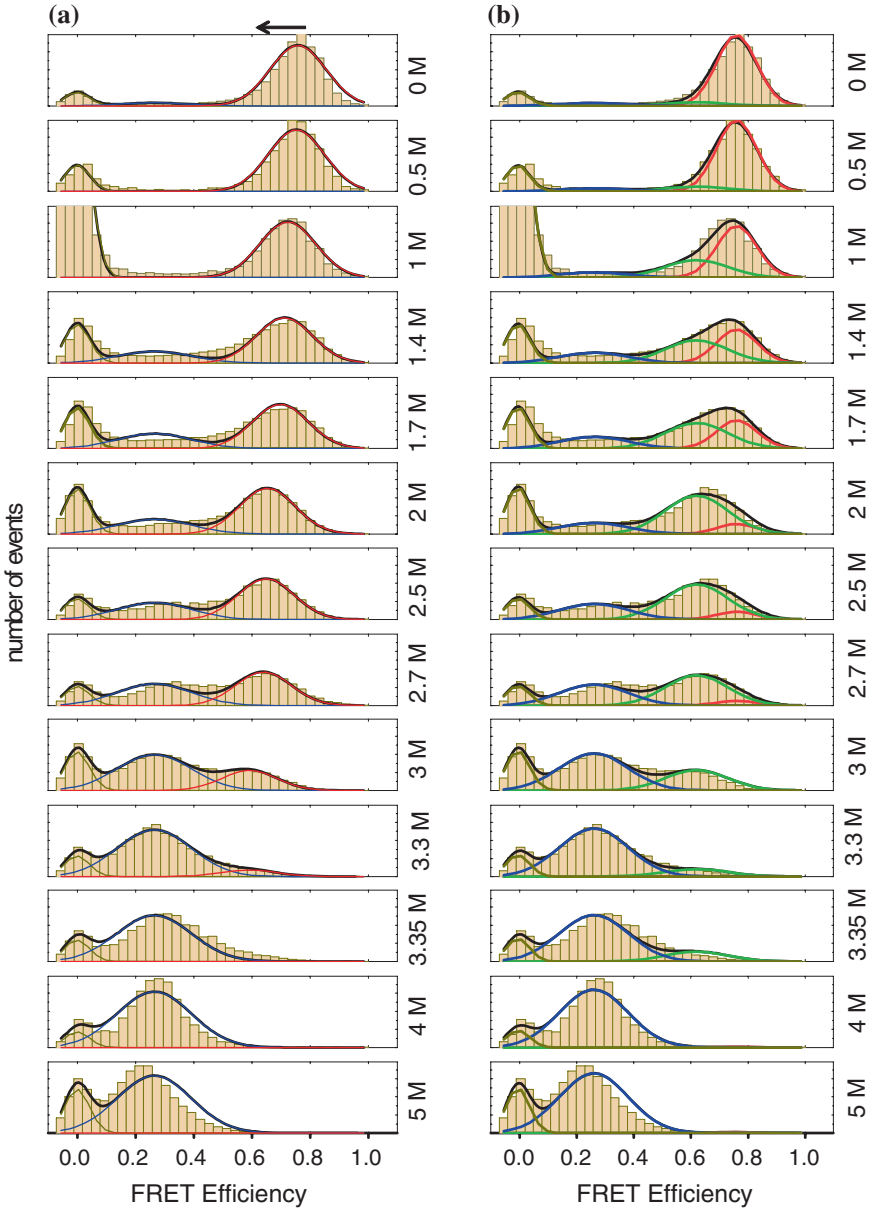

Fig. 3.1 The SM-FRET histograms of CTPR31C-Alexa at various GuHCl concentrations. Global fit of the histogram to (**a**) a model with three populations; a population with a varying FRET value (*red*), unfolded state (*blue*) and donor-only species (*black*), (**b**) to a model with four populations; folded (*red*), unfolded (*blue*), intermediate state (*olive green*) and donor only species (*olive green*)

where, $FRET_c$ are the mean FRET efficiencies of each state, and A and w are the relative propensities of the states and the widths of the distributions, respectively. Global fitting was used to fit all the histograms to the equation above with the shared parameters T_{Ci0}, $FRET_{C1}$, w_1, w_0, w_2, and free parameters A_0, A_1, A_2, $FRET_{C2}$ for all GuHCl concentrations. The result of the fit is shown in Fig. 3.1a. The shared parameters were found from the fit to be as follows: $FRET_{C_0} = 0\,\% \pm 0.1\,\%, FRET_{C_1} = 26.5\,\% \pm 0.3\,\%$, $w_0 = 8.5\,\% \pm 0.2\,\% w_1 = 25.5\,\% \pm 0.7\,\%$, $w_2 = 19.1\,\% \pm 4\,\%$, $FRET_{C_2}$ decreases from 76 to 64 % as the GuHCl concentration increases.

Testing for existence of a stable intermediate state

It is possible that the continuous shift of the high FRET efficiency peak hides an exchange between two populations: the folded state and an intermediate. To model the existence of a stable intermediate state, a four-Gaussian model was used (Fig. 3.1b). In this model Gaussian distribution functions (Gs) were used to model population by molecules with only donor fluorophores, folded, unfolded, and the intermediate state species:

$$Hist = \sum_{i=0}^{i=3} Gs(A_i, w_i, FRET_{Ci}) \tag{3.3}$$

where FRETc are the mean FRET efficiency of each state, and A and w are the relative propensities of the states and the width of its distributions, respectively. The histograms were fitted globally to the equation above with the shared parameters $FRET_{C_0}, FRET_{C_1}, FRET_{C_2}, FRET_{C_3}$, w_1, w_0, w_2, w_3 and free parameters A_0, A_1, A_2, A_3 for all different GuHCl concentrations. The shared parameters obtained from the fit are as follows: $FRET_{C_0} = 0\,\% \pm 0.1\,\%$, $FRET_{C_1} = 26.0\,\% \pm 0.3\,\%$, $FRET_{C_2} = 62.2\,\% \pm 1.9\,\%$, $FRET_{C_3} = 75.7\,\% \pm 0.3\,\%$ and $w_0 = 8.2\,\% \pm 0.2\,\%$, $w_1 = 24.1\,\% \pm 0.6\,\%$, $w_2 = 22\,\% \pm 2$ and $w_3 = 14.9\,\% \pm 0.6\,\%$.

As demonstrated in Fig. 3.1, and as is evident from the results of the fit, both models can explain the data at this stage. However, the two interpretations of the results are quite different in terms of energy landscape of the folding process. A continuous shift of the FRET efficiency implies a continuous conformational change of the folded state, which can be conceived as a shift of the free energy minimum due to the folded state. On the other hand, the existence of an intermediate state implies an additional local minimum on the free energy landscape. A comparison between the two model energy landscapes is shown in the cartoon of Fig. 3.2.

In the analysis on freely-diffusing molecules, a threshold of 35 photons was used as a criterion for burst selection. As a result of the low number of photons, the distributions are quite broad, and therefore difficult to interpret. To observe two possible separate peaks resulting from folded and intermediate states before the transition midpoint, a lower statistical noise/a higher number of photons per molecule is required. The following experiments were carried out to reduce statistical noise.

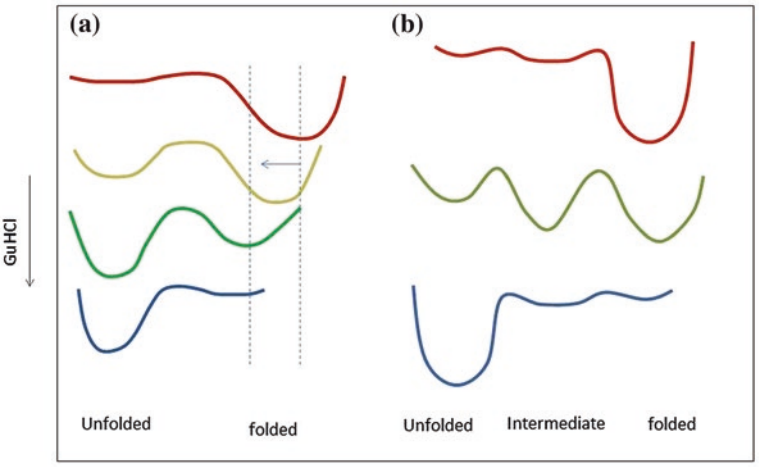

Fig. 3.2 The SM-FRET histograms of CTPR31C-Alexa at various GuHCl concentrations. Global fit of the histogram to (**a**) a model with three populations; a population with a varying FRET value (*red*), unfolded state (*blue*) and donor-only species (*black*), (**b**) to a model with four populations; folded (*red*), unfolded (*blue*), intermediate state (*olive green*) and donor only species (*olive green*)

3.2 Measurements on Diffusing Encapsulated CTPR31c-Alexa Molecules

To increase the observation time for each molecule and, as a result, to increase the number of photons per molecule, the diffusion of the protein was limited by encapsulating the individual molecules in 100 nm vesicles. Since the vesicles are much larger

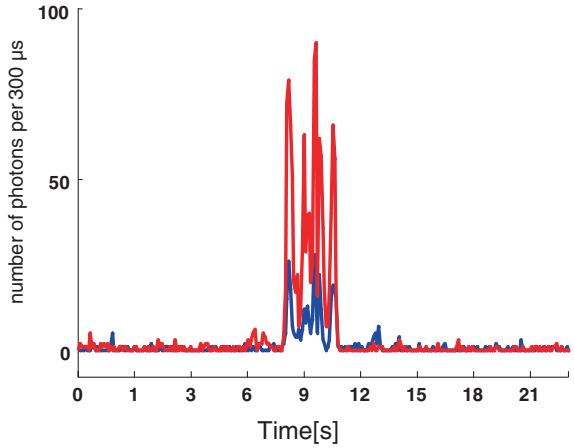

Fig. 3.3 Number of photons per 300 µs as a function of time for a vesicle-encapsulated protein molecule diffusing through the focus. *Blue* donor, *red* acceptor

than the protein, their diffusion rate is much slower. However, due to the Gaussian distributed laser intensity, the intensity of the bursts fluctuates even inside the vesicles. Setting the threshold based on the total intensity of the donor and acceptor channels results in a division of each burst into several smaller bursts (Fig. 3.3).

Nonetheless, this method results in a high number of photons per mini burst, enabling us to set the threshold as high as 100 photons. Such a high threshold decreases the statistical noise, making the distribution for each state narrower. The histograms obtained in these measurements (Fig. 3.4) are similar to those obtained from the freely-diffusing molecules, only narrower. If the broadening in the previous histograms were due to the existence of two populations and not one, then

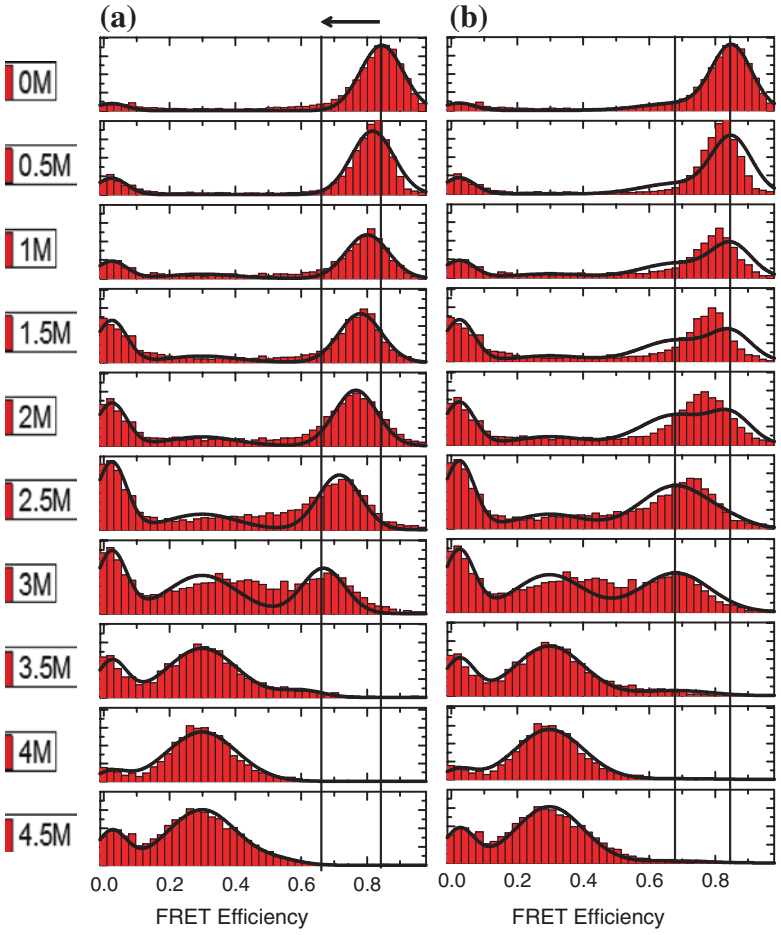

Fig. 3.4 The SM-FRET histograms of CTPR31C-Alexa at various GuHCl concentrations per 100 photons. Global fit of the histogram to (**a**) a model with three populations; a population with a varying FRET value, unfolded state, and donor-only species, (**b**) to a model with four populations; folded (FRET value of 85 %), unfolded, intermediate state (fret value of 68 %), and donor only species

decreasing shot noise should have resulted in two separate narrow distributions. Histograms of FRET efficiency, with bins of 100 photons, do not indicate the existence of two populations (Fig. 3.4b).

A continuous change of the FRET efficiency

Similarly to Sect. 3.1, Eq. 3.2, a three-Gaussian model with an unshared value for the folded state was used to fit the histograms globally (Fig. 3.4a).

$FRET_{C_2}$ of the folded state shifts from 85 to 68 % as a function of increasing GuHCl concentrations. The shared parameters from the fit are $FRET_{C_0} = 2.4\,\% \pm 0.1\,\%$, $FRET_{C_1} = 29.9\,\% \pm 0.2\,\%$, $w_0 = 9.1\ \% \pm 0.3\ \%$, $w_1 = 21.5\ \% \pm 0.5$, and $w_2 = 12.9\ \% \pm 0.2\ \%$.

Testing for the existence of a stable intermediate state

Similarly to Sect. 3.1, Eq. 3.3, a global fit of the histograms to a four-Gaussian model with shared values for folded and intermediate states was performed and is presented in Fig. 3.4b. The mean FRET values of the folded and intermediate states ($FRET_{C_3}$ and $FRET_{C_2}$), were fixed at 85 and 68 %, which are the peak centers of the FRET histogram at 0 M and 3 M GuHCl, respectively. The obtained shared parameters from the fit are: $FRET_{C_0} = 2.4\,\% \pm 0.2\,\%$, $FRET_{C_1} = 29.8\,\% \pm 0.4\,\%$, $w_0 = 9.2\ \% \pm 0.4\ \%$, $w_1 = 20.5\ \% \pm 0.7\ \%$, $w_2 = 20.9\ \% \pm 1.4\ \%$ and $w_3 = 12.1\ \% \pm 0.4\ \%$. A significant deviation of the fit from the histogram at intermediate GuHCl concentrations was found. This deviation is a strong indication of the existence of only one population and rules out the possibility of a stable intermediate state.

3.3 Ruling Out Motional Narrowing

Motional Narrowing model

A fast interchange between folded and intermediate states may result in a peak in-between the folded and intermediate peaks in the histograms. This peak is the outcome of averaging of folded and intermediate populations during the sampling time. This is similar to 'motional narrowing,' a well-known spectroscopic phenomenon. Figure 3.5 shows a simulation of FRET histogram for two interconverting populations for various time bins. Two populations with FRET values of 50 and 80 % where chosen to interconvert at a rate of 10,000/s. Binning the data in longer time bins results in averaging over the FRET populations during the bin time and appearance of an additional FRET population.

To rule out motional narrowing, we generated histograms with 300 μs time bins, shown in Fig. 3.6. Evidently, binning the data in 300 μs bins does not separate the two peaks, indicating, that if there were an interchange between populations, it would be much faster than this time. To slow down the folding dynamics, measurements were repeated at a lower temperature, 10 °C. A narrow distribution with a single peak was observed also under these conditions, demonstrating that

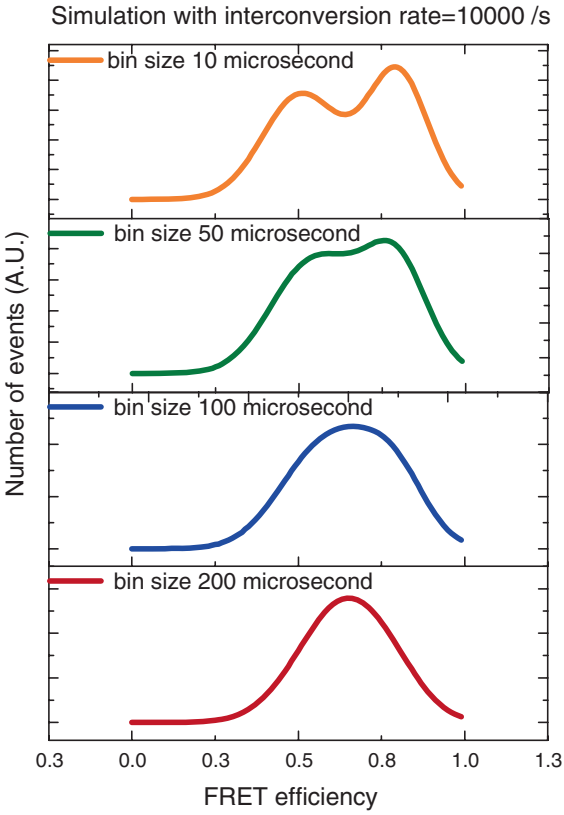

Fig. 3.5 Simulation of the histogram of FRET obtained from two interchanging populations with FRET values of 50 and 80 % using time bins of different lengths

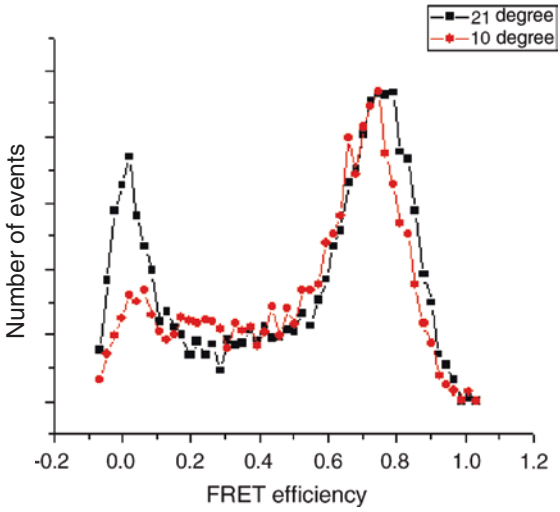

Fig. 3.6 Normalized histograms of FRET efficiency calculated with 300 μs time bins of diffusing vesicles at 21 °C s and 10 °C in 2 M GuHCl

if there is motional narrowing, it involves a rate that is much faster than 1/300 μs even at this temperature.

Gopich and Szabo analyzed the SM-FRET experiment in the presence of motional narrowing and showed that the histograms obtained under this situation can be represented as three Gaussians with *FRET* values of $FRET_{11}$, $FRET_{22}$, and $FRET_{12}$ [2]. The amplitudes of the three Gaussians are related to the rate constants of interconversion between the two conformations:

$$folded - Intermediate complex peak = \left(c_{11} \frac{e^{-\frac{(x-FRET_{11})^2}{2\sigma_{11}^2}}}{\sqrt{2\pi\sigma_{11}^2}} + c_{22} \frac{e^{-\frac{(x-FRET_{22})^2}{2\sigma_{22}^2}}}{\sqrt{2\pi\sigma_{22}^2}} + c_{12} \frac{e^{-\frac{(x-FRET_{12})^2}{2\sigma_{12}^2}}}{\sqrt{2\pi\sigma_{12}^2}} \right) \tag{3.4}$$

$FRET_{11}$ and $FRET_{22}$ are the FRET efficiencies of folded and intermediate states, with coefficients:

$$c_{22} = p_2 exp(-k_{12} * t), c_{11} = p_1 exp(-k_{21} * t), c_{12} = 1 - c_{11} - c_{22}; \tag{3.5}$$

where t is the time bin for the calculated FRET, k_{21} and k_{12} are the conformational rates from folded to intermediate, and from intermediate to folded state respectively, and p_1 and p_2 are propensities of folded and intermediate states.

$$p_2 = (1 - p_1), p_1 = \frac{k_{12}}{k_{12} + k_{21}} \tag{3.6}$$

$FRET_{12}$ is define by:

$$FRET_{12} = FRET_{11} f_{12} + FRET_{22} f_{21} \tag{3.7}$$

$$f_{21} = 1 - f_{12}, f_{12} = \frac{p_1 - c_{11}}{c_{12}} \tag{3.8}$$

Assuming that shot-noise is the only source of broadening of the histogram, the variance of each peak is given by:

$$\sigma_{ii}^2 = \frac{FRET_{ii}(1 - FRET_{ii})}{n} i = 1, 2, 3 \ldots \tag{3.9}$$

where, n is the number of photons per binning time.

$$\sigma_{12}^2 = FRET_{12} * (1 - FRET_{12})/n + \sigma c_{12}^2/n \tag{3.10}$$

$$\sigma c_{12}^2 = (FRET_{11} - FRET_{22})^2 \left(f_{12} f_{21} - \frac{p_1 p_2}{c_{12}} \Phi((k_{12} + k_{21})t) \right) \tag{3.11}$$

$$\Phi(\tau) = 1 - 2\frac{(\tau + exp(-\tau) - 1)}{\tau^2} \tag{3.12}$$

To find which inter-conversion rate between the folded and intermediate states leads to complete merging of the two peaks in the FRET histogram, due to averaging during the sampling time, we fit the histograms globally using the Gopich-Szabo motional narrowing model. To fit the FRET Efficiency Histogram (FEH), the peak representing the unfolded population, and the peak representing the molecules labeled only by donor dye, should be added to the Folded-intermediate complex peak.

$$FEH(E) = A\left(c_{11}\frac{e^{-\frac{(x-FRET_{11})^2}{2\sigma_{11}^2}}}{\sqrt{2\pi\sigma_{11}^2}} + c_{22}\frac{e^{-\frac{(x-FRET_{22})^2}{2\sigma_{22}^2}}}{\sqrt{2\pi\sigma_{22}^2}} + c_{12}\frac{e^{-\frac{(x-FRET_{12})^2}{2\sigma_{12}^2}}}{\sqrt{2\pi\sigma_{12}^2}}\right) + D\frac{e^{-\frac{(x-FRET_{33})^2}{2\sigma_{33}^2}}}{\sqrt{2\pi\sigma_{33}^2}} + G\frac{e^{-\frac{(x-FRET_{44})^2}{2\sigma_{44}^2}}}{\sqrt{2\pi\sigma_{44}^2}} \tag{3.13}$$

where, A, D, and G are the populations of folded-intermediate complex, the unfolded state, and the donor-only species, respectively.

Detailed balance between complex and unfolded states dictates:

$$P_U/(P_I + P_F) = D/A = K_3 \tag{3.14}$$

$$\frac{P_I}{P_F} = \frac{k_{21}}{k_{12}}, P_I + P_F = A \tag{3.15}$$

P_U, P_I and P_F are the populations of unfolded, intermediate state and folded state respectively.

K_3 is the equilibrium constant of the folded–intermediate complex and unfolded state.

Assuming exponential change of k_{12} and k_{21} and K_3 with GuHCl (gu):

$$k_{21} = k_{21_0} * \exp(-m_{21} * gu/RT)) \tag{3.16}$$

$$k_{12} = k_{12_0} * \exp(-m_{12} * gu/RT)) \tag{3.17}$$

k_{21_0}, k_{12_0} are the conformational rate constants from folded to intermediate, and

$$K_3 = K_{3_0} * \exp(-m_3 * gu/RT)) \tag{3.18}$$

from intermediate state to folded state at 0 M GuHCl respectively. K_{3_0} is the equilibrium constant between folded–intermediate complex and unfolded state at 0 M GuHCl. m_{21}, m_{12} are the constants of proportionality in the linear activation free energy relations for the kinetics of transition from folded to intermediate state and from intermediate to folded state respectively. m_3 is the constant of proportionality in the linear free energy of unfolding denaturation [3]. R is gas constant and T is the absolute temperature. Global fitting of all the histograms with the shared parameters $k_{21_0}, m_{21}, k_{12_0}, m_{12}, K_{3_0}, m_3, FRET_{11}, FRET_{22}, FRET_{33}, FRET_{44}, \sigma_{44}$, n, and the free parameters for components A and G in each histogram, is shown in

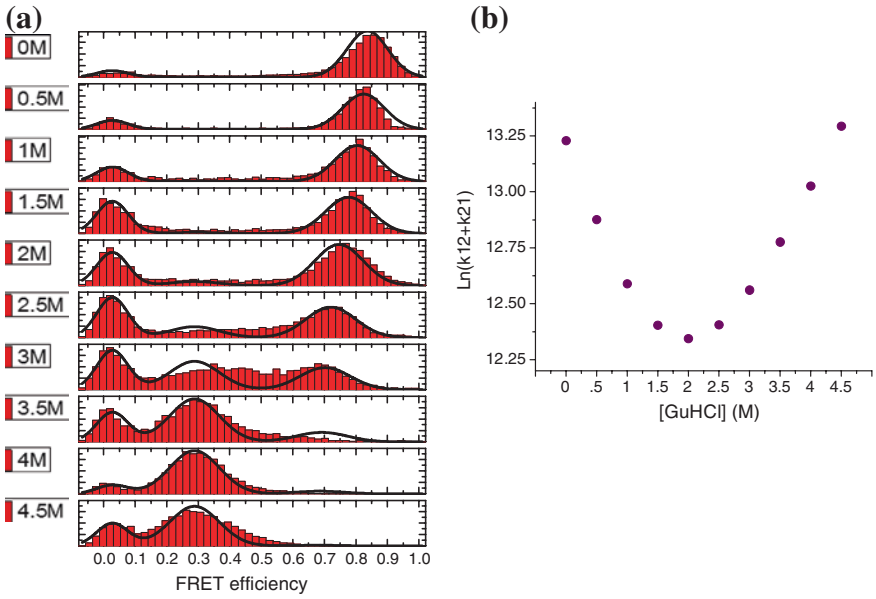

Fig. 3.7 **a** The fit of the histograms assuming fast dynamic. **b** The log of the rate of interchange of folded and possible intermediate state from the fit

Fig. 3.7a. In this case, data were binned according to constant number of photons, when average time (t) for obtaining 100 photons was 1.3 ms. n was chosen as a free parameter for considering the extra source of broadening in the histograms rather than the shot noise [4]. To simplify the fit, the FRET efficiency in 0 M GuHCl was forced to be 85 %.

Figure 3.7b shows the transition rates between folded and intermediate states obtained from the fit. At 2 M GuHCl, the sum of the rates ($k_{12} + k_{21}$) is higher than 0.23 MHz.

This result indicate that only dynamics as fast as 115 kHz or higher at 2 M can produce a single peak shift as obtained in our measurements. Using this rate, activation energy $|\Delta G_{\ddagger}|$ at the saddle point (2 M GuHCl) can be calculated from equation below.

$$k_{21} = k_{12} = \nu_{\kappa} exp(-\frac{|\Delta G_{\ddagger}|}{RT}) \tag{3.19}$$

where R is gas constant, T is the absolute temperature and ν_{κ} is characteristic vibration frequency along the reaction coordinate at the saddle point, with an empirical estimate for ν_{κ} of $\frac{10^6}{s}$ [3, 5]. Using the equation for the activation energy, we obtained $|\Delta G_{\ddagger}| < 2.16$ RT. We used the Boltzman equation $\exp(-\frac{|\Delta G_{\ddagger}|}{RT})$ to calculate the fraction of molecules in the transition state. The propensity of transition state at the saddle point of the energy landscape is calculated to be more than 10 %. This means that if there exist two separated folded and intermediate states,

the barrier between them is so shallow that these two states (folded and intermediate state) can exchange easily and can be counted as one state.

3.4 Free Diffusion Measurements on CTPR3NC-Alexa

To verify and substantiate the above results, we also carried out measurements on free diffusing molecules of the protein CTPR3NC-Alexa, which is labeled at "N" and "C" terminus. Bursts with more than 70 photons were chosen for analysis. Histograms of FRET efficiency per burst are shown in Fig. 3.8. Noticeably, this construct exhibits a similar shift of the population as in CTPR31C-Alexa. A fit to a three-Gaussian model (Eq. 3.2) is shown in Fig. 3.8.

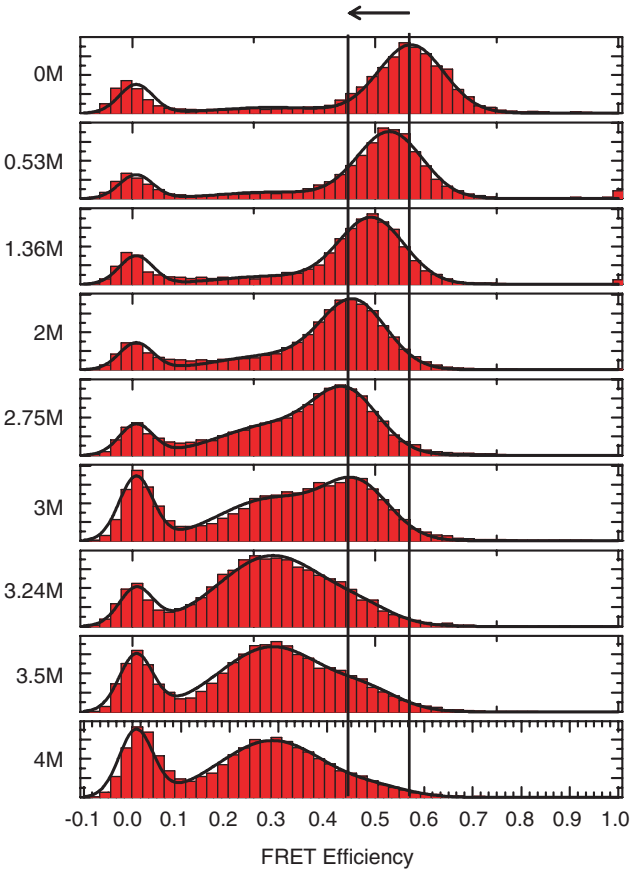

Fig. 3.8 The SM-FRET histograms of CTPR3NC-Alexa at various GuHCl concentrations per diffusing molecules at focus. Global fit of the histogram to (**a**) a model with three populations; a population with a varying FRET value, unfolded state and donor-only species

$FRET_{C2}$ (the folded state mean FRET value) shifts from 57 to 44 % following an increase in GuHCl concentration. The obtained shared parameters from the fit are $FRET_{C_0} = 0.7\,\% \pm 0.1\,\%$, $FRET_{C_1} = 28.7\,\% \pm 0.3\,\%$, $w_0 = 6.9\,\% \pm 0.1\,\%$, $w_1 = 21.8\,\% \pm 0.6\,\%$ and $w_2 = 13.4\,\% \pm 0.1\,\%$.

3.5 Measurements on Immobilized Molecules

Even though measurements on diffusing molecules already demonstrated the narrowness of the histograms, and ruled out the existence of two populations, we performed measurements on immobilized molecule to observe each molecule for even longer times, and detect any possible transition between states. CTPR31C-Alexa molecules were encapsulated in individual vesicles and immobilized using biotin-streptavidin chemistry on a lipid bilayer supported on glass in the presence of 2 M GuHCl. An automated single-molecule spectrometer was used for collection of large sets of single-molecule trajectories. During the acquisition, the laser power was set at 50 μWatt. Note that this is a very high power for such experiments, meant to obtain a large number of photons per second so as to push the time resolution to overlap that of the free diffusion experiment. The photon arrival times of both donor and acceptor fluorophores were registered in the experiment. These were later binned in 200 and 800 μs time bins, and the FRET efficiency was calculated bin by bin. Several sample trajectories are shown in Fig. 3.9. After systematic removal of various artifacts in the data set [6], 348 valid trajectories were selected for analysis. The average acceptor photobleaching time in these measurements was 15 ms.

Javadi and Main [7] performed kinetic measurements on CTPRa, a variant of the CTPR used by our laboratory, at 10 °C. The analysis of these measurements indicated a three-state folding behavior. The transition rates between the intermediate and folded states were found from a fit to a the three-state model to be 20 and 182 Hz, for the forward and backward direction, respectively. Based on the similarity between CTPR and CTPRa in structure and sequence, and considering that our measurements were performed at 25 °C, transitions between states were expected to be observed every few milliseconds. (Our analysis of the histograms already suggested that no such time scale should be seen in trajectories.)

To check if transitions between states with FRET efficiency higher than 50 % (*i.e.* not involving the unfolded state) occurred in our data, we analyzed each individual trajectory with 800 μs bin size, using a cumulative sum change-point algorithm [6]. Only in 9 molecules were transitions identified. An example of a FRET change detected by the change-point analysis is shown in Fig. 3.9a. The small changes in FRET value detected by change-point analysis in these nine molecules are clearly not the reason for the shift in the histograms.

To confirm that transitions were not overlooked because of binning of the data in 800 μs time bins, the correlation function (*expection value of* $(I(t)I(t+\tau))$) was calculated for each trajectory using a photon by photon algorithm developed by Jörg Enderlein [8]. The average autocorrelation (blue: donor, red: acceptor) and cross correlation (green) functions for all molecules are shown in Fig. 3.10.

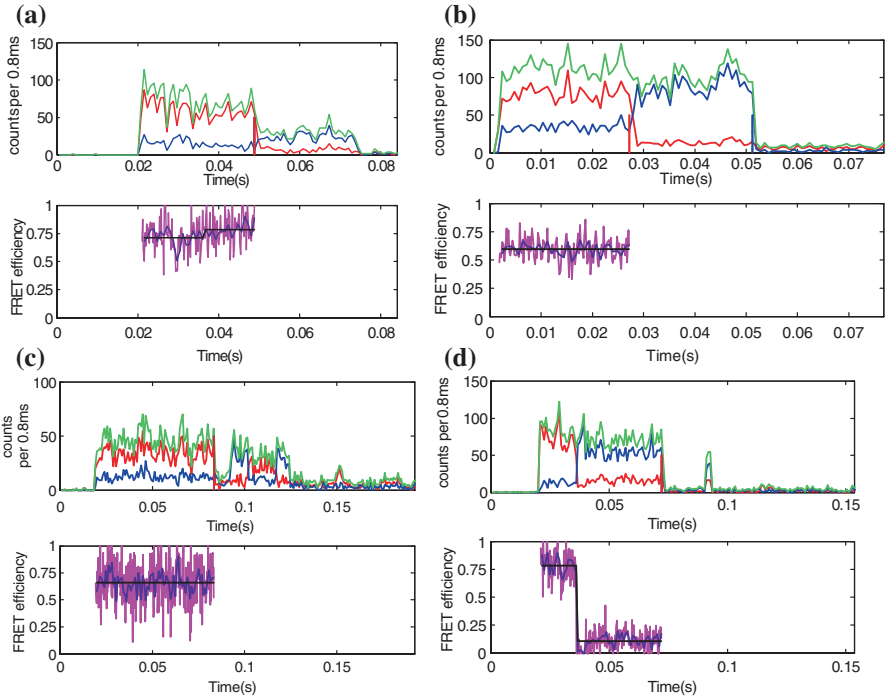

Fig. 3.9 Some examples of trajectories from measurements on immobilized single molecules at 2 M GuHCl. *Blue* donor, *red* acceptor, *green* total intensity per 800 µs bins, and the vertical *blue* and *red lines* represent the point of photobleaching of donor and acceptor, respectively. The FRET values are shown in the lower panel. *Purple* FRET per 800 µs time bin, *Magenta* FRET per 200 µs time bin, *black* Averages of FRET values between transition points identified by change-point analysis

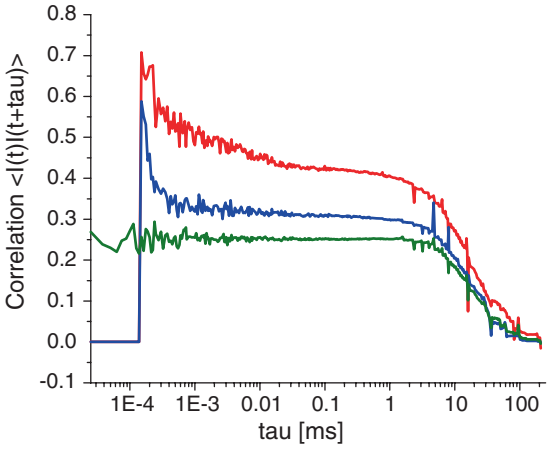

Fig. 3.10 Auto correlation of donor: *blue* & acceptor: *red*, and cross correlation (*green*) calculated photon by photon for each trajectory and averaged over all trajectories at 2 M GuHCl

FRET changes should result in a negative cross correlation between donor and acceptor channels. Therefore, fast conformational dynamics should appear as an increase in cross correlation function with time and a decrease in autocorrelation [9]. The absence of an increase in cross correlation function means that the FRET efficiency does not significantly fluctuate temporally in 2 M GuHCl (see Sect. 3.6).

3.6 FRET-FCS Measurements

In Sect. 3.5 we established, by correlation analysis of FRET trajectories of immobilized molecules, that fast dynamics involving the putative folded and intermediate states cannot be observed. However, in these measurements, the correlation functions were calculated relying on a relatively small number of selected molecules. Moreover, the auto-correlation functions presented there suffered from the well-known after-pulsing effect [10], as they were calculated from the data registered by a single detector. In order to follow FCS-FRET in a more robust approach, and to make sure the selection does not affect our results, FCS measurements were carried out on CTPR31C-Alexa variant molecules freely diffusing in the solution. The laser power was set to 30 μWatt. There are three processes affecting the correlation functions:

(1) The molecular diffusion, which is common to all functions.
(2) Formation of a triplet state, which is fluorophore specific and appears only in the auto-correlation function [11].
(3) Fluctuations of the FRET efficiency, which should appear in all functions. Such fluctuations lead to negative correlation between the donor and acceptor channels. Jumps between the two FRET values taking place in the time range of 1 μs and about 1 ms would result in a rise in the cross correlation [9]. The cross correlation function measured for CTPR3, however, did not indicate such fast dynamics (Fig. 3.11).

In accordance with the global fitting of the correlation curves to the Eqs. (3.20–3.22), when applying the diffusion time (τ_D) as a shared parameter, molecular diffusion alone can explain the shape of the cross correlation curve (Fig. 3.11). Other processes contributing to the cross correlation were not observed.

$$G(t) = \left(1 + \frac{t}{\tau_D}\right)^{-1} \tag{3.20}$$

$$AC_{DD}(t) = 1 + \frac{G(t)}{N} * \left(1 + \frac{A_{DD}}{1 - A_{DD}} * exp\left(-\frac{t}{\tau_{A_{DD}}}\right)\right) \tag{3.21}$$

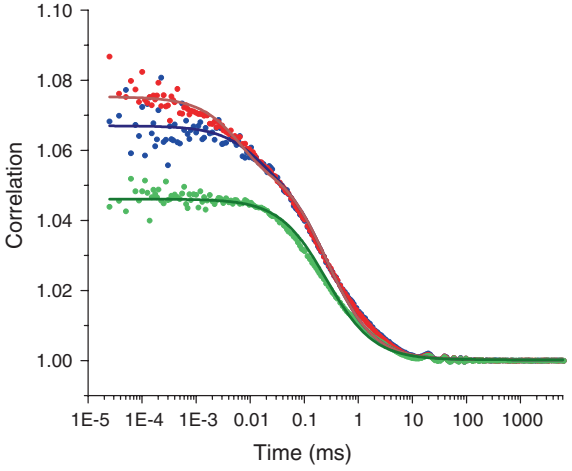

Fig. 3.11 FRET-FCS measurements. *Blue* donor auto correlation; *red* acceptor auto correlation; *green* donor acceptor cross-correlation. All *curves* are globally fitted with the assumption of lack of dynamics in the measured time range

$$AC_{AA}(t) = 1 + \frac{G(t)}{N} * \left(1 + \frac{A_{AA}}{1 - A_{AA}} * exp\left(-\frac{t}{\tau_{A_{AA}}}\right)\right) \tag{3.22}$$

$$CC(t) = 1 + \frac{1}{N} * G(t) \tag{3.23}$$

In these equations AC_{DD} is the donor auto-correlation, AC_{AA} is the acceptor auto-correlation, CC is cross correlation between the donor and acceptor, $\tau_{A_{DD}}$ *and* $\tau_{A_{AA}}$ are the relaxation times for singlet-triplet transition for donor and acceptor fluorophores respectively, A_{DD} and A_{AA} are the populations of the triplet state for donor and acceptor fluorophores respectively and N is the number of molecules within the focus.

If there is a dynamic change of FRET efficiency from 85 to 68 %, we would expect a rise of approximately 4 % in the cross correlation. This is calculated using the following equation [9]:

$$amplitude\ of\ change\ in\ cross\ correlation = \frac{(FRET_2 - FRET_1)^2}{(2 - FRET_1 - FRET_2)(FRET_1 + FRET_2)} \tag{3.24}$$

with the following assumptions: (1) the brightness for the two channels is identical and (2) the spectral leak from the donor emission to the acceptor channel is negligible.

In the correlation analysis on immobilized molecules (Fig. 3.10) even though the noise level is smaller than 4 %, no changes were observed in the cross correlation in the range of 1 and 100 ms, which indicates even faster dynamics than was estimated before in Sect. 3.3, Fig. 3.7b.

3.7 Correction of FRET Efficiency

To convert FRET efficiencies into distances, they should be corrected by the gamma factor in Eq. 3.2. The gamma factor is:

$$\gamma = \frac{Q_A C_A}{Q_D C_D}, \tag{3.25}$$

where Q_D, Q_A are the quantum yields of the acceptor and the donor dyes, respectively, and C_D and C_A are the collection efficiencies of the donor and the acceptor channels of the microscope, respectively.

We obtained the gamma factor using a set of measurements on immobilized molecules using the change in emission intensities before and after acceptor photobleaching:

$$\gamma = \frac{I_{A_{before}}}{I_{D_{after}} - I_{D_{before}}} \tag{3.26}$$

The average gamma factor obtained in these measurements is ~1.5.

The corrected value of the FRET efficiency can be written as

$$FRET_{corrected} = \frac{FRET_{measured}}{\gamma(1 - FRET_{measured}) + FRET_{measured}} \tag{3.27}$$

where

$$FRET_{measured} = I_A/(I_A + I_D). \tag{3.28}$$

The calculated corrected peak FRET efficiencies from the histograms are shown in Fig. 3.12.

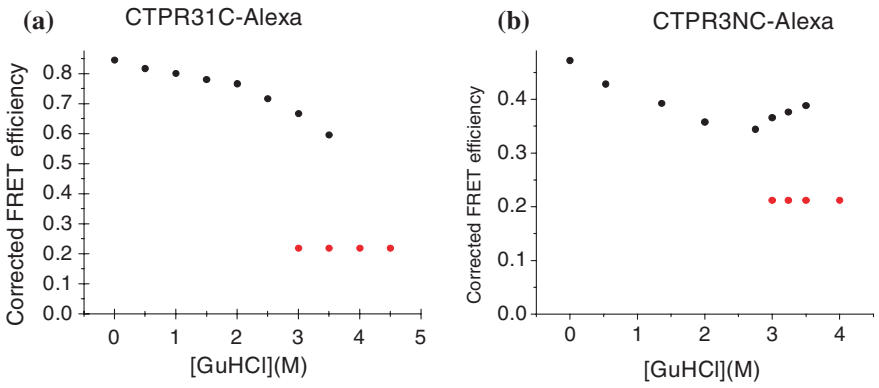

Fig. 3.12 Peak center for populations with a high FRET efficiency (*black*), and an unfolded population (*red*) from the global fit of the histograms to three populations for (**a**) CTPR31C-Alexa and (**b**) CTPR3NC-Alexa

3.8 Distance Calculations

To obtain some understanding of what changes in the distances can produce the changes in the FRET values that we obtained in the measurements, we translated the FRET efficiency to distances using the equation:

$$FRET_{corrected} = \frac{R_0^6}{R_0^6 + r^6} \tag{3.29}$$

Since the quantum yield of the dye might change when the dye is linked to the protein, the R_0 value of the dye pairs can be different for different proteins and even for different labeling positions. To calculate the R_0 value, first the quantum yield of the donor fluorophore linked to the protein at the C terminus was measured.

3.8.1 Determination of the Quantum Yield

The quantum yield of the fluorescence is defined as:

$$QY = Number\ of\ photons\ emitted / Number\ of\ photons\ \text{absorbed} \tag{3.30}$$

Since direct measurement of the quantum yield ('QY') of the sample is impractical, the sample was compared with a reference with a known quantum yield, and under identical excitation conditions.

$$QY = \alpha \times (Number\ of\ photons\ detected)/(\beta \times (Fraction\ of\ light\ absorbed)) \tag{3.31}$$

where α and β are constants related to the fraction of the emitted light entering the fluorimeter, and the intensity of the excitation source at the excitation wavelength, respectively.

In this equation, the number of detected photons is the integrated area under the fluorescence spectrum (F).

$$Fraction\ of\ light\ absorbed(A) = 1 - 10^{-OD} \tag{3.32}$$

where OD is the optical density at the excitation wavelength. For instance, for sample X:

$$QYx = \frac{\alpha F_X}{\beta A_X} \tag{3.33}$$

α and β are essentially identical for different samples except that the collection efficiency is proportional to the square of the index of refraction n^2.
Comparative measurements on sample X and a suitable standard reference 'R' can then be used to obtain the fluorescence quantum yield of X.

$$QYx = \frac{F_X}{F_R}\frac{A_R}{A_X}\left(\frac{n}{n_R}\right)^2 QY_R \tag{3.34}$$

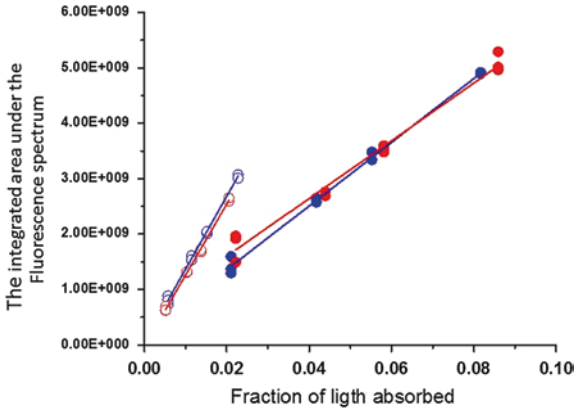

Fig. 3.13 Comparison of the slope of emission intensity as a function of absorption for CTPR3SL Alexa (*filled circle*), and fluorescein as a reference (*open circle*). Excitation wavelengths of 491 (*blue*) and 496 (*red*) were used. The lines are linear fits

Table 3.1 Distances and FRET efficiency for a protein with the last helix unfolded from lattice simulation

Excitation at	491 nm	496 nm
Slope of the fitted curve for fluorescein	1.31 E11	1.28 E11
Slope of the fitted curve for CTPRSL-Alexa 488	5.75 E10	5.21 E10
CTPRSL-Alexa488 slope/fluorescein slope	0.43	0.47
Quantum yield CTPRSL-Alexa488	0.41	0.45

Fluorescein at 0.1 M NAOH, 496 nm excitation, and at 22 °C, with a quantum yield of 95 % was used as a reference.

Preparation and measurements were repeated 3 times, and from each measurement, four dilutions of different concentrations were prepared. The emission spectrum of each sample was measured with a 1 nm excitation slit and a 5 nm emission slit in steps of 1 nm. The integrated area under a fluorescent emission spectrum as a function of OD was used to compare the samples with the reference (Fig. 3.13). Measurements were carried out with excitation wavelengths of 491 (blue) and 496 (red) on CTPR3SL-Alexa at 25 °C and on fluorescein at 22 °C.

A linear fit was drawn to find the slope of the fitted curve, and to compare the samples to the reference. Results of the fit were shown in the table below (Table 3.1).

3.8.2 Calculation of the Förster Distance

Förster Distance is R_0 can be calculated from equation below,

$$R_0^6 = \frac{9{,}000(\ln 10)\kappa^2 Q_D}{128\pi^5 N n^4} J(\lambda) \tag{3.35}$$

In these equations, κ^2 describes the orientation factor of the dyes. Q_D is the quantum yield of the donor fluorophore, N is the Avogadro number, and n is the index

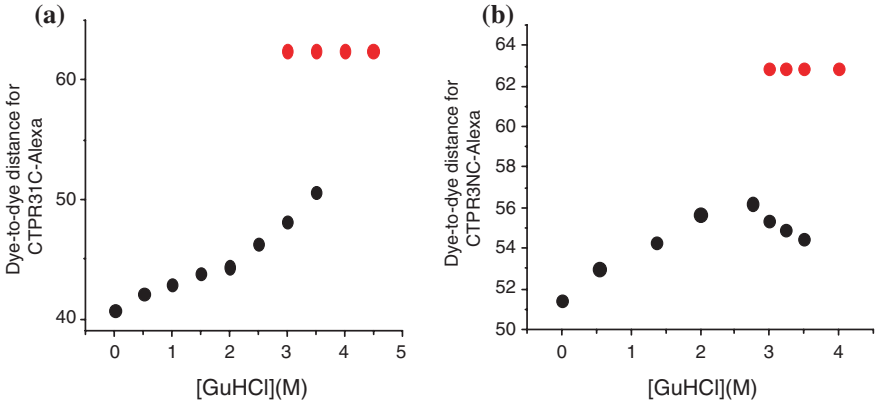

Fig. 3.14 Calculated dye-to-dye distance as a function of GuHCl for two populations in single-molecule measurements for (**a**) CTPR31C-Alexa and (**b**) CTPR3NC-Alexa

of refraction of the media. J is the overlap between the donor emission and the acceptor absorption spectrum.

$F_D(\lambda)$ is the emission spectrum of the donor fluorophore normalized by its area,

$$J(\lambda) = \int_0^\infty F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda \tag{3.36}$$

and $\varepsilon_A(\lambda)$ is the absorption coefficient spectrum of the acceptor fluorophore.
κ^2 is usually assumed to be 2/3, which is appropriate for dynamic random averaging of the donor and acceptor fluorophores [12]. Estimated error in distances due to nonrandom dynamic averaging of fluorophore estimated based on anisotropy of fluorophore is about 10 % [13]. The value of 0.43 was used for determining the quantum yield of Alexa 488 bound to CTPR. Therefore, a Förster distance of 50.5 Å was calculated for CTPR31C-Alexa and CTPR3NC-Alexa.

3.8.3 Distances Between the Dyes

The Förster distance was used to calculate the dye-to-dye distance, and the results are shown in Fig. 3.14 for the two protein samples. In this calculation we assumed that the variation in the distribution of distances was neglected.

3.9 Lattice Simulations of a Self-Avoiding Freely-Jointed Polymer Model

To obtain an estimate for the expected dye-to-dye distance changes in partially unfolded proteins predicted by the Ising model, we performed a simple computer simulation on a 3D cubic lattice.

Self-Avoiding Random Walk (SARW) simulation was carried out to simulate the unfolded portion of the partially unfolded proteins in the 3D lattice. In this simulation, biased modeling method was used to reduce the simulation time. Biased modeling excludes the self-intersection steps, and based on the number of the excluded steps assigns a weight factor to each walk [14].

To estimate the upper and lower limit of lattice unit size we used the simulation of denatured protein with 50 residues. A fit to the radius of gyration (R_g) for proteins in a 'good' solvent measured by small angle X ray scattering (SAXS) yielded the relation $R_g \approx 2\text{Å} \times N^{59}$ [15], where N is the number of units in the chain. For Gaussian random chains, the following relation exists between the end-to-end distance (R_{ee}) and the radius of gyration.

$$R_g^2 = \frac{\langle R_{ee} \rangle^2}{6} \tag{3.37}$$

Therefore, $\langle R_{ee} \rangle$ for 50 amino acid residues will be 49.3 Å.

The SARW simulation was carried out for a chain of 50 residues 10,000 times. The mean end-to-end distance obtained from simulation is 10.07 × cell units. By comparing the value obtained from simulation and the end-to-end distance from the model (equation above), an upper limit for the cell unit size was obtained, 4.9 Å. This size can be used to simulate the denatured state in a good solvent. In poor solvents, the denatured state of the protein collapses, [16] and smaller sized units should be used for simulation. The lower limit for the size of the cell unit is 3.8 Å, which is the average distance between adjacent C_α s in natural form.

The simulations were run for a series of lattice unit sizes to simulate the denatured state and partially unfolded state of CTPR3. The distances between the two points of the protein were obtained for each run, knowing that R_0 is 50.5 Å. The FRET efficiency was calculated for each simulation. The mean FRET value and the mean distance were calculated as the weighted average of the FRET values and the distances were obtained for 5,000 simulations and are presented in the tables below for various unit sizes of lattice.

The results of SARW simulation of the fully denatured CTPR are shown in Table 3.4. The most populated partially unfolded state predicted from the Ising model is the form with only the peripheral helix unfolded. To simulate the folded portion of the protein, the crystal structure of CTPR3 (PDB entry: 1NA0) was used, and C_α coordinates were projected onto the 3D lattice. The unfolded peripheral helix was simulated using SARW with an additional restriction: the occupied sites of C_α of the folded portion were not allowed to be reoccupied during the simulation of the unfolded peripheral helix. The results of this simulation are shown in Table 3.3. Comparing the result of the simulation with the expected FRET efficiency for each calculated from the distances between dyes in the crystal structure (Table 3.2), we concluded that unfolding of one helix changes the FRET efficiency of CTPR31C by about 27 %. However, the FRET efficiency of CTPR3NC does not change with the unfolding of one helix. This is due to the 3D structure of CTPR3.

Table 3.2 Mean distance values and mean FRET efficiency values measured from the crystal structure

	CTPR1C	CTPRNC
Mean dye-to-dye distances (Å)	36.5	50.2
Mean FRET efficiency	0.87	0.51

Table 3.3 Distances and FRET efficiency for a protein with the last helix unfolded from lattice simulation

Unit cell size (Å)	CTPR31C mean dye-to-dye distances (Å)	CTPR3NC mean dye-to-dye distances (Å)	CTPR31C mean FRET efficiency	CTPR3NC mean FRET efficiency
3.8	39.2	43.2	0.76	0.67
3.9	40.6	43.5	0.71	0.65
4	41.3	44.1	0.73	0.67
4.1	40	42.7	0.66	0.55
4.2	40.5	43.2	0.68	0.57
4.3	42.9	47.8	0.63	0.52
4.4	44.1	48.9	0.67	0.55
4.5	41.8	46.8	0.76	0.67
4.6	43	48.1	0.71	0.65
4.7	44.2	49.4	0.73	0.67
4.8	45.4	50.8	0.66	0.55
4.9	46.5	48.1	0.68	0.57
5	43.8	48.9	0.63	0.52

Table 3.4 Distances and FRET efficiency for the unfolded state from the lattice simulation

unit cell size (Å)	CTPR31C mean dye-to-dye distances (Å)	CTPR3NC mean dye-to-dye distances (Å)	CTPR31C mean FRET efficiency	CTPR3NC mean FRET efficiency
3.8	54.4	64	0.51	0.36
3.9	55.8	67	0.45	0.32
4	58.5	71	0.43	0.31
4.1	58.2	69.1	0.43	0.30
4.2	58.2	69.5	0.42	0.26
4.3	60.7	72.4	0.36	0.25
4.4	64.5	76.5	0.33	0.24
4.5	63.3	75.4	0.37	0.28
4.6	63	73.2	0.33	0.22
4.7	66.7	79.4	0.29	0.21
4.8	67.9	81.2	0.29	0.20
4.9	70.1	84.3	0.31	0.21
5	70.4	82.4	0.30	0.19

Remembering that in single-molecule measurements we observed a shift of the FRET population for both FRET pairs, the results of the simulation suggest that a continuous shift of the FRET value with low GuHCl concentrations for CTPR3 is not due to the unfolding of one helix.

3.10 Analysis of NMR Measurements on CTPR

These measurements were carried out by Dr. Aitziber Cortajarena and reanalyzed by the author of this thesis. NMR spectra were recorded on a Varian Unity Plus 600 MHz spectrometer, with 15 N-labeled protein at 1 mM protein concentration. 15 N labeled CTPR protein samples were lyophilized in water and resuspended in an equal volume of buffer in D_2O (Cambridge Isotope laboratories, Cambridge, UK) (150 mM NaCl, 50 mM phosphate, pH 6.8) at different Gu-HCl (Ultrapure, Fluka, Buchs, Switzerland) concentrations (0, 0.25, 0.75, 1, 1.25 and 1.5 M). HSQC spectra were recorded at 25 °C using 32 increments of 2,048 data points and 16 transients. The spectral widths were 7,000 Hz in the 1H dimension and 2,000 Hz in the 15 N dimension. The total acquisition time of each spectrum was 23 min. 1H-15 N HSQC spectral assignments of CTPR2 were previously published [17].

The 1H-15 N HSQC spectrum provides the correlations between the nitrogens and amide protons in the backbone peptide of the protein, and each amide provides a peak in the HSQC spectra. A continuous shift of the spectrum is observed

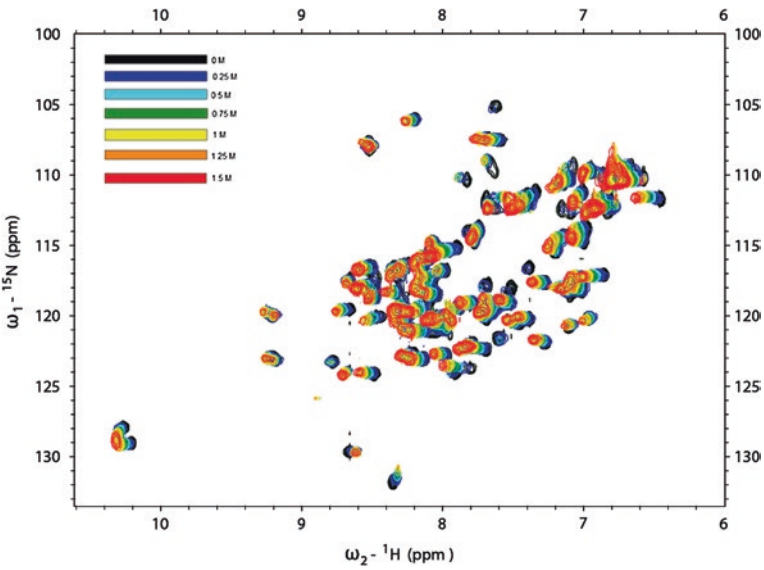

Fig. 3.15 Shift of HSQC spectrum of CTPR2 with GuHCl

in HSQC measurements of CTPR2 (Fig. 3.15). The shift of the spectra is indication of changes in the local environments of the Nitrogen-Hydrogen (N–H) bonds. As is evident in Fig. 3.15, the spectrum change is not particular to few amino acids but is a general feature for most peaks. Due to the overlap of peaks in 1.5 M GuHCl, it is not trivial to identify all of the amino acids in 1.5 M GuHCl for quantitative analysis.

To test whether there is any particular region of the protein that changes more than others, we used the hydrogen exchange data for CTPR3 previously reported by our collaborators [18]. In these measurements, Protection Factors (PFs) were calculated as the ratio of the observed exchange rate for a particular amino acid in the folded protein to the tabulated exchange rate of that amino acid in an unstructured peptide. We calculated the relative changes of PF for each residue using the equation below.

$$Relative\ Change\ of\ PF\ = \frac{PF_{0M} - PF_{1.5M}}{PF_{0M}} \tag{3.38}$$

The results are shown in Fig. 3.16, where they are also mapped on the crystal structure of CTPR3 using the following color code: dramatic (with a change of more than 90 %), high (with a change of more than 75 % and less than %90) and low (with a change of protection value less than %75) to red, green, and yellow color respectively (Fig. 3.16b).

As is evident, in some helices there is a dramatic change of protection factor for the central part of helix while residues in the cap have only a small change. This can be true only if the changes in the tertiary structure occur while the secondary structure is still preserved.

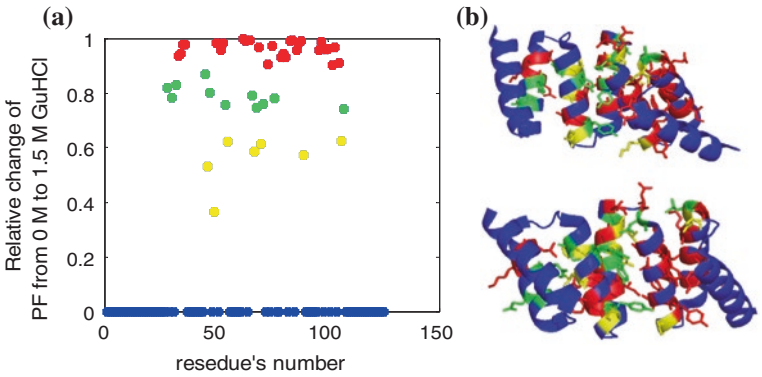

Fig. 3.16 **a** Relative changes of PF from 0 to 1.5 M GuHCl for each amino acid. **b** Map of the changes on the crystal structure of CTPR3, with the following color code: *red*-a change of more than 90 %, *green*-a change of more than 75 % and less than 90 %, and *yellow*-a change of protection value less than 57 %. The *blue* color marks residues whose rate of exchange is so fast that protection value could not be measured for them

3.11 Control Measurements

A set of control measurements was performed to rule out photo-physical effects as the reason for the observed shift of FRET efficiency from 85 to 68 %. Since $FRET = (1 + (I_D/I_A))\hat{}(-1)$, the continuous shift can be explained if the ratio of (I_D/I_A) is changed from 0.18 to 0.47. Such a change may be reasonably assigned to a photo-physical effect only if the quantum efficiency of one or both dyes changes dramatically.

In the measurements reported below, donor and acceptor quantum yields remained essentially unchanged as function of GuHCl concentration, ruling out the change of the quantum yield of one of the dyes as the reason for the shift.

3.11.1 Acceptor Quantum Yield

Figure 3.17a shows the emission spectra of CTPR31C-Alexa excited at 550 nm (direct acceptor excitation), for different GuHCl concentrations. Figure 3.17b demonstrates the peak intensity of the emission spectra as a function of GuHCl. The change of intensity with GuHCl indicates a mild quenching of the acceptor fluorophores by the change in the local environment between the folded state and unfolded states. However, in low GuHCl concentrations (0–2 M), where the shift in the FRET histograms occurs, the quantum efficiency of the fluorophore is essentially constant.

3.11.2 Donor Quantum Yield

Figure 3.18 demonstrates the emission spectra of CTPR3SL-Alexa excited at 465 nm, for different GuHCl concentrations. Figure 3.18b shows the peak intensity of the emission spectra as a function of GuHCl. Except for a quite small

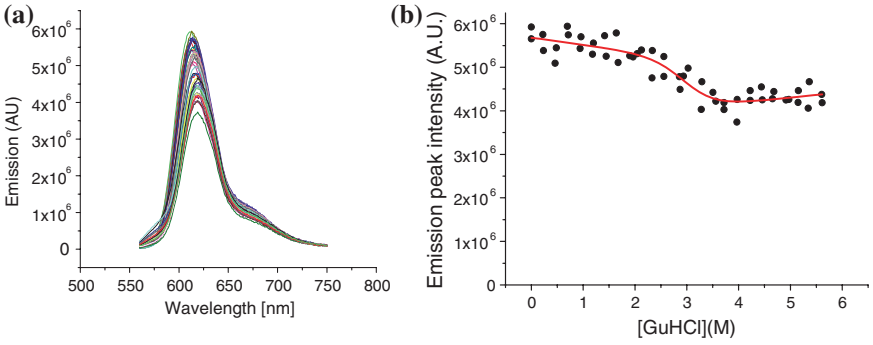

Fig. 3.17 (**a**) Emission spectra of CTPR31C-Alexa excited at 550 nm (direct acceptor excitation). (**b**) Peak intensity of (**a**) as function of GuHCl concentration fitted to sigmoidal curve

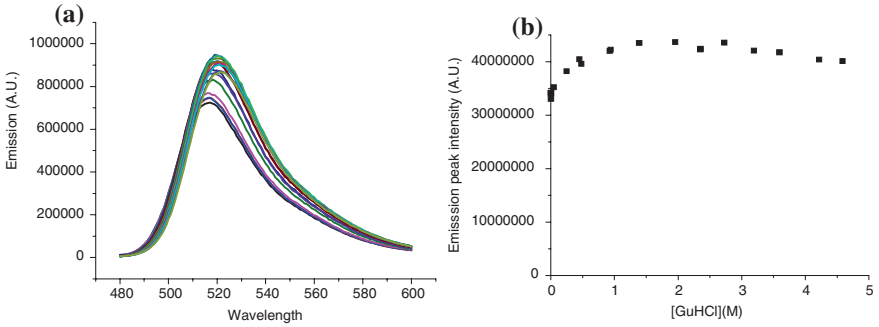

Fig. 3.18 (**a**) Emission spectra CTPR3SL-Alexa488 excited at 465. (**b**) Peak intensity of (**a**) as a function of GuHCl concentration remains relatively unchanged

increase in the quantum yield of the donor in 0–0.5 M GuHCl, the presence of the denaturant did not affect the donor quantum yield.

3.11.3 Fluorescence Anisotropy

The bulk fluorescence anisotropy of the dyes attached to the protein was measured at different GuHCl concentrations (Fig. 3.19). The low values of the fluorescence anisotropy, and the minimal changes induced by GuHCl, indicate a freedom of re-orientation of the dyes, independently of the denaturant concentration.

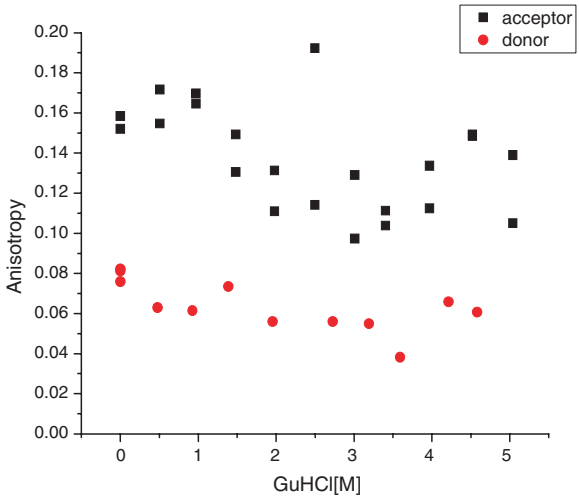

Fig. 3.19 Fluorescence anisotropy as a function of GuHCl concentration. *Blue dots* represent the measured values for CTPR3SL-Alexa 488 excited at 465 nm. *Red dots* represent the measured values of CTPR31C-Alexa protein molecules excited at 560 nm

3.11.4 Bulk Denaturation Curve

The denaturation curve of CTPR31c-Alexa was measured using FRET efficiency. The emission spectra were measured using donor excitation at 465 nm for different GuHCl concentrations.

FRET efficiency was calculated from Eq. 2.2, using the acceptor and donor emission peak intensities.

Figure 3.20 demonstrates the emission spectra of the double labeled protein with donor excitation. Figure 3.20b demonstrates the calculated FRET efficiency from Eq. 2.3, fitted to a two state folding model with linear baselines.

A continuous change of FRET in the denaturant range from 0 to 2 M was also clearly observed in the bulk measurements.

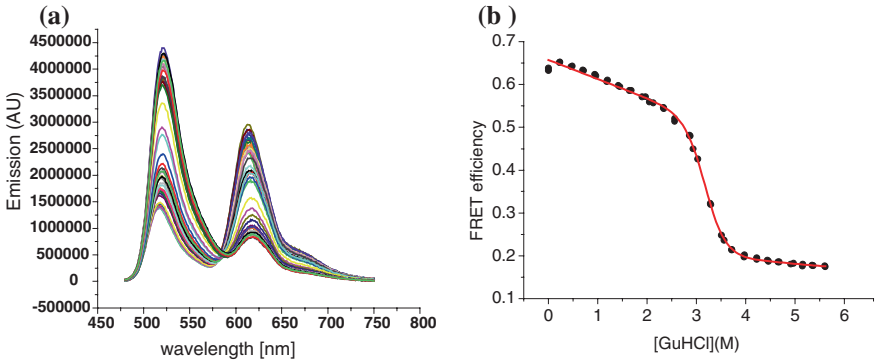

Fig. 3.20 **a** Emission spectra of CTPR31C-Alexa excited at 465 nm. **b** The calculated FRET efficiency from the emission spectrum (*black squares*) and sigmoidal curve fit (*red line*)

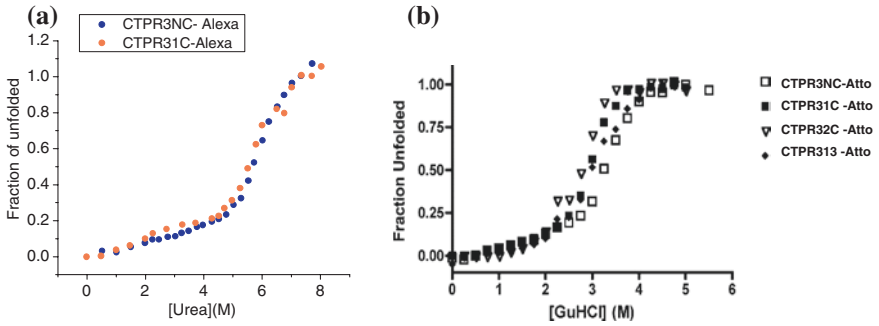

Fig. 3.21 Denaturation curves of CTPR3 labeled with different dyes and with different denaturant. All experiments confirm a significant increasing slope in the folding baseline

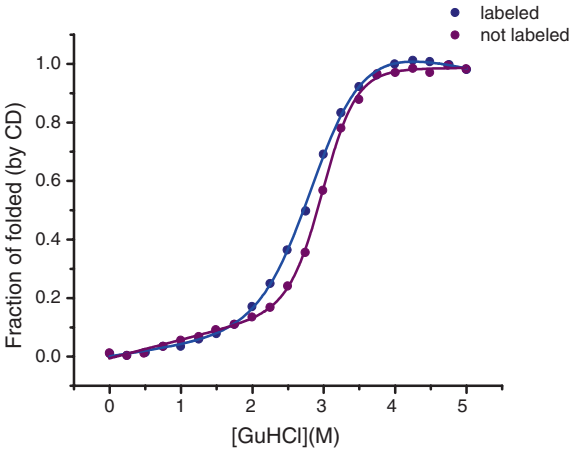

Fig. 3.22 Comparison of the unlabeled CTPR and CTPR31c_Alexa denaturation curves obtained using CD spectroscopy

3.11.5 FRET Measurements on CTPR Labeled by Different Dyes and at Different Sites

The following bulk measurements for other labeled CTPRs were carried out by Dr. Aitziber Cortajarena, and are shown in Fig. 3.21.

The change of FRET at low denaturant concentrations is clear for all species. Evidently, this effect is not denaturant-specific either.

3.11.6 Circular Dichroism Measurements

Denaturation curves were measured on CTPR31C-Alexa and on unlabeled CTPR3 using circular dichroism (Fig. 3.22). The similarity between CD curves of the labeled and unlabeled proteins confirms that the labeling did not affect the native structure of the protein significantly. However, labeling slightly destabilized the protein. A sigmoidal function describing a two-state folding was used to characterize the curves. The folding free energy is altered by 2.9 kcal/mol.

References

1. Schultz PG et al (2000) Single-molecule protein folding: Diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. Proc Natl Acad Sci USA 97(10):5179–5184
2. Gopich IV, Szabo A (2010) FRET efficiency distributions of multistate single molecules. J Phys Chem B 114(46):15221–15226

3. Fersht A (1999) Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. W.H. Freeman, New York. xxi, p 631
4. Kalinin S et al (2010) On the origin of broadening of single-molecule FRET efficiency distributions beyond shot noise limits. J Phys Chem B 114(18):6197–6206
5. Hagen SJ et al (1996) Diffusion-limited contact formation in unfolded cytochrome c: estimating the maximum rate of protein folding. Proc Natl Acad Sci USA 93(21):11615–11617
6. Pirchi M et al (2011) Single-molecule fluorescence spectroscopy maps the folding landscape of a large protein. Nat Commun 2:493
7. Javadi Y, Main ER (2009) Exploring the folding energy landscape of a series of designed consensus tetratricopeptide repeat proteins. Proc Natl Acad Sci USA 106(41):17383–17388
8. Wahl M et al (2003) Fast calculation of fluorescence correlation data with asynchronous time-correlated single-photon counting. Opt Express 11(26):3583–3591
9. Torres T, Levitus M (2007) Measuring conformational dynamics: a new FCS-FRET approach. J Phys Chem B 111(25):7392–7400
10. Krichevsky O, Bonnet G (2002) Fluorescence correlation spectroscopy: the technique and its applications. Rep Prog Phys 65:251–297
11. Price ES, DeVore MS, Johnson CK (2010) Detecting intramolecular dynamics and multiple Forster resonance energy transfer states by fluorescence correlation spectroscopy. J Phys Chem B 114(17):5895–5902
12. Lakowicz JR (1999) Principles of fluorescence spectroscopy. 2nd edn, Kluwer Academic/Plenum., New York. xxiii, p 698
13. Haas E, Katchalski-Katzir E, Steinberg IZ (1978) Effect of the orientation of donor and acceptor on the probability of energy transfer involving electronic transitions of mixed polarization. Biochemistry 17(23):5064–5070
14. Gould H, Tobochnik J (1996) An introduction to computer simulation methods. 2nd edn, Addison Wesley Publishing Company, Boston
15. Kohn JE et al (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. Proc Natl Acad Sci USA 101(34):12491–12496
16. Sherman E, Haran G (2006) Coil-globule transition in the denatured state of a small protein. Proc Natl Acad Sci USA 103(31):11539–11543
17. Main ER et al (2005) Local and long-range stability in tandemly arrayed tetratricopeptide repeats. Proc Natl Acad Sci USA 102(16):5721–5726
18. Cortajarena AL, Mochrie SG, Regan L (2008) Mapping the energy landscape of repeat proteins using NMR-detected hydrogen exchange. J Mol Biol 379(3):617–626

Chapter 4
Discussion

4.1 The Relevance of the Ising Model to Unfolding of CTPR

The Ising model was successfully applied by several authors to explain the folding of repeat proteins. For TPR proteins, each helix in the repeat is considered as an individual unit that may be folded or unfolded. Using single-molecule microscopy, we tested whether a partially unfolded state exists with one or more than one helix unfolded in the folding of CTPR3. In principle, the Ising model suggests the presence of 128 (2^7) states in the folding of 7 helices in CTPR3. However, many of these species may be unpopulated. The population of the folded or the unfolded state, or partially unfolded states, can be calculated using the free energy (G) [1]:

$$\mathrm{G} = k_b\mathrm{T} \sum_i (-\mathrm{J}\, S_i S_{\mathrm{i}+1} + \frac{\mathrm{m(gu - } x_c)}{2}\mathrm{S}_i), \tag{4.1}$$

where S_i is the spin state (1 or −1), and gu is the concentration of guanidine. The term $\frac{m(gu-x_c)}{2}$ refers to the stability of each helix. x_c is the guanidine concentration at the transition midpoint for each isolated helix, and J is the interaction between adjacent helices. The most populated intermediate state predicted by the Ising model is a structure whose peripheral helix is unfolded. In agreement with the Ising model, NMR hydrogen exchange measurements on CTPR show that the internal helices are more stable than the peripheral helices [2]. Kinetic measurements on a variant of CTPR show a roll-over in the Chevron plot [3], interpreted as the existence of an intermediate state. Calorimetric measurements on CTPR also show a deviation from the two-state models [4]. We used SM-FRET measurements on two CTPR variants, each labeled at a different pair of sites, to observe these intermediate states.

Histograms of FRET efficiency were obtained for each molecule during diffusion through the focus of a confocal microscope. These measurements were carried out in various concentrations of a chemical denaturant. We observed a gradual shift in the FRET efficiency of the folded state with the increase of denaturant

S. Cohen, *Single-Molecule Fluorescence Spectroscopy of the Folding of a Repeat Protein*, Springer Theses, DOI 10.1007/978-3-319-09558-5_4

concentration for both variants at mild denaturant concentrations, followed by an all-or-none transition to the unfolded state at higher denaturant concentrations. FCS measurements showed that there is no interchange between two FRET efficiency populations in the time scale slower than μs in mild denaturant concentrations. Thus the FRET efficiency shift can be interpreted as a continuous transition rather than a jump between two states.

To obtain the structural characteristics of the shift, we used NMR measurements. The NMR HSQC measurements on CTPR2 showed a global gradual shift of all amino-acid peaks in addition to the unfolding of the solvent helix following the increase of GuHCl concentration at mild denaturant concentration.

To quantify the relative change of each residue, we analyzed the results of hydrogen exchange measurements. It is known that unfolding of helices starts from the capping residues [5, 6]. However, analysis of hydrogen exchange measurements showed that the relative protection factor change of capping residues with denaturant is smaller than that of central residues. This finding indicates that the secondary structure is more conserved than tertiary structure in the first step of CTPR unfolding.

On the other hand, the Ising model, which does not consider the change of tertiary structure, predicts the existence of intermediate states which involve unfolded peripheral helix/helices. The population of these states depends on the parameters of the Ising model. Using the parameter set suggested previously for the folding of CTPR, based on fitting of experimental data [1], one predicts that less than 20 % of the population is in intermediate states. These intermediate states predicted by the Ising model were not observed as a separate peak in the smFRET efficiency histograms.. This might be due to the low population of the states or due to the similarity of the FRET efficiency in these states to that of the folded state.

The shift of FRET efficiency in low denaturant concentration was also observed in bulk denaturation curves generated using either urea or GuHCl as chemical denaturants (Fig. 3.21). This demonstrates that the shift of FRET efficiency in low denaturant concentrations is not specific to a particular labeling site, nor to a particular chemical denaturant. Even though the NMR spectrum of CTPR did not collapse in low guanidine concentrations, as one might expect for a molten globule intermediate [7, 8], we suggest that the continuous shift of the FRET efficiency distribution is related to the loosening of the tertiary structure, and the structure of this state is very similar to the molten globular state. Since loosening of the tertiary structure results in a change at the interface between helices, we propose modifying the Ising model to correct for this change. In the conventional model, the parameter J, which refers to the interface between helices, is a constant parameter and does not vary with denaturant concentration. We propose that the Ising model can be amended by assuming a continuous change of J with denaturant concentration.

4.2 Perturbation of Local Minima on the Free Energy Landscape

Expansion of the unfolded state has been seen in many proteins [8–10]. An unfolded protein in the absence of denaturant is only 10–30 % more expanded than its folded counterpart. These findings are in agreement with the expectation from polymer theory [11, 12]. In contrast to the unfolded state, most scientists agree that the folded state for small proteins has a unique and well-defined structure invariant with denaturants. Only in the last few years have some examples of continuous folding been reported [13, 14]. Continuous folding sometimes was termed "downhill folding" and is characterized by a barrier-less transition into the folded state. A very recent publication on the SH3 domain proposes an expansion of the folded state in the initial step of folding before the all-or-none transition to the unfolded state [15]. This finding suggests that even for a small protein the simple assumption of two-state folding needs to be investigated and probably justified.

A sigmoidal shape of the ensemble denaturation curve is observed for many proteins. This leads to the "two-state" model, which is used extensively to determine the stability of proteins. This model assumes a change in the populations of two distinctly preserved structures as the denaturant increases. However, a linear change of the baseline in the denaturation curve was detected for some observables (e.g. CD) in many different measurements. This is usually interpreted as an observable change that is independent of any structural changes [16]. A linear fit to that part of the curve is usually used to correct for this change. Regarding this finding, we demonstrated that a perturbation in a system not only changes the population of the states and the energy barrier in between them, but it also changes the structure of the state. This is manifested as a sloping baseline in the ensemble denaturation measurement, which is an indication of structural changes. This new finding advances our understanding of the complexity of protein folding.

4.3 Structural Flexibility in Repeat Proteins and Their Function

Usually long-range interactions in proteins stabilize the tertiary structure. Lack of long-range interactions in repeat proteins induces flexibility in the tertiary structure in comparison with globular proteins. It is suggested that repeat proteins are a structural class that falls in between typical globular structured proteins and intrinsically unstructured proteins [17].

In general, repeat proteins unfold at a very low mechanical force, ~ 50 piconewton [18–20]. Using atomic force microscopy, it was shown that the repeat protein ANK

incurs a linear extension at low forces before a step-like unfolding of the secondary structure [21, 22]. The flexibility of ANK is related to its function in mechanotransduction [23–25]. Molecular dynamics simulations were used to show the change in the mechanical properties of the ANK repeats after binding their partner [26].

Another example of an elastic helical repeat protein is PR65, the HEAT-repeat scaffold of phosphatase PP2A. Molecular dynamics simulation shows that smooth global flexural and torsional changes occur at low forces that mediate substrate binding and catalysis [27].

Flexibility in the tertiary structure of the importin-beta repeat, another helical repeat protein, was revealed by mutagenesis, small angle X-ray scattering, and molecular dynamics [17]. It is suggested that solenoid proteins, in general, use their structural properties to allow the specific binding of a single protein to a number of binding partners at different stages of a pathway.

The correlation between flexibility in the structure and the function of proteins is not specific to the repeat proteins. In opposition to the common belief that unfolding reduces the activity, the molten globular state in some helical proteins has been shown to have high activity [28, 29]. The link between folding and binding is also found in intrinsically disordered proteins [30, 31]. In practice, a correlation between the molten globular state and the binding activity has been observed for some helical proteins [32, 33].

In comparing the crystal structures of different TPRs, significant structural variations, due to their binding to the C-terminal peptides of Hsp90 and Hsp70, were not observed [34]. However, this study is based on the crystal structure, and sometimes the crystal can lock the protein to one specific structure. We show that the structure of CTPR with 3 repeats gradually expands under mild denaturant conditions. This finding is in agreement with the measurements and simulation regarding the flexibility of tertiary structures in repeat proteins [27]. We propose that this expansion is relevant for the function of the protein in binding target proteins. Future research needs to be carried out to investigate how loosening in tertiary structures of TPR repeats affects their binding capabilities, possibly leading to conformational selection for multiple partners.

References

1. Kajander T et al (2005) A new folding paradigm for repeat proteins. J Am Chem Soc 127(29):10188–10190
2. Cortajarena AL, Mochrie SG, Regan L (2008) Mapping the energy landscape of repeat proteins using NMR-detected hydrogen exchange. J Mol Biol 379(3):617–626
3. Javadi Y, Main ER (2009) Exploring the folding energy landscape of a series of designed consensus tetratricopeptide repeat proteins. Proc Natl Acad Sci USA 106(41):17383–17388
4. Cortajarena AL, Regan L (2011) Calorimetric study of a series of designed repeat proteins: modular structure and modular folding. Protein Sci 20(2):336–340
5. Neumaier S et al (2013) Testing the diffusing boundary model for the helix-coil transition in peptides. Proc Natl Acad Sci USA 110(32):12905–12910

6. Scott RA 3rd (1967) Statistical mechanical studies of polypeptides. 1. Theory of the helix-coil transition including right- and left-handed helical states. Biopolymers 5(10):931–951
7. Redfield C (2004) Using nuclear magnetic resonance spectroscopy to study molten globule states of proteins. Methods 34(1):121–132
8. Sherman E, Haran G (2006) Coil-globule transition in the denatured state of a small protein. Proc Natl Acad Sci USA 103(31):11539–11543
9. Schuler B, Eaton WA (2008) Protein folding studied by single-molecule FRET. Curr Opin Struct Biol 18(1):16–26
10. Ferreon ACM, Deniz AA (2011) Protein folding at single-molecule resolution. Biochimica Et Biophysica Acta-Proteins and Proteomics 1814(8):1021–1029
11. Haran G (2012) How, when and why proteins collapse: the relation to folding. Curr Opin Struct Biol 22(1):14–20
12. Ziv G, Thirumalai D, Haran G (2009) Collapse transition in proteins. Phys Chem Chem Phys 11(1):83–93
13. Liu J et al (2012) Exploring one-state downhill protein folding in single molecules. Proc Natl Acad Sci USA 109(1):179–184
14. Jha SK et al (2009) Continuous dissolution of structure during the unfolding of a small protein. Proc Natl Acad Sci USA 106(27):11113–11118
15. Campos LA et al (2013) Gradual disordering of the native state on a slow two-state folding protein monitored by single-molecule fluorescence spectroscopy and NMR. J Phys Chem B 117:13120
16. Horovitz A, Matthews JM, Fersht AR (1992) Alpha-Helix stability in proteins. 2. factors that influence stability at an internal position. J Mol Biol 227(2):560–568
17. Forwood JK et al (2010) Quantitative structural analysis of importin-beta flexibility: paradigm for solenoid protein structures. Structure 18(9):1171–1183
18. Li LW et al (2006) Stepwise unfolding of ankyrin repeats in a single protein revealed by atomic force microscopy. Biophys J 90(4):L30–L32
19. Serquera D et al (2010) Mechanical unfolding of an ankyrin repeat protein. Biophys J 98(7):1294–1301
20. Valbuena A et al (2012) Mechanical properties of beta-catenin revealed by single-molecule experiments. Biophys J 103(8):1744–1752
21. Lee G et al (2006) Nanospring behaviour of ankyrin repeats. Nature 440(7081):246–249
22. Sotomayor M, Schulten K (2007) Single-molecule experiments in vitro and in silico. Science 316(5828):1144–1148
23. Sotomayor M, Corey DP, Schulten K (2005) In search of the hair-cell gating spring: Elastic properties of ankyrin and cadherin repeats. Structure 13(4):669–682
24. Corey DP et al (2004) TRPA1 is a candidate for the mechanosensitive transduction channel of vertebrate hair cells. Nature 432(7018):723–730
25. Howard J, Bechstedt S (2004) Hypothesis: a helix of ankyrin repeats of the NOMPIC-TRP ion channel is the gating spring of mechanoreceptors. Curr Biol 14(6):R224–R226
26. Settanni G et al (2013) Effects of ligand binding on the mechanical properties of ankyrin repeat protein gankyrin. PLoS Comput Biol 9(1):e1002864
27. Grinthal A et al (2010) PR65, the HEAT-repeat scaffold of phosphatase PP2A, is an elastic connector that links force and catalysis. Proc Natl Acad Sci USA 107(6):2467–2472
28. Kristinsson HG (2002) Acid-induced unfolding of flounder hemoglobin: evidence for a molten globular state with enhanced pro-oxidative activity. J Agric Food Chem 50(26):7669–7676
29. Vamvaca K, Jelesarov I, Hilvert D (2008) Kinetics and thermodynamics of ligand binding to a molten globular enzyme and its native counterpart. J Mol Biol 382(4):971–977
30. Wright PE, Dyson HJ (2009) Linking folding and binding. Curr Opin Struct Biol 19(1):31–38
31. Sandhu KS, Dash D (2007) Dynamic alpha-helices: conformations that do not conform. Proteins 68(1):109–122

32. Naganathan AN, Orozco M (2011) The native ensemble and folding of a protein molten-globule: functional consequence of downhill folding. J Am Chem Soc 133(31):12154–12161
33. Soulages JL, Bendavid OJ (1998) The lipid binding activity of the exchangeable apolipoprotein apolipophorin-III correlates with the formation of a partially folded conformation. Biochemistry 37(28):10203–10210
34. Grove TZ, Cortajarena AL, Regan L (2008) Ligand binding by repeat proteins: natural and designed. Curr Opin Struct Biol 18(4):507–515

Chapter 5
Summary and Future Plans

We used the SM-FRET experiment to study the folding of CTPR proteins. Continuous changes in FRET efficiency at mild chemical denaturant concentrations, followed by a barrier crossing event that leads to a second peak at high denaturant concentrations, were observed for two different CTPR3 constructs. We showed that the observed shift at mild denaturant concentration is not the result of interchanging two stable populations. To support our findings, analysis of NMR measurements indicated perturbation on the environment of most residues at mild denaturant concentrations. From analysis of hydrogen exchange experiments, we obtained the protection factor for each residue as a function of the denaturant concentration. Residues in the interface between helices show dramatic changes in the relative protection factor, whereas the capping residues in the helix show smaller changes. This indicates a change in the tertiary structure with a preserved secondary structure.

A barrier crossing event was found in our SM-FRET experiments at higher concentrations of denaturant, and matched the CD curve which shows an inflection, indicating unfolding of the secondary structure. A schematic presentation of our model is shown in (Fig. 5.1). Both labeled variants used for single measurements on CTPR3 span the peripheral helix/helices. To confirm that a global opening of the tertiary structure occurs in low denaturant concentrations using single-molecule measurements, we should perform similar measurements on a FRET pair that does not span the peripheral helices. CTPR313 is a good candidate for this. Our collaborators in the group of Prof. Luca D'Andrea will prepare the CTPR313-Alexa variant. We expect to observe a linear shift of FRET efficiency at low denaturant concentrations also for this variant.

The flexibility of the tertiary structure in helical repeats was shown before [1, 2]. Here we suggest that the chemical environment of the protein can change the structure of CTPR proteins. The change in the chemical environment might also be induced by the ligand binding. We propose that this flexibility in the tertiary structure of TPR proteins is important for their selectivity and binding to different ligands. Probable structural changes induced by ligand binding to TPR proteins will be investigated in future measurements using SM-FRET.

S. Cohen, *Single-Molecule Fluorescence Spectroscopy of the Folding of a Repeat Protein*, Springer Theses, DOI 10.1007/978-3-319-09558-5_5

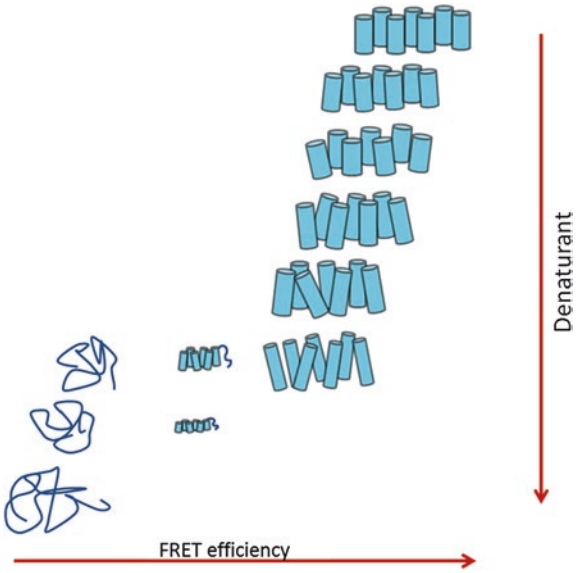

Fig. 5.1 Schematic representation of the folding model for CTPR3. A continuous loosening in the tertiary structure using mild denaturants, following by unfolding of the secondary structure in higher denaturant concentrations is shown. Sizes of the species show indicate their relative populations. The intermediate state with the peripheral helix unfolded is the most populated intermediate state according to the Ising model

SM-FRET will be used for other solenoid repeat proteins to study possible structural loosening in the presence of various perturbations (Temperature, pH, chemical denaturation).

References

1. Forwood JK et al (2010) Quantitative structural analysis of importin-beta flexibility: paradigm for solenoid protein structures. Structure 18(9):1171–1183
2. Grinthal A et al (2010) PR65, the HEAT-repeat scaffold of phosphatase PP2A, is an elastic connector that links force and catalysis. Proc Natl Acad Sci USA 107(6):2467–2472

Publications Authored by the Student

Sharona Sedghani, Vardit Eckhouse, Nir Davidson and Asher A. Friesem, "Suppression of thermal lensing effects in intra-cavity coherent combining of lasers", Optics Communications, Volume 276, 2007, Pages 139–144.

Sharona Sedghani Cohen, Vardit Eckhouse, Asher A. Friesem, and Nir Davidson, "Single frequency lasing using coherent combining", Optics Communications, Volume 282, 2009, Pages 1861–1866.

Menahem Pirchi, Guy Ziv, Inbal Riven, **Sharona Sedghani Cohen**, Nir Zohar, Yoav Barak, and Gilad Haran, *Single-molecule fluorescence spectroscopy maps the folding landscape of a large protein*. Nat Commun, 2011. 2: p. 493.

S. Cohen, *Single-Molecule Fluorescence Spectroscopy of the Folding of a Repeat Protein*, Springer Theses, DOI 10.1007/978-3-319-09558-5