INVESTIGATING THE FOLDING-UNFOLDING MECHANISMS OF METALLOPROTEINS WITH SINGLE-MOLECULE FORCE SPECTROSCOPY

by

Jiayu Li

B.Sc., Fudan University, 2016

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Chemistry)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

March 2022

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Investigating the Folding-Unfolding Mechanisms of Metalloproteins with Single-Molecule Force Spectroscopy

submitted by Jiayu Li in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Chemistry

Examining Committee:

Hongbin Li, Chemistry, UBC
Supervisor

Suzana Straus, Chemistry, UBC
Supervisory Committee Member

Yan Alexander Wang, Chemistry, UBC
University Examiner

Joerg Gsponer, Michael Smith Laboratories, UBC
University Examiner

Additional Supervisory Committee Members:

Chris Orvig, Chemistry, UBC
Supervisory Committee Member

Martin Tanner, Chemistry, UBC
Supervisory Committee Member

Abstract

Metalloproteins account for more than one-third of all proteins in nature and play important roles in biological processes. The folding process of metalloproteins is complicated, as it is driven by not only the polypeptide chain folding effect but also the metal coordination. Folding into the native structures with correctly assembled metal cofactors is a prerequisite for metalloproteins to perform their biological functions. Therefore, understanding the folding-unfolding mechanisms of metalloproteins is of critical importance. Over the past two decades, single-molecule force spectroscopy (SMFS) has evolved into a powerful method to investigate the folding-unfolding mechanisms of metalloproteins at the single-molecule level. This thesis presents the SMFS studies on the folding-unfolding mechanisms of four important metalloproteins, including three iron-sulfur proteins and one heme-containing protein.

First, we studied the mechanical unfolding behavior of a high potential iron-sulfur protein by atomic force microscopy (AFM)-based SMFS, and revealed a detailed mechanical unfolding mechanism. In combination with previous studies, we proposed a general mechanical unfolding mechanism for the iron-sulfur protein family.

We then investigated the folding behavior of the simplest iron-sulfur protein, rubredoxin, with optical tweezers (OT)-based SMFS. We discovered a novel binding-folding-reconstitution mechanism of the folding of rubredoxin, and highlighted the critical importance of the two-coordinate ferric site in the folding of rubredoxin.

We also explored the folding behavior of another iron-sulfur protein, ferredoxin, with OT-based SMFS. The unfolded ferredoxin was found to mostly misfold instead of folding back

to its native structure; however, the successful reconstitution of the β-sheet or the coordination center was observed in rare cases.

In addition, we studied the folding-unfolding behavior of a heme-containing protein, cytochrome c, using OT-based SMFS. We revealed a detailed folding-unfolding mechanism for holo-form cytochrome c, and identified the deviation from random coil behavior of apo-form cytochrome c, which was previously inaccessible by ensemble spectroscopic studies.

Finally, conclusions and future directions on investigating the folding-unfolding mechanisms of metalloproteins with SMFS were presented. Overall, this thesis advances our understanding of the folding-unfolding mechanisms of iron-sulfur proteins, heme-containing proteins, as well as metalloproteins in general, and the systematic studies pave the way for further research in this important area.

Lay Summary

Metalloproteins are an important class of proteins that play a variety of biological functions in organisms. Resolving the folding-unfolding mechanisms of metalloproteins is important for understanding the structure, function and dynamics of metalloproteins. Single-molecule force spectroscopy (SMFS), which enables direct observation of the folding-unfolding process of metalloproteins at the single-molecule level, has become a powerful method to study the folding-unfolding mechanisms of metalloproteins. Using SMFS, we investigated the folding-unfolding mechanisms of four important metalloproteins, including three iron-sulfur proteins and one heme-containing protein. Our studies revealed detailed folding-unfolding mechanisms of these metalloproteins at the single-molecule level, identified the common features in their folding-unfolding processes, and provided valuable insights into the folding-unfolding behaviors of metalloproteins in general.

Preface

I wrote chapter 1 and 6 independently.

A version of chapter 2 has been published. Li J. and Li H. (2018) Mechanical Unfolding Pathway of the High potential Iron-Sulfur Protein Revealed by Single-Molecule Atomic Force Microscopy: Toward a General Unfolding Mechanism for Iron-sulfur Proteins. J. Phys. Chem. B. 122(40), 9340-9349. This chapter is incorporated in this thesis with permission of the American Chemical Society, Copyright 2018. Dr. Li and I designed the experiments together. I conducted all the experiments and analyzed all the data. I wrote the manuscript together with Dr. Li.

A version of chapter 3 has been published. Li J. and Li H. (2020) Single molecule force spectroscopy reveals that a two-coordinate ferric site is critical for the folding of holo-rubredoxin. Nanoscale, 12(44), 22564-22573. This chapter is incorporated in this thesis with permission of The Royal Society of Chemistry, Copyright 2020. Dr. Li and I designed the experiments together. I conducted all the experiments and analyzed all the data. I wrote the manuscript together with Dr. Li.

A version of Chapter 4 is in preparation for publication. Li J. and Li H. The Mechanical Folding-Unfolding Behavior of a Plant-Type [2Fe-2S] Ferredoxin Characterized by Single-Molecule Force Spectroscopy. Dr. Li and I designed the experiments together. I conducted all the experiments and analyzed all the data. I wrote the chapter.

A version of Chapter 5 is in preparation for publication. Li J. and Li H. Single-molecule force spectroscopy revealed the folding-unfolding mechanisms of holo- and apo-form cytochrome c: a new understanding of the role of heme cofactor in folding. Dr. Li and I designed

the experiments together. I conducted all the experiments and analyzed all the data. I wrote the chapter.

Table of Contents

List of Tables

List of Figures

List of Symbols

ΔLc	Contour length change
$\langle z_C^2 \rangle$	Time-average square of the displacement of the cantilever in the *z* direction
$dP_f(t)/dt$	Folding probability density
$dP_u(t)/dt$	Unfolding probability density
F	Force
$F(x)$	Force at the extension of x
h	Plank constant
$k(F)$	Rate constant at a given force F
k_B	Boltzmann constant
k_c	Spring constant
Lc	Contour length
m	Mass
$M(F)$	Number of time windows at the force of F
$N(F)$	Number of the folding or unfolding events at the force of F
p	Persistence length
$P(F)$	Probability of protein folding/unfolding at the force of F
$P_f(t)$	Folding probability
$P_u(t)$	Unfolding probability
s	Second
T	Temperature in Kelvin
t	Time

x	Extension
$\alpha(F)$	Unfolding rate at the force of F
α_0	Unfolding rate at zero force
$\beta(F)$	Folding rate at the force of F
β_0	Folding rate at zero force
$\Delta G^0_{N\text{-}T}$	Unfolding Gibbs free energy barrier at zero force
$\Delta G^0_{U\text{-}T}$	Folding Gibbs free energy barrier at zero force
$\Delta G^F_{N\text{-}T}$	Unfolding Gibbs free energy barrier at the force of F
Δt	Time window
Δx_f	Folding distance
Δx_u	Unfolding distance
κ	Transmission coefficient
λ	Wavelength
ω_0	Angular resonant frequency

List of Abbreviations

aa	Amino acid
AFM	Atomic force microscopy
ALS	Amyotrophic lateral sclerosis
apo-	Apo-form
ATP	Adenosine triphosphate
BdpA	B domain of protein A
bp	Base pair
CaM	Calmodulin
Ccc2	Cation-chloride cotransporter 2
CD	Circular dichroism
cHiPIP	High potential iron-sulfur protein from *Chromatium tepidum*
cMyb	A transcription factor protein
CucA	Cu^{2+}-cupin A
CyaA	Adenylate cyclase
cytc	Cytochrome c
DNA	Deoxyribonucleic acid
DnaJ	A molecular chaperone protein
dsNDA	Double-strand deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
F	Force
FD	Ferredoxin

FKBP	FK506 binding protein
FRET	Forster resonance energy transfer
GB1	The B1 binding domain of protein G from *streptococcus*
GolB	A gold-specific binding protein
GuHCl	Guanidine hydrochloride
Hb	Hemoglobin
HiPIP	High potential iron-sulfur protein
holo-	Holo-form
hsp	Heat shock proteins
I	Intermediate state
IPTG	isopropyl-β-D-1-thiogalactopyranoside
isc	Iron-sulfur cluster
Jac1	A J-type co-chaperone
kDa	KiloDalton
LB	Luria-Bertani
m	Meter
Mb	Myoglobin
MncA	Mn^{2+}-cupin A
mV	Millivolt
N	Native state
nm	Nanometer
NMR	Nuclear magnetic resonance
nN	NanoNewton

NTA	Nitrilotriacetate
NuG2	A designed variant of protein G
OD	Optical density
OT	Optical tweezers
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
pfRD	Rubredoxin from *Pyrococcus furiosus*
pN	Piconewton
PrP	Prion protein
PSD	Position sensitive detector
RD	Rubredoxin
RNA	Ribonucleic acid
s	Second
SdrG	Serine-aspartate repeat protein G
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
sFD	Ferredoxin from *Spinacia oleracea*
SMCC	4-(N-Maleimidomethyl) cyclohexanecarboxylic acid N-hydroxysuccinimide ester
SMFS	Single-molecule force spectroscopy
SOD1	Superoxide dismutase 1
Ssq1	A stress-seventy subfamily Q protein
T	Temperature
TCEP	Tris(2-carboxyethyl)phosphine

Top7	An artificial protein
Tris	Tris(hydroxymethyl)aminomethane
U	Unfolded state
μm	Micrometer
μs	Microsecond
UV	Ultraviolet
UV-Vis	Ultraviolet-visible
WLC	Worm-like chain
wt	Wild-type

Acknowledgements

First and foremost, I would like to express my sincere gratitude and appreciation for my supervisor, Dr. Hongbin Li, whose dedicated support and guidance have been invaluable throughout my Ph.D. His patience, motivation, enthusiasm, and knowledge have deeply inspired me. He has taught me the methodology needed to carry out scientific research, and showed me how to become an independent, creative researcher. Every time I met problems or got frustrated by failures in my research, he encouraged me to take on the challenges and overcome the difficulties. He has also supported me to attend international conferences to present my work to a broader audience. I am extremely grateful for what he has offered me. It was a great privilege for me to work under his guidance.

I would like to extend my sincere thanks to Dr. Chris Orvig, Dr. Suzana Straus and Dr. Martin Tanner for being on my supervision committee and providing valuable suggestions and intellectual guidance.

I would like to thank all the past and present members of Li's group. Special thanks to Dr. Chengzhi He, Yabin Guo and Dr. Han Wang for teaching me all the experimental techniques when I first joined the group. Many thanks to Dr. Shanshan Lyu, Dr. Ao Xia, Dr. Xiaoqing Gao, Dr. Ruidi Wang, Dr. Yongliang Wang, Dr. Wenxu Sun, Dr. Jiahao Xia, Dr. Linglan Fu, Tianyu Duan, Adam Xiao, Denghuang Zhan, Xiaochen Xu, Jiacheng Zuo, Haoquan Lin, Qingyuan Bian, Jess Fong, Guojun Chen, Peiyun Li, Sena Arslan and Lucas Philip for their help and friendship.

Finally, I would like to express my deepest gratitude to my parents, for their unconditional love, support, understanding and sacrifice over the years, despite the long distance between us. I could not have done it without them. I would also like to thank my wife, who

experienced all my ups and downs by my side and provided me with continued love, support and encouragement.

To My Parents and Wife

Chapter 1: Introduction

Proteins are complex biomacromolecules that perform diverse cellular functions.[1-3] Newly synthesized proteins undergo a remarkable transformation from one-dimensional amino acid sequences into complex three-dimensional structures, which are biologically functional.[4-6] Protein folding is the most fundamental and universal example of biological self-assembly. Understanding this complex process will therefore provide a unique insight into the way that a molecular system evolves for functional advantage.[7] The investigation of the protein folding-unfolding process has heavily relied on traditional ensemble spectroscopic approaches, where folded proteins are perturbed by chemical denaturants or temperature change.[8-12] However, the spectroscopic signals from these ensemble experiments only describe the averaged properties of a large number of protein molecules, and rare events or short-lived intermediate states in the protein folding-unfolding process can often be masked.[13]

Over the past two decades, single-molecule force spectroscopy (SMFS) has evolved into a powerful tool to investigate the protein folding-unfolding process at the single-molecule level.[14] Taking advantage of the high-resolution single-molecule manipulation techniques including atomic force microscopy (AFM), optical tweezers (OT) and magnetic tweezers, SMFS allows one to apply force onto chosen residues on a single protein molecule, thus permitting the study of protein folding-unfolding along a specific reaction coordinate at the single-molecule level. So far, SMFS has revealed invaluable new insights into the folding-unfolding mechanisms for a wide range of proteins, and provided some new insights that are not available from traditional ensemble studies.[13, 15-18]

Metalloproteins account for more than one-third of all the proteins in cells and perform important biological functions.[19-21] The folding of metalloproteins is more complicated than that

of metal-free proteins, due to the intertwined effects of polypeptide folding and metal coordination. Traditional ensemble methods face great challenges in studying the folding-unfolding of metalloproteins, as the unfolding processes of metalloproteins are mostly irreversible *in vitro*.[22] SMFS has demonstrated superb advantages in probing metalloproteins in a near-native environment, as well as resolving folding-unfolding intermediates at the single-molecule level. To date, the folding-unfolding mechanisms of several important metalloproteins have been revealed by SMFS, yet many important questions in this area remain to be answered and call for further studies.[14, 23-39]

This chapter provides an overview of the research background of using SMFS to study the folding-unfolding mechanisms of metalloproteins. I will first introduce the basic knowledge of protein structure, folding, and some widely accepted protein folding theories. The traditional ensemble methods used to study protein folding-unfolding will also be briefly discussed. Then, I will introduce the method of SMFS, as well as the instrumental principles of AFM and OT, which SMFS frequently relies on. I will also introduce metalloproteins with a focus on how they differ from metal-free proteins in the biogenesis process, and how metal binding makes the folding process more complicated. Research progress made on studying the folding-unfolding of metalloproteins by SMFS will also be reviewed. Finally, the aims of this thesis will be stated.

1.1 Protein Structure and Folding

1.1.1 Protein Structure

Proteins are linear polypeptide chains synthesized with the help of ribosomes and are composed of amino acids. The synthesized proteins adopt different structures, and a four-level structural hierarchy is widely used to depict them. The sequence of amino acids in the polypeptide chain is known as the primary structure (Figure 1.1A). The primary structure is

determined by the DNA sequence of the gene that encodes the protein, and is hypothesized to contain all the necessary information for the protein to fold into the native structure as shown by Anfinsen in his pioneering works on ribonuclease.[40] The secondary structure refers to local folded structures stabilized by hydrogen bonds between adjacent amino acid residues in the polypeptide backbone (Figure 1.1B). The most common secondary structural motifs are the α-helix, featuring a right-handed helical structure, and the β-sheet, formed by parallel or anti-parallel β-strands. The tertiary structure refers to the overall geometry of a polypeptide chain with arranged secondary structural motifs (Figure 1.1C). The formation of the tertiary structure largely results from the interactions between the R groups of the amino acids, and for proteins containing cysteines and cofactors, the formation of disulfide bonds and the interaction between the polypeptide chains and the cofactors also contribute to the tertiary structure. For multi-subunit proteins, the quaternary structure refers to the assembly of multiple polypeptide chains with defined tertiary structures.

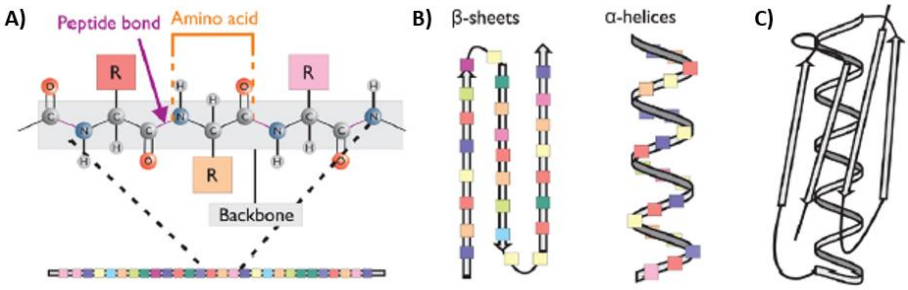

Figure 1.1 Schematic diagram of levels of protein structure. (A): primary structure, (B): secondary structure and (C): tertiary structure (protein L, PDB: 1HZ6). Images adapted with permission from IOP Publishing Ltd, Copyright © 2016.[18]

Most protein structures are determined experimentally by X-ray crystallography, nuclear magnetic resonance and cryo-electron microscopy.[41] For protein structures that could not be determined experimentally due to experimental challenges, computational methods can be used to predict the protein structures.[41-42] To date, more than 170000 protein structures have been

solved and deposited in Protein Data Bank (PDB), and the number is increasing rapidly every year.[43] The access to protein structures has critical importance and lays a foundation for studies on protein folding, function, interaction, etc.

1.1.2 Protein Folding

Protein folding is the physical process by which the synthesized polypeptide chain acquires a well-defined native three-dimensional structure that is biologically active. Generally, the native state is the thermodynamically most stable conformation, possessing a marginally higher stability than the unfolded states.[44] Protein folding is driven by multiple weak interactions, both local and long-range. The hydrophobic effect packs hydrophobic residues into the core of the protein and polar residues on the surface; hydrogen bonds can align in certain patterns and form secondary structures; charged amino acid residues attract or repel each other via electrostatic interactions; tightly packed atoms in the folded protein have van der Waals interactions; the polypeptide chain has its own backbone angle preferences, and the conformational entropy of the polypeptide chain also plays a role in protein structure; and some other interactions may exist in certain types of proteins, such as metal coordination interactions in metalloproteins.[45-47] Among these, hydrophobic interactions and conformational entropy of polypeptide chains are generally considered as the main stabilizing and destabilizing sources, respectively. These driving forces except for the conformational entropy of the polypeptide chain are described by potential energies in forcefield models that are widely used in computer simulations to study protein folding.[48]

Protein folding can happen both *in vivo* and *in vitro*, with both folding processes differing in many ways. *In vivo*, partially synthesized proteins may fold co-translationally before the entire chain has been made and extruded from the ribosomal exit tunnel, and ribosome, ribosome-

associated enzymes and chaperones create a special initial environment for protein folding.[49] In addition, many proteins undergo further modifications to be functional, including the formation of disulfide bonds, peptidyl-prolyl isomerization, cleavage, covalent attachment of carbohydrates and lipids, assembly of cofactors, etc., and all of them may potentially alter the protein folding pathway and the final native protein structure. The highly concentrated cellular milieu with concomitant macromolecular crowding, spatial organization and various weak molecular interactions also have a significant impact on the *in vivo* protein folding process.[50] The *in vitro* protein folding process, in contrast, mostly only involves purified, fully-expressed proteins in relatively simple aqueous environments.

1.1.2.1 Protein Folding Problem

The protein folding problem arose in the 1960s when the first atomic-resolution structures of globular proteins were determined.[51-53] The protein folding problem has been regarded as three main questions: (a) The folding code: How is the three-dimensional native structure of a protein determined by the physicochemical properties that are encoded in its one-dimensional amino-acid sequence? (b) The folding mechanism: A polypeptide chain can adopt infinite possible conformations. How can proteins fold so fast? (c) Protein structure prediction: Can we devise a computer algorithm to predict the native structure of a protein solely from its amino acid sequence?

The protein folding code problem arose upon Anfinsen's now-famous experiments on ribonuclease, where Anfinsen postulated that the native structure of a protein is the thermodynamically most stable structure; it depends only on the amino acid sequence and the conditions of the solution, but not the kinetic folding route.[40, 54] Although several exceptions have been found where the biologically active state is in fact kinetically trapped, it has been

widely accepted that most proteins fold into the thermodynamically most stable native structure, no matter if they fold *in vivo* or *in vitro*.[55-56] Anfinsen's thermodynamic hypothesis justifies the validity of studying protein folding *in vitro* and has had profound influence on research that followed his ground-breaking experiments.

1.1.2.2 Protein Folding Mechanism

The protein folding mechanism problem was first highlighted by Levinthal in the late 1960s, who pointed out that if all the possible conformations of the unfolded state are sampled randomly and unbiasedly, it would take longer than the age of the universe for a small protein to fold into its native structure.[57] Therefore, in order to acquire their native states in a biologically relevant time, proteins must somehow fold along certain pathways in sequential steps to avoid searching through all the possible conformations. The *in vitro* protein folding experiments have been carried out following this idea to characterize the protein folding kinetics and to find folding intermediates, which were then used as snapshots to depict the protein folding pathways.[58-59]

The search for a single, simple protein folding mechanism to describe the general protein folding behavior has never stopped but has not succeeded so far. Based on the generalization of the revealed protein folding pathways, several protein folding models have emerged. The framework model argues that the native structure is formed in a stepwise manner: the local secondary structures form rapidly and act as a scaffold, and the tertiary structure forms subsequently.[60] A related mechanism is the diffusion-collision mechanism, which states that the rapidly formed secondary structures diffuse, collide with each other and finally coalesce to form the tertiary structure.[61-62] These models gained experimental support mostly from small, helical proteins, which have relatively stable local interactions but no significant long-range tertiary

structure, such as engrailed homeodomain and protein A.[63-64] On the other extreme, the hydrophobic-collapse model suggests that protein folding is initiated by the polypeptide chain hydrophobically collapsing into a molten globule state with some degree of tertiary structure, followed by the formation of secondary structures and other interactions in a confined volume.[65] However, later it was found that the hydrophobic collapse is almost always accompanied by some secondary structure formation, and so far no protein has been shown to fold entirely through nonspecific hydrophobic collapse.[66] In contrast to the two extreme folding behaviors described in the two abovementioned models, i.e. either the secondary structure or the tertiary structure forms first, it has been found that for many proteins, the secondary and the tertiary structures form in a concerted and coupled way.[67] The nucleation-condensation model, which proposes that a structural nucleus first forms slowly on the polypeptide chain and the structure propagates rapidly from the nucleus to form the final structure, is considered to unify the features of both the framework and hydrophobic-collapse models to some degree, as it can shift to either of the two extreme models depending on the relative stabilities of the secondary and the tertiary structure.[66, 68] Many α, β and α/β proteins (proteins containing only α-helices, only β-sheets and both α-helices and β-sheets) are found to fold via the nucleation-condensation mechanism, such as cMyb, tenascin and FKBP.[69-70] In addition to these widely used models, there are also some models based on different structural notions such as the foldon and the topomer, which can also describe the folding behaviors of some proteins well.[71-72] These mechanisms are not mutually exclusive, namely, a certain protein can sometimes fold following the predictions of multiple models. Additionally, protein folding is such a complicated process that almost all the above-mentioned models have a potential risk of oversimplification.[73] It remains to be seen whether an

underlying unifying mechanism that governs the folding of all the proteins will be found in the future.

1.1.2.3 Protein Folding Energy Landscape

The above-mentioned pathway models are based on phenomenological kinetics using single symbols to represent the unfolded, folded, and intermediate states. They use structural intermediates to describe the folding pathway, and divide the protein folding process into separate steps. However, this may oversimplify the folding process, and the complex folding behaviors were hard to relate to any theoretical understanding of protein energetics.

Resulting from the advances in experimental methods with information at the atomic level (e.g., high-resolution hydrogen exchange, mass spectrometry and fast laser-triggered methods) and new conceptual models based on simplified statistical mechanics, protein folding energy landscape theory, which is also known as protein folding funnel theory, emerged in the late 1980s as an alternative viewpoint, treating the protein folding process from a global energetic surface perspective.[74] A protein folding energy landscape displays the energy of each conformation as a function of the degrees of freedom, with its vertical axis representing the internal energy of a given conformation, and its lateral axis representing the conformational entropy. According to this theory, the protein folding energy landscape has an overall funnel-shaped but rugged surface (Figure 1.2). The low-energy native conformation stays at the bottom of the funnel, and the high-energy disordered unfolded conformations reside at the edge of the funnel. In this theory, the protein folding process is viewed as a downward diffusive search over the landscape surface through an ensemble of structures to arrive at the native structure. During the folding process, proteins may encounter energy barriers and get kinetically trapped in non-native local minima, which slows down the folding process or causes misfolding.

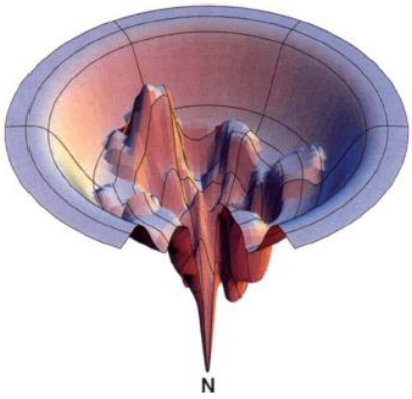

Figure 1.2 A rugged protein folding energy landscape. N: native state. Image adapted with permission from Nature Publishing Group, Copyright © 1997.[74]

The protein folding funnel theory suggests that the Levinthal paradox is not a real problem. Instead, it is more like a misconception of the process in which a macromolecule system reaches its thermodynamically stable state. The random searching idea in the Levinthal paradox corresponds to a flat energy landscape with one deep pit (the native state), which is in fact not the case for protein folding. The protein folding pathway theory, which is used to solve the Levinthal paradox, can be considered as a one-dimensional energy landscape profile along a certain reaction coordinate on the three-dimensional protein folding landscape. Proteins can adopt multiple pathways along the landscape surface to fold depending on their initial unfolded states and the folding environment.[75] In addition, this energy landscape concept treats protein folding as a parallel flow process of an ensemble of molecules, replacing the pathway concept with sequential folding events. The protein folding energy landscape theory provides a bridge between the protein folding pathway and protein folding thermodynamics and kinetics.[76-77] It has evolved rapidly over the past thirty years from both experimental and computational studies, and is now widely accepted as a conceptual framework to describe the protein folding process.[78]

1.1.3 Methods to Study Protein Folding

For decades, *in vivo* protein folding could only be studied by traditional biochemical methods such as cross-linking, enzymatic digestion, lysis and assaying.[79] However, newly-developed techniques, such as live-cell NMR and Förster resonance energy transfer (FRET), enable minimally invasive observation inside cells, thus making monitoring real-time folding, protein stability and even single protein molecules in cells possible.[80-84] Despite these, the majority of protein folding studies are carried out *in vitro*, mainly because the *in vitro* experimental systems are much simpler so that a broader range of techniques can be applied to extract more information of the process.[85] To study protein folding *in vitro*, proteins are diluted to very low concentration in solution. A variety of factors can be used as denaturants to trigger the protein folding-unfolding process, and spectroscopic signals of the ensemble or a single molecule can be monitored during the process.[86]

1.1.3.1 Using Traditional Denaturants

Extensive experimental efforts have been carried out to investigate *in vitro* protein folding using ensemble methods with traditional denaturants, including guanidine hydrochloride (GuHCl), urea, heat, extreme pH, etc.[87-88] These denaturants can alter the protein folding energy landscape to favor the unfolded state thermodynamically, thus can be introduced to protein solutions to induce protein unfolding. Removing such denaturants from the system (e.g., diluting the GuHCl and urea, removing the heat, neutralizing the pH) brings the system back to physiological conditions and triggers protein folding. A variety of spectroscopic signals reporting on different aspects of the protein conformation can be monitored during this protein folding-unfolding process, such as fluorescence (for the local environment of aromatic residues) and circular dichroism (for polypeptide secondary structures) (Figure 1.3A). To obtain folding or

unfolding kinetics, fast-mixing stopped-flow, laser-induced pH or temperature jump apparatus can be used to trigger reactions followed by time-resolved spectroscopic methods. The thermodynamic and kinetic parameters can be used to map the protein folding energy landscape (Figure 1.3A-B).

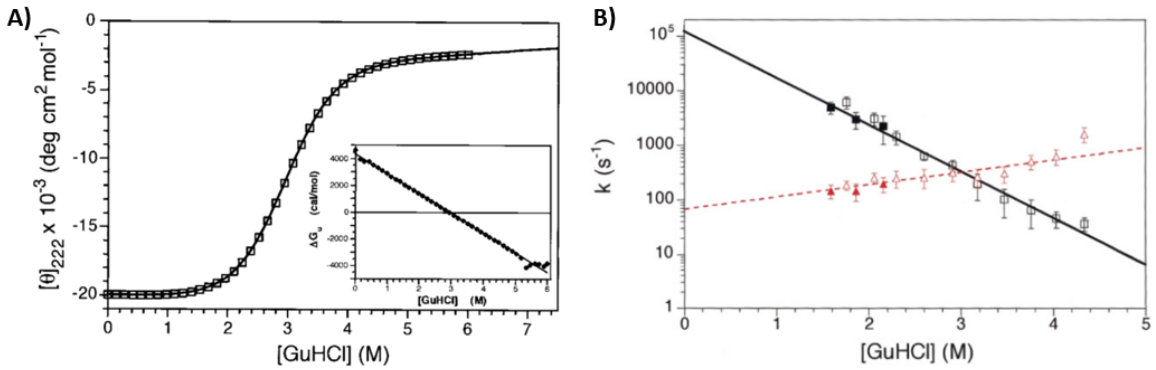

Figure 1.3 Studying protein folding with traditional denaturants. (A) GuHCl unfolding curve of B-domain of protein A (BdpA) monitored by circular dichroism. The solid curve is a fit using a two-state unfolding model. Inset: Free energy barrier of unfolding versus GuHCl concentration.[64] (B) Folding (black squares) and unfolding (red triangles) rate constants of BdpA as a function of GuHCl concentration. Lines are linear fits of lnk versus GuHCl concentration. Images adapted with permission from Nature Publishing Group, Copyright © 1969.[64]

Traditional denaturing methods have been proven powerful at answering many questions regarding the protein folding-unfolding process.[8-12] However, such spectroscopic methods can only measure the average signal of an ensemble, thus may result in some potential inaccuracies. For example, rare or transient events may be obscured, low-populated states may be masked, different subpopulations cannot be distinguished, and the reaction coordinate is not well-defined.[16] To overcome these obstacles, experimental tools that probe protein folding-unfolding at the single-molecule level were developed. Single-molecule fluorescence and SMFS are the two most commonly used single-molecule techniques for protein folding-unfolding study, and between them, SMFS exploits mechanical force as a novel protein denaturant, providing a new perspective to investigate the protein folding process.[13]

1.1.3.2 Unfolding Protein by Force

Applying a mechanical stretching force onto folded protein molecules will cause protein conformational changes (extending along the stretching direction), and as a result, a folded protein molecule can overcome its energy barrier and unfold. In addition, force is a common natural perturbation in biological systems. In many cellular processes such as the transport of cellular cargo by motors, protein degradation, DNA replication and RNA translation, biomacromolecules are subjected to or develop mechanical forces while functioning.[89-90] While extreme temperature, pH or high concentrations of denaturants are not commonly observed *in vivo*, force is involved in the *in vivo* folding-unfolding processes of some proteins, thus can be used as a physiologically relevant approach to trigger protein folding-unfolding and create previously unattainable experimental conditions.

The effect of force on the protein folding energy landscape can be easily demonstrated with a two-state system, where the folded state and unfolded state are separated by a single energy barrier (the transition state) on the free energy profile. As shown in Figure 1.4, while the folded state has a lower free energy than the unfolded state when no force is applied (grey curve), the force can tilt the free energy surface along the mechanical reaction coordinate, thereby lowering the free energies of both the transition state and the unfolded state (black curve). As a result, the unfolded state becomes more thermodynamically favorable and more populated under stretching force, and the unfolding kinetics increase due to the lowered energy barrier. Once the force is removed, the protein folding energy landscape will resume its original shape (with no force applied) with the folded state again being more stable than the unfolded state, and the protein molecule will tend to fold.

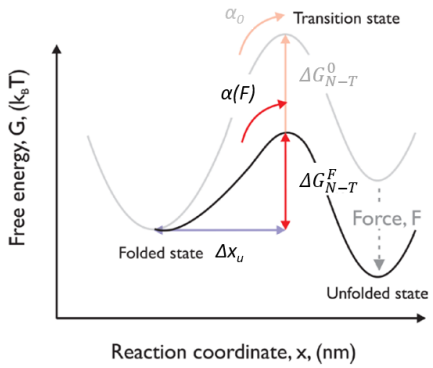

Figure 1.4 Schematic diagram showing the effect of force on the energy landscape of a two-state folding protein. α_0: unfolding rate under no force, $\alpha(F)$: unfolding rate under force, Δx_u: unfolding distance (the distance between the native state and the transition state), $\Delta G^0_{N\text{-}T}$: unfolding energy barrier under force, $\Delta G^F_{N\text{-}T}$: unfolding energy barrier under force. Image adapted with permission from IOP Publishing Ltd, Copyright © 2016.[18]

Different models have been developed to quantify the energy change induced by the applied force, and among them, the Bell-Evans model (or Bell-Evans-Zhurkov model) has been widely used due to its simplicity and validity in the protein/DNA/RNA folding-unfolding process. [91-93] It assumes that mechanical force does not change the unfolding and the folding distance (distances from the folded and the unfolded state to the transition state) on the protein folding energy landscape, but lowers the energy barrier by its work, $F\Delta x$. Incorporating Kramer's reaction rate theory, protein mechanical folding-unfolding is described as Brownian diffusion over a one-dimensional energy profile, where the unfolding rate and the folding rate under applied force, $\alpha(F)$ and $\beta(F)$, change exponentially with force, given by:

$$\alpha(F) = \kappa \frac{k_B T}{h} \exp\left(-\frac{\Delta G^0_{N-T} - F\Delta x_u}{k_B T}\right) = \alpha_0 \exp\left(\frac{F\Delta x_u}{k_B T}\right) \quad (1.1)$$

$$\beta(F) = \kappa \frac{k_B T}{h} \exp\left(-\frac{\Delta G^0_{U-T} + F\Delta x_f}{k_B T}\right) = \beta_0 \exp\left(\frac{-F\Delta x_f}{k_B T}\right) \quad (1.2)$$

where F is the applied mechanical force, $\Delta G^0_{N\text{-}T}$ and $\Delta G^0_{U\text{-}T}$ are the unfolding and folding energy barriers under zero force, Δx_u and Δx_f are the protein unfolding and folding distances, α_0 and β_0

are the protein unfolding and folding rate constants under zero force, k_B is the Boltzmann constant, T is the absolute temperature, κ is the transmission coefficient, and h is the Plank constant.[94] By fitting the force dependence of the observed rate constants, one can extract the folding/unfolding rate and distance parameters to reconstruct the protein folding energy landscape (Figure 1.4).[17] For experimental results that do not provide unfolding or folding rates explicitly, Monte Carlo simulations that simulate the stochastic protein unfolding events can be used to reproduce the experimental results and estimate the underlying kinetic parameters.[95-96]

Recent attempts to revise the potential oversimplification of the Bell-Evans model led to some other theoretical models. For example, Dudko *et al.* derived a force-dependent transition rate formula where it shows that force not only lowers the transition barriers but also changes the transition distances, leading to new free energy minimum and maximum positions.[97] It is proven to be suitable for unfolding events at high enough forces where entropic conformational fluctuation of biomolecules can be ignored.[98] More recently, Guo *et al.* developed a new model for one-dimensional transition taking the entropic conformational fluctuation of biomolecules at low forces into consideration.[98] When such force-dependent entropic fluctuation of biomolecules is considered, complex effects of force on the kinetics of protein unfolding or biomolecular complex dissociation emerge on a one-dimensional free energy landscape. Guo *et al.* also derived a transition pathway-independent expression for force-dependent unfolding/dissociation rate for a broader range of experiments.[99] Based on Arrhenius law and the structural-elastic differences between the transition and the native states, it shows that differential low-force entropic conformational fluctuations between the transition and native states can explain a wide range of complex kinetic behaviors such as catch-bond and catch-to-slip bond switch at low forces. Despite these advantages, these new models require some additional information about

the protein folding-unfolding process, which may not be readily available experimentally. Due to the simplicity and generality of Bell-Evans model as well as the clearly observed linear relationship between the logarithm of the protein folding-unfolding rates and the applied force, Bell-Evans model was still used in this thesis to extract the protein folding-unfolding kinetic parameters. This also enabled direct comparisons of the folding-unfolding kinetics with the previously studied metalloproteins, of which the folding-unfolding kinetics were elucidated using Bell-Evans model mostly.

Force as a denaturant can provide a defined reaction coordinate for protein folding and unfolding, and single-molecule force spectroscopy, which combines mechanical folding-unfolding with single-molecule measurements, overcomes the obstacles of traditional ensemble measurements and has become a powerful tool in investigating the protein folding-unfolding process. It is of note that different protein folding-unfolding methods may sample different parts of the folding energy landscape, therefore, direct comparisons between parameters obtained using different experimental methods may not have a definite physical meaning.[100]

1.2 Single-Molecule Force Spectroscopy

Single-molecule force spectroscopy (SMFS) uses mechanical force to stretch target objects and measures their response in real-time, in order to obtain the relationship between force, molecular extension and stretching time.[101-102] Taking advantages of high-resolution single-molecule techniques such as AFM, OT and magnetic tweezers, SMFS has acquired a manipulation capacity spanning six orders of magnitude in length (10^{-10}-10^{-4} m) and force (10^{-14}-10^{-8} N).[103] With this, SMFS has been used in a broad range of research areas, including cell adhesion, ligand-receptor interaction, DNA/RNA conformational changes, molecular motors, and very importantly, protein folding-unfolding.[103] In this section, I will focus on the principles

of the two most common techniques, AFM and OT, and introduce the application of SMFS in protein folding-unfolding studies. It is of note that magnetic tweezers-based SMFS has also been extensively used in protein folding-unfolding studies, and has revealed invaluable information regarding the protein folding-unfolding process with its unique advantages.[104-109] However, as this technique is not used in this thesis, it will not be discussed here in detail.

1.2.1 Atomic Force Microscopy-Based Single-Molecule Force Spectroscopy

The AFM, invented by Binnig, Quate and Gerber in 1986, is a high-resolution scanning probe microscope that allows mechanical manipulation of samples with atomic resolution. While it is mostly used as a high-resolution imaging tool, it can also be used in SMFS to probe and manipulate macromolecules, due to its high force sensitivity over a broad force range (10-10^4 pN), nanometer spatial resolution and sub-millisecond temporal resolution.

In an AFM instrument (Figure 1.5), a cantilever made of silicon or silicon nitrite that behaves as a Hookean spring is mounted onto a cantilever holder. A piezoelectric scanner controls the distance between the cantilever and the upper surface of the scanner. The cantilever has a sharp tip, and a focused laser beam is shined onto the back of the tip and reflected onto a position-sensitive detector (PSD). Whenever there are interactions between the tip of the cantilever and the surface, the cantilever will be deflected as a spring in the z-direction. The extension of the stretched molecule can be calculated with the position of the scanner and the deflection of the cantilever, which is reported by the PSD. The force on the cantilever can also be calculated with the deflection of the cantilever using Hook's Law.

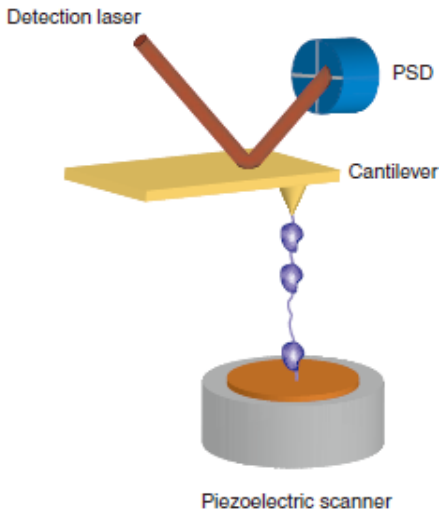

Figure 1.5 Schematic of the instrument of AFM. Image adapted with permission from Nature Publishing Group, Copyright © 2008.[103]

The cantilever is usually calibrated by the thermal fluctuation method, by which the cantilever tip is treated as a simple harmonic oscillator with one-dimensional fluctuation in response to thermal noise, so that,

$$\frac{1}{2}m\omega_0^2 < z_C^2 > = \frac{1}{2}k_c < z_C^2 > \quad (1.3)$$

where m is the mass of the cantilever, ω_0 is the angular resonant frequency, $<z_C^2>$ is the time-average square of the displacement of the cantilever in the z-direction, k_c is the spring constant of the cantilever.[110] And according to the equipartition theorem,

$$\frac{1}{2}k_c < z_C^2 > = \frac{1}{2}k_B T \quad (1.4)$$

where k_B is the Boltzmann constant and T is the absolute temperature. The cantilevers used in SMFS experiments usually have relatively low spring constants (< 100 pN/nm) comparing to those used for imaging purposes.

For protein folding-unfolding experiments carried out by AFM-based SMFS, the target protein is usually built into a long polyprotein construct, in order to improve the statistical evaluation, help identify single-molecule events and reduce nonspecific interferences.[111] During an AFM-based SMFS experiment, the scanner first keeps moving toward and away from the cantilever in order to pick up and tether a polyprotein molecule between the tip and the substrate. Nonspecific adsorption remains the simplest and most common attachment method for AFM-based SMFS, while sophisticated specific attachment schemes have been developed to increase the pickup rate of single molecules and to enable long-time manipulations.[112] Once a single molecule is picked up, various modes of experiments can be carried out, and high-resolution force-extension-time relation during the folding-unfolding process of single protein molecules can be recorded.

1.2.2 Optical Tweezers-Based Single-Molecule Force Spectroscopy

The effect of optical forces was first demonstrated in the early 1970s by Arthur Ashkin, and the instrument known as optical tweezers (OT) was invented in 1986.[113-114] Using a potential well formed by light, OT can trap dielectric particles with diameters ranging from 100 nm to 10 μm, including not only common experimental probes such as silica and plastic beads, but also living cells, bacteria, viruses and even functioning cell organelles within cells.[115]

The optical forces generated by OT are a result of the transfer of momentum from the beam to the dielectric object, which has a higher refractive index than that of its surrounding medium. As shown in Figure 1.6A, if the object is at the focus of the beam, the rays do not deviate, and the optical force is zero. If the object is in front of the focus, the beam converges, and the axial momentum flux of the beam decreases. If the object is behind the focus, the beam diverges, and the axial momentum flux of the beam increases. In these cases, momentum is

transferred between the beam and the object, generating a restoring force driving the object toward the focus. If the object is displaced sideways, the beam is deflected toward the center of the object and gains lateral momentum, resulting in a lateral reaction force on the object toward the beam axis. These forces moving the object to the focus of the beam in all three dimensions are called the gradient force. Another force that arises due to the reflection of the beam, called scattering force, can be either overcome by the strong gradient force, or cancelled out by using a counter-propagating laser in a dual-laser OT system.[116]

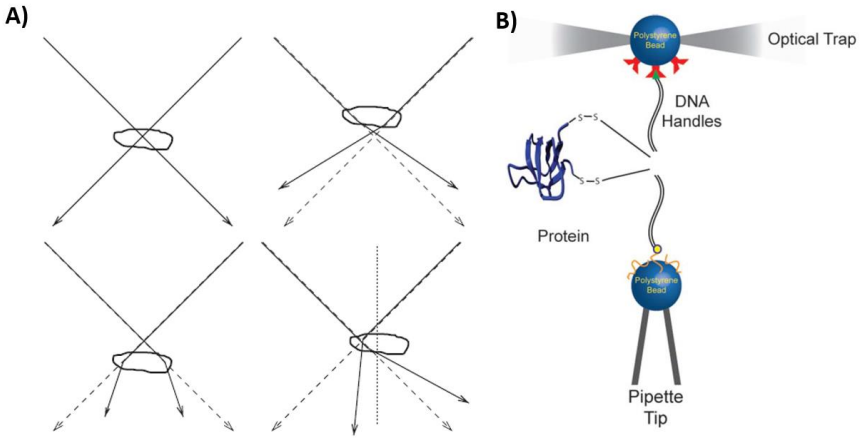

Figure 1.6 Schematic of OT experiments. (A) The force exerted on an object in OT arises from the transfer of momentum between the beam and the object. The gradient force drives the object to the focus of the beam in all three dimensions. Image adapted with permission from Elsevier Inc, Copyright © 2007.[116] (B) Schematic representation of the experimental setup used to apply force on single protein molecules with single-trap optical tweezers. Double-stranded DNA molecules are linked to specific cysteine residues on the protein via disulfide bonds, and act as handles to apply force on the protein. Image adapted with permission from Wiley Periodicals, Inc., Copyright © 2013.[13]

In practice, a high-intensity laser with a Gaussian output intensity profile is tightly focused by an objective lens with a high numerical aperture (>1.2) to create a large spatial gradient of light intensity, resulting in a strong gradient force. Most OT instruments dealing with biological samples use infrared lasers with a wavelength between 700 to 1300 nm, as the light in this region has minimal absorption by biological tissues.[117] In protein folding-unfolding experiments, spherical, isotropic polystyrene beads are frequently used as probes, with refractive

indices that are bigger than that of the surrounding aqueous solution. The size of the polystyrene beads is usually similar to the wavelength of the laser (0.1-10 λ). Within small displacements (<~150 nm) of the trapped object from its equilibrium position, the optical trap acts approximately as a Hookean spring.[118-119] The spring constant (the stiffness of the trap) depends on the steepness of the optical gradient, and can be calibrated by measuring and analyzing either the trajectory of the trapped object in a fixed trap, or the effect on the trapped object of a known force generated by fluid flow.[118, 120] The spring constant of OT is usually very low (<1 pN/nm) compared to that of an AFM cantilever, giving OT a higher precision in force measurements. The most versatile and sensitive position detection method is back-focal plane interferometry, where a PSD detects the interference between the light scattered by the trapped bead and the unscattered light, to measure the three-dimensional displacement of the bead relative to its equilibrium position.[121-122]

In most of the protein folding-unfolding studies using OT, the target protein molecule is sandwiched by two double-stranded DNA (dsDNA) handles at two solvent-exposed cysteine residues to form a DNA-protein-DNA chimera, via thiol reaction, click chemistry, protein tags or DNA oligonucleotide hybridization (Figure 1.6B).[123-127] The DNA handles act as spacers to prevent interference from nonspecific interactions, and the positions of the anchoring points can be adjusted as needed (by moving the cysteine residues to different positions) in order to apply force on specific parts of the protein or in specific directions.[128] The DNA-protein-DNA chimera is tethered between two polystyrene beads by ligand-receptor pairs or antibody-antigen recognition.[129] One of the beads is held and moved by the optical trap to apply force on the tethered molecule, while the other one can be sucked onto a fixed pipette or held by another optical trap. It is worth mentioning that dsDNA undergoes the B-S transition that leads to a force

plateau at ~65 pN, which can be used for identifying single-molecule events, but may also cause interference if protein unfolding events happen in that force region.

A brief comparison of the instrumental specifications of AFM and OT used in SMFS is summarized in Table 1.1.[103]

Table 1.1 Comparison of AFM and OT as SMFS techniques. Table adapted with permission from Nature Publishing Group, Copyright © 2008.[103]

	AFM	OT
Spatial resolution (nm)	0.5-1	0.1-2
Temporal resolution (s)	10^{-3}	10^{-4}
Stiffness (pN nm^{-1})	$10\text{-}10^{5}$	0.005-1
Force range (pN)	$10\text{-}10^{4}$	0.1-100
Displacement range (nm)	$0.5\text{-}10^{4}$	$0.1\text{-}10^{5}$
Probe size (μm)	100-250	0.25-5
Features	High-resolution imaging	Low-noise and low-drift dumbbell geometry
Limitations	Large probe; large minimal force; nonspecific interactions	Photodamage; sample heating; nonspecific interactions

1.2.3 Constant Force Mode of Single-Molecule Force Spectroscopy

Several experimental modes are used in SMFS experiments to study the protein folding-unfolding process, including constant force mode, constant velocity mode, constant distance mode, force ramp mode, etc. These experimental modes are complementary to each other and can access different information regarding the protein folding-unfolding behavior. As the two

most frequently used ones, the principles and implementations of constant force mode and constant velocity mode are introduced here in detail.

Constant force mode is also known as force clamp mode. In the constant force mode of SMFS, the target molecule is subjected to a constant force for a certain time, and the extension of the molecule is measured as a function of time. Protein unfolding is accompanied by an extension of its end-to-end distance, which can be clearly observed as a sharp increase of extension on the extension-time trace. The folding, on the contrary, leads to a decrease in extension. The increase and decrease of extension correlate to the number of amino acids being released and contracted, as well as the probing force. A high and a low force are usually applied alternately to trigger the protein unfolding and folding, but for mechanically labile proteins that unfold and fold around the same force, an unfolding-folding fluctuation can be observed around an equilibrium force (Figure 1.7A). Figure 1.7A shows a representative molecular extension-time trace of the unfolding-folding process of a protein probed by OT-based SMFS in constant force mode.

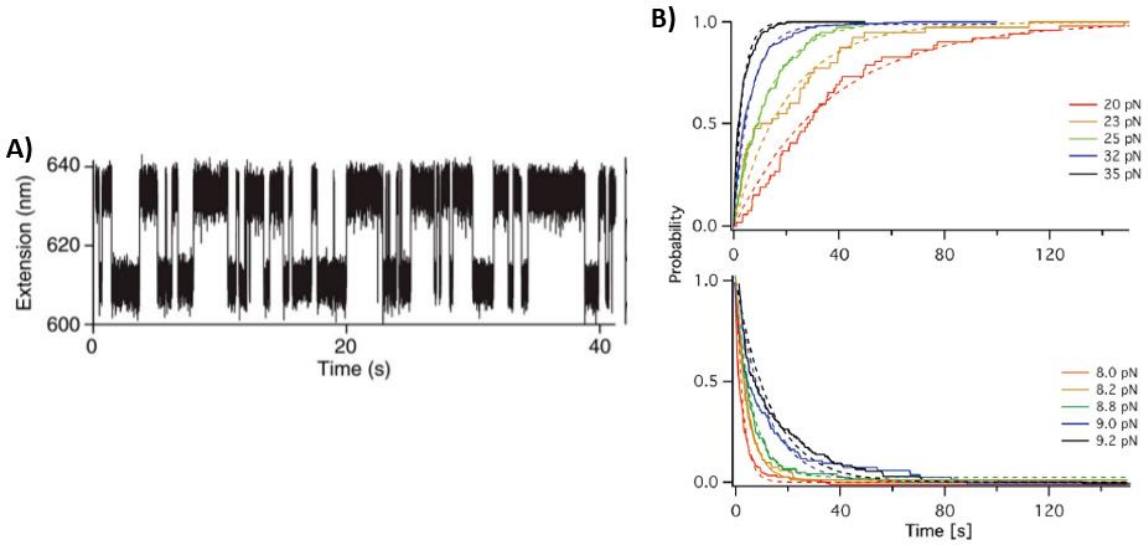

Figure 1.7 Constant force mode of SMFS. (A) A representative molecular extension-time trace of a prion protein PrP under a constant force probed by OT-based SMFS. Image adapted with permission from National Academy of

Sciences, Copyright © 2012.[130] (B) Survival time analysis of protein NuG2 unfolding (upper panel) and folding (lower panel) at different forces. Dotted lines are fits to the survival probability distribution. Image adapted with permission from John Wiley and Sons, Copyright © 2017.[131]

The time that a protein domain remains folded under a certain force, called survival time, is used to extract the unfolding kinetics. For a 2-state folding-unfolding protein that is stretched under a constant force, the unfolding probability is expressed as the following,

$$P_u(t) = 1 - \exp(-\alpha(F) \cdot t) \tag{1.5}$$

where $P_u(t)$ is the unfolding probability, $\alpha(F)$ is the protein unfolding rate constant at the force F, and t is the survival time.[132-133] Taking the derivative of probability with respect to time, the probability density of unfolding at time t is given by:

$$\frac{dP_u(t)}{dt} = \alpha(F) \cdot \exp(-\alpha(F) \cdot t) \tag{1.6}$$

where $dP_u(t)/dt$ is the unfolding probability density, which decays exponentially over time. As the distribution of survival time equals the probability density of unfolding at time t, by fitting the distribution of survival time, the unfolding rate constant at a given force F could be obtained (Figure 1.7B). Similarly, the folding kinetics of protein can be probed, and the folding probability density can be expressed as:

$$\frac{dP_f(t)}{dt} = \beta(F) \cdot \exp(-\beta(F) \cdot t) \tag{1.7}$$

where $dP_f(t)/dt$ is the folding probability density, and $\beta(F)$ is the folding rate constant under the constant force F (Figure 1.7B).[132-133] The unfolding and folding rates at different forces can be acquired by adjusting the force applied to the molecule, and fitting the rate-force relationship with the Bell-Evans model will reveal the intrinsic folding and unfolding rate constants at zero force. In practice, since protein folding usually happens at low forces with small changes of

extension, directly observing folding events in constant force mode requires high resolution and high signal-to-noise ratio of extension at low forces, which makes OT a better choice over AFM for this application.

1.2.4 Constant Velocity Mode of Single-Molecule Force Spectroscopy

Constant velocity mode, which is also known as the force-extension mode or distance ramp mode, is another frequently used experimental mode in SMFS. After the target molecule is tethered, the piezoelectric scanner in AFM or the optical trap in OT is moved back and forth at a constant velocity, so that the target molecule is stretched and relaxed cyclically. The force-distance relationship of the molecule is measured during the process and can also be converted to force-extension data of the molecule, by subtracting the movement of the probe (cantilever tip in AFM or trapped bead in OT) from the change of distance. In constant velocity mode, protein unfolding and folding events are visualized as sudden force drops and rises, respectively. Representative force-distance traces under the constant velocity mode by AFM and OT are shown in Figure 1.8A and Figure 1.9A.

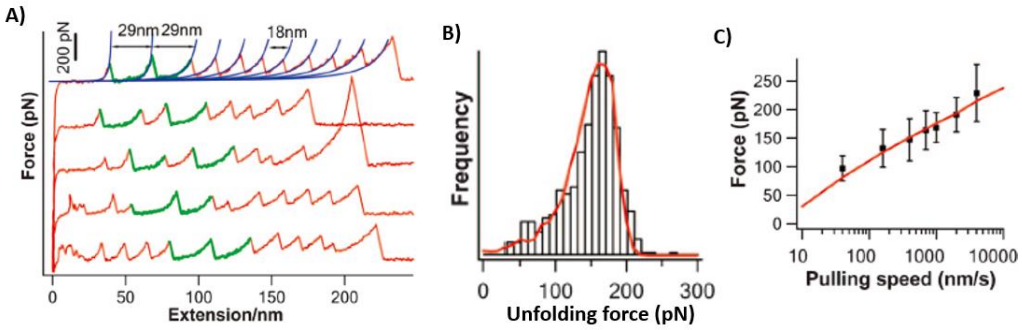

Figure 1.8 Unfolding of Top7 studied by AFM-based SMFS. (A) Force-extension curves of $(GB1)_4$-$(Top7)_2$-$(GB1)_4$. GB1 unfolding events are colored in red and Top7 in green. Blue curves are WLC fits to the force-extension curves. (B) Unfolding force histogram of Top7 at a pulling speed of 400 nm/s. The red line is the Monte Carlo simulation result of the mechanical unfolding of Top7. (C) Pulling speed dependence of the unfolding of Top7. Black symbols are experimental data and the red line is the Monte Carlo simulation result of the mechanical unfolding of Top7 using the same parameters. Images adapted with permission from National Academy of Sciences, Copyright © 2007.[134]

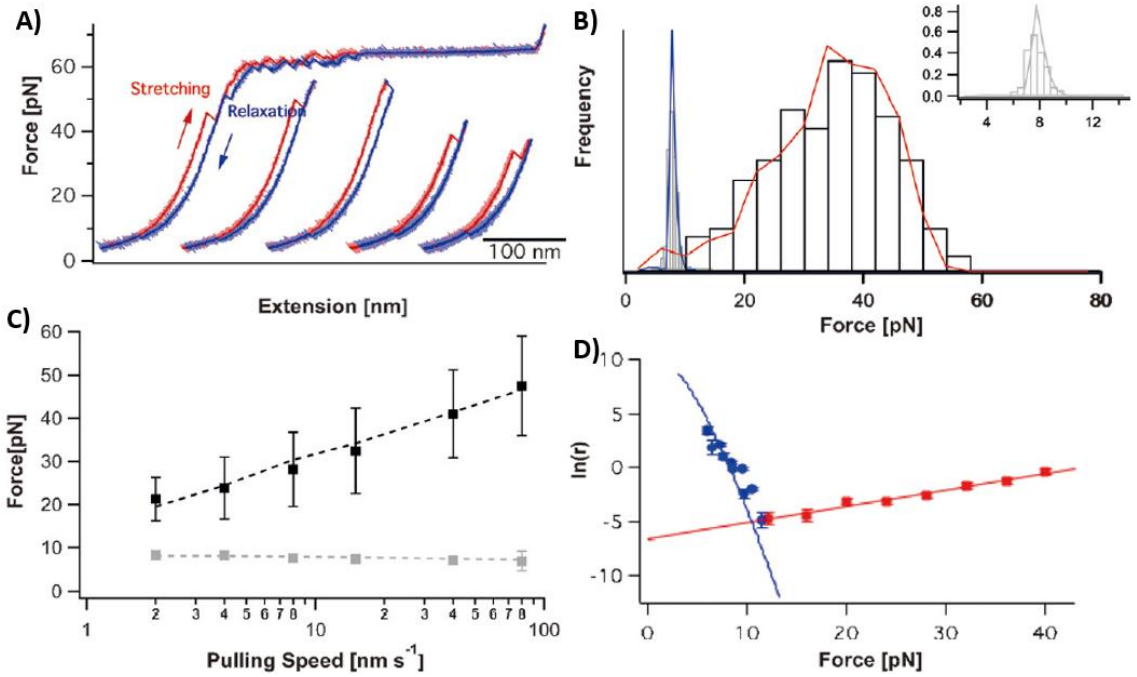

Figure 1.9 Unfolding-folding of NuG2 studied by OT-based SMFS. (A) Force-extension curves of NuG2 under constant pulling speed. (B) Force histograms of the folding-unfolding of NuG2 at a pulling speed of 15 nm/s. Inset is a zoomed view of the folding force histogram. (C) Pulling speed dependence of the unfolding (black) and folding (grey) forces of Top7. Dash lines are Monte Carlo simulation results. (D) Force-dependent folding-unfolding rates measured for NuG2. The solid lines are fits of the Bell-Evans model to the experimental data. Image adapted with permission from John Wiley and Sons, Copyright © 2017.[131]

The relationship between the force and the extension of proteins can be described by the worm-like chain (WLC) model of polymer elasticity:

$$F_{protein}(x) = \frac{k_B T}{p}\left[\frac{1}{4}\left(1-\frac{x}{L_C}\right)^{-2} - \frac{1}{4} + \frac{x}{L_C}\right] \quad (1.8)$$

where $F(x)$ is the entropic force at the extension of x, Lc is the contour length of the polymer, k_B is the Boltzmann constant, T is the absolute temperature, and p is the persistence length of the protein, which describes the rigidity of the polymer.[111, 135-137] As AFM-based SMFS directly probes polyproteins, the WLC model can be used to fit the sawtooth-like force-extension data, and the contour length increment (ΔLc) upon protein unfolding can be obtained, which contains important information about the structural change during protein unfolding.

The unfolding and folding forces of a protein stretched under constant velocity conditions, of which the distributions are usually displayed in histograms, contain underlying information about the protein unfolding and folding kinetics (Figure 1.8B, Figure 1.9B).[91, 95] For polyproteins stretched by AFM under constant velocity, there is no analytical solution for the most probable unfolding force. Therefore, a Monte Carlo simulation method is commonly used to extract the unfolding and folding kinetic parameters of the protein. The input kinetic parameters that lead to the best reproduction of the experimental results are considered as the kinetic parameters of the protein (Figure 1.8B).[96, 138] The pulling speed dependence data of unfolding and folding forces are also commonly used in Monte Carlo simulations, as it provides an additional and more straightforward comparison between the experimental and simulation results compared to only using the force histograms (Figure 1.8C).

For OT-based SMFS, as the target proteins are usually built into DNA-protein chimeras, the elastic response of the DNA handles in addition to the protein needs to be considered.[139] Since OT has a higher resolution at low forces, it is better than AFM at observing protein folding events which usually happen under these conditions. To extract the kinetic parameters from constant velocity data measured by OT, Monte Carlo simulation can also be used, with the only difference in generating force-extension curves for DNA-protein chimera instead of protein only in AFM experiments (Figure 1.9C). Another method developed by Oesterhelt *et al*. transforms constant velocity data to constant force data and can directly extract the dependency of unfolding and folding rate constants on force (Figure 1.9D).[140]

1.3 Metalloproteins

1.3.1 Types and Functions of Metalloproteins

Metalloproteins are defined as proteins containing at least one metal cofactor, including metal ions, metal clusters and larger chemical groups containing metals.[141] It is estimated that more than 30% of all the proteins in the cells rely on one or more metal cofactors to perform their biological functions.[142] A systematic bioinformatics survey of 1371 different enzymes showed that 47% of them required metals, with 41% containing metals at their catalytic centers.[143] To date, 13 metals have been found in naturally occurring metalloproteins, including 2 alkali metals (Na and K), 2 alkaline earth metals (Ca and Mg), and 9 d-block transition metals (V, Cr, Mo, Mn, Fe, Co, Ni, Cu and Zn).[20, 144] The common ligands of these metals in proteins are S of cysteine and methionine, N of histidine, and O of glutamic acid, aspartic acid and tyrosine. Potential ligands also include O of peptide carbonyl moieties, threonine and serine, N of deprotonated backbone amides and lysine side chains, and some modified amino acids.[145] The coordination interaction between metals and their ligands in proteins involves both electrostatic and covalent interactions.[22, 146] The combination of metal and ligand generally follows the prediction by the hard-soft theory of acids and bases, in which hard (small and non-polarizable) acids (Na, K, Ca, Mg and some of the transition metals) prefer hard bases (mostly carboxylates and backbone O), and soft (large and polarizable) or borderline acids (most of the transition metals) prefer soft or borderline bases (thiolates, imidazoles, phenolates, etc.). The exact coordination number and geometry of each metal site are also determined by the metal's oxidation state.[20]

The function of a metalloprotein depends subtly and elegantly on the specific chemical properties of the metal it contains, and the proteins' primary and secondary metal coordination

spheres.[147] Na and K are frequently incorporated in metalloenzymes, such as diol dehydratases.[148] Mg is the most prevalent metal in metalloenzymes, and several important enzymes including DNA/RNA polymerase and ATPase require Mg to perform their functions.[148] Ca sensing metalloproteins such as calmodulin are important for cellular signal transduction.[149] Metalloproteins containing redox-active metals, such as Fe and Cu, are perfect candidates for electron carriers (e.g., iron-sulfur proteins, heme-binding cytochromes and blue-copper proteins).[22] Some other Fe and Cu metalloproteins, such as hemoglobin, myoglobin and hemocyanin, are involved in O_2 storage and carriage due to their suitable affinity to O_2.[22] In contrast, Zn mostly serves as a superacid center in metalloproteins catalyzing hydrolysis and bond cleavage reactions (e.g., carboxypeptidases, carbonic anhydrase and alcohol dehydrogenase), and plays structural roles in metalloproteins such as superoxide dismutase and zinc finger motifs.[22] Metalloproteins containing other trace metals are mostly enzymes, including Co-vitamin B_{12} containing enzymes, Mn-containing superoxide dismutase and inorganic phosphatase, Ni-containing urease and hydrogenase, Mo- and V-containing nitrogenase, etc.[142, 144] Incorporating metals to build metalloproteins greatly expands the biological functions of proteins.

1.3.2 Biosynthesis of Metalloproteins

The biosynthesis of metalloproteins requires more than that of metal-free proteins. Apo-form proteins (proteins without cofactors) need to bind to the correct metals to become holo-form (proteins with cofactors), and some complicated metal cofactors such as heme or some metal clusters need to be synthesized first.[22] Few of the biosynthesis mechanisms of metalloproteins have been completely revealed so far, but several different scenarios can be outlined based on our current knowledge (Figure 1.10).[22, 146] The insertion of the metal cofactor

is likely to occur at different steps of the protein biogenesis process: during the translation; after the polypeptide is released but before the polypeptide completely folds; or, as the last step, after the polypeptide has adopted its folded conformation.

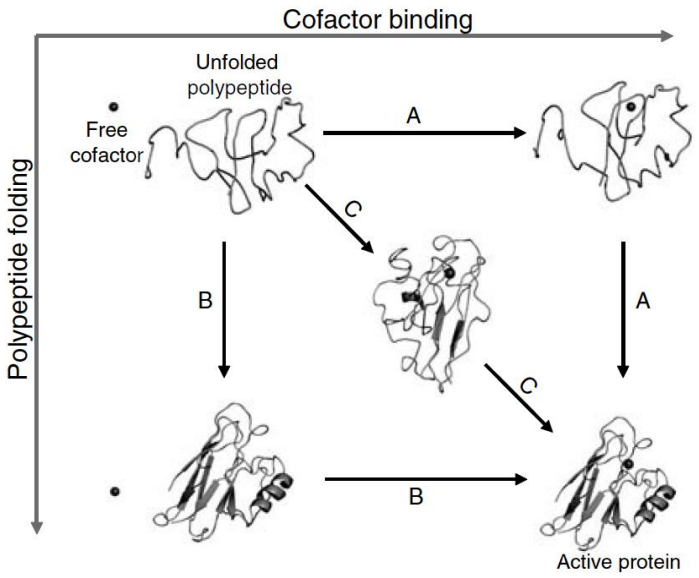

Figure 1.10 Illustration of possible metalloprotein biosynthesis mechanisms. The cofactor may bind before polypeptide folding (path A), after complete protein folding (path B), or to a partially folded protein structure (path C). Image adapted with permission from Cambridge University Press, Copyright © 2004.[150]

Metal sites on flexible polypeptide chains with imperfect steric selections tend to bind metals in a universal order of preference, which is described by the Irving-Williams series (Ca^{2+} < Mg^{2+} < Mn^{2+} < Fe^{2+} < Co^{2+} < Ni^{2+} < Cu^{2+} > Zn^{2+}).[21, 151] However, cells adopt multiple strategies to ensure that metalloproteins get the correct metal. Firstly, apo-form proteins fold in different cellular compartments with different metal concentrations, and the metal concentration sequence in that cellular compartment can sometimes override the metal binding affinity sequence, making the binding to metals with weaker binding affinities possible. For example, in the cyanobacterium, cupin CucA and MncA have identical sets of ligands but they bind Cu and Mn, respectively. The reason is that MncA folds in the cytoplasm where copper is either absent or tightly bound, and CucA folds in the periplasm after export thus acquiring the more

competitive copper.[152] Additionally, some metal cofactors are delivered to the apo-proteins by protein-based metal-delivery systems through sequential ligand-exchange reactions, where the specificity of protein-protein interactions determines which apo-form protein receives the metal.[153] For example, copper ions are mostly delivered to their destination by copper chaperones, for the redox properties of copper can cause oxidative damage to cells.[154-155] Iron can be transferred and released within the endosome by transferrins, and iron-sulfur clusters require a complex machinery encoded by the *isc* operon to be synthesized and transported.[156-158] For cytochrome c, the heme is covalently attached to the unfolded apo-form protein by a heme-lyase enzyme, and then the holo-form protein can fold into its native structure.[20, 144]

1.3.3 Folding of Metalloproteins

Upon metal binding, the protein conformation undergoes a series of adjustments for a compromise between protein conformation and metal-coordination (in terms of coordination number, bond length, bond angle, etc).[21] The holo-form metalloprotein is generally more stable than the apo-form under physiological conditions, however, to what degree metal binding modulates the folded conformation of protein differs case by case. In cases that the apo-form protein itself can already fold into an organized tertiary structure with strong hydrophobic interactions between secondary structures, such as in blue copper proteins and Cu transport protein Ccc2, the metal ion only introduces minimal conformational change at the binding site, while the overall protein structure is essentially maintained.[159-160] In addition, metal binding also rigidifies the structure and decreases its dynamics. In cases such as superoxide dismutase, zinc finger domains and most of the calcium binding proteins, metal ions play a key role in inducing or maintaining the correct arrangement of the structural elements.[161-163] In these cases, metal

binding drastically changes the folded apo-form conformation into a more stable folded holo-form conformation.

The folding of metalloproteins conforms to the same energy landscape principles, as does the folding of metal-free proteins. Essentially, the metal coordination introduces an additional coordination landscape, which interplays with the apo-form protein folding landscape. Depending on the metal binding mode, the coordination landscape could be either funnel-like (with specific binding sites) or rugged (with multiple competitive binding patterns), resulting in a composite energy landscape with different properties (Figure 1.11).[164] Specifically, nonspecific metal coordination increases the ruggedness of the composite landscape, which may contain multiple trapped conformations. Examples are many neurodegenerative diseases-causing proteins, which misfold and aggregate upon such nonspecific metal binding.[165] In contrast, the funnel-shaped coordination landscape could stabilize a certain region of the composite energy landscape. For polypeptide chains without a stable folded structure, the metal binding will lead to a stable folded structure (e.g, zinc-finger proteins); and for a polypeptide chain with a folding funnel, it will either further stabilize the funnel (e.g., azurin), or form a two-basin funnel landscape (e.g., calmodulin), depending on whether the positions of the two funnels overlap.[166-169]

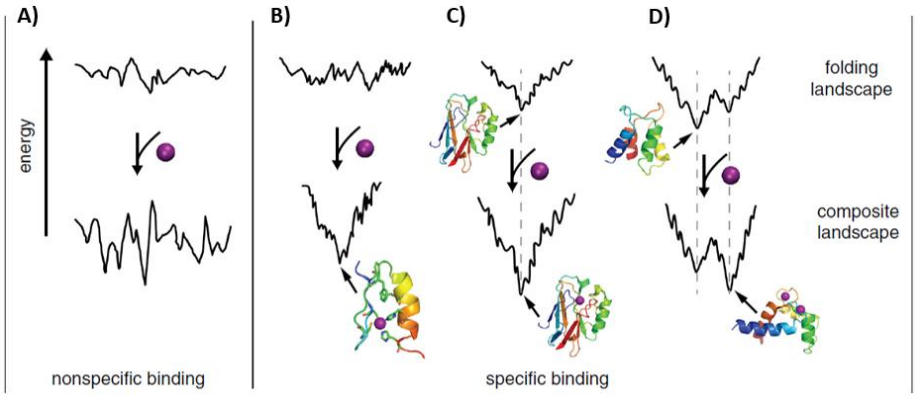

Figure 1.11 Schematic diagram of the effects of metal binding on the energy landscapes of proteins. (A) Nonspecific binding leads to a rugged composite landscape due to multiple competitive coordination modes; (B) binding creates a funnel on the composite landscape when the coordination landscape dominates the overall interactions; (C) binding induces a deeper composite landscape when the coordination landscape is consistent with the folding landscape; and (D) binding reshapes the folding landscape when the coordination landscape matches a metastable state on the folding funnel. Image adapted with permission from Elsevier Ltd., Copyright © 2014.[164]

1.3.4 Folding-Unfolding Mechanisms of Metalloproteins Revealed by Ensemble Spectroscopic Methods

The above-discussed folding energy landscape theory of metalloproteins only provides a qualitative energetic result of metal binding in protein folding. In terms of the folding pathway or mechanism of a specific metalloprotein, the problem becomes extremely complicated. *In vitro* spectroscopic protein folding-unfolding experiments face many challenges, such as the irreversibility of unfolding and metal dissociation. Nevertheless, extensive studies have been done on various categories of metalloproteins including heme-binding proteins, copper-binding proteins and iron-sulfur proteins, and ensemble spectroscopic experiments have provided important insights into the folding mechanisms of metalloproteins.[146] Some examples are as follows.

Cytochrome c (cytc). Cytc is a monomeric metalloprotein containing 80-120 residues and a covalently bound heme cofactor. Early chemical denaturation studies about the folding-unfolding behavior of cytc revealed a two-state folding-unfolding mechanism[170], while later

spectroscopic and calorimetric studies revealed the existence of an intermediate state with an altered helix[171-173]. Over and above standard spectroscopic methods, hydrogen exchange experiments demonstrated the presence of five folding-unfolding units (foldons) in cytc, and suggested that cytc folds by a stepwise assembly of the five foldon units.[174] It has also been found in ensemble experiments that the non-native coordination to the heme and proline isomerization can cause misfolding of cytc.[175] In addition, researchers found that the folding-unfolding behaviors of several cytc proteins with large differences in sequence, size and charge follow a similar three-state model under moderate denaturant concentrations, and the interaction between N- and C-terminal α-helices might act as a nucleus for folding in this protein family.[22, 176]

Copper-binding proteins. The folding-unfolding behavior of a copper-binding protein, azurin from *Pseudomonas aeruginosa*, has been extensively studied by ensemble methods, for it is a structurally simple single-domain metalloprotein with a common Greek-key fold (a protein fold in which four β-strands arrange into an antiparallel β-sheet) and one Cu^{2+} ion.[177] Different spectroscopic methods revealed a similar two-state folding-unfolding behavior, and it was found that the Cu^{2+} ion remained bound to the unfolded polypeptide chain.[178] In addition, modeled-peptide studies revealed that the β-like secondary structure of azurin formed upon the Cu^{2+} binding, thus it was proposed that the metal coordination may create a nucleation site for the folding of azurin.[179] Similar coordination-nucleation-folding behavior has also been observed in the folding processes of several other metalloproteins.[179-184] While the *in vivo* biogenesis process of azurin remains unclear, based on the observation in ensemble folding-unfolding experiments that an unfolded holo-form azurin reaches its native state (in milliseconds) much faster than a

folded apo-form azurin in the presence of copper (in minutes to hours), it was speculated that in cells the Cu^{2+} ion is delivered to an unfolded azurin before it folds.[185]

Iron-sulfur proteins. Iron-sulfur proteins are proteins containing iron ions or iron-sulfur clusters. The assembly and insertion of iron-sulfur clusters *in vivo* are not spontaneous processes but rather assisted by complex biogenesis systems.[186] The *in vitro* unfolding of iron-sulfur proteins by denaturants is generally irreversible as a result of the disintegration of the iron-sulfur centers, bringing a major obstacle to the folding-unfolding study of iron-sulfur proteins *in vitro*. To overcome this, some researchers focused on the unfolding process of iron-sulfur proteins, which may elicit a certain interaction or an intermediate state that may also be vital in the reverse folding process. The thermal denaturation of rubredoxin from *Methanocaldococcus jannaschii* was found to occur via a simple two-state process, and the folding processes of chemically denatured rubredoxins from *Clostridium pasteurianum* and *Pyrococcus furiosus* were found to follow an "iron-priming" mechanism.[187-188] More complex iron-sulfur proteins containing multiple iron-sulfur clusters have also been studied. For example, the unfolding of a di-cluster ferredoxin containing a [3Fe-4S] and a [4Fe-4S] cluster from *Acidianus ambivalens* has been found to follow different pathways under acidic, basic and thermal denaturing conditions.[189]

1.3.5 Investigating the Folding-Unfolding Mechanisms of Metalloproteins with Single-Molecule Force Spectroscopy

The folding-unfolding mechanisms of metalloproteins have also been studied by SMFS, a powerful tool in the protein folding-unfolding studies that is complementary to the ensemble spectroscopic methods. Significant progress in understanding the folding-unfolding mechanisms of metalloproteins has been achieved by SMFS over the past few years.[14, 23-39]

The application of SMFS to metal-coordination systems first started from non-biological metal-ligand systems, where the goal was to directly measure the mechanical strengths of coordination bonds. AFM-based SMFS were frequently used in such studies, as the rupture of mechanically stable coordination bonds sometimes requires forces >1 nN. In such experiments, the metal and the ligand were immobilized on the AFM tip and substrate respectively, usually with a polymer linker as a spacer to prevent short-range non-specific interactions. A variety of metal-ligand interactions have been studied, including Au-S bond, metal-NTA (nitrilotriacetate)/His-tag complexes, metal-terpyridine complexes, etc.[190-192] Large variations in rupture force were observed in these metal coordination systems (e.g., ~1.4 nN for Au-S and bonds and ~20 pN for Co^{2+}-NTA/His-tag complex), and factors such as the environmental pH and the redox state of the metal were found to modulate the mechanical stability of the coordination bonds.[190-192]

Researchers have also exploited the effects brought by metal binding to proteins by introducing ligands into metal-free proteins to construct artificial metalloproteins. The comparison between the unfolding behaviors of the artificial metalloproteins and the original metal-free proteins probed by SMFS provides a direct description of the metal binding effect on proteins at the single-molecule level. One such well-studied artificial metalloprotein is the bi-histidine mutant of protein GB1, which can bind various metal ions including Ni^{2+}, Co^{2+}, Co^{3+}, etc.[193-194] In the case of GB1, the metal binding can either mechanically stabilize or bring no obvious influence on the mechanical unfolding of the protein, depending on the type of the bound metal ion and the positions of the two introduced histidine ligands. This strategy was also used to modify the unfolding pathways of metal-free proteins in order to rationally tune the mechanical stability of proteins.[195]

Based on the invaluable results obtained from the above-mentioned studies, SMFS was soon applied to naturally occurring metalloproteins to study the folding-unfolding mechanisms of metalloproteins. The first two metalloproteins that were studied in detail by SMFS are calmodulin (CaM) and rubredoxin (RD), and they demonstrated two distinct mechanical folding-unfolding behaviors that are likely to be ubiquitous among metalloproteins.[14, 23, 34]

The first typical mechanical folding-unfolding behavior of metalloproteins, of which CaM is representative, features a synergistic effect between protein folding and metal binding. The rupture of the coordination bonds in this type of metalloprotein cannot be observed as individual events in SMFS. Generally, this type of behavior is usually from metalloproteins containing alkaline earth metals and likely alkali metals as well, which have metal coordination bonds that are more ionic in nature and of relatively low mechanical stabilities. The metal binding process for this type of metalloproteins is mostly reversible in solution, and the protein folding and metal binding usually happen in a cooperative way. CaM is the most important calcium binding protein in the human body that regulates calcium-dependent signaling cascades. Holo-CaM consists of two globular domains connected by an α-helical linker. Each of the globular domains contains two helix-loop-helix motifs, which can bind two calcium ions. The apo-CaM folds into a flexible and highly dynamic structure. The folding mechanisms of holo- and apo-CaM were carefully investigated with both AFM- and OT-based SMFS.[34-36] It was found that holo-CaM folds through a complex network of two on-pathway and two off-pathway intermediates, with different calcium-binding motifs folded. The metal coordination and the protein structure form and get disrupted in a cooperative fashion, so that no individual bond rupture event was observed in the SMFS experiment.[34] Apo-CaM, in contrast, folds in a simple two-state pathway.[36] A similar synergistic effect between the protein folding and metal binding

was also observed for several other calcium-binding proteins, including C-terminal repeats-in-toxin domain of CyaA, B1 domain of SdrG and biofilm-associated proteins.[37-39] While the mechanical strengths of the coordination bonds in this type of metalloproteins are considered to be relatively low, it is noteworthy that the metalloproteins themselves can be mechanically very stable. For example, the B1 domain of SdrG is the mechanically strongest protein fold discovered to date, which unfolds at ~2 nN.[38]

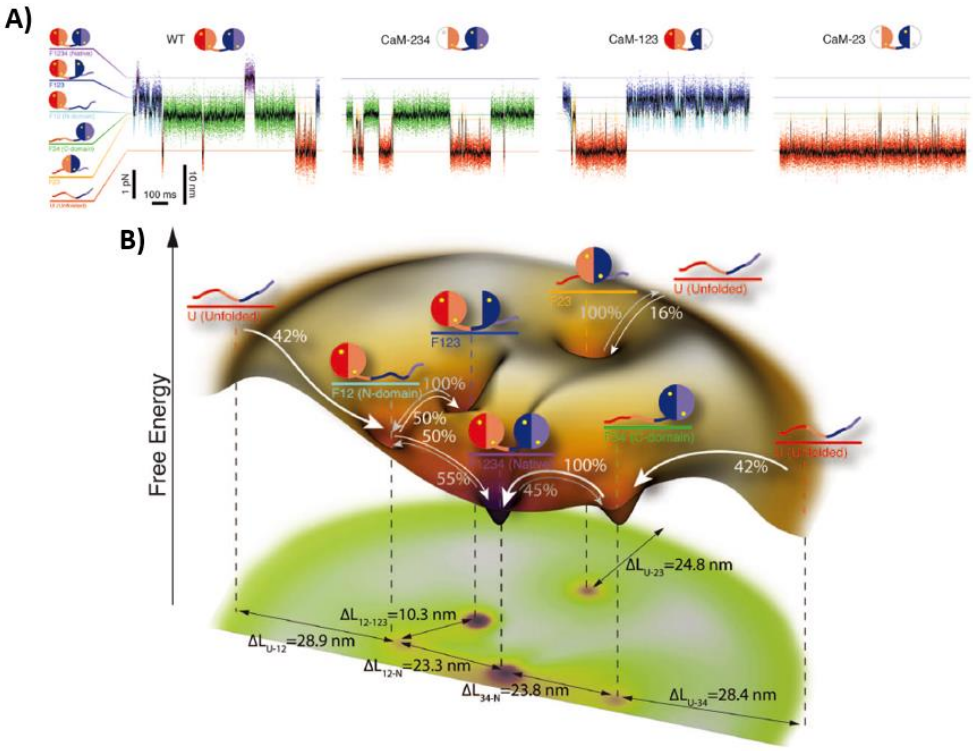

Figure 1.12 The folding mechanism of CaM revealed by SMFS. (A) Sample trace of holo-CaM fluctuation shows four intermediate states. (B) The kinetic network of holo-CaM folding and unfolding at zero load obtained from SMFS experiments. Images adapted with permission from the American Association for the Advancement of Science, Copyright © 2011.[34]

The second typical mechanical folding-unfolding behavior of metalloproteins is represented by RD, featuring mechanically stable metal coordination centers that can be retained despite the absence of a folded protein structure. In SMFS experiments, it is possible to observe the rupture of the coordination bonds in these metalloproteins as individual events in addition to

protein unfolding events. This type of mechanical unfolding behavior is usually found in metalloproteins containing transition metals, as they form coordination bonds of a more covalent nature and with relatively higher mechanical stabilities. The metal dissociation of these transition metal-containing metalloproteins is largely irreversible in solution due to the hydrolysis of the metals. RD is a typical metalloprotein with such mechanical unfolding behavior. It is a small iron-sulfur protein containing a ferric ion coordinated by four cysteines. Stretching from its N- and C-termini, the force will unfold the outer β-sheet and then rupture the Fe-S bonds. The rupture of the Fe-S bonds leads to an unfolding event of ΔLc of ~13 nm corresponding to the extension of the sequestered polypeptide sequence (residues 5-41), which can be clearly observed in SMFS.[23] Similar behaviors were observed in a variety of other transition metal-containing metalloproteins, including zinc-containing rubredoxin, ferredoxin, azurin, plastocyanin, gold-specific protein GolB, zinc finger protein, etc.[29-33] Detailed unfolding pathways with the unfolding of specific protein structures and the rupture of specific bonds were revealed based on SMFS unfolding experiments. In these metalloproteins, the rupture of the metal coordination bonds can be observed, indicating a relatively stable metal coordination center that can persist with some outer protein structures unfolded in the unfolding intermediate state. In addition, as the bond rupture event can be directed observed, studying the unfolding of these metalloproteins also provides a strategy to systematically measure the bond strength of a wide variety of metal-ligand coordination bonds in biological systems. For example, it was found that the highly covalent Fe-S bond in RD ruptures at ~210 pN, the Cu-S and Cu-N bond in azurin both rupture at ~50 pN, and the Au-S bond in GoIB ruptures at ~165 pN.[23, 29-30] For the same reason, the chemical reactivity of the metal site can also be investigated by inducing the unfolding of metalloproteins with SMFS in solutions containing reactive compounds.[27] It is of

note that whether the rupture of these coordination bonds can be observed in SMFS experiments also depends on the relative positions of the coordination bonds in the protein structure and is limited by the instrumental resolution.

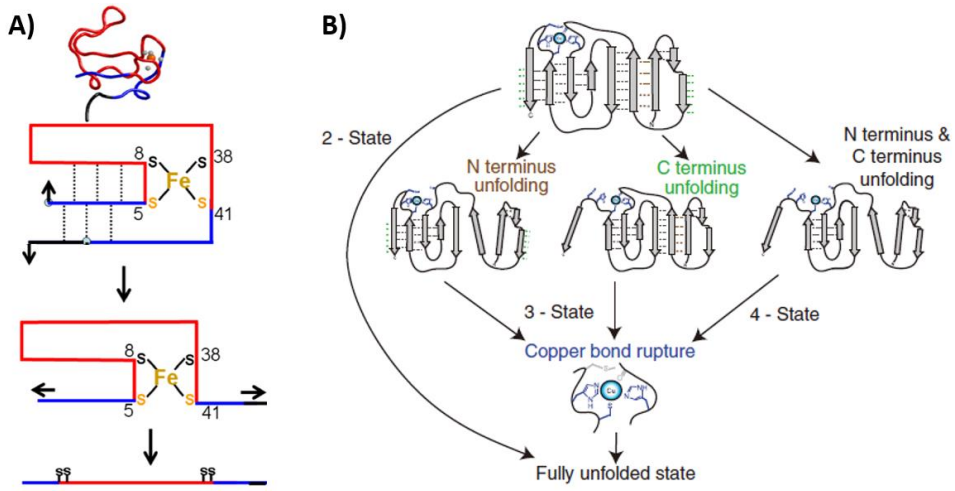

Figure 1.13 The unfolding mechanisms of some transition metal-containing metalloproteins revealed by SMFS. (A) The mechanical unfolding mechanism of RD revealed by SMFS. Image adapted from American Chemical Society, Copyright © 2013.[25] (B) The complex unfolding pathway revealed for azurin and plastocyanin by SMFS. The unfolding intermediate states have a (partially) retained copper coordination center. Image adapted with permission from Springer Nature, Copyright © 2015.[29]

As an emerging method in the toolbox for studying metalloproteins, SMFS has demonstrated its great potential in revealing new and complementary insights into metalloproteins that are otherwise difficult to obtain using traditional ensemble methods. However, more challenges are still ahead, including studying the folding process of transition metal-containing metalloproteins, investigating the folding-unfolding mechanisms of metalloproteins with complex coordination centers and protein structures, obtaining a general understanding of the folding-unfolding mechanisms for metalloproteins based on individual case studies, and revealing the misfolding process of disease-causing metalloproteins, etc.

1.4 Aim of This Thesis

As understanding the folding-unfolding mechanisms of metalloproteins has profound significance in biophysics, and SMFS has demonstrated its superb ability in studying the protein folding-unfolding process, in this thesis, I used SMFS to study the folding-unfolding mechanisms of metalloproteins. In Chapter 2, I used AFM-based SMFS to study the unfolding mechanism of a high potential iron-sulfur protein. I revealed its two paralleled unfolding pathways and proposed a general mechanical unfolding mechanism for the iron-sulfur protein family. In Chapter 3, I exploited OT-based SMFS to investigate the folding mechanism of RD. I discovered a novel "binding-folding-reconstitution" mechanism for the folding of RD, which may result in important implications to our understanding of the folding process of transition metal-containing metalloproteins. In Chapter 4, I used OT-based SMFS to study the folding-unfolding behavior of a [2Fe-2S] ferredoxin. Unfolded ferredoxin does not have well-defined folding pathways and it mostly folds into misfolded structures. The recovery of part of the native structure is possible but very challenging. In Chapter 5, I proceeded to study the folding-unfolding behavior of a heme-containing metalloprotein, cytc, with OT-based SMFS. The resolved folding-unfolding mechanisms provide some new insights into the folding of holo-form and apo-form cytc, and may have further implications on the folding process of other heme-containing metalloproteins. In Chapter 6, conclusions will be drawn based on the above research efforts, and some future research directions in studying the folding-unfolding mechanisms of metalloproteins will be discussed.

Chapter 2: Mechanical Unfolding Pathway of the High Potential Iron-Sulfur Protein Revealed by Single-Molecule Atomic Force Microscopy: Toward a General Unfolding Mechanism for Iron-sulfur Proteins

2.1 Synopsis

High potential iron-sulfur proteins (HiPIPs) are an important class of metalloproteins with a [4Fe-4S] cluster coordinated by four cysteine residues. Distinct from other iron-sulfur proteins, the cluster in HiPIP has a high reduction potential, making it an essential electron carrier in bacterial photosynthesis. Here we combined single-molecule atomic force microscopy and protein engineering techniques to investigate the mechanical unfolding mechanism of HiPIP from *Chromatium tepidum* (cHiPIP). We found that cHiPIP unfolds in a two-step fashion with the protein sequence sequestered by the iron-sulfur center as a stable unfolding intermediate state. The rupture of the iron-sulfur center of cHiPIP proceeds in two distinct parallel pathways; one pathway involves the concurrent rupture of multiple iron-thiolate bonds while the other one involves the sequential rupture of the iron-thiolate bonds. This mechanistic information was further confirmed by mutational studies. We found that the rupture of the iron-thiolate bonds in reduced and oxidized cHiPIP occurred in the range of 150-180 pN at a pulling speed of 400 nm/s, similar to that measured for iron-thiolate bonds in rubredoxin and ferredoxin. Our results may have important implications for understanding the general unfolding mechanism governing iron-sulfur proteins, as well as the mechanism governing the mechanical rupture of the iron-sulfur center.

2.2 Introduction

Metalloproteins are ubiquitous in nature and play important roles in a wide variety of biological processes.[196-197] Compared with proteins that do not contain metals, the metal-ligand bonds in metalloproteins are an important contributing factor to protein stability. Understanding the folding-unfolding mechanisms of metalloproteins as well as the role played by metal and metal-ligand bonds in the process is thus of great significance. Traditional ensemble methods have been widely used to investigate the folding-unfolding mechanism of metalloproteins, but their applications are often limited by the irreversible unfolding of metalloproteins in *in vitro* studies, and the inability to resolve multiple parallel unfolding pathways.[198-200] Over the last two decades, the development of atomic force spectroscopy (AFM)-based SMFS technique has evolved into a powerful tool to investigate the protein folding-unfolding mechanism at the single-molecule level, and the last few years have witnessed a fast development of using AFM to study the unfolding of a diverse range of metalloproteins.[201-208] These studies provided detailed mechanistic insights into the metal center rupture process as well as the protein unfolding pathway, and uncovered surprising mechanical strength of metal-thiolate bonds.

Iron-sulfur proteins constitute a large family of diverse metalloproteins and their structures show remarkable plasticity.[209] Because of the versatile chemical and electronic features of their Fe-S clusters, these proteins are involved in a diverse range of functions, such as electron transfer, regulation of gene expression and iron/sulfur storage.[186] In the vast majority of such iron-sulfur proteins, four cysteine residues are involved in the formation of iron or iron-sulfur cluster chelation sites.[197] Our detailed single-molecule AFM studies on the simplest iron-sulfur proteins,[210-215] i.e., [1Fe-0S] rubredoxin and [2Fe-2S] ferredoxin, have revealed their mechanical unfolding mechanisms explicitly, and provided a good starting point and solid

foundation for further systematic investigation of mechanical unfolding-folding mechanisms of more complexed iron-sulfur proteins. To investigate whether a general mechanism exists for the unfolding and folding mechanism for a broad range of iron-sulfur proteins, here we combine single-molecule AFM and protein engineering techniques to study the mechanical unfolding of a more complex iron-sulfur protein, the [4Fe-4S] high potential iron-sulfur protein (HiPIP).

HiPIPs are an important class of small (9-10 kDa) metalloproteins found mostly in photosynthetic bacteria.[216] HiPIPs have a $[4Fe\text{-}4S]^{2+/3+}$ cubane cluster coordinated by four cysteine residues. Distinct from other iron-sulfur proteins such as rubredoxin and low potential ferredoxin, HiPIPs have a high reduction potential (+50 to +450 mV), making them essential electron carriers in bacterial photosynthesis.[217] Here we use the HiPIP from thermophilic photosynthetic purple bacterium *Chromatium tepidum* (cHiPIP) as a model system for our single-molecule AFM studies. cHiPIP is 83 residues long and contains one [4Fe-4S] cluster at the center of the protein, which is coordinated by four sulfurs from cysteine residues (Cys 43, 46, 61, 75) in a tetrahedral geometry (Figure 2.1).[218-219] The polypeptide chain of cHiPIP wraps around the [4Fe-4S] cluster with several helices and β-strands punctuated by tight turns.[220] We used AFM to stretch cHiPIP from its N- and C-termini to investigate its mechanical unfolding pathway, and the role played by the [4Fe-4S] center in this process.

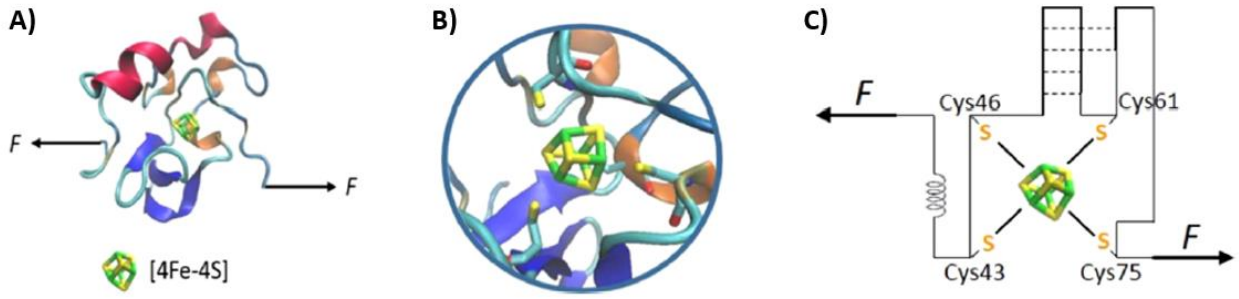

Figure 2.1 Structure of cHiPIP. (A, B) Three-dimensional structure of cHiPIP (PDB code: 2FLA). The [4Fe-4S] cubane cluster is highlighted, with irons colored in green and sulfurs in yellow. The irons are coordinated in a

tetrahedral geometry by both inorganic sulfurs and sulfurs from four cysteine residues. For simplicity, a schematic of cHiPIP is shown in (C), where the spiral indicates the helical structures outside the coordination center and the dotted lines represent backbone hydrogen bonds between β-strands.

2.3 Results

2.3.1 Protein Engineering and Characterization

Full-length holo-form recombinant polyprotein as well as holo-form of cHiPIP alone were successfully expressed and purified (see Experimental Section). On the SDS-PAGE, the purified $(GB1)_4$-cHiPIP-$(GB1)_4$ appeared as a dominant band with a molecular weight of ~60 kDa, consistent with the expected molecular weight of the polyprotein (Figure 2.2A). The purified polyprotein shows a UV absorbance at 380 nm, which is characteristic of the ligand-to-metal charge-transfer transitions, identical to that of the reduced cHiPIP alone (Figure 2.2B). These results indicate that the purified $(GB1)_4$-cHiPIP-$(GB1)_4$ corresponds to the reduced holo-form of cHiPIP.[221]

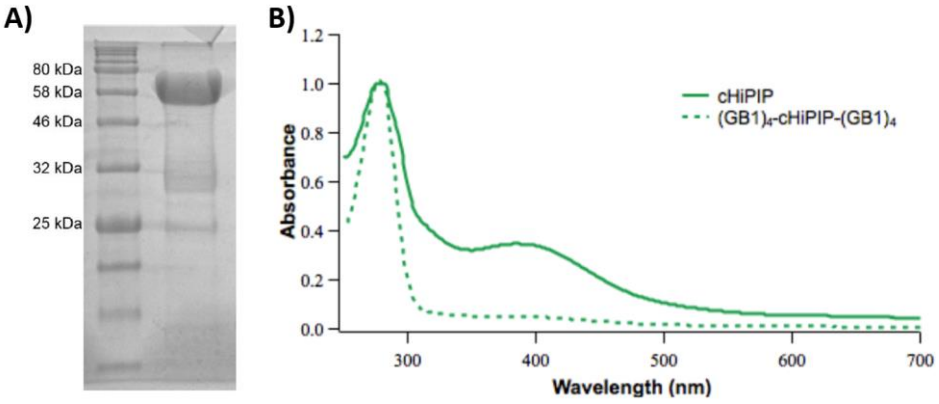

Figure 2.2 Characterizations of the recombinant protein $(GB1)_4$-cHiPIP-$(GB1)_4$. (A) SDS-PAGE photograph of the polyprotein $(GB1)_4$-cHiPIP-$(GB1)_4$. The polyprotein $(GB1)_4$-cHiPIP-$(GB1)_4$ shows a dominant band at ~60 kDa, consistent with the theoretical molecular weight of 62 kDa. The protein and ladders were visualized using the ultraviolet-induced reaction between the tryptophan residues and the trihalocompounds in the gel.[222] (B) UV-Vis spectra of purified cHiPIP and the polyprotein $(GB1)_4$-cHiPIP-$(GB1)_4$. The absorbance at 380 nm is characteristic of the ligand-to-metal charge-transfer transitions of the reduced cHiPIP.

2.3.2 Mechanical Unfolding of cHiPIP Occurs in a Two-Step Process and Results in the Rupture of the Iron-Sulfur Center

Stretching polyprotein $(GB1)_4$-cHiPIP-$(GB1)_4$ leads to characteristic sawtooth-like force-extension curves shown in Figure 2.3A, in which each individual sawtooth peak corresponds to the force induced unfolding of the individual domains in the polyprotein chain. The last force peak arises from the detachment of the fully unfolded polypeptide chain from either the glass substrate or the AFM tip. Stretching such a long polyprotein will lead to clear force-extension signals at a long extension of ~50-250 nm, which can be clearly distinguished from the interferences caused by short-ranged non-specific interactions which usually happen at <~50 nm. Since the stretching of a polyprotein in our experiments relies on non-specific interactions between the AFM tip and the polyprotein, the polyprotein will be picked up and stretched randomly along its contour, leading to different numbers of domains in the stretched protein chain and thus different numbers of unfolding events.

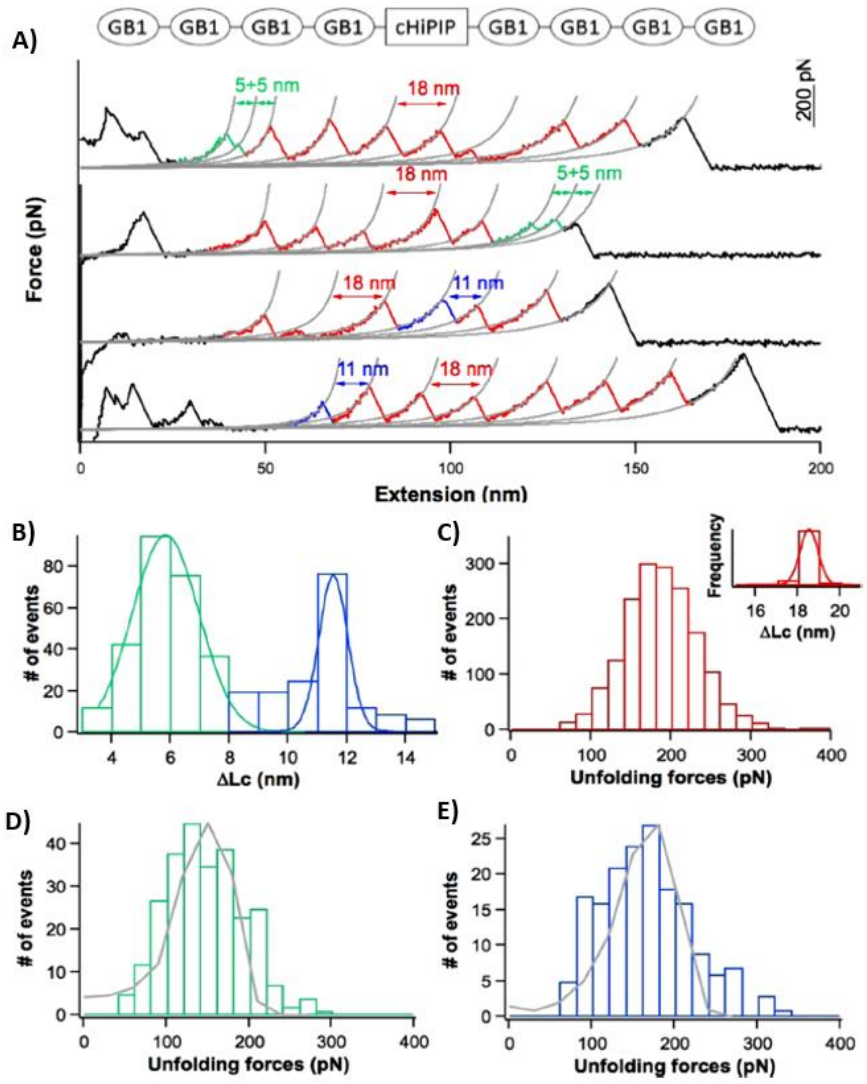

Figure 2.3 Mechanical unfolding of $(GB1)_4$-cHiPIP-$(GB1)_4$ reveals two distinct pathways for the rupture of the iron-sulfur center in cHiPIP. (A) Representative force-extension curves of $(GB1)_4$-cHiPIP-$(GB1)_4$. Force-extension curves display three distinct ΔLcs, signifying three different unfolding events. Mechanical unfolding events with ΔLc of ~18 nm (red) correspond to the unfolding of the fingerprint domain GB1. Unfolding events with a ΔLc of ~5 nm (green) and ~11 nm (blue) relate to the unfolding of cHiPIP. Gray solid lines are the wormlike chain (WLC) fits to the experimental data. A schematic of the polyprotein $(GB1)_4$-cHiPIP-$(GB1)_4$ is shown at the top panel. (B) ΔLc histogram of cHiPIP displays two distinct populations, which are centered around 5 and 11 nm, respectively. Gaussian fits (solid lines) to the experimental data show an average ΔLc of 5.3 ± 1.6 nm (avg ± standard deviation, n = 263) and 10.8 ± 0.7 nm (n = 170) for the two populations. (C) Unfolding force histogram of GB1 fingerprint domains in $(GB1)_4$-cHiPIP-$(GB1)_4$ at a pulling speed of 400 nm/s. The average unfolding force is ~188 ± 48 pN (n = 1736). Inset is the ΔLc histogram of GB1 domain with an average of 18.0 ± 1.5 nm (n = 1736). (D, E) Rupture force histograms for the two-step and one-step rupture pathways of cHiPIP. The average unfolding force for the iron-sulfur center of cHIPIP is ~147 ± 47 pN (n = 263) in the two-step pathway, and 165 ± 55 pN (n = 170) in the one-step pathway, respectively. Gray solid lines are Monte Carlo simulation results using a Δx_u of 0.14 nm and an α_0 of 1.2 s^{-1} for the two-step rupture and a Δx_u of 0.15 nm and an α_0 of 0.38 s^{-1} for the one-step rupture of cHiPIP.

Fitting the unfolding events to the worm-like chain (WLC) model of polymer elasticity revealed the contour length change of each individual unfolding event.[137] Unfolding events colored in red display a ΔLc of 18 nm and an average unfolding force of ~180 pN, which are characteristic of the unfolding of GB1 domains, and thus can be readily assigned as the unfolding of the fingerprint GB1 domains (Figure 2.3C). Since there are four GB1 domains at both ends of cHiPIP in the polyprotein $(GB1)_4$-cHiPIP-$(GB1)_4$, force-extension curves containing five or more GB1 unfolding events must contain the mechanical unfolding event of cHiPIP. Thus, the unfolding events colored in green and blue can be assigned as the unfolding of the reduced holo-cHiPIP domain (Figure 2.3A).

The unfolding of cHiPIP shows two distinct populations (Figure 2.3A-B). In one population, the unfolding of cHiPIP resulted in an unfolding event of ΔLc of ~10.8 ± 0.7 nm, while in the second population, the unfolding of cHiPIP occurred in a two-step fashion, with ΔLc_1 and ΔLc_2 of ~5.3 ± 1.6 nm and the sum of ΔLc_1 and ΔLc_2 being ~11 nm. It is of note that 79% of unfolding events with ~5 nm ΔLc (116 out of 147) appear in pairs in the force-extension curves, while those with ~11 nm ΔLc always appear individually. These results strongly indicate that the mechanical unfolding of cHiPIP follows two distinct pathways, one with a single unfolding step with ΔLc of ~11 nm and the other one with two sequential unfolding steps with similar ΔLc of ~5 nm.

cHiPIP contains 83 residues and the complete unfolding of cHiPIP would result in a ΔLc of ~26 nm (83 aa × 0.36 nm/aa - 2 nm, where 0.36 nm/aa is the length of an amino acid residue and 2 nm is the distance between the N- and C-termini of cHiPIP). However, we did not observe ΔLc of ~26 nm for cHiPIP in our experiments. Instead, we observed ΔLc of ~11 nm, which is much smaller than the complete unfolding of cHiPIP. It is likely that part of cHiPIP has unfolded

at lower forces that are below our detection limit in our AFM experiments (~20 pN). Indeed, the structure of the cHiPIP outside of the metal center (residues 1-43 and 75-83) consists of mechanically labile helical structures and non-structured sequences. Upon stretching, the unfolding of such sequences may occur at lower forces that are below the detection limit of our AFM experiments (~20 pN).

Our previous studies on rubredoxin and ferredoxin showed that the iron-thiolate bonds rupture at ~180 pN and thus the protein sequence enclosed by the iron-sulfur center will not experience the stretching force until the iron-sulfur center has been ruptured.[210-211] Iron-sulfur bonds in cHiPIP are similar to those in rubredoxin and ferredoxin, and should be mechanically stable.[223-225] Thus, they should be able to sequester the sequence enclosed by the iron-sulfur center (residues 43-75) and shield it from the stretching force until the iron-sulfur center has ruptured. Rupture of the iron-sulfur center will lead to the stretching of the sequestered sequence and result in a ΔLc of ~11 nm (32 aa × 0.36 nm/aa - 0.9 nm, where 0.9 nm is the distance between residue 43 and 75), in good agreement with our experimental observation. This result suggests that the mechanical unfolding of cHiPIP proceeds in two steps: the first step involves the unfolding of the mechanically labile sequence outside of the iron-sulfur center, which occurs at forces below 20 pN; the second step corresponds to the mechanical rupture of the iron-sulfur center and subsequent unfolding and extension of the protein structure sequestered by the iron-sulfur center.

2.3.3 Mechanical Rupture of the [4Fe-4S] Center Occurs in Two Parallel Pathways

Having established that the unfolding force events observed for cHiPIP correspond to the mechanical rupture of the iron-sulfur center, we now examine the rupture mechanism of the metal center. The [4Fe-4S] metal center is incorporated into cHiPIP via four cysteine residues

(Cys43, 46, 61 and 75) by forming four iron-thiolate bonds. Our experiments revealed that two parallel mechanisms exist for the mechanical rupture of the iron-sulfur center, ~51% of cHiPIPs followed a one-step mechanism (153 out of 300) with a ΔLc of ~11 nm, while 49% of cHiPIPs ruptured via a two-step mechanism (147 out of 300) with ΔLc_1 and ΔLc_2 of ~5 nm.

The one-step rupture of the iron-sulfur center, which results in ΔLc of ~10.8 nm, must involve the rupture of multiple iron-thiolate bonds which includes at least both Fe-S(Cys61) and Fe-S(Cys75) bonds. Different scenarios involving the rupture of different iron-thiolate bonds are shown in Table 2.1. The one-step rupture occurs as a sharp single-step event, suggesting that the rupture of these multiple iron-thiolate bonds occurs in a concurrent fashion, despite the large separation of the multiple iron-thiolate bonds by long polypeptide sequences (Figure 2.4). This observation is similar to what we observed for both rubredoxin and ferredoxin, where two ferric-thiolate bonds were observed to be ruptured concurrently.[210-211]

Table 2.1 Scenarios with Different ΔLcs Resulting from Rupturing Different Iron-Thiolate Bonds and Releasing Polypeptide Chain Segments Sequestered by Iron-Thiolate Bonds in cHiPIP.

	bond(s) ruptured	ΔLc (nm)
one-step pathway	Fe-S(Cys75) & Fe-S(Cys61)	9.6
	Fe-S(Cys75), Fe-S(Cys61) & Fe-S(Cys46)	10.9
	Fe-S(Cys43), Fe-S(Cys46) & Fe-S(Cys61)	10.9

	bond(s) ruptured	ΔLc_1 (nm)	bond(s) ruptured	ΔLc_2 (nm)
two-step pathway	Fe-S(Cys75)	4.0	Fe-S(Cys61)	4.6
	Fe-S(Cys75)	4.0	Fe-S(Cys61) & Fe-S(Cys46)	5.8
	Fe-S(Cys43) & Fe-S(Cys46)	5.8	Fe-S(Cys61)	4.0

(The ΔLcs are calculated by subtracting the original distance between cysteines from the length of the fully extended polypeptide chain segment.)

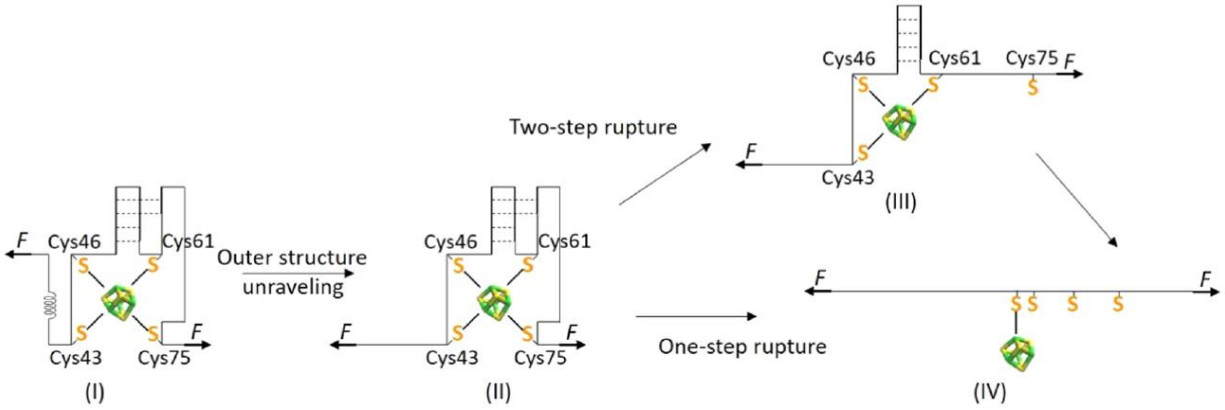

Figure 2.4 Schematic of the mechanical unfolding pathway of cHiPIP. Upon stretching, the protein sequence outside of the [4Fe-4S] coordination center (residues 1-43 and 75-83), which consists of helical and unstructured sequences, is unraveled first. The sequence sequestered by the iron-sulfur center serves as an unfolding intermediate state, which is shielded from the stretching force by the mechanically stable iron-thiolate bonds. Further stretching will cause the rupture of the iron-sulfur center and the complete unfolding of cHiPIP. The rupture of the metal center follows two parallel pathways: one-step and two-step unraveling. The one-step unraveling involves the concurrent rupture of multiple iron-thiolate bonds, whereas the two-step pathway involves the sequential rupture of the iron-thiolate bonds and subsequent complete unfolding of cHiPIP.

The two-step rupture of the iron-sulfur center yielded ΔLc_1 and ΔLc_2 of a similar amplitude of ~5.3 nm, suggesting that both steps involve the rupture of iron-thiolate bond(s) (Figure 2.4). Table 2.1 shows the different possible scenarios of the sequential rupture of the iron-thiolate bonds, which results in ΔLc ranging from 4 nm to 5.8 nm. Due to the limited resolution of our ΔLc measurements as well as the very close values of ΔLcs from different rupture scenarios, the ΔLc histogram we measured in Figure 2.3B is likely a convolution of the ΔLcs from the different rupture scenarios shown in Table 2.1.

2.3.4 Mechanical Rupture of the [4Fe-4S] Center Occurs at ~150 pN

The elucidation of the mechanical unfolding mechanism of cHiPIP allowed us to unambiguously determine the mechanical stability of the iron-thiolate bonds in cHiPIP. The rupture force for the iron-thiolate bonds in the sequential rupture scenario is around 147 pN at a pulling speed of 400 nm/s, while the rupture force in the concurrent pathway is slightly higher (165 pN at a pulling speed of 400 nm/s). Moreover, the rupture forces of the iron-thiolate bonds

show a broad distribution, indicative of a small distance between the bound state and mechanical transition state (Δx_u).[91, 226] To quantify the kinetic parameters characterizing the mechanical rupture free energy profile of the iron-thiolate bonds, we carried out pulling experiments at different pulling speeds (Figure 2.5). As expected, with the increase of the pulling speed, the mechanical unfolding/rupture force increases. Using well-established Monte Carlo simulation procedures[226-228], we reproduced the sawtooth-like force-extension results of $(GB1)_4$-cHiPIP-$(GB1)_4$. We found that the average unfolding forces and their pulling speed dependence for the concurrent mechanical rupture of the [4Fe-4S] center can be well reproduced using a Δx_u of 0.15 nm and a spontaneous off rate at zero force (α_0) of 0.38 s^{-1}, while the sequential rupture can be described using a Δx_u of 0.14 nm and an α_0 of 1.2 s^{-1} (Figure 2.3D-E and 2.5). These kinetic parameters are comparable to those for rubredoxin (Δx_u = 0.11 nm, α_0 = 0.15 s^{-1}) and ferredoxin (Δx_u = 0.13 nm, α_0 = 0.07 s^{-1})[210-211], indicating similar intrinsic mechanical stabilities and unfolding properties between these iron-sulfur proteins. The short Δx_us suggest a stiff structure of cHiPIP and may indicate transition states with a structure similar to the native cHiPIP.

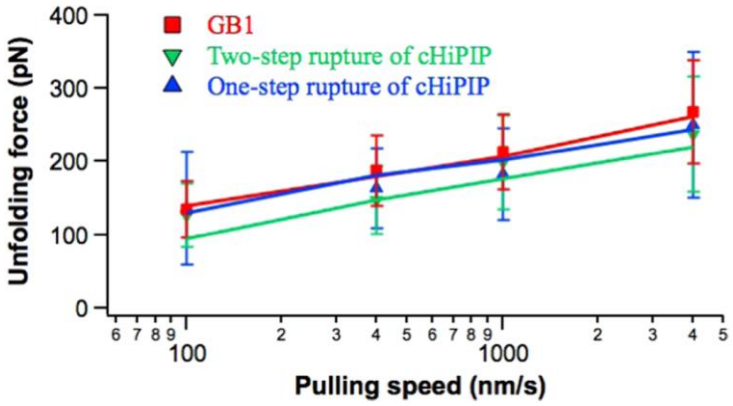

Figure 2.5 Pulling speed dependence of the unfolding/rupture forces for cHiPIP. (red: GB1, green: the two-step rupture of cHiPIP, blue: the one-step rupture of cHiPIP.) Solid lines are Monte Carlo simulation results using a Δx_u of 0.17 nm and an α_0 of 0.039 s^{-1} for GB1, a Δx_u of 0.14 nm and an α_0 of 1.2 s^{-1} for the two-step rupture of cHiPIP, and a Δx_u of 0.15 nm and an α_0 of 0.38 s^{-1} for the one-step rupture of cHiPIP.

2.3.5 Histidine Mutants of cHiPIP Corroborate the Mechanical Rupture Mechanism of the Iron-Sulfur Center

To further validate the proposed mechanical unfolding mechanism of cHiPIP, we selectively substituted cysteine residue(s) of cHiPIP in the metal center with histidine to engineer histidine mutants for single-molecule AFM experiments. Since the interaction between iron and nitrogen from histidine is much weaker than the interaction between iron and sulfur from cysteine, we anticipate that substituting the chelating cysteine to histidine will weaken the metal center and thus may alter the unfolding behavior of cHiPIP. For this purpose, we engineered the following histidine mutants: C61,75H, C75H and C61H. Compared with wt cHiPIP, the UV-Vis spectra of these cHiPIP histidine mutants showed weaker absorbance at ~380 nm, indicative of the changes brought by the histidine mutation to the iron-sulfur center of cHiPIP (Figure 2.6).

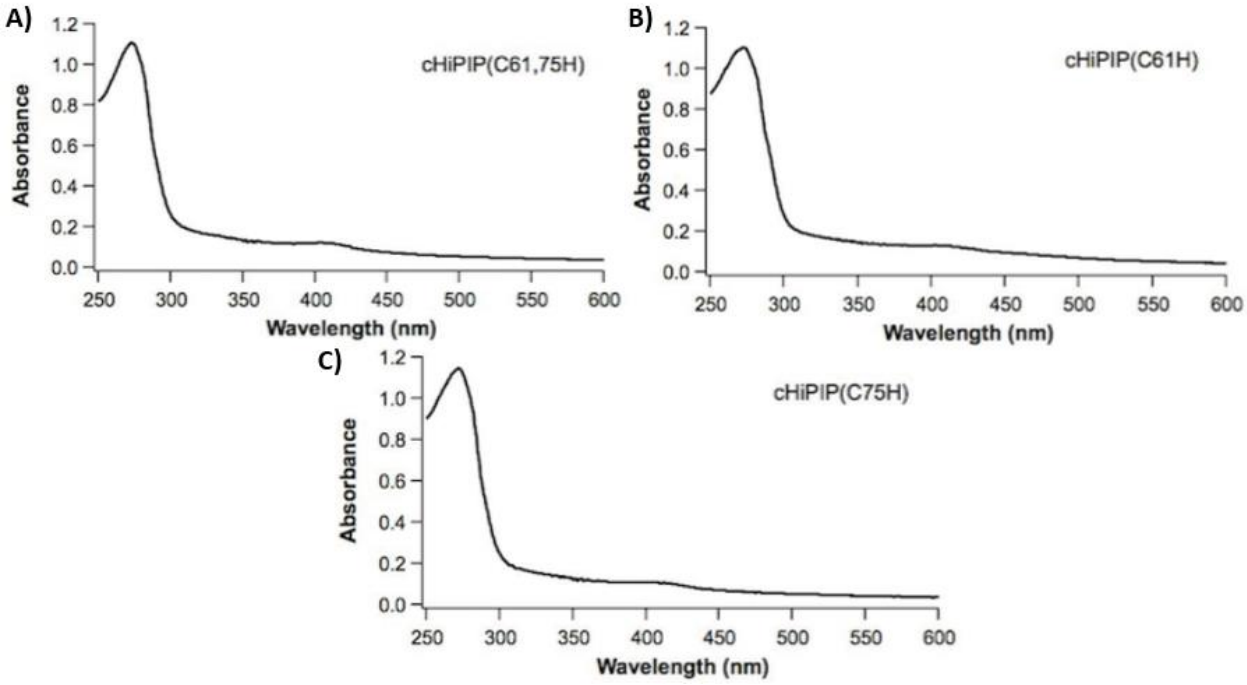

Figure 2.6 UV-Vis spectra of histidine variants of cHiPIP. The shift of absorbance indicates the influence of the histidine mutation on the ligand-to-metal charge-transfer bands.

The one-step rupture of the iron-sulfur center is due to concurrently rupturing several iron-thiolate bonds, including at least both Fe-S(Cys 61) and Fe-S(Cys75). We reasoned that if both Cys61 and Cys75 are mutated to histidines, the iron-sulfur center will be ruptured more readily at lower forces due to the lower stability of the Fe-N bonds. Indeed, in our single-molecule AFM experiments on $(GB1)_4$-cHiPIP(C61,75H)-$(GB1)_4$, about 40% force-extension curves that contain at least five GB1 unfolding events do not display any unfolding signature of cHiPIP, implying that the metal center in C61,75H mutant ruptured at forces lower than our AFM detection limit (which is ~20 pN) (Figure 2.7). This is in sharp contrast with the rupture of the metal center in wt cHiPIP, where only 7% force-extension curves do not show the unfolding signature of cHiPIP. Among the 60% curves (224 out of 365) that do show the mechanical unfolding signature of cHiPIP, both concurrent and sequential rupture pathways were observed and their occurrence is at a 1:1 ratio.

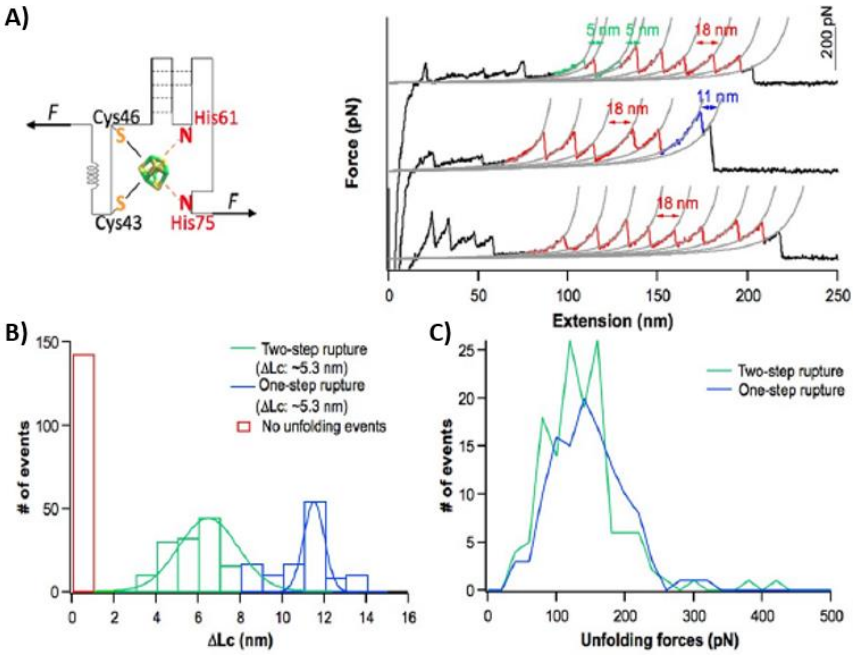

Figure 2.7 Unfolding behaviors of cHiPIP(C61,75H). (A) A schematic of cHiPIP(C61,75H) and representative force-extension curves of $(GB1)_4$-cHiPIP(C61,75H)-$(GB1)_4$. (B, C) Unfolding ΔLc and force histograms of cHiPIP(C61,75H). The ratio between one- and two-step rupture probabilities stays around 1:1 (112-112), but up to

39% of the polyprotein molecules we stretched show no detectable cHiPIP(C61,75H) unfolding events compared with only 7% for cHiPIP. The red bar in (B) indicates the number of molecules whose unfolding occurred at low forces (<20 pN) that are below the detection limit of our AFM.

In the two-step rupture mechanism, sequential rupture of iron-thiolate bond(s) leads to sequentially releasing two polypeptide chain segments separated by the iron-thiolate bond Fe-S(Cys61) (Table 2.1). Regardless of whether the rupture initiates from the Cys43/Cys46 end or Cys75 end, the Fe-S(Cys61) bond will remain intact after the first step in the two-step rupture pathway and serve as a stable intermediate state. If Cys61 is mutated to histidine, a much weaker Fe-N(His61) bond will likely be ruptured much more easily and cannot serve as a stable intermediate. In that case, it is likely that the two-step rupture mechanism may exhibit itself in a fashion similar to the one-step concurrent rupture mechanism. To test this hypothesis, we constructed C61H mutant and carried out single-molecule AFM experiments on $(GB1)_4$-cHiPIP(C61H)-$(GB1)_4$. Indeed, in the mechanical unfolding of the C61H mutant, the occurrence of the two-step rupture pathway decreased significantly (Figure 2.8). Only about 18% of C61H mutant (39 out of 218) was observed to rupture in the two-step pathway, as compared with ~50% occurrence for the wt cHiPIP.

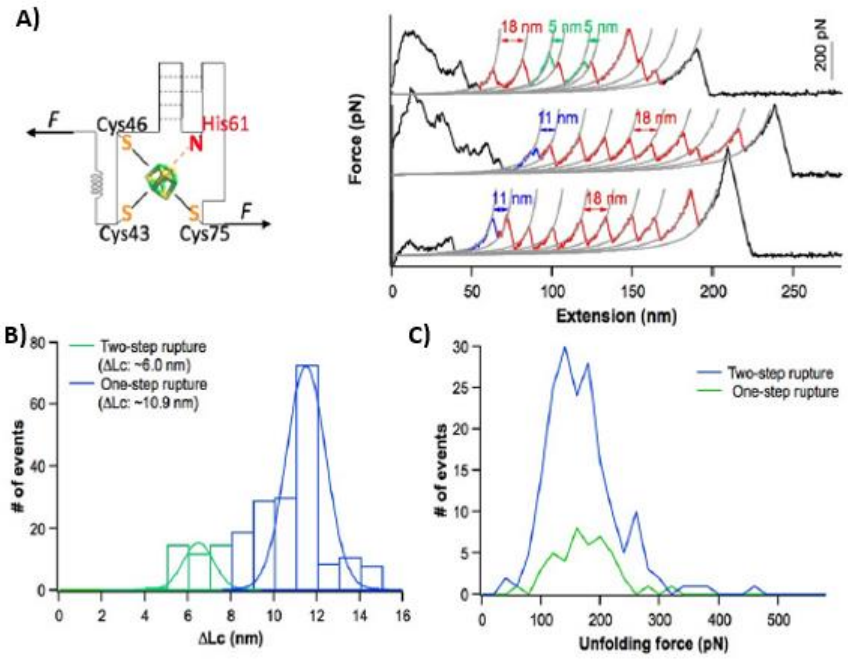

Figure 2.8 Unfolding behaviors of cHiPIP(C61H). (A) A schematic of cHiPIP(C61H) and representative force-extension curves of $(GB1)_4$-cHiPIP(C61H)-$(GB1)_4$. (B, C) Unfolding ΔLc and force histograms of cHiPIP(C61H). The percentage of two-step rupture events drops from 49 to 18% (39 out of 218).

To probe the mechanical rupture mechanism of cHiPIP further, we also constructed C75H mutant. Cys75 plays important roles in both one-step and two-step rupture mechanisms of cHiPIP. Weakening the chelation site by mutating Cys75 to histidine may affect both pathways. If Cys75 is mutated to histidine, the resultant Fe-N(His75) bond would be much weaker and may be ruptured at much lower forces, leading to interesting consequences to both unfolding pathways. On the one hand, the easier rupture of Fe-N(His75) may make the concurrent one-step rupture mechanism less likely. On the other hand, in the two-step rupture pathway, the rupture of the Fe-N(His75) bond may not appear as a clear rupture step, and thus the two rupture steps with ΔLc of ~5 nm would no longer occur in pairs. To test these possible consequences, we constructed $(GB1)_4$-cHiPIP(C75H)-$(GB1)_4$. As shown in Figure 2.9, the occurrence of the one-step concurrent rupture pathway decreased significantly from 51% for wt cHiPIP to about 19%

(55 out of 191) for C75H. More interestingly, almost all the two-step rupture events with ΔLc of ~5 nm (134 out of 136) occur by themselves in the force-extension curves (Figure 2.9A), in sharp contrast to the pairwise appearance of the unfolding events of ΔLc of ~5 nm in wild-type cHiPIP. Taken together, these results on histidine mutants largely corroborate the proposed unfolding/rupture mechanism of cHiPIP, and reveal interesting ways to fine tune the unfolding pathways of cHiPIP.

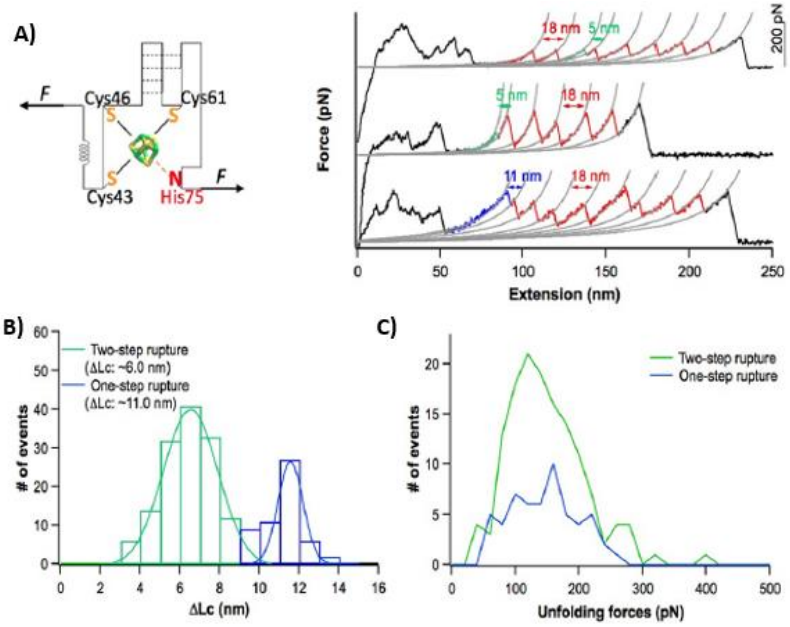

Figure 2.9 Unfolding behaviors of cHiPIP(C75H). (A) A schematic of cHiPIP(C75H) and representative force-extension curves of $(GB1)_4$-cHiPIP(C75H)-$(GB1)_4$. (B, C) Unfolding ΔLc and force histograms of cHiPIP(C75H). The percentage of one-step rupture events drops from 51 to 19% (55 out of 191), and 99% of two-step rupture molecules (134 out of 136) show only one two-step rupture event.

It is worth noting that the unfolding forces of these histidine mutants do not show clear trends in comparison with that of wild-type cHiPIP (Figure 2.7C, 2.8C, 2.9C). Since mechanical rupture is a stochastic process, the measured rupture forces for metal-ligand bonds (Fe-N and Fe-S) may simply reflect the high force end of a broad distribution for the rupture of a weak metal-ligand bond, while the low force end of the distribution is not properly sampled due to the

detection limit of the AFM (~20 pN). In addition, the protein environment may play an important role in modulating the rupture force of Fe-N bonds.

2.3.6 Mechanical Unfolding of Oxidized HiPIP

While most of the [4Fe-4S] ferredoxins show a lower redox potential range (from -250 to -650 mV) and a $[4Fe\text{-}4S]^{2+/1+}$ redox couple, the cluster in HiPIPs exchanges electrons at a high redox potential (+50 to +450 mV) using the $[4Fe\text{-}4S]^{3+/2+}$ redox couple.[229] HiPIPs serve as electron carriers in bacterial photosynthesis. It donates electrons to the photo-oxidized reaction center, operating via the membrane-bound *c*-type tetraheme cytochrome subunit.[217] In cHiPIP, the cluster exists as $[2Fe^{3+}, 2Fe^{2+}]$ ($[4Fe\text{-}4S]^{2+}$) in the reduced state and $[3Fe^{3+}, Fe^{2+}]$ ($[4Fe\text{-}4S]^{3+}$) in the oxidized state. The structure of cHiPIP remains almost the same in the two oxidation states,[219, 221] thus cHiPIP can serve as a good model system to study how the oxidation state of the coordinated cluster can affect metalloprotein's mechanical stability, if there is any. Such experiments do not require an anaerobic condition and can be readily carried out under ambient conditions, due to the high reduction potential of cHiPIP and the ease to oxidize cHiPIP *in vitro* by using chemical oxidizing agents, such as $K_3Fe(CN)_6$.[230] Following well-established methods, we treated cHiPIP with oxidizing agent $K_3Fe(CN)_6$ to obtain the oxidized form of cHiPIP.[230] The UV-Vis spectrum and the A_{280}/A_{380} ratio of the oxidized cHiPIP indicated a complete oxidation (Figure 2.10).[221]

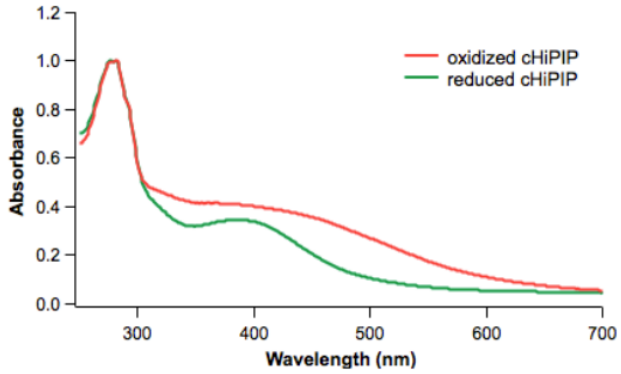

Figure 2.10 UV-Vis spectrum of oxidized cHiPIP. For comparison, the UV-Vis spectrum of reduced cHiPIP is also shown.

Single-molecule AFM experiments on the polyprotein $(GB1)_4$-(oxidized cHiPIP)-$(GB1)_4$ revealed that the oxidized cHiPIP showed the same ΔLc distribution as the reduced cHiPIP, with one population at ~5 nm and the other at ~11 nm (Figure 2.11A). Among the unfolding events of cHiPIP, the iron-thiolate bonds of about 61% (203 out of 332) were ruptured in the one-step manner, while those of the other 39% of them (129 out of 332) occurred following the two-step mechanism. The unfolding forces for the oxidized cHiPIP in the two-step and one-step pathway are similar (~180 pN), slightly higher than that of reduced cHiPIP (147 pN in the one-step pathway and 165 pN in the two-step pathway) (Figure 2.11B-C). Using Monte Carlo simulations to fit the unfolding force distribution as well as the pulling speed dependence of the unfolding forces of the oxidized cHiPIP revealed kinetic parameters that are similar to that of reduced cHiPIP (Δx_u of 0.16 nm and α_0 of 0.32 s^{-1} for two-step rupture and Δx_u of 0.15 nm and α_0 of 0.39 s^{-1} for the one-step rupture) (Figure 2.11B-D). These results suggest that the rupture forces of the iron-sulfur center in the oxidized cHiPIP are slightly higher than those of the reduced cHiPIP. This finding is in contrast with our previous studies on rubredoxin, which showed that the rupture force of the ferric-thiolate bond is significantly higher than that of ferrous-thiolate

bonds.[210] It is interesting to note that the rupture forces measured for cHiPIP (both oxidized and reduced) are lower than the rupture forces of the ferric-thiolate bonds in rubredoxin.[210]

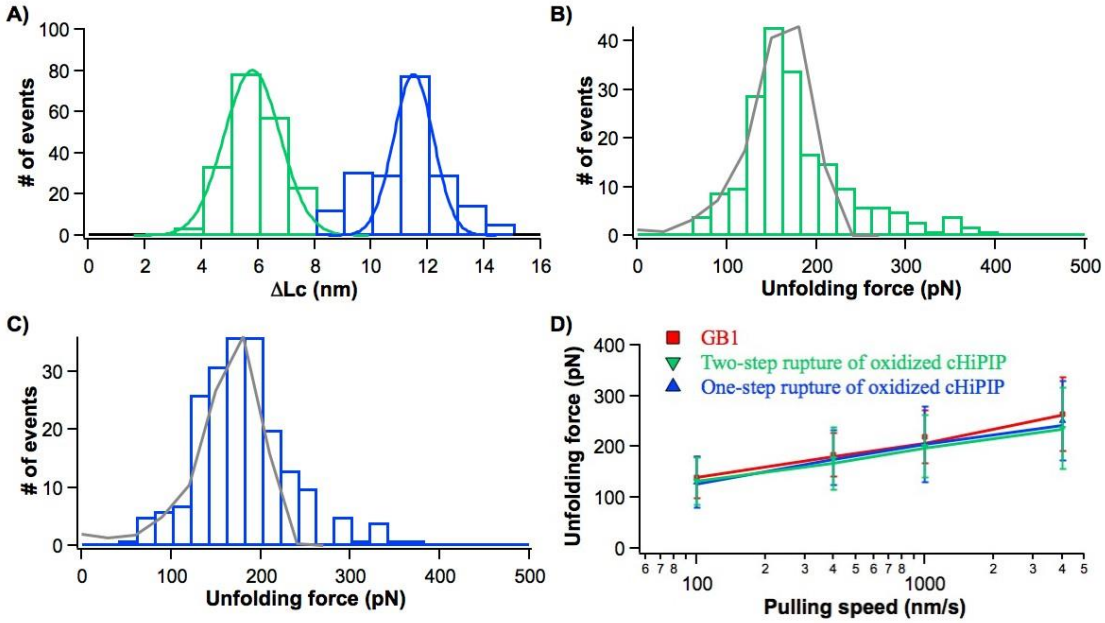

Figure 2.11 Mechanical unfolding of $(GB1)_4$-(oxidized cHiPIP)-$(GB1)_4$ reveals two unfolding pathways of oxidized cHiPIP similar to cHiPIP. (A) The histogram of ΔLc from oxidized cHiPIP. Gaussian fits (solid lines) to the experimental data show average ΔLcs of 5.3 ± 1.4 nm (n = 199) and 11.0 ± 1.0 nm (n = 203). (B, C) Unfolding force histograms of $(GB1)_4$-(oxidized cHiPIP)-$(GB1)_4$ at a pulling speed of 400 nm/s. Two-step rupture of the iron-sulfur center in oxidized cHiPIP occurs at ~177 ± 62 pN (n = 199) (B), and one-step rupture occurs at ~179 ± 54 pN (n =203) (C). Gray solid lines are Monte Carlo simulation results using a Δx_u of 0.16 nm and an α_0 of 0.32 s^{-1} for the two-step rupture of the iron-sulfur center in oxidized cHiPIP and a Δx_u of 0.15 nm and an α_0 of 0.39 s^{-1} for the one-step rupture. (D) Pulling speed dependence of the unfolding/rupture forces for GB1 (red), the two-step rupture (green), and the one-step rupture (blue) of oxidized cHiPIP. Solid lines are Monte Carlo simulation results using a Δx_u of 0.16 nm and an α_0 of 0.32 s^{-1} for the two-step rupture of the iron-sulfur center in oxidized cHiPIP, a Δx_u of 0.15 nm and an α_0 of 0.39 s^{-1} for the one-step rupture of the iron-sulfur center in oxidized cHiPIP, and a Δx_u of 0.17 nm and an α_0 of 0.039 s^{-1} for GB1.

The possible reason for this small difference in mechanical rupture force between the oxidized and reduced forms of cHiPIP may lie in the mixed-valence nature of the [4Fe-4S] cluster. The reduced $[4Fe\text{-}4S]^{2+}$ and the oxidized $[4Fe\text{-}4S]^{3+}$ are both mixed-valence clusters ($[2Fe^{3+}, 2Fe^{2+}]$ in the reduced state and $[3Fe^{3+}, Fe^{2+}]$ in the oxidized state) and the electrons are highly delocalized in the iron-sulfur center, making the four iron ions indistinguishable by spectroscopic methods.[231] It is thus likely that the measured rupture force for cHiPIP is an

average of the rupture force for both ferric- and ferrous-thiolate bonds in cHiPIP. The one positive charge added to the cluster by oxidation will not make significant changes to the bond strength. Indeed, crystallographic structures of reduced and oxidized cHiPIP also show that oxidation only causes less than 1.7% change to the bond lengths of the four Fe-S(Cys) bonds.[221] Thus, as a highly conjugated coordination system, the $[Fe_4S_4(SR)_4]$ coordination system in cHiPIP has a good capacity for one electron. This may contribute to the ability of HiPIP to transfer electrons in bacterial photosynthesis.

2.4 Discussion

2.4.1 Two-Step Mechanical Unfolding Mechanism May Be a General Feature Among Iron-Sulfur Proteins

By combining single-molecule AFM and protein engineering techniques, we have investigated the mechanical unfolding of a small [4Fe-4S] iron-sulfur protein cHiPIP in detail. Our results clearly demonstrated that the mechanical unfolding of cHiPIP is coupled with the mechanical rupture of the iron-sulfur center. Upon stretching from its N- and C-termini, the structure and sequence outside of the iron-sulfur center in cHiPIP unravels first, followed by the rupture of the iron-sulfur center and unfolding of the structure enclosed by the metal center. The mechanical stability of the iron-thiolate bonds allows the iron-sulfur center to sequester the sequence enclosed by the iron-sulfur center and prevent them from being subject to the stretching force until the iron-sulfur center ruptures. Thus, the overall mechanical unfolding process of cHiPIP can be considered as a two-step process, where the iron-sulfur center and its enclosed sequence serve as a stable on-pathway unfolding intermediate state. This unfolding mechanism was further corroborated by results from the cysteine to histidine mutants studies.

This two-step unfolding process is not unique to cHiPIP. Similar unfolding mechanisms were also observed for the unfolding of the two simpler iron-sulfur proteins we studied previously ([1Fe-0S] rubredoxin, [2Fe-2S] ferredoxin).[210-211] These results suggest that this two-step unfolding process may be a more general unfolding mechanism for iron-sulfur proteins, where the iron-sulfur center will serve as a stable unfolding intermediate state after the sequence outside the metal center has unfolded. Whether the unfolding of the sequence outside of the metal center for a given protein produces a measurable unfolding signature in the force spectroscopy experiments depends on the specific tertiary structure of the given iron-sulfur protein. For example, for both rubredoxin and cHiPIP, the structures of the protein sequence outside of the iron-sulfur center are mechanically labile and thus their unfolding occurs at low forces that are below our detection limit.[210] In contrast, the [2Fe-2S] ferredoxin from *Anabaena* assumes a mechanically stable fold outside the iron-sulfur center. Thus, its unfolding results in a clear mechanical unfolding signature prior to the mechanical rupture of the [2Fe-2S] center.[211] These experimental evidences clearly indicated that, for metalloproteins like iron-sulfur proteins, both metal coordination and protein secondary/tertiary structures contribute to protein stability and can be probed by single-molecule force spectroscopy method, thus highlighting the important structural roles of such iron-sulfur centers in the folding and function of metalloproteins. Moreover, this two-step unfolding mechanism may also be shared among other metalloproteins as long as the metal-ligand bonds are mechanically stable. For example, recent studies on cupredoxin and zinc finger protein showed similar two-step unfolding mechanism.[203, 206]

2.4.2 Concurrent Rupture of Multiple Iron-Thiolate Bonds Is a Common Feature for the Rupture of the Iron-Sulfur Center in Metalloproteins

Of the three iron-sulfur proteins we have studied so far (rubredoxin, ferredoxin and HiPIP), four cysteine residues form iron-thiolate bonds to constitute the specific iron-sulfur center in each of the iron-sulfur proteins. In all three cases, our results indicate that the mechanical rupture of the iron-sulfur center involves the concurrent rupture of two (or even three) iron-thiolate bonds, despite that the three iron-sulfur proteins showed very different structural patterns of the four iron-chelating cysteine residues, ranging from $CxxCx_nCxxC$ motif for rubredoxin, to Cx_4CxxCx_nC in ferredoxin and $CxxCx_{15-17}Cx_{14-15}C$ in cHiPIP.[216] It is of special note that the concurrent rupture of multiple iron-thiolate bonds occurs for the iron-thiolate bonds that are separated by long polypeptide sequences. For example, in cHiPIP, the two iron-thiolate bonds that were concurrently ruptured are Fe-S(Cys61) and F-S(Cys75), which are separated by 14 residues. This finding is in excellent agreement with our results on a loop insertion mutant of rubredoxin, where we inserted 19 residues in the middle of the iron chelation loop C38xxC41 and clearly observed the concurrent rupture of two ferric-thiolate bonds Fe-S(Cys38) and Fe-S(Cys41).[214] Collectively, these results highlight the structural plasticity of the iron chelation site formed by the four cysteine residues, which has been known for iron-sulfur protein. More importantly, although these four iron-thiolate bonds can be distant in space, they form part of the same iron-sulfur center. Upon stretching, they can act in a concerned (or cooperative) fashion to rupture concurrently on a time scale shorter than 100 μs. In other words, it may be more appropriate to view these iron-thiolate bonds as a concerned unit of the iron-sulfur center, rather than as individual uncorrelated iron-thiolate bonds. However, the origin of

this behavior remains unknown and calls for more systematic investigation, both experimentally and theoretically.

2.4.3 Forced Unfolding Mechanism of HiPIP Revealed by the AFM May Bear Potential Biological Relevance

It has been demonstrated that HiPIP from *Allochromatium vinosum* is translocated into the periplasmic space by the translocation system *in vivo*,[232] a process that involves a forced-unfolding of the target protein. Thus, the force spectroscopy method may provide a more biologically relevant tool to understand the forced-unfolding process of HiPIP during the translocation process *in vivo*. The forced unfolding pathway of a protein is dependent upon the pulling direction defined by the two residues along which the stretching force is applied.[233-234] In our AFM experiments, cHiPIP is stretched and unraveled from its N- and C-termini. Although this unfolding direction may be different from that occurring in the translocation process for HiPIP, our AFM results can nonetheless help better understand the forced-unfolding mechanism of HiPIP during its translocation process *in vivo*.

2.5 Conclusions

In this study, we have combined single-molecule AFM and protein engineering techniques to investigate the mechanical unfolding behavior of a [4Fe-4S] high potential iron-sulfur protein. Our results revealed that cHiPIP unfolds in a two-step manner, in which the sequence outside of the coordination center ruptures first, and the sequence sequestered by the iron-sulfur center serves as a stable unfolding intermediate state. The rupture of the iron-sulfur center proceeds in two distinct parallel pathways, involving concurrent and sequential ruptures of multiple iron-thiolate bonds. The rupture mechanism of the iron-sulfur center in cHiPIP is further corroborated by the mutational studies by substituting cysteine with histidine. Our results suggest

that the unfolding mechanism of cHiPIP and the rupture mechanism for the iron-sulfur center may be general among iron-sulfur proteins, thus providing a new tool to study iron-sulfur proteins from a new perspective.

2.6 Experimental Section

2.6.1 Protein Engineering

The plasmids encoding cHiPIP was a kind gift from Prof. Jiangyun Wang (Institute of Biophysics, Chinese Academy of Sciences). The gene of cHiPIP, which carries a 5' BamHI (G'GATCC), and 3' BglII (A'GATCT) and KpnI (G'GTACC) restriction sites, was amplified using polymerase chain reaction (PCR) and its sequence was confirmed by direct DNA sequencing. The gene of the polyprotein $(GB1)_4$-cHiPIP-$(GB1)_4$ was constructed following a well-established iterative digestion and ligation scheme, which is based on the identity of the sticky ends generated of the BamHI and BglII restriction sites. Histidine mutants of cHiPIP were constructed via standard site-directed mutagenesis methods using the cHiPIP gene as the template. The polyprotein gene was cloned into the expression vector pQE80L (Qiagen, Valencia, CA), which carries an N-terminal His6 tag.

The polyproteins were overexpressed in the *Escherichia coli* strain DH5α. After inoculation with 3 mL pre-culture, the cells were grown in 200 mL of LB media containing 100 μg/mL ampicillin at 37°C and 225 rpm without additional iron source added.[235] When the OD_{600} of the culture reached ~0.7, protein overexpression was induced with 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and the protein expression continued for 4 hours. Then the cells were pelleted by centrifugation at 4000 × *g* for 10 minutes at 4°C and resuspended in 10 mL of loading buffer (phosphate-buffered saline (PBS), 10 mM, pH 7.4). After adding 10 μL of protease inhibitor cocktail (Sigma-Aldrich, St. Louis, MO), 100 μL of 50 mg/mL lysozyme, 1

mL of 10% (w/v) Triton X-100, 50 µL of 1 mg/mL DNase I and RNase A to the loading buffer, cells were lysed for 40 minutes on ice. Cell debris was then removed by centrifugation at 22000 × *g* at 4°C, and the supernatant was loaded into Co^{2+} affinity chromatography column (TALON Metal Affinity Resins, Clontech, Mountain View, CA) with loading buffer. After washing the column with 50 mL of washing buffer (10 mM PBS, 300 mM NaCl, 7 mM imidazole, pH 7.4), the protein was eluted with 2 mL elution buffer (10 mM PBS, 300 mM NaCl, 250 mM imidazole, pH 7.4). The purified protein sample had a concentration of ~1 mg/mL. The oxidation of cHiPIP was achieved by adding $K_3Fe(CN)_6$ into the protein solution to a concentration of 10 mM.[230]

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed to confirm the molecular weight of the polyprotein. UV-Vis spectroscopy was used to quantitatively determine the purity of cHiPIP (Cary-100 UV-Vis Spectrometer, Agilent, Santa Clara, CA). The protein solution was desalted using a desalting column (Zeba Spin Desalting Columns, ThermoFisher Scientific, Waltham, MA) following the manufacturer's instructions before the UV-Vis measurement. The ratio between the absorbance maxima at 280 and 380 nm (A_{280}/A_{380}) was used to calculate the percentage of holo-protein.[221]

2.6.2 Single-Molecule AFM Experiment

Single-molecule AFM experiment was carried out on a custom-built AFM, which has been described previously,[236] as well as a Cypher AFM (Asylum Research, Santa Barbara, CA). Each Si_3N_4 cantilever (MLCT cantilever, Bruker, Santa Barbara, CA) was calibrated in buffer (PBS, 10 mM, pH 7.4) before each experiment using the equipartition theorem, yielding a spring constant of ~40 pN/nm. In a typical experiment, ~2 µL of the protein sample at a concentration of ~1 mg/mL was deposited onto a clean glass coverslip covered with ~100 µL buffer (PBS, 10

mM, pH 7.4). The AFM pulling experiments were carried out under a constant pulling velocity of 400 nm/s, unless otherwise noted.

Chapter 3: Single-Molecule Force Spectroscopy Reveals That Two-Coordinate Ferric Site Is Critical for the Folding of Holo-Rubredoxin

3.1 Synopsis

Metalloproteins play important roles in a wide range of biological processes. The folding process of metalloproteins is complex due to the synergistic effects of the folding of their polypeptide chains and the incorporation of metal cofactors. The folding mechanism of the simplest iron-sulfur protein rubredoxin, which contains one ferric ion coordinated by four cysteinyl sulfurs, is revealed by optical tweezers for the first time. The folding of the rubredoxin polypeptide chain is rapid and robust, while the reconstitution of the iron-sulfur center is greatly dependent upon the coordination state of the ferric ion on the unfolded polypeptide chain. If the ferric ion is coordinated by two neighboring cysteines, rubredoxin can readily fold with the iron-sulfur center fully reconstituted. However, if the ferric ion is only mono-coordinated, rubredoxin folds but the iron-sulfur center is not reconstituted. Our results suggested that the folding of holo-rubredoxin follows a novel binding-folding-reconstitution mechanism, which is distinct from the folding mechanisms proposed for the folding of metalloprotein. Our study highlights the critical importance of the two-coordinate ferric site on the folding of holo-rubredoxin, which may have some important implications to our understanding of the folding mechanism of more complex metalloproteins *in vivo*.

3.2 Introduction

Metalloproteins account for more than one-third of all proteins in nature and carry a diverse range of functions in biological processes.[19] To perform their biological functions, the folding of metalloproteins with their properly assembled metal centers is a prerequisite. The

folding of metalloproteins is complex, involving two intertwined processes: the folding of the polypeptide chain and the incorporation of the metal center. Understanding the folding, unfolding and misfolding mechanism of metalloproteins is of critical importance but very challenging. This is mainly due to the fact that, while living cells exploit dedicated machineries and chaperones in the biogenesis process to ensure the correct folding and metal-uptake of metalloproteins, the unfolding of many metalloproteins is irreversible *in vitro*, especially under near-native conditions.[153] For example, even for the smallest iron-sulfur protein rubredoxin, most insights were obtained from studies carried out in high concentrations of denaturants.[181, 187, 199]

SMFS, which has been proven to be a powerful tool for investigating the unfolding mechanisms of metalloproteins at the single-molecule level, has also demonstrated its great potential in probing the folding processes of metalloproteins, which are usually inaccessible using traditional ensemble methods.[23, 26, 32-33, 237-238] Using atomic force spectroscopy (AFM)-based SMFS techniques, we have investigated the simplest iron-sulfur protein, rubredoxin (RD) from *Pyrococcus furiosus* (*pf*RD) in great detail.[23, 26, 238] RD is a small electron transfer metalloprotein and contains 53 residues.[239] RD folds into an α/β structure, in which one ferric ion is coordinated by four cysteinyl sulfurs from two CXXC chelating motifs to form the $[Fe^{III}(SCys)_4]^-$ metal center (Figure 3.1A). Our AFM results have revealed detailed mechanistic insights into the unfolding and metal center rupture mechanism of RD, and demonstrated that mechanically unfolded RD can fold back to its native state.[23, 26, 238] However, due to the AFM's relatively limited force resolution, it has not been possible to directly monitor the folding behavior of RD in real-time. The folding mechanism of RD under the native condition remains elusive.

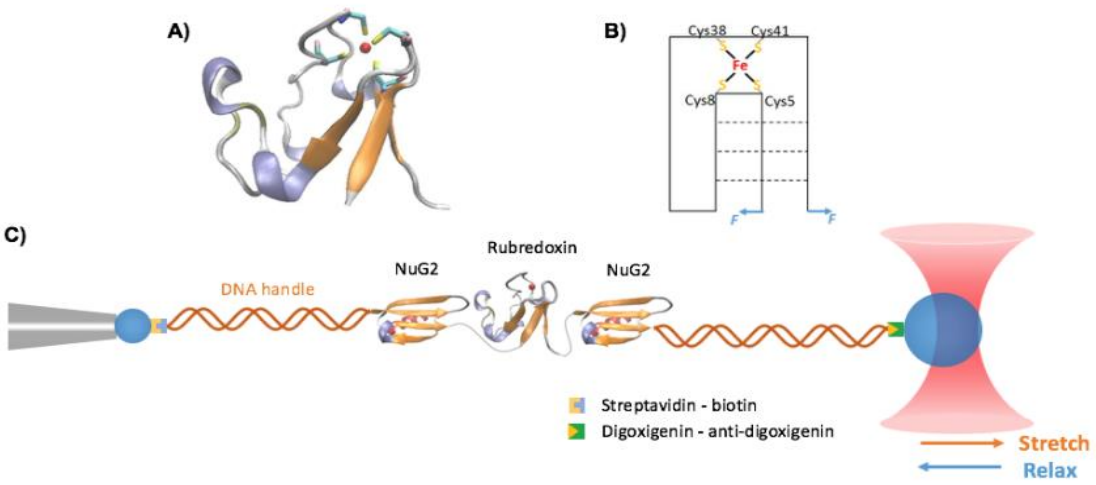

Figure 3.1 Studying the folding-unfolding mechanism of RD with OT-based SMFS. (A) Three-dimensional structure of Fe(III)-*pf*RD (PDB code: 1BRF). RD displays an α/β structure. The iron-sulfur center, in which the ferric ion is coordinated by cysteinyl sulfurs, shows a pseudo-tetrahedral geometry. The four cysteines belong to two CXXC chelating motifs. (B) Schematic of RD. The dotted lines represent backbone hydrogen bonds between β-strands and thick black lines indicate the ferric-thiolate bonds. (C) Schematic of the optical tweezers experiments to investigate the mechanical unfolding-folding of RD. Rubredoxin is flanked by two NuG2 domains and the polyprotein is covalently attached to DNA handles by the thio-maleimide reaction. The DNA-protein chimera is further linked to polystyrene beads held by a pipette tip and a laser beam via streptavidin-biotin and digoxigenin-anti-digoxigenin recognitions, respectively.

Taking advantage of the superb force resolution of optical tweezers (OT),[34, 123, 240] here we combine OT and protein engineering techniques to directly probe the folding mechanisms of both iron-containing form (the so-called holo-form) and iron-free form (apo-form) of *pf*RD under near-native condition. Our results showed that the folding of RD is initiated by the fast and robust folding of the polypeptide chain itself, followed by the reconstitution of the iron-sulfur center. Moreover, the formation of the two-coordinate ferric site by the neighboring cysteine residues is a prerequisite for the efficient reconstitution of the metal center, highlighting the critical role of the two-coordinate ferric site in the folding of holo-RD. Our results suggest a novel binding-folding-reconstitution mechanism for the folding of holo-RD, and have important implications for the folding of more complex iron-sulfur proteins *in vivo*.

3.3 Results

3.3.1 Mechanical Unfolding of Holo-RD by Optical Tweezers

To study the folding-unfolding mechanism of holo-RD with OT, we constructed a protein chimera Cys-NuG2-RD-NuG2-Cys, in which RD is flanked by the fingerprint domain NuG2 at its both termini. NuG2 is a computationally designed fast folding protein and its mechanical unfolding-folding has been well-characterized in our previous OT studies and is characterized by ΔLc of ~18 nm, unfolding force of 20-40 pN and refolding force of ~8 pN.[131, 241-242] Here NuG2 domains serve as both a fingerprint for identifying single-molecule stretching event of RD and an internal caliper for contour length increment (ΔLc). We then coupled two dsDNA handles to the protein to create the DNA-protein-DNA chimera, and used a MiniTweezer setup to measure its mechanical unfolding-folding.

Our previous AFM experiments showed that the holo-ferric RD is mechanically stable and unfolds at ~260 pN at a pulling velocity of 400 nm/s.[23] Indeed, in most force-distance curves from our constant velocity OT experiments, we only observed two NuG2 unfolding events with a ΔLc of ~17 nm prior to the DNA B-S transition at ~65 pN, and subsequent relaxation curves also only showed two NuG2 refolding events at ~8 pN (Figure 3.2A, Curve 1 and 2). Clearly, in these force-distance curves, holo-RD was not unfolded due to its high mechanical stability. Instead of trying to stretch the molecules to higher forces to unfold holo-RD, which often led to the rupture of the molecular tether mediated by noncovalent interactions, we held the molecule at a constant distance with a force of ~60 pN for an extended period of time to allow RD to unfold. After ~30 s, a sudden force drop was often observed, corresponding to the unfolding of the holo-RD in the protein construct (Figure 3.2A, Curve 3). In the subsequent relaxation curve, three refolding events were clearly observed. Two of them can be readily assigned to the refolding of

NuG2 domains, and the third one can thus be assigned to the folding of RD (Figure 3.2A, Curve 4).

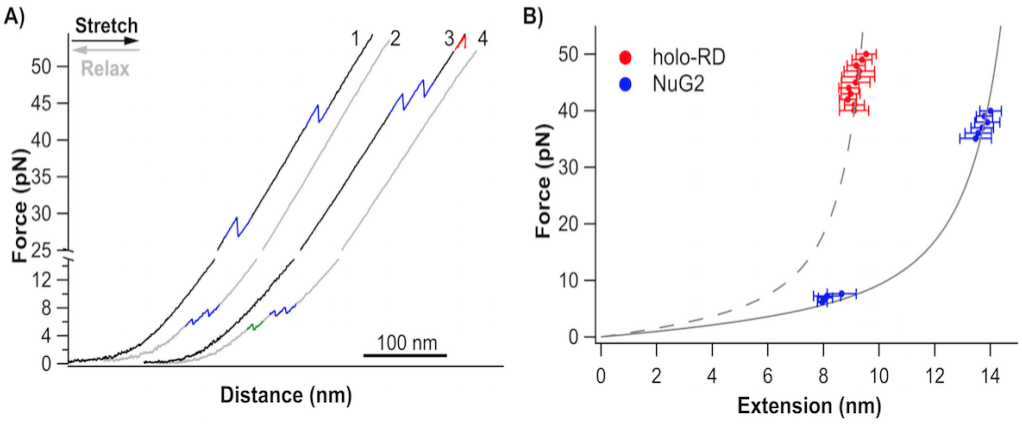

Figure 3.2 Unfolding signature of holo-RD in OT experiments. (A) Representative force-distance curves of NuG2-holo-RD-NuG2 at a pulling speed of 50 nm/s. The unfolding and folding events of the fingerprint domain NuG2 are colored in blue. The unfolding of holo-RD (colored in red) was realized by holding the molecule at ~55 pN until holo-RD unfolds. The refolding of RD is colored in green. (B) ΔLc analysis of the unfolding of the holo-RD. Each data point represents the average of the extension at a given force and the error bars indicate the standard deviation of the data. WLC fits to the experimental data revealed a persistence length of 0.8 nm and ΔLc of 11.2 ± 0.1 nm (grey dashes line) for the unfolding of holo-RD. For comparison, WLC fits to the NuG2 data revealed a ΔLc of 17.0 ± 0.1 nm (grey solid line). The data is presented as average ± standard deviation.

The force-extension relationship of the unfolding events can be obtained by measuring the increase in length versus force for the unfolding event (Figure 3.2B). Fitting the force-extension relationships of NuG2 and holo-RD unfolding events using the worm-like chain model (WLC) of polymer elasticity[137] revealed a ΔLc of 17.0 ± 0.1 nm (n = 386) for NuG2 and 11.2 ± 0.1 nm (n = 193) for RD, respectively, in good agreement with our AFM results[23, 131]. This result indicated that the unfolding events of holo-RD with a ΔLc of ~11 nm correspond to the rupture of the iron-sulfur center [Fe-$(SCys)_4$] and subsequent complete unfolding and extension of the polypeptide chain sequestered by the iron-sulfur center.[23] The unfolding of the β-hairpin outside the iron-sulfur center (residue 1-5 and 41-53) likely did not result in a clear unfolding signature. It is of note that during the rupture of the [Fe-$(SCys)_4$] center, at least two ferric thiolate bonds in one chelating motif (either Cys5,8 or Cys38,41) were ruptured.

3.3.2 Apo-RD Refolds Rapidly to Its Native Three-Dimensional Structure

Having established the unfolding signature of holo-RD in OT, we set to investigate the folding mechanism of holo-RD. Like other metalloproteins, the folding of RD involves two processes that are likely intertwined: the folding of the polypeptide chain as well as the constitution of the iron-sulfur center. To dissect the contribution of these two processes to the folding of holo-RD, we first investigated the folding process of RD polypeptide chain alone, i.e., the folding of apo-RD.

The iron-sulfur center of RD is largely buried in the folded protein structure and has low solvent accessibility. Only when RD is unfolded, the ferric ion can be chelated out by strong chelating agents such as EDTA.[243] To investigate the folding of apo-RD, we first unfolded the holo-RD in Tris buffer containing 100 mM EDTA to prepare the apo-RD. Holding the stretched NuG2-holo-RD-NuG2 molecule at ~60 pN for an extended period of time allowed us to mechanically unfold holo-RD (Figure 3.3A, Curve 1). We then held the unfolded protein at ~30 pN for ~20 s to facilitate EDTA to chelate the ferric ion from the unfolded RD chain. By doing so, we generated an apo-RD *in situ*, making it possible for us to investigate the folding and unfolding of apo-RD.

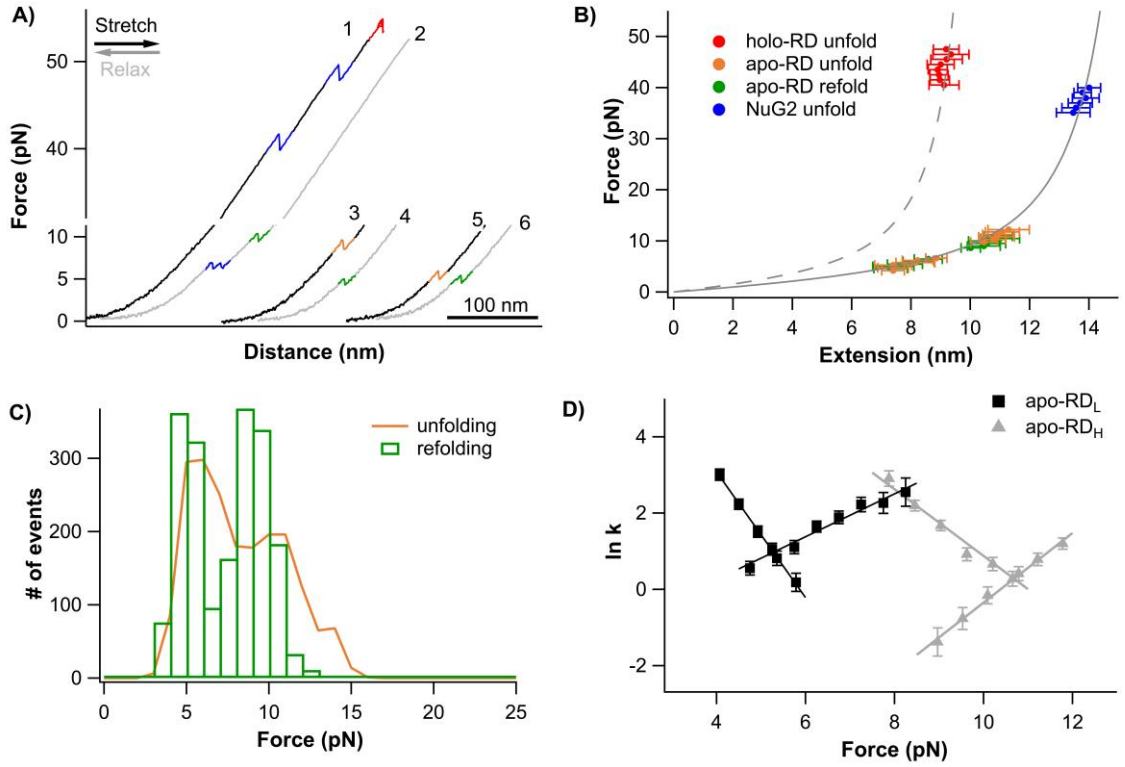

Figure 3.3 Mechanical unfolding and folding signatures of apo-RD. (A) Representative force-distance curves of NuG2-RD-NuG2 in the presence of EDTA at a pulling speed of 50 nm/s. Curve 1 was obtained by holding the molecule at ~55 pN until holo-RD unfolded. Unfolding event of holo-RD is colored in red, unfolding and refolding events of apo-RD are colored in orange and green, respectively. (B) Force-extension relationships of unfolding-refolding of apo-RD (orange and green) and the unfolding of NuG2 (blue). Grey curves are WLC fits to the experimental data. The WLC fitting measured a persistence length of 0.8 nm and ΔLc of 17.0 ± 0.1 nm for the unfolding-refolding of apoRD (solid line). For reference, the force-extension relationship of NuG2 is also shown (in blue). (C) Unfolding (orange) and refolding (green) force histograms of apo-RD at a pulling speed of 50 nm/s. The bin size is 1 pN for both unfolding and refolding histograms. For clarity, only the refolding data is displayed as a bar chart. Both histograms show a bimodal distribution (n = 1958 for each histogram). (D) Force-dependent unfolding-refolding rates for the two conformers of apo-RD (apo-RD$_L$ in black and apo-RD$_H$ in grey). Each data point represents the average lnk measured using the Oesterhelt method and the error bar indicates the standard deviation. Solid lines are Bell-Evans model fits to the experimental data. The fitting parameters for apo-RD$_L$ are: α_0 = 0.14 ± 0.04 s^{-1}, Δx_u = 2.3 ± 0.2 nm, β_0 = (1.6 ± 0.3) × 10^4 s^{-1} and Δx_f = 6.8 ± 0.2 nm; parameters for apo-RD$_H$ are: α_0 = (7.6 ± 2.7) × 10^{-5} s^{-1}, Δx_u = 3.8 ± 0.1 nm, β_0 = (1.5 ± 1.2) × 10^4 s^{-1} and Δx_f = 3.6 ± 0.3 nm. The data is presented as average ± S. D. Amongst 20 apo-RD molecules we measured, 7 adopt apo-RD$_H$ only, 2 adopt apo-RD$_H$ only and 11 displayed both conformations.

Relaxing this *in situ* prepared unfolded NuG2-apo-RD-NuG2 led to three distinct refolding events, suggesting that the two NuG2 and apo-RD had managed to refold during relaxation. Subsequent stretching revealed that the unfolding of apo-RD always occurred at very low forces (between 5-10 pN), significantly lower than that of NuG2 domains. Thus, we limited

the stretching to only trigger the unfolding of apo-RD. It is evident that the unfolded apo-RD can refold efficiently against a stretching force of 5-10 pN (Figure 3.3A). The unfolding and refolding of apo-RD occurred in a two-state fashion without the accumulation of any intermediate state. The folding of apo-RD is rapid and robust, with no fatigue or misfolding observed in more than 360 consecutive cycles, suggesting that the folding of the RD polypeptide chain itself is efficient and fast.

It has been reported that folded apo-RD has a virtually identical tertiary structure as holo-RD except for not having the iron-sulfur center.[244-245] Fitting the length change versus the force of apo-RD to the WLC model revealed that the unfolding and refolding of apo-RD showed a ΔLc of ~17 nm (Figure 3.3B), which agrees well with the ΔLc expected from the complete unfolding of apo-RD: 53 aa × 0.36 nm/aa - 1.2 nm = 17.9 nm, where 1.2 nm is the distance between the N- and C-termini in folded RD. This result strongly indicated that unfolded apo-RD is able to refold into its folded three-dimensional structure rapidly by itself without the assistance of the ferric ion.

Moreover, we observed that the unfolding and folding forces of apo-RD showed a clear bimodal distribution, suggesting the existence of two distinct populations of apo-RD. One population of apo-RD, termed as apo-RD_L, showed an average unfolding force of 5.6 ± 1.0 pN (average ± standard deviation) and a folding force of 5.1 ± 0.8 pN, while the other population, termed as apo-RD_H, was mechanically more stable and showed an unfolding force of 11.7 ± 1.0 pN and folding force of 9.3 ± 1.1 pN (n = 1958) (Figure 3.3C). ΔLc for these two populations is indistinguishable from each other, suggesting that both conformers may have structures that are similar to that of native apo-RD.

We used the method proposed by Oesterhelt[140] to extract the unfolding and refolding rate constants, and used the Bell-Evans model[95] to measure the kinetic parameters for the unfolding and refolding. This analysis revealed that these two conformers displayed different intrinsic unfolding-refolding rate constants α_0 and β_0 at zero forces (Figure 3.3D). Both conformers can refold fast, with a similar β_0 of around 1.5×10^4 s^{-1}, but show distinct unfolding (Δx_u) and refolding distance (Δx_f), respectively.

It is of note that the same apo-RD molecule can switch between the two conformations (apo-RD_H and apo-RD_L) during repeated stretching-relaxation cycles (Figure 3.4). As evidenced by the few inter-conversion events, it did not randomly fold into one of the conformations each time, and the interconversion of the two conformers was slow. These results suggested that the folded apo-RD displayed two distinct conformations, instead of having two paralleled folding pathways leading to the same conformation. Given the free energy difference of the two distinct conformers (~7 kT), these two conformers are likely kinetically trapped on the time scale of our OT experiments, leading to their slow inter-conversion. Interestingly, previous NMR studies also revealed two distinct conformations of apo-RD, and one of them was incapable of uptaking iron and folding into native holo-RD.[187] However, it is unknown if the distinct conformations observed in our OT experiments are the same as those observed in NMR experiments. Nonetheless, our results indicated that the folding of apo-RD itself is rapid and robust.

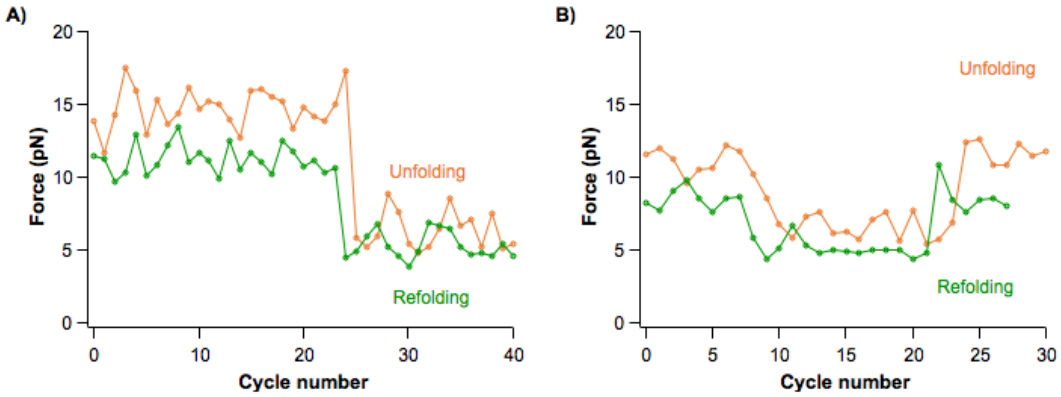

Figure 3.4 Unfolding (orange) and refolding (green) forces of two apo-RD molecules in consecutive unfolding-refolding cycles. The synchronous changes of unfolding and refolding forces indicate the switch of apo-RD between apo-RD_H and apo-RD_L.

3.3.3 Distinct Conformations of Apo-RD Does Not Originate from Proline Isomerization

The conformation of the proline residue, which can be either cis or trans, is a structural feature that can persist in the unfolded state. The structural difference between the cis and the trans conformation of the peptide bond is significant, thus the isomerization can guide the unfolded protein to fold into different conformations. In addition, the trans state is energetically only slightly preferred over the cis state, and the barrier between the states is high, resulting in a slow interconversion in equilibrium. RD has five prolines adopting the trans conformation in the native structure. To investigate whether the two distinct folded conformations, apo-RD_H and apo-RD_L, arise from the unfolded state or proline isomerization, we made a series of proline-to-glycine mutants of RD, constructed polyproteins Cys-NuG2-RD mutant-NuG2-Cys in the same way as we did on wild-type RD, and stretched them by OT.

We first mutated all the five prolines in RD to glycines to build a mutant of RD(no proline). Stretching NuG2-RD(no proline)-NuG2 in the presence of EDTA enabled us to study the folding-unfolding behavior of apo-form RD(no proline). As shown in Figure 3.5, the holo-RD(no proline) unfolded at ~50 pN in a similar way to wild-type RD (Cycle 1). However, once

unfolded, unfolded apo-RD(no proline) did not fold anymore (Cycle 2-3). This suggested that apo-RD with all the five prolines mutated to glycines lost its folding competence.

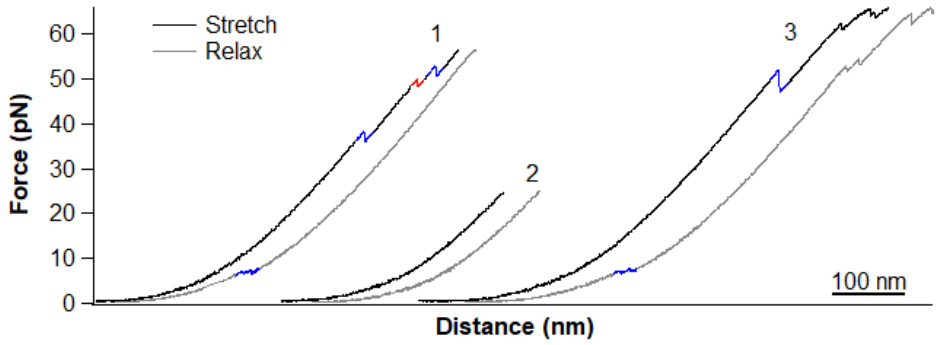

Figure 3.5 Representative force-distance curves of NuG2-RD(no proline)-NuG2 in the presence of EDTA at a pulling speed of 50 nm/s. The unfolding event of holo-RD(no proline) is colored in red, and the unfolding and refolding events of NuG2 are colored in blue. Curves are offset for clarity.

We then made another four RD mutants, RD(P19G), RD(P33G), RD(P39G) and RD(P25,44G), with one or two of the five prolines mutated to glycines. These mutants were observed to still be able to fold and unfold after the initial unfolding of the holo-form in the presence of EDTA, however, their folding and unfolding forces also showed bimodal distributions and similarly suggested two stable conformations as the wild-type apo-RD. This indicated that none of the five prolines in RD was responsible for the distinct conformations of folded apo-RD. Combining these results, we concluded that the apo-RD_H and apo-RD_L did not originate from proline isomerization, but may come from some other stable residual structures in unfolded apo-RD.

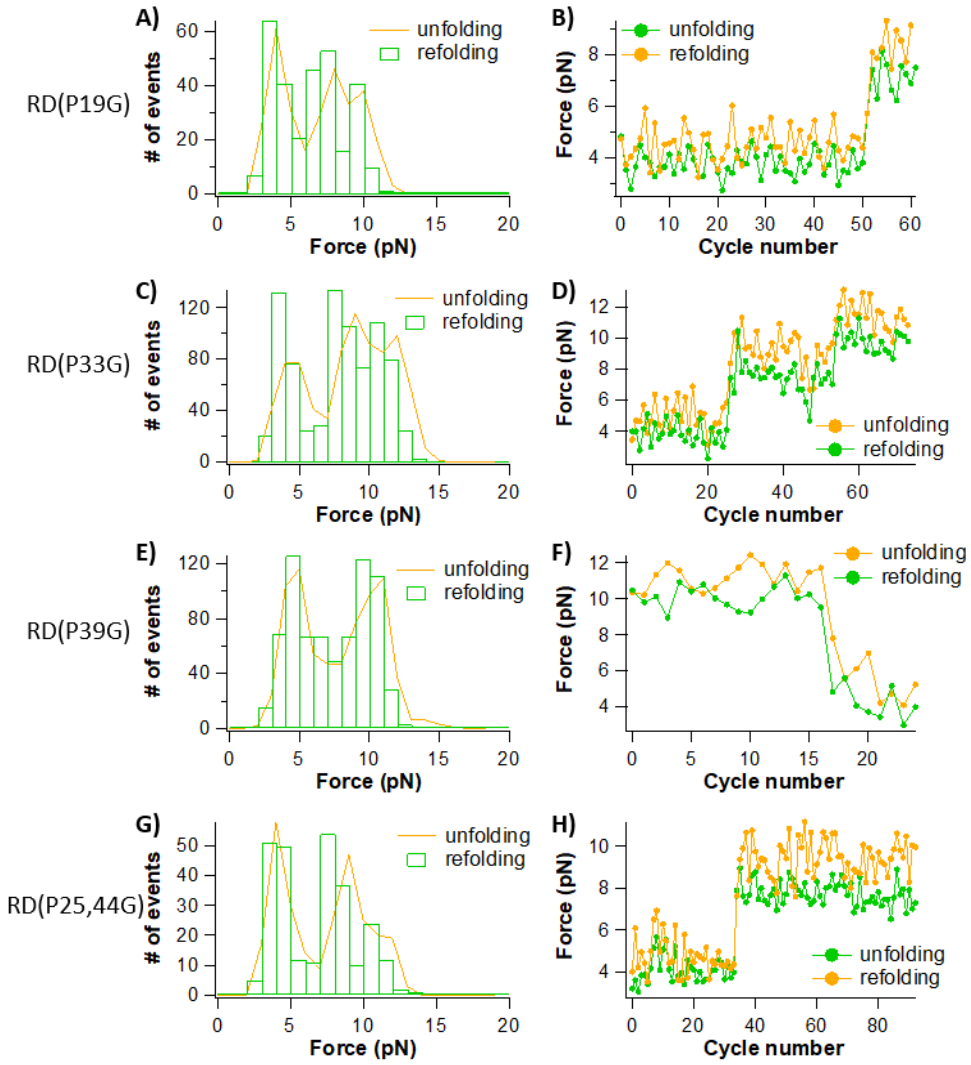

Figure 3.6 Mechanical folding-unfolding behaviors of single- and double-proline mutants of RD. (A), (C), (E) and (G) are histograms of the unfolding and folding forces, and (B), (D), (F) and (H) show the unfolding and folding forces of a molecule of RD mutants in consecutive unfolding-refolding cycles.

3.3.4 Direct Observation of the Folding of Holo-RD in the OT Experiments

To investigate the folding of holo-RD, we stretched holo-RD in Tris buffer. In these experiments, there was no additional iron source in the buffer and the only ferric ion was from the holo-RD itself. Our previous AFM studies showed that after the unfolding of holo-RD, the ferric ion may remain associated with the unfolded RD polypeptide chain for an extended period

of time,[238] leading to an unfolded holo-RD that can serve as an ideal system for investigating the folding mechanism of holo-RD.

Figure 3.7A shows representative unfolding and refolding curves of holo-RD in such experiments. The unfolding of the pristine holo-RD was achieved by holding the stretched molecule at ~55 pN for 30 s (Curve 1). Then the unfolded polypeptide chain was relaxed to zero force (Curve 2). In addition to the two NuG2 refolding events at ~8 pN, one additional refolding event was observed, which can be readily assigned to the refolding of RD. The folding of RD followed a two-state pathway with no detectable intermediate, and the resultant ΔL_c is ~17.0 nm, the same as that of apo-RD (Figure 3.7B). This result suggested that the folding event observed here for the unfolded holo-RD corresponds to the folding of the RD polypeptide chain itself.

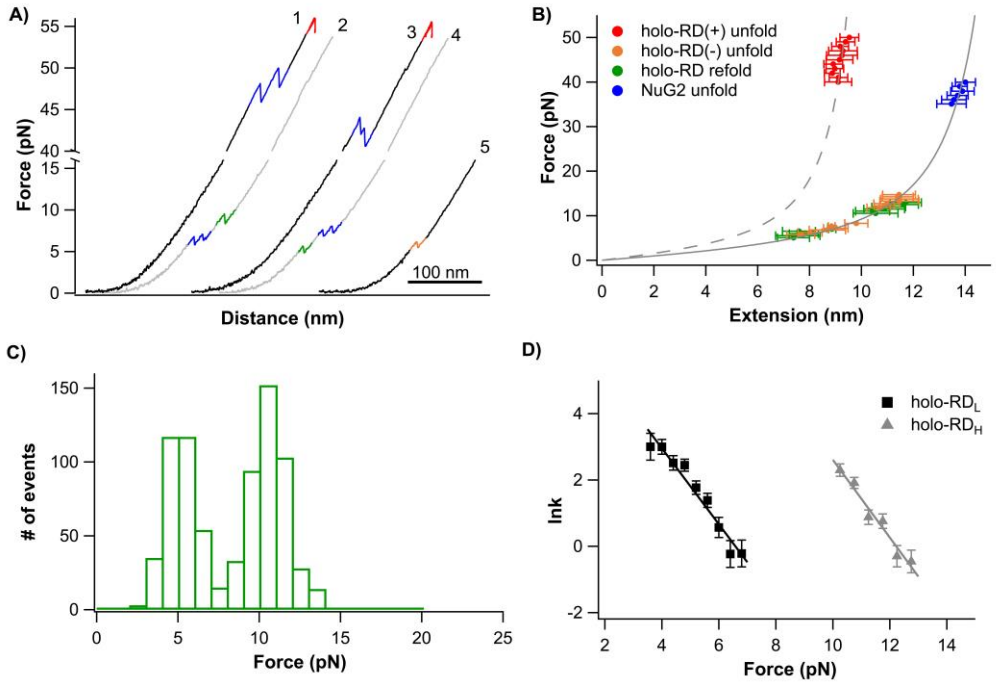

Figure 3.7 Unfolding-folding of holo-RD observed in OT experiments. (A) Representative force-distance curves of NuG2-holo-RD-NuG2 in Tris buffer. Curve 1 and 3 were obtained by holding the molecule at ~55 pN until holo-RD unfolds. (B) ΔLc analysis of the unfolding-refolding of holo-RD (orange and green) and the unfolding of NuG2 (blue). The grey curves are WLC fits to the experimental data. The high-force unfolding events of holo-RD can be fitted by a WLC with a persistence length of 0.8 nm and ΔLc of 12.2 ± 0.1 nm (grey dash curve). The low-force folding-unfolding events of holo-RD and the unfolding events of NuG2 can be fitted by a WLC with a persistence

length of 0.8 nm and ΔLc of 17.0 ± 0.1 nm (grey solid curve). (C) Refolding force histogram of unfolded holo-RD at a pulling speed of 50 nm/s (n = 767). The refolding force histogram shows a bimodal distribution, with the first peak centered around 5.1 pN and the second at 10.5 pN. (D) Force-dependent refolding rates of unfolded holo-RD. Fitting the experimental data to the Bell-Evans model yielded folding kinetic parameters for unfolded holo-RD. For holo-RD_L, $\beta_0 = (1.9 \pm 1.0) \times 10^3\ s^{-1}$ and $\Delta x_f = 4.7 \pm 0.4$ nm; and for holo-RD_H, $\beta_0 = (1.7 \pm 2.3) \times 10^6\ s^{-1}$ and $\Delta x_f = 4.8 \pm 0.5$ nm. In all 26 molecules we measured, 6 of them adopt holo-RD_L only, 12 adopt holo-RD_H only, and 8 displayed both conformations.

To check if the iron-sulfur center had been successfully reconstituted after the folding of the RD polypeptide chain, we stretched the refolded RD again in the subsequent cycle. Intriguingly, we observed two different unfolding behaviors of the refolded RD (Figure 3.7A, Curve 3-5). In curve 3, the unfolding of the refolded RD only occurred when the protein chain was held at ~55 pN for ~30 s, and the resultant ΔLc was ~12.2 nm (Figure 3.7B, red symbols). These are the exact signatures for the mechanical unfolding of the holo-RD, indicating that the iron-sulfur center had been successfully reconstituted. However, in other cases (Curve 5), the refolded RD unfolded at low forces (below 20 pN) and showed a ΔLc of ~17.0 nm (Figure 3.7B, orange symbols), suggesting that the unfolded RD managed to refold into its native conformation but the iron-sulfur center was not reconstituted. These two refolding behaviors can switch within the same RD molecule, suggesting that the ferric ion remained associated with the unfolded polypeptide chain during the whole experiment.[23] Therefore, starting from the unfolded holo-RD polypeptide chain with an associated ferric ion, the polypeptide chain of RD itself can fold rapidly, subsequently leading to two distinct conformations: holo-RD(+) with the fully reconstituted iron-sulfur center, and holo-RD(-) without the iron-sulfur center, where (+) indicate that the iron-sulfur center has been successfully reconstituted and (-) indicated otherwise.

The refolding forces of unfolded holo-RD showed a clear bimodal distribution, with the first peak centered at 5.1 ± 1.0 pN and the second at 10.5 ± 1.2 pN (n = 767) (Figure 3.7C). This bimodal distribution is similar to that of apo-RD, suggesting that the unfolded holo-RD also

shows two distinct populations: unfolded holo-RD_H and unfolded holo-RD_L, where H refers to the high refolding force and L to the low refolding force, and the binding of the ferric ion to the unfolded RD does not affect the two intrinsic conformations of RD polypeptide chain. Kinetic analysis revealed that the folding of the polypeptide chain itself with the associated ferric ion (both unfolded holo-RD_L and holo-RD_H) is rapid (Figure 3.7D). This result is similar to the folding of apo-RD_L and apo-RD_H.

Detailed analysis of the folding of the two populations revealed that unfolded holo-RD_L is largely incapable of folding into the holo-RD(+) with a fully reconstituted iron-sulfur center, as only less than 9% of the low force events led to the folding of holo-RD(+). This result suggested that the conformer apo-RD_L and its ferric loaded form holo-RD_L are kinetically trapped misfolded states of RD, which are not competent in producing the functional holo-RD(+), which the more stable conformer apo-RD_H and holo-RD_H are competent in folding into holo-RD(+).

For the unfolded holo-RD_H, ~39% of the events led to the successful folding of holo-RD(+), while 61% only folded into holo-RD(-). Further examination revealed that the refolding force distribution of unfolded holo-RD_H does not show any difference between the events leading to holo-RD(+) versus those leading to holo-RD(-) (Figure 3.8), suggesting that the reconstitution of the iron-sulfur center occurred after the RD polypeptide chain has refolded.

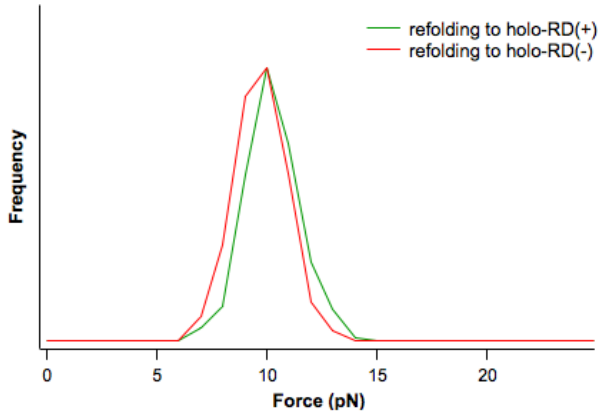

Figure 3.8 Refolding force distribution of unfolded holo-RD$_H$ leading to holo-RD(+) versus those leading to holo-RD(-).

3.3.5 Unfolded Holo-RD$_H$ Display Two Different Conformers with Different Competences to Reconstitute the Iron-Sulfur Center after Refolding

The successful refolding of the holo-RD(+) requires both correct folding of the polypeptide chain and a fully reconstituted iron-sulfur center. While the RD polypeptide chain can refold robustly, only some of the unfolded holo-RD$_H$ can successfully fold into holo-RD(+) with a fully reconstituted iron-sulfur center. In order to understand the folding mechanism of holo-RD(+) from the unfolded holo-RD$_H$, we investigated the conversion of these states during consecutive unfolding-refolding cycles. Figure 3.9A shows the unfolding and refolding forces in consecutive unfolding-refolding cycles for two RD molecules. The unfolding force of 55 pN indicates that the RD had refolded into holo-RD(+) with a fully reconstituted iron-sulfur center, while unfolding force below 20 pN indicates that RD refolded into holo-RD(-).

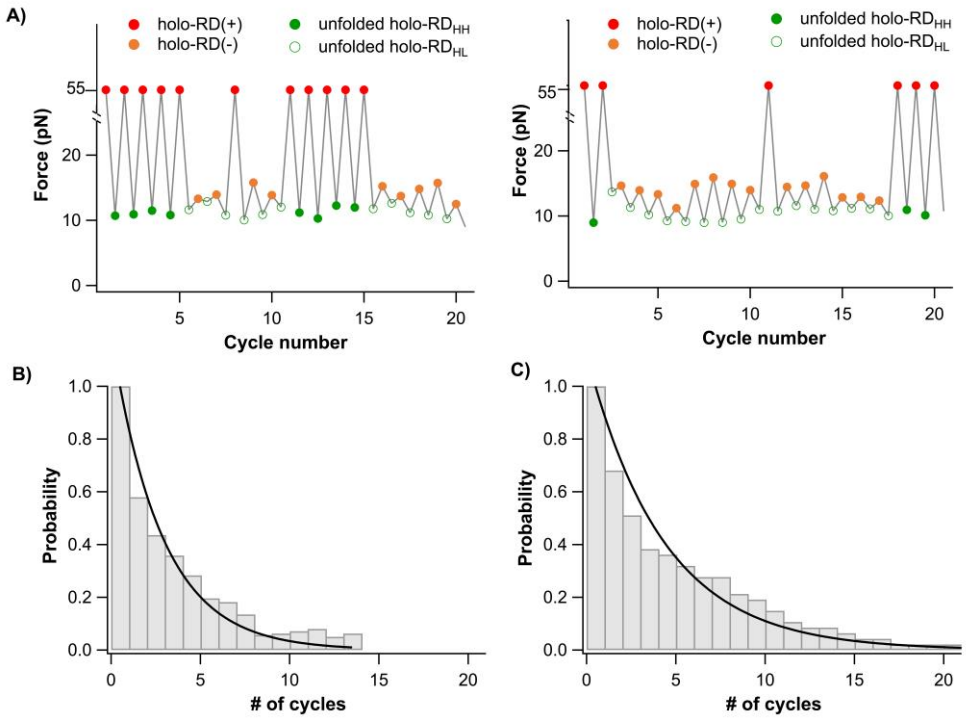

Figure 3.9 Folding competency is determined by the state of the unfolded holo-RD. (A) Unfolding and refolding forces of two holo-RD molecules in continuous unfolding-refolding cycles. Unfolding force of 55 pN indicates that the molecule has a completely reconstituted coordination center, while unfolding force below 20 pN indicates that the coordination center has not been reconstituted successfully. (B-C) The probability of observing n consecutive holo-RD(+) folding events (B) and n consecutive holo-RD(-) folding events. The probability distribution follows a binomial distribution. Solid lines are the fits of $f(n)=p^n$ to the experimental data with a p of 0.70 ± 0.01 and 0.79 ± 0.01 respectively.

From the conversion between different states of the single holo-RD molecule, it is evident that the unfolded holo-RD$_H$ displayed distinct probabilities of transition into holo-RD(+): the states indicated by green solid symbols had a higher probability of transition into holo-RD(+), while the states indicated by green open symbols had a lower probability of transition into holo-RD(+). This result strongly indicated that the states indicated by green solid and open symbols are two distinct states and display different competence towards reconstituting the metal center, despite that they displayed the same ΔLc. Thus, we termed the high competence species towards folding into the holo-RD(+) as unfolded holo-RD$_{HH}$ (green solid symbols), and the low competence state (green open symbols) was termed as unfolded holo-RD$_{HL}$ as most of which

fold into holo-RD(-). It is important to note that these two distinct states are solely determined by the unfolding of the holo-RD(+).

Evidently, four states were involved in the folding of holo-RD_H: unfolded holoRD_{HH} (green solid), unfolded holoRD_{HL} (green open), folded holo-RD(-) (orange open) and folded holo-RD(+) (orange solid). To quantitatively analyze the transition between different states, we modeled them as a Markov chain (Figure 3.10). The transition probability (*p*) between different states can be directly measured from the experimental data (Figure 3.9B-C). For example, the probability of observing n-consecutive folding events to holo-RD(+) measures the transition probability *p* of going from holo-RD(+) to unfolded holo-RD_{HH}, which was estimated to be 0.7 (Figure 3.9B), while the probability of observing n-consecutive folding events to holo-RD(-) measures the transition probability *p* from unfolded holo-RD_{HL} to holo-RD(-), which was estimated to be 0.79 (Figure 3.9C and Figure 3.10). Based on these transition probabilities, this model predicted that from the unfolded holo-RD_H, 47% trajectories should lead to the folding of holo-RD(+); and starting from holo-RD(+) and holo-RD(-), 67% trajectories should lead to the unfolded holo-RD_{HL}. This prediction is in close agreement with our experimentally measured values (39% and 70%, respectively) (Table 3.1).

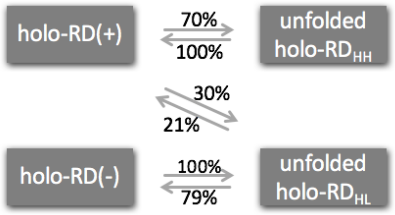

Figure 3.10 4-state Markov chain model for unfolding and refolding of holo-RD. The transition probability between states is calculated from experimental data and indicated in the schematics.

Table 3.1 Calculated and experimental occupancy of each species in the unfolding-refolding process of holo-RD.

	holo-RD(+)	holo-RD(-)	unfolded holo-RD_{HH}	unfolded holo-RD_{HL}
Calculated occupancy	0.47	0.53	0.33	0.67
Experimental occupancy	0.39	0.61	0.30	0.70

3.3.6 Different Competences of Reconstituting the Iron-Sulfur Center May Result from Different Coordination States of the Ferric Ion Associated with the Unfolded RD

In order to rupture the iron-sulfur center, at least two ferric-thiolate bonds need to be ruptured from the same CXXC chelation motif. In our previous AFM studies, we showed that the iron-sulfur center can be ruptured following different pathways.[26, 246] The ferric-thiolate bonds in the iron-sulfur center of RD may be ruptured concurrently or sequentially in a stochastic fashion, yielding the unfolded RD with two different ferric coordination states: two-coordinate and mono-coordinate ferric ion. Given the fact that the unfolding of the holo-RD(+) is the determining factor for the competence of the unfolded holo-RD_H, it is most likely that the competency is related to the two different ferric-coordination states in the unfolded protein. The unfolded species with two-coordinate ferric ion is likely of high competence, while the species with mono-coordinate ferric ion is largely of low competence to reconstitute the metal center.

Indeed, the unfolding of holo-RD(+) has a 70% likelihood of resulting in an unfolded holo-RD_{HH} form (Figure 3.9B). This is close to the probability of concurrent rupture of the iron-sulfur center (80%), where two ferric-thiolate bonds in the same CXXC chelation motif rupture concurrently, as determined from AFM experiments using a loop insertion variant of RD.[26] This result revealed the critical importance of the two-coordinate ferric site for the folding of holo-RD(+) under the near-native condition.

3.4 Discussion

3.4.1 The Complete Folding Pathway of Holo-Rubredoxin: The Critical Importance of the Two-Coordinated Ferric Site for the Folding of Holo-Rubredoxin

Using single-molecule optical tweezers techniques, here we investigated the folding mechanism of rubredoxin under the near-native condition at the single-molecule level. Our experimental data allowed us to provide a complete folding mechanism for the holo-ferric RD under the near-native condition for the first time (Figure 3.11).

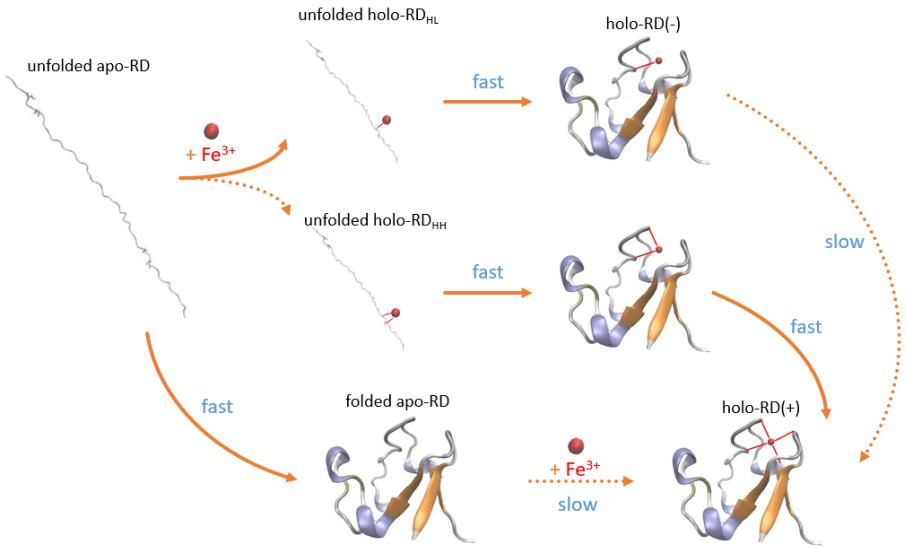

Figure 3.11 The productive folding mechanism of holo-RD under a quasi-native condition. In this folding mechanism, the misfolded state apo-RD$_L$ is not included.

Starting from the unfolded and apo-polypeptide chain, RD can spontaneously fold into the three-dimensional structure of the apo-state. A kinetically trapped misfolded state exists, which is not productive in the folding of the holo-RD with a fully reconstituted metal center. The structural origin of the misfolded state remains unclear. Since apo-*pf*RD showed high thermodynamic stability,[187, 199] it is likely that residual structure in unfolded apo-*pf*RD may play a role in the formation of the misfolded state.

For holo-RD, the folding mechanism is much more complex. In the presence of Fe^{3+}, the two processes, folding of RD and the binding of Fe^{3+}, are clearly intertwined. However, the folding of apo-RD (with a rate constant of ~10^5 s^{-1}) is much faster than the binding of Fe^{3+} to apo-RD (with a rate constant of ~100 s^{-1}). It is well-known from ensemble studies that reconstituting holo-RD from the folded apo-RD by adding iron source has a low efficiency,[187] suggesting that the binding of Fe^{3+} to RD prior to the folding of apo-RD is critical for the folding of holo-RD. Our OT results provided a detailed mechanistic picture for the folding of RD with preloaded Fe^{3+} (the unfolded holo-RD).

The coordination state of the bound ferric ion directly determines the outcome of the folding of unfolded holo-RD_H. The folding of holo-RD(+) always initiates from the folding of the polypeptide chain itself, which is a fast and robust process leading to a three-dimensional structure that is similar to the native holo-RD structure without the iron-sulfur center. If the ferric ion is coordinated by two neighboring cysteines (either Cys5/Cys8 or Cys38/Cys41), corresponding to the unfolded holo-RD_{HH}, the reconstitution of the iron-sulfur center in RD is fast after the polypeptide chain has refolded, leading to the complete folding of the holo-RD(+) with a fully reconstituted iron-sulfur center. However, if the ferric ion is mono-coordinated, which corresponds to the unfolded holo-RD_{HL}, the reconstitution of the iron-sulfur center is slow after RD has folded into the three-dimensional structure. Compared with the mono-coordinated state, the two-coordinate ferric site likely positions the ferric ion well enough to the vicinity of the other two cysteines. This arrangement may significantly reduce the conformational flexibility and freedom of the polypeptide sequence around the metal center to allow the two additional cysteines to react efficiently to form the tetrahedral iron-sulfur center. However, likely due to the

large conformation freedom in the metal center, the reconstitution of the iron-sulfur center from the mono-coordinate ferric ion in holo-RD(-) is slow.

The folding mechanism we discovered here highlights the critical role of the two-coordinated ferric site in the folding of holo-RD, and is the first folding mechanism of RD under a near-native condition. This folding mechanism is distinct from the previously proposed mechanisms that are commonly used to describe the folding of metalloproteins, i.e. binding before folding mechanism and binding after folding mechanisms.[146] The mechanism proposed here suggested a novel binding-folding-reconstitution mechanism, where the formation of the two coordinate iron species is a critical step. However, the two-coordinate iron site does not serve as a nucleus for folding as commonly seen in the binding before folding mechanism. Instead, the two-coordinate ferric site serves as a critical intermediate for the metal center reconstitution. The folding of the protein structure is rather an intrinsic property of RD polypeptide chain itself.

Previously, the iron priming mechanism was proposed for the folding of ferrous rubredoxin in the presence of chemical denaturants.[187, 199] In this mechanism, the formation of two-coordinate iron species is the first step, followed by the formation of the tri- and tetra-coordinated iron-sulfur center. The fully reconstituted iron-sulfur center can then drive the folding of the polypeptide chain into the holo-RD(+) in the presence of a high concentration of denaturants. [187, 199] However, different from these prior studies, we found that the folding of RD polypeptide chain precedes the reconstitution of the iron-sulfur center under near-native conditions, suggesting that the formation of the tri- and tetra-coordinated iron-sulfur center is not a prerequisite for the folding of holo-RD under near-native conditions. These differences suggest

that the folding of holo-RD may follow different pathways using different driving forces under different conditions.

Nonetheless, the iron-priming mechanism proposed for the folding of RD in the presence of denaturants[187, 199] bears relevance for the new mechanism we proposed here, as the formation of the two-coordinate Fe^{3+} can be considered as a form of iron-priming.

3.4.2 Biological Implication for the Folding of Metalloproteins *In Vivo*

Although RD is the simplest iron-sulfur protein, our single-molecule studies revealed a complex folding mechanism for RD. Our results may also have important implications for the folding of other complex iron-sulfur proteins *in vivo*. Inside the cell, the folding of more complex iron-sulfur proteins, such as ferredoxin, often requires biogenesis machinery and is much more complex.[186] This biogenesis process often involves the *de novo* assembly of an Fe-S cluster on a scaffold protein, followed by the transfer of the Fe-S cluster from the scaffold to a target apo-protein and its subsequent maturation into the holo-iron sulfur protein. Apart from the protection role offered by scaffold protein to the otherwise labile Fe-S cluster, it is intriguing why a scaffold is required to transfer the Fe-S cluster to the apo-protein. Our results and proposed folding mechanism may offer some possible hints. It is known that the uptake of iron for the apo-protein is slow. By forming the Fe-S cluster on a scaffold protein, it is possible to present the Fe-S cluster to the *bona fide* apo-protein to form a two-coordinated species, which can then react with the other chelating motif in the apo-protein efficiently to reconstitute the iron-sulfur center after the scaffold protein dissociates from the pre-assembled Fe-S cluster. By doing so, the two-coordinate site can effectively reduce the entropic barrier for the reconstitution of the iron-sulfur center and greatly facilitate the folding of complex iron-sulfur proteins *in vivo*.

3.5 Conclusions

Using optical tweezers, we demonstrated the complete folding mechanism of RD probed under a quasi-native condition for the first time. The refolding of holo-RD is initiated by the fast and robust folding of its polypeptide chain, followed by the reconstitution of the iron-sulfur center. The formation of the two-coordinate Fe^{3+} site is critical for the reconstitution of the iron-sulfur center. Our results suggest a new folding mechanism for metalloproteins: the binding-folding-reconstitution mechanism, which clearly highlights the synergistic effects between polypeptide chain folding and metal coordination reconstitution in the folding process of iron-sulfur proteins. Our results also provide a new general approach to investigate the folding process of metalloproteins, including complex iron-sulfur proteins at the single-molecule level, and may shed new insights into the folding of complex iron-sulfur proteins *in vivo*.

3.6 Experimental Section

3.6.1 Protein Engineering

The gene of RD carrying a 5' BamHI (G'GATCC) and a 3' KpnI (G'GTACC) restriction site was amplified using polymerase chain reaction, and its sequence was confirmed by direct DNA sequencing. The genes of proline-to-glycine mutants of RD, including RD(no proline), RD(P19G), RD(P33G), RD(P39G) and RD(P25,44G), were constructed via standard site-directed mutagenesis methods using the RD gene as a template. The genes of NuG2-RD-NuG2 and NuG2-RD mutant-NuG2 were constructed following a stepwise digestion and ligation scheme using restriction sites including BamHI, KpnI, HindIII and EcoRI, and further subcloned into a modified pET expression vector, which allows for adding a cysteine at both N- and C-termini of the protein. Protein expression and purification were performed using the same procedure as described in chapter 2 (2.6.1). NuG2-RD-NuG2 containing ferric-form RD was

purified by ion exchange chromatography using a Mono Q 5/50 anion exchange column (GE Healthcare).[23]

3.6.2 Preparation of DNA-Protein Chimera

Double-strand DNA (dsDNA) handles were prepared via the methods described previously. Two dsDNA handles of 802 and 558 bp were generated via regular PCR amplification. The template pGEMEX-1 plasmid was a generous gift from Professor X. F. Zhang of Lehigh University. The modified primers were purchased from Integrated DNA Technologies (IDT Inc, San Jose, CA). The reverse primer was 5'-NH_2 modified, and the forward primers were 5'-biotin and 5'-digoxigenin modified, respectively. The sequences of the primers were as follows:

Forward primer for dsDNA handles of 802 bp: 5'-Bio-CAA-AAA-ACC-CCT-CAA-GAC-CC

Reverse primer for dsDNA handles of 802 bp: 5'-NH_2-CGA-CGA-TAA-ACG-TAA-GGA-CAT-C

Forward primer for dsDNA handles of 558 bp: 5'-Dig-CAA-AAA-ACC-CCT-CAA-GAC-CC

Reverse primer for dsDNA handles of 558 bp: 5'-NH_2-GCT-ACC-GTA-ATT-GAG-ACC-AC

After the PCR amplification, QIAquick PCR purification kit (Qiagen, Germantown, MD) was used to purify the PCR products. Subsequently, dsDNA handles were allowed to react with 4-(N-Maleimidomethyl) cyclohexanecarboxylic acid N-hydroxysuccinimide ester (SMCC, Sigma-Aldrich, St. Louis, MO) overnight, which enabled the amine group at the end of the dsDNA handles replaced by the maleimide group. 50 μM of the freshly expressed proteins were reduced with 1 mM tris(2-carboxyethyl) phosphine (TCEP) (Sigma-Aldrich, St. Louis, MO) for 1 hour, and the remaining TCEP was removed by Zeba desalting columns (7kMW, Thermo Fisher Scientific, Waltham, MA). Then the reduced proteins were diluted to ~3 μM by Tris

buffer (20 mM Tris, 150 mM NaCl, pH 7.4). 1 μL of the diluted protein was added into 1 μL of mixed dsDNA handles (both are at 3 μM). The thiol-maleimide reaction was carried out at room temperature overnight. The formed dsDNA-protein chimera was diluted by Tris buffer (20 mM Tris, 150 mM NaCl, pH 7.4) to ~10 nM and ready for optical tweezers experiment.

3.6.3 Optical Tweezers Based-SMFS Experiment

The optical tweezers based-SMFS experiment was carried out with Minitweezers setup (http://tweezerslab.unipr.it/cgi-bin/mt/home.pl).[240] The liquid chamber of optical tweezers was filled with Tris buffer (20 mM Tris, 150 mM NaCl, pH 7.4) to provide the working environment, unless otherwise noted. In a typical experiment, 1 μL of 0.5% streptavidin modified polystyrene beads (1% w/v 1 μm, Spherotech Inc, Lake Forest, IL) was diluted to 3 mL and injected into the fluid chamber. A single streptavidin modified polystyrene bead was captured by a laser beam and then held by a micro pipette tip within the chamber. 1 μL of 5 nM DNA-protein chimera was allowed to react with 5 μL of 0.1% antidigoxigenin modified polystyrene beads (0.5% w/v, 2 μm, Spherotech Inc, Lake Forest, IL) for 30 min at room temperature. The mixture was then diluted to 3 mL and injected into the chamber. A single anti-digoxigenin (anti-Dig) modified polystyrene bead was captured by the laser trap. The laser trap controlled the movement of anti-dig bead against the streptavidin modified polystyrene bead fixed on the pipette tip to carry out the force-extension experiments (Figure 3.1C). In this work, the stretching and relaxing experiments were carried out at a constant speed of 50 nm/s.

3.6.4 Calculating the Kinetics of Unfolding/Folding of Proteins

We used the method proposed by Oesterhelt *et al.* to extract the folding and unfolding rate constants at different forces from the force-distance curves.[140] The curves were divided into time windows (Δt) that are small enough so that the force can be considered constant within the

time window. The probability of protein folding/unfolding within Δt can be calculated as $P(F) = N(F)/M(F)$, where the $N(F)$ is the total number of all the folding or unfolding events at the force of F and $M(F)$ is the total number of time windows at the force of F. The rate constant of protein folding/unfolding at the force of F can be calculated as $k(F) = P(F)/\Delta t$.

Chapter 4: The Mechanical Folding-Unfolding Behavior of a Plant-Type [2Fe-2S] Ferredoxin Characterized by Single-Molecule Force Spectroscopy

4.1 Synopsis

Metalloproteins are ubiquitous in nature and play important roles in organisms. Understanding the folding mechanisms of metalloproteins has profound significance, as correct folding is a prerequisite for proper functioning. As an important type of iron-sulfur protein, the folding behavior of [2Fe-2S] ferredoxin is of particular interest, and the folding process is inaccessible by ensemble spectroscopic experiments. Here we used optical tweezers-based single-molecule force spectroscopy to study the mechanical folding-unfolding mechanism of a [2Fe-2S] ferredoxin. Stretching ferredoxin from its N-terminus and residue 93 can unfold the protein in two steps, with its β-sheet being unfolded first and coordination center being ruptured subsequently. Once relaxed, the mechanically unfolded ferredoxin mostly refolds into an ensemble of misfolded states with weaker mechanical stabilities comparing to its native structure. In very rare cases, the refolded ferredoxin shows a correctly refolded β-sheet or a reconstituted coordination center, however, one with its whole native structure successfully reformed is not found. Our results point out the complex folding behavior of the polypeptide chain of ferredoxin and the challenges to recover the native structure *in vitro*, and may bring important implications on our understanding of the folding of metalloproteins *in vitro* and *in vivo*.

4.2 Introduction

It is well-known that proteins must fold into their specific three-dimensional structures to acquire their biological functions.[4] However, through what mechanisms proteins fold to ensure

the rapidness and reliability of folding has been inadequately understood until today. This is particularly the case for metalloproteins, of which the folding process is more complicated than that of metal-free proteins, as it is driven under a synergistic effect consisting of not only protein chain folding but also metal coordination.[150, 164]

The key to resolve the folding-unfolding mechanisms of metalloproteins is thus to dissect the intertwined effects of protein folding and metal coordination. In traditional ensemble spectroscopic experiments, this is usually achieved by monitoring two spectroscopic signals simultaneously during the protein folding-unfolding process, with one describing the protein chain structure (e.g., circular dichroism) and the other depicting the metal coordination effect (e.g., UV-Vis spectrum).[187, 199, 247-249] In SMFS, upon stretching, the force-bearing structures (including folded secondary structures and coordination centers) in a folded metalloprotein molecule will be unfolded or disrupted by the force one by one following the order that the force applies on, and the distinct force and the contour length change (ΔLc) indicate the structural change of the protein. Thus, for metalloproteins with structurally separated folded secondary structures and metal coordination centers, if the unfolding or rupture events are distinctive on SMFS, SMFS can distinguish the protein folding and metal coordination effects and depict the full folding-unfolding pathways.

Iron-sulfur proteins are an important class of metalloproteins in nature, and among them, the plant-type [2Fe-2S] ferredoxin is an electron carrier in photosynthesis widely existing in plants, algae and cyanobacteria.[250-251] It has a highly conserved folded structure consisting of a four-stranded β-sheet formed by its N- and C-termini (a β-grasp folding motif), and a coordination center in the middle of its sequence formed by a [2Fe-2S] cluster and four cysteinyl sulfurs in a -CxxxxCxxCx$_n$C- chelating motif (Figure 4.1A-B).[252] The mechanical unfolding

mechanism of a plant-type [2Fe-2S] ferredoxin has been studied by AFM-based SMFS, and it was found that stretching from its N- and C-termini could mechanically unfold this protein in two steps, by unfolding its outer β-sheet first and rupturing the inner coordination center subsequently.[28] Both of these force-bearing structures demonstrated high mechanical stabilities and the unfolding or rupture generated distinctive ΔLcs. In addition, the [2Fe-2S] cluster was found to remain attached to the mechanically unfolded ferredoxin for some time, and the mechanically unfolded protein was observed to refold in some cases in SMFS experiments.[28] These special structural and mechanical folding-unfolding properties make ferredoxin a perfect model system for studying the intertwined effects between the protein folding and the metal coordination during the folding-unfolding mechanisms of metalloproteins. However, due to the limited force resolution of AFM, the folding process was not directly observed, and a detailed folding mechanism of ferredoxin was not revealed.

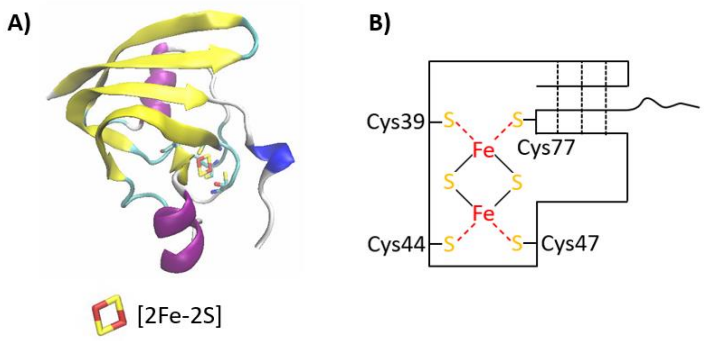

Figure 4.1 Structure of sFD. (A) Three-dimensional structure of the oxidized [2Fe-2S] ferredoxin Type I from Spinacia oleracea (sFD, PDB code: 1A70).[252] sFD is an α/β protein containing 97 amino acids and a [2Fe-2S] cluster. The [2Fe-2S] cluster is coordinated by four cysteinyl sulfurs in a pseudo-tetrahedral geometry. The [2Fe-2S] cluster is highlighted, where the ferric ions are colored in red, and inorganic sulfurs in yellow. (B) Schematic of sFD. The red dashed lines indicate the coordination bonds between the [2Fe-2S] cluster and the cysteinyl sulfurs, and the black dashed lines represent the hydrogen bonds between the β-strands.

Optical tweezers (OT)-based SMFS has a higher force resolution at the low-force regime, thus is better at monitoring protein folding processes than AFM-based SMFS. Combining with protein engineering techniques, here, we used OT-based SMFS to directly probe the folding

mechanism of ferredoxin from *Spinacia oleracea* (sFD). sFD is a typical plant-type ferredoxin containing 97 amino acids (aa) and a [2Fe-2S] cluster. We found that sFD can be mechanically unfolded by OT when stretched from its N-terminus and residue 93. When the mechanically unfolded sFD is relaxed, it mostly folds into an ensemble of misfolded states with much weaker mechanical stabilities comparing to that of the native state. In very rare cases, the refolded ferredoxin shows a correctly refolded β-sheet or a reconstituted coordination center, however, the whole native structure of sFD cannot be recovered even with extended folding time. Our results revealed the detailed folding-unfolding behavior of sFD at the single-molecule level, and highlighted the difficulties of recovering the native structure of sFD *in vitro*.

4.3 Results

4.3.1 The Mechanical Unfolding of sFD by Single-Molecule OT

To investigate the folding behavior of sFD by single-molecule OT, we first constructed a recombinant polyprotein Cys-NuG2-sFD-NuG2-Cys, which carries a cysteine at its both termini to facilitate the construction of dsDNA-protein-dsDNA for OT experiments. NuG2 is a computationally designed fast folding protein that is used here as a fingerprint domain for identifying the single-molecule stretching event, as its highly regular mechanical folding-unfolding behavior has been well characterized in our previous studies (unfolding at ~30-60 pN and folding at ~8 pN at 100 nm/s with a contour length change (ΔLc) of ~18 nm).[131] Stretching Cys-NuG2-sFD-NuG2-Cys by OT allowed us to stretch the sFD from its N- and C-termini (Figure 4.2A). In our previous AFM experiment, a ferredoxin with an identical folded structure was found to unfold in two steps with ΔLc of ~18 nm and ~13 nm, which correspond to the unfolding of the β-sheet and rupture of the coordination center, respectively.[28] A similar unfolding behavior was expected here for sFD stretched by OT. All the OT experiments were

carried out in Tris buffer containing 1M NaCl which has been proven to stabilize the structure of sFD.[253] Cycle 1 in Figure 4.2A is a representative stretching-relaxing cycle of Cys-NuG2-sFD-NuG2-Cys at a pulling speed of 100 nm/s, where the molecule was stretched to the force plateau of the DNA B-S transition and then relaxed. It is obvious that only the unfolding and folding events of the two NuG2 domains can be observed on the traces, and sFD, because of its high mechanical stability, was not unfolded during stretching. In order to unfold the mechanically stable sFD, we then stretched the molecule beyond the DNA B-S transition and held it at a very high force of ~80 pN for an extended period of time. However, as shown by Curve 2 in Figure 4.2B, the molecular tether mediated by noncovalent interactions tended to rupture before sFD unfolded. These results suggested that the mechanical resistance of sFD between its N- and C-termini is too high for OT experiments, therefore sFD cannot be unfolded efficiently in this way.

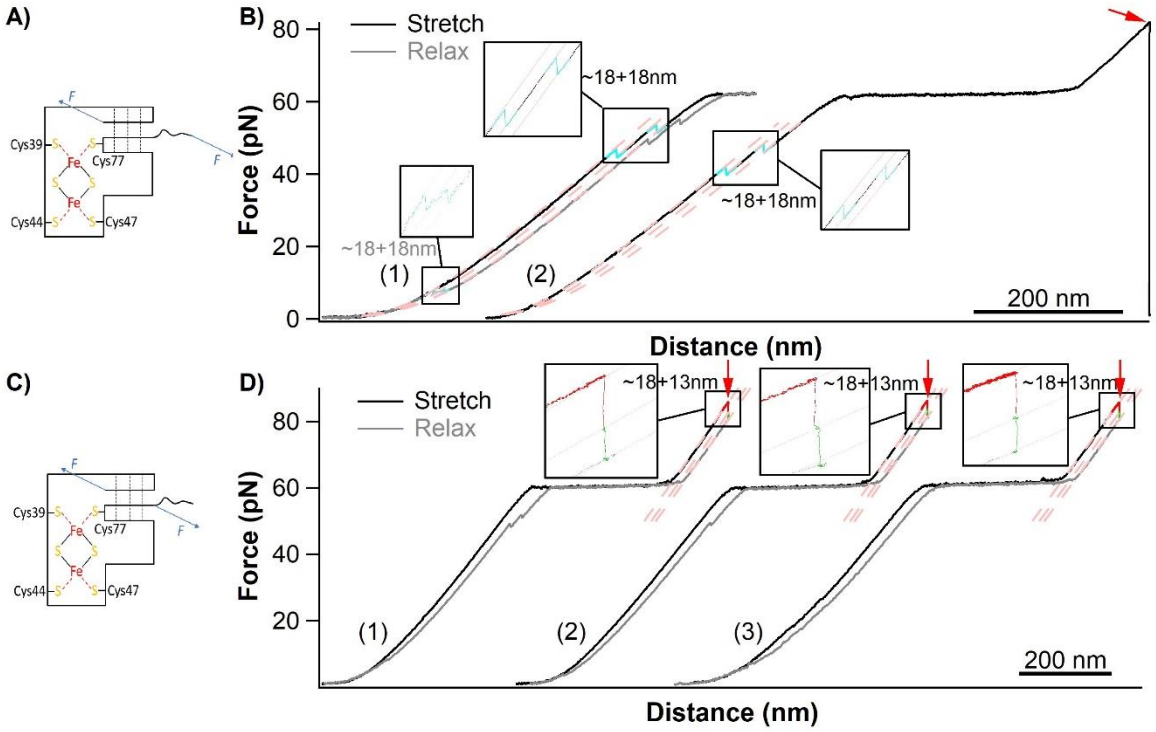

Figure 4.2 The mechanical unfolding of sFD with single-molecule OT. (A) Schematic of stretching sFD from its N- and C-termini. (B) Representative force-distance curves of cys-NuG2-sFD-NuG2-cys at a pulling speed of 100

nm/s. The red arrow indicates where the molecule was held for an extended period of time. The unfolding and folding of NuG2 are colored in cyan. Pink curves are pseudo-WLC fits to the force-distance data with ΔLc of ~18 nm in between. (C) Schematic of stretching sFD from its N-terminus and residue 93. (D) Representative force-distance curves of cys-sFD(E93C) at a pulling speed of 100 nm/s. The red arrow indicates where the molecule was held for an extended period of time. The unfolding of sFD is colored in red and green. Pink curves are pseudo-WLC fits to the force-distance data.

We then sought to move the DNA anchoring points to other positions, which may allow for unfolding the majority of the folded structure in sFD at a lower force. The residue 89-97 of sFD does not form significant folded structures but only a coil with a small helix outside of the β-sheet at the C-terminus. Our group previously found that, stretching sFD from its N-terminus and residue 93 could unfold the protein at ~180 pN (unpublished data), which is notably lower than the ~320 pN when a ferredoxin is stretched from its N- and C-termini (under a pulling speed of 400 nm/s by AFM). Therefore, we mutated the glutamic acid at residue position 93 to a cysteine and constructed a recombinant polyprotein Cys-sFD(E93C), which can let us stretch sFD from its N-terminus and residue 93 (Figure 4.2C).

Figure 4.2D shows three representative pairs of force-distance curves of Cys-sFD(E93C). The molecules were stretched beyond the DNA B-S transition and held at ~80 pN for extended periods of time.[137] After ~30s at ~80 pN, a force drop corresponding to a protein unfolding event was usually observed. Looking closely at these high-force unfolding events, a short-lived intermediate state can always be observed. Using a pseudo worm-like chain model (pseudo-WLC)[137] to fit the force-distance data, we found that the total ΔLc of the unfolding (from the native state to the unfolded state) is ~31 nm, and the intermediate state separates the unfolding into two steps with ΔLc of ~18 nm (from the native state to the intermediate state) and ~13 nm (form the intermediate state to the unfolded state), respectively. In theory, when sFD is stretched from its N-terminus and residue 93, unfolding the protein structure outside of the coordination center (residue 1-38 and 78-93) should lead to a ΔLc of ~17.0 nm ((38 + 16) aa × 0.36 nm/aa +

0.9 - 3.3 nm = 17.0 nm, where 0.36 nm/aa is the length of an aa residue, 0.9 nm is the distance between residue 39 and 77, and 3.3 nm is the distance between the N-terminus and residue 93). Rupturing the Fe-S bonds in the coordination center and extending the pre-sequestered protein sequence inside the coordination center (residue 39-77) should generate a ΔLc of ~12.8 nm ((77 - 39) aa × 0.36 nm/aa - 0.88 nm = 12.8 nm, where 0.88 nm is the distance between residue 39 and 77). These theoretical values of ΔLc are very close to the ~18 nm and ~13 nm that we observed in the experiment, suggesting that sFD was fully unfolded when it was stretched from its N-terminus and residue 93 by OT. sFD also followed the same mechanical unfolding mechanism as the ferredoxin stretched by AFM from its N- and C-termini, with its outer β-sheet being unfolded first and inner coordination center being ruptured subsequently.[28] The reason behind the lower mechanical stability of sFD when it is stretched from the N-terminus and residue 93 remains to be resolved. It is likely that changing the pulling sites leads to a subtly altered pulling direction in which sFD is mechanically weaker, or the glutamic acid at residue position 93 contributes to the stabilization of the β-sheet and the mutation disrupts some of the interactions.

4.3.2 The Unfolded sFD Mostly Misfolded into Non-Native Structures

Having successfully unfolded sFD by OT, we then set to investigate the folding mechanism of sFD. Mechanically unfolded ferredoxin was proven to possess a certain degree of folding competence, yet its detailed folding behavior is still mysterious.[28] To directly monitor the folding of unfolded sFD in real-time, once the sFD was unfolded, we relaxed the molecule back to 0 pN to trigger the folding process. As shown in Figure 4.3A, in most cases (354 out of the 419 folding processes), the mechanically unfolded sFD did show some folding events during relaxation (Cycle 1-3, Figure 4.3A), in the force range of ~5-10 pN. In rare cases (65 out of the

419 folding processes), the unfolded sFD did not show any observable folding event but a monotonic decreasing of force during relaxation (Cycle 4, Figure 4.3A). Looking closely at the folding events of sFD, we found the folding behavior of sFD quite complicated. The folding could happen in a two-state fashion or via different numbers of intermediate states (Figure 4.3A, inset), and the ΔLc of each folding event varied from ~5 to ~31 nm. Plotting the force against the ΔLc of each folding event, the data points on the figure scatter in a large range of value, and no cluster of data points, which may represent the folding of a specific structure, can be clearly identified (Figure 4.3B-D). These results implied that the mechanically unfolded sFD has some folding tendency, however, it does not have a well-defined folding pathway.

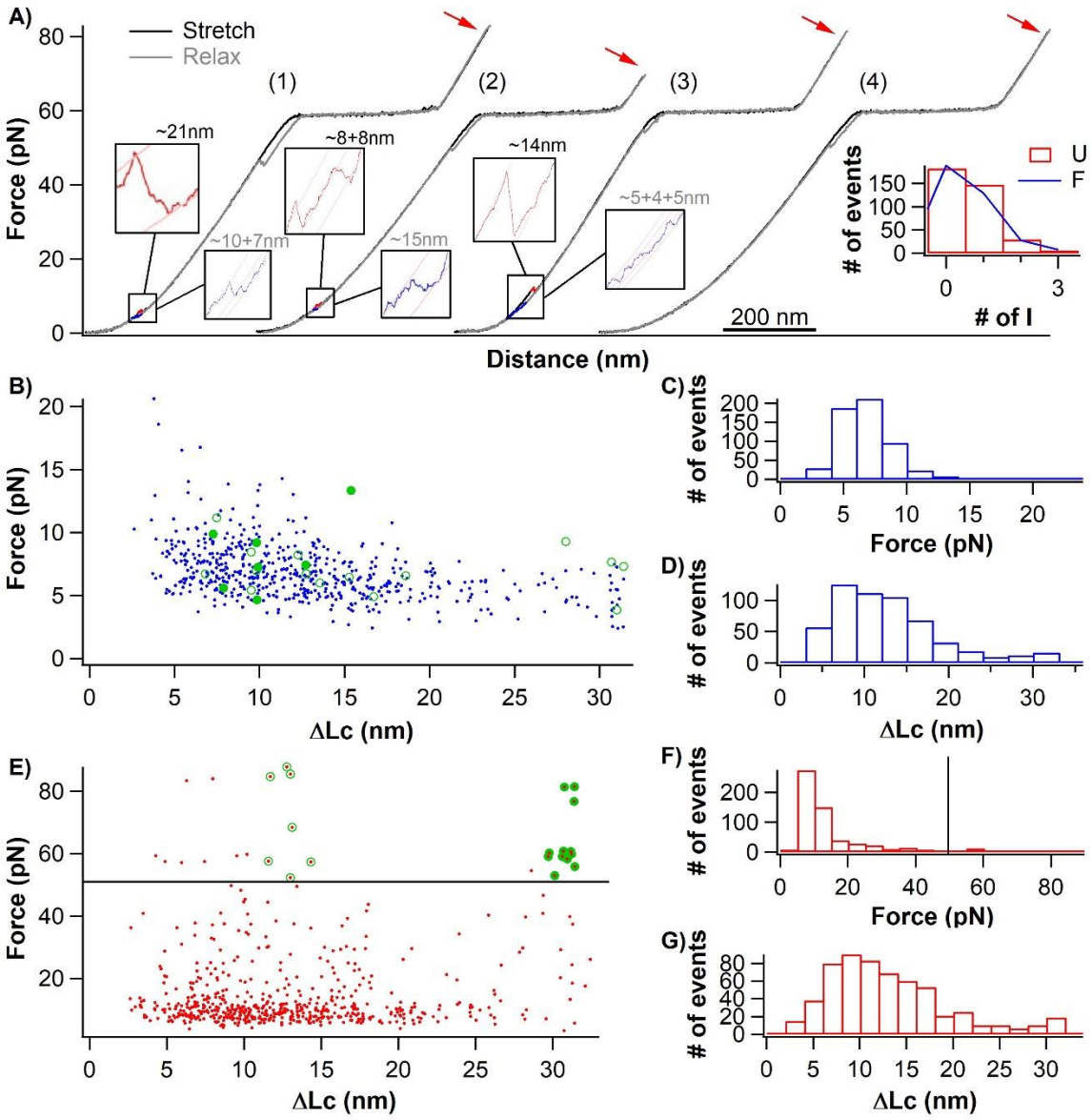

Figure 4.3 The folding-unfolding behavior of sFD after its initial unfolding. (A) Representative force-distance curves of cys-FD(E93C) after the initial unfolding of sFD at a pulling speed of 100 nm/s. The red arrow indicates where the molecule was held for an extended period of time. The unfolding of sFD is colored in red and the folding in blue. Pink curves are pseudo-WLC fits to the force-distance data. Inset: Histograms of the number of intermediates in the unfolding (red) and folding (blue) process of sFD (U: unfolding, F: folding, I: intermediate states). For clarity, only the unfolding histogram is shown as a bar chart. (B-D) Force-ΔLc relationships of each folding event of mechanically unfolded sFD, and their corresponding histograms. In (B), the green circled events led to possible refolded native structures (open circle: coordination center, filled circle: β-sheet). (E-G) Force-ΔLc relationships of each unfolding event of refolded sFD, and their corresponding histograms. In (E) and (F), the events below ~50 pN (indicated by the black line) happened during pulling before the DNA B-S transition, while those above happened at forces beyond DNA B-S transition or after holding the molecule for extended periods of time at high forces. In (E), the green circled events are likely the unfolding or rupture of refolded native structures (open circle: coordination center, filled circle: β-sheet).

To investigate whether such complicated folding events of unfolded sFD lead to recovery of the native structure of sFD, after relaxation, we stretched the refolded sFD again to high forces and monitored its unfolding behavior. If the whole native structure of sFD was successfully recovered in the previous folding process, the refolded sFD would unfold in the same way as it did initially, in two steps with ΔLcs of ~18 nm and ~13 nm at high forces. If only the coordination center or the β-sheet structure of sFD reformed, the refolded protein would unfold in a single step with ΔLc of ~13 nm or ~31 nm at high force. Looking at the unfolding events of refolded sFD (Cycle 1-3, Figure 4.3A), we found that it also happened in a complicated way involving different numbers of intermediate states (Figure 4.3A, inset). As illustrated in Figure 4.3E and F, while some unfolding events happened at relatively high forces, the majority of the refolded sFD unfolded during the constant speed pulling <~20 pN, which is significantly lower than the unfolding force of the initial native state, suggesting that most refolded structures were non-native, misfolded structures. In addition, for those low-force unfolding events, the ΔLc also scatters over a large range of value, which suggests a variety of different misfolded structures, and there is no cluster of data points on the unfolding force-ΔLc figure in the low-force regime, hinting that no specific folded structure frequently formed during refolding (Figure 4.3E). Altogether, these results indicated that, in most cases, mechanically unfolded sFD misfolded in a complicated way into an ensemble of non-native structures with much lower mechanical stability than its native state. In rare cases, the refolded sFD showed mechanically stability comparable to that of its native structure.

4.3.3 Possible Recovery of the Native Structure of sFD

While most of the refolded sFD molecules were mechanically unstable, we did observe some high-force unfolding events of refolded sFD, which happened either during the stretching

at forces above the DNA B-S transition, or after the molecule being held for extended periods of time at high forces (data points above the black line in Figure 4.3E). Some of those high-force unfolding events displayed a ΔLc of ~31 nm or ~13 nm (marked by closed and open circles in Figure 4.3E), which may potentially indicate structures with successfully refolded β-sheet or reconstituted coordination center. Cycle 1 and 2 in Figure 4.4A show two unfolding events with a ΔLc of ~31 nm that happened during the stretching at forces beyond the DNA B-S transition. These unfolding events happened in a two-state fashion without any intermediate state, which possibly implies a refolded sFD structure with only its β-sheet refolded but the coordination center still missing. Additionally, compared to the β-sheet in the native sFD structure which always requires some time at ~80 pN to be unfolded, the two unfolding events shown by Cycle 1 and 2 happened during stretching before the force reached ~80 pN, referring to a mechanical weaker β-sheet in the refolded sFD without the coordination center. Cycle 3 and 4 show the high-force unfolding events with a ΔLc of ~13 nm. As no other unfolding events happened prior to it, it can be inferred that the refolded sFD possibly possessed a reconstituted coordination center but no other structures outside of the coordination center. In the total of 419 refolding processes that we recorded based on 29 sFD molecules, the β-sheet was observed to successfully refold 12 times (success rate of ~2.9%) in 7 molecules, and the coordination center reconstituted for 7 times (success rate of ~1.7%) in 6 molecules (Figure 4.3E). Throughout our experiments, we did not find any case where the refolded sFD had both its β-sheet and coordination center successfully recovered. The closet case is the one illustrated by Cycle 5 in Figure 4.4A, where the unfolding event at ~80 pN confirms a reconstituted coordination center, and the unfolding at low forces suggests the formation of some folded structures outside of the coordination center, although not to the extent of forming the native β-sheet structure.

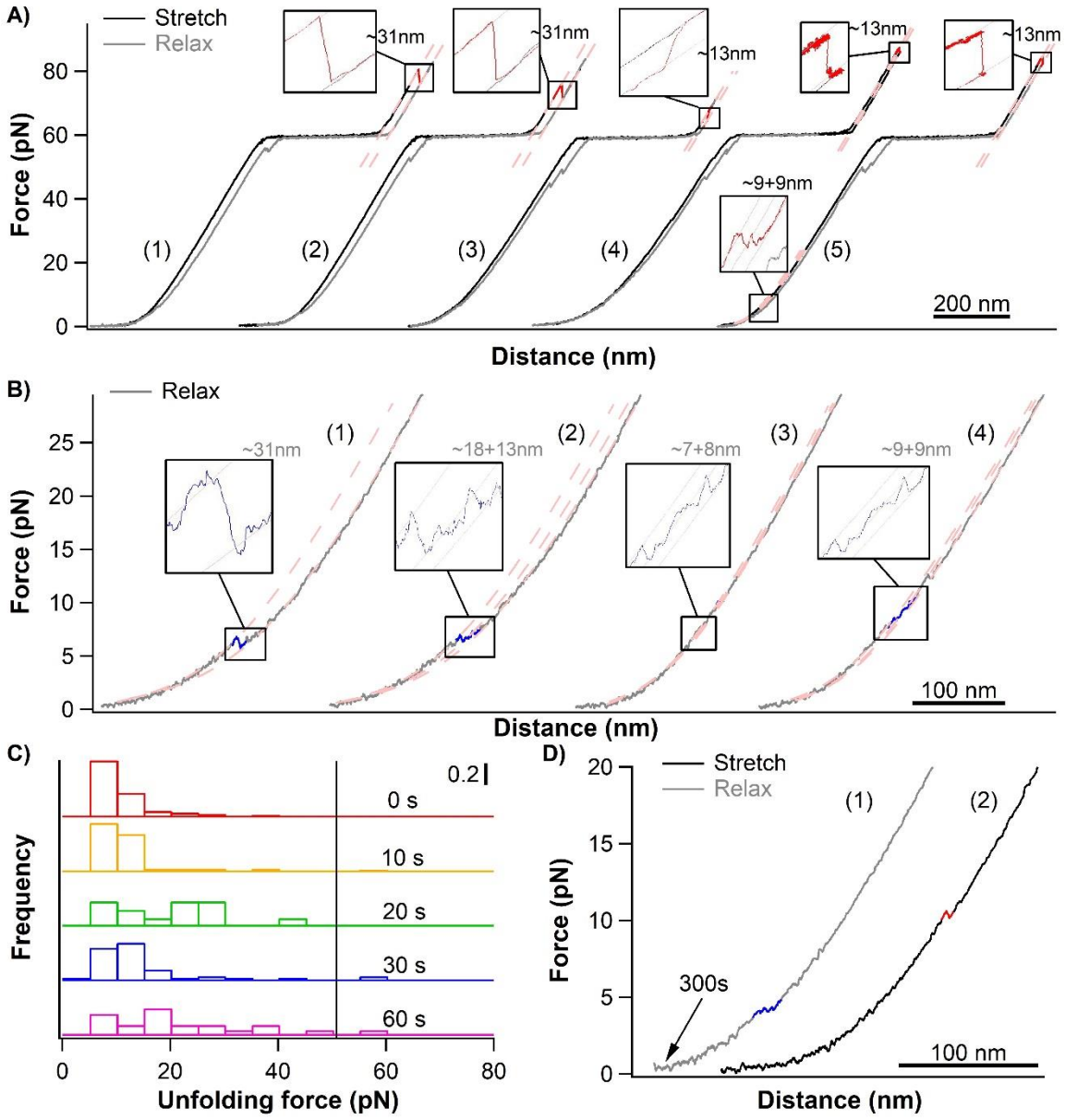

Figure 4.4 The recovery of the native structure of sFD. (A) Representative force-distance curves of cys-FD(E93C) with the unfolding of the possible recovered native structures at a pulling speed of 100 nm/s. (B) Representative force-distance curves of cys-FD(E93C) with folding events led to the possible recovery of native structures at a pulling speed of 100 nm/s. (C) Histograms of unfolding forces each unfolding event of refolded sFD after different folding times at 0 pN. The events below ~50 pN (indicated by the black line) happened during pulling before the DNA B-S transition, while those above happened at forces beyond DNA B-S transition or after holding the molecule for extended periods of time at high forces. (D) Force-distance curves of two consecutive relaxing and stretching processes of cys-FD(E93C) at a pulling speed of 100 nm/s with a folding time of 300s at 0 pN. For clarity, Curve 1 and 2 are horizontally offset relative to each other.

We then looked at the folding events of sFD that lead to either the folding of the β-sheet or the reconstitution of the coordination center, to see whether there is any specific common

folding behavior in these folding events. Figure 4.4B shows four representative refolding processes of these that lead to successful folding of the β-sheet (Curve 1-2) or reconstitution of the coordination center (Curve 3-4). We found that these folding events also happened without a well-defined pathway, via different numbers of intermediate states and with varied ΔLcs and forces (Figure 4.4B). Marking those folding events out in the force-ΔLc figure in Figure 4.3B (open circle for leading to coordination center reconstitution and filled circle for β-sheet folding), these events almost distribute evenly among all the data points, without providing any clue for a specific folding behavior. Additionally, we tried to extend the folding time at 0 pN to up to 300s, yet it did not significantly increase the success rate of recovering the native structure, and refolded sFD still mostly unfolded at relatively low forces (Figure 4.4C-D). Overall, these results demonstrated the possibilities as well as the difficulties of mechanically unfolded sFD to recover (part of) its native state.

4.4 Discussion

4.4.1 The *In Vitro* Folding Mechanism of sFD

Combining protein engineering and OT-based SMFS techniques, here we studied the folding-unfolding process of sFD *in vitro* and depicted the detailed folding behavior of sFD at the single-molecule level for the first time (Figure 4.5). Mechanically stable sFD cannot be unfolded by OT from its N- and C-termini due to its high mechanical stability, however, it can be unfolded by stretching from its N-terminus and residue 93 as a result of its mechanical anisotropy. Consistent to the observations in the previous AFM experiments, the mechanical unfolding of sFD occurs in two steps, with its outer four-stranded β-sheet being unfolded first and its inner coordination center being ruptured subsequently.[28] The unfolded sFD has a complicated folding behavior without well-defined folding pathways. It mostly misfolds in a

variety of seemingly random pathways into an ensemble of non-native structures with lower mechanical stabilities. Recovering the native structure of sFD *in vitro* in SMFS experiments is a possible but challenging process. Even with extended folding time at 0 pN, only in rare cases the unfolded sFD can recover its native β-sheet structure or reconstitute its coordination center. We did not find any case where mechanically unfolded sFD recovered its whole native structure. Our results clearly present the detailed folding-unfolding behavior of sFD at the single-molecule level, and demonstrate the possibility and difficulties for sFD to fold into its native structure *in vitro*.

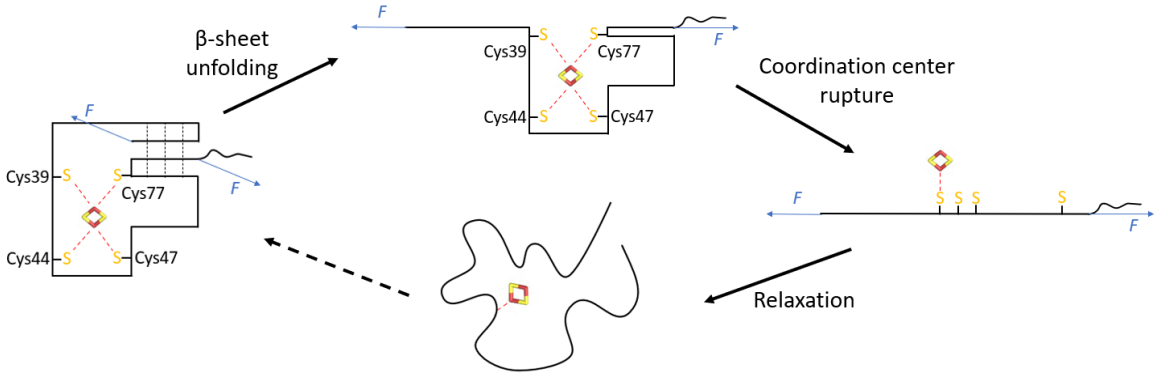

Figure 4.5 Schematic of the folding-unfolding pathway of sFD. sFD can be mechanically unfolded in two steps, with its β-sheet being unfolded first and its coordination center being ruptured subsequently. The mechanically unfolded sFD mostly folds into an ensemble of misfolded, non-native structures. The recovery of the native structure of sFD *in vitro* is challenging. The position on which the cluster is attached in the unfolded and misfolded sFD is chosen arbitrarily and only for the illustration purpose.

The conformations of sFD under different conditions have been studied by various spectroscopic studies previously. It was found that the [2Fe-2S] cluster in sFD plays a vital role in determining the native folded structure, by stabilizing not only the local structure near the coordination center but also the rest of the protein including the β-sheet.[254] Chemically removing the cluster from holo-form sFD leads to an apo-form sFD displaying largely disordered, molten-globule-like structures.[254] Enzymatic assembling and inserting the cluster back into sFD can drive the protein folding toward its native structure.[255] Our SMFS results corroborate the

disordered, molten-globule-like nature of apo-form sFD by revealing the seemingly random, complicated folding pathways of unfolded sFD, and identifying a large variety of mechanically weak, non-native folded structures without the reconstituted coordination center. Additionally, compared to the β-sheet in native sFD, the fact that the β-sheet in refolded sFD without the coordination center has relatively lower mechanical stability than that in the native sFD confirms the stabilizing effect of the coordination center on the protein structure. Besides, by examining the folding behavior of an unfolded sFD species with the cluster attached, which is not accessible under ensemble experimental conditions, our experiments further point out the difficulties of the reconstitution of the coordination center, as well as the folding of the rest of the structure subsequent to the reconstitution of the coordination center. Nevertheless, it is to be mentioned that the solvent-exposed iron-sulfur clusters on the unfolded polypeptide chains were found to have rates of degradation and dissociation on the level of 1×10^{-3}/s.[256-257] Since each molecule is usually stretched for minutes to tens of minutes in OT experiments, and the mechanically unfolded FD does not easily fold back to its native structure to encapsulate the cluster again, there is a potential possibility that the [2Fe-2S] cluster on the mechanically unfolded sFD degrades or dissociates during the experiments. This may result in even lower observed success rates of the reconstitution of the coordination center as well as the recovery of the whole native structure, comparing to those in AFM experiments where each molecule is mostly stretched for seconds only.[28]

4.4.2 Biological Implications on the Biogenesis of sFD *In Vivo*

The *in vivo* biogenesis of iron-sulfur proteins with clusters requires complex biosynthesis systems consisting of dedicated machineries and chaperones.[258] In general, the iron-sulfur cluster needs to be first synthesized *de novo* on a scaffold protein with the sulfurs provided by cysteine

desulfurases and the irons by iron donors, and then transferred to the target apo-form iron-sulfur protein and subsequently assembled into it. However, many processes have not been separated in experiments, and the conformational change of iron-sulfur proteins during the biogenesis process remains unclear.[186] The revealed *in vitro* folding behavior of sFD may provide some implications on the folding process of sFD during the biogenesis *in vivo*. For example, the challenging *in vitro* folding process of unfolded sFD implies that the newly synthesized apo-form ferredoxin *in vivo* may not easily fold. Instead, it may adopt some highly dynamic and flexible structures, which are better at accepting the iron-sulfur cluster delivered by the transfer protein than a rigid, folded structure with the cluster binding sites being enclosed inside of the folded protein. In addition, it was found that a dedicated chaperone system comprising the Hsp70 ATPase Ssq1 and the DnaJ-like co-chaperone Jac1 is evolved in the cluster transfer and assembly process.[259-261] While the details of the *in vivo* process are hard to obtain, it can be inferred that the folding of the ferredoxin is very likely to happen simultaneously under the assistance of these chaperones, as our SMFS experiments show that folding into the native structure is challenging for sFD by itself even with the coordination center reconstituted.

4.5 Conclusions

Using OT-based SMFS, we revealed the detailed folding-unfolding mechanism of sFD at the single-molecule level. sFD unfolds in a two-step fashion with its β-sheet being unfolded first and the coordination center being ruptured subsequently. The folding of sFD does not have a well-defined pathway. It mostly folds via complicated pathways into an ensemble of misfolded, non-native states. The recovery of the coordination center or the β-sheet during the folding of sFD is possible but very challenging, and mechanically unfolded sFD is unlikely to recover its whole native structure. Our results clearly demonstrate the difficulties for a complex iron-sulfur

protein to recover its native structure *in vitro*, and provide invaluable insights to our understanding of the folding behavior of metalloproteins in general.

4.6 Experimental Section

4.6.1 Protein Engineering

The gene encoding sFD was custom synthesized and designed to include a 5' BamHI (G'GATCC), a 3' BglII (A'GATCT) and a 3' KpnI (G'GTACC) restriction sites. (Genscript, Piscataway, NJ). The gene encoding the polyprotein Cys-NuG2-sFD-NuG2-Cys was constructed by digesting and ligating the insert of sFD into a modified pQE80L (Qiagen, Valencia, CA) expression vector, which allows for adding a cysteine and an NuG2 to both termini of the protein. The gene of sFD(E93C) was obtained by standard site-directed mutagenesis methods. The gene encoding the polyprotein Cys-sFD(E93C) was constructed by digesting and ligating the insert of sFD(E93C) into a modified pQE80L vector allowing for adding a cysteine to the N-terminus of the protein. All the sequences were confirmed by direct DNA sequencing.

Protein expression was performed using the same procedure as described in chapter 2 (2.6.1). Protein purification was performed using the same procedure as described in chapter 2 (2.6.1) but in Tris buffer with NaCl (100 mM Tris-HCl, 1 M NaCl, pH 7.4) instead of PBS buffer.

4.6.2 Preparation of DNA-Protein Chimera

DNA handles and DNA-protein chimera were prepared using the same procedure as described in chapter 3 (3.6.2) but in Tris buffer with 1M NaCl (100 mM Tris-HCl, 1 M NaCl, pH 7.4).

4.6.3 OT-Based SMFS Experiments

The single-molecule OT experiments were carried out in the same way as described in chapter 3 (3.6.3) but in Tris buffer with 1M NaCl (100 mM Tris-HCl, 1 M NaCl, pH 7.4). The stretching-relaxing experiments in this work were carried out at a constant speed of 100 nm/s.

Chapter 5: Single-Molecule Force Spectroscopy Revealed the Folding-Unfolding Mechanisms of Holo- and Apo-Form Cytochrome C: A New Understanding of the Role of Heme Cofactor in Folding

5.1 Synopsis

Metalloproteins account for over one-third of all proteins in nature and play important roles in biological processes. The formation of the native structures of metalloproteins requires not only the correct folding of the polypeptide chains, but also the proper incorporation of metal cofactors, which makes the folding process even more complicated. Heme proteins are one of the most common metalloprotein families in nature and are highly versatile in their biological roles. Among them, cytochrome c (cytc), which contains a c-type heme cofactor, acts as an important electron transfer protein involved in both aerobic and anaerobic respiration processes. The folding mechanism of horse heart cytc has been extensively studied by ensemble spectroscopic methods. It was found that the holo-form (heme-bound) cytc (holo-cytc) folded into a helical structure, and the apo-form (heme-free) cytc (apo-cytc) showed no spectroscopic signals typical of folded structures. The heme cofactor was thus believed to be decisive for the folding of cytc. Here, we used single-molecule optical tweezers to probe the mechanical folding-unfolding behaviors of holo-cytc and apo-cytc at the single-molecule level. Our results suggested that, holo-cytc demonstrates the typical folding-unfolding behaviors of a folded protein. Apo-cytc, which had been long been believed to be random coiled, intriguingly displays some intrachain interactions and may fold into an ensemble of molten globule states. Therefore, the interactions between the heme cofactor and the polypeptide chain can facilitate the polypeptide chain folding and guide the protein to fold into the native state. Our results unambiguously demonstrated the

mechanical folding-unfolding behaviors of holo- and apo-cytc, and brought new insights to our understanding of the folding mechanisms of heme proteins as well as the role of heme cofactors in the folding process.

5.2 Introduction

Metalloproteins account for more than one-third of all the proteins in nature and play a variety of important biological functions, ranging from electron transfer to catalyzing some of the most difficult chemical reactions that are out of reach of the current synthetic chemistry. The folding of metalloprotein involves not only the folding of the polypeptide chain, but also the constitution of the metal center through metal-ligand coordination.[147] Elucidating the folding mechanisms of metalloproteins has been challenging, as the unfolding of many metalloproteins is often irreversible *in vitro*. This is particularly true for transition metal-containing metalloproteins of which the metal or metal cluster cofactors have high hydrolysis constants.[262]

Heme proteins are an important family of metalloproteins that contain a prosthetic heme group and play critical roles in many important cellular processes, such as oxygen transportation, drug detoxification and signal transduction.[263] The cofactor heme, an Fe(II)-centered porphyrin, entails heme proteins with distinct spectroscopic characteristics. Among the heme proteins, horse heart cytochrome c (cytc) has long been used as a model system for protein folding study.[264] As an important electron transfer protein, the 104 amino acid (aa) residue protein is highly helical, and contains a highly soluble c-type heme cofactor bounded to the polypeptide chain by two thioether bonds and two coordination bonds (Figure 5.1).[265] These unique features prevent the metal cofactor from hydrolysis and dissociation upon protein unfolding, making the *in vitro* refolding studies possible. The folding-unfolding mechanism of holo-form cytc (holo-cytc) has been studied extensively using various experimental techniques at the ensemble level. Based on

hydrogen exchange experimental results, a foldon-dependent hierarchical multistep folding-unfolding mechanism was proposed for the folding of holo-cytc.[174, 266-267] According to this theory, the protein is divided into 5 structural subunits called foldons, which are small, cooperative folding units containing ~20 amino acids. These foldons fold and unfold sequentially in a strict order, resulting in multiple partially folded intermediate states along the folding-unfolding pathway.[174, 266-267] In contrast, the heme-free apo-form cytc (apo-cytc) has long been considered as a random coil with no folded structures based on its various properties, including sedimentation coefficient, intrinsic viscosity, CD, histidyl residues' reactivity, etc.[268-269] It is believed that the interactions between the cytc polypeptide chain and the heme play a decisive role in the folding and stabilization of the globular conformation of cytc. The flexible and disordered conformation of apo-cytc is believed to facilitate the accepting and enveloping of the heme cofactor during the folding of holo-cytc.[268-269] Despite these progresses, the nature of the folding mechanism of holo-cytc is still under debate, especially in light of the statistical mechanical view of the folding mechanism. In addition, the random coil conformation of the apo-cytc remains to be substantiated.

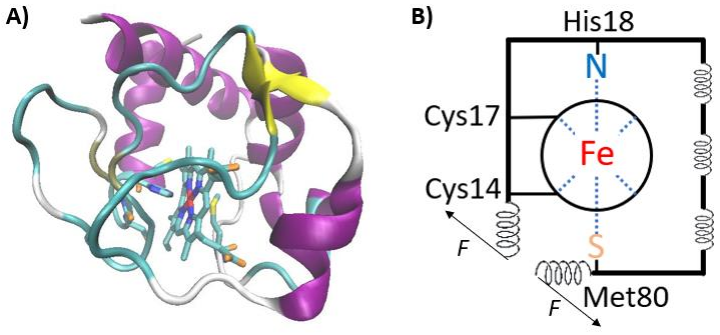

Figure 5.1 Structure of horse heart cytc. (A) Three-dimensional structure of horse heart cytc (PDB code: 1hrc). Cytc is a 104 aa helical protein containing a c-type heme cofactor. The heme is covalently bound to the polypeptide chain through two thioether bonds, and the heme iron forms axial Fe-N and Fe-S coordination bonds with a histidine and a methionine residue. (B) Schematic of cytc. The black circle indicates the porphyrin ring, the spirals indicate the helices, and the blue dotted lines represent the coordination bonds.

To address these questions, here we used OT-based SMFS technique, which has been used extensively to investigate the protein folding mechanism due to its superb force and distance resolution at low forces, to investigate the folding-unfolding of holo- and apo-cytc.[23, 28-29, 34, 37, 246, 270] We found that holo-cytc is mechanically stable and unfolds in either a two-state fashion, or a three-state fashion involving an intermediate with an intact coordination center and unfolded outer structure. In contrast, the folding of holo-cytc appeared to follow a two-state fashion without the accumulation of any intermediate state. Moreover, our results showed that apo-cytc demonstrates some intrachain interactions and may form a range of molten globule states with different structures that exhibit low mechanical resistance. Our results revealed some new insights into the conformation of apo-cytc and the folding-unfolding mechanism of holo-cytc, and help unfold the role played by the heme in the folding of cytc.

5.3 Results

5.3.1 Mechanical Unfolding of Holo-Cytc by Optical Tweezers

To investigate the unfolding and folding of holo-cytc using OT, we constructed a recombinant polyprotein Cys-holo-cytc-NuG2-Cys, which carries a cysteine at both termini to facilitate the construction of dsDNA-protein-dsDNA for OT experiments. NuG2 is a computationally designed fast folding protein and its mechanical folding-unfolding has been well characterized in our previous studies.[131, 271] The unfolding of NuG2 occurs at ~20-40 pN and folding occurs at ~8 pN with a contour length change ΔLc of ~18 nm, making it an ideal fingerprint domain for identifying the single-molecule stretching event in both OT and AFM experiments.[131] The pBTR vector encodes the gene of yeast heme lyase, allowing for the expressed Cys-holo-cytc-NuG2-Cys with a properly assembled heme prosthetic group.[272] Figure 5.2A shows the UV-Vis spectrum of the purified reduced form of Cys-holo-cytc-NuG2-Cys, as

well as the control holo-cytc. The distinctive absorbance at ~414, ~521, ~550 nm is characteristic of the holo-cytc. Based on the ratio of the absorbance at 410 nm and 280 nm,[273] we estimated that more than 70% of our Cys-cytc-NuG2-Cys sample contained holo-Cytc.

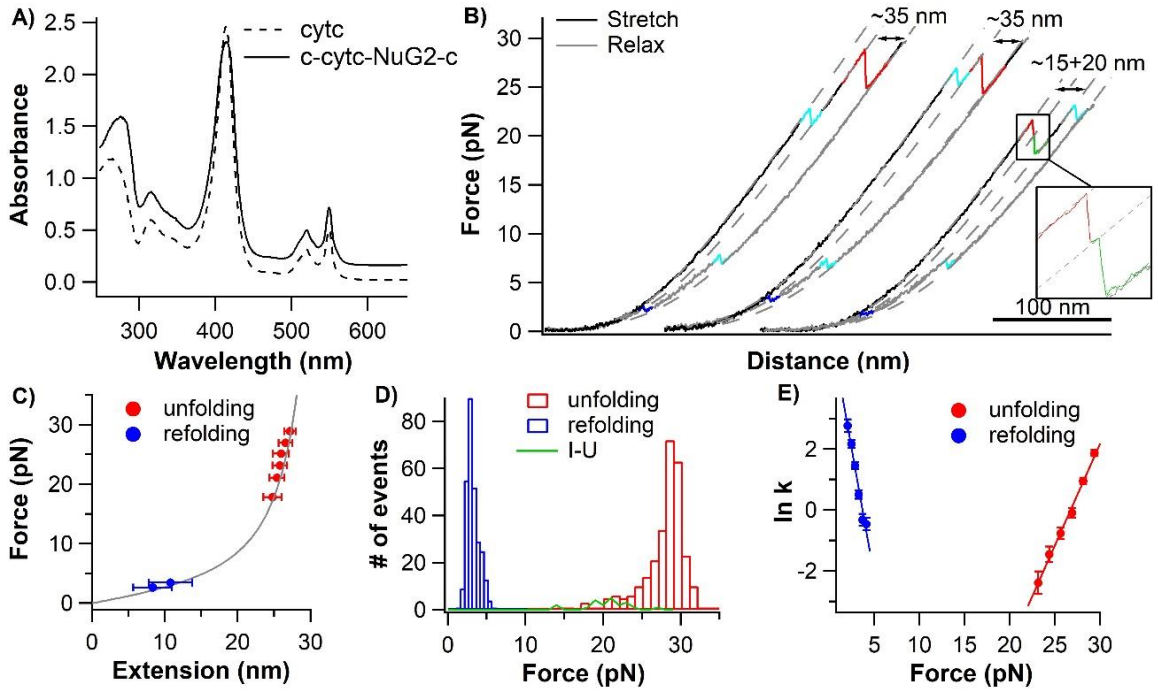

Figure 5.2 Mechanical folding-unfolding signatures of holo-cytc. (A) UV-Vis spectra of purified holo-cytc and Cys-holo-cytc-NuG2-Cys protein expressed using modified pBTR vector. The absorbance at ~410, ~520, ~550 nm is characteristic of the holo-cytc. (B) Representative force-distance curves of Cys-holo-cytc-NuG2-Cys at a pulling speed of 50 nm/s. The unfolding of holo-cytc is colored in red and green, the folding of holo-cytc is colored in blue, and the unfolding and folding of NuG2 are colored in cyan. An unfolding intermediate can be observed when holo-cytc unfolds at relatively low forces, while the folding is always two-state. (C) Force-extension relationships of unfolding and folding of holo-cytc (red and blue). WLC fit to the experimental data reveals a persistence length of 0.8 nm and a ΔLc of 34.6 ± 0.1 nm (red curve) between the folded and unfolded states of holo-form cytc. (D) Unfolding (red and green) and folding (blue) force histograms of holo-cytc at a pulling speed of 50 nm/s. The I-U (intermediate state to unfolded state) event (green) can only be observed when holo-cytc unfolds at relatively low forces (<~22 pN). The bin size is 1 pN for both unfolding histograms, and for clarity, the I-U unfolding data is shown as a line chart. (E) Force-dependent folding-unfolding rates for holo-cytc. Solid lines are the fits of Bell-Evans model to the experimental data with the kinetic parameters: spontaneous unfolding rate constant α_0 = (1.83 ± 1.13) × 10^{-8} s^{-1}, unfolding distance Δx_u = 2.7 ± 0.1 nm; spontaneous folding rate constant β_0 = (4.94 ± 2.00) × 102 s^{-1}, folding distance Δx_f = 6.9 ± 0.5 nm.

For OT experiments, we coupled two dsDNA handles to Cys-cytc-NuG2-Cys via thiol-maleimide coupling chemistry to create the DNA-protein-DNA chimera. Stretching Cys-holo-cytc-NuG2-Cys allowed us to stretch the reduced holo-cytc from its N- and C- termini direction.

Figure 5.2B shows the representative force-distance curves of the protein-DNA chimera at a pulling speed of 50 nm/s. During stretching and relaxation, holo-cytc displayed mechanical unfolding and folding events that are typical to folded proteins. The unfolding of cytc mostly happened between ~25-30 pN and displayed two parallel pathways: ~92% (258/279) of the unfolding occurred in a two-state fashion, and ~8% (21/279) via one short-lived intermediate state (Figure 5.2B). Fitting the force-extension relationships of holo-cytc using the worm-like chain model (WLC) of polymer elasticity yielded a contour length increment (ΔLc) of ~34.6 nm for the complete unfolding (from the native state to the unfolded state) of holo-cytc (Figure 5.2C).[137] In the unfolding events following the three-state pathway, cytc displayed ΔLc1 of ~15 nm (from the native to the intermediate state), and ΔLc2 of ~20 nm (from the intermediate state to the unfolded state).

The complete mechanical unfolding of holo-cytc should lead to a ΔLc of ~35.7 nm (104 aa × 0.36 nm/aa - 1.7 nm = 35.7 nm, where 0.36 nm/aa is the length of an aa residue and 1.7 nm is the distance between the N- and C-termini). This value is very close to the experimentally determined ΔLc of ~34.6 nm, confirming that the unfolding events we observed indeed correspond to the complete unfolding of holo-cytc.

We then tried to resolve the structure of the intermediate state in the two-step unfolding pathway. Looking at the structure of holo-cytc, we noticed that the heme coordination center sequesters the residue 14-80 by the covalent bonds and coordination bonds, leaving the residue 1-14 and 80-104 outside of the coordination center to form outer structures. The force on the N- and C-termini of holo-cytc should apply to the outer structures first and then the coordination center. In theory, the unfolding of the outer structure (residue 1-14 and 80-104) leads to a ΔLc of ~13.4 nm ((14 + 25) aa × 0.36 nm/aa + 1.1 nm - 1.7 nm = 13.4 nm, where 1.1 nm is the distance

between residue 14 and 80, and 1.7 nm is the distance between the N- and C-termini of holo-cytc), which is very close to the ΔLc of ~15 nm for the unfolding from the native state to the intermediate state that we observed in the experiments (Figure 5.2B). This suggests that in the two-step unfolding pathway, the structure outside of the metal center in holo-cytc unfolded first, resulting in an intermediate state with unfolded outer structure and an intact metal center, and the Fe-S(Met80) bond was ruptured subsequently, followed by the unfolding and extending of the protein structure inside of the metal center. The Fe-S(Met80) bond could only survive for a very short time at the forces where the unfolding of holo-cytc happened, resulting in a short-lived intermediate state.

5.3.2 Folding Behavior of Holo-Cytc

The mechanically unfolded holo-cytc consists of a fully extended polypeptide chain and a covalently attached heme cofactor. This makes the *in vitro* folding of holo-cytc possible, as the metal cofactor does not hydrolyze or dissociate up protein unfolding. After unfolding, we relaxed the mechanically unfolded holo-cytc back to 0 pN under a constant speed, in order to monitor the folding process of holo-cytc. In most cases, the unfolded holo-cytc folded via a 2-state process at ~3 pN with a ΔLc of ~35 nm (Figure 5.2B, D). Following the refolding, in most cases the subsequent unfolding of the happening again at ~20-30 pN with a ΔLc of ~35 nm, in the same way as the pristine holo-cytc unfolded, suggesting the mechanically unfolded holo-cytc was able to refold back to its native state. To obtain the spontaneous unfolding/folding rate constants and unfolding/folding distances, we measured the two-state unfolding and folding rate constants as a function of force using the Oesterhelt method, and fitted the experimental data with the Bell-Evans model (Figure 5.2E).[95, 140]

However, the refolding of holo-cyt was not always successful. In some cases, the unfolded holo-cytc was observed to either fold into a structure with a much weaker mechanical stability (Figure 5.3A, Cycle 1) or with a shorter unfolding ΔLc (Figure 5.3A, Cycle 2), and in some other cases, no folding events of unfolded holo-cytc could be observed during the relaxation process (Figure 5.3A, Cycle 3). These phenomena indicated that, besides folding back into its native state, unfolded holo-cytc may also misfold into non-native structures or do not fold at all during the relaxation. Plotting the unfolding force against the cycle number of a holo-cytc molecule during consecutive stretching-relaxing cycles (Figure 5.3B), we found that the unfolding events at ~28 pN (indicative of the native structure) can happen after the unfolding events at <15 pN (indicative of misfolded structures), suggesting that the misfolding of holo-cytc is not an irreversible process. Similarly, not being able to fold in a few cycles does not mean the unfolded holo-cytc permanently loses its folding competence. As shown in Figure 5.3C, Curve 1 shows no unfolding event of holo-cytc, indicating that the molecule did not fold before that. But after the relaxation shown by Curve 2, there is an unfolding event of holo-cytc on Curve 3, suggesting that the unfolded holo-cytc successfully folded back into its native structure again. It was also found that, after holo-cytc started to misfold or did not fold, extending the folding time at 0 pN to up to one minute does not significantly assist the molecule to fold back to its native state (Figure 5.3D).

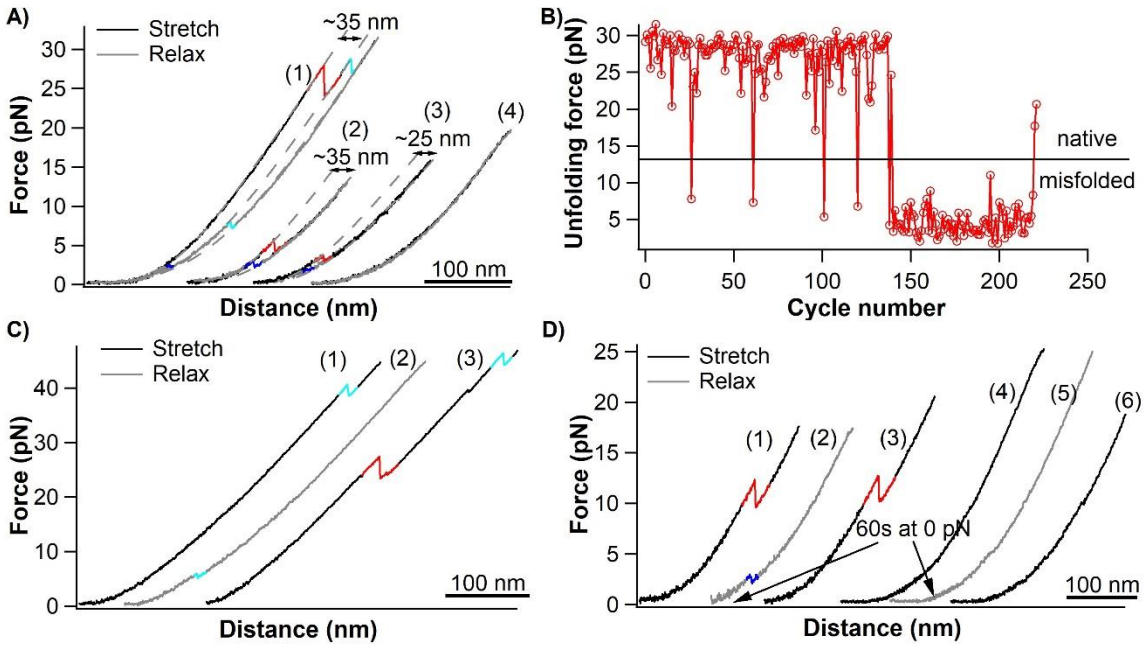

Figure 5.3 Unfolded holo-cyt may misfold or do not fold during consecutive stretching-relaxing cycles. (A) Representative force-distance curves of Cys-holo-cytc-NuG2-Cys that misfolds (Cycle 2-3) and does not fold (Cycle 4) at a pulling speed of 50 nm/s (cyan events: unfolding and folding of NuG2, red and blue events: unfolding and folding of holo-cytc). Cycle 1 shows an unfolding event of native holo-cytc for comparison. (B) Unfolding force of one holo-cytc molecules in consecutive stretching-relaxing cycles. The black line is drawn artificially to separate the unfolding events of natively folded (above the line) and misfolded (below the line) holo-cytc. (C) Three consecutive force-distance curves of a Cys-holo-cytc-NuG2-Cys molecule at a pulling speed of 50 nm/s (cyan events: unfolding and folding of NuG2, red and blue events: unfolding and folding of holo-cytc). Curves are offset horizontally relative to each other. (D) Representative force-distance curves of two Cys-holo-cytc-NuG2-Cys molecules which misfolded (curve 1-3) and did not fold (curve 4-6) at a pulling speed of 50 nm/s (red and blue events: unfolding and folding of holo-cytc). The molecules were held at 0 pN for 60s after relaxation. Curves are offset horizontally relative to each other.

5.3.3 Mechanical Response of Apo-Cytc

To further elucidate the role heme plays in the folding process of holo-cytc, we then set to investigate the folding and conformation of apo-cytc without the heme cofactor (in other words, the polypeptide chain of holo-cytc). We first cloned the gene of cytc alone into the pQE80L vector and expressed it. As no yeast heme lyase was co-expressed in this case, the heme moiety cannot be covalently attached to the protein, and the obtained protein should be apo-cytc. Indeed, the UV-Vis spectrum of the cytc expressed with pQE80L vector lacked the characteristic metal-ligand charge transfer absorbance peaks from the incorporation of heme at 414, 521 and

550 nm, confirming that the as-prepared protein was apo-cytc. In addition, we used CD spectroscopy to characterize its secondary structure. As shown in Figure 5.4B, the CD spectrum indicated that apo-cytc lacked secondary structures, consistent with what was reported in previous studies.[269]

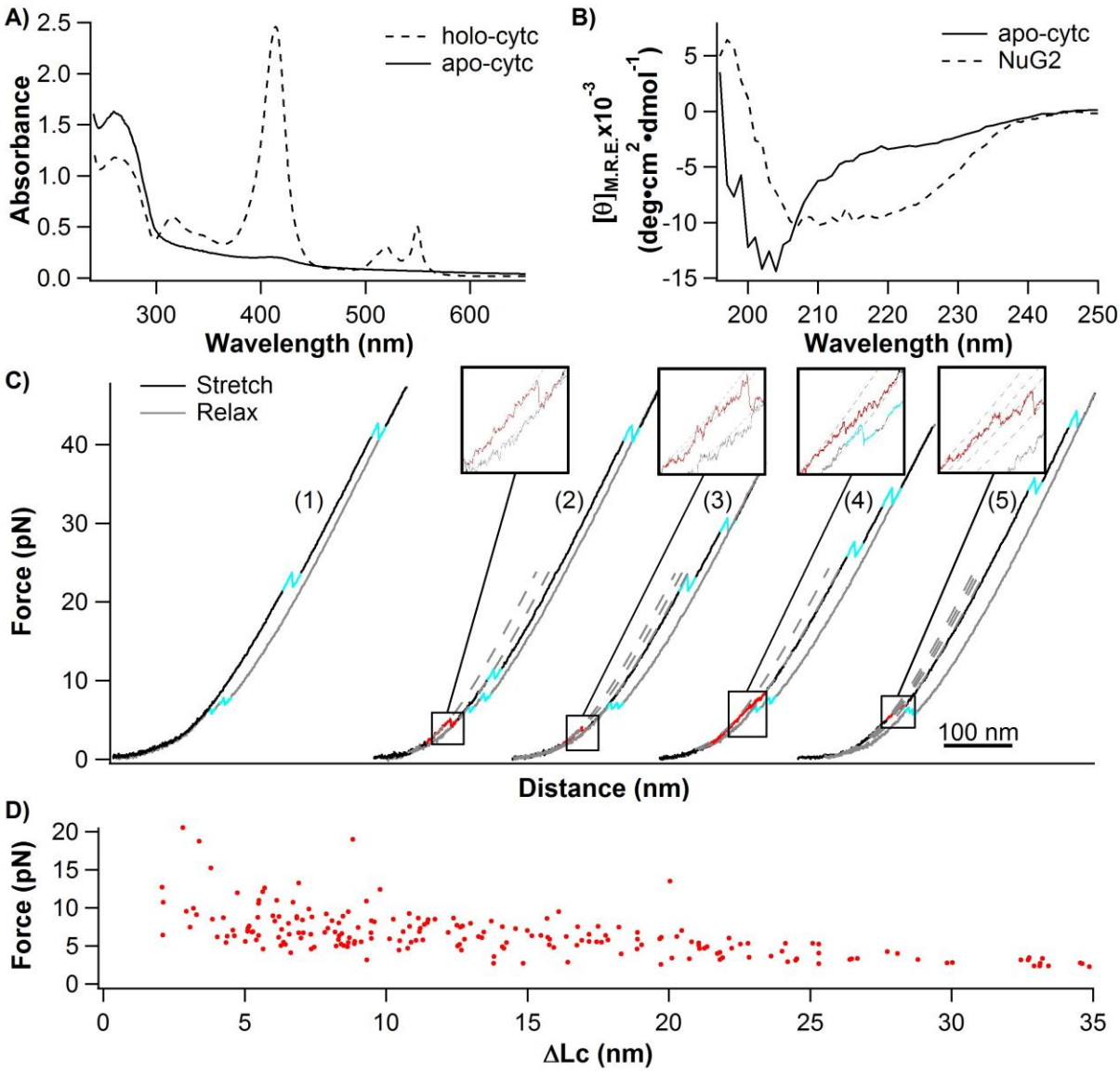

Figure 5.4 Mechanical response of apo-cytc. (A) UV-Vis spectrum of cytc expressed with pQE80L vector. The protein lacks the characteristic absorbance peak from the incorporation of heme and is in its apo-form. (B) CD spectrum of the apo-cytc. Apo-cytc shows a random coil-like CD feature in sharp contrast to that of a globular protein NuG2. (C) Representative force-distance curves of Cys-NuG2-apo-cytc-NuG2-Cys at a pulling speed of 50 nm/s (cyan events: unfolding and folding of NuG2, red events: rupturing of apo-cytc). ~77% of the cycles contain

only unfolding and folding events of 2 NuG2 domains (Cycle 1), and ~23% of the cycles show rupturing events of apo-cytc (Cycle 2-5). (D) Force-ΔLc relationship of individual rupture events of apo-cytc.

We then used OT to examine the mechanical response of apo-cytc to further examine this seemingly unstructured random coil. To do that, we first mutated the cysteine 14 and 17 in the gene of cytc to alanine, so that there would not be any free cysteine in the apo-cytc protein that may react with the maleimide group on the DNA handles and cause uncertainty of the positions from where the molecule is stretched. We then built a recombinant plasmid of Cys-NuG2-cytc(C14, 17A)-NuG2-Cys in the pQE80L vector, where the cytc(C14, 17A) was flanked by two NuG2 domains. A polyprotein Cys-NuG2-apo-cytc-NuG2-Cys can be expressed with this plasmid and used in our OT experiments, and force-extension curves displaying two unambiguous unfolding/folding events of the NuG2 fingerprint domains can be confirmed as containing an apo-cytc molecule.

Previous SMFS studies showed that stretching a true random coil resulted in monotonic rising force-extension curves, no “unfolding” or “folding”-like events were present. For many globular proteins studied in SMFS experiments, their unfolded polypeptide chains indeed behave like true random coils. In fact, in our OT experiments, ~77% (325/423) of the stretching-relaxing cycles of single Cys-NuG2-apo-cytc-NuG2-Cys molecules showed only the unfolding and folding events of two NuG2 domains (Figure 5.4C, Cycle 1), suggesting that apo-cytc did behave like a random coil in these stretching-relaxing cycles with no detectable intrachain interactions along the polypeptide chain at all. However, intriguingly, in the other 23% (98/423) of the cycles, some “unfolding”-like rupturing events can be visible mostly at forces below ~10 pN (Figure 5.4C, Cycle 2-5), in addition to the unfolding and folding events of two NuG2 domains. This suggests that apo-cytc, instead of being completely a random coil, sometimes demonstrates some intrachain interactions and may form some non-local structures. Moreover, these low-force

rupturing events were observed in all the 11 apo-cytc molecules that we captured, indicating that such deviation from a random coil is a general behavior among all apo-cytc molecules. Looking closely at the force region where these rupturing events happened, we found that the rupturing process happens in a seemingly random way with different ΔLc and behavior each time, and many different rupturing behaviors were observed, including one-step rupturing (Figure 5.4C, Cycle 2), fluctuation (Figure 5.4C, Cycle 3), hump-like rupturing (Figure 5.4C, Cycle 4) and multiple-step rupturing (Figure 5.4C, Cycle 5), etc. Plotting the force against the ΔLc of each rupturing event (Figure 5.4D), the data points on the figure scatter in a large range of value, and no cluster of data points, which may represent the rupturing of a specific interaction, can be clearly identified. This indicates that, although there are some subtle intrachain interactions, the amino acid residues in apo-cytc do not have a fixed interacting mode, which is consistent to the previous observations from ensemble experiments that apo-cytc lacks a folded structure.[269] Nevertheless, the formation of such interactions was barely visible as force-jump events during relaxation, implying that the weak intrachain interactions can only function at very low forces when the molecule is largely relaxed.

5.4 Discussion

5.4.1 The Folding-Unfolding Mechanism of Holo-Cytc at the Single-Molecule Level

By combining single-molecule OT and protein engineering techniques, here we investigated the mechanical folding-unfolding behavior of a heme-containing metalloprotein, horse heart cytochrome c. Holo-cytc stretched from its N- and C-termini unfolds in two paralleled pathways, either in one step from its native state directly to a fully unfolded state, or via an intermediate state containing an unfolded outer structure but an intact metal center and the Fe-S(Met80) bond. In most cases, the mechanically unfolded holo-cytc can fold back to its native

state in a two-state fashion, while occasional events of misfolding or not being able to fold were also observed (Figure 5.5). Our results clearly demonstrate the folding-unfolding mechanism of holo-cytc at the single-molecule level, which deepens our understanding of the folding behavior of heme-containing proteins as well as metalloproteins in general.

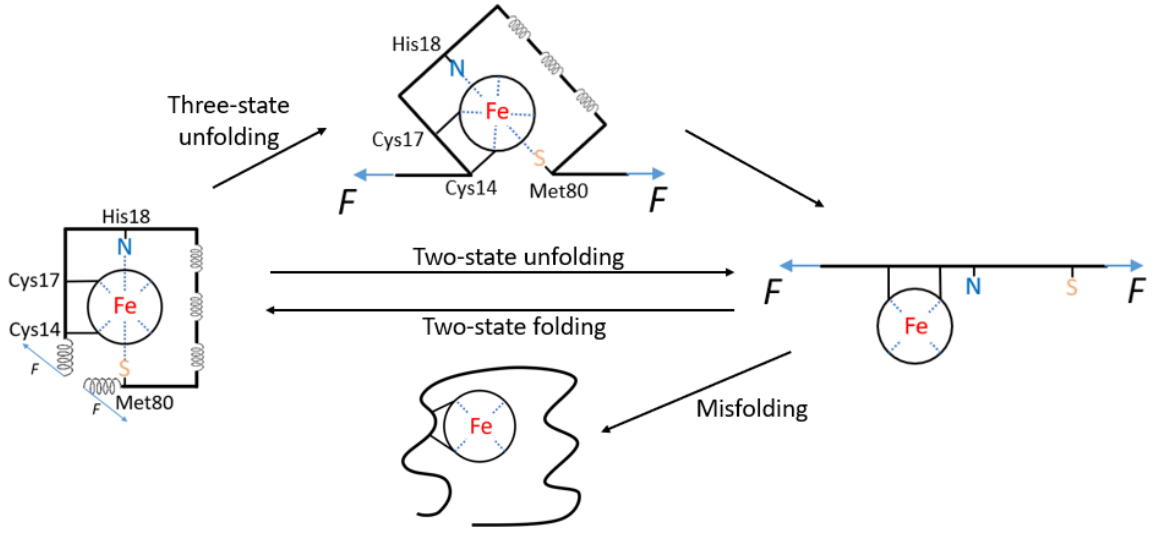

Figure 5.5 Schematic of the folding-unfolding mechanism of holo-cytc. Holo-cytc unfolds in either a two-state fashion, or a three-state fashion, with the outer protein structure being unfolded first and the coordination center being rupture subsequently. Mechanically unfolded holo-cytc mostly folds back to its native state in a two-state fashion, but misfolds or does not fold occasionally.

In our previous studies, a general mechanical unfolding mechanism was proposed for the iron-sulfur protein family based on the revealed unfolding behaviors of rubredoxin, ferredoxin and high potential iron-sulfur protein.[23, 28, 246] As the metal-coordination bonds (mostly Fe-S(Cys) and Fe-N(His) bonds) in iron-sulfur proteins are usually mechanically stable, the unfolding of these proteins usually proceeds via an intermediate state, which consists of unfolded protein structure outside of the coordination center but an intact coordination center. In the case of holo-cytc, the three-state unfolding pathway is in fact very similar to the unfolding pathway we proposed for iron-sulfur proteins, with its outer protein structure (1-14 and 80-104) being unfolded first, and its inner coordination center (the Fe-S(Met) bond) being ruptured

subsequently. However, probably due to the fact that the Fe-S(Met) bond in the coordination center of holo-cytc is not as mechanically stable as the Fe-S(Cys) bonds in iron-sulfur proteins, the unfolding of holo-cytc proceeds in a mixture of two-state and three-state pathways. Besides, compared to the rupture of the coordination centers in iron-sulfur proteins, where the tension is usually applied onto more than one identical coordination bond so that which exact bond is ruptured is uncertain[23, 28, 246], the rupture of the coordination center in holo-cytc always occur at the rupture of the Fe-S(Met80) bond, providing an explicit mechanical characterization of the rupture process of the Fe-S(Met80) bond as well as the coordination center in holo-cytc.

In terms of the occasional events where mechanically unfolded holo-cytc misfolds or does not fold, the reason probably lies in some subtle residual structures in the unfolded state that are not sensitive to force in the force range of OT, such as the conformation of proline residues (cytc contains four proline residues). Previous ensemble studies also suggested that proline isomerization and heme mis-ligation may limit the folding success rate of holo-cytc.[174]

5.4.2 Intrachain Interactions in Apo-Cytc, Which Is Long Believed as a Random Coil

In our SMFS experiments, apo-cytc, which has been long considered as a random coil, intriguingly demonstrates some subtle nonlocal intrachain interactions, and may fold into an ensemble of different conformations with low mechanical stabilities. Our results help us discover the deviation of random coil behavior of apo-cytc which was missed or overlooked in previous spectroscopic measurements. Random coil describes a protein conformation with its polypeptide backbone randomly coils in the three-dimensional space, and the backbone conformation of every aa residue independent of the conformations of neighboring residues, such as that of intrinsically disordered proteins and denatured proteins.[274-276] While a variety of spectroscopic techniques have been developed to characterize random coils, each of these measurements only

targets one specific aspect of the protein conformation (such as radius of gyration, longest linear dimension, or secondary structure). As a result, potential subtle folded structures may be omitted by one or even combined spectroscopic measurements, leading to mischaracterizing of random coils.[277-278] For example, the formation of hydrophobic clusters in the urea-denatured 434 repressor causes intensive medium-range interactions suggested by NMR nuclear Overhauser effects, hydrophobic interactions between aromatic sidechains keep a β-hairpin secondary structure in a 16-residue peptide from protein G, interactions between charged sidechains stabilizes helices in S-peptide from ribonuclease A, etc.[279-281] All these protein conformations exhibit random coil-like properties in several spectroscopic measurements but are later proved to possess folded structures. Misidentifying random coil conformation can lead to many problems, from misunderstanding the conformation and folding of a certain protein, to causing confusion in understanding the protein folding process in general.[277-278, 282]

In the case of apo-cytc, the conformation of it has been extensively studied with multiple spectroscopic methods since the early 1970s. The sedimentation coefficient and intrinsic viscosity of apo-cytc suggested a globally extended structure, the CD spectrum implied no local secondary structures, and the histidyl residues' reactivity, phenolic ionization of tyrosyl residues and UV absorbance of the tryptophanyl residue indicated the exposed nature of these residues, which are embedded in the folded holo-cytc.[268-269] Even though, our single-molecule OT results suggested that apo-cytc with such structural features is in fact still not completely random coiled. It is very likely that apo-cytc folds under strong hydrophobic interactions into an ensemble of collapsed molten globule conformations, which are observed to be folding intermediate in the folding process of many proteins, but without a well-defined structure.[283-284] Here, SMFS provides an unambiguous way of evaluating random coil behavior by exploring the mechanical

response at the single-molecule level, which may be promoted as a general method for random coil identification in the future.

5.4.3 The Role of Heme in the Folding of Holo-Cytc

The folding mechanism of holo-cytc has been extensively studied by hydrogen exchange experiments in denaturants. It was found that holo-cytc is composed of five cooperative folding units, and the folding and unfolding processes go through the same foldon-dependent native-like intermediates but in opposite orders. In the unfolding process, the substructures in the middle of the protein's sequence unfold first, followed by the unfolding of the foldon containing the N- and the C-terminal helices; and in the folding process, the N- and the C-terminal helices bind first in milliseconds, and the other structures reform subsequently in ~ 2s. It was concluded that the folded substructure formed by the N- and the C-terminal helices is more thermodynamically stable and has a faster folding kinetics than the other part of holo-cytc, and the folding of the substructures in the middle of the protein's sequence may rely on the correct folding of the N- and the C-terminal helices. Our SMFS results are consistent to such previous understandings gained from ensemble experiments in many ways. In our mechanical unfolding experiment, the folded substructure formed by the N- and the C-terminal helices is directly subjected to the force and acts as the force-bearing structural unit. Detachment of the two helices can directly lead to the complete unfolding of the whole protein (if the Fe-S(Met80) bond does not survive), indicating that the inner folded substructures do not persist in isolation without the folded N- and the C-terminal helices. The mechanical folding mechanism of holo-cytc is similar to the previously reported chemical folding mechanism, in the sense that the N- and the C-termini of unfolded holo-cytc re-bind at the beginning of the folding process.

As apo-cytc was believed to be a random coil, the interactions between the heme and the polypeptide chain were considered to play a decisive role in the folding of holo-cytc. By revealing the deviation from random coiled behavior of apo-cytc and identifying the molten globule state in apo-cytc, our results brought some new insights to the previous understanding of the folding process of holo-cytc. Instead of being a completely random coil, apo-cytc has a nonnegligible tendency to fold by itself, even though a well-defined folded structure cannot be reached in the end. The incorporation of heme greatly enhances the intrachain interactions and the folding tendency of the polypeptide chain, and helps guide the protein to successfully fold into the one native structure. The dominant but not decisive role that heme plays in the folding of holo-cytc distinguishes it from some other metalloproteins in respect of the folding mechanism, such as rubredoxin, whose polypeptide chain can fold to almost the native structure by itself, and the incorporation of ferric ion acts to enhance the conformational stability;[16] and the C-terminal repeats-in-toxin domain of adenylate cyclase toxin, whose polypeptide chain does not fold at all, and the incorporation of calcium ions enables its folding.[37] Our results highlight the complexity in the folding process of holo-cytc arising from the intertwining effects of polypeptide folding and metal coordination.

5.5 Conclusions

Using single-molecule OT, here we revealed the complete folding-unfolding mechanism of holo-cytc and demonstrated the mechanical response of apo-cytc. Holo-cytc mechanically unfolds in a two-state fashion or via an intermediate, and folds in a two-state fashion. Apo-cytc can fold into an ensemble of molten globule states, and the incorporation of heme cofactor greatly enhances the folding tendency and increases the folding fidelity to the native structure. Our results not only deepen and revise our understanding of the folding mechanism of an

important type of metalloprotein, holo-cytc, but also identify the deviation of random coil behavior of apo-cytc which was missed or overlooked in previous spectroscopic measurements.

5.6 Experimental Section

5.6.1 Protein Engineering

The original plasmid of pBTR(hCc), which was purchased from Addgene (Addgene plasmid # 61026), encodes both horse heart cytc and yeast heme lyase, thus enabling expression of holo-form horse heart cytc in *Escherichia coli*.[272] The pre-existing BglII (A'GATCT) restriction site in the plasmid was removed via standard site-directed mutagenesis methods, and the plasmid was modified by polymerase chain reaction (PCR) to carry a cysteine upstream to the cytc, as well as a new 3' BglII, a 3' KpnI (G'GTACC) restriction sites and a cysteine downstream to the cytc. The gene of protein NuG2 was then inserted between the BglII and KpnI restriction sites by digestion and ligation.

The gene of cytc (C14, 17A) was obtained by standard site-directed mutagenesis methods. The genes of cytc and cytc (C14, 17A) were amplified using standard PCR to carry a 5' BamHI (G'GATCC) and a 3' KpnI restriction site. They were then, subcloned into two modified pQE80L (Qiagen, Valencia, CA) expression vectors respectively, which allow for adding a cysteine, and a cysteine together with an NuG2, to both termini of the protein. All the sequences were confirmed by direct DNA sequencing.

All the recombinant proteins were expressed in the *Escherichia coli* strain BL21 (DE3) at 37 °C. To express holo-cytc, 5 mL of preculture was inoculated into 2 L of rich medium (12 g/L tryptone, 24 g/L yeast extract, 4mL/L glycerol, 2.3 g/L KH_2PO_4, 12.5 g/L K_2HPO_4) containing 100 μg/mL ampicillin, and the protein expression continued for 30 h.[272] To express apo-cytc, 3 mL of preculture was inoculated into 200 mL of 2.5% Luria-Bertani media containing 100 mg

mL/L ampicillin, and when the OD600 of the culture reached ~0.7, protein overexpression was induced with 0.5 mM isopropyl-b-D-1-thiogalactopyranoside (Thermo Fisher Scientific, Waltham, MA) and continued for 4 h. Protein purification was performed using the same procedure as described in chapter 2 (2.6.1).

5.6.2 Preparation of DNA-Protein Chimera

DNA handles and DNA-protein chimera were prepared using the same procedure as described in chapter 3 (3.6.2).

5.6.3 OT-Based SMFS Experiments

The single-molecule OT experiments were carried out in the same way as described in chapter 3 (3.6.3). The stretching-relaxing experiments in this work were carried out at a constant speed of 50 nm/s.

5.6.4 Calculating the Kinetics of Unfolding/Folding of Proteins

The force-dependent unfolding and folding rate constants were calculated as described in chapter 2 (3.6.4).

Chapter 6: Summary and Future Work

6.1 Thesis Summary

As an important class of proteins, metalloproteins play a wide variety of biological functions in organisms. Studying the folding-unfolding mechanisms of metalloproteins is of great importance to our understanding of the structural dynamics, function and biogenesis of metalloproteins in nature. SMFS, which employs mechanical force as a denaturant to trigger the real-time folding-unfolding process in proteins at the single-molecule level, has become a powerful tool in investigating the folding-unfolding mechanism of metalloproteins. In this thesis, I have used AFM-based and OT-based SMFS to investigate the mechanical folding-unfolding mechanisms of two important classes of metalloproteins, namely iron-sulfur proteins and heme-containing proteins. Many detailed mechanistic insights were elucidated based on the experimental results.

In Chapter 2, AFM-based SMFS was used to study the mechanical unfolding mechanism of an iron-sulfur protein, HiPIP, which contains a [4Fe-4S] cluster. We found that HiPIP unfolds in a similar way to RD and FD: firstly, the protein sequence outside the iron-sulfur center is unfolded; and secondly, the iron-sulfur center is ruptured. The rupture of the iron-sulfur center proceeds in two parallel pathways, with the iron-thiolate bonds being ruptured either concurrently or sequentially. The oxidized HiPIP has a similar mechanical stability to reduced HiPIP, likely due to the electron delocalization in the iron-sulfur center. Our results provide a detailed description of the mechanical unfolding process of HiPIP, and a previously unknown general mechanical unfolding mechanism for the iron-sulfur protein family was proposed.

In Chapter 3, the mechanical folding mechanism of the simplest iron-sulfur protein, RD, was investigated in detail by OT-based SMFS. RD contains only a ferric ion, and we found that

the folding of the RD polypeptide chain is rapid and robust, while the reconstitution of the iron-sulfur center is greatly dependent upon the coordination state of the ferric ion on the unfolded polypeptide chain. If the ferric ion is two-coordinated by two neighboring cysteines, the iron-sulfur center can fully reconstitute upon the folding of RD. However, if the ferric ion is only mono-coordinated, the iron-sulfur center cannot reconstitute. Based on this, we proposed that the folding of RD follows a novel binding-folding-reconstitution mechanism, which has not been observed previously in the folding of metalloproteins. Our study highlighted the critical importance of the two-coordinate ferric site in the folding of holo-rubredoxin, which may have some important implications to our understanding of the folding mechanism of more complex metalloproteins *in vivo*.

In Chapter 4, the *in vitro* folding-unfolding behavior of FD was studied by OT-based SMFS. We found that the mechanically unfolded ferredoxin mostly refolds into an ensemble of misfolded states with weaker mechanical stabilities comparing to its native structure. In very rare cases, the refolded ferredoxin shows a correctly refolded β-sheet or a reconstituted coordination center, however, one with its whole native structure successfully recovered is not found. Our results point out the complex folding behavior of the polypeptide chain of ferredoxin and the challenges to recover the native structure *in vitro*, and may bring important implications on our understanding of the folding of ferredoxin *in vivo*.

In Chapter 5, we studied the folding-unfolding mechanism of holo-cytc and demonstrated the mechanical response of apo-cytc using OT-based SMFS. Holo-cytc mechanically unfolds in a two-state fashion or via an intermediate, and folds in a two-state fashion. Apo-cytc can fold into an ensemble of molten globule states, and the incorporation of the heme cofactor greatly enhances the folding tendency and increases the folding fidelity to the native structure. Our

results not only deepen and revise our understanding of the folding mechanism of an important metalloprotein, cytc, but also highlight the necessity for characterizing random coiled conformations by a range of complementary techniques.

Overall, the work done in this thesis elucidated valuable new findings on the folding-unfolding behaviors of four important metalloproteins: three iron-sulfur proteins, HiPIP, RD and FD; and one heme-containing protein, cytc. The studies on HiPIP, RD and FD broaden and deepen our knowledge on the folding-unfolding behavior of iron-sulfur proteins, and the study on cytc resolves the single-molecule folding-unfolding mechanism of a heme-containing protein for the first time. These four metalloproteins have relatively simple structures and can be considered as model systems for the study of more complex metalloproteins, and therefore some of their folding-unfolding behaviors are very likely to be general among other metalloproteins. Our results further demonstrate the utility of using SMFS to study the folding-unfolding mechanisms of metalloproteins at the single-molecule level, and lay a solid foundation for further investigations in this area.

6.2 Future Work and Outlook

While significant progress has been made, using SMFS to explore the folding-unfolding mechanisms of metalloproteins remains an emerging area. Many questions in this area call for further studies, and more opportunities and challenges are still ahead.

6.2.1 Remaining Problems about the Folding-Unfolding Mechanisms of Iron-Sulfur Proteins

To date, iron-sulfur protein has become one of the most studied classes of metalloproteins in terms of the mechanical folding-unfolding behavior, yet many unsolved problems about iron-sulfur proteins call for future investigation. Chapter 2 introduces the revealed mechanical

unfolding mechanism of HiPIP. However, whether mechanically unfolded HiPIP can refold back to its native state, and if so, through what pathway it folds, remain unknown. Answering these questions can lead to a better understanding of the folding of iron-sulfur cluster-containing metalloproteins and may provide some implications on the biogenesis processes of iron-sulfur proteins.

Chapter 3 discusses the binding-folding-reconstitution mechanism that we proposed for the folding of RD. Nevertheless, there remain several important points that are unclear in the folding process of RD. For example, the existence of some residual folded structures in the unfolded apo-RD was speculated to be the reason for the two distinct conformations of the folded apo-RD. If so, what are these residual folded structures, and is it possible to use SMFS methods that can reach higher forces to disrupt these structures and verify the hypothesis? In addition, while it was surmised that different coordination states of the ferric ion on the unfolded RD led to different competences to reconstitute the iron-sulfur coordination center, which cysteine(s) the ferric ion binds in the unfolded apo-RD is still undisclosed. Computational studies may help resolve this issue, and bring us an unambiguous description of the folding process of RD.

Chapter 4 demonstrates some of the folding behaviors of FD that we observed in SMFS experiments. Nonetheless, which cysteine the cluster binds after FD is mechanically unfolded was still not clear. Does binding different cysteine lead to different competences to reconstitute the iron-sulfur coordination center, as it does in the case of RD? Does adding chaperones in the experiments assist the folding of FD, and if so, can this process be monitored at the single-molecule level by SMFS? On top of all of these questions, once a better understanding of the folding-unfolding processes of RD, FD and HiPIP is obtained, researchers can proceed to

investigate the folding-unfolding behaviors of more complicated iron-sulfur proteins, such as proteins containing multiple iron-sulfur clusters, or study the folding of iron-sulfur proteins in a more biologically relevant context, such as in the presence of iron-sulfur biogenesis proteins, which will further help people understand the structural dynamics, functions and biogenesis of iron-sulfur proteins.

6.2.2 The Folding-Unfolding Mechanisms of Other Heme-Containing Proteins

In Chapter 5, the folding-unfolding mechanism of a heme-containing protein, cytc, was revealed at the single-molecule level by OT-based SMFS. As the first holo-form heme-containing protein studied by SMFS in detail, cytc shows some distinct folding-unfolding properties that have not been observed in other metalloproteins, and the heme cofactor is proven to be important for the folding of cytc. Yet, whether these properties are unique to cytc, or to some degree general among the heme-containing protein family, remains a question. To answer this question, it is necessary to look at the folding-unfolding behaviors of other heme-containing proteins with SMFS.

Myoglobin (Mb) is a heme-containing protein that stores oxygen and facilitates the diffusion of oxygen in muscle tissues.[285] It is a monomeric protein consisting of 153 amino acids and adopts a compact, globular structure with eight helices (Figure 6.1A). The heme cofactor is non-covalently bound to the polypeptide chain via an Fe-N(His) bond, and the heme iron can also bind oxygen as its sixth ligand. Previous ensemble studies found that apo-Mb folds via an intermediate state, and the holo-Mb folds in a more cooperative way.[286-287] While the folding-unfolding behavior of the apo-Mb has been carefully characterized by OT-based SMFS, the folding-unfolding behavior of the holo-Mb at the single-molecule level has not been resolved yet, which calls for future investigations with SMFS.[288] As the heme cofactor is not covalently

bound to the polypeptide chain, the folding-unfolding of Mb may be accompanied by the association-dissociation of heme, and Mb may potentially demonstrate some synergistic effects between the protein folding and metal binding, which have only been observed in alkaline earth metal-containing metalloproteins so far (as discussed in Chapter 1.3.4). In addition, the influences brought by oxygen binding and heme orientation to the folding-unfolding behavior of Mb, which were studied by ensemble methods, can also be carefully examined at the single-molecule level.[289-290] Valuable information can be acquired regarding the folding-unfolding mechanism of myoglobin as well as heme-containing proteins in general.

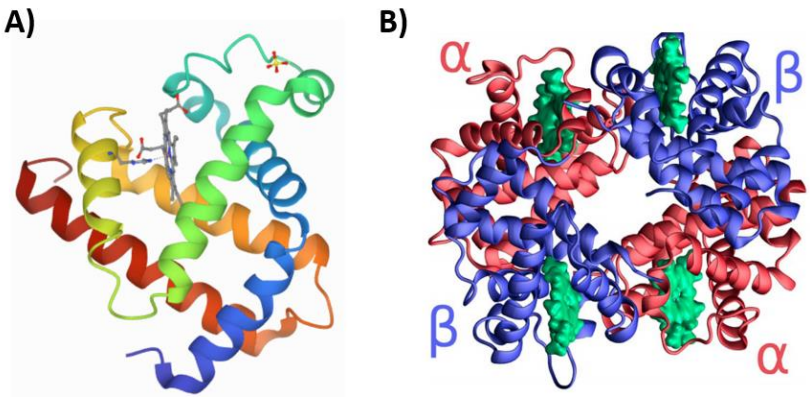

Figure 6.1 Structures of myoglobin and hemoglobin. (A) Three-dimensional structure of sperm whale myoglobin (PDB code: 1VXA).[291] The proximal histidine links the iron of the heme to the protein. (B) The tetrameric structure of hemoglobin. Image adapted with permission from Springer Nature, Copyright © 2017.[292]

The folding-unfolding behavior of another heme-containing protein, hemoglobin (Hb), is also worth careful investigations with SMFS in the future. Hb acts as a two-way respiratory carrier that transports oxygen from the lungs to the tissues and facilitates the return transport of carbon dioxide.[293] It is normally an $\alpha_2\beta_2$ tetramer consisting of four subunits, and each unit contains a non-covalently bound heme cofactor (Figure 6.1B). The α and β chains consist of 141 and 146 amino acids respectively, and their structures are very similar to that of Mb, although their amino acid compositions are very different.[294] Using ensemble methods, Hb was found to

fold and assemble in a stepwise sequential manner, where the folded monomeric α and β units form heterodimeric species and then assemble into tetramers.[249] It is of great interest to compare the folding-unfolding behavior of the α and β chain of Hb to that of Mb, and it is also possible to study the binding of the subunits with SMFS. Moreover, Hb has a large number of mutants, and some of them have compromised physiological functions, such as sickle cell Hb (glutamic acid-to-valine at position 6 in the β chain), which causes aggregation of the deoxy form of Hb and further anemia.[295] Investigating the folding-unfolding mechanisms of wild-type Hb units as well as the mutants could help us better understand not only the folding-unfolding behaviors of heme-containing proteins but also the structure-function relationship of Hb as well.

6.2.3 The Folding-Unfolding Mechanisms of Disease-Related Metalloproteins

So far, most of the studies on the folding-unfolding mechanisms of metalloproteins have been carried out on structurally simple metalloproteins. These simple metalloproteins are used as model systems, and their folding-unfolding behaviors potentially demonstrate some common features among other metalloproteins, which is invaluable in laying a foundation for understanding the folding-unfolding behaviors of metalloproteins in general. Once the fundamental knowledge has been acquired from studying the simple model systems, researchers can proceed to study metalloproteins with more complicated structures, of which the folding may be of more biological or clinical significance.

The biological function of a protein depends on its correct folding into the native structure. Misfolding into a non-native structure can result in a protein losing its biological function or gaining a toxic activity. Over the past few decades, diverse diseases have been shown to arise from misfolding and aggregation of certain proteins. Examples of diseases and the associated proteins include Alzheimer's disease and amyloid-β, Huntington disease and

Huntingtin, type II diabetes and amylin, amyotrophic lateral sclerosis (ALS) and Cu/Zn superoxide dismutase 1 (SOD1), etc.[296] Studying the folding-unfolding and misfolding mechanisms of these disease-related proteins is therefore of great significance for understanding, preventing, and developing therapeutic strategies for these diseases.

Among them, ALS is a fatal neurodegenerative disorder affecting motor neurons, caused by the misfolding and aggregation of a metalloprotein, SOD1.[297] Correctly folded SOD1 acts as a cytosolic antioxidant that protects cells against damage from superoxide. It has an immunoglobulin-like, β-barrel fold forming a homodimer, and each SOD1 monomer binds one Cu^{2+} and one Zn^{2+} ion (Figure 6.2A).[298] The folding of native, holo-form SOD1 has only been studied by ensemble experiments, in which the two monomers are found to fold in a two-state fashion or via intermediates first and then dimerize.[299-301] The folding-unfolding behavior of the reduced, apo-form of SOD1 monomer has been studied by OT-based SMFS, resolving numerous intermediate states with different numbers of native-like folded β-strands (Figure 6.2B).[302] Based on the current knowledge on the folding-unfolding mechanisms of transition metal-containing metalloproteins acquired from investigations on iron-sulfur proteins, cupredoxins, zinc finger proteins and heme-containing proteins (introduced in Chapter 1.3.4 and Chapter 2-5), it is possible and promising to use OT-based SMFS to study the folding-unfolding and misfolding mechanism of the wild-type, holo-form SOD1 at the single-molecule level. Resolving this issue could help people better understand the origin of ALS at a molecular level, as well as develop viable therapeutic strategies to treat this disease.

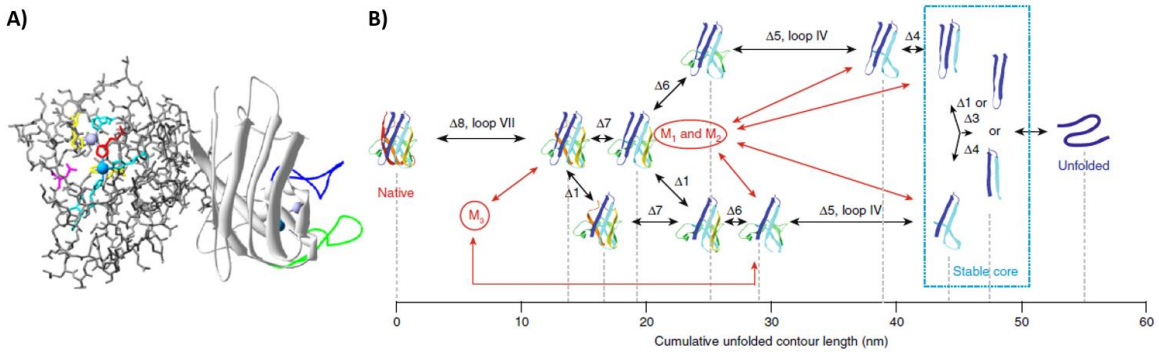

Figure 6.2 Structure of SOD1 and the folding-unfolding mechanism of apo-form SOD1 revealed by SMFS. (A) Structure of a native SOD1 dimer (pdb code: 1SPD). The left subunit is shown in stick representation and the right one in cartoon ribbon. The copper ion is colored in blue and the zinc ion in lavender. The colored residues in the left subunit are coordinating residues, and the metal binding loop and the electrostatic loop are colored in green and blue in the right subunit. Image adapted with permission from Elsevier, Copyright © 2006.[298] (B) The proposed structures of the folding-unfolding intermediates of reduced, apo-form SOD1 monomer revealed by OT-based SMFS. The drawn intermediates have native-like β-strands, and M_1, M_2 and M_3 are misfolded states. Image adapted with permission from Springer Nature, Copyright © 2017.[302]

Bibliography

1. Ouzounis, C. A.; Coulson, R. M.; Enright, A. J.; Kunin, V.; Pereira-Leal, J. B., Classification schemes for protein structure and function. *Nat Rev Genet* **2003,** *4* (7), 508-19.
2. McLachlan, A., Protein structure and function. *Annu Rev Phys Chem* **1972,** *23* (1), 165-192.
3. Robinson, P. K., Enzymes: principles and biotechnological applications. *Essays Biochem* **2015,** *59*, 1-41.
4. Onuchic, J. N.; Wolynes, P. G., Theory of protein folding. *Curr Opin Struct Biol* **2004,** *14* (1), 70-5.
5. Shakhnovich, E., Protein folding thermodynamics and dynamics: where physics, chemistry, and biology meet. *Chem Rev* **2006,** *106* (5), 1559-88.
6. Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G., Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* **1997,** *48*, 545-600.
7. Dobson, C. M., Protein folding and misfolding. *Nature* **2003,** *426* (6968), 884-90.
8. Zeeb, M.; Balbach, J., Protein folding studied by real-time NMR spectroscopy. *Methods* **2004,** *34* (1), 65-74.
9. Eaton, W. A.; Munoz, V.; Thompson, P. A.; Henry, E. R.; Hofrichter, J., Kinetics and dynamics of loops, α-helices, β-hairpins, and fast-folding proteins. *Acc Chem Res* **1998,** *31* (11), 745-753.
10. Hagihara, Y.; Aimoto, S.; Fink, A. L.; Goto, Y., Guanidine hydrochloride-induced folding of proteins. *J Mol Biol* **1993,** *231* (2), 180-4.
11. Schön, A.; Clarkson, B. R.; Jaime, M.; Freire, E., Temperature stability of proteins: Analysis of irreversible denaturation using isothermal calorimetry. *Proteins* **2017,** *85* (11), 2009-2016.
12. Kishore, D.; Kundu, S.; Kayastha, A. M., Thermal, chemical and pH induced denaturation of a multimeric β-galactosidase reveals multiple unfolding pathways. *PLoS One* **2012,** *7* (11), e50380.
13. Jagannathan, B.; Marqusee, S., Protein folding and unfolding under force. *Biopolymers* **2013,** *99* (11), 860-9.
14. Li, H.; Zheng, P., Single molecule force spectroscopy: a new tool for bioinorganic chemistry. *Curr Opin Chem Biol* **2018,** *43*, 58-67.
15. Carrion-Vazquez, M.; Oberhauser, A. F.; Fisher, T. E.; Marszalek, P. E.; Li, H.; Fernandez, J. M., Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Prog Biophys Mol Biol* **2000,** *74* (1-2), 63-91.
16. Javadi, Y.; Fernandez, J. M.; Perez-Jimenez, R., Protein folding under mechanical forces: a physiological view. *Physiology (Bethesda)* **2013,** *28* (1), 9-17.
17. Woodside, M. T.; Block, S. M., Reconstructing folding energy landscapes by single-molecule force spectroscopy. *Annu Rev Biophys* **2014,** *43*, 19-39.
18. Hughes, M. L.; Dougan, L., The physics of pulling polyproteins: a review of single molecule force spectroscopy using the AFM to study protein unfolding. *Rep Prog Phys* **2016,** *79* (7), 076601.
19. Holm, R. H.; Kennepohl, P.; Solomon, E. I., Structural and Functional Aspects of Metal Sites in Biology. *Chem Rev* **1996,** *96* (7), 2239-2314.

20. Bertini, I.; Gary, H.; Lippard, S.; Valentine, J., Bioinorganic Chemistry, Univ. Science Books, Mill Valley, CA: 1994.
21. Da Silva, J. F.; Williams, R. J. P., *The biological chemistry of the elements: the inorganic chemistry of life*. Oxford University Press: 2001.
22. Gomes, C. M.; Wittung-Stafshede, P., *Protein folding and metal ions: mechanisms, biology and disease*. CRC Press: 2016.
23. Zheng, P.; Li, H., Highly covalent ferric-thiolate bonds exhibit surprisingly low mechanical stability. *J Am Chem Soc* **2011,** *133* (17), 6791-8.
24. Zheng, P.; Li, H., Direct measurements of the mechanical stability of zinc-thiolate bonds in rubredoxin by single-molecule atomic force microscopy. *Biophys J* **2011,** *101* (6), 1467-73.
25. Zheng, P.; Chou, C. C.; Guo, Y.; Wang, Y.; Li, H., Single molecule force spectroscopy reveals the molecular mechanical anisotropy of the FeS4 metal center in rubredoxin. *J Am Chem Soc* **2013,** *135* (47), 17783-92.
26. Zheng, P.; Takayama, S. J.; Mauk, A. G.; Li, H., Single molecule force spectroscopy reveals that iron is released from the active site of rubredoxin by a stochastic mechanism. *J Am Chem Soc* **2013,** *135* (21), 7992-8000.
27. Zheng, P.; Arantes, G. M.; Field, M. J.; Li, H., Force-induced chemical reactions on the metal centre in a single metalloprotein molecule. *Nat Commun* **2015,** *6*, 7569.
28. Lei, H.; Guo, Y.; Hu, X.; Hu, C.; Hu, X.; Li, H., Reversible Unfolding and Folding of the Metalloprotein Ferredoxin Revealed by Single-Molecule Atomic Force Microscopy. *J Am Chem Soc* **2017,** *139* (4), 1538-1544.
29. Beedle, A. E. M.; Lezamiz, A.; Stirnemann, G.; Garcia-Manyes, S., The mechanochemistry of copper reports on the directionality of unfolding in model cupredoxin proteins. *Nat Commun* **2015,** *6*, 7894.
30. Wei, W.; Sun, Y.; Zhu, M.; Liu, X.; Sun, P.; Wang, F.; Gui, Q.; Meng, W.; Cao, Y.; Zhao, J., Structural Insights and the Surprisingly Low Mechanical Stability of the Au-S Bond in the Gold-Specific Protein GolB. *J Am Chem Soc* **2015,** *137* (49), 15358-61.
31. Yuan, G.; Liu, H.; Ma, Q.; Li, X.; Nie, J.; Zuo, J.; Zheng, P., Single-Molecule Force Spectroscopy Reveals that Iron-Ligand Bonds Modulate Proteins in Different Modes. *J Phys Chem Lett* **2019,** *10* (18), 5428-5433.
32. Perales-Calvo, J.; Lezamiz, A.; Garcia-Manyes, S., The Mechanochemistry of a Structural Zinc Finger. *J Phys Chem Lett* **2015,** *6* (17), 3335-40.
33. Yadav, A.; Paul, S.; Venkatramani, R.; Ainavarapu, S. R. K., Differences in the mechanical unfolding pathways of apo- and copper-bound azurins. *Sci Rep* **2018,** *8* (1), 1989.
34. Stigler, J.; Ziegler, F.; Gieseke, A.; Gebhardt, J. C.; Rief, M., The complex folding network of single calmodulin molecules. *Science* **2011,** *334* (6055), 512-6.
35. Junker, J. P.; Rief, M., Single-molecule force spectroscopy distinguishes target binding modes of calmodulin. *Proc Natl Acad Sci U S A* **2009,** *106* (34), 14361-6.
36. Stigler, J.; Rief, M., Calcium-dependent folding of single calmodulin molecules. *Proc Natl Acad Sci U S A* **2012,** *109* (44), 17814-9.
37. Wang, H.; Gao, X.; Li, H., Single Molecule Force Spectroscopy Reveals the Mechanical Design Governing the Efficient Translocation of the Bacterial Toxin Protein RTX. *J Am Chem Soc* **2019,** *141* (51), 20498-20506.
38. Milles, L. F.; Unterauer, E. M.; Nicolaus, T.; Gaub, H. E., Calcium stabilizes the strongest protein fold. *Nat Commun* **2018,** *9* (1), 4764.

39. Guttula, D.; Yao, M.; Baker, K.; Yang, L.; Goult, B. T.; Doyle, P. S.; Yan, J., Calcium-mediated Protein Folding and Stabilization of Salmonella Biofilm-associated Protein A. *J Mol Biol* **2019,** *431* (2), 433-443.
40. Anfinsen, C. B., Principles that govern the folding of protein chains. *Science* **1973,** *181* (4096), 223-30.
41. Mutharasappan, N.; Rao, G. R.; Mariadasse, R.; Poopandi, S.; Mathimaran, A.; Dhamodharan, P.; Sundarraj, R.; Pandian, C. J.; Jeyaraman, J., Experimental and Computational Methods to Determine Protein Structure and Stability. In *Frontiers in Protein Structure, Function, and Dynamics*, Springer: 2020; pp 23-55.
42. Wei, G.-W., Protein structure prediction beyond AlphaFold. *Nat Mach Intell* **2019,** *1* (8), 336-337.
43. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res* **2000,** *28* (1), 235-42.
44. Baker, D., Metastable states and folding free energy barriers. *Nat Struct Biol* **1998,** *5* (12), 1021-4.
45. Dill, K. A., Dominant forces in protein folding. *Biochemistry* **1990,** *29* (31), 7133-55.
46. Pace, C. N.; Shirley, B. A.; McNutt, M.; Gajiwala, K., Forces contributing to the conformational stability of proteins. *Faseb j* **1996,** *10* (1), 75-83.
47. Nick Pace, C.; Scholtz, J. M.; Grimsley, G. R., Forces stabilizing proteins. *FEBS Lett* **2014,** *588* (14), 2177-84.
48. Ponder, J. W.; Case, D. A., Force fields for protein simulations. *Adv Protein Chem* **2003,** *66*, 27-85.
49. Kim, Y. E.; Hipp, M. S.; Bracher, A.; Hayer-Hartl, M.; Hartl, F. U., Molecular chaperone functions in protein folding and proteostasis. *Annu Rev Biochem* **2013,** *82*, 323-55.
50. Gershenson, A.; Gierasch, L. M., Protein folding in the cell: challenges and progress. *Curr Opin Struct Biol* **2011,** *21* (1), 32-41.
51. Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C., A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **1958,** *181* (4610), 662-6.
52. Kendrew, J. C.; Dickerson, R. E.; Strandberg, B. E.; Hart, R. G.; Davies, D. R.; Phillips, D. C.; Shore, V. C., Structure of myoglobin: A three-dimensional Fourier synthesis at 2 A. resolution. *Nature* **1960,** *185* (4711), 422-7.
53. Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Muirhead, H.; Will, G.; North, A. C., Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. *Nature* **1960,** *185* (4711), 416-22.
54. Haber, E.; Anfinsen, C. B., Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. *J Biol Chem* **1962,** *237*, 1839-44.
55. Sohl, J. L.; Jaswal, S. S.; Agard, D. A., Unfolded conformations of alpha-lytic protease are more stable than its native state. *Nature* **1998,** *395* (6704), 817-9.
56. Wang, Z.; Mottonen, J.; Goldsmith, E. J., Kinetically controlled folding of the serpin plasminogen activator inhibitor 1. *Biochemistry* **1996,** *35* (51), 16443-8.
57. Levinthal, C., Are there pathways for protein folding? *J chim phys* **1968,** *65*, 44-45.
58. Kim, P. S.; Baldwin, R. L., Intermediates in the folding reactions of small proteins. *Annu Rev Biochem* **1990,** *59*, 631-60.
59. Matthews, C. R., Pathways of protein folding. *Annu Rev Biochem* **1993,** *62*, 653-83.

60. Kim, P. S.; Baldwin, R. L., Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu Rev Biochem* **1982,** *51*, 459-89.
61. Karplus, M.; Weaver, D. L., Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci* **1994,** *3* (4), 650-68.
62. Karplus, M.; Weaver, D. L., Diffusion–collision model for protein folding. *Biopolymers: Original Research on Biomolecules* **1979,** *18* (6), 1421-1437.
63. Mayor, U.; Johnson, C. M.; Daggett, V.; Fersht, A. R., Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc Natl Acad Sci U S A* **2000,** *97* (25), 13518-22.
64. Myers, J. K.; Oas, T. G., Preorganized secondary structure as an important determinant of fast protein folding. *Nat Struct Biol* **2001,** *8* (6), 552-8.
65. Ptitsyn, O. B., Protein folding: hypotheses and experiments. *J Protein Chem* **1987,** *6* (4), 273-293.
66. Daggett, V.; Fersht, A. R., Is there a unifying mechanism for protein folding? *Trends Biochem Sci* **2003,** *28* (1), 18-25.
67. Uversky, V. N.; Fink, A. L., The chicken-egg scenario of protein folding revisited. *FEBS Lett* **2002,** *515* (1-3), 79-83.
68. Fersht, A. R., Nucleation mechanisms in protein folding. *Curr Opin Struct Biol* **1997,** *7* (1), 3-9.
69. Clarke, J.; Hamill, S. J.; Johnson, C. M., Folding and stability of a fibronectin type III domain of human tenascin. *J Mol Biol* **1997,** *270* (5), 771-8.
70. Fulton, K. F.; Main, E. R.; Daggett, V.; Jackson, S. E., Mapping the interactions present in the transition state for unfolding/folding of FKBP12. *J Mol Biol* **1999,** *291* (2), 445-61.
71. Krishna, M. M.; Maity, H.; Rumbley, J. N.; Lin, Y.; Englander, S. W., Order of steps in the cytochrome C folding pathway: evidence for a sequential stabilization mechanism. *J Mol Biol* **2006,** *359* (5), 1410-9.
72. Debe, D. A.; Carlson, M. J.; Goddard, W. A., 3rd, The topomer-sampling model of protein folding. *Proc Natl Acad Sci U S A* **1999,** *96* (6), 2596-601.
73. Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R., The protein folding problem. *Annu Rev Biophys* **2008,** *37*, 289-316.
74. Dill, K. A.; Chan, H. S., From Levinthal to pathways to funnels. *Nat Struct Biol* **1997,** *4* (1), 10-9.
75. Varela, A. E.; England, K. A.; Cavagnero, S., Kinetic trapping in protein folding. *Protein Eng Des Sel* **2019,** *32* (2), 103-108.
76. Leopold, P. E.; Montal, M.; Onuchic, J. N., Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc Natl Acad Sci U S A* **1992,** *89* (18), 8721-5.
77. Wang, J.; Oliveira, R. J.; Chu, X.; Whitford, P. C.; Chahine, J.; Han, W.; Wang, E.; Onuchic, J. N.; Leite, V. B., Topography of funneled landscapes determines the thermodynamics and kinetics of protein folding. *Proc Natl Acad Sci U S A* **2012,** *109* (39), 15763-8.
78. Wolynes, P. G., Recent successes of the energy landscape theory of protein folding and function. *Q Rev Biophys* **2005,** *38* (4), 405-10.
79. Gruebele, M.; Dave, K.; Sukenik, S., Globular Protein Folding In Vitro and In Vivo. *Annu Rev Biophys* **2016,** *45*, 233-51.
80. Smith, A. E.; Zhang, Z.; Pielak, G. J.; Li, C., NMR studies of protein folding and binding in cells and cell-like environments. *Curr Opin Struct Biol* **2015,** *30*, 7-16.

81. Ebbinghaus, S.; Dhar, A.; McDonald, J. D.; Gruebele, M., Protein folding stability and dynamics imaged in a living cell. *Nat Methods* **2010,** *7* (4), 319-23.
82. Ebbinghaus, S.; Gruebele, M., Protein folding landscapes in the living cell. *J Phys Chem Lett* **2011,** *2* (4), 314-319.
83. Ignatova, Z.; Gierasch, L. M., Inhibition of protein aggregation in vitro and in vivo by a natural osmoprotectant. *Proc Natl Acad Sci U S A* **2006,** *103* (36), 13357-61.
84. König, I.; Zarrine-Afsar, A.; Aznauryan, M.; Soranno, A.; Wunderlich, B.; Dingfelder, F.; Stüber, J. C.; Plückthun, A.; Nettels, D.; Schuler, B., Single-molecule spectroscopy of protein conformational dynamics in live eukaryotic cells. *Nat Methods* **2015,** *12* (8), 773-9.
85. Hingorani, K. S.; Gierasch, L. M., Comparing protein folding in vitro and in vivo: foldability meets the fitness challenge. *Curr Opin Struct Biol* **2014,** *24*, 81-90.
86. Street, T. O.; Courtemanche, N.; Barrick, D., Protein folding and stability using denaturants. *Methods Cell Biol* **2008,** *84*, 295-325.
87. Gupta, R.; Yadav, S.; Ahmad, F., Protein stability: urea-induced versus guanidine-induced unfolding of metmyoglobin. *Biochemistry* **1996,** *35* (36), 11925-30.
88. Sánchez-Ruiz, J. M.; López-Lacomba, J. L.; Cortijo, M.; Mateo, P. L., Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin. *Biochemistry* **1988,** *27* (5), 1648-52.
89. Robinson, A.; van Oijen, A. M., Bacterial replication, transcription and translation: mechanistic insights from single-molecule biochemical studies. *Nat Rev Microbiol* **2013,** *11* (5), 303-15.
90. Olivares, A. O.; Baker, T. A.; Sauer, R. T., Mechanical Protein Unfolding and Degradation. *Annu Rev Physiol* **2018,** *80*, 413-429.
91. Bell, G. I., Models for the specific adhesion of cells to cells. *Science* **1978,** *200* (4342), 618-27.
92. Evans, E.; Ritchie, K., Dynamic strength of molecular adhesion bonds. *Biophys J* **1997,** *72* (4), 1541-55.
93. Evans, E.; Ritchie, K., Strength of a weak bond connecting flexible polymer chains. *Biophys J* **1999,** *76* (5), 2439-47.
94. Kramers, H. A., Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940,** *7* (4), 284-304.
95. Evans, E., Probing the relation between force--lifetime--and chemistry in single molecular bonds. *Annu Rev Biophys Biomol Struct* **2001,** *30*, 105-28.
96. King, W. T.; Su, M.; Yang, G., Monte Carlo simulation of mechanical unfolding of proteins based on a simple two-state model. *Int J Biol Macromol* **2010,** *46* (2), 159-66.
97. Dudko, O. K.; Hummer, G.; Szabo, A., Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc Natl Acad Sci U S A* **2008,** *105* (41), 15755-60.
98. Guo, S.; Efremov, A. K.; Yan, J., Understanding the catch-bond kinetics of biomolecules on a one-dimensional energy landscape. *Communications Chemistry* **2019,** *2* (1), 1-9.
99. Guo, S.; Tang, Q.; Yao, M.; You, H.; Le, S.; Chen, H.; Yan, J., Structural-elastic determination of the force-dependent transition rate of biomolecules. *Chem Sci* **2018,** *9* (27), 5871-5882.
100. Stirnemann, G.; Kang, S. G.; Zhou, R.; Berne, B. J., How force unfolding differs from chemical denaturation. *Proc Natl Acad Sci U S A* **2014,** *111* (9), 3413-8.

101. Ritort, F., Single-molecule experiments in biological physics: methods and applications. *J Phys Condens Matter* **2006,** *18* (32), R531-83.
102. Oberhauser, A. F.; Carrión-Vázquez, M., Mechanical biochemistry of proteins one molecule at a time. *J Biol Chem* **2008,** *283* (11), 6617-21.
103. Neuman, K. C.; Nagy, A., Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat Methods* **2008,** *5* (6), 491-505.
104. Chen, H.; Fu, H.; Zhu, X.; Cong, P.; Nakamura, F.; Yan, J., Improved high-force magnetic tweezers for stretching and refolding of proteins and short DNA. *Biophys J* **2011,** *100* (2), 517-23.
105. Zhao, X.; Zeng, X.; Lu, C.; Yan, J., Studying the mechanical responses of proteins using magnetic tweezers. *Nanotechnology* **2017,** *28* (41), 414002.
106. Le, S.; Liu, R.; Lim, C. T.; Yan, J., Uncovering mechanosensing mechanisms at the single protein level using magnetic tweezers. *Methods* **2016,** *94*, 13-8.
107. Tapia-Rojo, R.; Eckels, E. C.; Fernández, J. M., Ephemeral states in protein folding under force captured with a magnetic tweezers design. *Proc Natl Acad Sci U S A* **2019,** *116* (16), 7873-7878.
108. Popa, I.; Rivas-Pardo, J. A.; Eckels, E. C.; Echelman, D. J.; Badilla, C. L.; Valle-Orero, J.; Fernández, J. M., A HaloTag Anchored Ruler for Week-Long Studies of Protein Dynamics. *J Am Chem Soc* **2016,** *138* (33), 10546-53.
109. Chen, H.; Yuan, G.; Winardhi, R. S.; Yao, M.; Popa, I.; Fernandez, J. M.; Yan, J., Dynamics of equilibrium folding and unfolding transitions of titin immunoglobulin domain under constant forces. *J Am Chem Soc* **2015,** *137* (10), 3540-6.
110. Hutter, J. L.; Bechhoefer, J., Calibration of atomic‐force microscope tips. *Rev Sci Instrum* **1993,** *64* (7), 1868-1873.
111. Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E., Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **1997,** *276* (5315), 1109-12.
112. Hinterdorfer, P.; Dufrêne, Y. F., Detection and localization of single molecular recognition events using atomic force microscopy. *Nat Methods* **2006,** *3* (5), 347-55.
113. Ashkin, A., Acceleration and trapping of particles by radiation pressure. *Phys Rev Lett* **1970,** *24* (4), 156.
114. Ashkin, A.; Dziedzic, J. M.; Bjorkholm, J. E.; Chu, S., Observation of a single-beam gradient force optical trap for dielectric particles. *Opt Lett* **1986,** *11* (5), 288.
115. Favre-Bulle, I. A.; Stilgoe, A. B.; Scott, E. K.; Rubinsztein-Dunlop, H., Optical trapping in vivo: theory, practice, and applications. *Nanophotonics* **2019,** *8* (6), 1023-1040.
116. Nieminen, T. A.; Knöner, G.; Heckenberg, N. R.; Rubinsztein-Dunlop, H., Physics of optical tweezers. *Methods Cell Biol* **2007,** *82*, 207-36.
117. Svoboda, K.; Block, S. M., Biological applications of optical forces. *Annu Rev Biophys Biomol Struct* **1994,** *23*, 247-85.
118. Smith, S. B.; Cui, Y.; Bustamante, C., Optical-trap force transducer that operates by direct measurement of light momentum. *Methods Enzymol* **2003,** *361*, 134-62.
119. Neuman, K. C.; Block, S. M., Optical trapping. *Rev Sci Instrum* **2004,** *75* (9), 2787-809.
120. Sarshar, M.; Wong, W. T.; Anvari, B., Comparative study of methods to calibrate the stiffness of a single-beam gradient-force optical tweezers over various laser trapping powers. *J Biomed Opt* **2014,** *19* (11), 115001.

121. Rohrbach, A.; Stelzer, E. H., Three-dimensional position detection of optically trapped dielectric particles. *J Appl Phys* **2002,** *91* (8), 5474-5488.
122. Gittes, F.; Schmidt, C. F., Interference model for back-focal-plane displacement detection in optical tweezers. *Opt Lett* **1998,** *23* (1), 7-9.
123. Cecconi, C.; Shank, E. A.; Bustamante, C.; Marqusee, S., Direct observation of the three-state folding of a single protein molecule. *Science* **2005,** *309* (5743), 2057-60.
124. Yang, Y. R.; Liu, Y.; Yan, H., DNA Nanostructures as Programmable Biomolecular Scaffolds. *Bioconjug Chem* **2015,** *26* (8), 1381-95.
125. Mukhortava, A.; Schlierf, M., Efficient Formation of Site-Specific Protein-DNA Hybrids Using Copper-Free Click Chemistry. *Bioconjug Chem* **2016,** *27* (7), 1559-63.
126. Min, D.; Arbing, M. A.; Jefferson, R. E.; Bowie, J. U., A simple DNA handle attachment method for single molecule mechanical manipulation experiments. *Protein Sci* **2016,** *25* (8), 1535-44.
127. Jiao, J.; Rebane, A. A.; Ma, L.; Zhang, Y., Single-Molecule Protein Folding Experiments Using High-Precision Optical Tweezers. *Methods Mol Biol* **2017,** *1486*, 357-390.
128. Cecconi, C.; Shank, E. A.; Dahlquist, F. W.; Marqusee, S.; Bustamante, C., Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers. *Eur Biophys J* **2008,** *37* (6), 729-38.
129. Hermanson, G. T., *Bioconjugate techniques*. Academic press: 2013.
130. Yu, H.; Liu, X.; Neupane, K.; Gupta, A. N.; Brigley, A. M.; Solanki, A.; Sosova, I.; Woodside, M. T., Direct observation of multiple misfolding pathways in a single prion protein molecule. *Proc Natl Acad Sci U S A* **2012,** *109* (14), 5283-8.
131. Lei, H.; He, C.; Hu, C.; Li, J.; Hu, X.; Hu, X.; Li, H., Single-Molecule Force Spectroscopy Trajectories of a Single Protein and Its Polyproteins Are Equivalent: A Direct Experimental Validation Based on A Small Protein NuG2. *Angew Chem Int Ed Engl* **2017,** *56* (22), 6117-6121.
132. Brujic, J.; Hermans, R. I.; Garcia-Manyes, S.; Walther, K. A.; Fernandez, J. M., Dwell-time distribution analysis of polyprotein unfolding using force-clamp spectroscopy. *Biophys J* **2007,** *92* (8), 2896-903.
133. Cao, Y.; Li, H., Single-molecule force-clamp spectroscopy: dwell time analysis and practical considerations. *Langmuir* **2011,** *27* (4), 1440-7.
134. Sharma, D.; Perisic, O.; Peng, Q.; Cao, Y.; Lam, C.; Lu, H.; Li, H., Single-molecule force spectroscopy reveals a mechanically stable protein fold and the rational tuning of its mechanical stability. *Proc Natl Acad Sci U S A* **2007,** *104* (22), 9278-83.
135. Stirnemann, G.; Giganti, D.; Fernandez, J. M.; Berne, B. J., Elasticity, structure, and relaxation of extended proteins under force. *Proc Natl Acad Sci U S A* **2013,** *110* (10), 3847-52.
136. Bouchiat, C.; Wang, M. D.; Allemand, J.; Strick, T.; Block, S. M.; Croquette, V., Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys J* **1999,** *76* (1 Pt 1), 409-13.
137. Marko, J. F.; Siggia, E. D., Stretching dna. *Macromolecules* **1995,** *28* (26), 8759-8770.
138. Rief, M.; Fernandez, J. M.; Gaub, H. E., Elastically coupled two-level systems as a model for biopolymer extensibility. *Phys Rev Lett* **1998,** *81* (21), 4764.
139. Yao, M.; Chen, H.; Yan, J., Thermodynamics of force-dependent folding and unfolding of small protein and nucleic acid structures. *Integr Biol (Camb)* **2015,** *7* (10), 1154-60.

140. Oberbarnscheidt, L.; Janissen, R.; Oesterhelt, F., Direct and model free calculation of force-dependent dissociation rates from force spectroscopic data. *Biophys J* **2009,** *97* (9), L19-21.
141. Bertini, I.; Bertini, G.; Gray, H.; Gray, H. B.; Stiefel, E.; Valentine, J. S.; Stiefel, E. I., *Biological inorganic chemistry: structure and reactivity*. University Science Books: 2007.
142. Gray, H. B., Biological inorganic chemistry at the beginning of the 21st century. *Proc Natl Acad Sci U S A* **2003,** *100* (7), 3563-8.
143. Andreini, C.; Bertini, I.; Cavallaro, G.; Holliday, G. L.; Thornton, J. M., Metal ions in biological catalysis: from enzyme databases to general principles. *J Biol Inorg Chem* **2008,** *13* (8), 1205-18.
144. Lippard, S. J.; Berg, J. M., *Principles of bioinorganic chemistry*. University Science Books Mill Valley, CA: 1994; Vol. 70.
145. Warder, S. E.; Prorok, M.; Chen, Z.; Li, L.; Zhu, Y.; Pedersen, L. G.; Ni, F.; Castellino, F. J., The roles of individual gamma-carboxyglutamate residues in the solution structure and cation-dependent properties of conantokin-T. *J Biol Chem* **1998,** *273* (13), 7512-22.
146. Wittung-Stafshede, P., Role of cofactors in protein folding. *Acc Chem Res* **2002,** *35* (4), 201-8.
147. Waldron, K. J.; Rutherford, J. C.; Ford, D.; Robinson, N. J., Metalloproteins and metal sensing. *Nature* **2009,** *460* (7257), 823-30.
148. Gohara, D. W.; Di Cera, E., Molecular Mechanisms of Enzyme Activation by Monovalent Cations. *J Biol Chem* **2016,** *291* (40), 20840-20848.
149. Carafoli, E.; Krebs, J., Why Calcium? How Calcium Became the Best Communicator. *J Biol Chem* **2016,** *291* (40), 20849-20857.
150. Wilson, C. J.; Apiyo, D.; Wittung-Stafshede, P., Role of cofactors in metalloprotein folding. *Q Rev Biophys* **2004,** *37* (3-4), 285-314.
151. Irving, H.; Williams, R., Order of stability of metal complexes. *Nature* **1948,** *162* (4123), 746-747.
152. Tottey, S.; Waldron, K. J.; Firbank, S. J.; Reale, B.; Bessant, C.; Sato, K.; Cheek, T. R.; Gray, J.; Banfield, M. J.; Dennison, C.; Robinson, N. J., Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. *Nature* **2008,** *455* (7216), 1138-42.
153. Waldron, K. J.; Robinson, N. J., How do bacterial cells ensure that metalloproteins get the correct metal? *Nat Rev Microbiol* **2009,** *7* (1), 25-35.
154. Lamb, A. L.; Wernimont, A. K.; Pufahl, R. A.; Culotta, V. C.; O'Halloran, T. V.; Rosenzweig, A. C., Crystal structure of the copper chaperone for superoxide dismutase. *Nat Struct Biol* **1999,** *6* (8), 724-9.
155. Lamb, A. L.; Torres, A. S.; O'Halloran, T. V.; Rosenzweig, A. C., Heterodimeric structure of superoxide dismutase in complex with its metallochaperone. *Nat Struct Biol* **2001,** *8* (9), 751-5.
156. Agar, J. N.; Krebs, C.; Frazzon, J.; Huynh, B. H.; Dean, D. R.; Johnson, M. K., IscU as a scaffold for iron-sulfur cluster biosynthesis: sequential assembly of [2Fe-2S] and [4Fe-4S] clusters in IscU. *Biochemistry* **2000,** *39* (27), 7856-62.
157. Ollagnier-de-Choudens, S.; Mattioli, T.; Takahashi, Y.; Fontecave, M., Iron-sulfur cluster assembly: characterization of IscA and evidence for a specific and functional complex with ferredoxin. *J Biol Chem* **2001,** *276* (25), 22604-7.

158. Ollagnier-De Choudens, S.; Sanakis, Y.; Hewitson, K. S.; Roach, P.; Baldwin, J. E.; Münck, E.; Fontecave, M., Iron-sulfur center of biotin synthase and lipoate synthase. *Biochemistry* **2000,** *39* (14), 4165-73.
159. Banci, L.; Bertini, I.; Ciofi-Baffoni, S.; Huffman, D. L.; O'Halloran, T. V., Solution structure of the yeast copper transporter domain Ccc2a in the apo and Cu(I)-loaded states. *J Biol Chem* **2001,** *276* (11), 8415-26.
160. Garrett, T. P.; Clingeleffer, D. J.; Guss, J. M.; Rogers, S. J.; Freeman, H. C., The crystal structure of poplar apoplastocyanin at 1.8-A resolution. The geometry of the copper-binding site is created by the polypeptide. *J Biol Chem* **1984,** *259* (5), 2822-5.
161. Banci, L.; Bertini, I.; Cramaro, F.; Del Conte, R.; Viezzoli, M. S., Solution structure of Apo Cu,Zn superoxide dismutase: role of metal ions in protein folding. *Biochemistry* **2003,** *42* (32), 9543-53.
162. Tuffery, P.; Etchebest, C.; Popot, J. L.; Lavery, R., Prediction of the positioning of the seven transmembrane alpha-helices of bacteriorhodopsin. A molecular simulation study. *J Mol Biol* **1994,** *236* (4), 1105-22.
163. Bayley, P.; Martin, S.; Jones, G., The conformation of calmodulin: a substantial environmentally sensitive helical transition in Ca4-calmodulin with potential mechanistic function. *FEBS Lett* **1988,** *238* (1), 61-6.
164. Li, W.; Wang, J.; Zhang, J.; Wang, W., Molecular simulations of metal-coupled protein folding. *Curr Opin Struct Biol* **2015,** *30*, 25-31.
165. Leal, S. S.; Botelho, H. M.; Gomes, C. M., Metal ions as modulators of protein conformation and misfolding in neurodegeneration. *Coord Chem Rev* **2012,** *256* (19-20), 2253-2270.
166. Li, W.; Zhang, J.; Wang, J.; Wang, W., Metal-coupled folding of Cys2His2 zinc-finger. *J Am Chem Soc* **2008,** *130* (3), 892-900.
167. Weinkam, P.; Romesberg, F. E.; Wolynes, P. G., Chemical frustration in the protein folding landscape: grand canonical ensemble simulations of cytochrome c. *Biochemistry* **2009,** *48* (11), 2394-402.
168. Zong, C.; Wilson, C. J.; Shen, T.; Wittung-Stafshede, P.; Mayo, S. L.; Wolynes, P. G., Establishing the entatic state in folding metallated Pseudomonas aeruginosa azurin. *Proc Natl Acad Sci U S A* **2007,** *104* (9), 3159-64.
169. Li, W.; Wang, W.; Takada, S., Energy landscape views for interplays among folding, binding, and allostery of calmodulin domains. *Proc Natl Acad Sci U S A* **2014,** *111* (29), 10550-5.
170. Knapp, J. A.; Pace, C. N., Guanidine hydrochloride and acid denaturation of horse, cow, and Candida krusei cytochromes c. *Biochemistry* **1974,** *13* (6), 1289-94.
171. Russell, B. S.; Bren, K. L., Denaturant dependence of equilibrium unfolding intermediates and denatured state structure of horse ferricytochrome c. *J Biol Inorg Chem* **2002,** *7* (7-8), 909-16.
172. Latypov, R. F.; Maki, K.; Cheng, H.; Luck, S. D.; Roder, H., Folding mechanism of reduced Cytochrome c: equilibrium and kinetic properties in the presence of carbon monoxide. *J Mol Biol* **2008,** *383* (2), 437-53.
173. Sagle, L. B.; Zimmermann, J.; Dawson, P. E.; Romesberg, F. E., Direct and high resolution characterization of cytochrome c equilibrium folding. *J Am Chem Soc* **2006,** *128* (44), 14232-3.

174. Hu, W.; Kan, Z. Y.; Mayne, L.; Englander, S. W., Cytochrome c folds through foldon-dependent native-like intermediates in an ordered pathway. *Proc Natl Acad Sci U S A* **2016,** *113* (14), 3809-14.
175. Pierce, M. M.; Nall, B. T., Coupled kinetic traps in cytochrome c folding: His-heme misligation and proline isomerization. *J Mol Biol* **2000,** *298* (5), 955-69.
176. Ptitsyn, O. B., Protein folding and protein evolution: common folding nucleus in different subfamilies of c-type cytochromes? *J Mol Biol* **1998,** *278* (3), 655-66.
177. Adman, E. T., Copper protein structures. *Adv Protein Chem* **1991,** *42*, 145-97.
178. Leckner, J.; Wittung, P.; Bonander, N.; Karlsson, B. G.; Malmström, B. G., The effect of redox state on the folding free energy of azurin. *J Biol Inorg Chem* **1997,** *2* (3), 368-371.
179. Pozdnyakova, I.; Guidry, J.; Wittung-Stafshede, P., Copper-triggered β-hairpin formation: initiation site for azurin folding? *J Am Chem Soc* **2000,** *122* (26), 6337-6338.
180. Robinson, C. R.; Liu, Y.; Thomson, J. A.; Sturtevant, J. M.; Sligar, S. G., Energetics of heme binding to native and denatured states of cytochrome b562. *Biochemistry* **1997,** *36* (51), 16141-6.
181. Bertini, I.; Cowan, J. A.; Luchinat, C.; Natarajan, K.; Piccioli, M., Characterization of a partially unfolded high potential iron protein. *Biochemistry* **1997,** *36* (31), 9332-9.
182. Wittung-Stafshede, P.; Malmstrom, B.; Winkler, J.; Gray, H., Electron-transfer triggered folding of deoxymyoglobin. *J. Phys. Chem* **1998,** *102*, 5599-5601.
183. Wittung-Stafshede, P.; Lee, J. C.; Winkler, J. R.; Gray, H. B., Cytochrome b562 folding triggered by electron transfer: approaching the speed limit for formation of a four-helix-bundle protein. *Proc Natl Acad Sci U S A* **1999,** *96* (12), 6587-90.
184. Luisi, D. L.; Wu, W. J.; Raleigh, D. P., Conformational analysis of a set of peptides corresponding to the entire primary sequence of the N-terminal domain of the ribosomal protein L9: evidence for stable native-like secondary structure in the unfolded state. *J Mol Biol* **1999,** *287* (2), 395-407.
185. Pozdnyakova, I.; Wittung-Stafshede, P., Biological relevance of metal binding before protein folding. *J Am Chem Soc* **2001,** *123* (41), 10135-6.
186. Lill, R., Function and biogenesis of iron-sulphur proteins. *Nature* **2009,** *460* (7257), 831-8.
187. Bonomi, F.; Iametti, S.; Ferranti, P.; Kurtz, D. M., Jr.; Morleo, A.; Ragg, E. M., "Iron priming" guides folding of denatured aporubredoxins. *J Biol Inorg Chem* **2008,** *13* (6), 981-91.
188. Henriques, B. J.; Saraiva, L. M.; Gomes, C. M., Probing the mechanism of rubredoxin thermal unfolding in the absence of salt bridges by temperature jump experiments. *Biochem Biophys Res Commun* **2005,** *333* (3), 839-44.
189. Leal, S. S.; Gomes, C. M., On the relative contribution of ionic interactions over iron-sulfur clusters to ferredoxin stability. *Biochim Biophys Acta* **2008,** *1784* (11), 1596-600.
190. Grandbois, M.; Beyer, M.; Rief, M.; Clausen-Schaumann, H.; Gaub, H. E., How strong is a covalent bond? *Science* **1999,** *283* (5408), 1727-30.
191. Conti, M.; Falini, G.; Samorì, B., How Strong Is the Coordination Bond between a Histidine Tag and Ni - Nitrilotriacetate? An Experiment of Mechanochemistry on Single Molecules. *Angew Chem Int Ed Engl* **2000,** *39* (1), 215-218.
192. Hao, X.; Zhu, N.; Gschneidtner, T.; Jonsson, E.; Zhang, J.; Moth-Poulsen, K.; Wang, H.; Thygesen, K. S.; Jacobsen, K. W.; Ulstrup, J.; Chi, Q., Direct measurement and modulation of

single-molecule coordinative bonding forces in a transition metal complex. *Nat Commun* **2013,** *4*, 2121.
193. Cao, Y.; Yoo, T.; Li, H., Single molecule force spectroscopy reveals engineered metal chelation is a general approach to enhance mechanical stability of proteins. *Proc Natl Acad Sci U S A* **2008,** *105* (32), 11152-7.
194. Xia, J.; Zuo, J.; Li, H., Single molecule force spectroscopy reveals that the oxidation state of cobalt ions plays an important role in enhancing the mechanical stability of proteins. *Nanoscale* **2019,** *11* (42), 19791-19796.
195. Cao, Y.; Li, Y. D.; Li, H., Enhancing the mechanical stability of proteins through a cocktail approach. *Biophys J* **2011,** *100* (7), 1794-9.
196. Holm, R. H.; Kennepohl, P.; Solomon, E. I., Structural and Functional Aspects of Metal Sites in Biology. *Chem. Rev.* **1996,** *96* (7), 2239-314.
197. Johnson, D. C.; Dean, D. R.; Smith, A. D.; Johnson, M. K., Structure, function, and formation of biological iron-sulfur clusters. *Annu. Rev. Biochem.* **2005,** *74*, 247-81.
198. Wittung-Stafshede, P., Role of cofactors in protein folding. *Acc. Chem. Res.* **2002,** *35* (4), 201-8.
199. Morleo, A.; Bonomi, F.; Iametti, S.; Huang, V. W.; Kurtz, D. M., Jr., Iron-nucleated folding of a metalloprotein in high urea: resolution of metal binding and protein folding events. *Biochemistry* **2010,** *49* (31), 6627-34.
200. Palm-Espling, M. E.; Niemiec, M. S.; Wittung-Stafshede, P., Role of metal in folding and stability of copper proteins in vitro. *Biochim. Biophys. Acta* **2012,** *1823* (9), 1594-603.
201. Li, H.; Zheng, P., Single molecule force spectroscopy: a new tool for bioinorganic chemistry. *Curr. Opin. Chem. Biol.* **2018,** *43*, 58-67.
202. Zheng, P.; Li, H., Direct measurements of the mechanical stability of zinc-thiolate bonds in rubredoxin by single-molecule atomic force microscopy. *Biophys. J.* **2011,** *101* (6), 1467-73.
203. Beedle, A. E.; Lezamiz, A.; Stirnemann, G.; Garcia-Manyes, S., The mechanochemistry of copper reports on the directionality of unfolding in model cupredoxin proteins. *Nat. Commun.* **2015,** *6*, 7894.
204. Yadav, A.; Paul, S.; Venkatramani, R.; Ainavarapu, S. R. K., Differences in the mechanical unfolding pathways of apo- and copper-bound azurins. *Sci. Rep.* **2018,** *8* (1), 1989.
205. Giannotti, M. I.; Cabeza de Vaca, I.; Artes, J. M.; Sanz, F.; Guallar, V.; Gorostiza, P., Direct Measurement of the Nanomechanical Stability of a Redox Protein Active Site and Its Dependence upon Metal Binding. *J. Phys. Chem. B* **2015,** *119* (36), 12050-8.
206. Perales-Calvo, J.; Lezamiz, A.; Garcia-Manyes, S., The Mechanochemistry of a Structural Zinc Finger. *J. Phys. Chem. Lett.* **2015,** *6* (17), 3335-40.
207. Wei, W.; Sun, Y.; Zhu, M.; Liu, X.; Sun, P.; Wang, F.; Gui, Q.; Meng, W.; Cao, Y.; Zhao, J., Structural Insights and the Surprisingly Low Mechanical Stability of the Au-S Bond in the Gold-Specific Protein GolB. *J. Am. Chem. Soc.* **2015,** *137* (49), 15358-61.
208. Cao, Y.; Li, H., Dynamics of protein folding and cofactor binding monitored by single-molecule force spectroscopy. *Biophys. J.* **2011,** *101* (8), 2009-17.
209. Beinert, H., Iron-sulfur proteins: ancient structures, still full of surprises. *JBIC, J. Biol. Inorg. Chem.* **2000,** *5* (1), 2-15.
210. Zheng, P.; Li, H., Highly covalent ferric-thiolate bonds exhibit surprisingly low mechanical stability. *J. Am. Chem. Soc.* **2011,** *133* (17), 6791-8.

211. Lei, H.; Guo, Y.; Hu, X.; Hu, C.; Hu, X.; Li, H., Reversible Unfolding and Folding of the Metalloprotein Ferredoxin Revealed by Single-Molecule Atomic Force Microscopy. *J. Am. Chem. Soc.* **2017,** *139* (4), 1538-44.
212. Zheng, P.; Takayama, S. J.; Mauk, A. G.; Li, H., Hydrogen bond strength modulates the mechanical strength of ferric-thiolate bonds in rubredoxin. *J. Am. Chem. Soc.* **2012,** *134* (9), 4124-31.
213. Zheng, P.; Chou, C. C.; Guo, Y.; Wang, Y.; Li, H., Single molecule force spectroscopy reveals the molecular mechanical anisotropy of the FeS4 metal center in rubredoxin. *J. Am. Chem. Soc.* **2013,** *135* (47), 17783-92.
214. Zheng, P.; Takayama, S. J.; Mauk, A. G.; Li, H., Single molecule force spectroscopy reveals that iron is released from the active site of rubredoxin by a stochastic mechanism. *J. Am. Chem. Soc.* **2013,** *135* (21), 7992-8000.
215. Zheng, P.; Arantes, G. M.; Field, M. J.; Li, H., Force-induced chemical reactions on the metal centre in a single metalloprotein molecule. *Nat. Commun.* **2015,** *6*, 7569.
216. Bartsch, R. G., Purification of (4Fe-4S)1--2--ferredoxins (high-potential iron--sulfur proteins) from bacteria. *Methods Enzymol.* **1978,** *53*, 329-40.
217. Ciurli, S.; Musiani, F., High potential iron-sulfur proteins and their role as soluble electron carriers in bacterial photosynthesis: tale of a discovery. *Photosynth. Res.* **2005,** *85* (1), 115-31.
218. Liu, L.; Nogi, T.; Kobayashi, M.; Nozawa, T.; Miki, K., Ultrahigh-resolution structure of high-potential iron-sulfur protein from Thermochromatium tepidum. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002,** *58* (Pt 7), 1085-91.
219. Hirano, Y.; Takeda, K.; Miki, K., Charge-density analysis of an iron-sulfur protein at an ultra-high resolution of 0.48 A. *Nature* **2016,** *534* (7606), 281-4.
220. Carter, C. W., Jr.; Kraut, J.; Freer, S. T.; Nguyen Huu, X.; Alden, R. A.; Bartsch, R. G., Two-Angstrom crystal structure of oxidized Chromatium high potential iron protein. *J Biol Chem* **1974,** *249* (13), 4212-25.
221. Ohno, H.; Takeda, K.; Niwa, S.; Tsujinaka, T.; Hanazono, Y.; Hirano, Y.; Miki, K., Crystallographic characterization of the high-potential iron-sulfur protein in the oxidized state at 0.8 A resolution. *PLoS One* **2017,** *12* (5), e0178183.
222. Ladner, C. L.; Yang, J.; Turner, R. J.; Edwards, R. A., Visible fluorescent detection of proteins in polyacrylamide gels without staining. *Anal Biochem* **2004,** *326* (1), 13-20.
223. Dey, A.; Jenney, F. E., Jr.; Adams, M. W.; Babini, E.; Takahashi, Y.; Fukuyama, K.; Hodgson, K. O.; Hedman, B.; Solomon, E. I., Solvent tuning of electrochemical potentials in the active sites of HiPIP versus ferredoxin. *Science* **2007,** *318* (5855), 1464-8.
224. Rose, K.; Shadle, S. E.; Eidsness, M. K.; Kurtz, D. M.; Scott, R. A.; Hedman, B.; Hodgson, K. O.; Solomon, E. I., Investigation of Iron− Sulfur Covalency in Rubredoxins and a Model System Using Sulfur K-Edge X-ray Absorption Spectroscopy. *J. Am. Chem. Soc.* **1998,** *120* (41), 10743-7.
225. Anxolabehere-Mallart, E.; Glaser, T.; Frank, P.; Aliverti, A.; Zanetti, G.; Hedman, B.; Hodgson, K. O.; Solomon, E. I., Sulfur K-edge X-ray absorption spectroscopy of 2Fe-2S ferredoxin: covalency of the oxidized and reduced 2Fe forms and comparison to model complexes. *J. Am. Chem. Soc.* **2001,** *123* (23), 5444-52.
226. Evans, E., Probing the relation between force--lifetime--and chemistry in single molecular bonds. *Annu. Rev. Biophys. Biomol. Struct.* **2001,** *30*, 105-28.

227. Oberhauser, A. F.; Marszalek, P. E.; Erickson, H. P.; Fernandez, J. M., The molecular elasticity of the extracellular matrix protein tenascin. *Nature* **1998,** *393* (6681), 181-5.
228. Rief, M.; Fernandez, J. M.; Gaub, H. E., Elastically coupled two-level systems as a model for biopolymer extensibility. *Phys. Rev. Lett.* **1998,** *81* (21), 4764-67.
229. Carter, C. W., Jr.; Kraut, J.; Freer, S. T.; Alden, R. A.; Sieker, L. C.; Adman, E.; Jensen, L. H., A comparison of Fe 4 S 4 clusters in high-potential iron protein and in ferredoxin. *Proc. Natl. Acad. Sci. U. S. A.* **1972,** *69* (12), 3526-9.
230. Mizrahi, I. A.; Wood, F. E.; Cusanovich, M. A., Oxidation-reduction properties of Chromatium vinosum high potential iron-sulfur protein. *Biochemistry* **1976,** *15* (2), 343-8.
231. Stephens, P. J.; Thomson, A. J.; Keiderling, T. A.; Rawlings, J.; Rao, K. K.; Hall, D. O., Cluster characterization in iron-sulfur proteins by magnetic circular dichroism. *Proc. Natl. Acad. Sci. U. S. A.* **1978,** *75* (11), 5273-5.
232. Bruser, T.; Yano, T.; Brune, D. C.; Daldal, F., Membrane targeting of a folded and cofactor-containing protein. *Eur J Biochem* **2003,** *270* (6), 1211-21.
233. Brockwell DJ, P. E., Zinober RC, Beddard GS, Olmsted PD, Smith DA, Perham RN, Radford SE., Pulling geometry defines the mechanical resistance of a beta-sheet protein. *Nat Struct Biol.* **2003,** *10*, 731-737.
234. Carrion-Vazquez M, L. H., Lu H, Marszalek PE, Oberhauser AF, Fernandez JM., The mechanical stability of ubiquitin is linkage dependent. *Nat Struct Biol.* **2003,** *10*, 738-743.
235. Kurth, J. M.; Brito, J. A.; Reuter, J.; Flegler, A.; Koch, T.; Franke, T.; Klein, E. M.; Rowe, S. F.; Butt, J. N.; Denkmann, K.; Pereira, I. A.; Archer, M.; Dahl, C., Electron Accepting Units of the Diheme Cytochrome c TsdA, a Bifunctional Thiosulfate Dehydrogenase/Tetrathionate Reductase. *J. Biol. Chem.* **2016,** *291* (48), 24804-24818.
236. Cao, Y.; Li, H., Polyprotein of GB1 is an ideal artificial elastomeric protein. *Nat. Mater.* **2007,** *6* (2), 109-14.
237. Beedle, A. E.; Lezamiz, A.; Stirnemann, G.; Garcia-Manyes, S., The mechanochemistry of copper reports on the directionality of unfolding in model cupredoxin proteins. *Nat Commun* **2015,** *6*, 7894.
238. Zheng, P.; Wang, Y.; Li, H., Reversible unfolding-refolding of rubredoxin: a single-molecule force spectroscopy study. *Angew Chem Int Ed Engl* **2014,** *53* (51), 14060-3.
239. Beinert, H., Iron-sulfur proteins: ancient structures, still full of surprises. *J Biol Inorg Chem* **2000,** *5* (1), 2-15.
240. Zhang, X.; Halvorsen, K.; Zhang, C. Z.; Wong, W. P.; Springer, T. A., Mechanoenzymatic cleavage of the ultralarge vascular protein von Willebrand factor. *Science* **2009,** *324* (5932), 1330-4.
241. Nauli, S.; Kuhlman, B.; Le Trong, I.; Stenkamp, R. E.; Teller, D.; Baker, D., Crystal structures and increased stabilization of the protein G variants with switched folding pathways NuG1 and NuG2. *Protein Sci* **2002,** *11* (12), 2924-31.
242. He, C.; Hu, C.; Hu, X.; Hu, X.; Xiao, A.; Perkins, T. T.; Li, H., Direct Observation of the Reversible Two-State Unfolding and Refolding of an alpha/beta Protein by Single-Molecule Atomic Force Microscopy. *Angew Chem Int Ed Engl* **2015,** *54* (34), 9921-5.
243. Prakash, S.; Sundd, M.; Guptasarma, P., The key to the extraordinary thermal stability of P. furiosus holo-rubredoxin: iron binding-guided packing of a core aromatic cluster responsible for high kinetic stability of the native structure. *PLoS One* **2014,** *9* (3), e89703.

244. Zartler, E. R.; Jenney, F. E., Jr.; Terrell, M.; Eidsness, M. K.; Adams, M. W.; Prestegard, J. H., Structural basis for thermostability in aporubredoxins from Pyrococcus furiosus and Clostridium pasteurianum. *Biochemistry* **2001,** *40* (24), 7279-90.
245. Strop, P.; Mayo, S. L., Rubredoxin variant folds without iron. *J Am Chem Soc* **1999,** *121* (11), 2341-2345.
246. Li, J.; Li, H., Mechanical Unfolding Pathway of the High-Potential Iron-Sulfur Protein Revealed by Single-Molecule Atomic Force Microscopy: Toward a General Unfolding Mechanism for Iron-sulfur Proteins. *J Phys Chem B* **2018,** *122* (40), 9340-9349.
247. Yeh, S. R.; Rousseau, D. L., Folding intermediates in cytochrome c. *Nat Struct Biol* **1998,** *5* (3), 222-8.
248. Palm-Espling, M. E.; Niemiec, M. S.; Wittung-Stafshede, P., Role of metal in folding and stability of copper proteins in vitro. *Biochim Biophys Acta* **2012,** *1823* (9), 1594-603.
249. Boys, B. L.; Konermann, L., Folding and assembly of hemoglobin monitored by electrospray mass spectrometry using an on-line dialysis system. *J Am Soc Mass Spectrom* **2007,** *18* (1), 8-16.
250. Meyer, J., Iron-sulfur protein folds, iron-sulfur chemistry, and evolution. *J Biol Inorg Chem* **2008,** *13* (2), 157-70.
251. Fukuyama, K., Structure and function of plant-type ferredoxins. *Photosynth Res* **2004,** *81* (3), 289-301.
252. Binda, C.; Coda, A.; Aliverti, A.; Zanetti, G.; Mattevi, A., Structure of the mutant E92K of [2Fe-2S] ferredoxin I from Spinacia oleracea at 1.7 A resolution. *Acta Crystallogr D Biol Crystallogr* **1998,** *54* (Pt 6 Pt 2), 1353-8.
253. Hasumi, H.; Nakamura, S.; Koga, K.; Yoshizumi, H., Effects of neutral salts on thermal stability of spinach ferredoxin. *Biochem Biophys Res Commun* **1979,** *87* (4), 1095-101.
254. Pagani, S.; Vecchio, G.; Iametti, S.; Bianchi, R.; Bonomi, F., On the role of the 2Fe-2S cluster in the formation of the structure of spinach ferredoxin. *Biochim Biophys Acta* **1986,** *870* (3), 538-544.
255. Pagani, S.; Bonomi, F.; Cerletti, P., Enzymic synthesis of the iron-sulfur cluster of spinach ferredoxin. *Eur J Biochem* **1984,** *142* (2), 361-6.
256. Wu, S. P.; Cowan, J. A., Iron-sulfur cluster stability. Kinetics and mechanism of ligand-promoted cluster degradation. *Chem Commun (Camb)* **2007,** (1), 82-4.
257. Teixeira, M. H.; Curtolo, F.; Camilo, S. R. G.; Field, M. J.; Zheng, P.; Li, H.; Arantes, G. M., Modeling the Hydrolysis of Iron-Sulfur Clusters. *J Chem Inf Model* **2020,** *60* (2), 653-660.
258. Balk, J.; Lobréaux, S., Biogenesis of iron-sulfur proteins in plants. *Trends Plant Sci* **2005,** *10* (7), 324-31.
259. Dutkiewicz, R.; Schilke, B.; Cheng, S.; Knieszner, H.; Craig, E. A.; Marszalek, J., Sequence-specific interaction between mitochondrial Fe-S scaffold protein Isu and Hsp70 Ssq1 is essential for their in vivo function. *J Biol Chem* **2004,** *279* (28), 29167-74.
260. Vickery, L. E.; Cupp-Vickery, J. R., Molecular chaperones HscA/Ssq1 and HscB/Jac1 and their roles in iron-sulfur protein maturation. *Crit Rev Biochem Mol Biol* **2007,** *42* (2), 95-111.
261. Andrew, A. J.; Dutkiewicz, R.; Knieszner, H.; Craig, E. A.; Marszalek, J., Characterization of the interaction between the J-protein Jac1p and the scaffold for Fe-S cluster biogenesis, Isu1p. *J Biol Chem* **2006,** *281* (21), 14580-7.

262. Barnum, D. W., Hydrolysis of cations. Formation constants and standard free energies of formation of hydroxy complexes. *Inorganic Chemistry* **1983,** *22* (16), 2297-2305.
263. Shimizu, T.; Lengalova, A.; Martínek, V.; Martínková, M., Heme: emergent roles of heme in signal transduction, functional regulation and as catalytic centres. *Chem Soc Rev* **2019,** *48* (24), 5624-5657.
264. Salemme, F. R., Structure and function of cytochromes c. *Annu Rev Biochem* **1977,** *46*, 299-329.
265. Bertini, I.; Cavallaro, G.; Rosato, A., Cytochrome c: occurrence and functions. *Chem Rev* **2006,** *106* (1), 90-115.
266. Maity, H.; Maity, M.; Englander, S. W., How cytochrome c folds, and why: submolecular foldon units and their stepwise sequential stabilization. *J Mol Biol* **2004,** *343* (1), 223-33.
267. Bai, Y.; Sosnick, T. R.; Mayne, L.; Englander, S. W., Protein folding intermediates: native-state hydrogen exchange. *Science* **1995,** *269* (5221), 192-7.
268. Fisher, W. R.; Taniuchi, H.; Anfinsen, C. B., On the role of heme in the formation of the structure of cytochrome c. *J Biol Chem* **1973,** *248* (9), 3188-95.
269. Stellwagen, E.; Rysavy, R.; Babul, G., The conformation of horse heart apocytochrome c. *J Biol Chem* **1972,** *247* (24), 8074-7.
270. Li, J.; Li, H., Single molecule force spectroscopy reveals that a two-coordinate ferric site is critical for the folding of holo-rubredoxin. *Nanoscale* **2020,** *12* (44), 22564-22573.
271. He, C.; Hu, C.; Hu, X.; Hu, X.; Xiao, A.; Perkins, T. T.; Li, H., Direct Observation of the Reversible Two-State Unfolding and Refolding of an α/β Protein by Single-Molecule Atomic Force Microscopy. *Angew Chem Int Ed Engl* **2015,** *54* (34), 9921-5.
272. Patel, C. N.; Lind, M. C.; Pielak, G. J., Characterization of horse cytochrome c expressed in Escherichia coli. *Protein Expr Purif* **2001,** *22* (2), 220-4.
273. Abriata, L. A.; Cassina, A.; Tórtora, V.; Marín, M.; Souza, J. M.; Castro, L.; Vila, A. J.; Radi, R., Nitration of solvent-exposed tyrosine 74 on cytochrome c triggers heme iron-methionine 80 bond disruption. Nuclear magnetic resonance and optical spectroscopy studies. *J Biol Chem* **2009,** *284* (1), 17-26.
274. Oldfield, C. J.; Dunker, A. K., Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* **2014,** *83*, 553-84.
275. Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N.; Dothager, R. S.; Seifert, S.; Thiyagarajan, P.; Sosnick, T. R.; Hasan, M. Z.; Pande, V. S.; Ruczinski, I.; Doniach, S.; Plaxco, K. W., Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc Natl Acad Sci U S A* **2004,** *101* (34), 12491-6.
276. Smith, L. J.; Fiebig, K. M.; Schwalbe, H.; Dobson, C. M., The concept of a random coil. Residual structure in peptides and denatured proteins. *Fold Des* **1996,** *1* (5), R95-106.
277. Choi, U. B.; McCann, J. J.; Weninger, K. R.; Bowen, M. E., Beyond the random coil: stochastic conformational switching in intrinsically disordered proteins. *Structure* **2011,** *19* (4), 566-76.
278. Plaxco, K. W.; Gross, M., Unfolded, yes, but random? Never! *Nat Struct Biol* **2001,** *8* (8), 659-60.
279. Neri, D.; Billeter, M.; Wider, G.; Wüthrich, K., NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science* **1992,** *257* (5076), 1559-63.
280. Blanco, F. J.; Rivas, G.; Serrano, L., A short linear peptide that folds into a native stable beta-hairpin in aqueous solution. *Nat Struct Biol* **1994,** *1* (9), 584-90.

281. Baldwin, R. L., Alpha-helix formation by peptides of defined sequence. *Biophys Chem* **1995,** *55* (1-2), 127-35.
282. Baldwin, R. L.; Zimm, B. H., Are denatured proteins ever random coils? *Proc Natl Acad Sci U S A* **2000,** *97* (23), 12391-2.
283. Arai, M.; Kuwajima, K., Role of the molten globule state in protein folding. *Adv Protein Chem* **2000,** *53*, 209-82.
284. Judy, E.; Kishore, N., A look back at the molten globule state of proteins: thermodynamic aspects. *Biophys Rev* **2019,** *11* (3), 365-375.
285. Gros, G.; Wittenberg, B. A.; Jue, T., Myoglobin's old and new clothes: from molecular structure to function in living cells. *J Exp Biol* **2010,** *213* (Pt 16), 2713-25.
286. Balestrieri, C.; Colonna, G.; Giovane, A.; Irace, G.; Servillo, L., Equilibrium evidence of non-single step transition during guanidine unfolding of apomyoglobins. *FEBS Lett* **1976,** *66* (1), 60-4.
287. He, E.; Ren, W.; Wang, J.; Li, W.; Wang, W., Effects of heme binding on myoglobin folding: Coarse grained molecular simulations. *J Theor Comput Chem* **2015,** *14* (08), 1550059.
288. Elms, P. J.; Chodera, J. D.; Bustamante, C.; Marqusee, S., The molten globule state is unusually deformable under mechanical force. *Proc Natl Acad Sci U S A* **2012,** *109* (10), 3796-801.
289. Moczygemba, C.; Guidry, J.; Wittung-Stafshede, P., Heme orientation affects holo-myoglobin folding and unfolding kinetics. *FEBS Lett* **2000,** *470* (2), 203-6.
290. Wittung-Stafshede, P.; Malmström, B. G.; Winkler, J. R.; Gray, H. B., Folding of deoxymyoglobin triggered by electron transfer. *J Phys Chem A* **1998,** *102* (28), 5599-5601.
291. Yang, F.; Phillips, G. N., Jr., Crystal structures of CO-, deoxy- and met-myoglobins at various pH values. *J Mol Biol* **1996,** *256* (4), 762-74.
292. Bringas, M.; Petruk, A. A.; Estrin, D. A.; Capece, L.; Martí, M. A., Tertiary and quaternary structural basis of oxygen affinity in human hemoglobin as revealed by multiscale simulations. *Sci Rep* **2017,** *7* (1), 10926.
293. Ahmed, M. H.; Ghatge, M. S.; Safo, M. K., Hemoglobin: Structure, Function and Allostery. *Subcell Biochem* **2020,** *94*, 345-382.
294. Paoli, M.; Marles-Wright, J.; Smith, A., Structure-function relationships in heme-proteins. *DNA Cell Biol* **2002,** *21* (4), 271-80.
295. Dean, J.; Schechter, A. N., Sickle-cell anemia: molecular and cellular bases of therapeutic approaches (first of three parts). *N Engl J Med* **1978,** *299* (14), 752-63.
296. Soto, C., Protein misfolding and disease; protein refolding and therapy. *FEBS Lett* **2001,** *498* (2-3), 204-7.
297. Crapo, J. D.; Oury, T.; Rabouille, C.; Slot, J. W.; Chang, L. Y., Copper,zinc superoxide dismutase is primarily a cytosolic protein in human cells. *Proc Natl Acad Sci U S A* **1992,** *89* (21), 10405-9.
298. Rakhit, R.; Chakrabartty, A., Structure, folding, and misfolding of Cu,Zn superoxide dismutase in amyotrophic lateral sclerosis. *Biochim Biophys Acta* **2006,** *1762* (11-12), 1025-37.
299. Rumfeldt, J. A.; Lepock, J. R.; Meiering, E. M., Unfolding and folding kinetics of amyotrophic lateral sclerosis-associated mutant Cu,Zn superoxide dismutases. *J Mol Biol* **2009,** *385* (1), 278-98.

300. Lindberg, M. J.; Normark, J.; Holmgren, A.; Oliveberg, M., Folding of human superoxide dismutase: disulfide reduction prevents dimerization and produces marginally stable monomers. *Proc Natl Acad Sci U S A* **2004,** *101* (45), 15893-8.
301. Assfalg, M.; Banci, L.; Bertini, I.; Turano, P.; Vasos, P. R., Superoxide dismutase folding/unfolding pathway: role of the metal ions in modulating structural and dynamical features. *J Mol Biol* **2003,** *330* (1), 145-58.
302. Sen Mojumdar, S.; Z, N. S.; Dee, D. R.; Rouleau, L.; Anand, U.; Garen, C.; Woodside, M. T., Partially native intermediates mediate misfolding of SOD1 in single-molecule folding trajectories. *Nat Commun* **2017,** *8* (1), 1881.

Appendices

Amino acid sequences of the proteins

GB1

MDTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTE

HiPIP

AAPANAVTADDPTAIALKYNQDATKSERVAAARPGLPPEEQHCANCQFMQANVGEGD

WKGCQLFPGKLINVNGWCASWTLKAG

HiPIP(C61,75H)

AAPANAVTADDPTAIALKYNQDATKSERVAAARPGLPPEEQHCANCQFMQANVGEGD

WKGHQLFPGKLINVNGWHASWTLKAG

HiPIP(C75H)

AAPANAVTADDPTAIALKYNQDATKSERVAAARPGLPPEEQHCANCQFMQANVGEGD

WKGCQLFPGKLINVNGWHASWTLKAG

HiPIP(C61H)

AAPANAVTADDPTAIALKYNQDATKSERVAAARPGLPPEEQHCANCQFMQANVGEGD

WKGHQLFPGKLINVNGWCASWTLKAG

NuG2

MDTYKLVIVLNGTTFTYTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTE

RD

AKWVCKICGYIYDEDAGDPDNGISPGTKFEELPDDWVCPICGAPKSEFEKLED

RD(no proline)

AKWVCKICGYIYDEDAGDGDNGISGGTKFEELGDDWVCGICGAGKSEFEKLED

RD(P19G)

AKWVCKICGYIYDEDAGDGDNGISPGTKFEELPDDWVCPICGAPKSEFEKLED

RD(P33G)

AKWVCKICGYIYDEDAGDPDNGISPGTKFEELGDDWVCPICGAPKSEFEKLED

RD(P39G)

AKWVCKICGYIYDEDAGDPDNGISPGTKFEELPDDWVCGICGAPKSEFEKLED

RD(P25,44G)

AKWVCKICGYIYDEDAGDPDNGISGGTKFEELPDDWVCPICGAGKSEFEKLED

sFD

AAYKVTLVTPTGNVEFQCPDDVYILDAAEEEGIDLPYSCRAGSCSSCAGKLKTGSLNQD

DQSFLDDDQIDEGWVLTCAAYPVSDVTIETHKEEELTA

sFD(E93C)

AAYKVTLVTPTGNVEFQCPDDVYILDAAEEEGIDLPYSCRAGSCSSCAGKLKTGSLNQD

DQSFLDDDQIDEGWVLTCAAYPVSDVTIETHKECELTA

cytc

GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGFTYTDANKNKGIT

WKEETLMEYLENPKKYIPGTKMIFAGIKKKTEREDLIAYLKKATNE

cytc(C14, 17A)

GDVEKGKKIFVQKAAQAHTVEKGGKHKTGPNLHGLFGRKTGQAPGFTYTDANKNKGI

TWKEETLMEYLENPKKYIPGTKMIFAGIKKKTEREDLIAYLKKATNE