

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



Tailoring protein nanomechanics through chemical reactivity with single bond resolution

Beedle, Amy Elizabeth McIver

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

TAILORING PROTEIN NANOMECHANICS THROUGH CHEMICAL REACTIVITY WITH SINGLE BOND RESOLUTION

Amy. E. M. Beedle

Thesis submitted for the degree of Doctor of Philosophy

2017

First Supervisor: Professor Sergi Garcia-Manyes

Second Supervisor: Professor David Richards

King's College London

DECLARATION

I declare that the work undertaken in this thesis is my own original work, except where mentioned in the text.

Amy Beedle

Life may be chemistry, but it's a special circumstance of chemistry. Organisms exist not because of reactions that are possible, but because of reactions that are barely possible. Too much reactivity and we would spontaneously combust. Too little, and we would turn cold and die. Proteins enable these barely possible reactions, allowing us to live on the edges of chemical entropy – skating perilously, but never falling in.

Siddhartha Mukherjee (The Gene: An Intimate History)

ABSTRACT

It is now well established that the overall mechanical properties of a wide variety of tissues can be described by the cumulative nanomechanical properties of each of the individual constituent proteins. However, on a deeper, atomistic level, it is far less understood how the chemical reactivity of the individual chemical bonds forming the protein's architecture can tune its nanomechanics. The amino acid cysteine is unique among all the protein amino acids due to its reactive sulfhydryl group. The particular chemical properties of the sulfur atom enable cysteine to exhibit a wide range of oxidation states, resulting in a rich and varied chemical reactivity that spans from the formation of organo-metallic bonds when coordinating a plethora of metal ions to the formation of the highly covalent and mechanically rigid disulfide bonds. It is therefore tempting to speculate that the incorporation of an individual cysteine residue into a protein sequence is a natural strategy for modulating chemical reactivity, mostly through post-translational modifications, with knock-on effects for protein mechanics. In this thesis work I employ single molecule force spectroscopy AFM to investigate how the different chemical reactivity of cryptic cysteines affects the folding and the nanomechanical properties of individual proteins. Firstly, I characterize the mechanical stability of the copper-thiolate bond in the metalloproteins azurin and plastocyanin. Furthermore, I use a constant force approach to probe two chemically distinct strategies to achieve non-enzymatic disulfide bond formation within the context of oxidative folding; through a transient sulfenic acid species or through the creation of a mixed disulfide species between a protein cysteine and a low molecular weight thiol. Altogether this work highlights that subtle adjustments to the chemical reactivity of a single thiol-bond within a protein, has unanticipated yet profound effects on the overall protein nanomechanics.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the help and support from a huge number of people.

First and foremost, I need to thank my supervisor, Sergi Garcia-Manyes. Five years ago, you took a leap of faith and accepted me into your lab. I was extremely naïve and could barely use a pipette, but your time, guidance, knowledge and encouragement has not only transformed my approach to scientific research, but the future of my life. My sincerest thank you for everything you have done over the years, from teaching me to be a perfectionist to occasionally correcting my English. Sempre tindrà la meua gratitud i el meu suport.

I am infinitely grateful to Ainhoa and Palma who are protein-making-magicians, and have truly made the work in this thesis possible! And an extra special thanks needs to go to Bill Luckhurst for making sure that everything in the lab always runs smoothly.

I need to also thank all of the people that I have shared many hours in the basement of King's with; Josep (especially for teaching me the in's and out's of the AFM), Judit (especially for training me to work with proteins), Marta (especially for building the most beautiful new lab toy), Marc (especially for not laughing at all my stupid chemistry questions), Elvira (especially for not laughing at my stupid cell questions), Jim, Bob, David, Andy, and Vinoth. One basement buddy in particular is Ash, you kept me laughing no matter what.

I need to thank someone that has not actively helped during the course of this PhD, but was fundamental in starting me on this journey. Mr Pattison, without your teaching and enthusiasm, I am not sure what I would be doing, but it certainly wouldn't be this.

I owe everything to my family. You are my rocks. Mum and dad, you have inspired me in countless ways and have always encouraged me to follow my gut instincts. I am not sure I will never be able to show you how thankful I am for everything you have done, and I certainly won't be able to repay you. Mum, thanks for the cake! Dad, thanks for the bubbles!

My Hayes crew, Anjali, David, Laura and Warren, are my non-blood family. Being friends since we were 12 means you know far too many embarrassing stories, but you can always make me laugh until I cry. To Will and Sam, thanks for the COD after a long day! And finally, thanks to Douglas. You have kept me sane and insane in equal measures, always for the good. I will always look forward to the countryside walks and tales of golf.

LIST OF PUBLICATIONS ARISING FROM THIS PhD

- *Forcing the reversibility of mechanochemical reactions*
Beedle, AEM*, Mora, M*, Davis, CT., Ambrosius, SP., Strinemann, G., Garcia-Manyes, S. (* denotes equal contribution)
Under Review.
- *Force triggers YAP entry by mechanically regulating transport across nuclear pores*
Elosegui-Artola, A., Andreu, I., Beedle, A.E.M., Lezamiz, A., Uroz, M., Kosmalska, A., Oria, R., Trepát, X., Navajas, D., Garcia-Manyes, S., Roca-Cusachs, P.
Cell 171 (6), 1397 – 1410 (2017)
- *Steering chemical reactions with force*
Garcia-Manyes, S., Beedle, A.E.M.
Nature Reviews Chemistry 1 (11), 0083 (2017)
- *Tailoring protein nanomechanics with chemical reactivity*
Beedle, AEM., Mora, M., Lynman, S., Strinemann, G., Garcia-Manyes, S.
Nature Communications 8:15658 (2017)

CHAPTER 5 (APPENDIX VI)

- *Binding of myomesin to obscurin-like-1 at the muscle M-band provides a strategy for isoform-specific mechanical protection.*
Pernigo, S., Fukuzawa, A., Beedle, AEM., Holt, M., Round, A., Pandini, A., Garcia-Manyes, S., Gautel, M., Steiner, SA.
Structure 25, 1-14 (2017)

- *Protein S-sulfenylation is a fleeting molecular switch that regulates non-enzymatic oxidative folding.* Beedle, AEM., Lynham, S., Garcia-Manyes, S. *Nature Communications* 7:12490 (2016)

CHAPTER 4 (APPENDIX V)

- *Single molecule identification of a misfolded, domain-swapped conformation in the cataract-forming human γ D-crystallin.*
Garcia-Manyes, S., Giganti, D., Badilla., C., Lezamiz, A., Perales-Calvo, J., Beedle, AEM., Fernandez, J. *The Journal of Biological Chemistry*. 291(8): 4226 – 4235 (2016)
- *Mechanobiology – chemical origin of membrane mechanical resistance and force-dependent signalling.*
Beedle, AEM., Williams, A., Relat-Goberna, J., Garcia-Manyes, S. *Current Opinion in Chemical Biology* 29, 87-93 (2015)
- *The mechanochemistry of copper reports on the directionality of unfolding in model cupredoxin proteins.*
Beedle, AEM., Lezamiz, A., Stirnemann, G., Garcia-Manyes, S. *Nature Communications* 6:7894 (2015)

CHAPTER 3 (APPENDIX IV)

Contents

DECLARATION	2
ABSTRACT	4
ACKNOWLEDGEMENTS	5
LIST OF PUBLICATIONS ARISING FROM THIS PhD	7
TABLE OF FIGURES.....	13
NOMENCLATURE	18
1. SINGLE MOLECULE BIOPHYSICS	20
1.1 CHEMISTRY UNDER FORCE	21
1.2 SINGLE MOLECULE FORCE TECHNIQUES.....	26
FORCE SPECTROSCOPY OPERATIONAL MODES	28
CONSTANT VELOCITY MODE	29
CONSTANT FORCE MODE.....	32
FUTURE OF SINGLE MOLECULE TECHNIQUES.....	36
1.3 PROTEIN MECHANOCHEMISTRY	38
INDIVIDUAL PROTEIN UNFOLDING	39
BINDING MEDIATES MECHANICS	42
MECHANICAL PROTEIN FOLDING	48
MECHANICS MEDIATES BINDING	52
DISULFIDE BOND RUPTURE	54
DISULFIDE BOND REFORMATION.....	61
1.4 SCOPE AND OBJECTIVES.....	65
2. EXPERIMENTAL TECHNIQUES	67
2.1 PROTEIN PREPARATION	68

PROTEIN ENGINEERING & PURIFICATION	68
2.2 SINGLE MOLECULE FORCE SPECTROSCOPY ATOMIC FORCE MICROSCOPE	70
BACKGROUND	70
SETTING UP AN EXPERIMENT	71
EXPERIMENTAL CONFIGURATION	72
EXPERIMENTAL CALIBRATION.....	74
USER SOFTWARE INTERFACE.....	76
FORCE EXTENSION MODE	77
FORCE CLAMP MODE	79
DATA ANALYSIS	80
FORCE EXTENSION PROTEIN UNFOLDING ANALYSIS.....	80
FORCE CLAMP ANALYSIS	80
3. THE MECHANOCHEMISTRY OF COPPER REPORTS ON THE DIRECTIONALITY OF UNFOLDING IN MODEL CUPREDOXIN PROTEINS.....	82
3.1 INTRODUCTION	83
3.2 RESULTS	86
THE NANOMECHANICS OF AZURIN UNFOLDING	86
THE DISTINCT UNFOLDING PATHWAYS REVEALED BY MUTAGENESIS.....	90
THE PLASTOCYANIN PROTEIN EXHIBITS A SIMILAR UNFOLDING SCHEME	93
MOLECULAR DYNAMICS SIMULATIONS.....	96
3.3 DISCUSSION	99

4. PROTEIN S-SULFENYLATION IS A FLEETING MOLECULAR SWITCH THAT REGULATES NON-ENZYMATIC OXIDATIVE FOLDING.....	105
4.1 INTRODUCTION	106
4.2 RESULTS	109
SULFENIC ACID TRIGGERS DISULFIDE BOND FORMATION.....	109
THE KINETICS OF DISULFIDE BOND REFORMATION	116
THE REACTIVITY OF SULFENIC ACID DEPENDS ON SOLVENT ACCESSIBILITY.....	117
DIMEDONE BINDING CONFIRMS THE PRESENCE OF SULFENIC ACID.....	122
4.3 DISCUSSION	125
5. TAILORING PROTEIN NANOMECHANICS WITH CHEMICAL REACTIVITY.....	135
5.1 INTRODUCTION	136
5.2 RESULTS	140
S-CYSTEINYLLATION PROMOTES NON-ENZYMATIC OXIDATIVE FOLDING	140
S-HOMOCYSTEINYLLATION BLOCKS OXIDATIVE FOLDING.....	143
THE MIXED DISULFIDE PLAYS A KEY ROLE IN DISULFIDE FORMATION	145
HOMOCYSTEINE EXHIBITS AN ANOMALOUS HIGH pK_a	146
DFT CALCULATIONS PROVIDE A THERMODYNAMIC EXPLANATION	148
THE DISTINCT NUCLEOPHILIC POWER OF L-CYS AND HCYS.....	150

THE EXPERIMENTAL APPROACH PREDICTS THE NATURE OF NAC	152
5.3 DISCUSSION	154
6. CONCLUSIONS & FUTURE DIRECTIONS	160
6.1 CONCLUSIONS	161
CHAPTER 3 CONCLUSIONS	161
CHAPTER 4 CONCLUSIONS	162
CHAPTER 5 CONCLUSIONS	163
6.2 FUTURE DIRECTIONS	165
APPENDIX I.....	175
APPENDIX II	199
APPENDIX III.....	218
APPENDIX IV	235
APPENDIX V	246
APPENDIX VI.....	258

TABLE OF FIGURES

CHAPTER 1

Fig 1.1	The activation energy of a chemical reaction can be lowered by the addition of a catalyst.
Fig 1.2	Single molecule force spectroscopy approaches have the power to probe a single bond within the context of a larger molecule.
Fig 1.3	The amalgamation of single molecule force spectroscopy approaches permits investigation into proteins with a wide range of nanomechanical properties.
Fig 1.4	Constant velocity measurements monitor the force as a function of extension.
Fig 1.5	Force clamp mode monitors protein elongation as a function of time at a constant force giving rise to a staircase-like trajectory.
Fig 1.6	A single protein can house a variety of covalent and non-covalent interactions that span a range of mechanical forces.
Fig 1.7	Protein structure is a key indicator of protein nanomechanics.
Fig 1.8	Ligand binding has the power to modify the mechanical stability of a protein.
Fig 1.9	Force spectroscopy can monitor a single organo-metallic bond within the context of protein unfolding.
Fig 1.10	A force-quench protocol captures individual protein folding.
Fig 1.11	Direct observation of protein folding.

- Fig 1.12** Force mediates the exposure of cryptic binding sites that can impact protein folding.
- Fig 1.13** Capturing a single bimolecular chemical reaction within the core of a protein.
- Fig 1.14** A combination of experimental and computational approaches has captured the complexity of force mediated disulfide scission.
- Fig 1.15** Single molecule force clamp spectroscopy captures the formation of an individual covalent disulfide bond.

CHAPTER 2

- Fig 2.1** Schematic representation of the molecular biology steps to produce a polyprotein for single molecule experiments.
- Fig 2.2** The atomic force microscope.
- Fig 2.3** Magnification of key components of the atomic force microscope.
- Fig 2.4** Cantilever calibration steps.
- Fig 2.5** The user controls for operating force-extension experiments.
- Fig 2.6** The user controls for operating force-clamp experiments.

CHAPTER 3

- Fig 3.1** Wt-azurin unfolds through two mechanically labile intermediates.
- Fig 3.2** Comparison of multiple azurin unfolding events containing a single intermediate.
- Fig 3.3** The Azu_{C112A} and the Azu_{C26A} mutations confirm the existence of parallel unfolding pathways.
- Fig 3.4** Structural similarity between plastocyanin and azurin.
- Fig 3.5** Plastocyanin unfolds through parallel pathways fingerprinted by the rupture of individual Cu-S and Cu-N bonds.
- Fig 3.6** MD simulations capture the distinct unfolding pathways.
- Fig 3.7** A generic kinetic partition scheme for the complex mechanical unfolding of wt-azurin and wt-plastocyanin.

CHAPTER 4

- Fig 4.1** Single molecule force-clamp spectroscopy captures the non-enzymatic reformation of individual disulfide bonds, which drive protein folding.
- Fig 4.2** Capturing misfolding events triggered by the formation of non-native disulfide bonds between contiguous Ig domains within the polyprotein chain.
- Fig 4.3** The (I27_{E24C-K55C})₈ polyprotein is able to successfully refold at pH = 12.8 when the disulfide bond remains oxidised.
- Fig 4.4** The sulfenic acid-induced reformation of a disulfide bond is independent on the position of the disulfide bond within the protein structure.
- Fig 4.5** Kinetics of disulfide bond reformation.
- Fig 4.6** S-sulfenylation of cryptic cysteines inhibits oxidative folding.

- Fig 4.7** The force-dependent conformation of the protein modulates the chemical evolution of sulfenic acid.
- Fig 4.8** Dimedone selectively targets sulfenic acid and blocks oxidative folding.
- Fig 4.9** The refolding of the oxidised (I27_{E24C-K55C})₈ polyprotein is not blocked by the presence of dimedone.
- Fig 4.10** Successful non-enzymatic oxidative folding can only be observed after the presence of an alkali-mediated reduction event, which introduces a sulfenic acid moiety in the protein in a controlled manner.
- Fig 4.11** Hydrogen peroxide induces S-sulfenylation in cryptic cysteines and promotes the formation of non-native disulfide bonds in wt-I27.
- Fig 4.12** A generic kinetic scheme for the force-triggered, conformation-dependent reactivity of sulfenic acid and its implications for oxidative protein folding.

CHAPTER 5

- Fig 5.1** Single molecule force-clamp spectroscopy captures non-enzymatic disulfide bond reformation.
- Fig 5.2** DL-homocysteine prevents successful non-enzymatic oxidative folding.
- Fig 5.3** The absence of the mixed disulfide results in a refolded protein void of the disulfide bond.
- Fig 5.4** The surprisingly high pK_a of Homocysteine leads to a thermodynamically stable asymmetric mixed-disulfide.
- Fig 5.5** The kinetics of disulfide reduction is largely accelerated by homocysteine due to its high nucleophilic power.

Fig 5.6 *N*-acetylcysteine creates a stable protein mixed-disulfide that largely blocks oxidative folding.

Fig 5.7 Protein reactivity induced by S-thiolation regulates protein nanomechanics.

CHAPTER 6

Fig 6.1 Testing the free energy of disulfide reduction for a number of chemically similar sulfur based nucleophiles.

Fig 6.2 Mechanical force can induce thermodynamically forbidden reactions.

Fig 6.3 The thermodynamic favourability of disulfide reduction has profound implications on the reverse reformation process.

Fig 6.4 A simple one dimensional energy landscape can explain the observed reaction reversibility.

NOMENCLATURE

AFM	Atomic force microscope
DFT	Density functional theory
DNA	Deoxyribonucleic acid
ER	Endoplasmic reticulum
F	Force
FRET	Förster resonance energy transfer
GFP	Green fluorescent protein
Hcys	Homocysteine
I27	27 th immunoglobulin domain of titin
I27 _{24/55}	27 th immunoglobulin domain of titin with a disulfide bond between residues 24 and 55
I27 _{32/75}	27 th immunoglobulin domain of titin with a disulfide bond between residues 32 and 75
L-cys	l-cysteine
LMW	Low molecular weight
MD	Molecular dynamics
MS	Mass spectrometry
ms	Millisecond (10^{-3} seconds)
MT	Magnetic tweezers
N	Newton
NAC	N-acetyl-cysteine
nm	Nanometre (10^{-9} metres)
nN	Nanonewton (10^{-9} newtons)
OH ⁻	Hydroxide anion
OT	Optical tweezers
<i>P</i>	Persistence length
PBS	Phosphate-buffered saline

PCR	Polymerase chain reaction
pN	Piconewton (10^{-12} newtons)
PTM	Post translational modification
s	Seconds
SH	Protonated sulfur atom
SM	Single molecule
SMD	Steered molecular dynamics
SOH	Sulfenic acid
SO ₂ H	Sulfinic acid
SO ₃ H	Sulfonic acid
SS	Disulfide bond
t _{ext}	Extended state time
t _q	Quench time
Trx	Thioredoxin

1. SINGLE MOLECULE BIOPHYSICS

1.1 CHEMISTRY UNDER FORCE

Chemical reactivity underlies our fundamental understanding of many physical and biological phenomena. In a reductionist way, a chemical reaction can be defined as the breaking and formation of individual chemical bonds. Despite such apparent simplicity, the reactivity and outcome of a given chemical reaction is governed by a hyper-dimensional energy landscape where the initial components - the reactants – and the final form - the products, are separated by high energy transition state conformations. Generally, a given reaction will only proceed once the system has enough energy to traverse this energy barrier. While statistical mechanics demonstrates that there is always the possibility to spontaneously cross this barrier, it follows that an increase in the barrier height leads to an exponential increase in reaction times¹. Consequently, these long timescales render many important reactions inaccessible in the laboratory.

A classical way to activate a chemical reaction is through the addition of a catalyst, which can speed up reaction kinetics - typically by reducing the transition state free energy (Fig. 1.1). In particular, a wide variety of investigations have utilized heat, light and electricity – underlying classical thermochemistry, photochemistry and electrochemistry – to reduce the activation free energy and accelerate the reaction. These external perturbations modify the topology of the energy landscape, such that the reaction can be studied on feasible timescales. Collectively, these experiments have enabled to establish the foundation of chemical reactivity by unveiling the reaction mechanisms and kinetics that define modern chemistry.

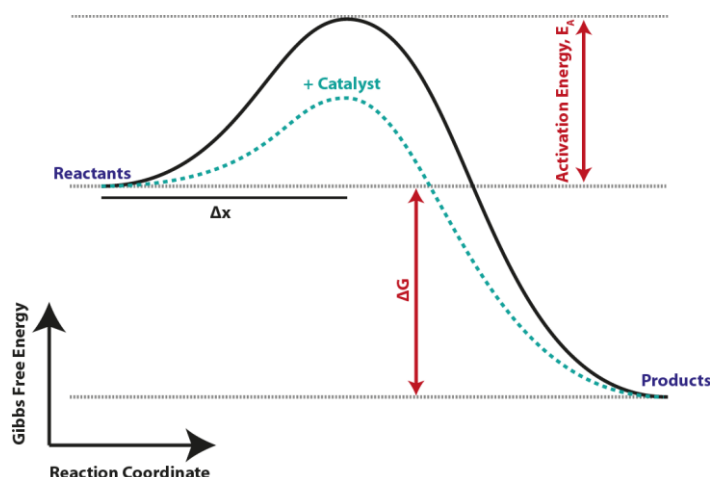


Figure 1.1. The activation energy of a chemical reaction can be lowered by the addition of a catalyst. For reactants to convert into products, they need to overcome an activation energy barrier, E_A , along a given reaction coordinate. The height of this energy barrier can be reduced if a catalyst is added which accelerates the kinetics of the reaction.

An alternative, albeit far less studied, route towards accelerating a reaction is via the application of mechanical force. While it is clear that a mechanical perturbation can accelerate a reaction – for example the grinding force generated by our teeth to break down food - very little is known about how this vectorial perturbation imparts energy to individual chemical bonds, which ultimately dictates the outcome of a reaction.

The principal reason underpinning our limited knowledge of mechanochemistry stems from the associated experimental challenges. Firstly, we require an experimental technique with the capability to apply force in a defined direction with respect to the orientation of the bond. While bulk mechanical approaches, such as sonication or ultrasound, have offered insight into the power of mechanical force at accelerating reaction rates², the molecules are distributed in random orientations and thus the vectorial component of force with respect to the individual chemical bonds is largely unknown. This problem was circumvented with the advent of single molecule force spectroscopy techniques, where force is applied to an individual molecule that is tethered between two well-defined anchoring points. This technological achievement utilizes force to tilt the free energy landscape along

a one-dimensional reaction coordinate, lowering the height of the energy barrier and accelerating the reaction. As such, this approach offers the unique and exciting possibility to apply a well-calibrated force to a highly localized set of known chemical bonds. This novel energy input drives the system to conformations that cannot be sampled in classical bulk experiments, and even has the power to sample reaction pathways that are forbidden by the laws of thermodynamics^{3, 4, 5, 6, 7}.

While a simple reaction between well-defined reactants and products can be sufficiently described by a single energy barrier, for more complex systems this is no longer satisfactory⁸. In particular, a prominent question that remains; does the reactivity of a bond in isolation recapitulate the behaviour when it is embedded within a larger, more complex environment? (Fig. 1.2). One of the most prominent approaches towards answering this question involves mechanically perturbing naturally occurring biopolymers such as DNA, polysaccharides and proteins that contain a multiplex of chemical bonds that are prone to be exposed to mechanical force^{9, 10, 11, 12}. This approach has begun to elucidate the relationship between mechanical force, polymer behaviour and the network of bonds.

In particular, proteins are ideal candidates for performing mechanochemistry studies. Firstly, proteins are rich in a wide variety of chemical bonds due to the large number of amino acid side chains. Furthermore, thanks to advances in molecular biology techniques, it is now, in principle, straightforward to either include or delete specific chemical bonds by introducing point amino acid mutations. In practice, this can prove challenging, with certain mutations causing severe structural changes or even preventing the protein from folding¹³. Finally, with the rapidly expanding number of crystal structures available with angstrom resolution, and advances in computational approaches, it is possible to infer an educated guess regarding the orientation

of the bond with respect to the direction of the force propagation. Taken together, these protein features provide a platform to study a wide variety of mechanically activated chemical reactions, ranging from non-covalent to covalent interactions and their direct connection to protein nanomechanics.

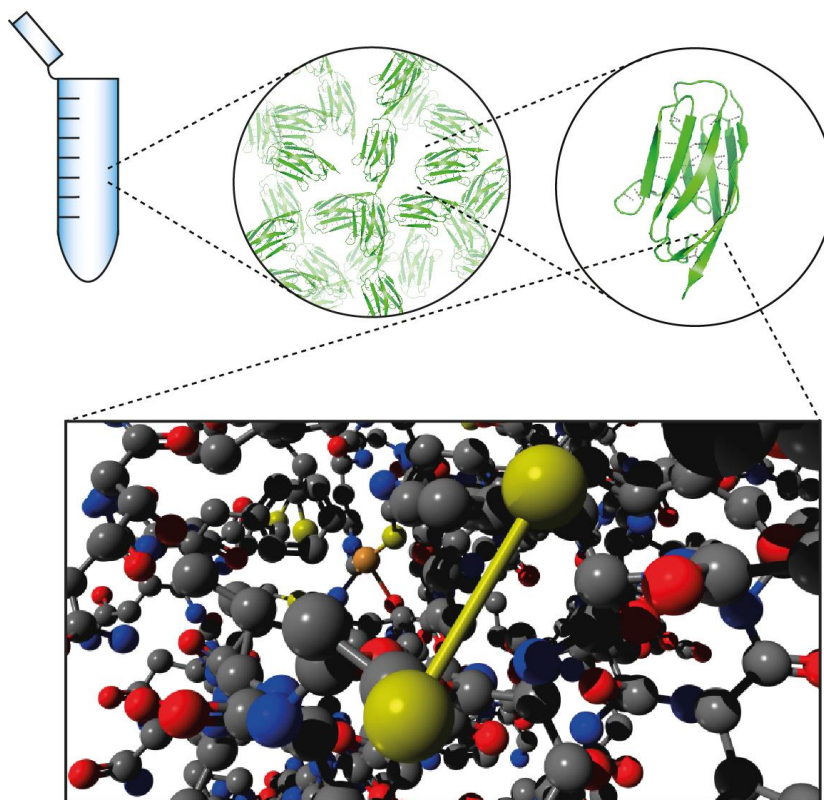


Figure 1.2. Single molecule force spectroscopy approaches have the power to probe a single bond within the context of a larger molecule. Proteins are formed of hundreds of chemical bonds that maintain the native structure. Single molecule force spectroscopy can directly monitor the mechanical response of an individual bond within this complex network.

This major achievement in biopolymer mechanochemistry seems to have important repercussions for our understanding of the nanoscale biological world. There is now accumulating evidence to imply that biology might employ a mechano-chemical strategy to regulate the mechanical properties of key individual proteins by incorporating specific chemical bonds. The giant protein titin is a prime example; as one of the main protein constituents of muscle, it is under constant mechanical stress and is primarily responsible for the passive elastic response. In particular, unstructured PEVK and N2B

regions, void of mechanically stabilizing chemical bonds, act as entropic springs extending under the applied force according to the worm-like chain model of polymer elasticity¹⁴. However, it has been shown that in times of oxidative stress, the N2B region will generate highly covalent molecular staples that transform the nanomechanics of this elastic segment into a rigid and inextensible region, demonstrating that the chemical reactivity of one bond ultimately tunes the nanomechanics of the whole protein¹⁵. Furthermore, many protein interactions that are involved in cell and bacteria adhesion are mediated by catch bonds^{16, 17}. These non-covalent bonds undergo conformational rearrangements, leading to the formation of new hydrogen bonds that ultimately resists dissociation by increasing their lifetime under an increasing force¹⁸. Finally, pathogenic bacteria incorporate stiff, covalent isopeptide bonds into the protein that constitutes the fibrilous anchors, resisting any extension under tensile force and maintaining a tight cellular adhesion to promote infection^{19, 20}. These examples point to a synergistic relationship between chemistry, mechanics and biology. Therefore by understanding how force mediates the reactivity of key chemical bonds, there is potential to shine light on the naturally employed mechanisms that regulate mechanically driven biological processes at larger length scales.

In the remainder of this introduction I shall cover the various experimental approaches used to apply forces to individual proteins. I shall then discuss the major achievements these single molecule methods have led to in the field of protein mechanochemistry.

1.2 SINGLE MOLECULE FORCE TECHNIQUES

The advent of single molecule techniques has dramatically accelerated our understanding of the biological world at the nanoscale. As such there are continuously emerging biophysical methods towards deepening our understanding of the relationship between mechanical force, chemical reactions and biological processes. Some of these techniques include the use of acoustic waves to simultaneously unravel multiple molecules, nanopores which force proteins to unfold during translocation and even a newly developed method based on DNA origami whereby the assembly of a specially designed nanostructures can stretch single molecules at a constant force^{21, 22, 23, 24, 25}.

While these approaches push the boundaries of single molecule techniques^a, the three most conventional single molecule force spectroscopy techniques remain the atomic force microscope (AFM), the magnetic tweezers (MT) and the optical tweezers (OT)²⁶. Despite their conceptual similarities, each technique has some key features that make it more suitable for studying a particular force regime, and the amalgamation of the three permits investigation spanning a wide range of forces (Fig. 1.3). In the particular case of proteins, these three experimental approaches all exploit force as the denaturant along a well-defined one-dimensional reaction coordinate, triggering the rupture of weak hydrophobic interactions in the $\sim$ pN regime

^a The single molecule “umbrella” term extends to include fluorescent approaches. Fluorescent processes promotes an electron to a higher energy state upon interaction with a photon. Upon returning to the ground state, the acquired energy dissipates in the form of a lower energy photon. These spectroscopic methods are extremely sensitive to changes in the local chemical environment and are also able to gain insight into structural information. In particular, Förster resonance energy transfer (FRET), where the excited electron of one fluorophore transfers the energy to a second fluorophore when the distance between them is sufficiently close, can be described as a molecular ruler. This single molecule distance sensitivity makes FRET an extremely powerful alternative method to investigate protein folding and protein-protein interactions.

up to the scission of more mechanically resilient covalent bonds ($\sim$ nN) that reside in their structure^{27, 28}.

These experimental methods are now routinely performed alongside computational approaches to gain an atomistic insight into the unfolding process. In particular, steered molecular dynamics simulations (SMD) are often referred to as “computational microscopes” and replicate the force spectroscopy approach by mechanically perturbing proteins²⁹. This *in silico* approach is capable of directly visualising the structural conformations visited by the molecule along the force-denatured unfolding pathway.

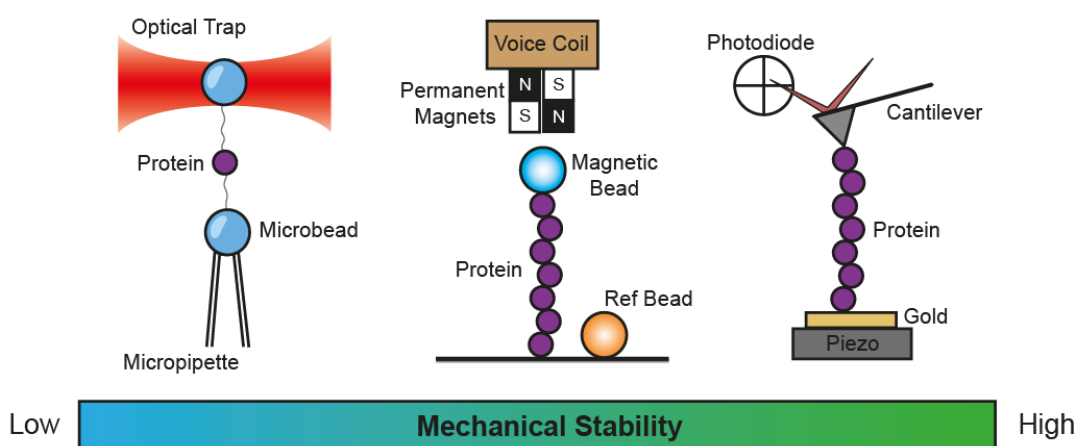


Figure 1.3. The amalgamation of single molecule force spectroscopy approaches permits investigation into proteins with a wide range of nanomechanical properties. (from left to right; optical tweezers, magnetic tweezers and the atomic force microscope)

FORCE SPECTROSCOPY OPERATIONAL MODES

Despite the atomic force microscope, magnetic tweezers and optical tweezers all using a different approach to apply force – mechanical, magnetic and radiation pressure respectively – the general working principles are all surprisingly similar²⁶. Typically, one end of the molecule is attached to a surface and the other end is attached to the ‘probe’ – the AFM cantilever tip, the magnetic bead or the bead stuck in the optical trap in each case. As the distance between these two surfaces is increased, the tethered molecule experiences a mechanical perturbation. At some point, the bonds that stabilize the tertiary structure can no longer resist the force, which typically elicits an increase in the proteins contour length as a consequence of bond rupture. The outcome of this event is measured by tracking the elongation of the molecule along the direction of the applied force. This mechanical perturbation can be applied in two distinct ways, enabling two distinct experimental read-outs. The first and most common method is stretching the protein by separating the two tethering surfaces at a constant velocity – coined ‘force extension’^{30,31}. The second approach is to maintain a constant applied force throughout the entirety of the trajectory, typically referred to as ‘force clamp’³².

Given that each protein unfolding event typically presents itself as a single ‘rip’ in the experimental trajectory, a common strategy for the acquisition of robust and reliable single molecule data is achieved by specifically designing and engineering polyprotein constructs, containing repeats of a single protein or an arrangement of a variety of proteins³³. This method provides a clear repetitive pattern; not only increasing data acquisition and giving improved statistics, but enabling the user to distinguish protein unfolding events from any unwanted spurious interactions between the cantilever tip and the gold coated surface.

CONSTANT VELOCITY MODE

Constant velocity mode applies a mechanical force to a molecule by separating the tethering surfaces at a constant velocity and measuring the force as a function of extension. As the two surfaces separate, a mechanical force propagates along the polypeptide backbone of each individual protein domain. At some point, the bonds stabilizing the protein structure are no longer able to withstand the mechanical stress and rupture, giving rise to protein unfolding hallmarked by a sharp peak in the trajectory. The apex of the peak corresponds to the protein unfolding force at that particular pulling velocity and the distance between consecutive peaks is indicative of the number of amino acids that have been released and extended upon unfolding. This cycle repeats until all domains have fully unfolded and are extended to their full contour length, giving rise to the now common ‘saw-tooth’ pattern unfolding trajectory (Fig. 1.4). The resulting force-extension trajectory is typically fitted with the worm-like chain (WLC) model of polymer elasticity; a model that describes the behaviour of semi-flexible polymers under force³⁴. It assumes that in the absence of force, the polymer will reside in a collapsed, random configuration. However as the polymer is stretched, there is a reduction in the number of possible conformations which results in an entropic force resisting the external elongation. The force dependent behaviour of a stretched polymer is well described by eq. 1.1:

$$F(x) = \frac{k_B T}{p} \left[\frac{1}{4} \left(1 - \frac{x}{l_c} \right)^{-2} + \frac{x}{l_c} - \frac{1}{4} \right] \quad (1.1)$$

where F represents force, x is the distance, p represents the persistence length - which indicates the flexibility of the protein, l_c is the contour length, k_B is the Boltzmann constant and T is temperature.

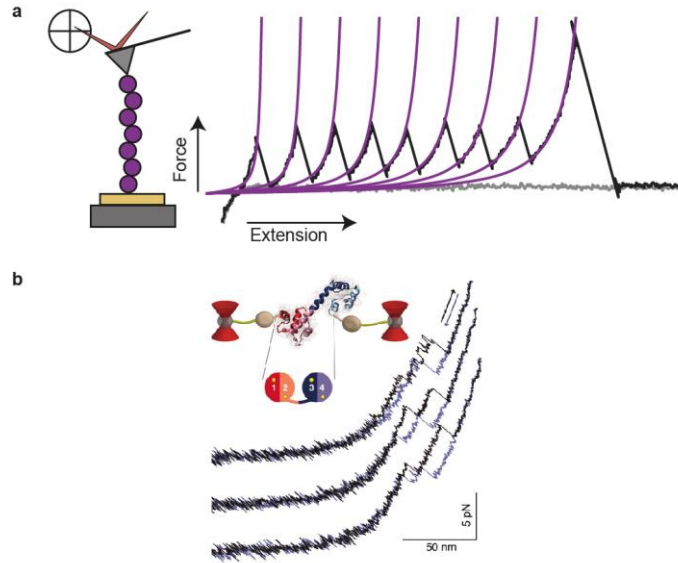


Figure 1.4. Constant velocity measurements monitor the force as a function of extension. (a) A typical saw-tooth pattern unfolding trajectory obtained in the AFM, where each peak corresponds to the unfolding of an individual I27 protein in the engineered polypeptide chain. Each unfolding event is fitted with the worm like chain model of polymer elasticity (purple lines). (b) A constant velocity measurement obtained by optical tweezers. The covalent tethers permits repeated unfolding cycles of the same calmodulin protein (reproduced from ³⁵).

This force extension mode provides valuable insight into the mechanical response of individual proteins. In particular it has captured the unfolding and folding behaviour of a range of proteins with a plethora of different biological functions³⁶. Furthermore, the versatility of the approach has made this an incredibly powerful tool for probing the nanomechanical properties of a broad range chemical bonds that are housed within the protein, including metallo-organic bonds and disulfide bonds^{37, 38, 39}. I shall expand on these exciting discoveries later in this introduction.

Despite the plethora of information that the force-extension mode has enabled, there are a number of fundamental drawbacks. Firstly, in this approach, protein unfolding is a probabilistic event that depends on the pulling velocity, where higher unfolding forces are observed at faster pulling velocities according to equation 1.2.

$$F(v) = \frac{k_B T}{\Delta x} \ln \frac{v}{\alpha_0 \Delta x} \quad (1.2)$$

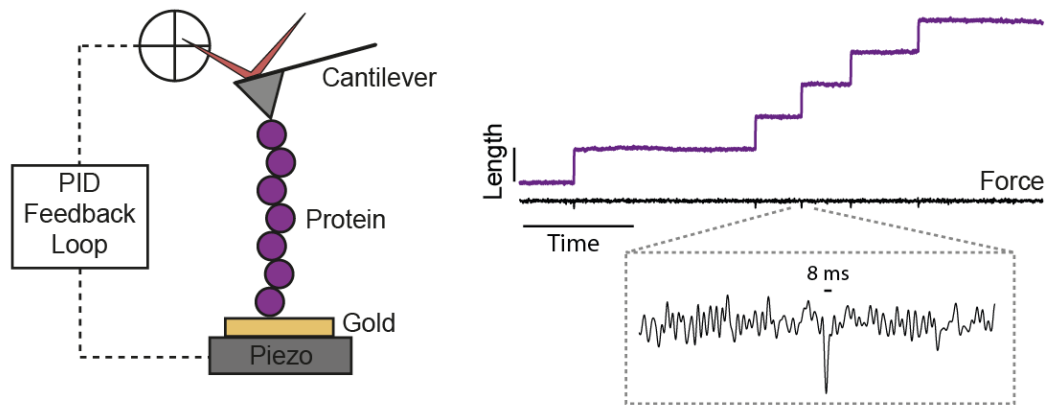
where v is the pulling velocity, Δx is the distance to the unfolding transition state and α_0 is the spontaneous rate of protein folding in the absence of force. Therefore, depending on Δx , the mechanical properties of the protein can substantially vary depending on the loading rate. Furthermore, throughout a single trajectory, the force and loading rate experienced by each protein changes unpredictably over very short timescales landscape, making it challenging to infer information regarding the unfolding energy and virtually impossible to obtain force dependent kinetics as there is little control of how the molecular conformation changes with time. Finally, in this constant velocity approach, the force induced conformational changes of the protein are typically extremely quick, and as such they can be challenging to observe. These limitations present a clear hindrance to the progression of our understanding of protein mechanics, triggering the Fernandez lab to develop a method capable of applying a constant force to a tethered protein which circumvents these issues³².

CONSTANT FORCE MODE

In constant force mode - as the name suggests - the mechanical force applied to the protein is held constant throughout the experiment. This approach follows the structural changes of a protein in time along a reaction coordinate – the end-to-end length of the molecule. The main method of applying a constant force is through a constant-force-feedback^b. Initially, the two surfaces that tether the protein separate until the applied force matches the user-defined set-point, and at this point the position of the surfaces are held constant. When one of the proteins in the chain unfolds, the force across the whole molecule decreases. Therefore, in order to compensate for this drop in force, the feedback moves the position of one of the tethering surfaces until the original force is restored. This increase in the surface separation distance is a direct reflection of the force-induced conformational change of the protein, and as such each step in the resulting ‘staircase-like’ unfolding trajectory hallmarks the unfolding of a single protein (fig. 1.5).

^b An alternative to the constant-force-feedback approach is the constant-trap-position approach used in optical tweezers. In this case, the position of the laser trap remains constant, and when the tethered molecule unfolds/folds the force decreases/increases. The resulting trajectory follows the resulting force across the molecule as a function of time, and is a direct reflection of the conformational state of the protein. While this technique has been used for a number of studies investigating the fast conformational switching of proteins due to its high temporal resolution, it is not strictly force-clamp, and as such I shall not discuss it further in this section.

Active Feedback Force-Clamp:



Passive Force-Clamp:

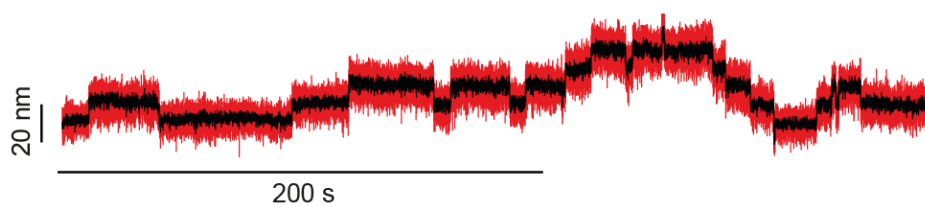


Figure 1.5. Force clamp mode monitors protein elongation as a function of time at a constant force giving rise to a staircase-like trajectory. (Active Feedback Force-Clamp) The atomic force microscope operates with an active feedback loop between the photodiode and the piezo that compensates for the molecular elongation upon a protein folding event. In this configuration, the unfolding event leads to a momentary deviation from the set-point force (inset) that typically lasts $\sim 2 - 8$ ms. During this time, the protein no longer experiences the constant force. **(Passive Force-Clamp)** In passive force-clamp, a constant force is maintained upon small elongations of the protein. This is ideal for equilibrium experiments, where the protein hops between the unfolded and folded state.

The slight drawback of this active feedback approach is that during the mechanical movement of the surface, there is a small amount of time ($\sim 2 - 8$ ms) that the protein is not under a constant force, clearly shown in the force channel in figure 1.5. While in many unfolding and folding experiments this short feedback response time does not pose any hindrance to the measurement, in some cases this reduced temporal resolution of the configuration can become an issue. In particular, capturing the rapid conformational changes of a protein in equilibrium requires accurate measurement of the lifetime of the unfolded and folded state, especially when extracting information about the underlying free energy landscape⁴⁰.

The temporal resolution can be improved with the use of a passive-force-clamp rather than an active-force-clamp. This approach does not require any

mechanical movement of the set-up to compensate for a protein unfolding event. Instead, the force remains constant upon a short elongation of the protein. For example, in optical tweezers, there is an anharmonic region of the trapping potential where the force is constant for small displacements (~50 nm) of the bead⁴¹. This passive force clamp approach allows very precise measurements of fast molecular conformational changes, which can be challenging in experimental configurations that require a feedback response loop. The slight drawback of this approach is that the 'allowed elongation' region which maintains the constant force is very narrow, which excludes a large number of proteins from being studied with this approach. Altogether, there is always a balance between the temporal and spatial resolution in force clamp, and the suitability of an approach depends on the timescales involved in the measurement²⁶.

These experimental developments, with the capability to apply a precise well-defined force to a single molecule, open a multiplex of opportunities. Firstly, contrasting the constant velocity mode where each molecule is extended to its full contour length, force clamp gives precise control over the structural conformation of the protein brought by the force, determined by the worm-like chain model⁴². This means that the degree of protein extension can be modulated, which gives precise control over the exposure of cryptic binding sites to the surrounding aqueous solution^{43, 44}. Furthermore, such a precise control over the force provides the ability to slow down the rate of force-dependent conformational changes and separate events in time, providing additional resolution into the unfolding pathway of proteins⁴⁵. Finally, the ability to monitor the molecular extension as a function of time at a particular force, grants access to the force dependent kinetics of a process, including protein unfolding, binding interactions and chemical reactions, which in turn

gives important insight into the underlying free-energy landscape of the protein^{28, 46, 47, 48}.

One of the most common approaches towards extracting the underlying free energy landscape along the one-dimensional reaction from constant force data is provided by the Bell-model^{28, 49, 50}. The Bell-model was first developed in 1978 by George Bell to describe how the lifetime of individual bonds is modulated by an external applied force. It assumes that the mechanical force reduces the unfolding barrier by $\Delta E - F\Delta x$ and that a mechanically ruptured bond cannot reform because the force will move them too far apart - therefore the only relevant kinetics for the process is the rate of dissociation. It follows that the dissociation kinetics are exponentially dependent on the force following equation 1.3:

$$\alpha(F) = \alpha(0) \exp\left(\frac{-F \Delta x}{k_B T}\right) \quad (1.3)$$

where $\alpha(F)$ is the rate of rupture at force F , $\alpha(0)$ is the rate of rupture in the absence of force, Δx is the distance to the reaction transition state along the reaction coordinate, k_B is Boltzmann's constant and T is absolute temperature. This simple model assumes that the Δx is invariant to the applied force, and while there have been considerable efforts to devise more complex descriptions of how force can modify the distance to the transition state⁵¹, the Bell model remains a popular approach towards analysing single molecule mechanical protein unfolding. Furthermore, the conjunction with Arrhenius equation - that describes the kinetics of a simple two-state reaction that needs to overcome an activation energy barrier - it is possible to obtain the height of the activation energy barrier, E_a , in the absence of force (equation 1.4).

$$\alpha(0) = A \exp\left(\frac{-E_a}{k_B T}\right) \quad (1.4)$$

where A is the pre-exponential factor / the attempt frequency. A value for A was experimentally determined by measuring the rate of mechanical protein unfolding at a range of temperatures (5 – 45 °C), finding a value of $4 \times 10^9 \text{ s}^{-1}$ for the model protein ubiquitin⁵².

Altogether, this constant force approach has transformed our understanding of the dynamics of individual proteins under the effect of an external mechanical perturbation, with fundamental insights into the mechanisms that govern mechanical protein unfolding and folding^{47, 53, 54}.

FUTURE OF SINGLE MOLECULE TECHNIQUES

While single molecule force spectroscopy techniques facilitate capturing rare and/or transitory events during protein unfolding/folding with unprecedented resolution, the next challenge is to put these discoveries into the broader and larger-scale context. On one hand, the results at the single molecule level provide insight into the relationship between molecular architecture and mechanical stability, and as such they have the potential to aid in rationally designing mechanically responsive materials^{55, 56, 57, 58, 59}. Intriguingly, recent studies using single molecule approaches have demonstrated that overall mechanical properties of a material can be adequately explained by considering the nanomechanical properties of the individual components. Pioneering single molecule experiments from the Fernandez lab demonstrated that the overall mechanics of the giant muscle protein titin can be well described by considering only the mechanics of the individual immunoglobulin domains⁶⁰, this concept was further reinforced in novel experiments from the Li lab where the mechanical properties of muscles were recapitulated macroscopically by protein hydrogels where the subunits were carefully selected⁶¹. Finally, Chung and co-workers used engineered titin-mimicking modular polymers to directly probe the relationship between the single molecule mechanics and those observed in the bulk⁶². Altogether,

these concepts highlight that SM approaches have the power to inspire new applications towards the construction of new and novel materials.

On the other hand, another challenging endeavour is to understand whether the results at the single molecule level recapitulate the behaviour of proteins that are physiologically exposed to mechanical stress *in vivo*⁶³. While SM approaches have successfully revealed the mechanical behaviour of isolated proteins in simple chemical buffers, the chemical environment of the cell is complex. Therefore it is not straightforward that the nanomechanical properties of a protein at the SM level will directly translate into the behaviour within a cell. There are a number of approaches that are now aiming to tackle this question, including force sensors that utilize Förster resonance energy transfer (FRET) to elucidate the magnitude of the forces that are applied *in vivo*^{64, 65}. These experiments have demonstrated that some proteins are driven to partially or fully unfolded conformations due to physiological forces^{66, 67}. The next major challenge towards truly understanding how proteins respond to mechanical stress is to integrate a single molecule stretching device within a cell in order to directly probe the nanomechanical of proteins in their native environment.

1.3 PROTEIN MECHANOCHEMISTRY

The nanomechanics of a single protein is dictated by a variety of distinct molecular mechanisms. Here I shall cover a few of the most prevalent strategies by which the nanomechanical properties of a single protein can be modified. First of all the (i) **unfolding of a single protein by force** switches the rigid native state into an elastic and compliant unfolded conformation. Secondly, a strategy to tune a proteins nanomechanical response is through (ii) the **binding of a cofactor** to the native state of the protein that can tune the overall resistance to a mechanical perturbation. Furthermore, the reverse process of (iii) **protein folding**, transitions the extended polymer back to a folded structure with mechanical resilience. This protein folding process can be disrupted upon (iv) **cofactor binding to cryptic sites**, which become exposed upon a mechanical perturbation, with serious implications for the subsequent mechanical properties of the folding protein. Finally, a combination of mechanical force and chemical reactivity dictates both the (v) **rupture** and (vi) **reformation of individual covalent disulfide bonds** within the core of a protein, with profound knock on effects for both protein mechanics and function (Fig. 1.6).

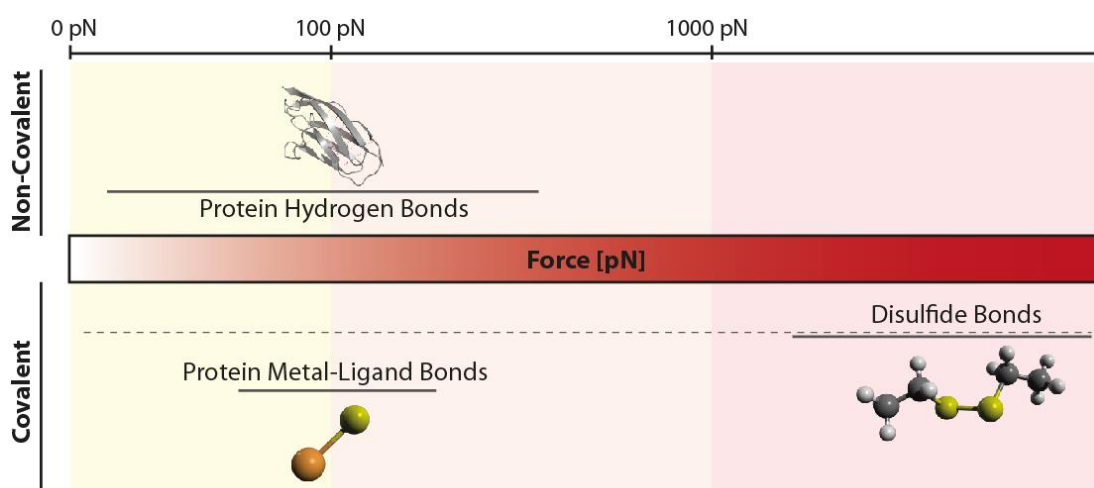


Figure 1.6. A single protein can house a variety of covalent and non-covalent interactions that span a range of mechanical forces. The dashed line for disulfide bonds represents that the addition of a nucleophile can modulate the apparent mechanical properties of the bond.

INDIVIDUAL PROTEIN UNFOLDING

A simple inspection of a protein's native structure can give rough indications on the mechanical properties^{36, 68} (Fig. 1.7). In particular, the least mechanically stable folded conformation is the random coil, an unstructured flexible chain with minimal stabilizing enthalpic contributions, and as such its mechanical properties at low forces are dominated by entropic contributions^{69, 70}. In other protein conformations, the enthalpic contributions - in particular hydrogen bonds - induce mechanical resistance to an applied stress. Proteins formed predominantly of alpha-helices, where the hydrogen bonds are formed within each helix, are generally mechanically weak, requiring ~10-30pN to unfold. On the contrary, beta-sheet proteins, defined by a dense network of hydrogen bonds spanning neighbouring strands, generally unfold at much higher forces^{71, 72}. In this case, it is typically a single set of hydrogen bonds (coined the 'mechanical clamp') that bears the brunt of the force and is the rate-limiting step for protein unfolding - generally considered to be the Achilles heel of protein unfolding by mechanical force^{73, 74}.

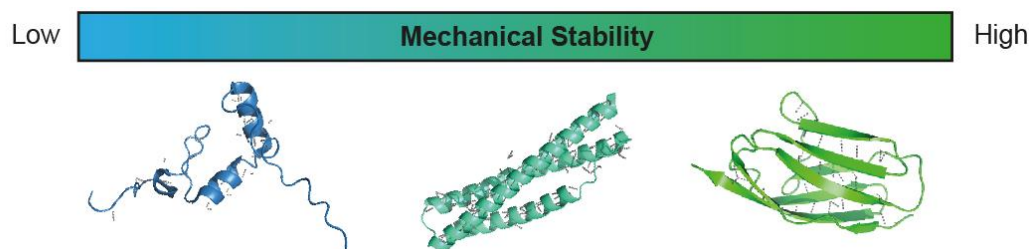


Figure 1.7. Protein structure is a key indicator of protein nanomechanics. The least mechanically stable conformation is a random coil, whereas the dense hydrogen bond network of β -sheet proteins commonly gives rise to a mechanically stable structure.

This mechanical clamp is fundamentally important for the overall nanomechanics of the protein. It is now established that the topology of the clamp and the number of hydrogen bonds within it are fundamentally important for understanding, and even predicting, the mechanical properties of the protein⁷⁵. In particular, the unfolding force can be maximized by

increasing the number of hydrogen bonds in the clamp region and with a parallel strand topology that gives rise to a shearing motion when exposed to a mechanical perturbation. In contrast, the mechanical stability of the protein is minimised with fewer hydrogen bonds in the clamp and an antiparallel topology which easily unzips with force. The importance of the clamp is reinforced with examples demonstrating that the introduction of point amino acid mutations into this region can modulate the overall mechanics by either strengthening or disrupting the bonding network^{13, 76}. Despite these findings, there have been a handful of studies demonstrating that it is not sufficient to merely consider the protein structure and mechanical clamp when predicting the mechanical response of the protein⁷⁷. In particular, two structurally homologous Staphylokinase proteins have shown considerable disparity in their unfolding forces ($\Delta F_u \sim 180$ pN), suggesting that there might be additional protein features located outside of the clamp that ‘allosterically’ contribute to the chemical origin of mechanical stability that we are yet to fully understand⁷⁸.

The idea that the mechanical stability of the protein is intrinsically linked to the structure, is cemented by studies where the direction of the applied force was altered. While the majority of force spectroscopy studies constrain the reaction coordinate to the N-C-terminal direction, molecular biology techniques have facilitated alternative anchoring points that allow a different pulling geometry. This approach has revealed the mechanical anisotropic nature of a number of proteins, including ubiquitin, ankyrin, and src SH3^{79, 80, 81, 82}. This is due to the complex secondary structure and expansive hydrogen bond networks, which are not isotropic, and therefore the pulling direction has a profound influence on the mechanical response.

In addition to changing the mechanical properties of the stretched protein, pulling from different directions in some cases reveals signatures of

complexity in the unfolding energy landscape that cannot be captured by pulling from the N-C termini. For example, pulling the green fluorescent protein (GFP) from different anchoring points leads, in the majority of the cases, to an all-or-none unfolding pathway whereby the unfolding force spans from 100 pN – 600 pN. However, pulling from a particular direction reveals that the protein unfolds through a well-defined intermediate⁸³. These findings clearly demonstrate that the direction of the force transmission along the protein backbone modulates the appearance of additional mechanical energy barriers, which are likely due to local energy minima in the energy landscape corresponding to partially folded structures. Hence, the change in the pulling direction reveals, for some proteins, previously hidden complexities in their underlying landscape.

Interestingly, a number of proteins have been shown to stochastically exhibit mechanical unfolding intermediates after crossing the main transition state even when the force vector remains the same. This process, known as kinetic partitioning, is likely to originate from the protein having a number of different energy barriers in the unfolding pathway, and the application of mechanical force ‘traps’ a particular conformation^{84, 85, 86}. Recent work has demonstrated that these ‘on-pathway’ intermediates can be modulated by changes in the chemical environment and by the magnitude of the applied force^{45, 87}. It is therefore tempting to speculate that this might be a naturally employed mechanism to tune the number of mechanical energy barriers in order to control the extent of molecular extension in conditions where it would be un/favourable to fully unfold.

Altogether, these investigations demonstrate that mechanical force can trigger protein unfolding, transitioning a single protein from a rigid structure into an unfolded and elastic string of amino acids. The next step is to investigate

whether there is a way of rationally tuning the mechanical properties of the native state of a molecule.

BINDING MEDIATES MECHANICS

A ubiquitous biological strategy to tune or enhance the repertoire of protein function is through binding with an external co-factor. In particular, the interaction between a protein and a ligand/protein is a fundamental biological process in almost all living organisms, where this chemical interaction is capable of switching protein function with profound knock on effects for biological signalling. While the concepts of mechanical force and chemical reactions have generally been considered as two distinct topics, it is becoming apparent that this is not the case and in fact the two are, in some cases, fundamentally intertwined.

Astonishingly, single molecule force spectroscopy approaches have revealed ligand-induced mechanical changes in a wide variety of proteins, including the enzyme dihydrofolate reductase, the molecular chaperone DnaJ, and staphylococcal nuclease^{38, 88, 89, 90, 91, 92, 93, 94, 95}. In all reported cases, the ligand either increases the mechanical stability of the native state of the protein, or imposes an additional mechanical energy barrier during the unfolding pathway (Fig. 1.8 a, b). In some of these cases the binding interaction is located away from the mechanical clamp of the protein, implying that the ligand may introduce allosteric conformational changes that shift the mechanical properties. This seemingly ubiquitous effect implies that ligand binding may be a natural protection strategy against mechanical stress.

This ability to modulate the mechanical properties of a protein has lead to a number of studies utilizing this approach as a method for rationally designing proteins with mechanical tunability⁹⁶. In particular, protein GB1 is converted from an entropic string to an elastomeric mechanically stable protein in the

presence of the Fc fragment of human IgG⁹⁷. This observed ‘chameleon’ behaviour upon the addition of an external stimuli, offers a tantalizing opportunity to create biomaterials with tuneable mechanical behaviour^{56, 98}.

Although these studies provide a comprehensive understanding of how ligands modify the mechanics of the folded substrate, a direct mechanical quantification of the binding interaction remains far more mysterious. This is primarily due to the experimental difficulties associated with acquiring reliable data. To quantify the strength of binding, the protein must be attached to the probe and the ligand attached to the surface (fig. 1.8 c). In this configuration, the dissociation of the protein-ligand interaction leads to a loss in the physical connection between the surface and the probe, which results in the end of the trajectory. Unfortunately, without a well-defined fingerprint it is not possible to unambiguously determine the mechanical origin the bond that has been broken by force, as the possibilities can include the protein-ligand interaction or the protein-surface / ligand-probe interaction⁹⁹. A number of groups have made significant biotechnological advances in order to obtain reliable results. The Gaub group has used a number of approaches covalently immobilizing one half of a binding interaction onto the AFM cantilever and the other is engineered onto the end of the molecule on the surface –in this configuration each trajectory is a result of a specific interaction between the two binding partners^{100, 101, 102}. Recently, the Carrión-Vázquez group developed a ‘molecular Velcro’ system to create a long covalently bonded linker that encompasses the whole interaction, therefore upon dissociation, the connection is not broken and the rupture is hallmarked by the extension of the flexible linker (fig. 1.8 d)¹⁰³. These approaches create a new vista for directly quantifying the mechanical strength of the bonds governing protein-protein and protein-ligand interactions.

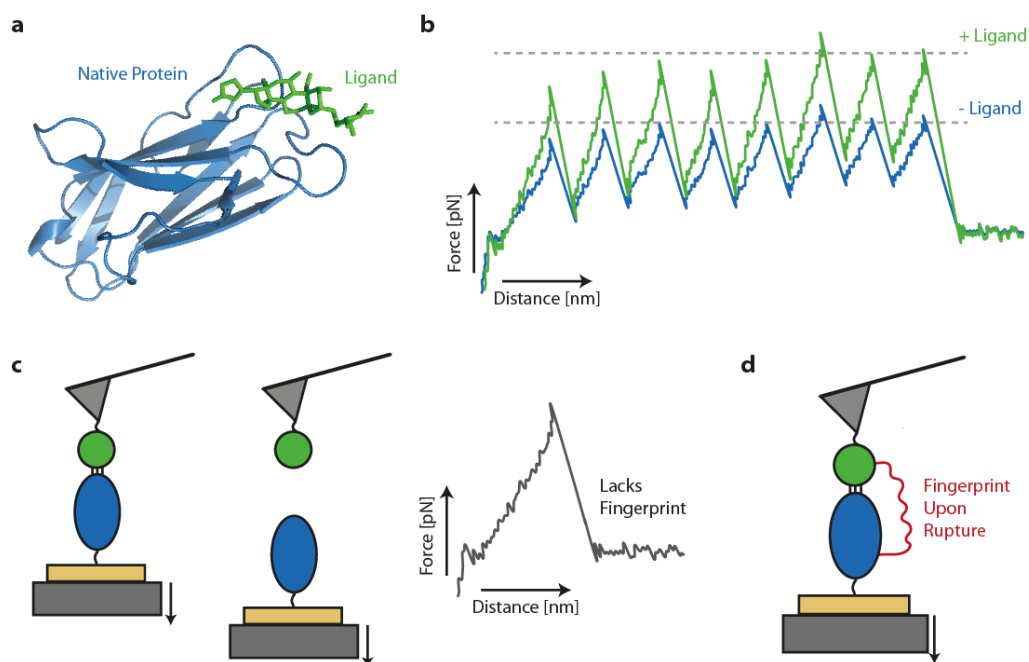


Figure 1.8. Ligand binding has the power to modify the mechanical stability of a protein. (a) A ubiquitous biological strategy to tune protein function is through a ligand (green) binding to a protein (blue). (b) Ligand binding has the power to modify the nanomechanical properties of the native state of the protein. (c) To probe the rupture forces of the protein-ligand interaction, force must be applied across the binding interaction. Upon rupture, the connection is lost, resulting in a single peak in the force extension trajectory. The mechanical origins of this trajectory lacks a clear fingerprint and thus is open to interpretation. (d) This problem is circumvented if a flexible linker encompasses the protein and the ligand – providing a reliable and reproducible fingerprint.

A parallel strategy to enhance the protein's mechanical stability is through metal chelation; Cao *et al* demonstrated that Nickel binding in an engineered bi-histidine site is enough to enhance the mechanical stability of the GB1 protein in as much as 100 pN¹⁰⁴. In fact, the incorporation of a metal into the protein structure is an in-vivo strategy used by a large number of proteins to expand their functionality, including proteins related e.g. in electron and oxygen transport. Despite the importance of the bound ion, very little is known regarding the mechanochemistry of these co-ordinating bonds within the context of a protein. The Li laboratory was the first to measure the mechanical disruption of organometallic bonds embedded within metalloproteins¹⁰⁵. This novel study looked at the nanomechanics of rubredoxin, containing a conserved metallic centre (FeS₄) composed by two pairs of non-equivalent Fe(III)-thiolate bonds (fig 1.9 a). Despite the seemingly

high covalent character of the Fe(III)-S interaction, the mechanical stability of an individual Fe(III)-thiolate bond is surprisingly low (~ 210 pN) compared to the nanoNewton bond strength defining typical covalent bonds ⁹⁹ (fig. 1.9 b). Within the same iron centre, the Fe-S bond pairs with shorter distance ($2.25 \text{ \AA}$) require a higher force to disrupt than the longer ones ($2.31 \text{ \AA}$). Moreover, the rupture force of the reduced Fe(II)-thiolate is 30% lower (~ 152 pN) than that corresponding to the oxidised form. Altogether, it seems then evident that there is a direct relationship between a higher covalency and higher bond strength. Notably, by varying the pulling velocity the distance to the transition state was measured, revealing a much shorter distance ($\Delta x = 0.11 \text{ nm}$) than that typically associated to protein unfolding ($\Delta x \sim 0.25 \text{ nm}$), and corresponds to about half of the Fe-S bond length. The unexpected mechanical lability of the Fe-thiolate organometallic bond was further confirmed upon Zn substitution in the rubredoxin centre¹⁰⁶ (fig. 1.9 c). In that case, the mechanical stability of an individual Zn-S bond was found to be even lower (~ 170 pN) than that corresponding to the Fe-S bond (~ 210 pN). The relative decrease in mechanical stability for the Zn-S bond compared to the Fe-S bond may be again rationalized in terms of the different degree of covalency in both cases.

After the pioneering work on rubredoxin, a recent set of reports on different metalloproteins seems to generally conclude that individual organometallic bonds within proteins exhibit a surprisingly low mechanical stability despite their high covalent character. For example, the low mechanical resistance of the Zn-S bond was further confirmed in the case of the zinc finger of the DnaJ chaperone, for which a rupture force as low as ~ 90 pN was measured¹⁰⁷. In addition, the force required to disrupt the Cu-S bond was found to be ~ 171 pN within the context of the GolB protein¹⁰⁸. A similar force value (~ 165) pN was obtained in the same GolB protein when the nanomechanical stability of the S-Au was measured instead.

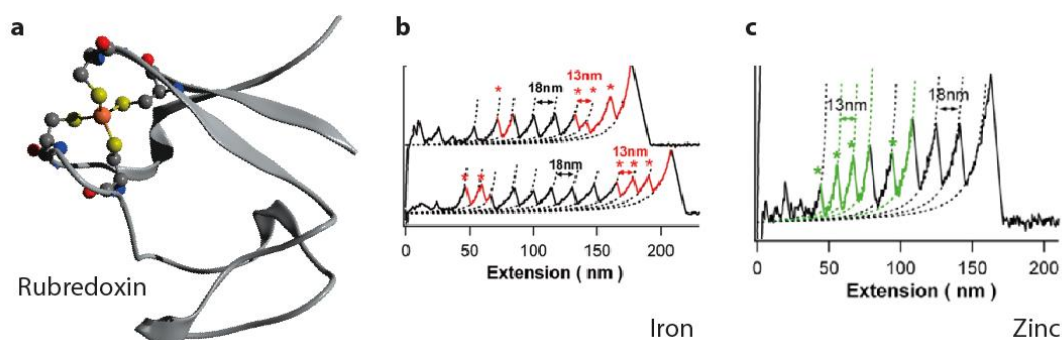


Figure 1.9. Force spectroscopy can monitor a single organo-metallic bond within the context of protein unfolding. (a) The protein rubredoxin houses a single metal within its structure. (b) The rupture and release of iron from rubredoxin is fingerprinted by steps of 13nm, occurring at ~ 210 pN. (c) Replacing the iron centre with zinc still elicits steps of 13 nm, however the rupture force is now observed at 170 pN. (Reproduced from references ¹⁰⁵ and ¹⁰⁶)

In general, different reasons could partially explain the significant (and quite counterintuitive) lower mechanical stability of the highly covalent organometallic bonds within the protein framework. Firstly, in some of the cases such as the S-Au interaction, the studied organometallic bond length is larger in the protein environment, thus rendering the bond weaker. Moreover, further chemical interactions occurring in the metal centre, might also contribute to regulate the strength of the organometallic bond. This was clearly demonstrated with the effect that hydrogen bonding has on the modulation of the ferric-thiolate bonds in rubredoxin³⁹. Finally, perhaps the most important factor is the alignment of the force vector with the studied chemical bond. In fact, in most of the studied metalloproteins the topology of the metal centre involves the formation of multiple bonds between the metal and the coordinating ligands. Hence, in the most plausible scenario, the pulling force is not perfectly aligned with the direction of the first organometallic bond to be mechanically disrupted. In fact, similar to mechanical unfolding, stretching the organometallic bonds from different directions is likely to exhibit mechanical anisotropy.

Altogether, these novel experiments on the nanomechanics of individual organometallic bonds have already provided non-expected results, which call

for additional experimental and theoretical efforts. Interestingly, recent work from the Li lab has demonstrated that when the metalloprotein is pulled at neutral pH, the rupture process is homolytic. By contrast, under high pH conditions or in the presence of poor nucleophilic ligands, the ferric– thiolate bond rupture follows instead a heterolytic reaction mechanism¹⁰⁹. This highlights that the local chemical environment and the chemical reactivity of the metal centre plays a fundamental role in modulating the mechanical properties of the coordinating bonds and thus the mechanical response of the protein.

Interestingly, whereas in most cases the addition of a ligand seems to switch the mechanical response of the protein in an all-or-none fashion, the mechanical response of the protein in the presence of a metal can be dictated by the chemical reactivity and coordination structure of the metal centre. Furthermore, the location of the metal within the protein has an important role in protein unfolding. In proteins such as rubredoxin, where the metal is at the protein termini, the unfolding force of the protein is directly dictated by the nanomechanics of iron, therefore Fe-S bond rupture is the rate-limiting step for unfolding^{106, 110}. Alternatively, in proteins such as DNAJ, the metal is located after the hydrogen bond clamp and is ‘hidden’ from the force until the protein unfolds¹⁰⁷. In this instance, the metal does not modulate the initial nanomechanical properties of the native state, but it imparts an additional mechanical energy barrier separating the native and unfolded conformations. This offers a unique opportunity to rationally and reversibly enhance the stability of a protein by engineering metal coordination sites at different locations to preferentially stabilize molecular conformations - a strategy that is now being expanded to tailor the spatial and temporal properties of materials^{111, 112}.

Altogether, the control of the mechanical properties of proteins has provided direct insight into the molecular mechanisms controlling their unfolding pathways and the associated kinetics. A key discovery in protein mechanics was the observation that mechanical unfolding of elastic proteins that are physiologically exposed to mechanical force is, in general, reversible^{30, 113, 114}. As such, understanding the molecular mechanisms underlying the folding process of mechanically unfolded proteins as they recover their natively folded (and mechanically stable) state has become a pressing need in the advancement of the protein nanomechanics field.

MECHANICAL PROTEIN FOLDING

According to the Levinthal paradox, if an unfolded protein samples every possible conformation it would take longer than the age of the universe to reach its native state¹¹⁵. While ensemble approaches have clearly demonstrated this is not the case, they lack the resolution to gain mechanistic and structural insight into this complex process that guides the protein along this journey^{116, 117}. Single molecule force spectroscopy has proven a powerful tool to study unfolding, however folding has proved far more challenging due to the low force that this event occurs at. A force-quench approach using force-clamp spectroscopy circumvents the need for high resolution at low forces. An initial high force pulse triggers each of the proteins in the polypeptide chain to unfold, each marked by a step in the trajectory (Fig. 1.10). This unfolding pulse brings each protein into a highly extended state conformation. Quenching the force down to 0 pN brings the amino acid chain close in conformational space and initiates the protein folding process¹¹⁸. The success of folding is then probed with a subsequent high force pulse, where the re-unfolding and the full recovery of mechanical stability is the signature of successful protein folding (fig. 1.10 a, b). This approach has delineated the steps governing protein folding from a mechanically extended state, where

first the protein will undergo elastic recoil of the polymer chain, followed by the hydrophobic collapse of the protein to the molten globule conformation and finally the acquisition of the native fold^{119, 120, 121, 122}.

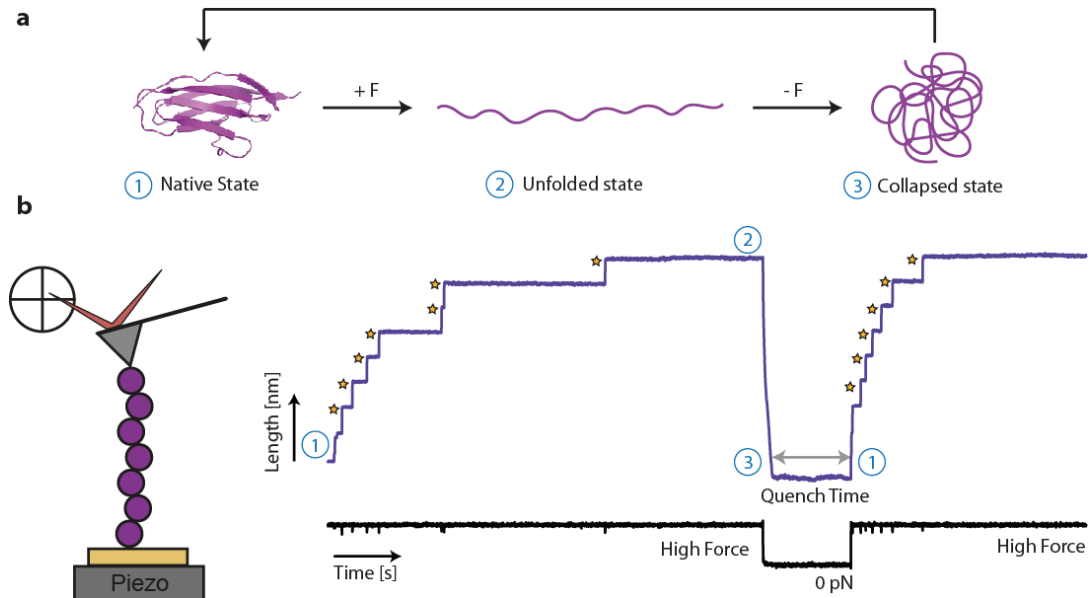


Figure. 1.10. A force-quench protocol captures individual protein folding. (a) Mechanical force drives the native protein to a fully extended, unfolded conformation. The subsequent release of the force triggers the collapse of the protein which eventually leads to the re-acquisition of the native conformation. (b) The signature of protein folding in the force quench approach is the reacquisition of mechanical stability in the second high force pulse (yellow star). In this trajectory, 7 proteins have been unfolded by the first high force pulse. A short quench time allows the protein to collapse and subsequently fold. In the second high force pulse, the success of folding is tested – in this case 6 proteins have regained their native state.

In this AFM force-quench approach, the precise moment that folding occurs cannot be captured, instead the success of folding can only be inferred from the subsequent unfolding. Alternatively, magnetic tweezers, due to their high resolution at low forces, can capture the precise moment that a protein transitions from the unfolded to the collapsed/native state (Fig. 1.11). Once the protein has been unfolded with a high force pulse, the force is quenched down to the lower force ($\sim 3 - 10$ pN). During this low force pulse, individual steps can be observed (Fig. 1.11 a, blue arrows) that correspond to the folding of each individual protein. This approach has already provided insight into the molecular mechanisms governing protein folding and the underlying free

energy landscape^{53, 123, 124}, and will undoubtedly continue to provide important contributions towards our understanding of protein folding.

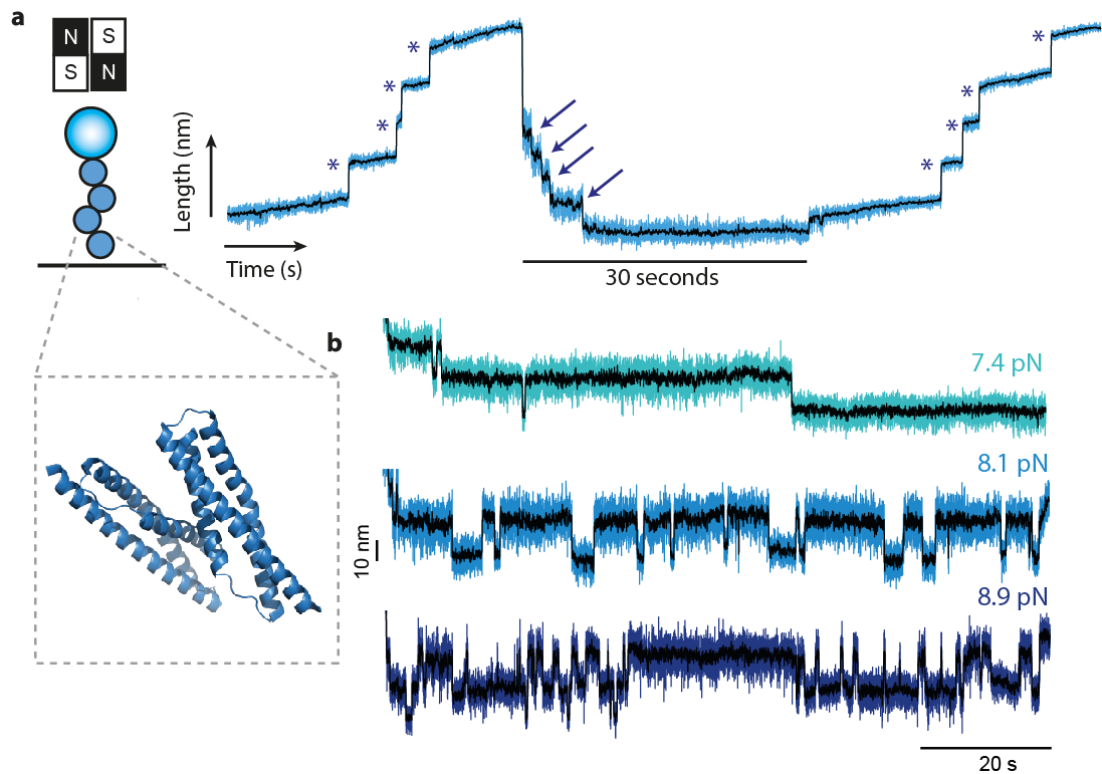


Figure 1.11. Direct observation of protein folding. (a) The enhanced low force resolution of the magnetic tweezers reveals the precise moment of protein folding. In this trajectory, each asterisk marks the unfolding of a single spectrin repeat 73 protein. Upon quenching the force down to 8 pN, it is possible to directly capture the protein transitioning from the unfolded to the folded state, blue arrows. (b) Holding the spectrin repeat 73 protein at a low force allows the protein to exist in an equilibrium state, hopping between the unfolded and folded conformation. The life time of each state is strongly influenced by the applied force.

While this force-quench approach tracks a protein from an out of equilibrium highly extended state back to the native, other efforts have focused on looking at protein unfolding/folding at equilibrium^{35, 53, 125}. This is typically performed using passive force clamp due to the enhanced temporal resolution⁴⁰, and monitors the dynamics of protein unfolding-folding at a constant force (fig. 1.11 b). Such an approach captures the precise moment that the protein ‘hops’ between the unfolded and folded state, and has revealed both the force dependence of protein folding and how external co-factors can modulate folding dynamics^{126, 127, 128}. All of the aforementioned single molecule folding approaches have provided significant insight into both the mechanisms and

kinetics that govern how the protein recovers its mechanical properties, switching from a compliant unfolded polymer to a mechanically stable folded protein^{126, 129, 130}.

While there are numerous examples demonstrating the successful reversibility of protein unfolding/folding, it remains challenging to capture the rare occasions when something goes awry during the folding process¹³¹. Protein misfolding occurs when a protein is no longer able to attain its native fold and instead adopts an abnormal structural conformation that lacks function. Often this misfolded state becomes toxic and has been implicated as the initial stages of a wide range of pathologies, including Huntington's, Parkinson's and Alzheimer's¹³². Despite the critical importance of this process, very little is known about the precise molecular mechanisms that give rise to this damaging conformation. An emerging tactic to elucidate this information is using single molecule force spectroscopy to compare the underlying folding free energy landscape in cases where the protein folds and misfolds^{133, 134}. This approach revealed a local roughening of the landscape for misfolding of the prion protein PrP¹³⁵. In many cases the nanomechanical properties of the resulting misfolded structure are completely divergent from the native – either reduced or enhanced^{136, 137, 138, 139}.

Altogether, these results demonstrate that the protein folding process is a crucial step in modulating protein nanomechanics, transitioning from an entropic spring back to a mechanically stable structure. The next crucial step is to understand how external contributions, such as post translational modifications or molecular chaperones, can modulate the protein folding pathway with knock on effects for molecular compliancy.

MECHANICS MEDIATES BINDING

The protein folding process is governed by a careful interplay between entropic and enthalpic contributions¹⁴⁰. There are a wide variety of factors - such as post translational modifications, disulfide bonds and protein binding - that can disrupt this balance and alter protein folding, especially when this occurs during the folding process. The native state of a protein conceals many amino acids within the core which protects them from the aqueous environment and prevents them from participating in binding. However, mechanical force has the power to modulate the structural conformation of the protein, which in turn exposes these previously cryptic binding sites^{141, 142, 143, 144}. This has led to a surprising and unexpected relationship between mechanical activation and biochemical reactions^{37, 145, 146, 147, 148}.

In particular, an unsolved dichotomy is whether this force-mediated binding that occurs after protein unfolding has repercussions for the subsequent folding fate of the protein. A handful of publications have investigated this question in biologically relevant systems. In particular, the two focal adhesion proteins talin and vinculin, which are essential players in cellular signalling, adhesion and migration, have been studied using a variety of single molecule force spectroscopy approaches. Talin acts as a molecular force sensor *in vivo*, where the force dependent unfolding and folding pathways make it an excellent buffer to the biological forces experienced circa ~5-10 pN¹⁴⁹. Single molecule magnetic tweezers have beautifully demonstrated that upon mechanical unfolding, talin exposes a previously cryptic vinculin binding site¹⁵⁰. This initial demonstration of a mechanically mediated biochemical reaction is thought to be an important mechanism for the transduction of force in the cell^{149, 151}. Furthermore, follow up work has demonstrated that this binding interaction locks talin in an unfolded state, thus maintaining a mechanically compliant conformation¹⁵². The next challenging step is to

understand why this unfolded conformation is important within the context of a cell.

Along this vein of force revealing buried binding sites, Alegre-Cebollada *et al* investigated the folding aftermath upon the force induced post-translational modification s-glutathionylation⁴³. S-glutathionylation is a modification that involved the addition of glutathione to a protein cysteines. Glutathione is one of the most prominent low molecular weight thiols in biology with the potential to act as both an oxidising and reducing agent depending on its oxidation state. In particular, glutathione is the end product in a number of redox driven signalling pathways and has been implicated in a number of cardiac pathologies¹⁵³. Single molecule AFM revealed that upon glutathionylation, the 27th immunoglobulin domain of titin cannot obtain its original mechanically stable state. Interestingly, the mechanical stability of the folded protein seems to depend on the location of the modified cysteine. The substrate protein in this study has two native cysteine residues; the modification of one leads to a weakly folded state, whereas modification of the other gives a completely misfolded state. Given that the overall elasticity of titin can be well described by the mechanical properties of each individual domain⁶⁰, it is tempting to speculate that this variation in the folding kinetics has important implication in the overall elasticity of muscles.

These poignant examples highlight the importance of this underexplored concept that force can mediate biochemical reactions that ultimately dictate the folding fate and mechanics of the protein (Fig. 1.12). There are a wide variety of proteins in a range of biological scenarios that are believed to experience to mechanical perturbations, including extracellular matrix proteins, focal adhesion complexes and muscle protein complexes^{154, 155, 156, 157}. Biochemical experiments are not able to bring the protein to these extended conformations, and therefore it is possible that there are a wide variety of

important *in vivo* binding interactions that have thus far eluded discovery. In this context, single molecule force spectroscopy approaches are the most capable of capturing and quantifying these interactions.

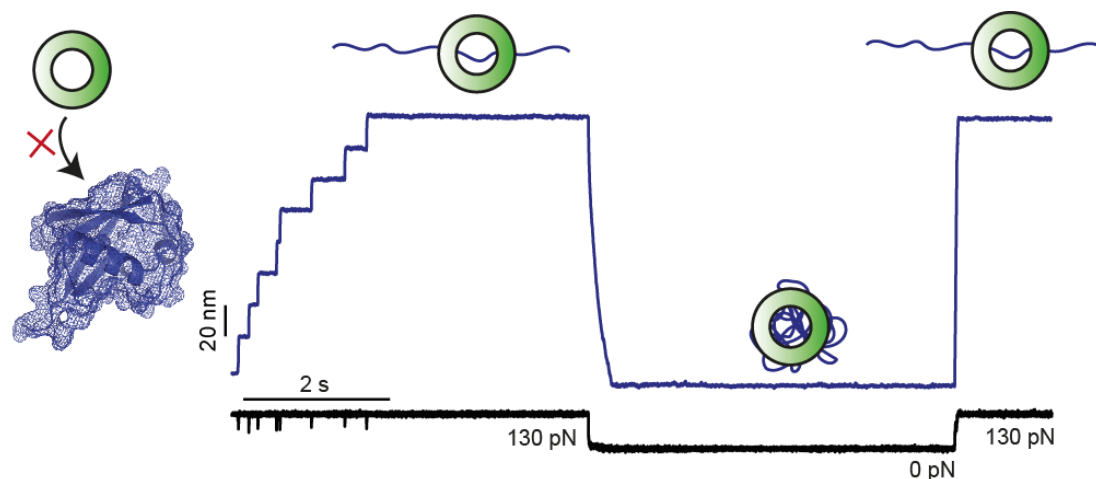
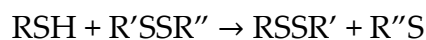


Figure 1.12. Force mediates the exposure of cryptic binding sites that can impact protein folding. The native folded state of the protein can act as a cage, shielding binding sites within the core. The application of mechanical force unfolds the protein, ultimately modulating the exposure of the previously concealed sites and permitting binding interactions. If this interaction remains bound/attached to the protein, the protein is prevented from folding and thus will remain in a compliant elastic conformation. This inability to fold is hallmarked by the absence of steps in the second high force pulse.

DISULFIDE BOND RUPTURE

One of the most prominent approaches to modulate the mechanics and the overall extensibility of a protein is through the incorporation of a disulfide bond; a highly covalent bond formed between two sulfur atoms¹⁵⁸. From a thermodynamic perspective, the incorporation of a disulfide into a protein structure reduces the entropy of the unfolded state and stabilises the folded conformation, however these covalent molecular staples have additional pivotal functional roles as they are able to dynamically form, rupture and shuffle in response to environmental chemical changes, altering both protein structure and function^{159, 160}. The molecular mechanism that predominantly governs this process is thiol-disulfide interchange. This occurs when a deprotonated thiol initializes an S_N2 nucleophilic substitution on one of the thiols participating in the disulfide, resulting in the cleavage of the original disulfide and the formation of a new mixed disulfide moiety.



This bimolecular reaction can be treated as a simple chemical reaction that proceeds via a transitory transition state:



where the reaction transition state is a fleeting molecular conformation ~picoseconds before evolving to more stable conformations. This ephemeral nature has rendered the characterization of the transition state conformation experimentally inaccessible for even simple chemical reactions. While the advent of femtosecond lasers, capable of freezing molecular arrangements to gain unprecedented access to atomic motions, have revolutionized our approach towards ‘visualizing’ chemicals reactions¹⁶¹, it is not possible to use this approach to study reactions governed by mechanical force, and in particular, reactions occurring within the core of individual proteins. Fortunately, the advent of single molecule force spectroscopy in combination with powerful computational tools has circumvented this problem and provides an orthogonal method for ‘visualizing’ the transition state of a mechanically activated chemical reaction¹⁶².

It is well established that the rupture of a disulfide bond by a thiolate is a concerted $\text{S}_{\text{N}}2$ reaction, with a single smooth energy barrier leading to a single transition state and no intermediate species. Therefore this becomes an ideal, simple candidate to start to understand how force perturbs the free energy landscape of a chemical reaction. One of the primary methods towards this endeavour involves embedding a single disulfide bond within the core of an individual protein¹⁶³. In this configuration the protein acts as a cage, shielding the bond from the aqueous environment and preventing the initialization of the reaction until the protein is mechanically unfolded, giving precise control over the start time of the reaction kinetics^{164, 165} (Fig, 1.13 a).

Single molecule force spectroscopy allows direct observation of disulfide cleavage by monitoring protein elongation over time, providing a new biophysical vista for studying chemical reactivity. In this configuration, first the protein is first unfolded up to the disulfide bond (generally in a short time pulse), exposing the disulfide to the solution (Fig. 1.13 b, grey steps). Secondly, a higher force pulse in conjunction with a nucleophile leads to the rupture of the disulfide and this is characterised by a step that corresponds to the elongation of the amino acids previously trapped behind the bond^{166, 167, 168} (Fig. 1.13 b, blue steps).

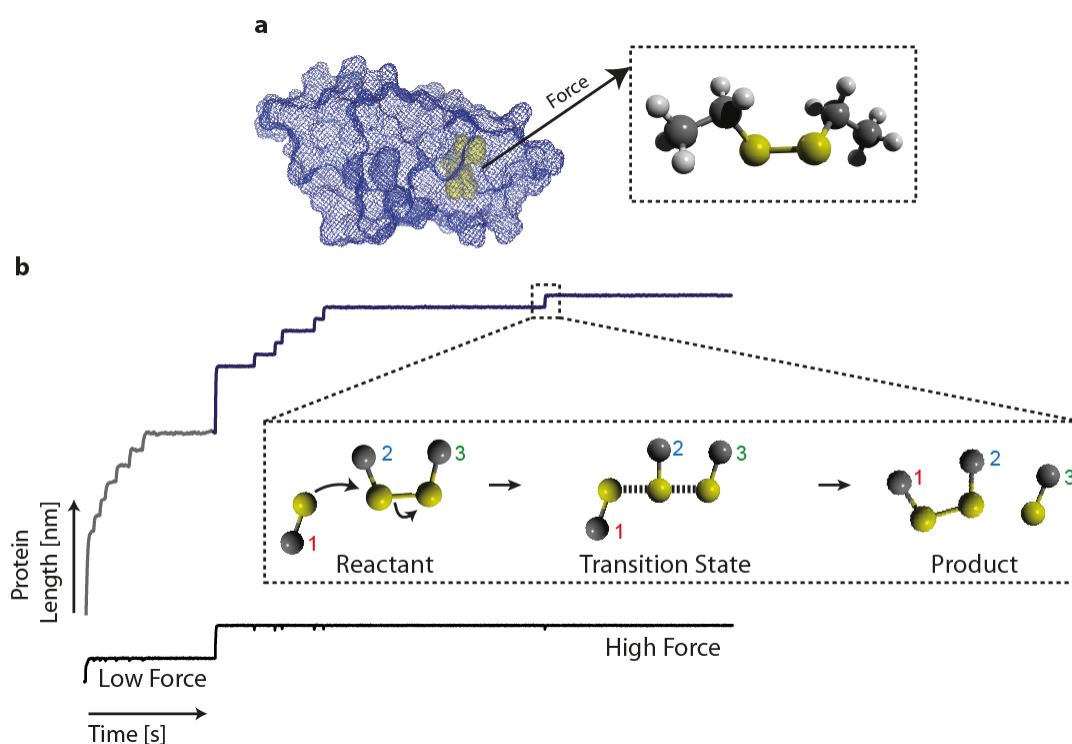


Figure 1.13. Capturing a single bimolecular chemical reaction within the core of a protein. (a) It is possible to engineer a disulfide bond into the core of a protein. In this configuration, the protein acts as a cage, shielding the disulfide bond from solution until force triggers protein unfolding and the subsequent bond exposure. (b) In a force clamp experiment, an initial low force pulse unfolds and extends all of the amino acids outside of the disulfide (grey steps). A second high force pulse in conjunction with a nucleophile in solution triggers disulfide bond rupture. Each rupture event is hallmarked by the extension of the amino acids that were previously trapped behind the bond.

In seminal work, Wiita *et al* demonstrated that the application of force imparts energy to the reaction, lowering the activation energy barrier and thus increasing the rate of the reaction¹⁶². The observed rate of reduction, r , of this

force-moderated process is exponentially dependent on the applied force and linearly dependent on the concentration of deprotonated nucleophile.

$$r = [Nu] \alpha(F) \quad (1.5)$$

Where $\alpha(F)$ depends exponentially on the applied force according to the Bell model

$$\alpha(F) = A \exp\left(\frac{F\Delta x - E_a}{k_B T}\right) \quad (1.6)$$

The rate of rupture at an applied force follows a single exponential, which is indicative of a two-state process with smooth and highly localized energy landscape¹⁶⁹, consistent with an S_N2 disulfide reduction proceeding through a linear S-S-S Walden transition state^c. Similar to the force-clamp experiments of protein unfolding, a plot of the logarithmic rate of disulfide reduction covering the entire range of applied forces allows direct quantification of the Δx values associated to the force-dependent disulfide bond cleavage⁴⁸. Remarkably, the obtained values for a variety of low molecular weight thiols falls within a narrow range of $\Delta x = 0.29 - 0.35 \text{ \AA}$ ¹⁷⁰ (Fig. 1.14 a, b), roughly an order of magnitude shorter than that associated to protein unfolding. These sub-angstrom values demonstrate that the mechanical reduction of a protein disulfide is significantly less force-dependent than mechanical unfolding. The next natural question is to investigate whether the measured Δx values has a direct physical meaning. The first test is to elucidate if this value is directly consistent with that obtained through bond stretching. Considering that the vibrational frequency of a disulfide bond stretching mode is $\sim 522 \text{ cm}^{-1}$, corresponding to a bond spring constant of $\sim 255 \text{ N.m}^{-1}$, forces as high as 500 pN would only increase the sulfur-sulfur separation distance by $\sim 0.02 \text{ \AA}$. This

^c The energy barrier of the S_N2 reaction is minimized when the attacking sulfur, the attacked sulfur, and the leaving group sulfur are arranged in a straight line, such that the S-S-S angle in the transition state is close to 180°. This is known as the linear Walden transition state.

extension value is clearly well below the experimentally observed Δx value, suggesting that brute mechanical force cannot be solely explain the experimental observations¹⁶⁶. Instead, ab initio quantum mechanical simulations reveal that the $\Delta x = 0.32 \text{ \AA}$ value correlates surprisingly well with the elongation of the S-S bond at the S_N2 transition state¹⁷¹. Hence, it is tempting to speculate that the slope of semi-log plot between the rate of the reaction and the pulling force provides direct information on the conformation of the reaction transition state. Follow up experiments changing the chemistry of the central atom reveal that phosphor- and oxygen-based nucleophiles display larger transition state distances of 0.42-0.46 $\text{\AA}$ and 0.5 $\text{\AA}$ respectively, compared to sulfur ($\sim 0.35 \text{ \AA}$), suggesting that the central atom of the attacking nucleophile seems to play a pivotal role in the obtained Δx value¹⁷⁰. Strikingly, this trend is also confirmed in related quantum mechanical simulations¹⁷⁰. Taken together, these results strongly suggest that the chemical properties of the nucleophile have a profound influence the mechanisms and reactivity governing this force-activated reaction¹⁷⁰. While it is clear that the chemistry of the central atom affects Δx , the chemical factors influencing the height of the activation energy barrier remains an unanswered conundrum.

The force-dependent reactivity of all previous experiments have been efficiently fitted with a simple Bell model, however at high forces there is a surprising sharp deviation. In a number of disulfide containing Ig domains there is an abrupt change in the logarithmic rate of reduction as a function of force at $\sim 500 \text{ pN}$. Treating the two regimes separately with the Arrhenius equation reveals a shift in the transition state distance towards the reactant from 0.46 $\text{\AA}$ to 0.07 $\text{\AA}$, meaning that at higher forces the reduction rate is considerably less sensitive to the application of force¹⁶⁶. This reactivity switch can be rationalized by the force-mediated rearrangement of the backbone dihedral angles resulting in an emergent second energy barrier that dominates

the reaction at higher forces. Computational approaches reveal that while forces *circa* 500pN do not dramatically distort the C-S-S-C dihedral angles away from their preferred 90° angle, the S-S-C-C dihedral is rearranged leading to a steric constraint limiting nucleophilic accessibility and thus slowing the reaction rate¹⁷² (Fig. 1.14 c, d). These computational approaches have enhanced magnification of the underlying mechanisms governing force mediated reactivity, but also have provided access to the reactivity at forces that are currently not experimentally accessible due to limitations with molecular tethers. In particular, isotensile experiments performed by the Marx lab have investigated the force dependent activation free energies of six reaction pathways. While the S_N2 reaction at the sulfur remains the predominant pathway for forces lower than 2 nN, at forces > 2 nN the reaction switches from a bimolecular reaction to a unimolecular reaction dominated by the brute force rupture of the S-S or C_α-S bonds. Despite the activation energies of the S_N2 reaction at C and the α-elimination not favouring these pathways, they still display force dependent behaviour. This is contrary to the behaviour of the β-elimination which is invariant with force and is purely thermally limited, thus demonstrating that there is a disulfide rupture mechanism that is thermodynamically rather than kinetically governed^{173, 174}.

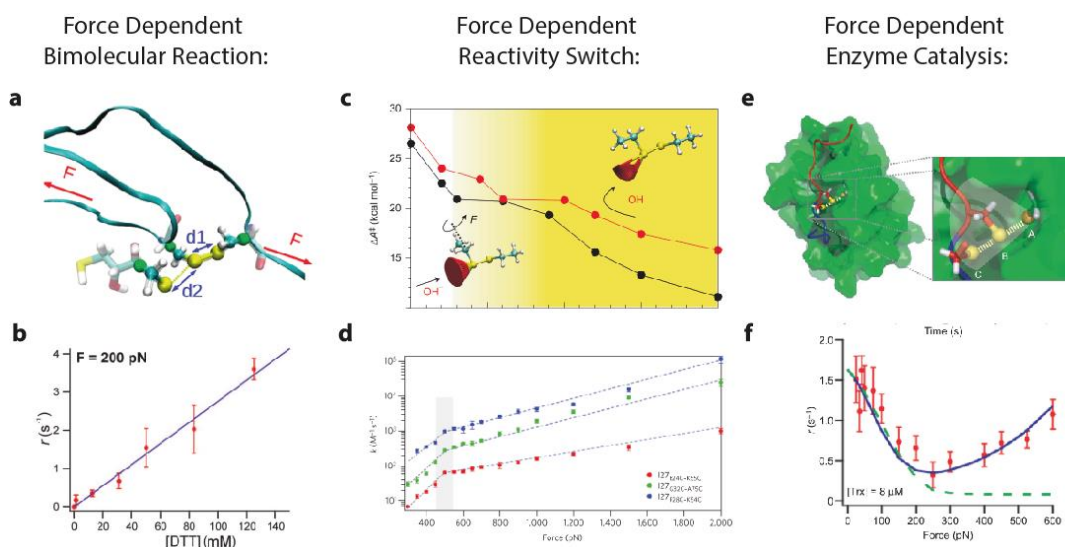


Figure 1.14. A combination of experimental and computational approaches has captured the complexity of force mediated disulfide scission. (a, b) The reduction of a disulfide by a nucleophile is accelerated upon the application of force. The rate of rupture is exponentially increased by a mechanical perturbation and is linearly dependent on the concentration of deprotonated attacking species. (Reproduced from references ¹⁷⁵ and ⁴⁸) (c, d) There is an abrupt force-mediated switch in the reactivity of disulfide bond rupture at 500 pN. A combination of experimental and computational approaches has revealed that this originates from the rotation of the protein backbone angles, which sterically hinders attack. (Reproduced from references ¹⁷² and ¹⁶⁶) (e, f) The reduction of a protein disulfide by the naturally occurring enzyme thioredoxin displays a biphasic force dependent nature. This is caused by the geometric restrictions imposed by the active site of the enzyme, preventing the formation of the linear transition state of the reduction reaction. (Reproduced from reference ⁴⁸)

The aforementioned investigations have all focused on understanding disulfide reduction using small chemical nucleophiles. However within the cell, disulfide reduction is primarily mediated by a set of dedicated enzymes known as oxidoreductases^{159, 176, 177}. To establish the molecular mechanisms governing enzyme mediated disulfide reduction, Wiita *et al* used this force clamp approach to probe the molecular mechanisms governing thioredoxin mediated disulfide reduction¹⁷⁸. Converse to the chemically equivalent reactions with simple nucleophiles that always demonstrate an increasing reduction rate with force, the enzyme revealed a biphasic dependency with an initial decrease in reduction rate with increasing force before a shift to the expected behaviour. Treating the two regimes as two independent curves following Arrhenius equation, reveals an initial negative transition state distance $\Delta x \sim 0.79$ nm followed by a positive transition state distance $\Delta x \sim 0.17$ nm. From a structural perspective this corresponds to bond shortening at low

forces and bond elongation at higher forces. This behaviour is highly unexpected given that the attacking atom on the enzyme is a deprotonated cysteine, and experiments using isolated cysteine residues follow the simple single barrier behaviour. These results can be rationalized considering the linearity of the Walden transition state and the conformational constraints imposed by the active enzyme site; depending on the force applied the bond, an atomic rearrangement is required to obtain the linear transition state within the active site of the enzyme (Fig. 1.14 e, f).

This battery of work demonstrates that single molecule approaches have the power to reveal the rich force dependent reactivity of a bimolecular reaction^{174, 179, 180, 181}. Furthermore, this complex relationship between mechanical force and chemical reactivity occurring within the core of a single protein has profound repercussions for the nanomechanical response of proteins^{165, 182}. It is therefore tempting to speculate that, given there are number of oxidized proteins that are naturally exposed to mechanical stress, this intrinsic link between chemical reactivity and mechanical force may be a naturally employed strategy to tune protein extensibility¹⁸³.

DISULFIDE BOND REFORMATION

Given that the incorporation of a single disulfide into a protein has important knock-on effects on its extensibility, it is imperative to understand the molecular mechanisms governing disulfide formation within the context of protein folding – a process known as oxidative protein folding^{184, 185, 186}. Towards this endeavour there have been a handful of innovative single molecule force spectroscopy studies that have targeted understanding both the pathway and the kinetics that govern this process.

One route to achieve an oxidised protein is via disulfide isomerization. In this instance, the dynamic nature of the disulfide permits thiol-disulfide

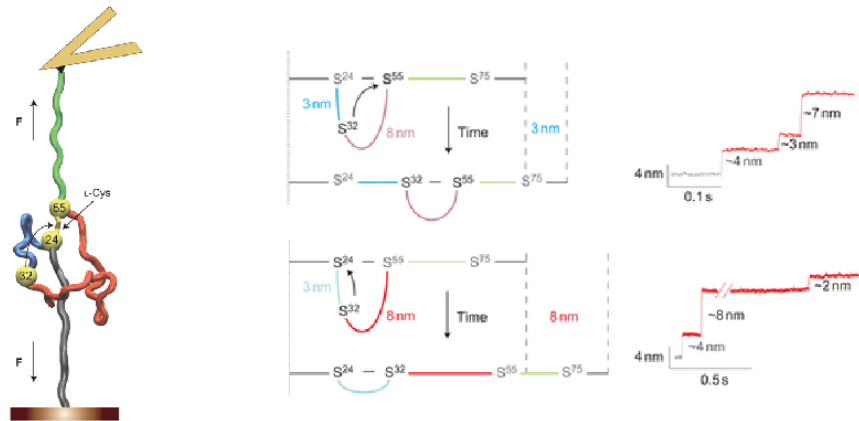
interchanges, whereby a free thiolate attacks an existing disulfide bond, leading to the simultaneous rupture of the existing bond and formation of a new one in a different location (fig. 1.15 a). It is believed that this shuffling process is a key pathway for proteins to acquire their native disulfide bond, especially in cases where the protein contains large numbers of cysteine residues. Alegre-Cebollada *et al* engineered a protein to contain two disulfide bonds at specific locations¹⁸⁷. Combining a mechanical perturbation with a low molecular weight thiol caused the rupture of one of the disulfide bonds, generating a protein with one remaining disulfide and at least one free thiolate. As long as the free thiol was not under force, it was able to attack the disulfide leading to a newly formed disulfide in a new location. By carefully monitoring the molecular elongation, it was possible to infer the location of both the *attacking* and the *attacked* thiol. This provided a platform to directly measure the kinetics of protein isomerization with single bond resolution for the first time. While this shuffling process does not alter the number of covalent bonds within the protein, it does modulate the extensibility of the protein under force depending on the precise location of the bond and the number of amino acids hidden behind the bond.

An alternative prominent biological strategy to achieve successful oxidative folding is through a dedicated enzyme system. In particular, the protein disulfide isomerase (PDI) catalyses the formation (and rupture) of disulfide bonds in both eukaryotes and bacteria¹⁸⁴. In order to capture this process using single molecule approaches, Kosuri *et al* designed a five-pulse-force protocol, capable of probing both protein folding and disulfide reformation (fig. 1.15 b). These experiments revealed that the enzyme plays a passive role, serving as a molecular placeholder during the protein folding process that then drives the formation of the bond when in a near native conformation¹⁸⁸. An analogous experiment conducted with the bacterial enzyme DsbA reveals a similar

mechanism, albeit differing kinetics. These studies appear to show a consistent molecular mechanism presiding over enzyme mediated catalysis, demonstrating that this ubiquitous interaction plays a fundamental role in regulating protein mechanics¹⁷⁹.

These experiments demonstrate that the outcome of mechanical protein folding is intricately linked to chemical reactivity with profound implications on the nanomechanical properties of the resulting folded protein. Despite this phenomenal progress, it remains to be seen whether the reactivity of disulfide rupture is in any way linked to the reverse reformation process. In particular, while it has been shown that the chemical properties of the attacking nucleophile can alter the energy landscape of disulfide reduction, it is yet to be seen whether this will then have repercussions for the reverse oxidative protein folding reaction.

Disulfide Isomerization:



Enzymatic Oxidative Protein Folding:

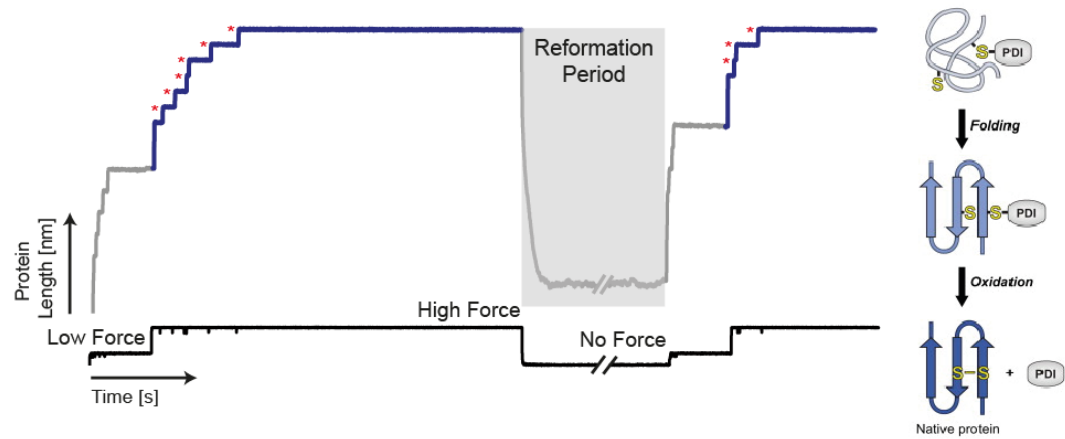


Figure 1.15. Single molecule force clamp spectroscopy captures the formation of an individual covalent disulfide bond. (a) An Ig domain of titin is mutated to contain two disulfide bonds in specific locations. A combination of mechanical force and l-cysteine as a nucleophile leads to the rupture of the first disulfide. This rupture results in a newly created free thiol that is available to attack the remaining disulfide. This nucleophilic attack results in the simultaneous rupture of a disulfide and formation of a new one in a new location. This force clamp approach allows the direct observation of isomerization and thus disulfide formation. **(b)** A five-pulse force protocol elucidates the disulfide reformation mechanism by the enzyme protein disulfide isomerase (PDI). The first two mechanical pulses unfold the protein (grey) and rupture the disulfide (blue). The subsequent quench pulse at $F=0$ pN collapses the protein and brings all amino acids close in conformational space. The subsequent two high force pulses probe the outcome of the quench period – revealing that PDI acts as a molecular placeholder, only reforming the disulfide bond once the protein has attained its native state. (Adapted from reference ¹⁸⁷ and ¹⁸⁸)

1.4 SCOPE AND OBJECTIVES

In this thesis I aim to build upon this arsenal of investigations to gain a deeper and more comprehensive understanding of the relationship between mechanical force and chemical reactivity within the context of a single protein. In particular, I will use single molecule force spectroscopy AFM to address two key topics of protein mechanochemistry; organometallic bonds and non-enzymatic oxidative folding.

Firstly I will probe the mechanical properties of the bonds that coordinate a single metal ion into the core of a single protein. The majority of single molecule experiments on metalloproteins have used the protein rubredoxin as the platform to study the mechanics and mechanisms of Fe-S and Zn-S bond rupture. While this approach has transformed our understanding of the mechanical properties of organometallic bonds, we are far from a comprehensive picture that links the metal chemistry and the local protein environment to the nanomechanical properties. In **chapter 3**, I shall use single molecule force spectroscopy in force-extension mode to study the mechanical unfolding of two copper binding proteins from the cupredoxin family, azurin and plastocyanin.

Secondly, I shall use single molecule force clamp spectroscopy to investigate two distinct pathways to achieve and modulate non-enzymatic oxidative protein folding. While there have been a number of revolutionary studies using single molecule force clamp spectroscopy to elucidate the molecular mechanisms that give rise to successful enzyme mediated disulfide bond formation, our understanding of alternative non-enzymatic routes has significantly lagged behind. On one hand, the addition of reactive oxygen species can cause an increase in the oxidation state of cysteine to a moiety known as sulfenic acid (SOH). It has been postulated that this chemical species can lead to the formation of a disulfide bond, however this process has largely

evaded characterisation due to the lack of approaches that can follow its fleeting and reactive nature. In **chapter 4** I will use force clamp spectroscopy to follow the folding of a single protein upon the post translational modification (PTM) by a single sulfenic acid moiety.

On the other hand, a non-redox means to achieve the formation of a disulfide bond is through thiol-disulfide exchange. In this case, a free protein thiol attacks and ruptures an existing disulfide bond, simultaneously forming a new disulfide bond in a different location. Inspired by this mechanism, in **chapter 5** I rupture an existing protein disulfide bond using a low molecular weight thiol. This rupture leads to the formation of a newly formed mixed disulfide between one of the protein thiols and the previously attacking thiol, and a single free protein thiol. I use this approach as a platform to study the link between the chemical reactivity of the mixed disulfide species and the outcome of oxidative folding.

Altogether, this thesis aims at providing new insights into the molecular mechanisms that relate the chemical reactivity of the cysteine residue with protein nanomechanics.

2. EXPERIMENTAL TECHNIQUES

Single molecule force spectroscopy atomic force microscope is a powerful tool for studying the behaviour of individual molecules upon application of an external mechanical perturbation. The success of this technique is reliant on a combination molecular biology techniques and precise experimental configuration. This chapter explains the principles of both these vital components used to collect the data presented in this thesis.

2.1 PROTEIN PREPARATION

All protein engineering and purifications required to conduct the single molecule experiments were kindly conducted by Miss Ainhoa Lezamiz and Dr Palma Rico. Below is a brief description of the protocols used for the construction and purification of all polyproteins (Fig. 2.1). I have provided the exact details of the molecular biology techniques for each chapter in the corresponding appendix.

PROTEIN ENGINEERING & PURIFICATION

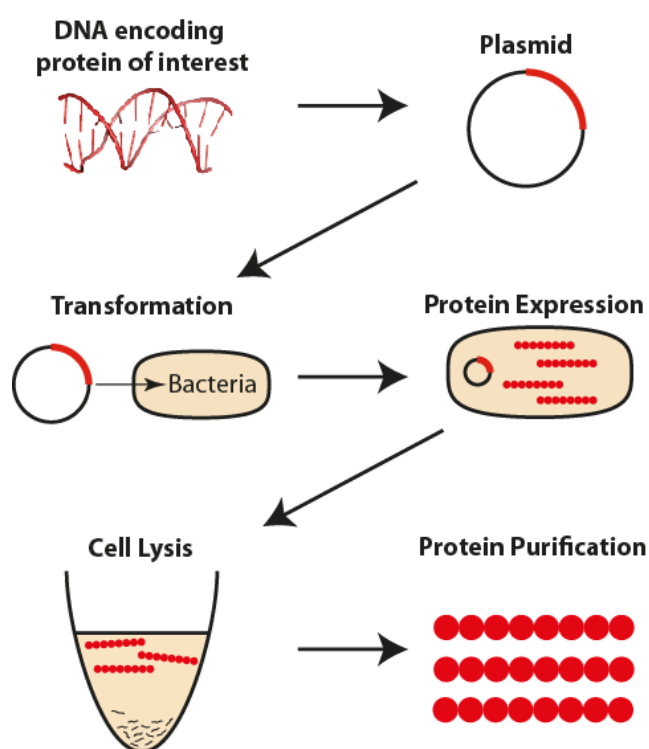


Figure 2.1 Schematic representation of the molecular biology steps to produce a polyprotein for single molecule experiments. The DNA encoding the protein of interest is inserted into a plasmid. The plasmid is transformed into bacteria which leads to the expression of the protein. The cells are disrupted such that the protein of interest can be accessed. The proteins are then subjected to two rounds of purifications to obtain a pure protein fraction.

The first step towards polyproteins for single molecule experiments is to obtain the DNA sequence that encodes for the protein of interest. In all of the single molecule work conducted in this thesis we used polyproteins composed of 8 proteins; either the repeats of the same domain or repetitive pattern of two different domains. These multidomain constructs are created using

recombination molecule biology techniques. In this approach, the DNA of interest is amplified by polymerase chain reaction (PCR). Then restriction enzymes are added that leave 'sticky ends' on the DNA that are able to fuse with complimentary DNA ends. This DNA ligation process is conducted within a bacterial expression vector – known as a plasmid. In all protein constructs, a histidine tag is encoded into the gene at one termini for purification purposes, and a cysteine tag is encoded at the other termini to provide a chemical anchor in the single molecule experiments.

Once the complete gene is encoded in the plasmid, the plasmids are transformed into bacteria, causing the expression of the encoded polyprotein. To access these expressed proteins, the cells need to be disrupted. This process can be either induced with mechanical force or chemical denaturants – a process known as cell lysis. The disruption method used in this thesis is with mechanical force is performed with a French press machine. The resulting lysate is centrifuged at high speed and the protein is extracted from the supernatant and the pellet is discarded.

In order to obtain clean and uncontaminated protein for the single molecule experiments, the proteins are subjected to two rounds of purification. The first purification process is with HisTrap cobalt affinity column. The histidine tag that is engineered at the N-terminus of the polyprotein specifically binds to the cobalt affinity resin, and all other contaminants are eluted. The protein of interest can then be eluted with the addition of imidazol, which binds to the resin with high affinity and displaces the bound protein. A second purification is then performed with ÄKTA fast protein liquid chromatography on a superdex 200 column, which separates the proteins according to the size of the molecule. The protein concentration is typically estimated using a Bradford assay and the remaining sample is stored at 4°C in PBS buffer until needed for single molecule experiments.

2.2 SINGLE MOLECULE FORCE SPECTROSCOPY ATOMIC FORCE MICROSCOPE

BACKGROUND

The first atomic force microscope (AFM) were performed in 1986 by Binnig, Quate and Gerber. This first AFM was used to obtain topographic images in contact mode. Since then, there have been considerable efforts towards developing the AFM into a versatile cross-disciplinary approach used for a range of applications, including material science, polymer chemistry and pharmacology. In particular, a major advance in both technology and molecular biology techniques has propelled this set-up to the forefront of single molecule force spectroscopy techniques. In this novel application, an individual molecule is tethered between the AFM tip and a highly sensitive piezo electric actuator. As the piezo actuator retracts from the tip, the tethered molecule experiences a mechanical perturbation. A collimated laser reflected off the back of the cantilever onto a four-quadrant photodiode monitors any changes in cantilever deflection. Calibration of the cantilever spring constant, k , enables the force to be deduced from the deflection via Hooke's law, $F = -kx$.

The AFM is arguably the most versatile of all the single molecule force spectroscopy approaches. The set-up can readily operate in two different modes (i) constant velocity mode and (ii) the addition of a feedback between the photodiode and the piezo allows the maintenance of a constant cantilever deflection and hence facilitates force clamp. Furthermore, the AFM can apply and measure a wide range of forces spanning from piconewton to nanonewton force depending on the spring constant of the chosen cantilever.

Despite these advantages, the intrinsic high stiffness of any cantilever limits the low force resolution of the approach. The lower force limit is fundamentally coupled to the spring constant of the probe given by eqn. (2.1)

$$\delta F = \sqrt{k k_B T} \quad (2.1)$$

where k_B is the Boltzmann constant, T is the absolute temperature and k is the cantilever spring constant. The typical spring constant used in single molecule protein experiments is $\sim 17 \text{ pN.nm}^{-1}$ giving a theoretical lower force limit of $\sim 10 \text{ pN}$. Given the tremendous number of biological processes occur at low forces, additional techniques, such as magnetic and optical tweezers, have been developed with high resolution in the low force regime in order to circumvent this limitation.

SETTING UP AN EXPERIMENT

The single molecule AFM work carried out in this thesis was obtained on either a commercially available set-up (Luig's and Neumann GmbH) or a home-built system (Fig. 2.2). In the following section I shall detail how to set-up an experimental on the Luig's and Neumann GmbH system, however both the working principles and the software interface are very similar on both configurations.

The experimental buffers used in each experiment are detailed in the appropriate chapter. Prior to every experiment the buffer was filtered to ensure it was free of contaminants (VWR, $0.2 \mu\text{m}$ pore-size).

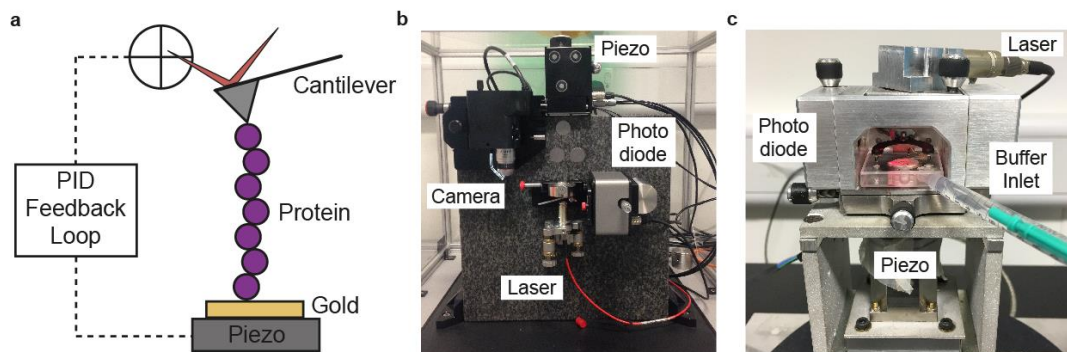


Figure 2.2. The atomic force microscope. (a) Schematic design of the atomic force microscope. The polypeptide construct (purple circles) is tethered between the cantilever tip and a gold substrate. The protein has a cysteine tag engineered at the end to allow specific binding to the gold. (b) Picture of the Luig's and Neumann AFM. (c) Picture of the home-built AFM.

EXPERIMENTAL CONFIGURATION

The following list details the necessary components to perform single molecule measurements:

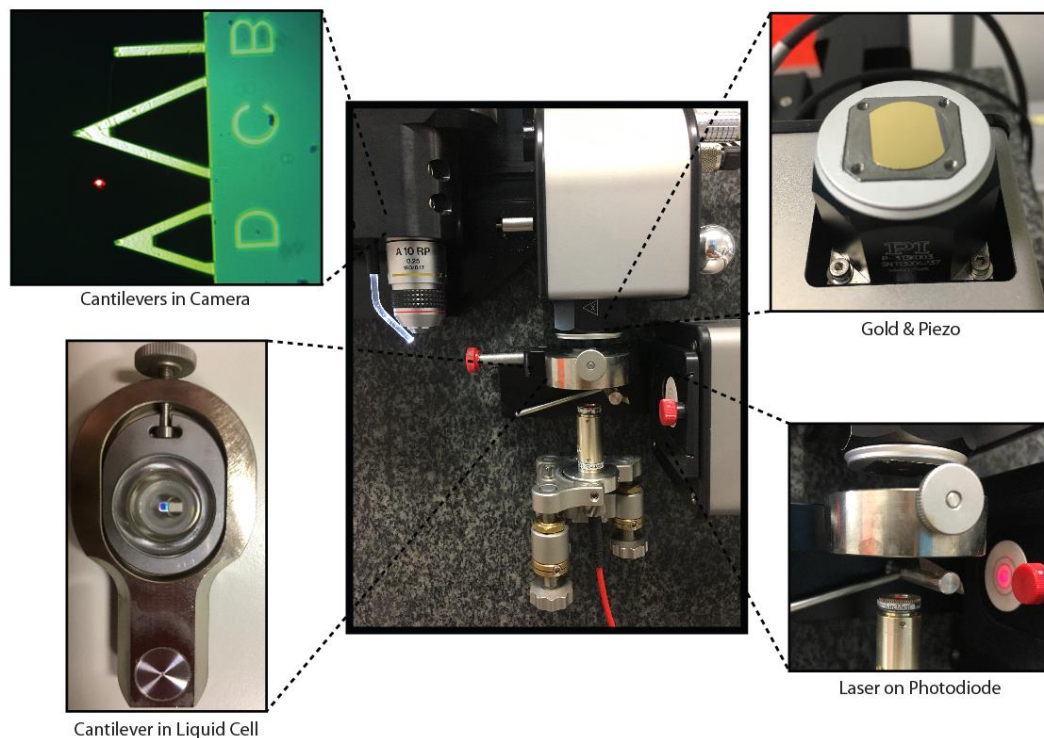


Figure 2.3. Magnification of key components of the atomic force microscope. *From bottom left clockwise:* The cantilever is clamped into a liquid cell that holds ~25 μ l volume of buffer. A camera allows the user to visualise the cantilevers and thus align the laser onto the back of the tip. A gold coated cover slide is attached to the piezo actuator using vacuum grease. When properly aligned the laser is reflected into the centre of the photodiode.

The cantilever: The cantilever typically has a reflective surface on the back-side which reflects the laser, and a sharp probe underside that makes contact with the sample. The spring constant properties of the cantilever determine its suitability to perform the experiment. In single molecule experiments the protein is relatively soft and therefore cantilevers with a low spring constant are selected. In this thesis all measurements were made with MLCT tips C or D (Fig. 2.3) (Bruker) with a typical spring constant of 17-35 pN.nm⁻¹. In order to perform experiments the cantilever is clamped into a liquid chamber. The cantilevers can be visualised by a camera that is positioned above the chamber (Fig. 2.3).

The piezoelectric actuator: A piezoelectric material will both (i) generate an electric charge as a result of mechanical stress, and (ii) undergo a mechanical strain in response to an applied electric field. In the AFM, we have a quartz crystal piezo (PI instruments) that undergoes nanometer movements upon an applied voltage (Fig. 2.3). Prior to the first experiments conducted on this machine, we calibrated the relationship between voltage and distance for the piezo to ensure the reliability of all acquired data. An additional particular feature of this piezo is that the travel range is short (~ 500nm), however this small travelling distance affords very high spatial (< nm) and temporal (<2 ms) resolution for protein unfolding events.

The gold cover slide: A gold-coated glass cover slide is mounted atop of the piezo with vacuum grease (Fig. 2.3), and facilitates protein binding via a cysteine tag engineered at the termini of each polyprotein. The cover slides are prepared by sonicating glass cover slides (Ted Pella) in a 1% cleaning solution (Hellmanex™ III, Sigma-Aldrich) for 15 minutes. The slides are then washed thoroughly in MilliQ water, before undergoing 3 sonication cycles of isopropanol (Sigma-Aldrich) followed by MilliQ water. The slides are then individually dried with nitrogen and loaded into a specially designed holder. The slides are mounted into a thermal evaporator (Leybold U300). The desired deposition materials – in this case nickel/chromium and gold (Ted Pella) - are loaded into two separate tungsten boats. In this evaporator configuration, a high current (~ 1 Amp) is applied across the boats thus heating the sample, leading to the evaporation and deposition of the material into the chamber. Between the Ni/Cr deposition (~10-20 nm) and the gold deposition (~40-60 nm) the system is left to cool and return to an efficient vacuum for a minimum for 5 hours. Each evaporation produces ~50 cover slides, these are stored in a desiccator until needed.

The protein: A small volume (~ 1-3 µl) of protein is spread onto a freshly prepared gold cover slide and left to absorb for ~ 5mins.

The laser: The collimated laser spot (Schaffter + Kirchhoff, $\lambda=660$ nm, $P=5$ mW) is positioned onto the back of the cantilever and a mirror then deflects the beam onto the centre of the photodiode. The intensity of the laser beam can be adjusted in the programme software.

The photodiode: A photodiode is a device that converts incident light into electrical signal. The four-quadrant photodiode in the AFM tracks the movement of the laser, which is intrinsically linked to the cantilever deflection.

EXPERIMENTAL CALIBRATION

The spring constant of each cantilever slightly differs, and as such it is imperative to measure the exact properties to achieve correct force measurements. Prior to every experiment we perform a two-step calibration process that enables the forces to be accurately deduced from Hooke's law.

The first step relies on the equipartition theorem, which states that the total energy of a system is equally shared amongst all degrees of freedom. In terms of the cantilever, this means that the kinetic energy can be related to the thermal energy through equation (2.2)

$$\frac{1}{2}k\langle x^2 \rangle = \frac{1}{2}k_B T \quad (2.2)$$

where k_B is the Boltzmann constant, T is the temperature, k is the cantilever spring constant and x is the displacement of the cantilever. Rearranging eqn 2.2 demonstrates that the cantilever can related to the mean square of the displacement:

$$k = \frac{k_B T}{\langle x^2 \rangle} \quad (2.3)$$

This is measured by taking a power density spectrum of the cantilever when it is far away from the surface to ensure that any movement is purely down to the thermal fluctuations. The first harmonic peak of the power density spectrum is due to the fluctuations of the cantilever and thus the shape of the peak is related to k (Fig. 2.4).

The second step of the calibration process is to calibrate the deflection sensitivity. This step calibrates the photodiode response by measuring the change in voltage on the photodiode with respect to the movement of the piezo actuator when the tip is pushing against a hard surface. The slope of the force extension curve is thus:

$$slope = \frac{\Delta V}{\Delta z} \quad (2.4)$$

where Δz is the displacement of the piezo actuator and ΔV is the change in voltage.

Once these two steps have been completed the calibration process is done and the everything is prepared to being experiments. However, if at any time during the experiment the position of the laser on the cantilever is adjusted it is imperative to repeat the calibration steps.

USER SOFTWARE INTERFACE

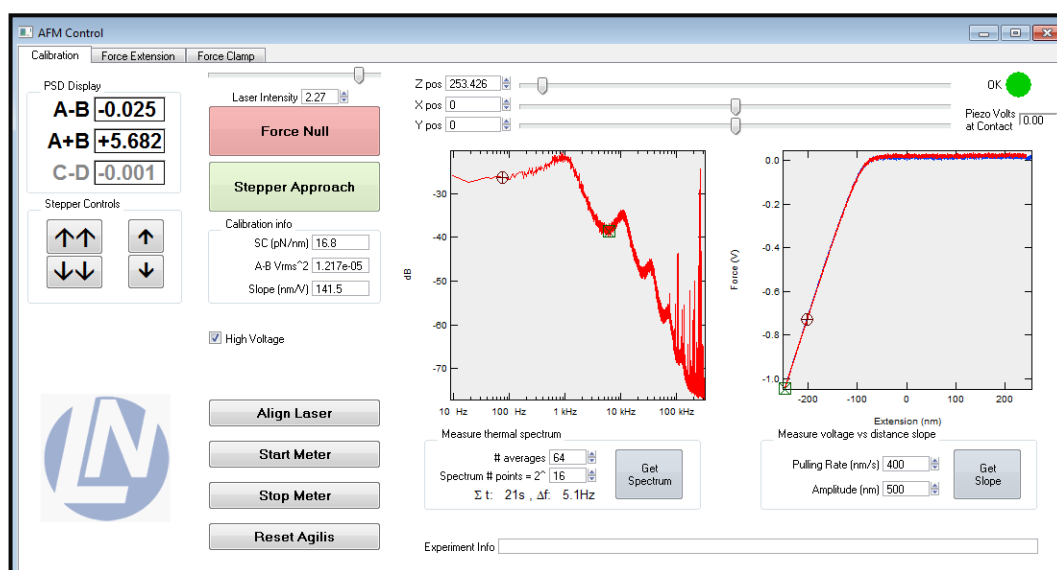


Figure 2.4. Cantilever calibration steps. A picture of the operating software that allows cantilever calibration.

Align Laser: This initialises the camera, allowing the user to visualise both the cantilever tips and the laser spot.

Laser Intensity: The slider allows the user to set the laser intensity. Typically this value is set to give a typical A+B reading of 5-6 V.

PSD display: The PSD is the output display of the position of the laser on the photodiode. The position where the (A+B) is maximum, and both the (A-B) and the (C-D) is minimum, indicates that the laser spot is central on the photodiode. The photodiode output is most accurate at deviations close to the centre.

Force Null: Given that the photodiode sensitivity is highest closest to the centre, it is important to try and maintain this throughout the course of the experiment. In this set-up the photodiode is mounted on a stepper motor. The force null button causes the stepper motor to move the photodiode until the laser spot is returned to the central position. This not only ensures accurate

measurements, but also helps to compensate for any drift present in the experiment.

Stepper Approach: The piezo actuator is mounted on top of a stepper motor. This motor allows the piezo to be moved much longer travel distances than the 500 nm range. The stepper approach button controls the piezo stepper motor and causes the piezo to approach the surface in preparation for starting experiments.

Stepper Controls: These stepper buttons control the movement of the piezo stepper motor and allow the user to approach/retract the cantilever and the surface.

Z, X, Y positions: The positioning controls allow the user to move the X, Y, Z positions of the piezo. The X and Y axis control the position of the cantilever on the surface and the Z axis controls the height with respect to the surface.

Measure Thermal Spectrum: This button starts the acquisition of the power density spectrum. The user is required to select the first harmonic using the cursors.

Measure voltage vs distance slope: This performs a single force-extension trajectory and determines the slope of the region when the cantilever is pushing against the surface.

Calibration Info: This box displays the results of the calibration process and allows the user to verify that the spring constant is in-line with the manufacturer's values.

FORCE EXTENSION MODE

Below is displayed the software interface for performing force extension measurements in the Luigi's and Neumann AFM (Fig. 2.5).

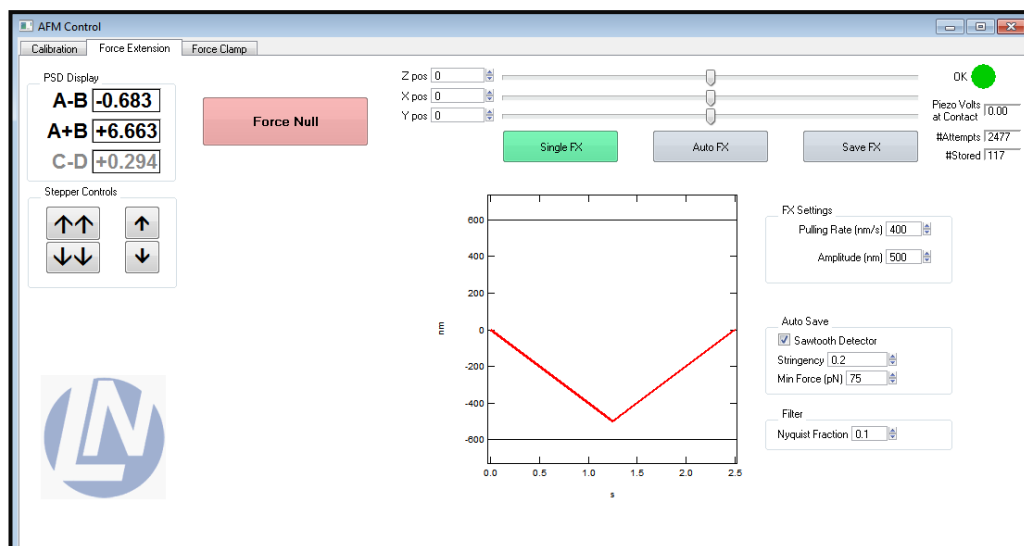


Figure 2.5. The user controls for operating force-extension experiments. A picture of the operating software used to perform constant velocity measurements.

Single FX, Auto FX, Save FX: These buttons let the user acquire and save force extension trajectories.

Pulling Rate: The pulling rate determines how quickly the piezo will approach and retract during a trajectory. The pulling velocity will influence the protein unfolding / bond rupture force according to eq. 1.1, therefore it is imperative to only compare molecular mechanical stability data at the same pulling velocity. I have conducted all the experiments in this thesis at an arbitrary pulling velocity of 400 nm.s^{-1} , as this is the velocity most commonly used in the literature and thus is better for force data comparison.

Amplitude: The amplitude controls the total travel distance of the piezo during each trajectory. In this Luig's and Neumann set-up the total length of the piezo is 500 nm. The total length of the protein-of-interest dictates the amplitude.

Auto save: The auto save function limits the saved number of trajectories, discarding traces containing nothing and saving trajectories of protein unfolding. The function works by detecting repetitive events, and therefore

this is not suitable for proteins featuring an pathway does not give rise to a regular, equal sized saw tooth pattern.

FORCE CLAMP MODE

Below is displayed the software interface for performing force clamp measurements in the Luig's and Neumann AFM (Fig. 2.6).

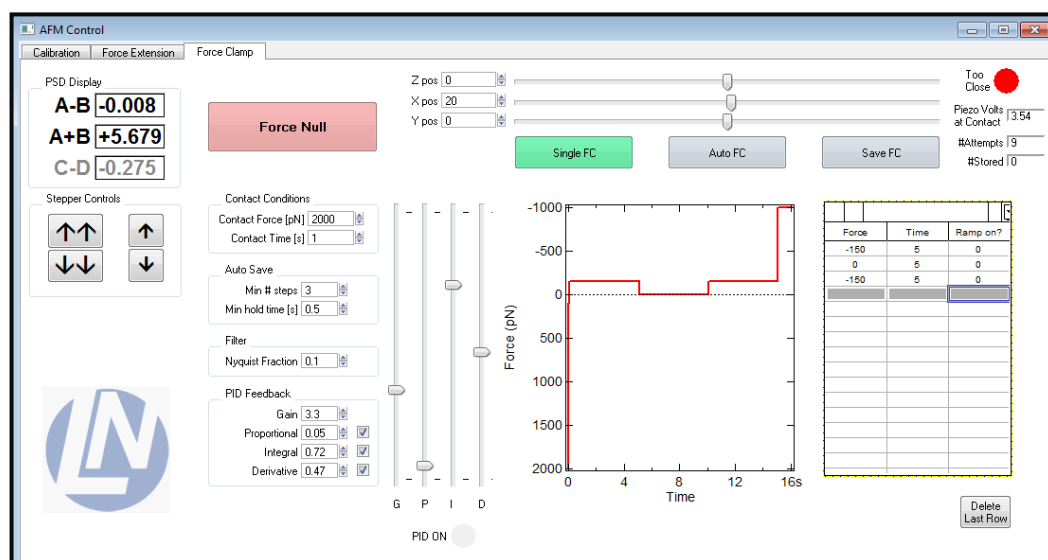


Figure 2.6. The user controls for operating force-clamp experiments. A picture of the operating software used to perform constant force measurements.

Contact force & time: These parameters control how long and for how much force the cantilever tip should push against the surface. A longer time and more force enhances the probability of 'fishing' a protein, however also enhances the probability of spurious interactions, so should be varied depending on the state of the experiment.

PID feedback: These controls allow the user to adjust the PID feedback settings in order to minimise the feedback response time of the experiment. The goal is to have the settings as fast as possible without causing the piezo to try to move faster than its resonant frequency, generating a ringing noise which will ultimately damage the piezo. This calibration process is done prior to every force clamp experiment.

Force table: The force table allows the user to define the force protocol.

DATA ANALYSIS

All data analysis was performed using software custom written in Igor pro (courtesy of Prof. Fernandez lab, Columbia University).

FORCE EXTENSION PROTEIN UNFOLDING ANALYSIS

Each protein unfolding event is manually fitted with the worm like chain equation³⁴ (eq. 2.5)

$$F(x) = \frac{k_B T}{p} \left[\frac{1}{4} \left(1 - \frac{x}{l_c} \right)^{-2} + \frac{x}{l_c} - \frac{1}{4} \right] \quad (2.5)$$

Where F is force, x is extension, k_B is the Boltzmann constant, T is the absolute temperature, p is the persistence length and l_c is the contour length of the molecule.

Typically, when fitting the data with the WLC model the persistence length is maintained between 0.2-0.6nm to reflect the average size of a single amino acid (~0.36nm).

In all experiments conducted on polyproteins containing repeats of I27 separated by a protein of interest, the trace is selected according to the number of observed I27 unfolding events. Only traces exhibiting 3 or 4 I27 unfolding events are selected. In the case of 3 I27 events, this ensures a minimum of 2 unfolding events of the protein of interest. In experiments conducted on eight repeats of the same protein, the trajectory is only selected if it contains five or more unfolding events.

FORCE CLAMP ANALYSIS

DISULFIDE RUPTURE KINETIC STUDIES

The experiments probing the disulfide bond rupture kinetics in the presence of different nucleophiles, only recordings showing the signature of at least 5

unfolding events (15 nm steps) followed by 5 reduction events (10 nm steps) were analysed. If a trace contained an unfolding event in the high force pulse it was not further considered for analysis. To obtain the reduction rate at each particular pulling force, we summed and normalized numerous ($n = 15-35$) reduction recordings for each particular pulling force. To obtain the reduction rate at the pulling force, we fitted the averaged traces with single an exponential fit.

DISULFIDE REFORMATION STUDIES

For the oxidative folding experiments only recordings showing the signature of at least 5 unfolding events (15 nm steps) followed by 5 reduction events (10 nm steps) were analysed. Furthermore, the trajectory was only considered if the probe pulse reached the same length as the initial pulse. The number of independent observations, n , for a particular condition was counted as the total number of successful trajectories. Standard error of the mean (SEM) for the refolding fraction was estimated through the bootstrap method, where each recording was treated as an independent data point.

3. THE MECHANOCHEMISTRY OF COPPER REPORTS ON THE DIRECTIONALITY OF UNFOLDING IN MODEL CUPREDOXIN PROTEINS

Understanding the directionality and sequence of protein unfolding is crucial to elucidate the underlying folding free energy landscape. An extra layer of complexity is added in metalloproteins, where a metal cofactor participates in the correct, functional fold of the protein. However, the precise mechanisms by which organometallic interactions are dynamically broken and reformed upon (un)folding are largely unknown. Here we use single molecule force spectroscopy AFM combined with protein engineering and MD simulations to study the individual unfolding pathways of the blue copper proteins azurin and plastocyanin. Using the nanomechanical properties of the native copper centre as a structurally-embedded molecular reporter, we demonstrate that both proteins unfold via two independent, competing pathways. Our results provide experimental evidence of a novel kinetic partitioning scenario whereby the protein can stochastically unfold through two distinct main transition states placed at the N- and C- termini that dictate the direction in which unfolding occurs.

*This chapter is based on the article published 3 August 2015 in Nature
Communications (Attached in appendix IV)*

3.1 INTRODUCTION

During folding, unfolded polypeptides progressively narrow down their conformational space search, often briefly visiting intermediate conformations reminiscent of localized energy minima^{189, 190, 191} until they eventually reach the functional natively folded structure^{192, 193, 194, 195, 196}. Understanding the molecular mechanisms underpinning each intermediate as well as identifying the sequential order statistics in which each conformation occurs along an individual (un)folding pathway, remains a challenge in the protein folding field¹⁹⁰.

Single molecule experiments are ideally suited to detect individual short-lived protein conformations, often occurring through statistically rare pathways that would be otherwise smeared out in the ensemble. In particular, force experiments excel at precisely mapping out individual mechanical (un)folding pathways by monitoring the protein extension along a well defined 1D reaction coordinate; the protein's end-to-end length. These experiments have revealed signatures of complexity in the unfolding trajectories of a wide variety of proteins. While earlier works reported the presence of well-defined mechanical intermediates for a number of topologically distinct small proteins^{197, 198, 199, 200, 201}, more recent studies on larger multidomain proteins have experimentally demonstrated the concept of kinetic partitioning^{202, 203, 204, 205}, where both the number and the position of these intermediates stochastically change after the main unfolding event, which remains the same in all trajectories^{86, 88, 110, 206, 207, 208, 209}. A further layer of complexity would be added if a protein also exhibited stochastic behaviour in the rupture of the main mechanical clamp, defining a multiplicity of pathways from the early stages of unfolding.

A key advantage of single molecule atomic force microscopy (AFM) mechanical experiments is the ability to directly measure the mechanical

activation of individual chemical bonds. These measurements enable a direct correlation between the strength and nature of the bond and its mechanical stability. For example, while mechanical protein unfolding involves the disruption of a key set of hydrogen bonds, typically requiring forces ranging from ~20-200 pN²¹⁰, the mechanical stability of covalent bonds is significantly higher, up to ~3 nN². In particular, covalent disulfide bonds cannot be broken upon the application of force, unless a nucleophile is added to the solution²¹¹. The repertoire of chemical interactions that are ruptured during the mechanical unfolding trajectories have recently been extended to include organometallic bonds^{105, 110}. Metals play an important structural and biological role in a vast variety of proteins, spanning one-half of all known protein crystal structures in the Protein Data Bank. Hence, understanding the mechanochemistry of individual organometallic bonds and the kinetics of metal release from the protein core is essential when studying the mechanical (un)folding of metalloproteins. While pioneering recent studies have mainly focused on the mechanical properties of iron-binding proteins such as rubredoxin^{39, 105, 110}, the nanomechanics of other organometallic bonds such as copper, which is an essential metal in biological processes involved in a variety of redox reactions, remains vastly elusive.

Here we use a combination of protein engineering and single molecule force experiments to study the individual unfolding pathways of two model blue copper proteins: azurin, involved in electron transfer processes in the redox system of *Pseudomonas aeruginosa* and other bacteria²¹², and plastocyanin^{213, 214, 215}, which functions as an electron carrier in the photosynthetic electron-transfer chain, shuttling electrons between the cytochrome *b6f* complex and Photosystem-I in all higher plants as well as in a number of green algae and cyanobacteria such as *Phormidium laminosum*²¹⁶. Azurin is topologically rich, exhibiting a β -barrel structure organised in a Greek key motif (Fig. 3.1 a, b)²¹⁷,

²¹⁸. Crucially, its structure harbours the three main families of chemical determinants that confer mechanical stability to proteins, namely the presence of two sets of key hydrogen bonds at each of its termini, a penta-coordinated copper ion and a native covalent disulfide bond. Such interesting structural diversity renders azurin particularly attractive from the mechanochemistry perspective. By using the native copper as an internal “built-in” mechanical intermediate reporter, we demonstrate that azurin can unfold via two competing independent pathways. We show that the protein can stochastically unfold along either the N-C or the C-N direction, suggesting that the barriers to unfolding from both termini are largely equivalent at the probed pulling velocity. Each individual unfolding pathway is unambiguously distinguished by the position of the subsequent rupture of the copper-ligand bond and further supported by directed mutagenesis and molecular dynamics (MD) simulations. Crucially, the mechanical stability of each individual organo-copper bond is surprisingly low despite their highly covalent character. These observations are expanded to the topologically similar plastocyanin protein, exhibiting analogous behaviour. Our results provide experimental evidence of a novel kinetic partitioning scenario where the stochastic nature of the unfolding process originates already at the very early stages, as the protein chooses between two competing main unfolding transition states that precisely dictate the unravelling sense along the termini direction.

3.2 RESULTS

THE NANOMECHANICS OF AZURIN UNFOLDING

Using single molecule spectroscopy AFM we stretched a single azurin polyprotein whereby each azurin monomer is flanked by two modules of the titin 27th Ig domain, which provide an internal molecular fingerprint²¹⁹ (Fig. 3.1 c). A typical force-extension unfolding trajectory of the (Azu-I27)₄ polyprotein marks first the unfolding of the azurin proteins (orange), occurring at ~56 pN, followed by the unfolding of the I27 modules (purple), requiring ~210 pN to unfold (Fig. 3.1 d) and exhibiting an increase in contour length of $\Delta L \sim 28$ nm²¹⁹. Fitting the WLC of polymer elasticity to the data reveals that each azurin unfolding event entails a concomitant increase in contour length of $\Delta L \sim 38$ nm (orange fits). Assuming a contour length per residue of 0.38 nm²²⁰, unfolding and extending each azurin monomer right from the termini would release $\Delta L \sim 48$ nm. However, a disulfide bond between the amino acids C3-C26 shortcuts the protein (Fig. 3.1 a, b, white), and effectively shortens it by ~10 nm. Hence, the increase of $\Delta L \sim 38$ nm that we observed is perfectly compatible with the unfolding of the whole azurin protein from a position very close to the termini in the presence of the disulfide bond (orange region in Fig. 3.1 a, Appendix I Fig. 1 and Table 1).

Close inspection of each azurin unfolding event (Fig. 3.1 d, inset) reveals, in ~80% of the occurrences, the presence of a clear mechanical intermediate (Fig. 3.2). Remarkably, the position of this intermediate is not unique, since we measured (with similar probability) events occurring either ~6 nm (dashed green fit) or ~9 nm (dashed brown fit, Fig. 3.1 d, inset) after the main unfolding event, occurring at ~54 pN (Fig. 3.1 g, grey). Despite the close proximity in the measured length defining both intermediates, the grey histogram in Fig. 3.1 f is bimodal. The mechanical disruption of such intermediate conformations occurs at forces of ~43 pN (Fig. 3.1 g, blue), with a concomitant protein

elongation of $\Delta L \sim 32$ nm or $\Delta L \sim 28$ nm, respectively (Fig. 3.1 f, blue histogram). Remarkably, in 8% ($n = 25$) of the total unfolding trajectories showing an intermediate, the protein unfolds via a 4-state unfolding pathway where both intermediates (~ 6 nm and ~ 9 nm) are readily captured (Fig. 3.1 e). Crucially, in these cases the order of occurrence of both ~ 6 nm and ~ 9 nm events is stochastic. In the remaining $\sim 20\%$ of events, the protein unfolds through a simple two-state scenario, extending a total length of ~ 38 nm, requiring forces of ~ 56 pN (Fig. 3.1 g, orange). In order to further validate the presence of the two distinct mechanical intermediates, we conducted experiments under force-clamp conditions, whereby the protein extension is recorded as a function of time (Appendix I Fig. 2). These experiments confirmed the presence of two statistically distinct mechanical intermediates.

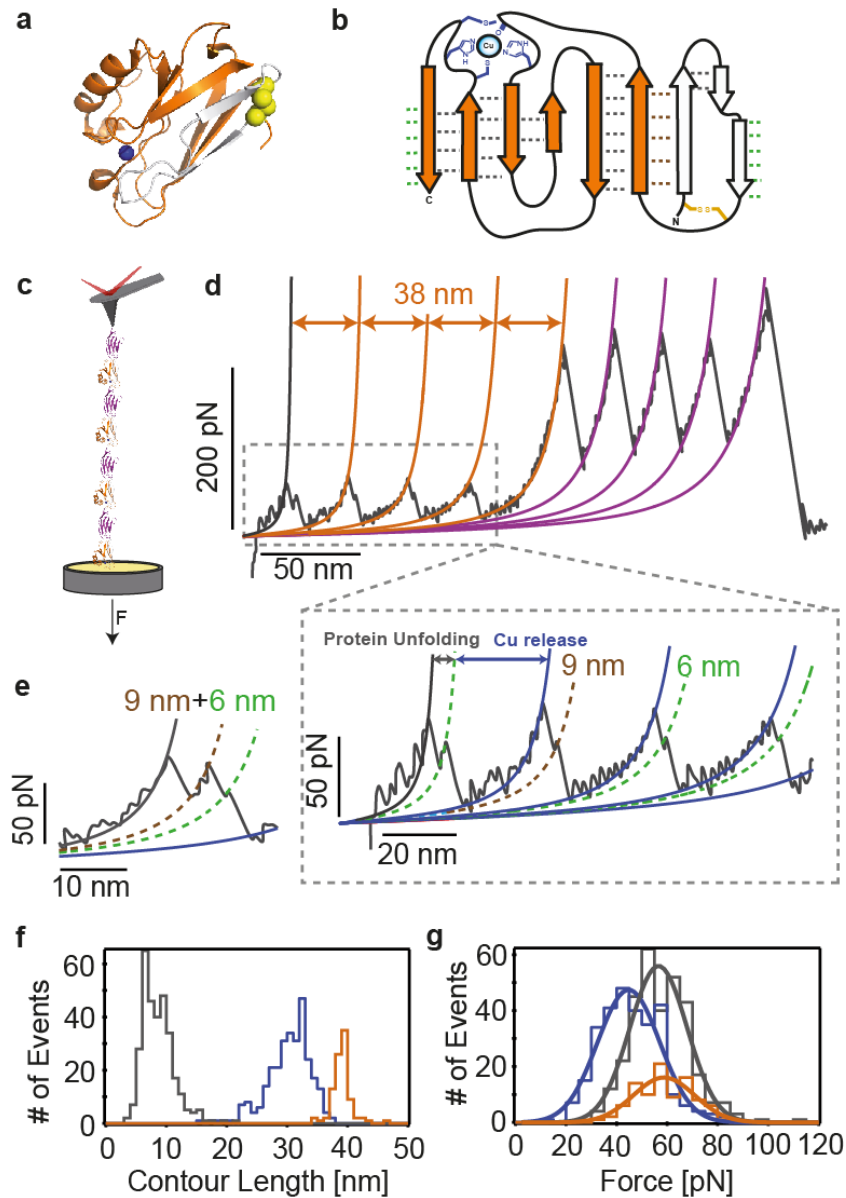


Figure 3.1 Wt-azurin unfolds through two mechanically labile intermediates. (a) Schematic structure of wt-azurin (PDB: 4AZU). (b) The residues trapped after the disulfide bond (yellow) are marked in white. The rest are represented in orange. (c) Scheme of the engineered (Azu-I27)₄ polypeptide being pulled in our AFM set-up. (d) Pulling on an individual (Azu-I27)₄ polypeptide results in an unfolding trajectory where the unfolding of azurin (orange) occurs at ~55pN, followed by the unfolding of the I27 domains, distinguished by an unfolding force of ~210pN and an increment in contour length of $\Delta L = 28$ nm (purple). *Inset:* The unfolding of wt-azurin occurs via two well-defined mechanically stable intermediates occurring at ~6nm (green dashed line) or ~9nm (brown dashed line). (e) In a few trajectories ($n=25$), a 4-state unfolding pathway is measured, where both intermediates (~6nm and ~9nm) are captured within the same trajectory. (f, g) WLC fits to the data reveal that each complete azurin unfolding event occurs concomitant to an increment in contour length of $\Delta L = 38.6 \pm 0.9$ nm, $n = 93$ (f, orange histogram), requiring an unfolding force of 56.3 ± 12.1 pN, $n = 93$ (g, orange). When unfolding occurs via a mechanical intermediate, the main unfolding event (dashed lines), occurring at 54.1 ± 11.3 pN, $n = 323$ (g, grey), elicits a protein length of <10 nm (f, grey histogram) before it reaches a mechanically resistant intermediate. The mechanical disruption of such intermediate form occurs at forces of $\sim 42.3 \pm 12.3$ pN, $n = 289$ (g, blue), with a concomitant protein elongation of $\Delta L \sim 30$ nm (f, blue histogram).

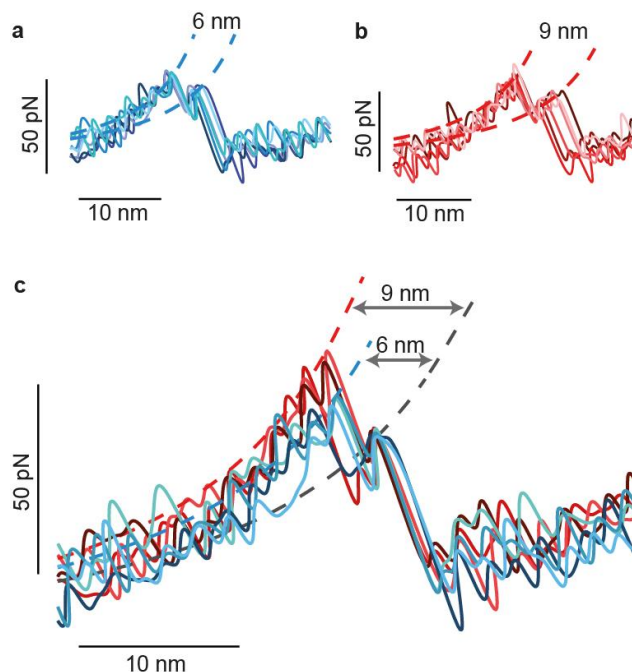


Figure 3.2 Comparison of multiple azurin unfolding events containing a single intermediate (a) Multiple trajectories displaying a ~6nm (dashed blue WLC fit) corresponding to the C-terminus unfolding up to the C112-Cu bond ($n=6$). **(b)** Overlay of $n=6$ events where azurin unfolding occurs from the N-terminus, extending the protein by ~9nm up to the Cu-N_{His46} bond (dashed red WLC fit, $n=6$). **(c)** Direct comparison of the increment in contour length associated to the unfolding from either the N-terminus (red), or the C-terminus (blue). In the representation, each trajectory is aligned at the point of metallobond rupture.

According to the presence of the copper site in the tertiary structure of azurin, we speculate that the mechanical intermediates that we measured are associated with the mechanical rupture of the copper centre. In azurin (PDB: 4AZU), copper exhibits a characteristic trigonal bipyramidal geometry, whereby the metal interacts with five key residues; Met-121, His-117, Cys-112, His-46 and Gly-45 (Fig. 3.1 b). Of these, structural and spectroscopic studies reveal that the three equatorial residues Cys-112, His-117 and His-46 exhibit the strongest interactions with the metal²¹³. Taking this into consideration, the ~6 nm intermediate would be compatible with an unfolding scenario where the protein unfolds from the C-terminus starting by disrupting the β -sheets β 2b- β 8 (Fig. 3.1 b, green hydrogen bonds) up to the Cu-S_{Cys112} bond, hence releasing the amino acids contained between the C-terminus (Lys-128) and Cys-112 (Appendix I Fig. 1). However, the ~9 nm intermediate event cannot be accounted for by this unfolding sequence. By contrast, it could be readily

explained assuming that the protein unfolds from the N-terminus instead (Fig. 3.1 b, brown hydrogen bonds). In that case, the position of the mechanical intermediate would match perfectly with the unfolding up to the Cu-N^{His46} bond. Indeed, extending the amino acids released upon unfolding the protein from the N-terminus up to His-46 (after deducting those residues that are trapped behind the disulfide bond formed between Cys-3 and Cys-26, white region in Fig. 3.1 a, b) results in a length increment of ~11 nm (Appendix I Fig. 1), consistent with the observed ~9 nm intermediate. Due to the weak character of the backbone carbonyl of Gly-45²²¹, it is highly unlikely that the rupture of such interaction with the Cu centre can account for the mechanical intermediate observed in our experiments.

THE DISTINCT UNFOLDING PATHWAYS REVEALED BY MUTAGENESIS

In order to verify the hypothesis that azurin unfolds via two competing pathways fingerprinted by the rupture of the Cu-S and Cu-N individual bonds, we first engineered a polyprotein containing the Cys112Ala mutant²²², thus removing the Cu-S_{Cys112} bond (Fig. 3.3 a). Crucially, circular dichroism experiments certified the correct folding of the protein (Appendix I Fig 4), and ICP-OES measurements revealed that, although only partially, Cu was still embedded within the azurin structure (Appendix I notes). Stretching the (AzUC_{112A}-I27)₄ polyprotein revealed the presence of only the ~9 nm intermediate (brown fits, Fig. 3.3 b; and brown histogram, Fig. 3.3 d). The absence of the ~6nm intermediate in the AzUC_{112A} case firmly suggests that the ~6nm intermediate observed in the unfolding trajectories of wt-Azu (Fig. 3.3 d) is indeed triggered by the mechanical disruption of the Cu-S_{Cys112} bond.

To further investigate the origin of the ~9nm intermediate, we studied the unfolding properties of the (AzUC_{26A}-I27)₄ polyprotein, lacking the native disulfide bond (Fig. 3.3 c). The absence of the disulfide bond was further

checked by the Ellman's assay^{223, 224}, revealing an increase in the number of free sulfhydryls in the polypeptide chain in the mutated protein as opposed to the wild-type form (Appendix I notes). As expected, each azurin unfolding event showed a larger associated increment in contour length of $\Delta L \sim 48$ nm, corresponding to the full protein length (orange fits). Akin to the wild-type case, the unfolding trajectories showed two well-defined intermediates; the first one occurring at ~ 6 nm and the second one now occurring at ~ 19 nm (dashed light brown fits). The distribution of events is shown in the contour length histogram of Fig. 3.3 e (light brown). Hence, upon removal of the disulfide bond, only the N-terminus unfolding pathway was modified by adding to the length required to reach the His-46 the protein length previously trapped behind the disulfide bond resulting in the total length increase of ~ 19 nm, thus separating both populations of intermediate events. Interestingly, the relative abundance of both pathways for this disulfide bond-depleted construct is not $\sim 50\%$ as for the wild-type construct; rather, the "N-terminus" pathway, eliciting the 19 nm intermediate, is observed in $\sim 80\%$ of the events featuring an intermediate, as observed in the normalized histogram shown in Fig. 3.3 e.

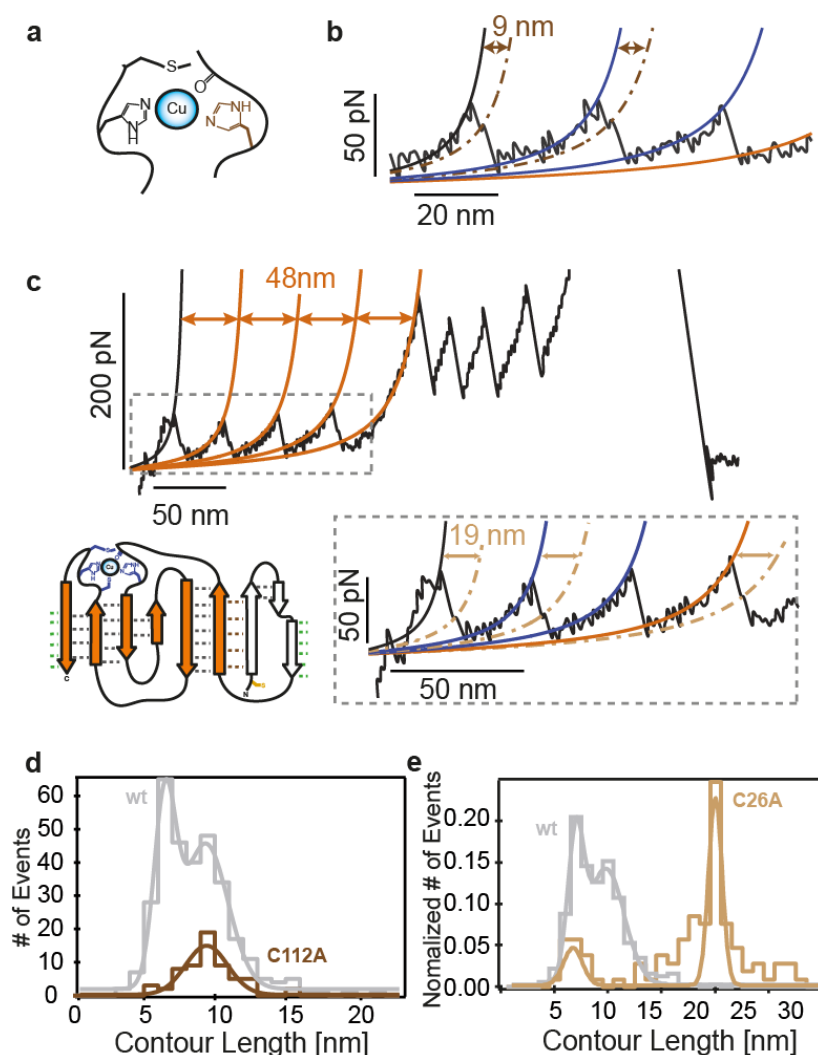


Figure 3.3 The Azu_{C112A} and the Azu_{C26A} mutations confirm the existence of parallel unfolding pathways. (a) The Azu_{C112A} mutation prevents the formation of the Cu-SCys₁₁₂ bond. (b) Upon unfolding, the Azu_{C112A} protein extends by ~9 nm (dashed brown fit) before encountering a mechanical intermediate that corresponds to the disruption of the Cu-N_{His46} bond (dark blue fit). (c) The C26A mutation prevents the formation of the disulfide bond. In the absence of the disulfide bond, each azurin unfolding event occurs concomitant to an increase in contour length of $\Delta L = 47.6 \pm 1.2$ nm, $n = 77$ (orange fits). The mechanical intermediate is now mostly observed (80% of the occurrences) at $\Delta L = 19.0 \pm 0.5$ nm (dashed light brown fits), $n = 78$, consistent with the unfolding and extension of the total protein length from the N-terminus up to His-46. In the remaining 20% of the cases in which the protein unfolds via a 3-state pattern, the protein unfolds via the C-terminus, as fingerprinted by the intermediate occurring at $\Delta L = 5.6 \pm 1.1$ nm, $n = 18$. (d) The histogram corresponding to the increment in contour length elicited after the main unfolding event measured for the Azu_{C112A} mutant reveals a main unfolding event occurring at $\Delta L = 9.4 \pm 1.5$ nm, $n = 62$ nm, which matches rather well the broader histogram of the wt-Azu data shown in Fig. 3.1f (background light grey bars), containing both intermediate events. (e) The normalized histogram corresponding to the unfolding of the Azu_{C26A} mutant shows well defined intermediates, occurring at $\Delta L = 6.2 \pm 1.0$ nm, $n = 18$ and $\Delta L = 19.5 \pm 0.7$ nm, $n = 81$. Crucially, the N-terminus pathway is highly favoured in the absence of the native disulfide bond.

THE PLASTOCYANIN PROTEIN EXHIBITS A SIMILAR UNFOLDING SCHEME



Figure 3.4 Structural similarity between plastocyanin and azurin. Alignment of the azurin (PDB:1AZU, orange) and plastocyanin (PDB:3BQV, cyan) structures highlights the similar topology of the two proteins, only differing in the length of the alpha helix (encompassing 23 amino acids) that is present in azurin but absent in plastocyanin.

Azurin is a model protein for Type 1 (T1) copper centres. T1 copper proteins are distinct from other copper proteins because of their unique geometry and ligand sets²¹². Despite the fact that the sequence identity between the T1 copper proteins is less than 20%²²⁵, the overall structural fold of different T1 copper proteins is highly conserved²¹³. Such common cupredoxin fold consists of eight β -strands arranged into a Greek β -barrel (Fig. 3.4). Inspired by such large structural homology, we investigated whether the intriguing features observed in the mechanical unfolding pathway of azurin could be generalized to other structurally similar T1 model blue copper proteins with the cupredoxin fold, such as plastocyanin (Fig. 3.5 a). To this end, we studied the mechanical unfolding pathways of the (Plastocyanin-I27)₄ polyprotein (Fig. 3.5 b). Complete unfolding of plastocyanin, occurring at ~ 75 pN (Fig. 3.5 e), results in an increase of contour length of $\Delta L \sim 37$ nm (Fig. 3.5 c, cyan fits and Appendix I Fig. 5), which is consistent with the extension of the full protein. Remarkably, in 60% of the unfolding events a mechanical intermediate was also observed (Fig. 3.5 c, inset), which, as in the case of azurin, clustered in two main length populations at ~ 6 nm and ~ 14 nm (Appendix I Fig. 6, 7), giving rise to the —

now well-separated— bimodal histogram shown in Fig. 3.5 d. Akin to azurin, in plastocyanin the copper ion is bound to the sulfur of Cys-89 and to two nitrogens (from His-39 and His-92) in the trigonal plane, and to a methionyl sulfur (from Met-97) in the axial position at a relatively longer distance (Fig. 3.5 f). Taking this structure into account, the ~6nm intermediate can be readily explained in this case by the unfolding of the protein from the C-terminus until Cys-89. In this case, after unfolding, the protein would be expected to extend 7.0 nm (Appendix I Table 2). By contrast, unfolding the protein from the N-terminus up to His-39 would extend 16.2 nm, thus consistent with the second intermediate that we observed. Crucially, mechanical disruption of the copper organometallic bonds requires forces of ~45 pN (Fig. 3.5 g).

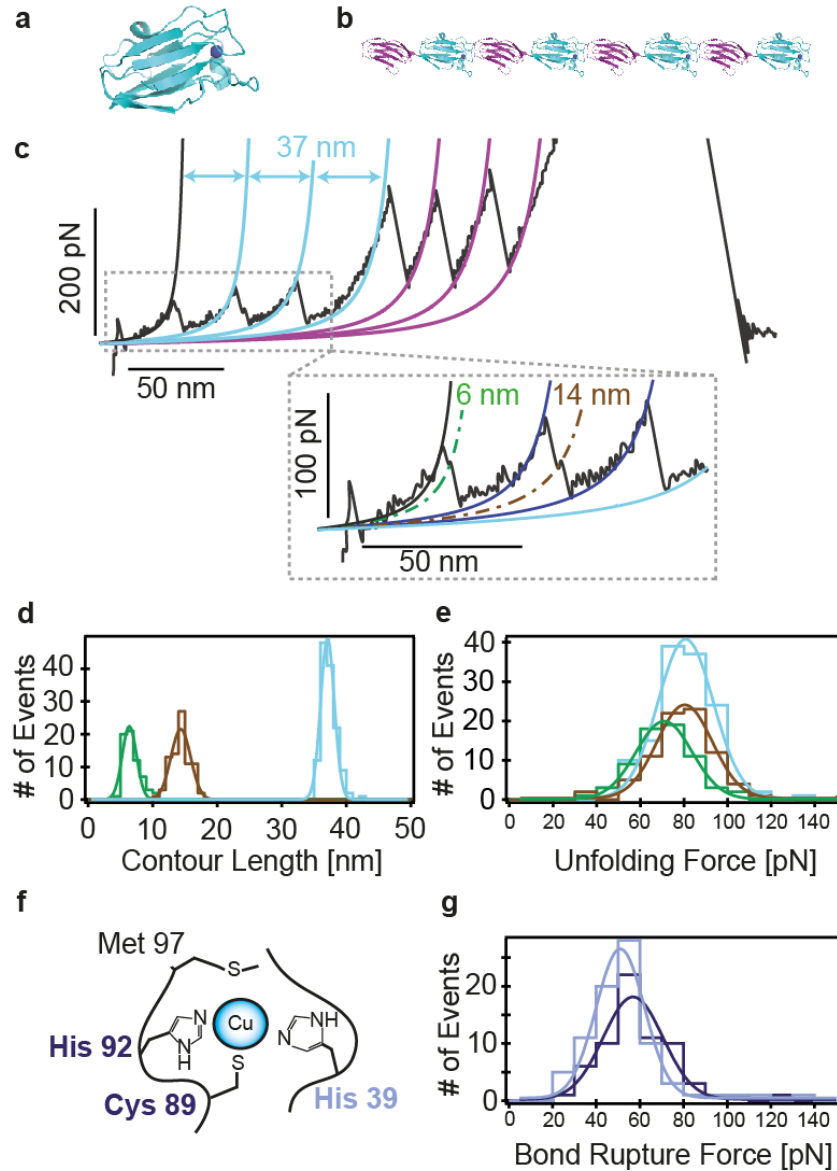


Figure 3.5 Plastocyanin unfolds through parallel pathways fingerprinted by the rupture of individual Cu-S and Cu-N bonds. (a) Plastocyanin structure (PDB: 3BQV). (b) Scheme of the (Plastocyanin-I27)₄ polypeptide. (c) Individual constant velocity unfolding trajectory of a single (Plastocyanin-I27)₄ polypeptide, whereby the unfolding of each plastocyanin domain (cyan WLC fits) occurs prior to the unfolding of the I27 markers (purple WLC fits). *Inset*: Close inspection to each plastocyanin unfolding event reveals that the unfolding process occurs through two distinct mechanical intermediates, which are characterized (d) by an increase in contour length of either $\Delta L = 5.9 \pm 1.1$ nm, $n = 66$ (green) or $\Delta L = 13.8 \pm 1.5$ nm, $n = 82$ (brown). In the cases where no intermediate is observed, suggestive of an apo-plastocyanin form, unfolding triggers the extension of the complete length of the protein, $\Delta L = 36.8 \pm 1.1$ nm, $n = 135$ (cyan). (e) Remarkably, the unfolding forces measured for each individual pathway (N-terminus, C-terminus and two-state unfolding) are very similar (75.3 ± 12 pN, $n = 82$, brown; 65.9 ± 13 pN, $n = 66$, green and 75.8 ± 12 pN, $n = 135$, cyan), respectively. In all cases, protein unfolding occurs at higher forces ~ 76 pN than the rupture of (f) the Cu-N_{His39} bond (46.0 ± 11 pN, $n = 82$, light blue) or the Cu-S_{Cys89} individual bond, occurring at 51.8 ± 13 pN, $n = 66$ (dark blue), as shown in (g).

MOLECULAR DYNAMICS SIMULATIONS

All molecular dynamics simulations were performed by Dr. Guillaume Stirnemann (CRNS – Institut de Biologie Physico-Chimique, Paris)

In order to gain further molecular insight into such enticing unfolding scheme, we conducted molecular dynamics (MD) simulations of azurin and plastocyanin unfolding under a constant force²²⁰. While present force fields are not yet adequate to reliably measure the rupture of organometallic bonds, which are constrained with harmonic potentials during the course of the simulation, the simulations provide a detailed molecular picture of the mechanisms underlying the first stages of protein unfolding, where a set of key hydrogen bonds are disrupted by the stretching force. Individual azurin unfolding trajectories were generated by pulling on one terminal C α (the other being fixed) under a constant force in the range of 250-350 pN (Fig. 3.6 a) that guarantees that unfolding occurs on a timescale accessible to the simulation (~10 ns) as shown in Fig. 3.6 b. In 90% of the cases ($n=19$), the protein unfolded from the C-terminus while in a single trajectory ($n=1$), the protein was found to unfold from the N-terminus instead (Fig. 3.6 c and solid bars in Fig. 3.6d). In another individual trajectory the protein was observed to unfold from both ends simultaneously. When unfolding from the C-terminus, the $\beta 2b$ - $\beta 8$ strands unzip first, followed by the disruption of $\beta 7$ - $\beta 8$. On the contrary, unfolding from the N-terminus encompasses the rupture of the hydrogen bonds between 3 pairs of β -sheets, namely $\beta 2b$ - $\beta 8$, $\beta 1$ - $\beta 3$ and $\beta 3$ - $\beta 6$. In sharp contrast with these results, pulling on the reduced protein (containing free cysteines, without the disulfide bond present) revealed unfolding trajectories whereby the protein mainly unfolded (80%, $n=7$) from the N-terminus (patterned bars in Fig. 3.6 d), in close agreement with the experimental results.

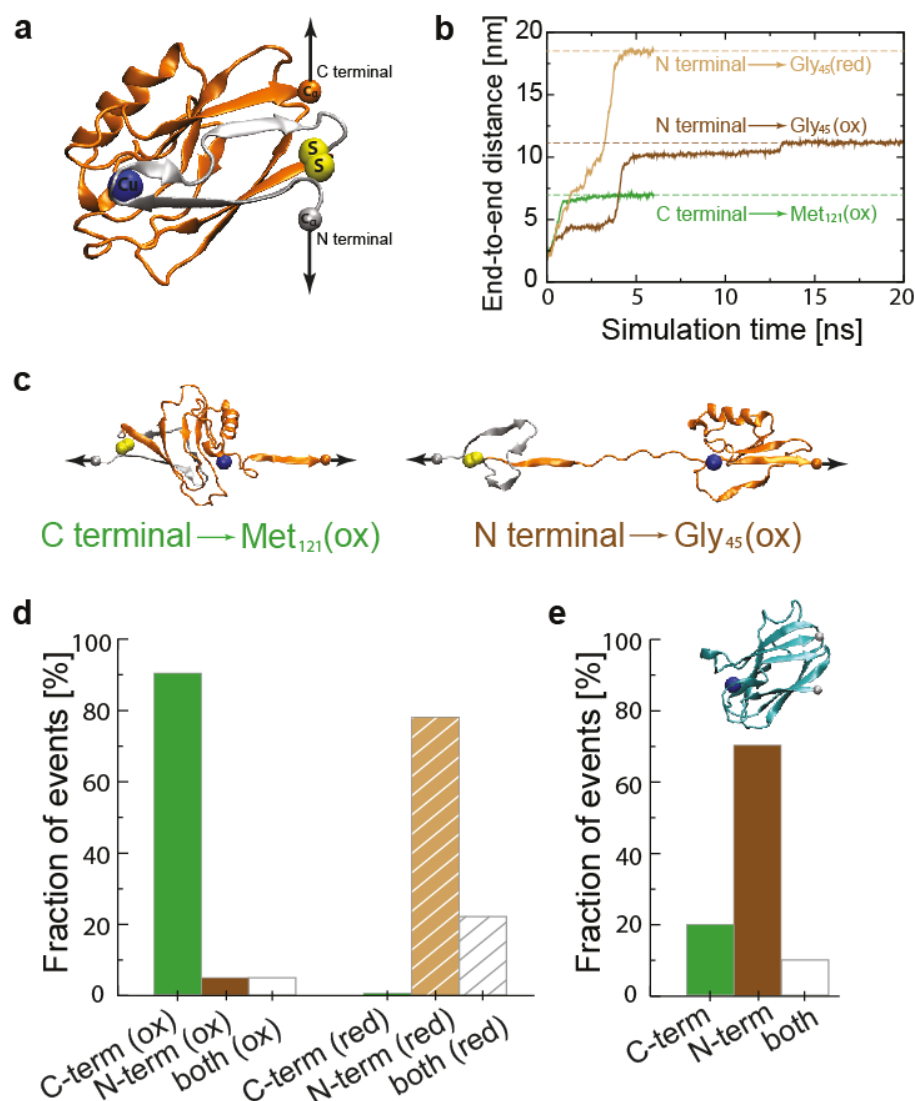


Figure 3.6 MD simulations capture the distinct unfolding pathways. (a) Azurin PDB structure (4AZU). Copper is explicitly shown as a blue sphere, the two sulfur atoms of the disulfide bond Cys3-Cys26 in yellow, the residues trapped after the disulfide bond in white and the rest, in orange. In a steered MD setup, one of the termini is fixed in the laboratory frame and a constant force is applied to the other one. (b) Unfolding trajectories represented by the time evolution of the protein's end-to-end distance for three representative scenarios: wt-Azu unfolding from the N-terminal up to Gly-45 (brown), from the C-terminal up to Met-121 (green) and unfolding of wt-Azu lacking its disulfide bond from the N-terminal side (light brown). Each simulation was stopped once the end-to-end distance had reached the maximum possible extension at this force, the organometallic bonds being intact. (c) Representative structures following unfolding on the C-terminal side and up to Met-121 (left) and unfolding on the N-terminal side up to Gly-45 (right). (d) Percentage of the population of events falling in each scenario (green: C-terminal unfolding; brown: N-terminal; grey: both simultaneously), for both the oxidized form of wt-Azu (solid bars) and the reduced version lacking its disulfide bond (patterned bars). (e) Unfolding trajectories of wt-plastocyanin (3BQV). In ~70% of the trajectories ($n = 7$), plastocyanin unfolds through the N-terminus (brown). In 20% of the occurrences ($n = 2$), the protein unfolds via the C-terminus instead (green), and in the remaining trajectory the protein unfolds from both termini simultaneously (grey). Hence, from a qualitative viewpoint, these *in-silico* results support the experimental observations.

Similarly, pulling on the wild-type plastocyanin protein under a constant force of 250-350 pN revealed that, in 70% of the trajectories ($n=7$), the protein unfolded from the N-terminus, while in 20% of the trajectories ($n=2$), the unfolding started at the C-terminus instead (Fig. 3.6 e). In an individual trajectory ($n=1$), unfolding from both termini occurred concomitantly. Very similar ratios were observed for the case of apo-plastocyanin (80/20/0%, respectively, $n = 10$). Altogether, the *in-silico* unfolding trajectories demonstrate the unusual capability of both azurin and plastocyanin to mechanically unfold from both termini, lending strong support to the experimental data.

3.3 DISCUSSION

Here we employed the mechanochemistry of copper to elucidate, at the single molecule level, the distinct mechanical unfolding pathways of the model T1 copper proteins azurin and plastocyanin. The active site of azurin contains a copper centre in a trigonally distorted tetrahedral geometry with a short Cu-S_{Cys112} bond (~ 2.1 Å) and two typical Cu-N_{His} bonds (~ 1.95 Å) in the equatorial plane²²¹. The axial position is occupied by a methionine at a relatively longer distance Cu-S_{Met121} (2.9 Å)²¹³. Finally, a weak ionic interaction from an axial carbonyl oxygen (Gly-45) is also present. Similarly, in plastocyanin Cu(II) is coordinated by four atoms, the N δ atoms of His-39 (~ 2.2 Å) and His-92 (~ 2.0 Å), the S γ atom of Cys-89 (~ 2.1 Å), and the S δ atom of Met-97 (~ 2.6 Å), forming a distorted tetrahedral geometry²¹⁶. Blue copper proteins owe their name to a strong (~ 2.1 eV)²²⁶ absorption around 600 nm associated with a S_{Cys112}, $p\pi \rightarrow$ Cu ligand to metal charge transfer (LMCT) transition. These particular spectral features are related to the highly covalent nature of the Cu-S bond^{227, 228}, resulting from the strong interaction between the Cu $3d_{x^2-y^2}$ acceptor orbital and the 3S $p\pi$ donor orbital^{213, 229}. The high Cu-S stretching frequency of 400 cm^{-1} measured in Resonance Raman experiments supports the strong Cu-S_{Cys112} bond^{221, 230}. S K-edge X-ray absorption spectroscopy (XAS) results show that $\sim 38\%$ S character is mixed into the Cu $3d_{x^2-y^2}$ orbital in the ground state²²¹. Moreover, the highly covalent copper-thiolate bond imparts a narrow four-line hyperfine splitting in the EPR spectra²¹³. Such strong covalent character notwithstanding, the forces required to break the Cu-S_{Cys112} and the Cu-N_{His46} bonds (Cu-S_{Cys89} and Cu-N_{His39} in plastocyanin) are surprisingly low, of only ~ 45 pN (Appendix I Fig. 8). Interestingly, while the forces to unfold plastocyanin are slightly higher (~ 75 pN) than those required to unfold azurin (~ 55 pN), the mechanical stability of the organocopper bonds in both proteins is ~ 45 pN. These values are in sharp contrast with the high mechanical forces

required to break homolytic covalent Si-Si bonds (2.1 nN) or heterolytic C-Si (2.0 nN) or Au-S (> 2.5nN) bonds^{2, 99, 211}. The low mechanical forces required to disrupt the copper-ligand bonds in azurin and plastocyanin, together with the relatively low mechanical stability measured for the Fe-S (~211 pN)¹⁰⁵ and Zn-S bonds (~170 pN)¹⁰⁶ in rubredoxin, suggest that, from a mechanical perspective, the labile covalent organometallic bonds are significantly distinct from the mechanically resilient non-metallic covalent bonds such as Si-Si, S-S or Si-C bonds⁹⁹, despite their similar dissociation energies. These experimental findings call for computational efforts aiming at developing novel force fields that couple organometallic bonds with mechanical strain to help elucidation of the quantum-mechanical origin of the force-induced organometallic bond rupture.

The precise novel characterization of the nanomechanical properties of the copper-based organometallic bonds allowed us to uncover the complexity in the azurin and plastocyanin unfolding pathways when placed under mechanical force. In the case of azurin, we attribute the rupture of the unfolding intermediate to the breaking of the Cu-S_{Cys112} and Cu-N_{His46} bonds, however, no obvious intermediate was detected that was compatible with the rupture of the Cu-N_{His117} bond, which has been described using chemical denaturation techniques to be still formed even after azurin unfolding²³¹. It is possible that due to the short distance (5 residues) within Cys-112 and His-117 within the structure, the length resolution of our instrumentation makes them indistinguishable, being all included in the data defining 6 nm intermediate. However, the observation that for the (Azu_{C112A-I27})₄ mutant the population of intermediate events at ~6 nm (or lower) was vanishingly small renders this option unlikely. Hence, it is plausible that upon unfolding from the C-terminus, the disruption of the β 2b- β 8 sheets induces a conformational change

in the protein loop connecting $\beta 7$ - $\beta 8$ strands that concomitantly distorts the active site structure thus perturbing the Cu-N_{His117} bond.

A remarkable finding in our experiments was the stochastic rupture of the main unfolding barrier in both studied blue copper proteins, azurin and plastocyanin. Due to the particular β -barrel topology of the cupredoxin fold, two independent mechanical transition states, each presumably placed close to both of the termini, govern the mechanical stability of the proteins. Fortuitously, the height of the energy barrier to unfold both termini in each protein was found to be almost equivalent when a pulling velocity of 400 nm s⁻¹ was employed, requiring a relatively low force of ~55 pN to unfold both termini in azurin (Appendix I, Fig. 8), and ~75 pN to unfold both C- and N-termini in the case of plastocyanin (Fig. 3.5 e). In fact, MD simulations on azurin clearly revealed that for both transition states, the rupture of the hydrogen bonds between the involved β -sheets occurs through an unzipping mechanism because the force is applied parallel to the hydrogen bonds axes (Fig. 3.6 a).

These results put forward a rather unique global mechanical unfolding scenario for wt-azurin and wt-plastocyanin (Fig. 3.7). In 20% of the cases (40% in the case of plastocyanin), azurin unfolds in a single two-state event. In this case, we cannot pinpoint which mechanical transition state, either the N- or C-termini, breaks first. It is plausible that in some of these events, the resolution of our instrumentation was not able to detect the fast mechanical intermediates. However, by pulling in force clamp at forces as low as 20 pN (Appendix 1 Fig. 9), and hence readily expanding the time-span of each trajectory, we could still capture both populations of events; those exhibiting a clear mechanical intermediate and those events where the unfolding was two-state, lacking any measurable intermediate.. We hence hypothesise that the observed two-state unfolding trajectories correspond to the unfolding of

individual apo-azurin (or apo-plastocyanin) modules. In 80% of the occurrences, azurin unfolds through a complex pathway involving at least one mechanical intermediate. Of these, in ~45% of the trajectories mechanical unfolding starts from the C-terminus (green pathway), whereas in ~45% of the cases, unfolding proceeds instead through the N-terminus (brown pathway). Finally, in the remainder ~10% of the occurrences, the protein unfolds first from one of the termini, then from the other. On a few occasions ($n=16$) a unique 15 nm step corresponding to the concomitant unfolding of both termini is observed. Similarly, wt-plastocyanin follows a similar global unfolding scheme, with slightly different percentages corresponding to each unfolding pathway.

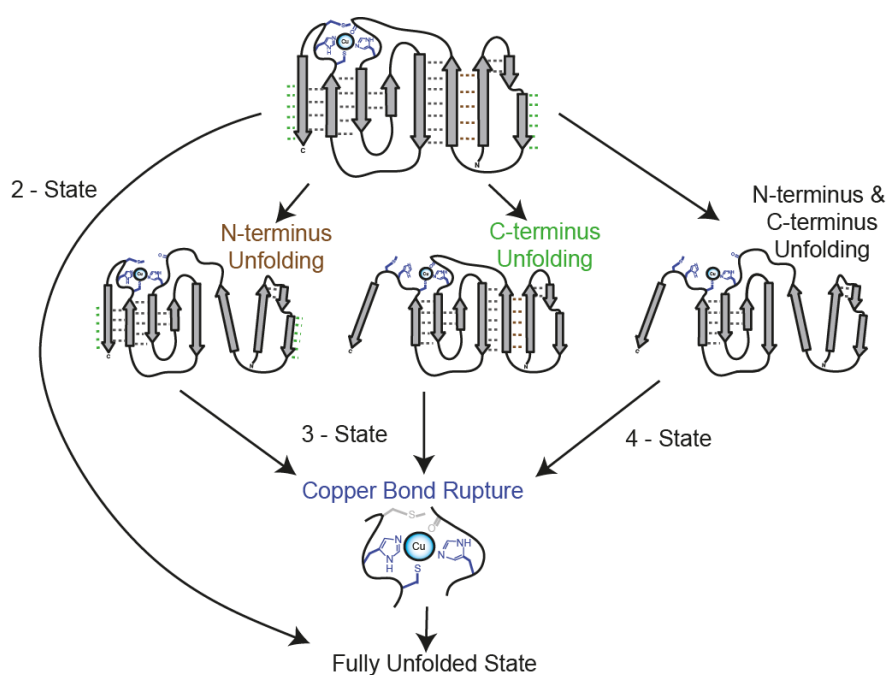


Figure 3.7 A generic kinetic partition scheme for the complex mechanical unfolding of wt-azurin and wt-plastocyanin. Upon the application of mechanical force, azurin can unfold following multiple pathways. In 20% of the occurrences (40% in the case of plastocyanin), the protein unfolds according to an all-or-none process. By contrast, in 75% of the cases (50 % for plastocyanin), azurin unfolds via one mechanically stable intermediate after unfolding from either the N-terminus (brown) or the C-terminus (green). In the remainder 5% (10%) of the cases, a 4-state unfolding pathway involving the simultaneous unfolding of the N- and C- termini is observed.

The observed similar probability of unfolding occurring from either the C- or the N- termini in wt-azurin is severely biased when the native disulfide bond

is absent (Fig. 3.3 e). In this case, unfolding mostly occurs (~80%) from the N-terminus. MD simulations further confirm the experimental results by showing that, in 80% of the cases, the reduced wt-Azu lacking its disulfide bond (equivalent to the C26A mutant) unfolds indeed by breaking the N-terminus β -sheets. Crucially, in our experiments we observed that wt-azurin can unfold from the N- or C- termini with almost equal probability. By contrast, our MD simulations show a drastic change in the pathway preference since 90% of the trajectories unfold from the C-terminus. It is likely that such discrepancy can be explained by the slightly different distance to the transition state, Δx , associated with the unfolding from both C- and N-termini, giving rise to distinct force dependencies of unfolding. Since the force range probed in the *in-vitro* and *in-silico* experiments is significantly different, it is plausible that both approaches capture different relative unfolding probabilities at the distinct sampled pulling forces.

Altogether, the whole unfolding cascade is reminiscent of a complex kinetic partition scheme. In the recently reported examples for T4 Lysozyme⁸⁶, Maltose Binding protein⁸⁸ and Leucin Binding Protein²⁰⁸ the main unfolding event was unique and always present, while mechanical intermediates further along the unfolding pathway were found to be stochastic and occurring in different positions. In the case of GFP, a bifurcation in the unfolding pathway was observed after the detachment of an N-terminal α -helix, fingerprinted by internal engineered disulfide bonds²⁰⁹ or circular permutations²³². By contrast, azurin and plastocyanin exhibit two distinct mechanical clamps placed very close to each terminus that exhibit a similar stability. Hence, upon application of force, the protein stochastically crosses one of the two main barriers to unfolding. The directionality of the entire unfolding pathway is therefore determined by this early random event. Once the barrier is crossed, the protein unravels and gets held “on pathway” at well-defined positions around the

copper site. Taken together, the new unfolding scheme depicted for both blue copper proteins follows a complex kinetic partition scheme in which the stochastic nature of the process is highlighted at the very beginning of the unfolding process with the rupture of the hydrogen bonds determining the main barrier for unfolding.

Our experiments have allowed us to pinpoint the unfolding directionality in the studied blue copper proteins, occurring either from the C- or N-termini independently. The choice in the terminus that initiates unfolding might be of biological relevance, as demonstrated by recent nanopore-based experiments, which have successfully recorded the distinct co-translocation unfolding dynamics depending on the terminus that is first presented to the nanopore mouth²³³. Similarly, during degradation in the proteasome, proteins are mechanically unfolded before entering the narrow proteolytic chamber^{234, 235}. The direction of entry in both processes depends on the force needed to unfold its two termini²³⁶. When the N- and C- termini are mechanically non-equivalent, the protein will presumably tend to search for the weak link to initiate unfolding at the pore mouth²³⁷. In this regard, proteins with dual mechanical transition states might provide an interesting case study for these applications, especially if one of the N- or C- termini can be tuned at will to act as the protein's Achilles' heel⁷³.

Altogether, our experiments demonstrate the power of single molecule techniques at revealing complex unfolding pathways in topologically rich proteins such as the case of metalloproteins, of high relevance in nature.

4. PROTEIN S-SULFENYLATION IS A FLEETING MOLECULAR SWITCH THAT REGULATES NON-ENZYMATIC OXIDATIVE FOLDING

The post-translational modification S-sulfenylation functions as a key sensor of oxidative stress. Yet the dynamics of sulfenic acid in proteins remains largely elusive due to its fleeting nature. Here we use single-molecule force-clamp spectroscopy and mass spectrometry to directly capture the reactivity of an individual sulfenic acid embedded within the core of a single Ig domain of the titin protein. Our results demonstrate that sulfenic acid is a crucial short-lived intermediate that dictates the protein's fate in a conformation-dependent manner. When exposed to the solution, sulfenic acid rapidly undergoes further chemical modification, leading to irreversible protein misfolding; when cryptic in the protein's microenvironment, it readily condenses with a neighbouring thiol to create a protective disulfide bond, which assists the functional folding of the protein. This mechanism for non-enzymatic oxidative folding provides a plausible explanation for redox-modulated stiffness of proteins that are physiologically exposed to mechanical forces, such as cardiac titin.

*This chapter is based on the article published 22 August 2016 in Nature
Communications (Attached in appendix V)*

4.1 INTRODUCTION

Protein folding defines a molecular self-assembly process that is finely tuned by a subtle interplay between enthalpic and entropic contributions^{238, 239, 240}. An extra layer of complexity is added when a native covalent disulfide bond needs to reform along the folding pathway^{241, 242}. Other chemical alterations, such as post-translational modifications, pose further challenges to the successful completion of the folding process.

Cysteine is a principal target for post-translational modifications, especially those related to oxidative stress²⁴³. Due to the presence of empty *d*-orbitals ready for bonding, the highly polarizable sulfur atom can display oxidation states ranging from -2 to +6^{244, 245}. Cysteine sulfenic acid (Cys-SOH) is the first oxidation step of cysteinyl thiols (Cys-SH)²⁴⁶ and the resulting post-translational modification, cysteine S-sulfenylation, has been identified as a redox sensor in an increasing number of proteins^{247, 248}. However, the unusually high reactivity of sulfenic acid—a consequence of its dual electrophilic and nucleophilic character^{244, 247, 249}—, together with its invisibility to UV-vis and fluorescent spectroscopic detection²⁵⁰, has rendered its characterization challenging. In general, stabilization of protein sulfenic acid is highly dependent on the protein microenvironment²⁴⁸, following the general notion that solvent-exposed cysteines feature increased chemical lability²⁵¹.

Identification of protein sulfenic acids has been mostly indirect, through chemical labelling with 5,5-dimethyl-1,3-cyclohexadione (dimedone)²⁵², a mass-shifting molecular probe that specifically traps –SOH moieties and is easily identifiable with mass spectrometry (MS)^{253, 254, 255}. Despite this phenomenal progress²⁴⁹, given the irreversible nature of the detection methods, which interfere with the intrinsically fast reactivity of Cys-SOH,

direct identification of the dynamics of a particular protein sulfenic acid and its implication on folding remains largely elusive.

Under oxidative conditions, Cys-SOH can be readily oxidized to more stable products, namely sulfinic (Cys-SO₂H) or sulfonic (Cys-SO₃H) acids²⁴³. The transition to these higher oxidation state species has been generally regarded as irreversible, and a marker for oxidative damage. Such deleterious pathway can be avoided if the electrophilic sulfur atom in sulfenic acid reacts with an intramolecular neighbouring protein thiol nucleophile to form a stable disulfide bond^{256, 257}. Such oxidant-mediated disulfide bond formation process, typically occurring within the reducing conditions of the cytosol upon a transient increase in oxidative stress, differs from the enzyme-catalysed formation of structural disulfide bonds. This latter process takes place in the oxidising milieu of the endoplasmic reticulum (ER) or the mitochondrial inter membrane space (IMS) in eukaryotes or in the periplasmic space in bacteria^{256, 258} via the formation of a mixed-disulfide intermediate between the catalytic enzyme and the protein substrate^{259, 260}. Despite its *in-vivo* relevance, a comprehensive molecular description of the formation of disulfide bonds within the context of non-enzymatic oxidative folding is still missing.

Here we use a combination of single molecule force-clamp spectroscopy, protein engineering techniques and mass spectrometry (MS) to directly monitor the reactivity of an individual force-induced sulfenic acid moiety occurring within a single Ig domain of the giant titin protein. Our results demonstrate that the fleeting sulfenic acid intermediate, the life-time of which is largely conformation-dependent, acts as a molecular switch that directly modulates protein function. In a short window of opportunity in the second timescale, the -SOH moiety is able to avoid irreversible protein misfolding—induced by cysteine hyperoxidation and/or aldehyde condensation—by readily forming a disulfide bond that guarantees the functional final folding

of the protein. We hypothesize that this mechanism for non-enzymatic oxidative folding may be commonplace for oxidation-induced post-translational modifications occurring on solvent-exposed cysteines, drastically affecting the elastic properties of proteins that are exposed to mechanical forces.

4.2 RESULTS

SULFENIC ACID TRIGGERS DISULFIDE BOND FORMATION

Sulfenic acid is typically induced by exposing a cysteine residue to high concentrations of hydrogen peroxide (H_2O_2)²⁶¹. However, its fast kinetics ($10\text{--}10^7 \text{ M}^{-1} \text{ s}^{-1}$)^{262, 263} and its further reactivity to higher oxidation state species often precludes the capture of this transient sulfenic acid intermediate. An alternative means to create a protein sulfenic acid in a much more controlled manner is based on the alkaline hydrolysis of an individual disulfide bond, resulting in the stoichiometric formation of a sulfenic acid and a thiolate^{264, 265, 266} (Fig. 4.1 a). Albeit admittedly less physiologically relevant, this strategy is also found in nature^{267, 268}, as reported for phosphoglycerate kinase (PGK) reductase^{269, 270}.

Using an experimental single molecule mechanical assay designed previously²¹¹, an individual sulfenic acid moiety can be readily created by the hydrolysis of a single protein disulfide bond under high pH conditions (Fig. 4.1 a). The atomistic detail of this process has been provided by highly revealing molecular quantum mechanics/molecular mechanics (QM/MM) simulations^{172, 271}. We now set out to extend our experimental approach to probe whether each cleaved disulfide bond can be reformed in the absence of a catalytic enzyme, and how such putative reformation process affects the entire protein folding phenomena.

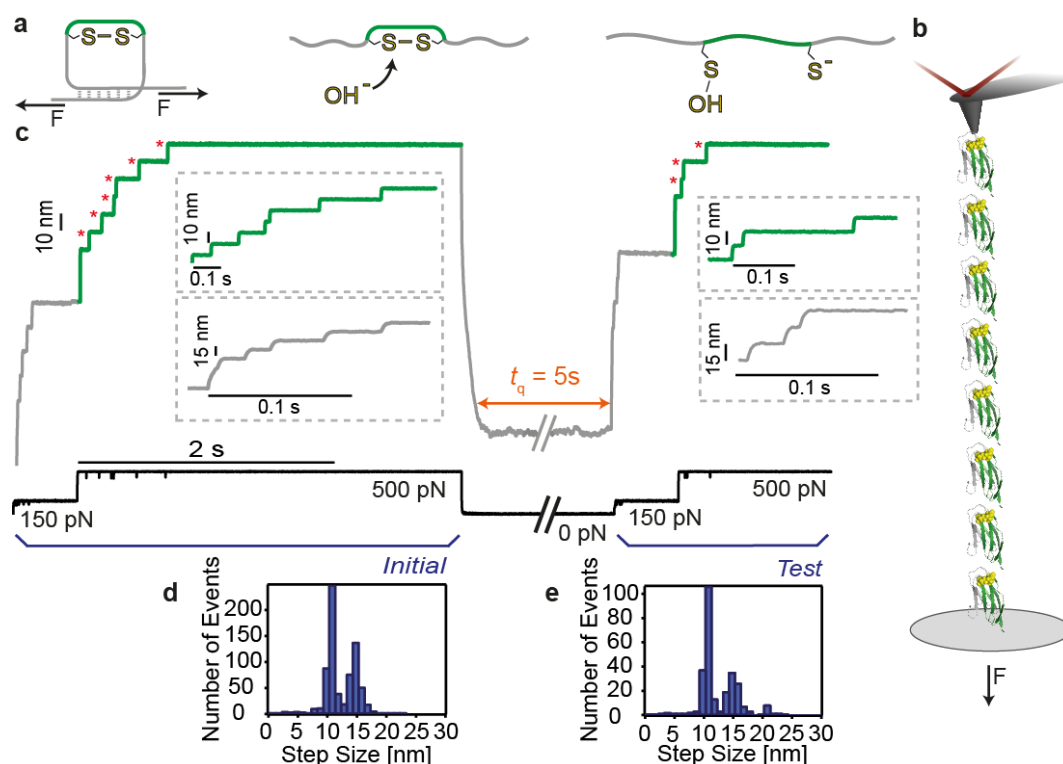


Figure 4.1 Single molecule force-clamp spectroscopy captures the non-enzymatic reformation of individual disulfide bonds, which drive protein folding. (a) The reduction of the protein disulfide bond via a S_N2 attack by hydroxide anions results in the creation of a sulfenic acid and thiolate. (b) Schematics of the engineered $(I27^{E24C-K55C})_8$ polyprotein being stretched in the AFM set-up. (c) Pulling on an individual polyprotein at constant force using a specifically designed pulse-protocol captures oxidative protein folding. A first *initial* pulse $F = 150$ pN unfolds and extends all the modules of the $(I27^{E24C-K55C})_8$ polyprotein to the mechanical clamp created by the disulfide bonds, which exposes them to the solution (grey line). Each unfolding event is marked by a step increase of ~ 15 nm (inset, grey line). A second stronger force pulse ($F = 500$ pN) triggers disulfide reduction (green line), fingerprinted by equally spaced ~ 10 nm steps (inset, green line) that correspond to the extension of the peptide chain released on the reduction of each disulfide bond (red asterisks). Releasing the pulling force ($F = 0$ pN) triggers the protein to collapse and fold for a time quench $t_q = 5$ s. The subsequent *test* pulse probes the folding status of the protein. Mirroring the *initial* pulse, a force pulse of $F = 150$ pN unfolds and extends the refolded protein up to the newly created disulfide bond (inset, grey line) whereas the last $F = 500$ pN pulse is able to reduce it again (inset, green line). The step-size histograms corresponding to the (d) *initial* and (e) *test* pulse highlight in both cases the presence of ~ 10 nm and ~ 15 nm steps.

In our experiments, we used a single molecule force-clamp atomic force (AFM) spectrometer (Fig. 4.1 b) to stretch under high alkali conditions (pH=12.8) a polyprotein containing eight identical domains of a mutated titin Ig27 domain containing two cysteines in positions 24 and 55, which readily form a structurally buried disulfide bond (Fig. 4.1 a). We designed a five-pulse force protocol to measure the protein elongation over time, which enables independent capture of each individual cleavage and reformation event, with single bond resolution (Fig. 4.1 c). Applying an initial pulse of 150 pN for 0.5

seconds to the (I27_{E24C-K55C})₈ polyprotein resulted first in the presence of fast elongation steps of ~15 nm (inset, grey), corresponding to the unfolding and extension of each individual protein domain up to the rigid disulfide bond, which became exposed to the solvent environment (Fig. 4.1 c, grey). A second force pulse at a much higher force (500 pN) triggered the forced-induced scission of each individual disulfide bond via a S_N2 nucleophilic attack by the hydroxyl (OH⁻) anion. Each individual hydrolysis event is fingerprinted by a step of ~10 nm (inset, green), corresponding to the extension of the amino acids that were trapped behind the disulfide bridge (Fig. 4.1 c, green). The two distinctive step size populations in these first two high force pulses (*initial* pulse) are clearly observed in the bimodal histogram shown in Fig. 4.1 d. Notably, after the cleavage of all disulfide bonds at 500 pN, the fully unfolded and extended polyprotein is extended to almost its contour length and all of its backbone residues are fully exposed to the chemical environment of the measuring solution.

Shortly after the last individual rupture event, the stretching force was completely withdrawn for $t_q = 5$ seconds in order to trigger the folding of the polypeptide. Such force-quench approach allows dissection of the distinct phases involving mechanical folding from highly extended states^{118, 121, 188, 272}. After the quenching pulse, a *test* pulse —composed again of a stretching sequence of 0.5 s at 150 pN followed by a pulse at a higher force of 500 pN— was applied to the protein to probe its folding status.

The elongation pattern of the protein during the *test* pulse enables discrimination amongst three different possible oxidative folding scenarios; (i) if the protein has not been able to fold within the quench time, it will extend featuring the absence of well-defined steps^{121, 272}. Instead, (ii) if the protein is able to refold —thus regaining mechanical stability— but the disulfide bond has not been reformed, the protein will extend featuring steps of 25 nm,

corresponding to the unfolding and stretching of the fully folded protein in the absence of the mechanically rigid, covalent disulfide bond¹¹⁸. Finally, (iii) if the protein has been able to both successfully refold and reform the disulfide bond, the *test* pulse will be then composed of the unfolding ~15 nm steps followed by the ~10 nm steps corresponding to the (re)-rupture of each successfully reformed disulfide bond. In summary, a 25 nm step provides the signature of a properly folded domain with no disulfide, whereas a 15 nm step (followed by a 10 nm step) indicates a properly folded domain with a formed disulfide.

The latter scenario was indeed observed in our folding trajectories (Fig. 4.1 c). The histogram corresponding to the distribution of step sizes populating the *test* pulse hallmarks the presence of both ~15 nm and ~10 nm steps (Fig. 4.1 e). Surprisingly, not a single event corresponding to the ~25 nm steps was observed. Interestingly, a small population (~4%) of ~21 nm events was also present in the high force pulse, corresponding to the scission of a non-native disulfide bond formed between two contiguous domains within the polyprotein chain¹⁸⁸ (Fig. 4.2). As a control experiment, we verified that the oxidised protein was able to independently refold under high pH conditions, as shown in Fig. 4.3. Altogether, our experiments qualitatively show that an individual disulfide bond is able to reform in the absence of catalytic enzymes.

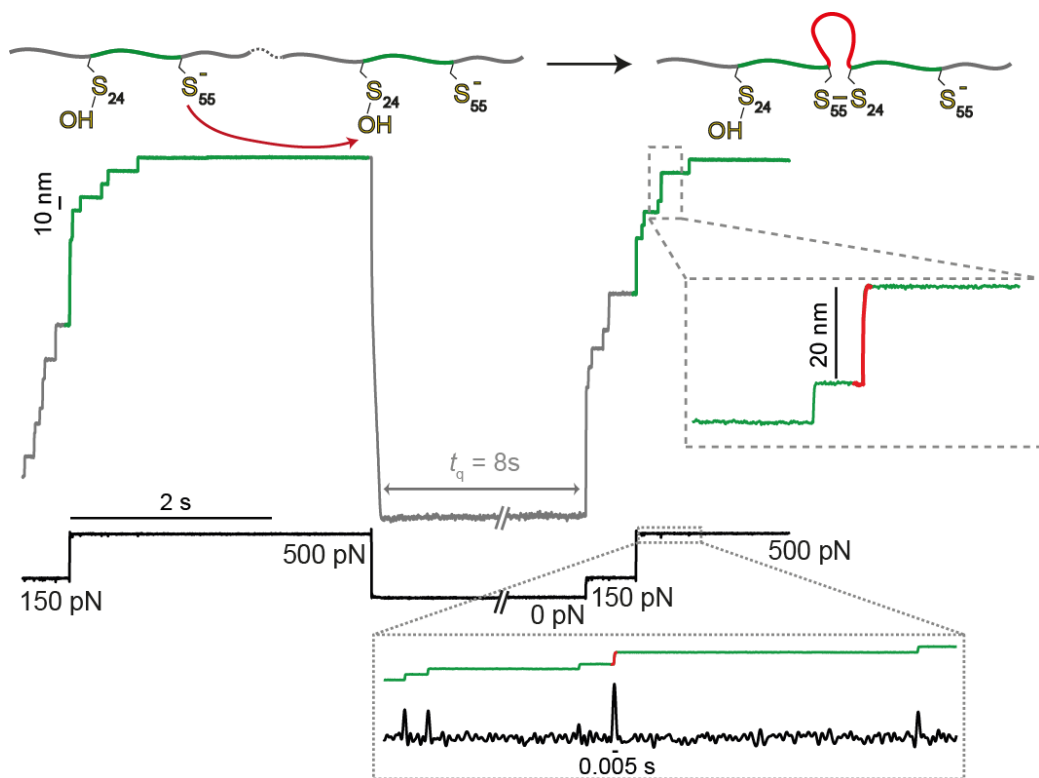


Figure 4.2 Capturing misfolding events triggered by the formation of non-native disulfide bonds between contiguous Ig domains within the polyprotein chain. The low-probability events (~4%) observed in the step size histogram corresponding to the *test* pulse (Fig. 4.1 e) can be explained by the formation of a disulfide bond between two cysteines belonging to two independent, neighbouring domains of the (I27_{E24C-K55C})₈ polyprotein. The individual oxidative folding trajectory fingerprints the presence of such new, ill-formed disulfide bond by the presence of a 20 nm step (red) in the *test* pulse occurring at 500 pN. This step size corresponds to the extension of the 58 amino acids trapped behind a disulfide bond formed between cysteine 55 and cysteine 24 of two neighbouring Ig domains. *Inset:* Magnification of the force channel demonstrating that the feedback response time of the instrument (upward spikes) is less than 5 ms.

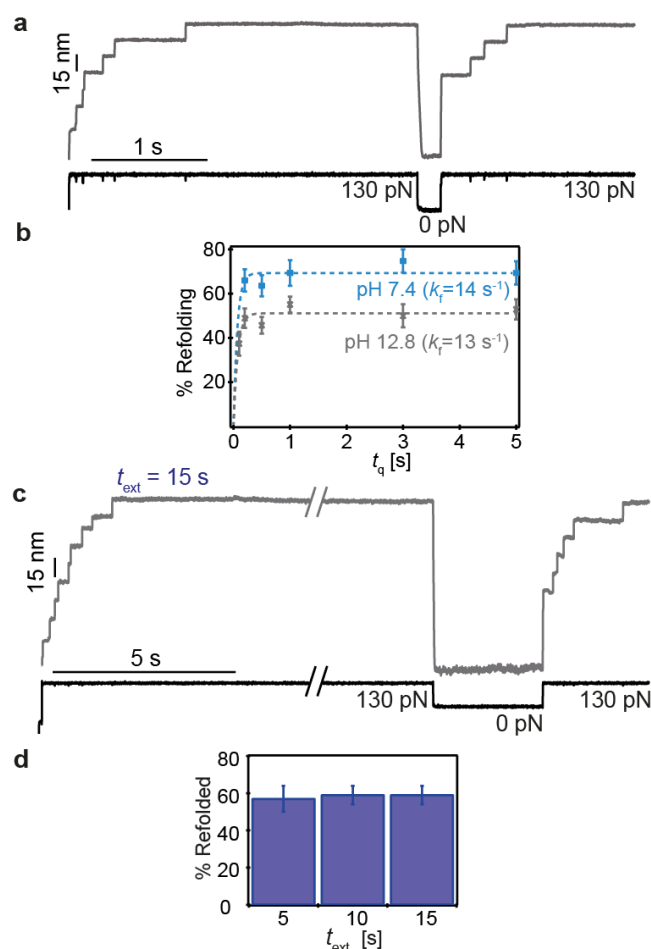


Figure 4.3 The (I27^{E24C-K55C})₈ polyprotein is able to successfully refold at pH = 12.8 when the disulfide bond remains oxidised. (a) Using a force quench protocol that lacks the high-force pulses employed to trigger disulfide bond reduction, we monitored the refolding of the (I27^{E24C-K55C})₈ polyprotein when it is extended up to the rigid disulfide bond. A first pulse of 130 pN induces the unfolding of the protein up to the disulfide bond, marked by a step-wise increase in length of 15 nm. After 3 seconds, the force is removed for t_q in order to trigger mechanical folding. A *test* pulse back at 130 pN exhibits the re-unfolding of the protein in steps of 15 nm. The recovery of the mechanical stability of the previously unfolded modules unambiguously certifies that the protein has successfully refolded under high pH = 12.8 conditions. (b) Measuring the percentage of successfully refolded modules during the test pulse compared to those occurring in the initial unfolding pulse for distinct quench times (t_q) measures the time-course evolution of mechanical folding. Fitting a single exponential fit to the data yields a folding rate $k_f = 13 \text{ s}^{-1}$ (grey discontinuous line), which closely resembles the folding rate observed in analogous experiments conducted at neutral pH = 7.4, exhibiting a folding rate of $k_f = 14 \text{ s}^{-1}$ (blue discontinuous line). While the experiments conducted at high pH exhibit a slightly lower (19%) efficiency of refolding, probably due to an alkali induced modification of the side chain of any of the amino acids, these results directly demonstrate that (I27^{E24C-K55C})₈ polyprotein is able to successfully refold when the disulfide bond has not been reduced under high pH = 12.8 conditions. (c) Crucially, the time the protein is left unfolded and extended at high force, t_{ext} , does not have any negative impact on the folding efficiency, (d), which is kept constant at ~ 58%.

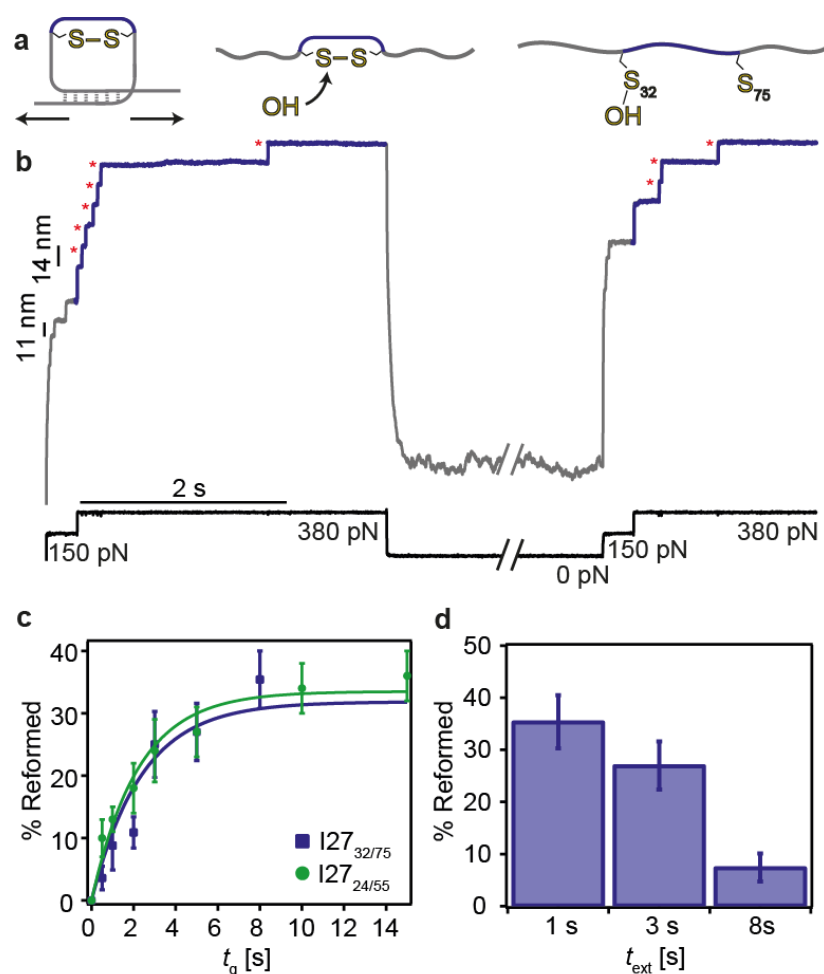


Figure 4.4 The sulfenic acid–induced reformation of a disulfide bond is independent on the position of the disulfide bond within the protein structure. (a) We repeated the non-enzymatic oxidative folding experiments on the (I27_{E24C-K55C})₈ polyprotein described in Fig. 4.1 in the main text with a similar construct whereby the disulfide bond position is changed, occurring between cysteines 32-75. (b) Stretching the resulting (I27_{G32C-A75C})₈ polyprotein construct at 150 pN first results in the unfolding of the protein up to the position of the disulfide bond, eliciting steps of 11 nm (grey). A second pulse at a higher force of 380 pN triggers the hydroxyl-induced cleavage of the individual disulfide bonds (red asterisks), fingerprinted by a length increase of 14 nm (purple). (c) The percentage of reformed disulfide bonds increases exponentially with the t_q with an associated rate constant of $k_{ref} = 0.42 \text{ s}^{-1}$ (purple), very similar to that measured for (I27_{E24C-K55C})₈ with $k_{ref} = 0.45 \text{ s}^{-1}$ (green). (d) Increasing the exposure time, t_{ext} , of the sulfenic acid to the solution, while maintaining $t_q = 5 \text{ s}$, decreases the percentage of reformed disulfide bonds, similar to the behaviour observed for (I27_{E24C-K55C})₈. These experiments demonstrate that the position of the disulfide bond within the protein structure does not significantly affect the sulfenic acid-mediated mechanism of oxidative folding.

We next examined whether such oxidative folding scenario could also be reproduced when the specific location of the disulfide bond within the structure of the I27 protein was changed. We repeated the experiments reported in Fig. 4.1 on a distinct polyprotein construct, (I27_{G32C-A75C})₈, in which the cysteine mutations were engineered in different positions of the I27

sequence. The resulting trajectories (Fig. 4.4) qualitatively certified the successful reformation of the disulfide bond along the folding pathway.

THE KINETICS OF DISULFIDE BOND REFORMATION

To quantify the kinetics of disulfide bond reformation, we repeated the experiments shown in Fig. 4.1 and varied the quench time (t_q) within the range spanning $t_q = 0.5$ -15 s (Fig. 4.5). As observed in the individual trajectories shown in Fig. 4.5 a-b, the longer the protein was allowed to refold in the absence of force, the higher the percentage of disulfide bonds that reformed (hallmarked by the appearance of ~ 10 nm steps in the *test* pulse). As shown in Fig. 4.5 c, the % of disulfide bond reformation increased exponentially with the quench duration, suggesting that the reformation of the protein disulfide bond is a two-state process with an associated characteristic time of $k_{\text{ref}} = 0.45 \text{ s}^{-1}$. This is consistent with a scenario where the nucleophilic thiol would attack the electrophilic $-\text{SOH}$ centre via a concerted $\text{S}_{\text{N}}2$ reaction, occurring through a well-defined transition state^{273, 274}, to reform the disulfide bond. Noteworthy, the measured reformation rate is expected to be pH-dependent, since the nucleophilic capacity of the free thiolate to attack the electrophilic sulfenic acid will be largely dependent on the thiolate being completely deprotonated. However, given the expected $\text{pK}_{\text{a}} \sim 8$ for each free cysteine, we do not anticipate these rates to be significantly affected in our high-pH experiments, although this might have an effect at lower pH values. Remarkably, the measured rate of disulfide reformation for the (I27_{G32C-A75C})₈ construct ($k_{\text{ref}} = 0.42 \text{ s}^{-1}$) was found to be almost identical to that measured for the (I27_{E24C-K55C})₈ polyprotein, thus demonstrating that the disulfide bond reformation kinetics is not significantly altered by changing the position of the disulfide bond within the protein structure (Fig. 4.4).

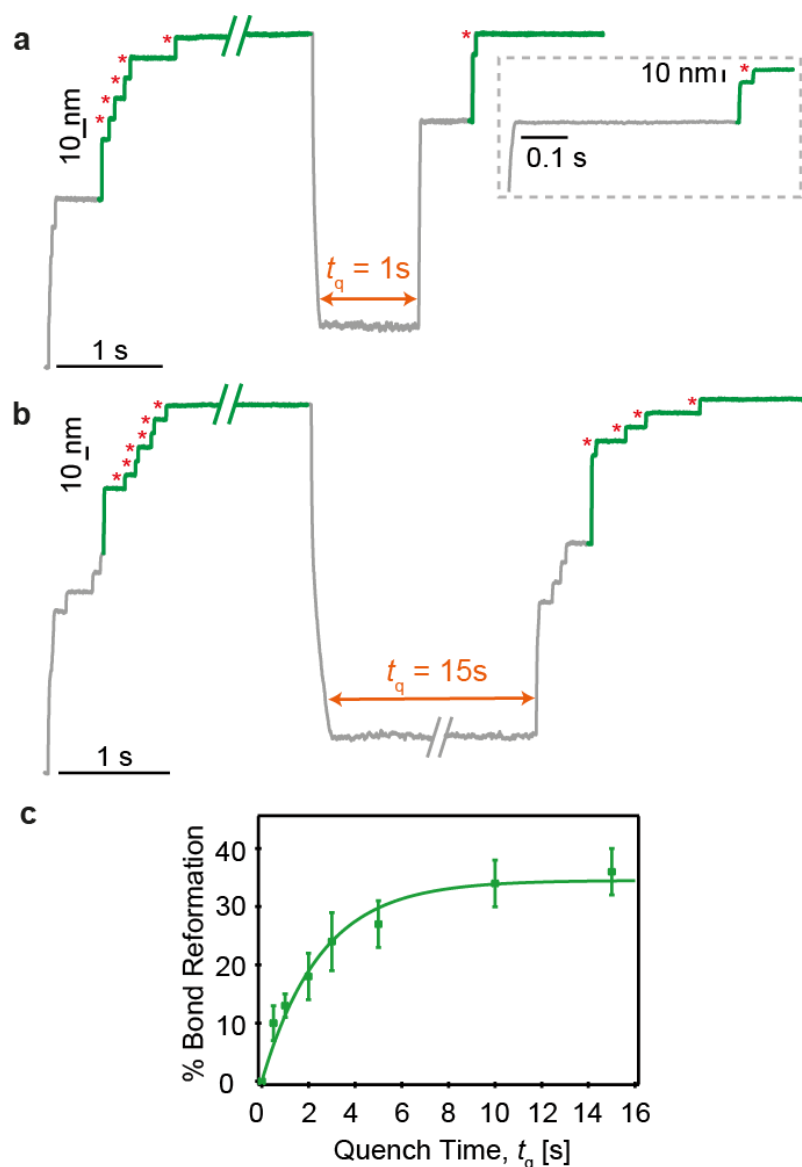


Figure 4.5 Kinetics of disulfide bond reformation. (a, b) The percentage of disulfide reformation, fingerprinted by the presence of ~ 10 nm steps (red asterisks), increases with t_q . Figure 4.5 a illustrates a particular trajectory showing that the reformation of the disulfide bond (marked by the 10 nm step, red arrow) occurred in the absence of protein refolding. (c) The percentage of disulfide bond reformation increases exponentially with t_q with an associated rate constant of $k_{ox} = 0.45$ s $^{-1}$ (solid green line, $\chi^2_{red} = 0.513$).

THE REACTIVITY OF SULFENIC ACID DEPENDS ON SOLVENT ACCESSIBILITY

While generally short, the life-time of -SOH has been described to crucially depend on the degree of solvent exposure^{257, 275}. Force-clamp AFM allows to explore new protein conformations that are not usually sampled in classical ensemble experiments conducted in the bulk²⁷⁶. In particular, the force-

triggered low-entropy, highly extended unfolded state fully exposes the backbone residues to the solvent environment²⁷⁷. To interrogate the conformation-dependent reactivity of sulfenic acid, we systematically changed the time the fully reduced and mechanically stretched protein was exposed to the solvent, t_{ext} , thereby increasing the probability to be chemically modified, while keeping the quench time constant ($t_{\text{q}} = 8$ s, Fig. 4.6 a,b). Leaving the protein stretched for $t_{\text{ext}} = 0.5$ s resulted in a ~55% disulfide bond reformation. Substantially increasing the extended time up to $t_{\text{ext}} = 30$ s considerably decreased the reformation probability down to ~5%. The absence of disulfide bond reformation inevitably resulted in protein misfolding, since we never observed the presence of a 15 nm step without the companion 10 nm step in the *test* pulse. The evolution of the % of disulfide bond reformation as a function of t_{ext} showed a dramatic, fast decrease (Fig.4.6 c). While fitting a single exponential to the data qualitatively described the time-course of the process with an associated decaying rate of $k = 0.17$ s⁻¹ ($\chi^2_{\text{red}} = 5.02$), a double exponential ($k_1 = 0.065$ s⁻¹ and $k_2 = 0.695$ s⁻¹) slightly increased the goodness of the fit ($\chi^2_{\text{red}} = 2.78$, Appendix II Fig. 1), reminiscent of two independent reaction pathways occurring concomitantly.

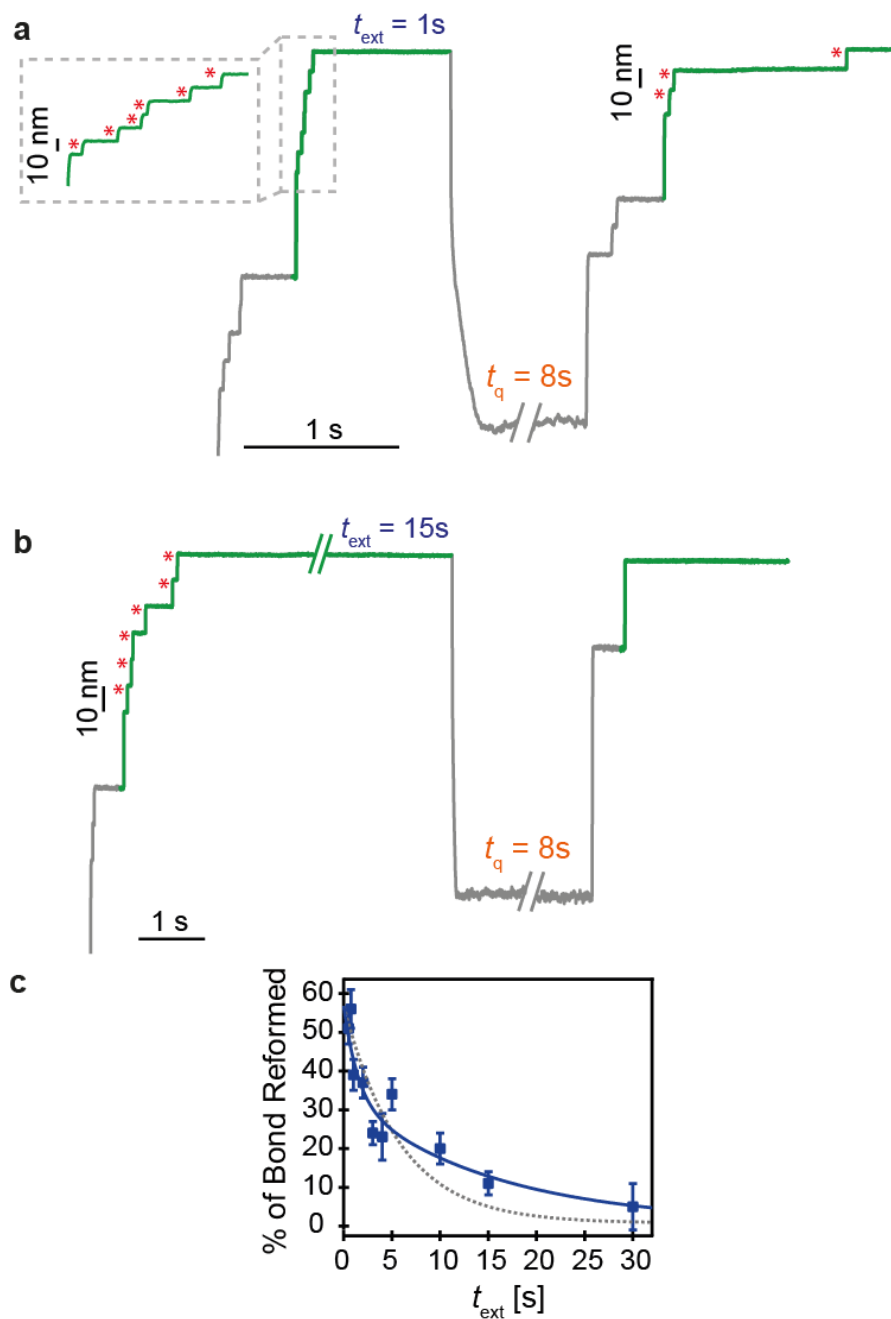


Figure 4.6 The evolution of sulfenic acid inhibits oxidative folding. (a, b) Increasing the time, t_{ext} , the protein is reduced and stretched at high force ($F = 500$ pN) results in a drastic decrease of successful oxidative folding (marked by the presence of both ~ 10 nm, red asterisks, and ~ 15 nm in the *test* pulse). (c) The decrease in the % of disulfide bond reformation as a function of t_{ext} can be best described by a double exponential fit (solid blue line), with associated rate constants $k_1 = 0.065$ s $^{-1}$ and $k_2 = 0.695$ s $^{-1}$, reminiscent of two competing processes occurring concomitantly.

To directly quantify the relationship between the degree of solvent exposure and sulfenic acid reactivity, we systematically varied the stretching force after triggering the formation of sulfenic acid. Using a six-pulse force protocol (Fig. 4.7), the *initial* two pulses unfolded the protein (150 pN, 0.3 s) and triggered the rupture of the disulfide bond within a short time (500 pN, 2 s). The force was then varied within a range spanning 0–500 pN during 4 s (grey box) before being withdrawn (0 pN, $t_q = 8$ s) to trigger collapse¹²¹ and oxidative folding. As before, the folding success was probed by two high force *test* pulses (150 pN followed by 500 pN). The folding trajectories shown in Fig. 4.7 a qualitatively show that the number of 10 nm steps observed in the test pulse (red asterisks) significantly decreased as the stretching force was increased. This tendency was quantitatively verified in Fig. 4.7 b, displaying the % of disulfide bond reformation as a function of the stretching force. Crucially, the observed trend can be nicely reproduced by the worm-like chain (WLC) model of polymer elasticity (using a persistence length $p = 1.2$ nm)²⁷⁷, which relates the protein's end-to-end length with the pulling force. At each defined force, the WLC model defines an equilibrium end-to-end conformation that is characterized by a different degree of solvent exposure of the protein residues²⁷⁸. Altogether, these experiments demonstrate that the reactivity of sulfenic acid—intimately related to the oxidative folding fate of the protein— is largely determined by the protein's conformation.

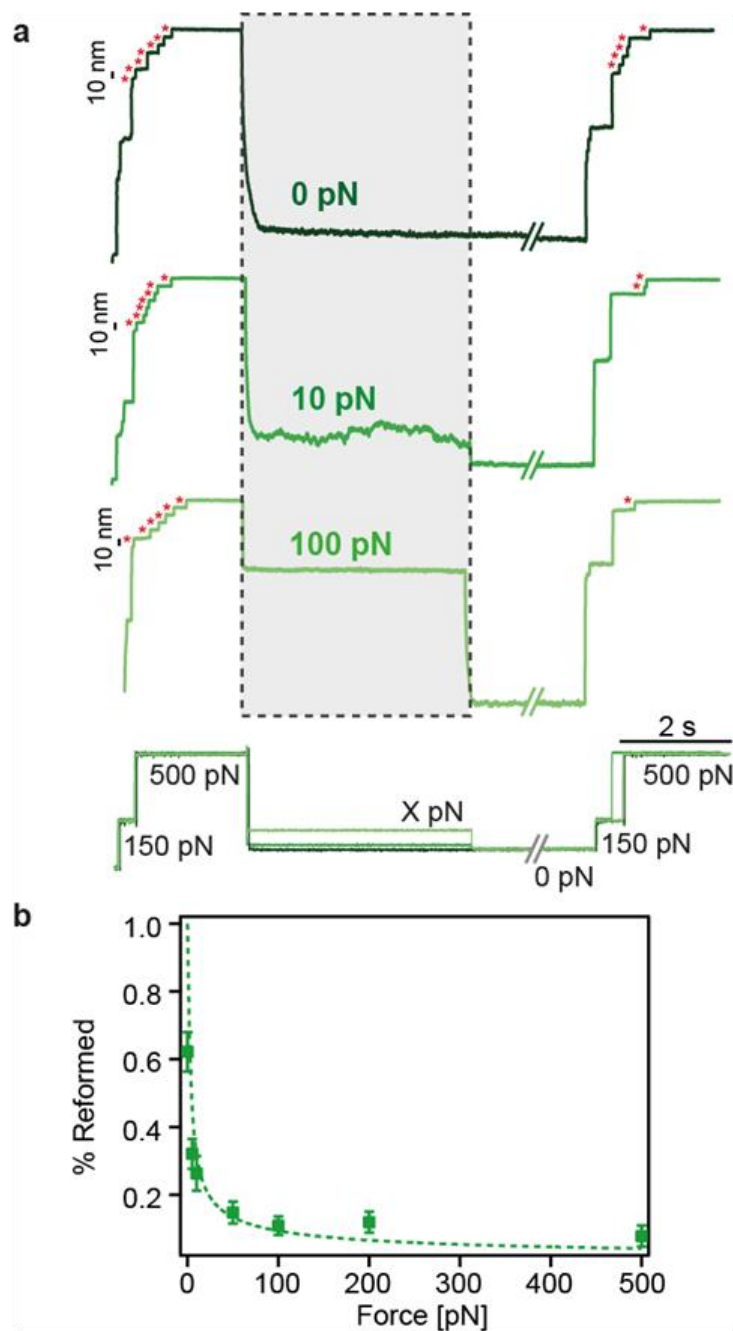


Figure 4.7 The force-dependent conformation of the protein modulates the chemical evolution of sulfenic acid. The degree of solvent exposure is tuned by the pulling force, which determines the (I27^{E24C}-K55C)₈ protein end-to-end length. **(a)** A six-pulse force protocol probes the effect of force and solvent exposure on the outcome of oxidative folding. The number of 10 nm steps (red asterisks) significantly decreases with increasing the force applied to the protein after the sulfenic acid is formed (grey box), varying from 0 pN (dark green) to 100 pN (light green). **(b)** The % of disulfide bond reformation shows a drastic decrease with an increase in the stretching force, and the evolution can be captured by the worm-like chain (WLC) model of polymer elasticity ($p=1.2$ nm). This trend is indicative of the distinct reactivity of sulfenic acid as a function of its exposure to the solvent. At low forces the local protein environment shields the sulfenic acid, allowing it to condense into a reformed disulfide bond that ensures successful oxidative folding. By contrast, at higher forces the sulfenic acid is fully exposed to the solution and undergoes further chemical modification, ultimately leading to protein misfolding, hallmarked by the inability of the protein to recover its native mechanical stability.

In order to investigate the molecular origin of the chemical modification that prevented the reformation of the disulfide bond at high pulling forces, we conducted Mass Spectrometry (MS) measurements on the monomeric I27^{E24C-K55C} protein under high pH conditions (*All mass spectrometry measurements performed by Steven Lynham, King's College London*). MS spectra revealed the presence of both sulfinic and sulfonic acid, thus suggesting the successive irreversible oxidation of sulfenic acid (Appendix II Fig. 2-3 and Table 1). Notably, the presence of aldehyde (with a concomitant loss of H₂S) was also detected with good evidence (Appendix II Fig. 4 and Table 1). Indeed, aldehyde formation has been long reported as a sub-product of the decomposition of sulfenic acid under high alkali conditions^{265, 279, 280}. While both the protein conformation and the timescales sampled in the Mass Spectrometry and single molecule mechanics experimental approaches are certainly different, the MS data provides excellent complementary information that illustrates the most plausible chemical processes responsible for the misfolding events captured in the nanomechanical experiments, greatly affecting the folding and elastic properties of the individual titin polypeptides.

DIMEDONE BINDING CONFIRMS THE PRESENCE OF SULFENIC ACID

To provide a final, unambiguous proof that –SOH is actively involved in the molecular mechanism driving the reformation of disulfide bonds, we directly tested the effect of dimedone on the outcome of our oxidative folding assay. Dimedone has been vastly demonstrated to effectively react with sulfenic acid (but not with thiols) with high specificity (Fig. 4.8 a)^{244, 252}. Addition of 30 mM of dimedone to the measuring solution dramatically blocked both the reformation of the disulfide bond and the folding of the protein (Fig.4.8 b), decreasing the bond reformation efficiency down to ~15% (Fig. 4.8 c).

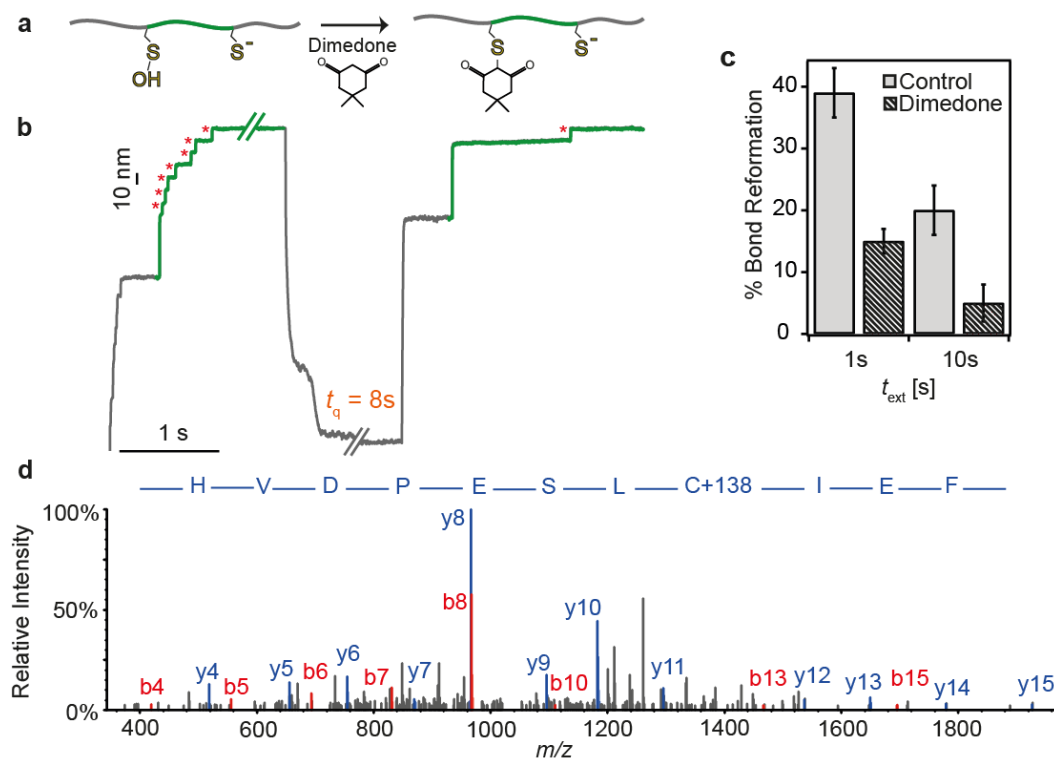


Figure 4.8 Dimedone selectively targets sulfenic acid and blocks oxidative folding. (a) The newly formed sulfenic acid moiety can be selectively trapped by dimedone. (b) When an individual (I27^{E24C-K55C})₈ polypeptide is reduced and stretched in the presence of a 30 mM solution of dimedone, the subsequent *test* pulse shows a dramatic decrease in the presence of both ~10 nm (red asterisks) and ~15 nm steps, implying that oxidative folding is hindered. (c) The percentage of disulfide bond reformation (and hence of successful folding) decreases with t_{ext} , the time the protein remains extended and stretched at a high force. For $t_{ext} = 1s$, the % reformation decreases down to $15.6 \pm 2.3\%$ in the presence of dimedone. Upon increasing $t_{ext} = 10s$, the % of reformation is further reduced ($6.0 \pm 3.3\%$). (d) The MS/MS fragmentation spectra of an I27^{E24C-K55C} monomer under high pH conditions provided unambiguous evidence of the presence of the dimedone, hallmarked by a cysteine mass shift of $\Delta m/z = +138$ Da.

This drastic decrease in the reformation (and subsequent refolding) efficiency was further exacerbated (down to a residual ~6%) when the protein was left fully extended at high forces for a longer period of time, $t_{ext} = 10s$, promoting dimedone binding (Fig. 4.8 c). By contrast, dimedone did not prevent the protein from successfully refolding when the disulfide bond remained oxidised (Fig. 4.9). In order to indisputably verify the presence of the covalent S-dimedone adduct, we conducted Mass Spectrometry measurements on the monomeric I27^{E24C-K55C} protein under high pH conditions. The resulting MS/MS fragmentation spectra identified, for both 24 and 55 cysteines, a shift in $\Delta m/z = +138$ Da (Fig. 4.8 d, Appendix II Fig. 5 and Table 2), which confirmed the presence of dimedone.

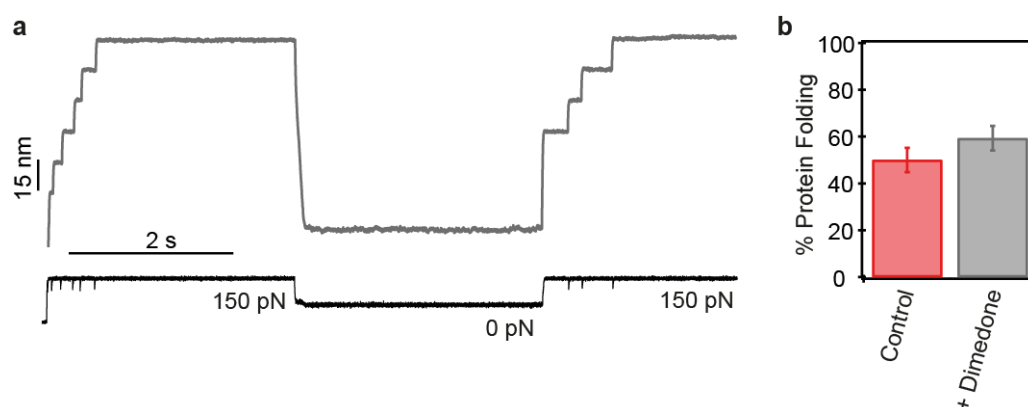


Figure 4.9 The refolding of the oxidised (I27_{E24C-K55C})₈ polyprotein is not blocked by the presence of dimedone. In order to indisputably prove that dimedone only interacts with the sulfenic acid moiety created upon the alkali-mediated reduction of the disulfide bond, we monitored the folding efficiency of the oxidised (I27_{E24C-K55C})₈ polyprotein at pH = 12.8 and under 30 mM of dimedone. As shown in the folding trajectory (a) and bar diagram (b), dimedone does not impede the successful refolding of the oxidised protein, hallmarked by the presence of 15 nm steps in the *test* pulse. Therefore, these experiments demonstrate that dimedone alone does not interfere with the refolding of the (I27_{E24C-K55C})₈ polyprotein.

4.3 DISCUSSION

Oxidative stress defines elevated intracellular levels of reactive oxygen species (ROS) —mostly including superoxide anion, hydrogen peroxide and hydroxyl radicals— that cause damage to DNA, lipids and proteins²⁸¹. While ROS were originally known for oxidizing various cellular compartments and promoting aging and a broad range of pathologies, more recent evidence shows that ROS also acts as signalling molecules that regulate basic cellular processes including growth, differentiation and cell migration^{263, 282}. It is increasingly clear that a major mechanism of redox signalling is the dynamic regulation of protein function by the chemiselective oxidation of cysteine residues. The first product of oxidation is cysteine sulfenic acid, exhibiting an oxidation state of (0). Despite its ephemeral nature, sulfenic acid has been trapped in a number of crystallised proteins^{249, 283}. These carry a wide range of functionalities²⁴⁹, encompassing enzyme catalysis (such as protein tyrosine phosphatases²⁸⁴, kinases²⁸⁵ and cysteine proteases²⁸⁶), transcription factors (including OxyR²⁸⁷, Yap1²⁸⁸ and NF- κ B²⁸⁹), and ion channels (such as ORAI1²⁹⁰ and Kv1.5²⁹¹). This progress notwithstanding, these structures provide a static snapshot, and the fast dynamics of such important post-translational modification often evaded characterization.

The relatively small number of crystallized proteins contrasts with the increasingly larger number of proteins where sulfenic acid has been identified *in-vivo*. Indeed, the dimedone-based cell-permeable chemical probe DAz-2, capable of detecting sulfenic acid modified proteins directly in living cells²⁸², has allowed for a comprehensive analysis of the cellular “sulfenome”²⁵⁵, revealing the presence of (>1000) S-sulfenylation sites on more than 700 proteins in intact cells, highly prevalent in different cancer cell lines^{251, 292}. Importantly, these newly identified protein targets largely expand the diversity of functionalities of the already identified sulfenylated proteins.

Therefore, unveiling the molecular mechanisms defining sulfenic acid reactivity is an essential requirement for the understanding of redox regulation in cells.

Here we employed a combination of single molecule force spectroscopy and MS measurements to capture the reactivity of sulfenic acid within the core of a single molecule. Our results demonstrate that the fine regulation of protein functionality provided by sulfenic acid is largely conformation-dependent and intrinsically linked to the complexity of its reactivity. When the reduced protein is mechanically unfolded and fully stretched, the newly formed –SOH moiety is completely exposed to the solvent environment. Under these circumstances, sulfenic acid, which is traditionally viewed as an intermediate conformation *en route* to other higher oxidation states²⁶², undergoes further modification to sulfinic (Cys-SO₂H) and sulfonic (Cys-SO₃H) acids as well as aldehyde condensation, as revealed by MS measurements (Appendix II Fig. 2-4). Crucially, the chemical evolution towards these distinct chemical moieties inevitably results in protein misfolding (Fig. 4.6 and 4.7). Such irreversible redox damage is evaded after protein collapse when the transient sulfenic acid condenses with the neighbouring free cysteine in the I27 mutant, to create a rescuing disulfide bond that introduces a conformational constraint to the protein with remarkable implications in the context of the oxidative folding.

Our single molecule approach allows independent measurement of the protein folding and disulfide bond formation processes, enabling us to directly visualise the molecular events defining the complex kinetics of the oxidative folding mechanism. The crucial role of disulfide reformation in protein refolding was further probed in experiments on the pre-reduced (I27^{E24C-K55C})₈ polyprotein at pH =12.8. While the oxidised protein was able to refold under high pH conditions (Fig. 4.3), the folding efficiency of the pre-reduced protein was significantly decreased down to ~20%, exhibiting a seemingly slow

folding kinetics, $k_{\text{fold}} = 0.10 \text{ s}^{-1}$ (Appendix II Fig. 6). This result contrasts with the much higher folding efficiency of the sulfenylated protein, which folds upon the reformation of the disulfide bond ($k_{\text{ref}} = 0.45 \text{ s}^{-1}$), ultimately setting the associated folding rate ($k_{\text{fold}} = 0.46 \text{ s}^{-1}$). These experiments demonstrate that while the protein is mostly unable to successfully refold on its own under high pH, the reformation of the disulfide bond dramatically increases its folding efficiency. In this sense, sulfenic acid emerges as a key molecular player able to rescue the protein from misfolding by forming a protective disulfide bond that allows the protein to successfully refold. This kinetic scheme seems to differ from the enzyme-mediated oxidative folding findings catalysed by PDI¹⁸⁸- and DsbA²⁹³, whereby the same I27 protein mutant was able to refold on its own while keeping the catalytic enzyme attached through a mixed disulfide conformation before the disulfide bond was successfully reformed at a later stage, the overall kinetics being largely controlled by the off-rate dynamics of the catalytic enzyme.

We postulate that the oxidative folding mechanism promoted by -SOH condensation with a neighbouring intramolecular thiolate is mechanistically different from the process where the reformation of the disulfide bond occurs through a mixed disulfide intermediate, which characterises the enzyme-catalysed oxidative folding process²⁵⁹. To test this hypothesis, we repeated the force-quench experiments at $\text{pH} = 7.2$ using the small thiol 3-mercapto-1-propanol as the attacking nucleophile²⁷³ (Fig. 4.10). In this case, the rupture of each individual disulfide bond resulted in the formation of a thiolate and the newly formed mixed disulfide. Unlike the case where sulfenic acid could readily reform the cleaved disulfide, thus triggering the protein to correctly fold, when 3-mercapto-1-propanol was used as a nucleophile, the number of reformation events was vanishingly small. By contrast, these experiments highlighted (albeit with a relatively low abundance) the presence of 25 nm

steps in the test pulse. Altogether, our experiments demonstrate the unique capabilities of sulfenic acid to form a disulfide bond employing a chemical mechanism that is fundamentally different from that occurring during enzyme-mediated protein folding^{188, 259}. In this novel –SOH mediated scenario, the newly formed disulfide reduces the conformational space of the protein, and subsequently triggers almost “instantaneous” folding of the remaining protein in the timescale of our experiments. Thus, the kinetics of disulfide bond formation is the rate-limiting step that defines the overall kinetics of the non-enzymatic oxidative folding process. It is certainly remarkable that enzyme-free oxidative folding occurs under high pH conditions at a similar rate ($k_{\text{ref}} = 0.45 \text{ s}^{-1}$) than that catalysed by a dedicated enzyme such as PDI ($k_{\text{ref}} = 0.35 \text{ s}^{-1}$) and DsbA ($k_{\text{ref}} = 0.70 \text{ s}^{-1}$)^{188, 293}.

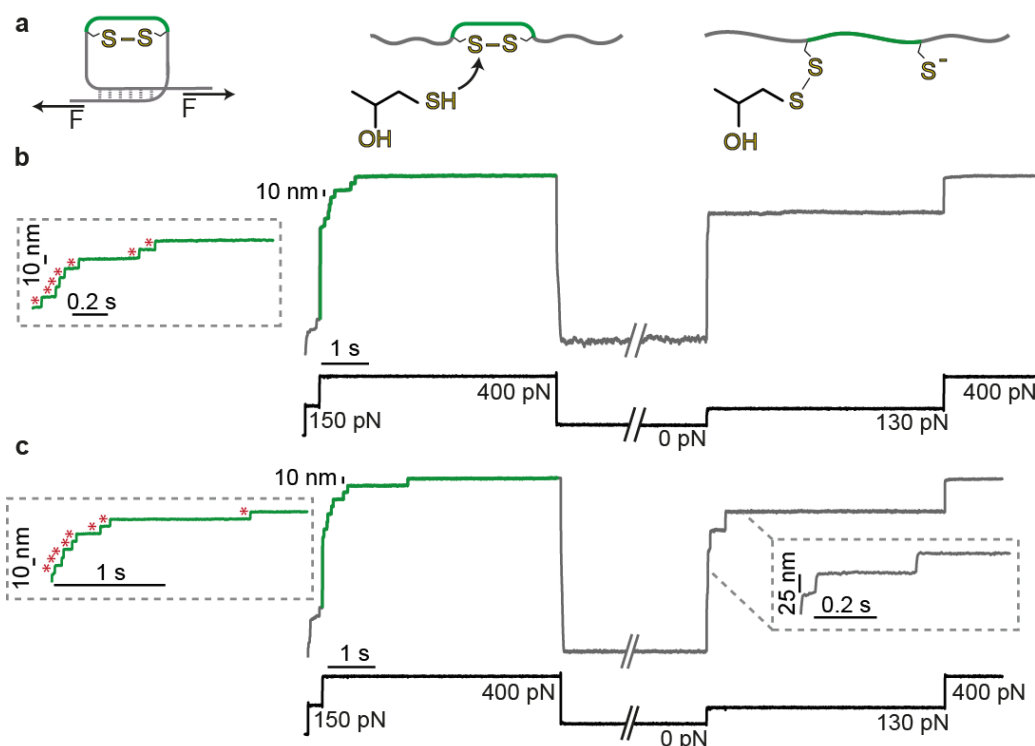


Figure 4.10 Successful non-enzymatic oxidative folding can only be observed after the presence of an alkali-mediated reduction event, which introduces a sulfenic acid moiety in the protein in a controlled manner. By contrast, when another nucleophile such as 3-mercapto-1-propanol is used as a disulfide bond reducing agent, disulfide bond cleavage results in the presence of a mixed disulfide and a thiol. The force-quench trajectories that do not exhibit the reformation of the disulfide bond. (a) While in most of the trajectories (~65%) we observed that refolding was blocked (b), in the remaining (~35%) of the trajectories at least one 25 nm step was observed (c), fingerprinting the correct folding of the protein lacking a disulfide bond. We hypothesize that this last scenario might correspond to the reduction of the mixed disulfide by another thiol from the solution, leaving the substrate with two free thiols. The presence of the low occurrence 25 nm steps indicates that the reformation of the protein disulfide is discouraged, supporting the idea that two free thiols will not form a disulfide bond unless an oxidant agent (such as sulfenic acid) activates the thiols by catalysing the process.

Our experiments demonstrate that the probability of reforming the disulfide bond increases exponentially with time (Fig.4.5). This implies that, when the protein lies within the non-folded collapsed conformation after the force-quench, the highly reactive $-SOH$ moiety is sterically stabilized within the microenvironment of the collapsed protein, most likely through becoming cryptic to the solvent. This situation contrasts with the fast evolution of the $-SOH$ in the fully extended state (Fig. 4.6). At the present stage it is difficult to unambiguously pinpoint the exact origin of the fast reactivity undergone by the solvent-exposed sulfenic acid. Due to the relatively short timescale of our single molecule experiments, it is unlikely that the progression to higher

oxidation state species is favoured, and hence we hypothesise that the aldehyde formation mechanism might be prevalent^{265, 279, 280}, which would account for the main pathway described by the single exponential fit. Given the slightly higher accuracy of the double exponential fit, we postulate a possible competition between the high oxidation pathway (promoted by the dissolved oxygen in solution) and the aldehyde formation mechanism, exhibiting similar rate constants. Noteworthy, we cannot rule out the possibility that a fraction of the aldehyde population found in our MS experiments is a result of the direct alkali-mediated cleavage of the disulfide bond through an α -elimination mechanism, which competes with a β -elimination and the hydroxyl-mediated S_N2 nucleophilic attack^{264, 265, 266, 294} (Appendix II Fig. 7). Indeed, our MS spectra show evidence of the three competing pathways (Appendix II Fig. 2-4 and Table 1). However, it is very likely that, in the presence of force, the hydrolysis (S_N2) pathway is largely favoured over the elimination route, as suggested by recent quantum mechanical simulations under force¹⁷³, in which case the presence of aldehyde is likely to stem mostly from the evolution of sulfenic acid. Further verification of the plausible competing reaction pathways under force, probably guided by ab initio simulations, might provide a comprehensive understanding of this complex reactivity landscape. In any case, while the high pH conditions tested in our experiments probably trigger a faster $-SOH$ reactivity towards the aldehyde product than one would perhaps expect under more physiologically relevant conditions, our experiments readily demonstrate that the reactivity of sulfenic acid is largely conformation-dependent, crucially contributing to the folding fate of the protein. We speculate that this mechanism might hold true in situations where sulfenic acid is triggered by other biologically-occurring oxidative species such as H_2O_2 . To test this hypothesis, we conducted folding experiments on the wt-I27 Ig module (containing two cryptic and native cysteines in positions 47 and 63 that do not form a disulfide bond) in the

presence of 300 μM H_2O_2 at neutral pH (Fig. 4.11). Stretching and unfolding the (wt-I27)₈ polyprotein exposed the cryptic cysteines to the oxidative conditions. Quenching the force and re-unfolding back again resulted in (i) the misfolding of the protein, fingerprinted by the absence of recovery of mechanical stability, and (ii) the inability of the protein to completely re-extend up to its unfolded contour length, suggestive of the presence of a stiff, non-native (probably interdomain) disulfide bond. Controls were performed with the double mutant (I27_{C47A-C63A})₈ polyprotein (devoid of both cysteines) showing full refolding and complete protein extensibility. These experiments, conducted in the presence of the less controllable yet physiologically more relevant H_2O_2 oxidative agent, strongly suggest that, when solvent exposed upon mechanical unfolding⁴³, cryptic cysteines undergo post-translational modifications that dramatically block protein folding, while lending further support to the sulfenic acid-mediated mechanism of disulfide bond formation that we investigated using the high pH experimental strategy.

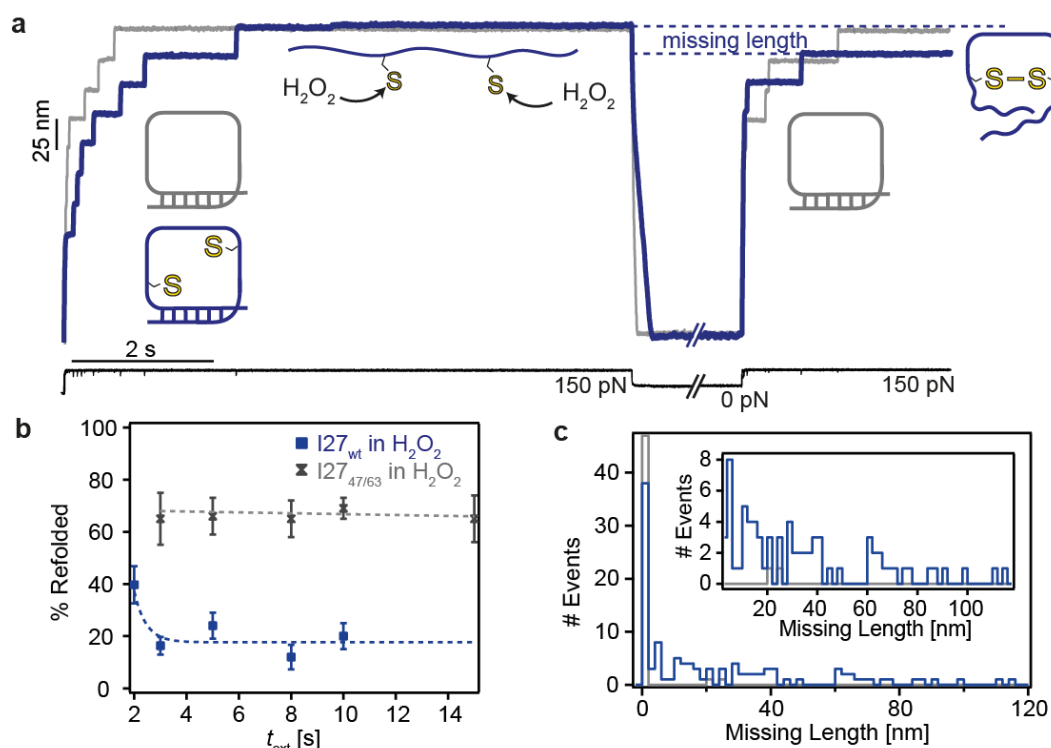


Figure 4.11 Hydrogen peroxide induces S-sulfenylation in cryptic cysteines and promotes the formation of non-native disulfide bonds in wt-I27. (a) Mechanically unfolding wt-I27 (blue), containing two buried, native cysteines that do not form a disulfide bond, in the presence of 300 μ M hydrogen peroxide (H_2O_2) gives rise to the regular unfolding pattern where each domain unfolds in steps of ~ 25 nm. Mechanical unfolding exposes the two previously cryptic cysteines to the oxidative aqueous environment. The force is subsequently quenched down to 0 pN for $t_q = 5$ s to allow the proteins to refold, before the force is re-applied back again to probe folding success. The folding dynamics of the (wt-I27)₈ polyprotein is compared to that of the (I27_{C47A-C63A})₈ polyprotein, devoid of native cysteines (grey). (b) Under these oxidative conditions (wt-I27)₈ dramatically decreases its folding efficiency, an effect that is further exacerbated with t_{ext} . By contrast, the folding success of the (I27_{C47A-C63A})₈ polyprotein is not affected by its exposure to H_2O_2 . (c) A recursive feature of the (wt-I27)₈ folding trajectories is its inability to recover its initial unfolded length upon being stretched in the *test* pulse. Such a systematic shortening of the protein, spanning a wide range of lengths, is not observed in the case of the (I27_{47/63})₈ polyprotein. These 'missing length' events are suggestive of the formation of non-native disulfide bonds, which increase the stiffness of the protein. Altogether, these experiments demonstrate that cryptic cysteines in the wild-type I27 protein undergo H_2O_2 -mediated post-translational modifications (S-sulfenylation), which dramatically affect the fate of the mechanical folding efficiency; if sulfenic acid undergoes 'uncontrolled' evolution to hyperoxidised species such as sulfinic and sulfonic acid, a decrease the folding efficiency of the wt-I27 protein is measured. By contrast, upon quenching the force, sulfenic can promote the formation of non-native disulfide bonds that short-cut the protein and diminishes its extensibility.

Taken together, our nanomechanical experiments put forward a rather unique kinetic scheme (Fig. 4.12) that captures the intricate conformation-dependent reactivity of the transitory sulfenic acid intermediate, directly linked with the folding fate of the protein. In the case of titin, a protein that is physiologically exposed to mechanical force, such –SOH regulated folding scenario might have important consequences for the functional protein elasticity.

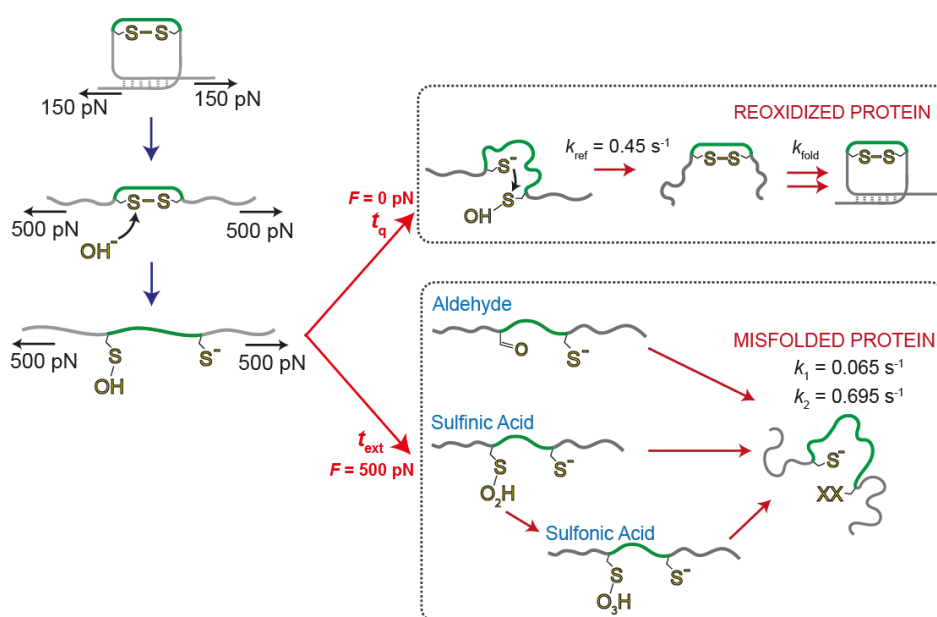


Figure 4.12 A generic kinetic scheme for the force-triggered, conformation-dependent reactivity of sulfenic acid and its implications for oxidative protein folding. On the application of mechanical force, the hydroxide anions rupture a protein disulfide bond, creating a sulfenic acid and a thiolate. In the absence of force ($F = 0 \text{ pN}$) during the quench time (t_q) the sulfenic acid moiety remains cryptic in the protein's collapsed state. Under these conditions, the longer t_q , the higher the probability the thiolate anion can attack the sulfenic acid moiety (in the absence of dimedone), thus creating a disulfide bond with an associated rate of $k_{ref} = 0.45 \text{ s}^{-1}$. The reformation of the disulfide bond triggers the 'instantaneous' refolding of the protein (essentially set by the disulfide bond reformation rate). By contrast, when the protein is reduced and stretched at high force ($F = 500 \text{ pN}$), sulfenic acid is fully exposed to the solvent and undergoes further modification either to the formation of sulfinic acid (which may be further oxidised to sulfinic acid) or to aldehyde. Such further reactivity of sulfenic acid results in irreversible misfolding of the reduced (I27^{E24C-K55C})₈ polypeptide, with a consequent decrease in its stiffness properties.

Titin is a molecular ruler that determines the passive elasticity of muscle, and is key to ensure, for example, cardiac function²⁹⁵. Indeed, alterations in the stiffness of cardiac titin are related to a number of myopathies, such as diastolic dysfunction²⁹⁶. The I-band region of titin is formed by a compliant, unstructured region (PEVK and N2B) intercalated between folded Ig domains, which are mechanically resistant²⁹⁷. In order to adapt to the mechanical perturbations induced by the heartbeats, titin needs to reversibly modulate its length by entropic extension/recoil of the compliant region and unfolding/refolding (at least of the mechanically weak) Ig domains⁴³. Upon Ig domain unfolding, the stiffness of the protein is greatly reduced, and it is increased back again upon refolding. Recent pioneering evidence shows that S-glutathionylation of cryptic cysteines blocks titin refolding⁴³. Our results are likely to add S-sulfenylation to the toolbox of post-translational modifications that regulate titin elasticity; whereas cysteine hyperoxidation irreversibly suppresses Ig27 refolding, thus severely decreasing its stiffness, the reformation of a protective disulfide bond has the opposing effect. In this sense, sulfenic acid would act as “mechanical switch”, helping titin adapt its elasticity in response to the fluctuating oxidative stress conditions. We hence speculate that sulfenic acid might work as an active redox sensor that directly regulates protein elasticity.

5. TAILORING PROTEIN NANOMECHANICS WITH CHEMICAL REACTIVITY

The nanomechanical properties of elastomeric proteins determine the elasticity of a variety of tissues. A widespread natural tactic to regulate protein extensibility lies in the presence of covalent disulfide bonds, which significantly enhance protein stiffness. The prevalent in-vivo strategy to form disulfide bonds requires the presence of dedicated enzymes. Here we propose an alternative chemical route to promote non-enzymatic oxidative protein folding via disulfide isomerization based on naturally occurring small molecules. Using single molecule force-clamp spectroscopy, supported by DFT calculations and mass spectrometry measurements, we demonstrate that subtle changes in the chemical structure of a transient mixed-disulfide intermediate adduct between a protein cysteine and an attacking low molecular-weight thiol have a dramatic effect on the protein's mechanical stability. This approach provides a general tool to rationalize the dynamics of S-thiolation and its role in modulating protein nanomechanics, offering molecular insights on how chemical reactivity regulates protein elasticity.

This chapter is based on the article published 6 June 2017 in Nature Communications

(Attached in appendix VI)

5.1 INTRODUCTION

Modular proteins represent a natural strategy to achieve highly elastic materials with enhanced mechanical properties²⁹⁸. For example, the giant protein titin is formed by a large number of stiff immunoglobulin (Ig) and Fibronectin-like (Fn) domains intercalated between non-structured, extensible sequences (PEVK, N2B) that are mechanically compliant^{30, 60, 299}. Combined, the distinct mechanical properties of both structurally diverse components regulate the large-scale passive elasticity of muscle under physiological conditions⁶⁰. At the local scale, modulation of the nanomechanics of a single protein is mainly achieved through four specific molecular mechanisms; first and foremost (i), forced unfolding converts the mechanically rigid native state into an elastic unfolded conformation devoid of mechanical stability²⁷⁸. (ii) Secondly, a simple strategy to alter the mechanical properties of the protein's native state is through site-directed mutagenesis in well-defined positions within the force-bearing ("mechanical clamp") structural motif¹³. Alternatively, (iii) ligand binding can drastically increase the mechanical stability of a folded protein^{38, 89, 300, 301}, or prevent an unfolded polypeptide from recovering its native mechanical stability by blocking successful refolding³⁰². Both strategies often result in an all-or-none switch between both extreme mechanical stabilities⁹⁴. (iv) A far less explored mechanism to modulate protein mechanics is achieved via post-translational modifications, through subtle yet crucial changes in chemical reactivity within the protein core^{44, 303}. The most relevant protein modification with a mechanical impact is arguably the formation of disulfide bonds, which are crucial modulators of protein extensibility. Disulfide bonds establish a rigid, molecular shortcut that impairs the complete force-induced unfolding of a variety of proteins including titin³⁰⁴, and the extracellular CAM (Cell Adhesion Molecules) protein superfamily¹⁸³. In these cases, modulation of protein elasticity is generally dynamic and

occurs under redox control^{305, 306}; while oxidising conditions guarantee the presence of the covalently rigid disulfide bond, reducing conditions induce its rupture, thereby triggering full protein unfolding under force. Hence, the ability of the protein to form a disulfide bond during its folding route, a process known as oxidative folding²⁴¹, emerges as a crucial functional determinant of protein mechanics.

In vivo, the formation of disulfide bonds occurs through two independent pathways³⁰⁷. Within the ‘redox route’, a cysteine thiol can be oxidised via one or two electrons to a higher oxidation state species. Most typically, the free thiol is first oxidised to sulfenic acid, which then condenses with a neighbour protein thiol to form a protective disulfide bond that prevents irreversible protein damage due to cysteine overoxidation^{44, 244, 249}. The second and most prevalent means of disulfide bond formation is catalysed by a superfamily of dedicated oxidoreductases—including thioredoxin, glutaredoxin, the bacterial Dsb and the eukaryotic protein disulfide isomerase (PDI) enzymes—that share a common active site containing two highly conserved catalytic cysteines^{177, 184}. Their mechanism of action employs a thiol-exchange nucleophilic reaction, creating a transient yet obligatory mixed-disulfide intermediate between a substrate cysteine and the N-terminal cysteine of the enzyme’s active site³⁰⁸. The determinant role of the mixed disulfide conformation to orchestrate oxidative folding has deserved great attention^{309, 310}. While the static conformational picture has been accessed through a variety of protein structures^{311, 312, 313, 314, 315}, the dynamics of this crucial fleeting mixed disulfide intermediate has only now begun to be understood^{188, 293}.

Inspired by the functioning of oxidoreductases enzymes, we now aim to acquire a deep understanding of the naturally employed chemical tactics to rationalise the molecular mechanisms underpinning other closely related protein S-modifications. We will use this *ad-hoc* acquired knowledge to design

new experimental strategies that rationally exploit the chemical properties of naturally occurring small molecules to modulate protein elasticity. A main advantage of this experimental approach is that it is not restricted to enzymatic activity, thus having the potential to be scaled up to large quantities, it has a predictive character and it offers exquisite control over the chemical selectivity of the small molecules as it avoids the inextricable conformational changes involved during protein-protein interactions that define enzymatic catalysis.

The most obvious candidates for the small chemical modulators are low-weight molecular thiols (LMW-SH)³¹⁶, especially encompassing cysteine (Cys), cysteinylglycine (CysGly), homocysteine (Hcys) and glutathione (GSH). Present both intracellularly and also in the human plasma, these small thiols are often regarded as biomarkers for oxidative stress³¹⁷, and their presence as mixed-disulfide with proteins increases with age^{318, 319}. While GSH reactivity has been the focus of extensive research^{320, 321}, the reactivity of Cys and Hcys and their related protein post-translational modifications, namely S-cysteinylation and S-homocysteinylation, have comparatively received much less attention. In particular, high levels of L-Hcys have been associated to reduced muscle function and integrity³²². In fact, hyperhomocysteinemia is considered as an independent risk factor for cardiovascular disease^{323, 324, 325, 326}, and has been linked to pathological alterations in mechanically functional proteins such as the impaired fibrillin-fibronectin assembly³²⁷. Despite important efforts, a direct molecular link between the presence of S-homocysteinylation and the pathological mechanical effect is not completely understood.

Here we use single molecule force-clamp spectroscopy, supported by Mass Spectrometry (MS) and Density Functional Theory (DFT) calculations, to provide mechanistic insights into the direct link between the chemical reactivity of biologically relevant LMW thiols and their effect on mechanical

protein folding. Our results demonstrate that the life-time of the ephemeral mixed disulfide intermediate structure is an essential modulator of the nanomechanics of cardiac titin, switching between two successfully refolded conformations with markedly different stiffness properties.

5.2 RESULTS

S-CYSTEINYLATION PROMOTES NON-ENZYMATIC OXIDATIVE FOLDING

We applied a constant force to a polyprotein composed of eight identical domains of a mutated titin Ig27 domain containing two cysteines in positions 24 and 55, which readily form a structurally buried disulfide bond (Fig. 5.1 a). A specifically designed five-pulse force protocol measures the protein elongation over time in the presence of 13.3 mM of L-cysteine (Fig. 5.1 b), enabling the independent capture of each individual disulfide bond rupture and reformation event, with single bond resolution (Fig. 5.1 c)⁴⁴. Applying a short (500 ms) pulse of 150 pN to the (I27_{E24C-K55C})₈ polyprotein elicits steps of ~15 nm, corresponding to the unfolding of the protein up to the position of the rigid disulfide bond (grey), which becomes exposed to the solution. A second pulse at a higher force (300 pN) triggers a ~10 nm stepwise increase in the protein length (red), corresponding to the force-catalysed rupture of each individual disulfide bond by the attacking L-cysteine nucleophile. The initial pulse, composed of ~10 nm and ~15 nm steps (Fig. 5.1 d), results in a fully stretched and reduced protein form exhibiting a free cysteine thiolate and a mixed disulfide conformation between the native cysteine and the attacking L-cysteine (Fig. 5.1 b).

Removal of the pulling force for $t_q = 8$ s (a period of time that is long enough to ensure quantitative protein refolding) triggered the collapse and folding of the protein. A second test pulse that mirrors the initial pulse probes the folding success of the folding reaction for a given t_q . Similar to the initial pulse, the elongation of the protein in the test pulse occurs in steps of ~15 nm (unfolding) followed by the sequential reduction of each individual disulfide bond (~10 nm steps, Fig. 5.1 e). This sequence of events fingerprints the success of non-enzymatic oxidative folding process. In particular, the presence of ~10 nm

steps in the test pulse unambiguously hallmarks the non-enzymatic reformation of individual disulfide bonds, surprisingly occurring in the absence of any oxidising agent, and its intricate link to the successful refolding of the protein.

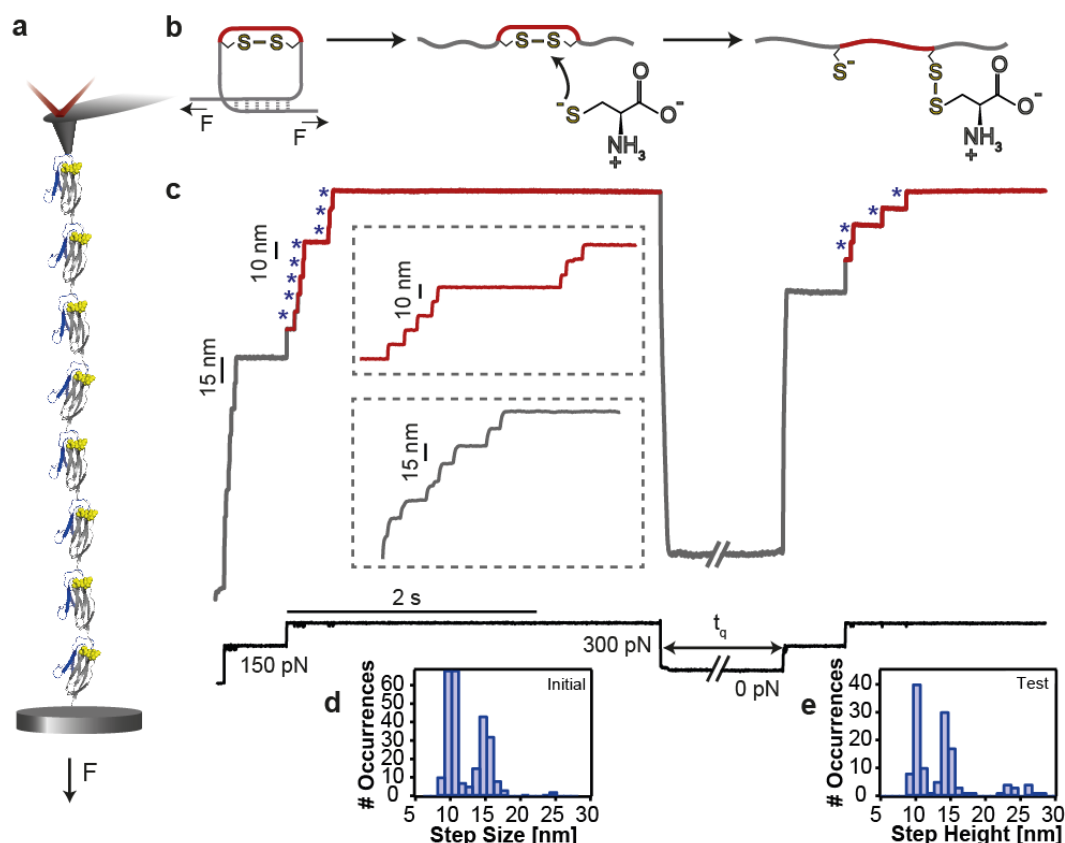


Figure 5.1 Single molecule force-clamp spectroscopy captures non-enzymatic disulfide bond reformation. (a) A polypeptide construct formed of eight identical repeats of a mutated I27th Ig domain of titin containing a buried disulfide bond (I27^{E24C-K55C})₈ is tethered between a gold substrate and an AFM cantilever tip, which applied a constant stretching force throughout the experiment. (b) Upon the application of force, the protein unfolds (grey) and exposes the disulfide bond to a L-cysteine solution (13.3 mM, pH = 7.2), which is able to attack the disulfide bond through a S_N2 nucleophilic attack, marked by the extension of the amino acids that were trapped behind the disulfide bond (red). The force-induced cleavage of the protein disulfide results in a stretched protein conformation containing a reduced protein thiol and a mixed disulfide bond between the attacking L-cysteine and a native cysteine. (c) Experimental folding trajectory of the (I27^{E24C-K55C})₈ revealed by a 5-pulse force quench protocol. The initial pulse unfolds the protein and ruptures the disulfide bond; the first pulse $F=150$ pN unfolds and extends all the modules of the (I27^{E24C-K55C})₈ in steps of ~15 nm (grey inset), exposing the disulfide bonds to the L-cysteine solution. The second pulse at $F = 300$ pN catalyses the rupture of each disulfide bond (blue asterisks), giving rise to a staircase-like increase of protein length occurring in ~10 nm steps (red inset). Quenching the pulling force ($F=0$ pN) triggers the protein to collapse and fold for a time quench $t_q=8$ s. The subsequent test pulse probes the folding status of the protein. Mirroring the initial pulse, a force pulse of $F=150$ pN unfolds and extends the refolded protein up to the newly created disulfide bond (grey line) whereas the last $F=300$ pN pulse is able to rupture it again (red line). The step-size histograms corresponding to the (d) initial and (e) test pulse highlight in both cases the presence of ~10 nm and ~15 nm steps.

S-HOMOCYSTEINYLATION BLOCKS OXIDATIVE FOLDING

To probe the generality of the small-thiol mediated non-enzymatic oxidative folding process, we next probed the outcome of oxidative folding in the presence of the chemically analogous homocysteine nucleophile (Fig. 5.2 a), only differing from L-cysteine in the presence of an extra methylene group next to the nucleophilic thiolate (Fig. 5.2 b). Surprisingly, in this case the test pulse was mostly populated by steps of ~25 nm, and the proportion of ~10 nm steps, hallmark of disulfide reformation, was vanishingly small (Fig. 5.2 a). The presence of ~25 nm in the 150 pN test pulse stands for a mechanically weaker protein that has properly folded but has not reformed its internal stiff disulfide bond. The percentage of reduced, folded protein (~25 nm steps) is dramatically higher (~40%) than in the case of L-cysteine (~3%), whereas the situation is completely reversed in the case of disulfide bond reformation, which is favoured with L-cysteine (~30%) and significantly reduced (~10%) when Hcys is the attacking nucleophile (Fig. 5.2 c,d). Altogether, these results conclude that the attacking nucleophile has a profound effect on the nanomechanics of the refolded protein; while L-cysteine promotes a stiff, refolded protein with a properly reformed disulfide bond, S-homocysteinylation blocks oxidative folding but eventually allows the protein to fold into a much compliant native state conformation void of the reformed disulfide bond. The dramatic contrast in the effect that both analogous post-translational modifications have on the oxidative folding fate of the protein and on its nanomechanical properties is certainly surprising given the closely related chemical structures of both low molecular weight thiols.

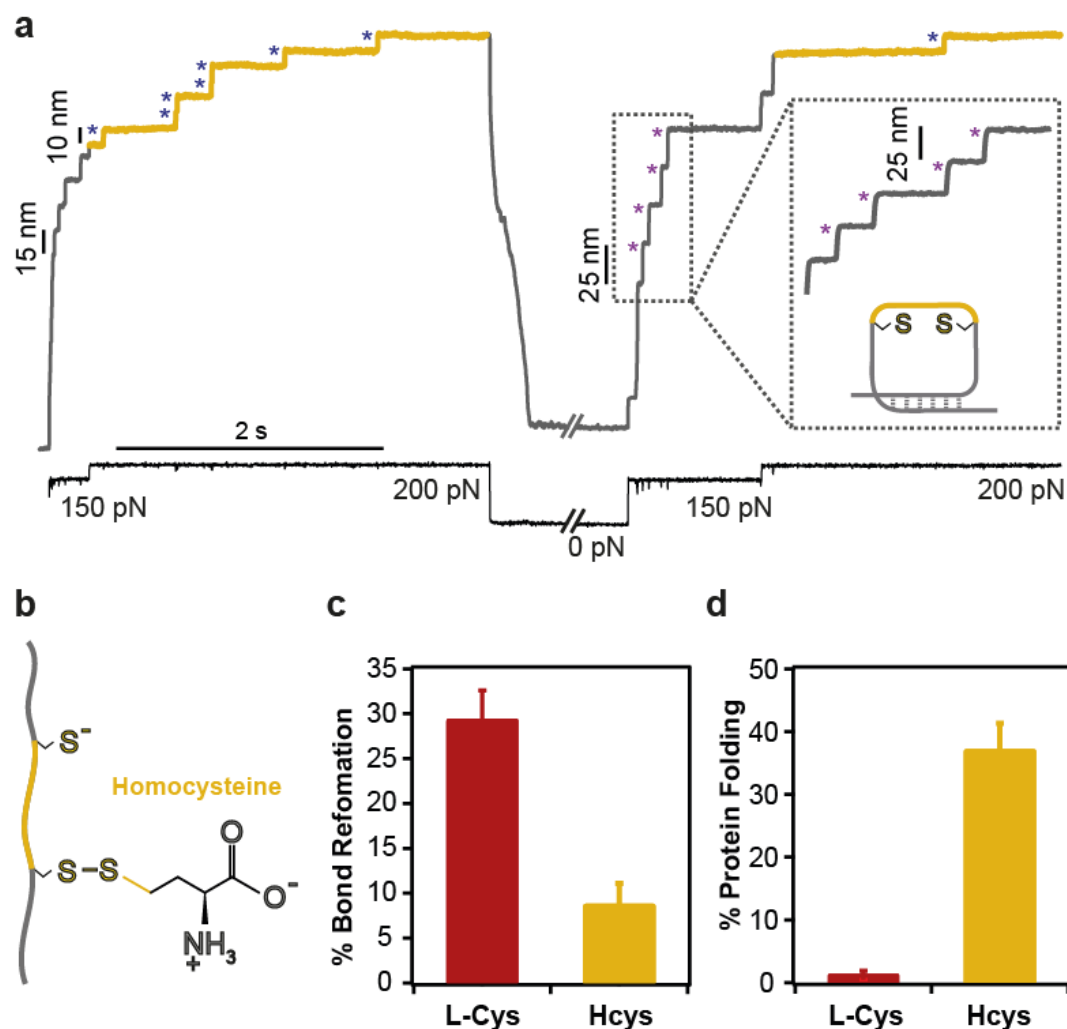


Figure 5.2. DL-homocysteine prevents successful non-enzymatic oxidative folding. (a) The oxidative folding trajectory in the presence of homocysteine (111 mM, pH = 7.6) exhibits a test pulse that is composed of steps of ~25 nm (marking the refolding of a reduced protein form, purple asterisks) followed by a vanishingly small population of ~10 nm steps, fingerprinting disulfide bond reformation (blue asterisks). (b) The asymmetric mixed disulfide between a native cysteine and the attacking homocysteine, which only differs from L-cysteine in the presence of a methylene group (yellow) next to the nucleophilic thiol plays a crucial role in the outcome of the folding reaction. (c) While the percentage of disulfide bond reformation in the presence of homocysteine (yellow) is negligible compared to that of L-cysteine (red), (d) the situation is reversed for the population of folded and reduced proteins (~25 nm steps), which is largely predominant for homocysteine (yellow) when compared to the residual fraction measured for L-cysteine (red) (error bar $\pm$ s.d.).

THE MIXED DISULFIDE PLAYS A KEY ROLE IN DISULFIDE FORMATION

It is very plausible that the mechanism underpinning the non-enzymatic reformation of a protein disulfide bond observed for L-cysteine and also—to a much lower extent—for homocysteine, involves the reactivity of the mixed disulfide intermediate. Mass spectrometry (MS) measurements on the I27_{E24C-K55C} monomer detected with high evidence the post-translational modification of both 24 and 55 cysteines by both L-cysteine and homocysteine (Appendix III Fig. 1, Appendix III Fig. 2 and Appendix III Table 1) (*All mass spectrometry measurements were conducted by Steven Lynham, King's College London*), which unambiguously confirmed the presence of the mixed disulfide. Noteworthy, given the different degree of solvent exposure of the protein disulfide bond obtained in MS and mechanical measurements, the yield of protein modification is not comparable in both experimental approaches. To prove the necessary participation of the mixed disulfide adduct in the dynamics of disulfide bond reoxidation, we conducted analogous experiments on an equivalent (I27_{G32C-A75C})₈ polyprotein using the phosphorus-based TCEP nucleophile (we chose the I27_{G32C-A75C} mutant, harbouring the buried disulfide bond in a different position within the protein structure, because TCEP can partially reduce the disulfide bond of the (I27_{E24C-K55C})₈ polyprotein in the absence of force¹⁶³). In the case of the P-based nucleophile TCEP, the force-induced S_N2 disulfide cleavage¹⁷⁰ obviously does not yield a stable mixed disulfide, but rather a highly reactive mixed S-P thioalkoxyphosphonium cation intermediate³²⁸ (Appendix III Fig. 3), which can be easily displaced by water molecules to render a completely unfolded protein with two reduced thiolates (Fig. 5.3 a). Analysis of the test pulse in our refolding trajectories (Fig. 5.3 b) reveals the absolute absence of ~10 nm steps, indicating that no disulfide bond is present in the otherwise successfully refolded protein (confirmed by

the presence of ~25 nm steps, Fig. 5.3 c). These experiments demonstrated the crucial implication of the mixed disulfide species in the thiol-mediated mechanism of oxidative folding.

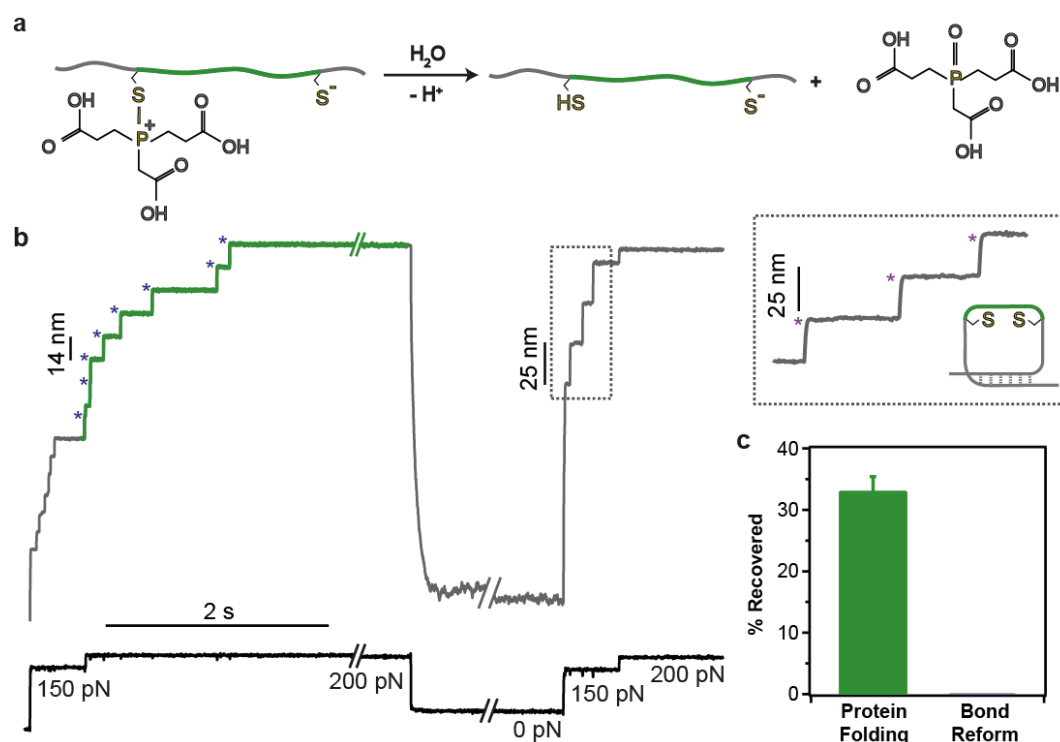


Figure 5.3 The absence of the mixed disulfide results in a refolded protein void of the disulfide bond. (a) When a phosphorous nucleophile such as TCEP is used to induce the force-activated cleavage of the disulfide bond on a structurally analogous (I27_{G32C}-A75C)₈ polyprotein (4mM, pH 7.2), a highly unstable mixed S-P thioalkoxyphosphonium cation intermediate is formed, which can be rapidly resolved by water solvolysis. TCEP-mediated disulfide bond rupture results in a stretched protein harbouring two reduced cysteines. (b) The test pulse in the folding trajectories is composed of steps of ~25 nm (purple asterisks), fingerprinting a natively folded protein devoid of the rigid disulfide bond. (c) Crucially, the absence of a mixed disulfide precludes the reformation of the native disulfide bond in the context of non-enzymatic oxidative folding (error bar $\pm$ s.d.).

HOMOCYSTEINE EXHIBITS AN ANOMALOUS HIGH pK_a

To elucidate the molecular origin of the markedly different behaviour observed for L-cysteine and homocysteine, we investigated how the physicochemical properties of both small thiols could determine their reactivity within the context of a nucleophilic substitution (S_N2) reaction. It is well established that a charged thiolate will be a better nucleophile than a protonated thiol³²⁹. Hence, there is a general direct connection between the goodness of a particular nucleophile/leaving group and its acid-base

properties. Although basicity is a thermodynamic property and nucleophilicity is a kinetic phenomenon, a good correlation between both properties has been observed, generally following the rule of thumb stating that good leaving groups are the conjugate bases of strong acids³³⁰. While L-cysteine is reported to exhibit a $pK_a = 8.15^{331}$ - 8.3^{332} , there is a surprising disagreement in the literature regarding the pK_a value for homocysteine, for which two largely disparate values ($pK_a = 8.7^{333}$ and $pK_a = 10^{332, 334, 335}$) are reported. To resolve such a dichotomy, we independently measured the pK_a of both compounds by monitoring the free thiol absorbance at $\lambda = 240\text{ nm}^{336, 337}$ as a function of pH for both L-cysteine and homocysteine compounds (Fig. 5.4 a) (*All pK_a determination measurements were conducted by Marc Mora, King's College London*). While an expected $pK_a = 8.4$ was measured for L-cysteine, the pK_a of homocysteine was found to be significantly higher ($pK_a = 9.9$) in our experimental conditions, supporting previous findings³³². Crucially, our homocysteine measurements required reducing conditions to attenuate its fast dimerization, especially under basic conditions (Appendix III Fig. 4). The dissimilar pK_a value found for both isolated compounds in solution is likely to account for their different reactivity when selectively bound to a protein via a protein post-translational modification.

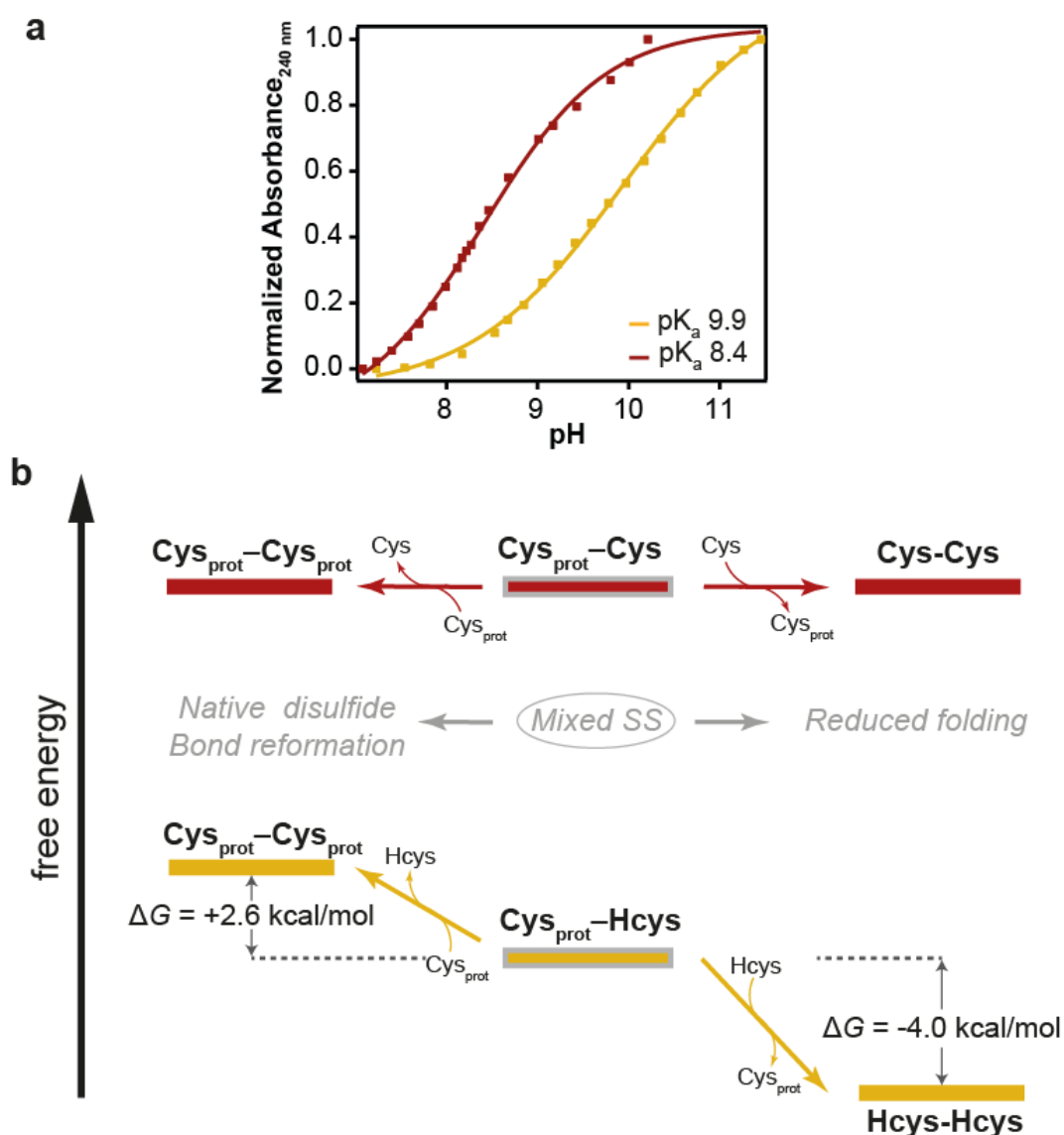


Figure 5.4 The surprisingly high pK_a of Homocysteine leads to a thermodynamically stable asymmetric mixed-disulfide. (a) Free homocysteine exhibits a surprisingly higher pK_a = 9.9 (red) than the homologous L-cysteine (pK_a = 8.4, yellow) revealed by absorbance measurement at λ = 240 nm in solution. (b) DFT calculations predict the high stability of the Cys-Hcys mixed disulfide, which cannot be attacked by a neighboring protein free thiol (ΔG = +2.6 kcal mol⁻¹) to reform the native disulfide.

DFT CALCULATIONS PROVIDE A THERMODYNAMIC EXPLANATION

To obtain a more precise understanding of the thermodynamically allowed reactivity between the different thiol species involved in our experiments (namely (i) the initial disulfide bond between two structurally close cysteines, (ii) the newly mixed disulfide bond species; (iii) the resulting free cysteine thiol

after disulfide bond cleavage and (iv) the free small thiol nucleophile species in solution), we performed DFT calculations on model thiol and disulfide species using the expanded 6-311+G(d,p) basis set B3LYP functional, which has been used in the context of thiol-disulfide exchange³³⁸ (Table 1) (*All DFT calculations were performed by Guillaume Stirnemann, CRNS – Institut de Biologie Physico-Chimique, Paris*). Noteworthy, in this simplified approach any free-energy contributions arising from a protein conformational change upon formation of a native or mixed disulfide are neglected (only the free-energy associated with an isolated disulfide bond rupture is computed).

Table 1. DFT free-energy calculations corresponding to the disulfide-exchange reactions induced by the studied low molecular weight thiols.

	Reaction	ΔG [kcal mol ⁻¹]
1	$\text{Cys-Cys} + \text{Cys}^- \rightarrow \text{Cys-Cys} + \text{Cys}^-$	0.0
2	$\text{Cys-Hcys} + \text{Hcys}^- \rightarrow \text{Hcys-Hcys} + \text{Cys}^-$	-4.0
3	$\text{Cys-Hcys} + \text{Cys}^- \rightarrow \text{Cys-Cys} + \text{Hcys}^-$	+2.6
4	$\text{Cys-NAC}^- + \text{Cys}^- \rightarrow \text{Cys-Cys} + \text{NAC}^{2-}$	+7.0
5	$\text{Cys-NAC}^- + \text{NAC}^{2-} \rightarrow \text{NAC-NAC}^{2-} + \text{Cys}^-$	-7.7

The attack of the Cys_{prot}-Cys mixed disulfide bond by a solution L-cysteine is, from a free-energy perspective, an equilibrium process with no net energy contribution ($\Delta G = 0$ kcal mol⁻¹ reaction 1, Table 1), whereas the rupture of the mixed non-equivalent disulfide between Cys_{prot}-Hcys by free Hcys in solution (reaction 2, Table 1) is favourable ($\Delta G = -4$ kcal mol⁻¹). By contrast, the attack of the same mixed disulfide (Cys_{prot}-Hcys) by the freed protein cysteine (reaction 3, Table 1) is not favourable ($\Delta G = +2.6$ kcal mol⁻¹). Hence, these results can rationalise a few key non-expected experimental observations, namely that: (i) the symmetric mixed disulfide formed by 2 equivalent cysteines can be attacked by the freed protein thiolate when in close physical proximity to reform a disulfide bond (Fig. 5.4 b); by contrast, (ii) the asymmetric Cys_{prot}-Hcys is not quantitatively attacked by the free thiol protein

cysteine to reform the previously cleaved disulfide bond but (iii) can be attacked by free homocysteine molecules in solution, thus resolving the mixed disulfide and leaving the protein with two reduced thiolates able to successfully refold in the absence of a formed disulfide bond giving rise to the ~25 nm steps.

THE DISTINCT NUCLEOPHILIC POWER OF L-CYS AND HCYS

While free-energy DFT calculations agree with our (thermodynamic) acid-base measurements, full understanding of the S_N2 nucleophilic substitution reaction would also require kinetic considerations. At the fundamental level, thiol-disulfide exchange is a classical bimolecular nucleophilic exchange S_N2 reaction that occurs through a linear (S-S-S) Walden transition state complex^{169, 329, 338}. Crucially, mechanical force accelerates disulfide cleavage by lowering the kinetic barrier^{162, 166, 170}. To compare the nucleophilic power of both small thiols, we measured the rate of disulfide reduction at the same stretching force ($F = 200$ pN) for two solutions containing the same concentration of active, deprotonated thiolate species at the same pH (Fig. 5.5 a). Our results demonstrate that the rate of disulfide reduction is clearly higher for homocysteine ($r = 1.11$ s⁻¹) than for cysteine ($r = 0.08$ s⁻¹), thus highlighting the better nucleophilic nature of homocysteine (Fig. 5.5 b). From the computational perspective, a very crude estimate of the relative nucleophilic character of both cysteine and homocysteine compounds can be obtained by comparing the energies of the HOMO orbitals of the isolated small compound. Our calculations revealed (Appendix III Fig. 5) that the HOMO energy for Hcys is higher than that of Cys, suggesting an enhanced reactivity toward electrophiles. Similarly, the effective electrostatic negative charge localised on the attacking sulfur atom in homocysteine is more negative (-0.81 e) than in L-cysteine (-0.77 e). Altogether, these results suggest that Hcys is a better nucleophile (and thus, a worse leaving group) than Cys, qualitatively agreeing

with the experimental results and the DFT free-energy calculations (Table 1).

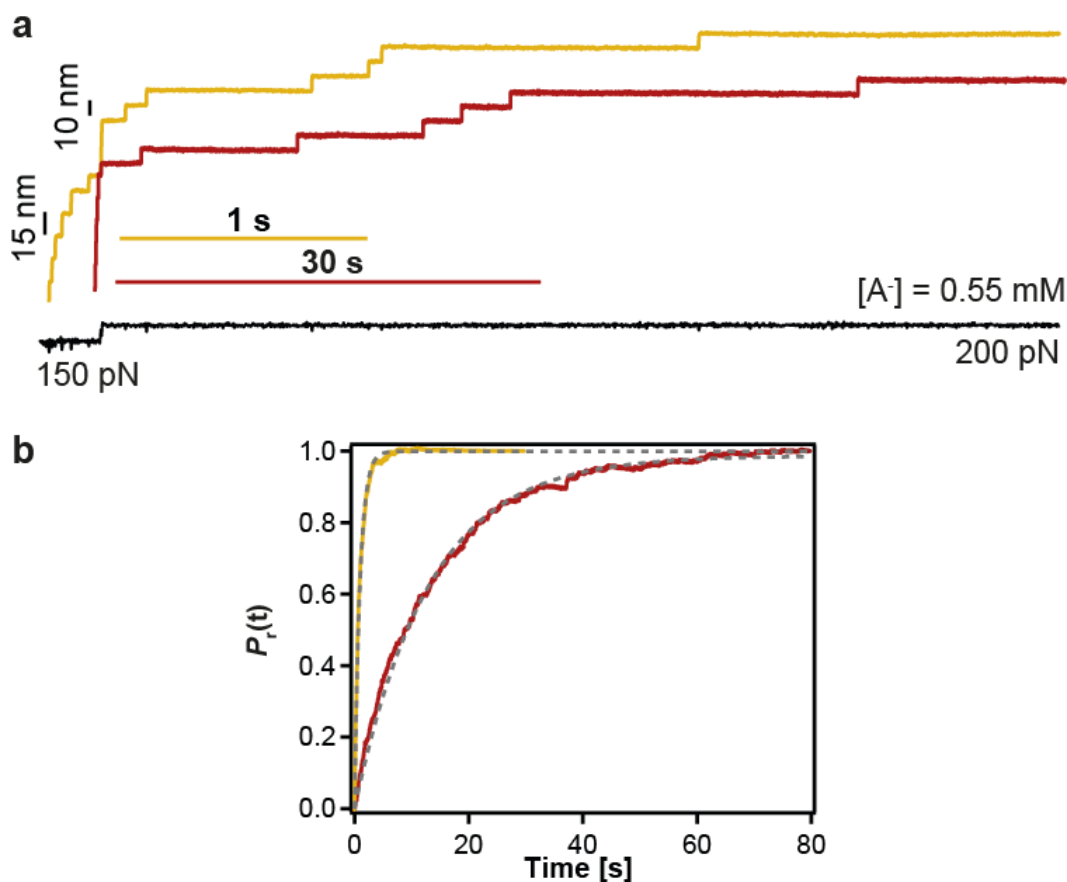


Figure 5.5 The kinetics of disulfide reduction is largely accelerated by homocysteine due to its high nucleophilic power. (a) A two-pulse force protocol measures the kinetics of disulfide reduction when the protein is exposed to the same concentration of active deprotonated species of L-cysteine (red) and homocysteine (yellow). The time-course of disulfide bond reduction is dramatically longer for L-cysteine as illustrated by the drastically different scale bars **(b)** The rate of disulfide bond reduction at $F = 200 \text{ pN}$ is ~14-fold faster for homocysteine ($r = 1.11 \text{ s}^{-1}$) than for L-Cys ($r = 0.08 \text{ s}^{-1}$) under the same experimental conditions, demonstrating that homocysteine is a better nucleophile than L-Cysteine.

Noteworthy, precise quantification of both thermodynamic and kinetic parameters defining the system would require taking into account the subtle yet important effects of the larger and complex protein environment, and also the non-trivial, catalytic effect that the pulling force has on the reaction kinetics^{162, 166}. Further free-energy calculations at the QM/MM level with explicit solvent, hopefully coupled with a force perturbation^{172, 271, 339} (which goes beyond the scope of the current work), will help elucidate the finer parameters (e.g. solvation, bond orientation³³⁹) regarding the mechanistic dynamics of the process. In any case, our simple estimates regarding

thermodynamic and kinetic considerations of the isolated reactants provide a good general qualitative description of the experimental results, describing a localised chemical reaction occurring in the core of a much complex protein environment.

THE EXPERIMENTAL APPROACH PREDICTS THE NATURE OF NAC

To further test if these considerations are sufficient predictors of disulfide reactivity and thus, of the final outcome of the oxidative folding reaction, we conducted similar nanomechanical experiments using *N*-acetylcysteine (NAC) as a nucleophile (Fig. 5.6 a, b). NAC is a drug derivative of cysteine with an acetyl group attached to its nitrogen (Fig. 5.6 a), which has been used in clinical practice for decades as a mucolytic agent and also employed in the treatment of numerous disorders including paracetamol intoxication and vascular and cardiac injury³⁴⁰. The uniqueness in the action of NAC stems from its particular physicochemical properties, which render it an optimal reducing agent of protein disulfide bonds³⁴¹. Measurement of the pK_a for NAC through our absorbance experiments yielded a value $pK_a = 9.6$ (Fig. 5.6 c). Similarly, DFT calculations predicted that, from a thermodynamics perspective, the reaction whereby a free protein cysteine attacks a NAC-Cys_{prot} mixed disulfide bond is energetically impaired ($\Delta G = +7 \text{ kcal mol}^{-1}$, reaction 4, Table 1). By contrast, the NAC-Cys_{prot} mixed disulfide can be attacked by free solution NAC molecules ($\Delta G = -7.7 \text{ kcal mol}^{-1}$, reaction 5, Table 1). Moreover, HOMO and electrostatic calculations (Appendix III Fig. 5) highlight the high nucleophilic power of NAC with respect to L-cysteine. Hence, from the physicochemical viewpoint, NAC is closely related to Hcys. In effect, our single molecule folding trajectories in the presence of NAC (Fig. 5.6 b) certify that while the protein can fold into its reduced form (Fig. 5.6 d), its ability to effectively reform the previously reduced disulfide bond is vanishingly small (showing

an even smaller percentage than homocysteine, Fig. 5.6 e). The presence of the NAC-protein mixed disulfide was further certified through mass spectrometry analysis (Appendix III Fig. 6).

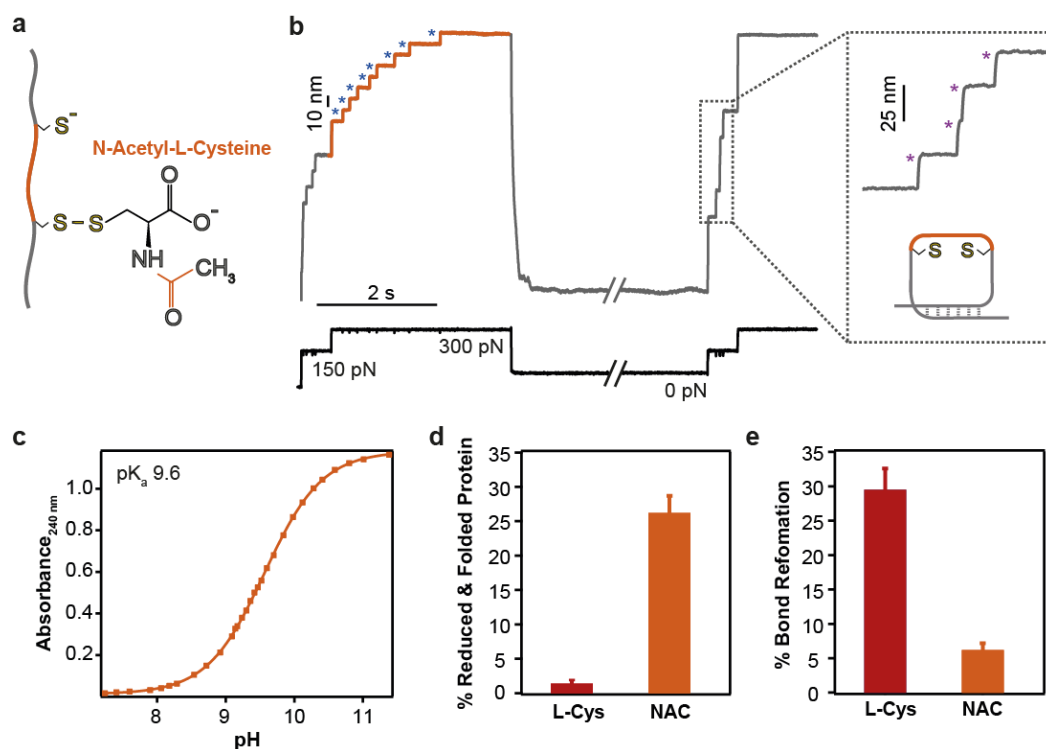


Figure 5.6 N-acetylcysteine creates a stable protein mixed-disulfide that largely blocks oxidative folding. (a) The force-induced cleavage of a protein disulfide bond results in an asymmetrical mixed disulfide intermediate of a protein cysteine with NAC, a cysteine derivative with an acetyl group (red) attached to its nitrogen. (b) The individual single molecule trajectories track the success of the oxidative folding assay. The initial pulse results in the unfolding of the (I27_{E24C}-K55C)₈ polyprotein (15 nm steps, grey) followed by the force-induced cleavage of the solvent-exposed disulfide bonds, marked by the 10 nm step-wise increase of the protein length (blue asterisks). By contrast, the test pulse is mostly composed of 25nm steps (purple asterisks), which hallmark that the protein has successfully refolded into a reduced conformation that is void of the native disulfide bond. (c) pK_a determination through absorbance measurements at $\lambda = 240$ nm yields a pK_a = 9.6. These results demonstrate that NAC is a weak acid, and hence its conjugated base will be a poor leaving group, in agreement with the high stability of the NAC-mediated mixed disulfide. (d) While the % of successful folding to a reduced conformation is largely prevalent in the presence of NAC, (e) the % of disulfide bond reformation is vanishingly small for NAC as compared to the structurally akin L-cysteine (error bar $\pm$ s.d).

5.3 DISCUSSION

Detection of chemical intermediates is the Holy Grail in the elucidation of reaction mechanisms. However, due to their ephemeral nature, their capture is frequently challenging. An extra layer of complexity is added when the chemical reactions occur within the core of an individual protein —giving rise to the generally called post-translational modifications— mainly because the reactivity of the newly created intermediate adduct is intrinsically coupled with the conformational motions of the protein. Hence, direct observation of the chemical evolution of a protein-based chemical reaction and its crucial effect on the folding fate of a protein remains largely elusive.

Due to the highly polarizable sulfur atom, the amino acid cysteine is a particular target for numerous chemical modifications in the cellular milieu³⁴², of which the most commonly known is the formation of disulfide bonds. Such thiol-disulfide exchange reactions have crucial roles in biology, both stabilizing structural protein motifs and also as essential regulators of the dynamic functional properties of many enzymes, including most oxidoreductases³⁰⁷. Albeit perhaps less studied, another important natural example of disulfide bond formation involves small plasma thiols. Indeed, an increase in plasma disulfides, including low-molecular-mass protein-mixed disulfides, has been postulated as a biomarker of oxidative stress and aging. In particular, direct correlations were found between the disulfide forms of Cys and Hcys and age^{317, 318, 319}. Several pathologies, including cognitive impairment, osteoporosis, eye's disorders, complications in pregnancy, reduced skeletal muscle function and integrity and cardiovascular disease have been directly correlated with high homocysteine levels³³². Hence, it is enticing to speculate about the presence of a set of distinct chemical properties of Hcys that underlie its ability to dimerize and to form stable disulfide bonds with protein cysteine residues that impair or alter normal protein function.

Several proteins with distinct biological function such as albumin³³¹, fibronectin, fibrillin, tropoelastin³²⁷, transthyretin, alpha-crystallin³⁴³ and metallothionein have been identified as natural Hcys targets³³². While a few pioneering studies postulated a molecular-targeting hypothesis that link the S-homocysteinylation post-translational modification with protein function^{332, 334}, direct experimental evidence remained elusive.

Our single-molecule approach allows independent measurement of the protein folding and disulfide bond formation processes, enabling us to directly visualize the molecular events defining the complex kinetics of the oxidative folding mechanism and the effect that closely related post-translational modifications have on the protein's mechanical function. Here we present a chemical strategy to promote successful non-enzymatic oxidative protein folding that is based on the high reactivity of a symmetric Cys_{prot}-Cys mixed disulfide obtained through protein S-cysteinylation. In this scenario, being in energetic equilibrium with a neighbouring protein thiol, the shuffling between three equivalent cysteines when in close spatial proximity allows disulfide isomerization¹⁸⁷, eventually leading to the spontaneous reformation of the native disulfide bond. By contrast, homocysteine forms a highly stable asymmetrical disulfide bond with a protein cysteine that can hardly be re-attacked by the previously freed neighbour thiol. We hypothesise that the high propensity of Hcys to form stable mixed disulfides with protein cysteines can be a generic mechanism that underlies the lack of protein functionality upon protein S-homocysteinylation. Similarly, the high affinity of NAC (and in general, all the tested small thiols) to create disulfide bonds might explain its important role in protein protection under oxidative stress³⁴⁴, avoiding irreversible thiol overoxidation to sulfinic and sulfonic acid.

A particular characteristic of our experiments is that chemical modifications occur on cryptic cysteines, which only become solvent-exposed after

mechanical unfolding of the protein. In this sense, the outcome of the folding process is not only limited by the chemical reactivity of the protein, but also by protein conformation, which is precisely modulated by mechanical force. Hence, our findings suggest a combined mechanochemical scheme that reveals how the distinct physicochemical properties and reactivity of closely related small thiols dictates the folding fate of the protein, giving rise to two successfully refolded forms, yet with very distinct mechanical properties (Fig 5.7 a). In our approach, the native disulfide bond is first reduced by the attacking nucleophile thiol upon application of force. While the redox potential is suggested to be slightly modulated by the pulling force³³⁹, force-activated reduction of a protein disulfide bond is mostly under kinetic control¹⁶², resulting in a protein stretched conformation containing a completely solvent-exposed mixed disulfide which is *a priori* susceptible to be attacked by the thiol nucleophiles present in solution.

Given that in the presence of TCEP (where no mixed-disulfide bond can occur) the protein successfully refolds (certified by the presence of 25 nm steps in the test pulse), we speculate that, in the presence of L-Cys, Hcys and NAC, the observed 25 nm steps (especially obvious with the experiments with Hcys and NAC) correspond to a scenario whereby free thiolates in solution are able to attack the mixed disulfide conformation (namely Cys-Cys_{Sprot}, Hcys-Cys_{Sprot} and NAC-Cys_{Sprot}), therefore resulting in a protein containing two reduced thiolates that are able to properly refold, void of the disulfide bond (Fig. 5.7 b). Indeed, DFT calculations show that the attack of solution free Hcys to the solvent-exposed, mixed asymmetrical Hcys-Cys mixed disulfide is thermodynamically very favoured ($\Delta G = -4.0 \text{ kcal mol}^{-1}$, reaction 2, Table 1), and it is very likely that this equilibrium is further displaced due to the high insolubility of the Hcys-Hcys homodimer (additional small molecule MS experiments reveal that monomeric Hcys in solution undergoes oxygen-

mediated oxidation to its dimeric, insoluble form³²⁶ within the timescale of ~1 hour, Appendix III Fig. 7). Similarly, the attack of NAC free molecules in solution to the NAC-Cys_{prot} heterodimer is even more favourable from the thermodynamic perspective ($\Delta G = -7.7$ kcal mol⁻¹, reaction 5, Table 1). These free-energy considerations are in very good agreement with our experimental observations, showing that while the percentage of reduced folding is vanishingly small for L-Cys (Fig. 5.2 d), it is instead predominant for both Hcys (Fig. 5.2 d) and also for NAC (Fig. 5.6 d). Further experiments where the time the reduced protein is left extended and exposed to NAC molecules in the solvent (Appendix III Fig. 8) confirmed that the percentage of reduced folding (requiring a solution NAC molecule attacking the asymmetric disulfide bond) remained invariable over time, suggesting that the overall reactivity process seems to occur under thermodynamic control, probably reflecting the simplicity of the reaction and the lack of conformational and steric hindrance.

Supported by recent findings using a cysteine-defective thioredoxin mutant¹⁸⁸ and reduced glutathione³⁰³, we speculate that the presence of the mixed disulfide blocks protein folding (giving rise to a compliant protein form), and that it needs to be resolved, either through the attack of a neighboring protein cysteine to reform the native disulfide bond after protein collapse or via reaction with thiols free in solution. However, we cannot completely rule out at this stage the possibility that the protein is able to refold around its mixed-disulfide and regain mechanical stability.

Upon withdrawal of the applied force, the protein collapses into a molten globule conformation³⁴⁵ where reactive cysteines are physically close and sterically buried from the protein environment. In the case of L-cysteine, disulfide bond isomerization¹⁸⁷ allows the reformation of the native disulfide bond, giving rise to a rigid protein form (Fig. 5.7 b). By contrast, in the case of

Hcys and NAC where the mixed disulfide bond has been resolved by reacting with free solution nucleophiles while the protein is stretched under force, quenching the force results in a refolded and reduced protein form that displays a mechanical stability that is intermediate between the mechanically stiff oxidised protein containing the disulfide bond, and the mechanically compliant misfolded protein that harbours a stable, non-resolved mixed disulfide species (Fig. 5.7 b).

The measurements presented here are conducted on an Ig domain of cardiac titin, a protein that carries an important mechanical role in vivo. A key requirement of elastomeric proteins, always subject to multiple stretching-relaxation cycles, is their reversible functionality and reliability over time³⁴⁶. Our approach provides an extra level of understanding on the molecular mechanisms of disulfide bond reformation under force, and its direct link to mechanical folding. Crucially, our measurements reflect the largely distinct role of the closely related S-cysteinylation and S-homocysteinylation on protein elasticity, and add to the emerging research field that highlights the importance of force-induced exposure of cryptic sites to regulate protein nanomechanics, both via post-translational modifications, such as S-glutathionylation³⁰³ and S-sulfenylation⁴⁴, or through the unravelling of metal binding sites³⁷. More generally, our proof-of-principle approach demonstrates the use of highly localized chemical reactivity—which can be potentially expanded to other distinct chemical mechanisms—to rationally tailor protein elasticity, and offer a new approach to engineer reversible redox-responsive protein-based biomaterials with adaptable mechanical properties.

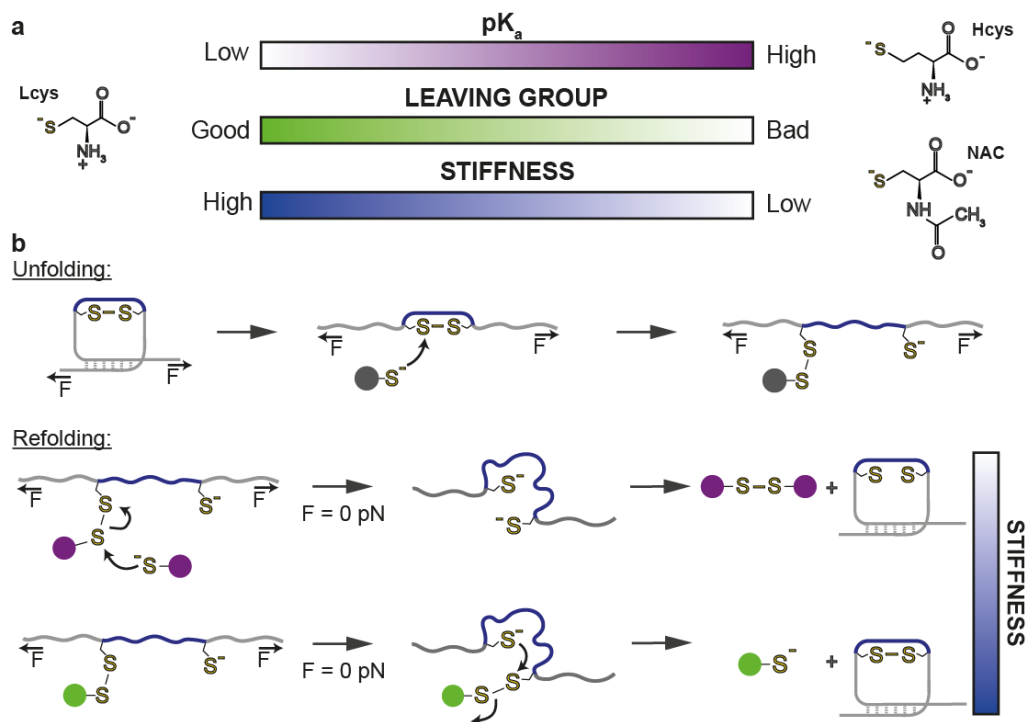


Figure 5.7 Protein reactivity induced by S-thiolation regulates protein nanomechanics. (a) The physicochemical properties of the different small thiols govern the life-time of the mixed disulfide, which largely controls the outcome of the oxidative folding reaction and determines protein nanomechanics. (b) Schematics of the reactivity of the S-thiolated protein under force. On the application of mechanical force, the different small thiolates rupture a protein disulfide bond, creating a mixed disulfide and a reduced protein thiolate. In such extended and reduced protein conformation, the protein thiol that is bound to a strong nucleophile acts as the leaving group when attacked by free thiols in solution (purple spheres, e.g. homocysteine and NAC) resulting in two reduced protein cysteines. In the absence of force ($F = 0$ pN), the reduced protein can recover a more compliant folded form that is devoid of the native disulfide bond. By contrast, when a softer nucleophile is used (green sphere, e.g. L-cysteine), the highly reactive symmetric mixed disulfide is in equilibrium with the free protein cysteine thiol. In this scenario, disulfide bond isomerization occurs, resulting in a stiff folded protein characterized by the reformation of the native disulfide bond.

6. CONCLUSIONS & FUTURE DIRECTIONS

6.1 CONCLUSIONS

I will hereby proceed to summarise the main conclusions of each experimental chapter of the thesis.

CHAPTER 3 CONCLUSIONS

- Understanding the directionality and sequence of protein unfolding is critical to reveal the underlying folding free energy landscape. This process is complicated in metalloproteins, where a metal cofactor participates in the native fold. In chapter 3, I used single molecule force spectroscopy AFM combined with protein engineering and MD simulations to study the unfolding pathways of the blue copper proteins azurin and plastocyanin.
 - The unfolding pathway of azurin reveals that after the main unfolding barrier, there are two mechanically labile intermediates. The mechanical origin of these on-pathway unfolding intermediates can be assigned to the rupture of a single bond that coordinates the copper into the protein structure. These experiments revealed that the organometallic bond was disrupted at forces of ~ 45 pN at a pulling velocity of 400 nm.s^{-1} , which is surprisingly low given the covalent nature of this bond.
 - Utilizing the nanomechanical properties of the copper centre as a structurally-embedded molecular reporter, we demonstrate that both proteins azurin and plastocyanin unfold via two independent, competing pathways – one unfolding from the C-terminus and the other from the N-terminus. This finding was supported by protein mutagenesis and steered molecular dynamics simulations. These results reveal a novel kinetic partitioning scenario whereby the protein can stochastically unfold through two distinct main transition states placed at each of the termini that dictate the directionality of unfolding.

CHAPTER 4 CONCLUSIONS

- The post-translational modification *S*-sulfenylation functions as a key sensor of oxidative stress. However, the dynamics of sulfenic acid in proteins remains largely elusive due to its highly reactive nature. In chapter 6, I used single-molecule force-clamp spectroscopy to directly capture the reactivity of an individual sulfenic acid embedded within the core of a single immunoglobulin domain of titin.
 - To create a stable sulfenic acid species, I ruptured an existing protein disulfide bond using OH^- , thus creating a sulfenic acid moiety and a free protein thiol. A force-quench protocol captured the unambiguous reformation of the disulfide bonds upon protein *S*-sulfenylation, confirming that sulfenic acid is the first step of a non-enzymatic route that gives rise to successful oxidative folding in the absence of any enzyme.
 - The sulfenic acid moiety can undergo further chemical modification if exposed to the protein solution for a long time - ie in a conformational dependent manner- , with profound knock-on implications for the folding fate of the protein. The evolution of sulfenic acid to sulfinic acid, sulfonic acid or an aldehyde species leads to complete protein misfolding and a resulting conformation void of mechanical stability.
 - The most prominent cellular route towards the formation of sulfenic acid is via the presence of reactive oxygen species. To determine whether the results with high pH recapitulate this redox route, I performed mechanical folding experiments in the presence of H_2O_2 on the 27th Ig domain of titin, which contains two free thiols that do not spontaneously form a disulfide bond. Albeit less controlled, these experiments clearly demonstrated a reduced mechanical extensibility after the exposure of the cryptic cysteines, indicating the formation of a rigid disulfide bond. This calls for further experiments to determine

whether transient increases in reactive oxygen species within the cell can trigger the formation of both native and non-native disulfide bonds in proteins, especially in cases where the protein experiences mechanical force, such as titin.

- Altogether, the results in chapter 4 demonstrate that sulfenic acid is a crucial short-lived chemical moiety that dictates the protein's folding fate.

CHAPTER 5 CONCLUSIONS

- The mechanical properties of elastomeric proteins dictate the elasticity of a multitude of tissues. A natural strategy to modulate protein extensibility lies in the presence of covalent disulfide bonds, which enhance protein stiffness. The prevalent *in-vivo* strategy to create disulfide bonds requires the assistance of dedicated enzymes. In chapter 5, I used single molecule force clamp spectroscopy to investigate an alternative chemical route to promote non-enzymatic oxidative protein folding via disulfide isomerization with naturally occurring low molecular weight thiols.
 - The mixed disulfide created by L-cysteine leads to the successful the non-enzymatic reformation of disulfide bonds. On the contrary, when a mixed disulfide is created by the chemically analogous homocysteine, the formation of disulfide bonds is dramatically impaired, and instead protein successfully refolds, yet into a reduced form harbouring intermediate mechanical properties.
 - This stark difference in the outcome of oxidative folding can be rationalised by looking at the physicochemical properties of the attacking thiols. Keeping in mind the rule of thumb that a good leaving group is the conjugate base of a strong acid, the high pK_a of homocysteine implies it is a very bad leaving group. Therefore upon protein S-homocysteinylation, the highly stable mixed disulfide

cannot be re-attacked by the remaining free protein thiol, thus leading to protein misfolding.

- Density functional theory calculations provide a thermodynamic perspective into the chemical reactivity. These calculations reveal that for the case of disulfide scission by cysteine, the newly generated mixed disulfide is in thermodynamic equilibrium and therefore there is no energy cost for the protein thiol to attack the disulfide and reform the bond. On the contrary, for homocysteine the most thermodynamically favourable outcome of the cysteine-homocysteine dimer is an attack from a free homocysteine in solution, thus leading to a protein void of a reformation mechanism.
- Altogether the results in chapter 5 demonstrate that subtle changes in the chemical structure of a transient mixed-disulfide intermediate adduct between a protein cysteine and an attacking low molecular-weight thiol have a dramatic effect on the protein's mechanical stability. This approach offers molecular insights on how chemical reactivity regulates protein elasticity.

6.2 FUTURE DIRECTIONS

The work presented throughout this thesis utilises the power of mechanical force to manipulate a single protein. In particular, the single molecule force clamp approach used in chapters 4 and 5 demonstrates that force can trigger the rupture of a single disulfide bond and consequently introduces a specific post-translational modification into a well-defined location within the protein structure. Furthermore, this force-clamp approach unambiguously captures how the PTM impacts the folding fate of the protein, with important implications for both the functional state and nanomechanical properties of the molecule. Altogether this work provides an experimental platform to gain insight into a number of fundamental mechanochemistry questions, expanding our understanding of the relationship between force and reactivity.

In particular, previous mechanochemical investigations have mainly focused on understanding how force modulates (i) the kinetics of the reactant-to-product conversion^{48, 167, 172, 347} or (ii) the outcome of the reaction, giving rise to novel products^{3, 4, 348}. This progress notwithstanding, the subsequent reversibility of such a mechanically induced reaction, especially in cases where a thermodynamically forbidden product has been created, has so far evaded experimental investigation. The solution of such a dichotomy calls for experimental approaches capable of applying a mechanical perturbation to a single bond while tracking both the forward and reverse reaction.

Force spectroscopy techniques, capable of applying a well-calibrated force to an individual tethered molecule, have granted access to the rupture and reformation of individual chemical bonds such as organo-metallic^{349, 350} and hydrogen bonds^{43, 117, 121, 351}, typically occurring within the context of a larger biomolecule such as a protein. Moreover, the disulfide bond approach used throughout this thesis, has afforded access to the rupture and reformation of a single disulfide bond upon a bimolecular reaction^{179, 187, 188, 352, 353}. In this

configuration, the disulfide bond is generally engineered into the core of a protein, such that when the protein is folded, the bond is buried and inaccessible to the surrounding solvent. Upon the application of mechanical force, the protein unfolds and exposes the previously cryptic disulfide bond. At this moment, the bond is aligned in the direction of the force and the clock begins for the start time of the reduction kinetics. In particular, in chapter 5 I have shown that disulfide scission by a nucleophile with sulfur as the central attacking atom, leads to the stoichiometric formation of a free protein-thiolate and a mixed disulfide between the remaining protein thiol and the attacker. Importantly, this experimental approach is useful at inducing a battery of well-defined post-translational modifications in the protein cysteines. In the absence of force, this mixed disulfide moiety is susceptible to an S_N2 nucleophilic attack by the free protein thiol, giving rise to successful non-enzymatic oxidative protein folding. This direct observation of disulfide reformation provides a platform to gain a comprehensive understanding of the relationship between the force-activated reactivity (kinetics), redox potential and stability (thermodynamics) and chemical reversibility.

We have first set out to investigate whether it is possible to mechanically induce the rupture of a disulfide that cannot occur under purely thermodynamic control. Initially, and as a first simple approximation, we have investigated the thermodynamic favourability of disulfide bond rupture with a range chemically similar sulfur-nucleophiles using the Ellman's assay. The Ellman's assay is typically used to quantify the number of free thiols in a solution, where a thiolate will cleave the oxidized 5,5'-dithiobis-2-nitrobenzoic acid (DTNB) compound producing 2-nitro-5-thiobenzoate (TNB^-) which is yellow in colour. Therefore a more intense yellow colour indicates a higher degree of solution thiols. Instead here we have tweaked this approach to test the thermodynamic favourability of disulfide rupture. Firstly, all 9 S-based

nucleophiles were prepared to the same active deprotonated concentration and then incubated for 5 minutes with the Ellman's reagent. A strong absorbance indicates a very thermodynamically favourable reaction, whereas little or no absorbance indicates that the sulfur compound was not successful at reducing the oxidized compound. This experiment revealed a strong correlation between the pKa of the sulfur and of the intensity of absorbance (Fig. 6.1 a). In particular, the compounds penicillamine and cysteine-methyl-ester show none, or a very weak, absorbance indicating that the reduction reaction did not proceed. Furthermore, density functional theory calculations on a disulfide bond formed between two cysteine residues provides insight into the Gibbs free energy of disulfide reduction, corroborating with these bulk spectroscopic results. Astonishingly, the compounds penicillamine and cysteine-methyl-ester have a positive ΔG for disulfide reduction, demonstrating this reaction is not permitted according to thermodynamics (Fig. 6.1 b). Altogether these findings give us a unique opportunity to understand whether mechanical force can induce a thermodynamically forbidden reaction within the core of a protein, and furthermore to understand whether this impacts on the subsequent reaction reversibility in the absence of force.

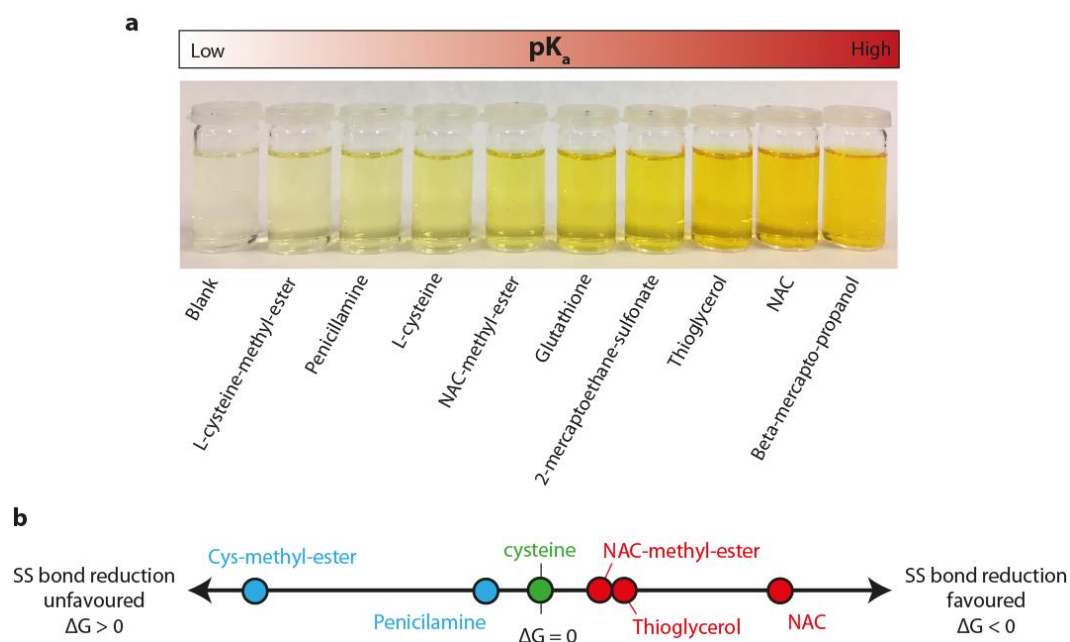


Figure 6.1. Testing the free energy of disulfide reduction for a number of chemically similar sulfur based nucleophiles. (a) The Ellman's assay provides insight into the thermodynamic favourability of disulfide reduction. The intense yellow colour is indicative a high population of cleaved DTNB, indicating that the reduction reaction is very favourable. In contrast, a weak yellow colour indicated that the majority of the DTNB has remained oxidized and thus the nucleophilic reduction is not favoured. **(b)** Density functional theory calculations on the free energy of the reduction reaction for a select number of nucleophiles corroborate with the results obtained with the Ellman's assay, and highlight that in the case of Penicillamine and cysteine-methyl-ester, the reduction of the disulfide is not thermodynamically permitted.

To test whether it is possible to induce the rupture of the disulfide bond with penicillamine and cysteine-methyl-ester, we performed single molecule force clamp spectroscopy experiments on the protein I27_{24/55} in the presence of these nucleophiles (Fig. 6.2). Using a two-pulse force protocol, an initial pulse (150 pN, 0.4 s) unfolds the protein in steps of 15 nm and exposes the disulfide to the reducing environment. A second high force pulse (350 – 450 pN, < 15 s) in conjunction with the nucleophile tests whether scission of the disulfide can be triggered by mechanical force. Indeed, these experiments reveal 10 nm steps in the high force pulse, fingerprinting the mechanically activated reduction of the disulfide bonds. Altogether, these results highlight that mechanical force can distort the energy landscape to overcome the thermodynamic impediment.

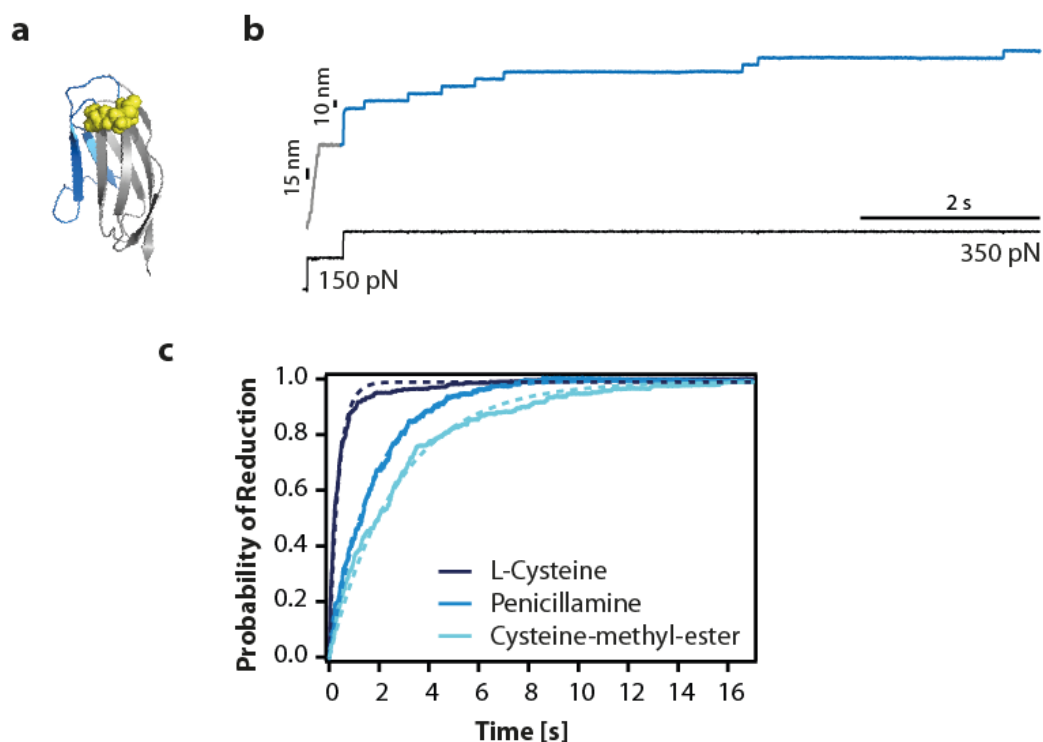


Figure 6.2. Mechanical force can induce thermodynamically forbidden reactions. (a) The mutated 27th Ig domain of titin contains a single cryptic disulfide bond between residues 24 and 55. (b) A two pulse force protocol tests whether mechanical force can trigger the scission of a disulfide bond that is thermodynamically forbidden. A first pulse unfolds the protein (grey) and exposes the bond to solution. A second pulse in the presence of penicillamine ($[s-] = 1\text{mM}$, $F = 350\text{ pN}$) triggers the rupture of each individual disulfide bond in the polypeptide chain, hallmarked by steps of 10 nm. (c) Similar experiments in the presence of cysteine-methyl-ester similarly show that force can trigger disulfide bond reduction.

Finally, to test the hypothesis that the reactivity of the forward reaction will influence the reversibility, we have performed five-pulse force protocols in the presence sulfur-nucleophiles that exhibit a wide variety of pK_a values, which we are using as a first approximation for the ΔG of reduction. This approach can unambiguously determine whether the disulfide bond has been reformed in the absence of force. In the case of 1-mercapto-2-propanol, which displays the highest pK_a value and the strongest absorbance in the Ellman's assay - indicating an extremely favourable forward reaction - the disulfide only reforms in ~1% of cases. In stark contrast, cysteine-methyl-ester, which was shown to only induce disulfide bond rupture when in conjunction with a mechanical perturbation, displays an exceptionally high reformation percentage ~40% (Fig. 6.3). The remaining S-thiols tested in these experiments

(Fig.6.3) exhibit pKa values lying within these two extremes, and remarkably all correlate with the expected trend of disulfide bond reformation. These preliminary findings demonstrate that there is a strong correlation between the favourability of the forward reaction and the subsequent reversibility.

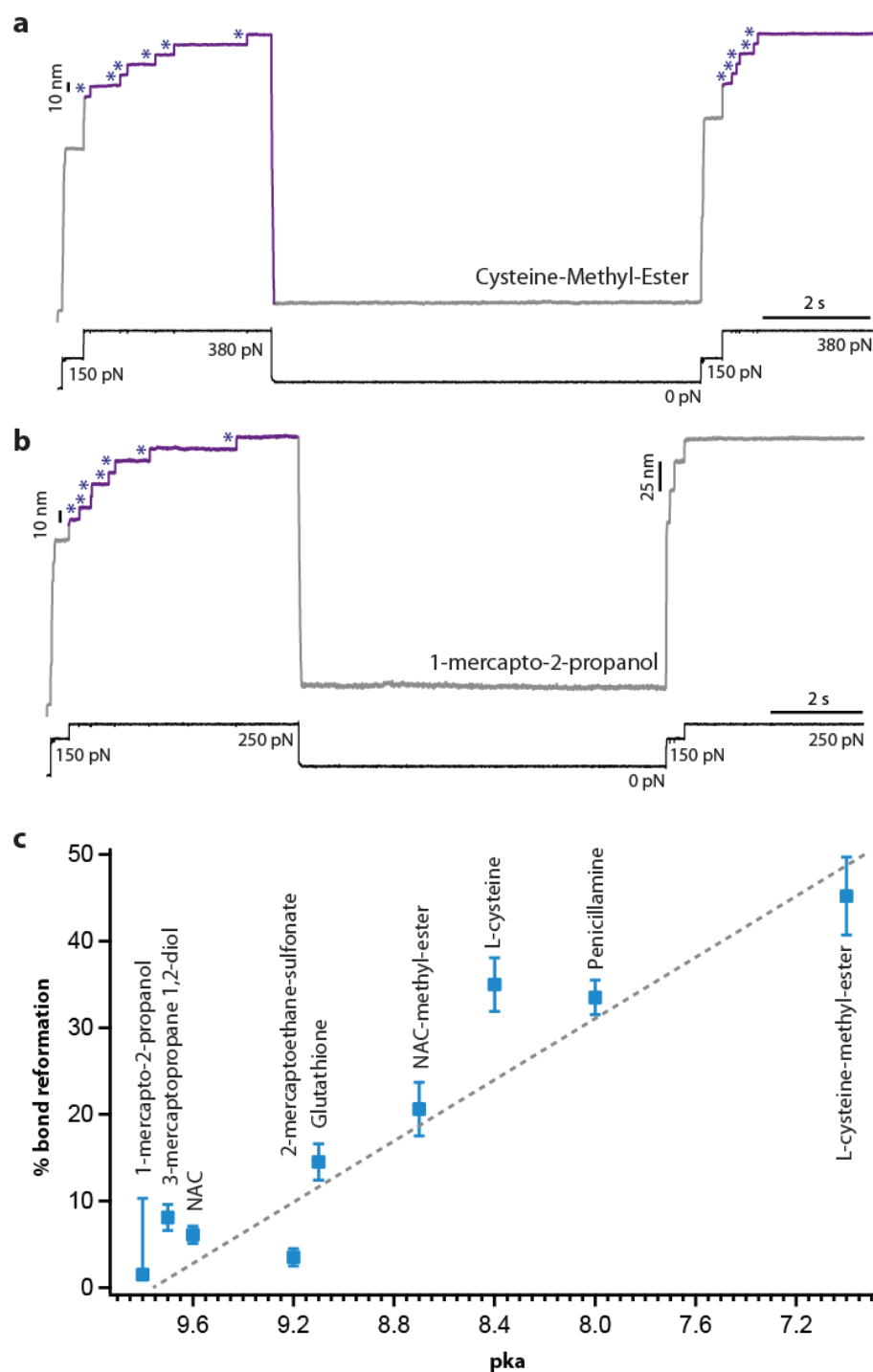


Figure 6.3. The thermodynamic favourability of disulfide reduction has profound implications on the reverse reformation process. (a) A five-pulse force protocol tests the outcome of non-enzymatic oxidative protein folding. An initial two pulses unfolds the protein and ruptures the disulfide bond. Quenching the force down to 0 pN collapses the protein and brings the mixed disulfide and the free protein thiol close in conformational space. The reapplication of force pulses, tests the reversibility of the reaction. In the case of cysteine-methyl-ester, there is a high yield of reformation. (b) Whereas in the case of 1-mercapto-2-propanol there is a vanishingly small population of disulfide reformation events. (c) The observed reaction reversibility scales with the pK_a of the attacking thiol of the forward reaction.

For the moment we have rationalized these surprising results using a simplistic 1-dimensional representation of the free-energy landscape of a reaction. If the energy of the products is lower than the reactants, the forward reaction will be thermodynamically favored. However, upon release of the force, which triggers the reverse reaction, the free energy difference to revert back to the original reactants will be positive and therefore disulfide reformation is extremely unlikely. In contrast, if the energy of the products is higher than the reactants, the forward reaction will not be favored and mechanical force is required to tilt the landscape, lowering the height of the energy barrier and catalyzing the reaction. In this case, when the force is released, the reverse reaction from the products to the reactants will be downhill with a $\Delta G < 0$ (Fig. 6.4), thus rendering the reaction highly reversible. The major limitation of this interpretation is that it does not account for any kinetic barrier in the reverse pathway. While the time-dependent reformation kinetics will provide insight into whether such a barrier exists, it is not straightforward to extract information regarding the barrier height.

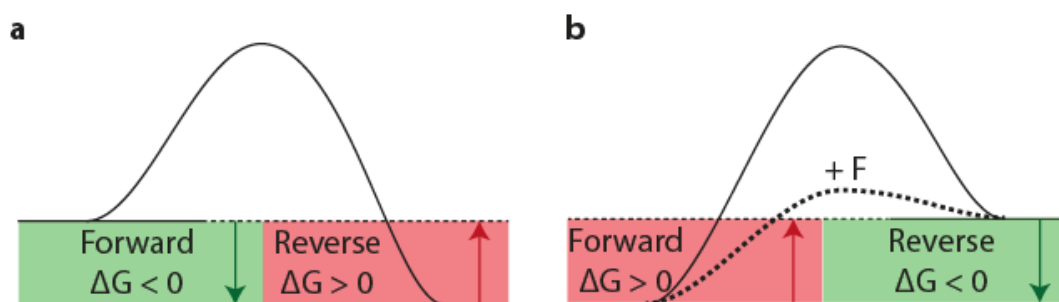


Figure 6.4. A simple one dimensional energy landscape can explain the observed reaction reversibility. (a) If the energy of the products is lower than the energy of the reactants, then the forward reaction is favoured. However, this gives rise to an unfavourable reverse reaction. **(b)** If the energy of the products is higher than the reactants, then the forward reaction is not favoured. The reaction can be catalysed by the addition of mechanical force that reduces the height of the energy barrier. In the case that force has triggered this thermodynamically forbidden reaction, the reverse reaction in the absence of force is downhill and thus extremely favourable.

Crucially, these preliminary results seem to put forward a global scheme that links chemical reactivity with reaction reversibility. This calls for a battery of experimental and theoretical approaches, which I will conduct in the near

future, to gain a comprehensive understanding of how chemical kinetics can be finely coupled with thermodynamics to predict and modulate and the reversibility of individual mechanochemical reactions.

APPENDIX I

Supporting information for chapter 3: The mechanochemistry of copper reports on the directionality of unfolding in model cupredoxin proteins.

Methods

Protein engineering

Protein engineering was conducted by Ainhoa Lezamiz (King's College London) using the following procedure:

The (Azu-I27)₄, (Azu_{C112A}-I27)₄, (Azu_{C26A}-I27)₄ and (Plasto-I27)₄ polyproteins were subcloned using the BamHI, BglII and KpnI restriction sites²¹⁹. Synthetic gene for plastocyanin was produced by Life Technologies (Paisley, UK). The multidomain proteins were cloned into the pQE80L (Qiagen) expression vector, and transformed into the BLR(DE3) Escherichia coli expression strain. All mutations were introduced by using the QuikChange Site-Directed Mutagenesis kit (Agilent). The cells were grown in LB broth supplemented with 100 µg/mL ampicillin at 37°C. After reaching an OD₆₀₀ of ~0.6 the cultures were induced with Isopropyl β-D-1-thiogalactopyranoside (1 mM) and incubated overnight at 20°C (constructs containing Azurin) or 16°C (Plastocyanin construct). For some constructs, CuSO₄ (0.5 mM) was supplemented to the cultures right before induction to increase copper intake. After harvesting the cells, disruption using a French Press was performed. The polyproteins from the lysate were purified by metal affinity chromatography on Talon resin (Clontech) followed by gel-filtration using a Superdex 200 10/300 GL column (GE Biosciences). Protein concentration in the samples was estimated using the Bradford protein assay.

Force spectroscopy

Constant velocity Atomic Force Microscopy (AFM) experiments were conducted at room temperature using both a home-made set-up³⁵⁴ and a commercial Luigs and Neumann force spectrometer¹⁶⁷. In all cases, the sample was prepared by depositing 1-10 µL of protein in PBS solution (at a concentration of 1-10 mg ml⁻¹) onto a freshly evaporated gold coverslide. Each

cantilever (Si_3N_4 Bruker MLCT-AUHW) was individually calibrated using the equipartition theorem, giving rise to a typical spring constant of $\sim 12\text{-}35$ pN nm^{-1} . Single proteins were picked up from the surface and pulled at a constant velocity of 400 nm s^{-1} . Experiments were carried out in a sodium phosphate buffer solution, specifically, 50 mM sodium phosphate (Na_2HPO_4 and NaH_2PO_4), 150 mM NaCl, $\text{pH}=7.2$.

Data analysis

All data were recorded and analysed using custom software written in Igor Pro 6.0 (Wavemetrics). For all polyproteins, only recordings showing the signature of at least 2 events corresponding to the unfolding of the I27 fingerprints were analysed.

Circular dichroism experiments

Experiments were conducted by the facilities of King's College London according to the following procedure:

UV & CD spectra of the 4 polyproteins in phosphate buffer were acquired on the Applied Photophysics Chirscan Plus spectrometer (Leatherhead, UK). 10mm and 0.5mm Quartz Suprasil rectangular cells (Hellma UK Ltd) were employed in the region $400\text{-}190\text{nm}$. The instrument was flushed continuously with pure evaporated nitrogen throughout the experiment. The following parameters were employed: 2nm spectral bandwidth, 1nm stepsize and 1.5s instrument time per point. UV & CD spectra were buffer baseline corrected and measured at 23°C . The far-UV CD spectra were smoothed with a window factor of 4 using the Savitzky-Golay method for better presentation. The far-UV CD spectra of the 4 polyproteins were then corrected for concentration and path-length and expressed in terms of $\Delta\epsilon$ ($\text{M}^{-1} \text{cm}^{-1}$) per amino acid residue ($\text{MWt.}=113$). Protein secondary structure contents were assessed using the Principle Component Regression method based on 16 known protein

structures.

Ellmann's reagent experiments (determination of free protein sulfhydryls)

Experiments were conducted by Ainhoa Lezamiz (King's College London) according to the following procedure:

Protein samples were dissolved in 0.1M phosphate buffer, pH 7.4, containing 4M guanidine hydrochloride and adjusted to a final volume of 950 μ l. A reagent blank (without sample) was prepared in a similar manner. After the addition of 50 μ l of 20 mM 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB) the sample was mixed, incubated at RT for 15 minutes and the absorbance at 412 nm was recorded. The amount of sulfhydryl was calculated using the extinction coefficient of DTNB in the presence of GuHCl ($13,700 \text{ M}^{-1} \text{ cm}^{-1}$). The number of free cysteines was determined using the protein and sulfhydryl molar concentration from each sample.

ICP-OES (determination of copper content in the protein samples)

Experiments were conducted by the facilities of the University of Barcelona and the University of Edinburgh according to the following procedure:

The content of each sample eppendorf tube was totally digested by using a 1ml HNO_3 and 0.5ml H_2O_2 in a closed, Teflon reactor overnight at 90°C . Samples were then cooled down to RT and 15 ml of ultrapure water were added to each individual eppendorf containing the protein samples. The Cu content of the resulting solution was performed using an ICP-OES instrument (Perkin Elmer, Optima 8300) at working wavelengths of 324.752 nm and 327.393 nm. Calibration was performed using 5 standard solutions prepared by dilution of a 1000 ppm certified standard solution.

MD simulations

Experiments where conducted by Guillaume Stirnemann (CRNS – Institut de Biologie Physico-Chimique, Paris) according to the following procedure:

Constant-force steered molecular dynamics (MD) were performed on azurin starting from the PDB structure 4AZU solvated in a water box large enough in the direction of pulling to accommodate the partially unfolded protein (6*6*18 or 6*6*21 nm). The NAMD 2.10 software was used to run the simulations on pure CPU or CPU/GPU nodes. Atomic parameters correspond to those in the CHARMM36 force-field³⁵⁵, while atomic charges for copper II and atoms of the 5 ligands were taken from quantum calculations³⁵⁶. We also constrained the organometallic bond distances as described elsewhere³⁵⁶. Starting from a previously equilibrated protein whose CN axis was aligned along the longer edge of the box (direction of pulling), the C α atom of the first residue was fixed while a constant force (250, 310 or 350 pN) was applied to the C α atom of the last residue. In half of the trajectories, C α from residue #1 was fixed and C α from residue #128 was pulled; for the other half, C α of residue #128 was fixed and C α from residue #1 was pulled in the opposite direction. 21 such trajectories, lasting between 6 and 20 ns (until full possible extension was reached), were generated for the protein containing the disulfide bond (10 or 11 in each direction) in a ~18-nm long box; 9 trajectories total were generated in the absence of disulfide bond, in a slightly larger box (~21 nm). A very similar methodology was employed to simulate plastocyanin (3BQV, $n=10$ trajectories – 5 in each direction) and apoplastocyanin (2PCY, $n=10$ trajectories – 5 in each direction). As the active site of plastocyanin is very similar to that of azurin, we used the same set of atomic charges and bond potentials, albeit at slightly different equilibrium distances corresponding to that measured in the crystal structure.

- Generalities

Constant-force steered molecular dynamics (MD) were performed on azurin starting from the PDB structure 4AZU solvated in a neutralized water box large enough in the direction of pulling to accommodate the partially unfolded protein. Simulations were performed both on the native, WT protein and on a reduced form lacking the disulfide bond between Cys3 and Cys26. In the former case, the box size was approximately 6*6*18 nm and contained 61,482 atoms. In the latter case, the ~ 6*6*21-nm box contained 71,879 atoms. The NAMD 2.10 software was used to run the simulations on pure CPU or CPU/GPU nodes. Atomic parameters correspond to those in the CHARMM36 force-field³⁵⁵, while atomic charges for copper (II) and atoms of the 5 ligands were taken from quantum calculations³⁵⁶. We also constrained the organometallic bond distances following reference ³⁵⁶.

- Simulation procedure

The atomic positions were first minimized for 2,000 steps using the steepest-descent algorithm of NAMD. All subsequent simulations were run in the NPT ensemble, using the Langevin thermostat (0.2/ps) and the Langevin barostat of NAMD. Long-range electrostatics were treated using particle-mesh Ewald with a grid size of 1.2 Å⁻¹. Simulation timestep was 2 fs and bonds between hydrogen and heavy atoms were maintained rigid. After the minimization, the water was equilibrated around the protein, which was maintained fixed for 1 ns. All protein atoms except terminal C_α's were then allowed to move and equilibrated for 1 more nanosecond.

Starting from this previously equilibrated protein whose CN axis was aligned along the longer edge of the box (direction of pulling), the C_α atom of the first residue was fixed while a constant force (250, 310 or 350 pN) was applied to the C_α atom of the last residue. In half of the trajectories, C_α from residue #1 was fixed and C_α from residue #128 was pulled; for the other half, C_α of residue #128 was fixed and C_α from residue #1 was pulled in the opposite direction. 20

such trajectories, lasting between 6 and 20 ns (until full possible extension was reached), were generated for the protein containing the disulfide bond (10 in each direction) in a ~18-nm long box; 9 trajectories total were generated in the absence of disulfide bond, in a slightly larger box (~21 nm). Simulation results are presented in Table S1 and Table S2.

- *Description of the unfolding dynamics in azurin*

MD simulations clearly revealed that, for both transition states, the rupture of the hydrogen bonds (HBs) between the involved β -sheets occurred through an unzipping mechanism because the force is applied parallel to the HBs axes (Fig. 3.6A). In contrast to a protein containing such HBs perpendicular to the force, HB rupture occurs here sequentially, which explains why a relatively low force is sufficient to partially unfold the protein from one side. In the C-terminal unfolding scenario, the unfolding involves the rupture of the $\beta 7$ - $\beta 8$ and $\beta 2b$ - $\beta 8$ hydrogen-bonds, which occur at the same time; when the protein unfolds on the N-terminal side, it unzips by rupturing the hydrogen-bonds between $\beta 2b$ - $\beta 8$, then $\beta 3$ - $\beta 6$ and $\beta 1$ - $\beta 3$ partially, before the remaining hydrogen-bonds of $\beta 1$ - $\beta 3$ break. Interestingly, after one of these partial unfolding events, the protein flips and now exhibits β -sheets HBs perpendicular to the force on the side that is still folded (snapshots in Fig. 3.6 C). Crucially, the key HBs whose rupture would lead to unfolding are now maintained perpendicular to the force and can resist to it in a cooperative manner. This suggests an explanation to the small occurrence (10% of the cases) of consecutive 6- and 9-nm steps. Indeed the low mechanical resistance of the organometallic bonds (Cu-Cys112 and Cu-His46, ~45 pN) makes them more likely to break before the HBs between β -sheets as long as they can cooperatively resist to the force.

The unfolding of the protein lacking its disulfide bond occurs preferentially from the N-terminal side. Interestingly, the time-course of β -sheets rupture

events leading to unfolding slightly differs from that of the regular WT protein (detailed above). In this case, the protein generally unfolds as follows: first the hydrogen-bonds between $\beta 1$ - $\beta 3$ break, followed by $\beta 1$ - $\beta 2a$, $\beta 2b$ - $\beta 8$, and finally $\beta 3$ - $\beta 6$. This suggests that the presence of the SS bond locally rigidifies the protein and prevents the rupture of $\beta 1$ - $\beta 3$, which clearly initiates the unzipping of the all the other β -sheets when the disulfide bond is not present. Instead, the main unfolding event in wt-Azu is the rupture of $\beta 3$ - $\beta 6$ (which gives rise in the simulations to the presence of a short lived conformation around $L=10$ nm; Figure 3.6 in the main text).

- *MD simulations protocol of plastocyanin (holo- and apo- forms)*

Simulations of plastocyanin (PDB #ID 3BQV) and apo-plastocyanin (#2PCY) were performed as previously described for azurin. As the active site of plastocyanin is very similar to that of azurin, with the exception of the fifth weak interaction with Gly45 of azurin being replaced by Pro38 in plastocyanin, we employed the same atomic charges as previously determined for azurin for His39, Cys89, His92 and Met97, and we took the charges on the carbonyl group and the C_α of Gly45 for the corresponding atoms of Pro38. This approximation is not expected to have any impact on the mechanical stability of the β -sheets located on the other side of the protein. As ligand-metal bonds display slightly different lengths than in azurin, we employed the equilibrium distances of the PDB structure to constrain these bonds during the simulations, with the exception of Pro38-Cu bond, which, since it is very long in the PDB (4.32 Å), we did not constrain as the interaction with the metal is probably extremely weak.

The rest of the simulation protocol was identical to that of azurin; solvation in a $\sim 6 \times 6 \times 18$ nm box, minimization and equilibration, and then pulling at a constant force in the range of 250-310 pN by fixing the C_α from the first residue and pulling on the C_α at the other end of the protein in the z-direction, and vice

versa. 10 such trajectories (5/5) for plastocyanin and 10 for apo-plastocyanin were generated following this protocol, lasting a total of 5 or 10 ns each.

- *Effect of the organometallic bonds*

In classical MD simulations, organometallic bonds are usually very challenging to describe as they usually have a covalent character that is very sensitive to the local geometry and to the environment. In usual forcefields, an interaction between 2 atoms is either described by non-bonded interactions (electrostatic and vdW forces) or by a bonded term, usually as an harmonic potential if they form a covalent bond. Organometallic bonds cannot be described by a non-bonded term only as the metal ions would probably not stay in the active site (which we have checked for azurin). Therefore, additional harmonic potentials are necessary to keep the geometry of the active site correct, the main drawback being that they cannot be broken during the simulation, nor changed in energy due to local changes of the environment or the ligand-field geometry.

However, for the purpose of the current work, such an approximation is probably very reasonable, for 2 main reasons. First, the mechanical unfolding events studied here all start with the breaking of key sets of hydrogen-bonds that are located on β -sheets placed on the other side of the protein, next to both termini. The breaking of metallic bonds is only involved at the very end of the unfolding events, when one portion of the protein is extended. This bond-breaking is not possible in the simulations but can be artificially accounted for when comparing the simulation observed extensions to that of the experiments (see Simulation section of azurin).

Second, we have simulated plastocyanin in its regular, metal-bonded form, but also in its apo- structure, lacking the metal. As seen in Tab SI5 and Tab SI6, in both cases we observe a very similar ratio of N-terminal to C-terminal unfolding events (around 80/20%, respectively), suggesting that the metal

active site has actually none or very little influence on the mechanical stability of plastocyanin. Moreover, the first step of the mechanical unfolding in both scenarios is identical in each case (rupture of $\beta 1$ - $\beta 3$ for N-terminal unfolding; rupture of $\beta 2$ - $\beta 8$ and $\beta 7$ - $\beta 8$ for C-terminal unfolding). The main role of the metal in the current investigation is thus to reveal the intermediate states of the protein unfolding that would normally totally unfold in the absence of metal.

- *Predictions of released contour lengths from MD trajectories*

MD simulations can be used to predict the contour lengths of unfolding in the two scenarios (N-terminal and C-terminal unfolding). To this end, we have followed the following procedure (Table SI7): we first measured, for each simulation, the extension of the protein during unfolding, i.e., the difference between the final and original end-to-end distances projected onto the z-axis. Then, since in the simulations unfolding does not occur up to Cys112 for C-terminal unfolding (rather, up to Met121) and up to His46 for N-terminal unfolding (rather up to Gly45), we artificially added the extra-lengths corresponding to the release of the amino-acids in between (Cys112-Met121 and Gly45-His46), at each simulation force, using the worm-like chain model with a contour length per residue of 0.38 nm and a persistence length of 0.4 nm. This correction is very moderate for N-terminal unfolding but more substantial for C-terminal (about 2 nm), and it is force-dependent. By doing so, we obtained the average extension in each scenario and at each force. The last step is to convert these extensions at finite forces into that measured in the experiments, which correspond to the contour lengths obtained by applying a WLC model fit of the data. This conversion is done once again using the WLC model on the « free » amino-acids, i.e. those that are extended and that do not belong to the folded structure anymore in each intermediate (between each free-termini and the active site, minus the portion between the disulfide bond, if applicable). Following this procedure, we obtained the following values: 7.3

nm (N-terminal, with SS bond present); 10.1 nm (C-terminal, with SS bond present); 18.0 nm (C terminal, without SS bond). The agreement with the experimental measurements (respectively 6, 9 and 19 nm) and the schematic predictions is very good.

Results

Biochemical characterization of the samples:

a) Ellman's test:

The results of the Ellman's test for the three polyproteins (I27-Azu)₄, (I27-Azu_{C26A})₄ and (I27-Azu_{C112A})₄ is summarized below:

	Measured free thiols	Free thiols if SS bond is present in azurin	Free thiols if reduced cysteines in azurin
(I27-Azu) ₄	14.7	14	22
(I27-Azu _{C26A}) ₄	16.8	18	18
(I27-Azu _{C112A}) ₄	11.6	10	18

The fact that the results obtained by the Ellman's assay (red) clearly follow the trend of the expected number of free cysteines present in the sample, and in particular that the wt protein (with the expected disulfide bond) exhibits a lower value than the C26 mutant (where the disulfide bond is not present) strongly suggests that the disulfide bond is preserved in the wt-azurin construct.

b) ICP-OES experiments:

Measurement of the copper content in the analysed protein samples provides quantification of the % of copper uptake for each sample (assuming that 100% of the proteins are correctly folded).

(I27-Azu)₄: 38.58%

(I27-Azu_{C112A})₄: 11.38 %

(I27-Plastocyanin-wt)₄: 55.94%

The copper uptake for each of these proteins is, as expected, lower than 100%. This implies that some of the proteins within the polyprotein context did not bind the copper cofactor. This is likely to explain the two-state trajectories, lacking the mechanical intermediate, which we observed in the single molecule mechanical experiments. Crucially, the C112A mutation, while showing a lower percentage of copper uptake, does not completely abolish copper binding. These results are in line with our experimental observations showing that the C112A mutant exhibited a larger proportion of two-state trajectories.

Figures and tables

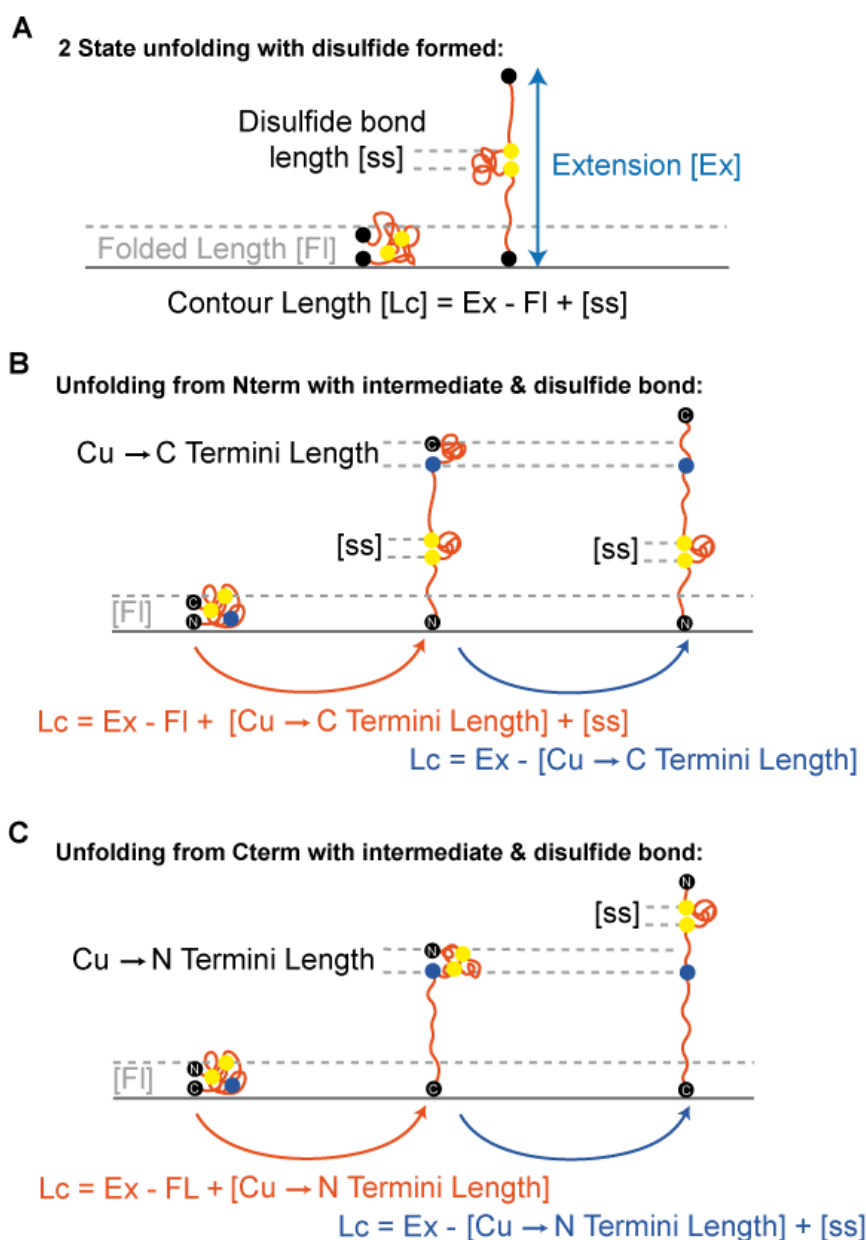


Figure 1. Schematic representation of the two- and three-state (via N- or C- terminus) unfolding of a single azurin molecule containing a disulfide bond. The corresponding equations to calculate the increment in contour length for each event are provided. (A) When no intermediate is observed, the expected increment in contour length (L_c) is calculated by subtracting the initial folded length ($FI = 1.12$ nm, which is the distance between the N- and C- termini obtained from the crystal structure) to the protein extension, ($Ex = \# \text{ of amino acids} \times 0.38 \text{ nm/aa} - FI$) and adding the length of the formed disulfide bond (ss). (B) When the unfolding process occurs via an intermediate, the contribution of the disulfide bond length is dictated by which termini unfolds. In addition, the remaining folded portion of the protein must be accounted for. As the disulfide bond is located between the N-terminus and His-46, when the N-terminus unfolds the bond length is accounted for in the initial unfolding event (orange equation). (C) Conversely, when the C-terminus unfolds first, the disulfide bond remains in the folded region of the protein, and is only extended, once the metallo-bond has been ruptured (blue equation).

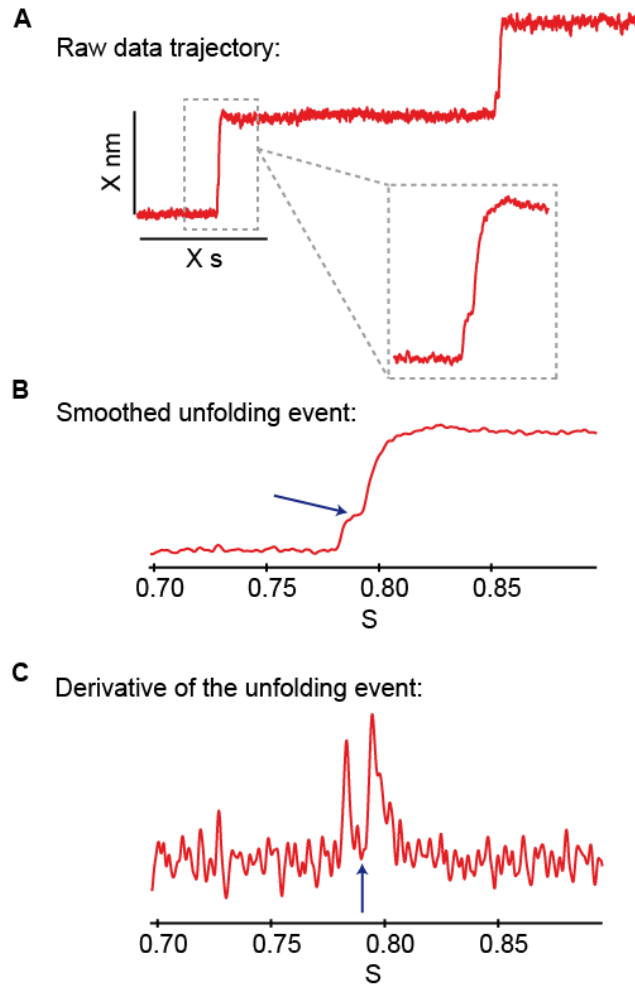


Figure 2. Detection method for the unfolding intermediates observed in force clamp spectroscopy. (A) Raw single molecule force clamp trajectory displaying the unfolding of two consecutive wt-azurin monomers at a constant force of 30pN. Within the main unfolding step, an intermediate corresponding to the rupture of the metallobond can also be observed. In some cases the intermediate occurs quickly (~6 ms, dashed inset). (B) In order to unambiguously determine the step height in these instances, we firstly smooth the trajectory by applying a 20-point box filter, which plots the average point of the neighbouring 20 data points (C). We then took the derivative of the filtered trace to highlight the point of inflection, which provides precise location of the intermediate event.

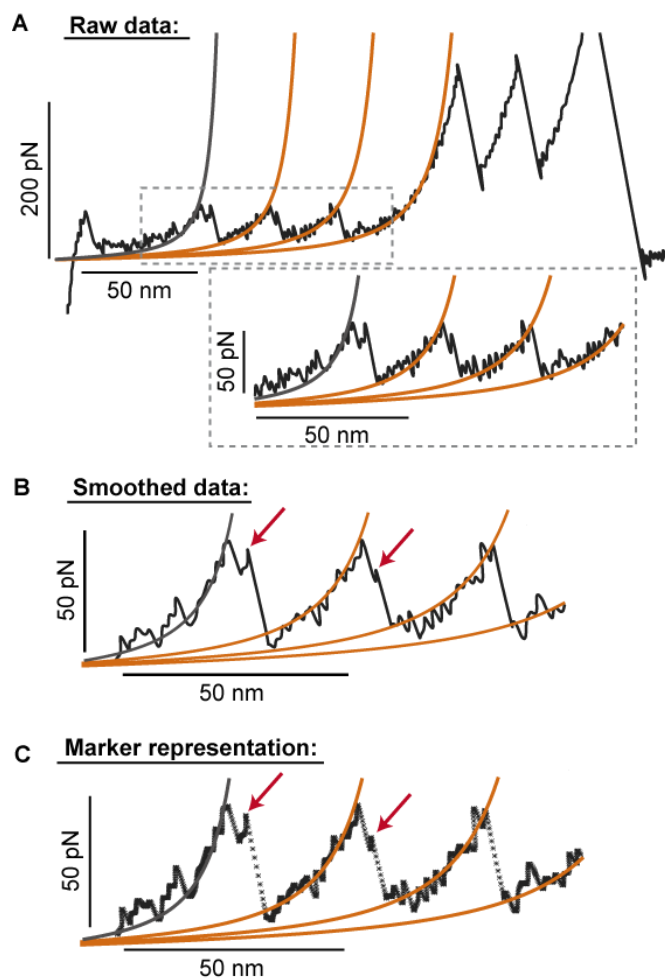


Figure 3. Detection method for intermediate events in the unfolding trajectory of a single molecule in force extension mode. (A) Raw data trajectory of three azurin unfolding events (Orange WLC) followed by the unfolding of two I27 marker proteins. Trajectories are typically acquired using a high Nyquist fraction (~ 0.5) and a high number of points per trace ($\sim 5,000$ - $9,000$). (B) Reducing the Nyquist fraction down to (~ 0.1) and applying a low pass filter, which removes small fluctuations to reveal the unaltered form of the data, provides further evidence of the presence of a clear mechanical intermediate (red arrow) that occurs after the main protein unfolding event. (C) Representing the trajectory as markers instead of a continuous line, where each marker corresponds to an acquired data point, reveals a high density of data points defining the intermediate. The localisation of data points indicates that the cantilever gets temporarily 'held' at a mechanically resilient event, signature of barrier crossing. In the case of azurin, such mechanical resilience originates from one of the metallo bonds coordinating the copper ion.

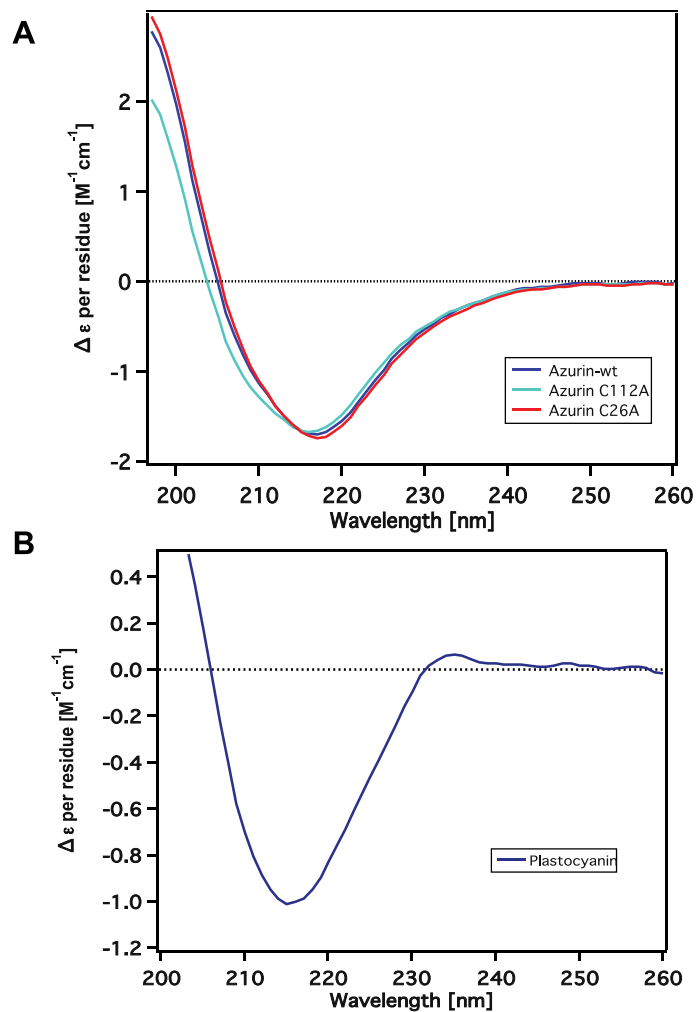
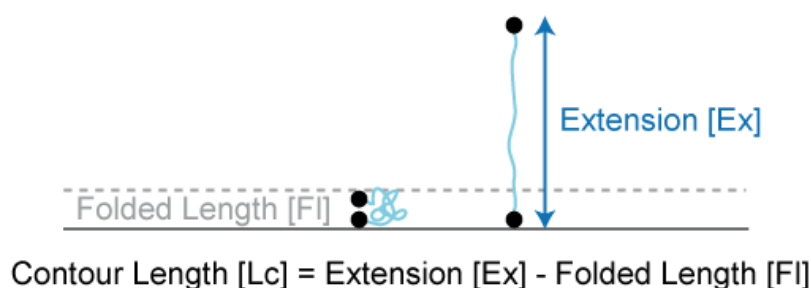


Figure 4. CD spectra of the different polyproteins reveal that in all cases the protein is correctly folded. (A) CD spectra comparing the folding of the three $(Azu-I27)_4$, $(Azu_{C112A}-I27)_4$, $(Azu_{C26A}-I27)_4$ polyproteins used in our single molecule experiments. Crucially, the spectrum of the Azu_{C112A} mutant displays a similar behaviour as the Azu_{wt} and Azu_{C26A} forms, that suggesting that the protein is correctly folded. (B) Similarly, the CD spectrum of the $(Plasto-I27)_4$ polyprotein ensures the correct folding of the wt-plastocyanin protein.

A 2 State Unfolding:



B 3 State Unfolding due to intermediate:

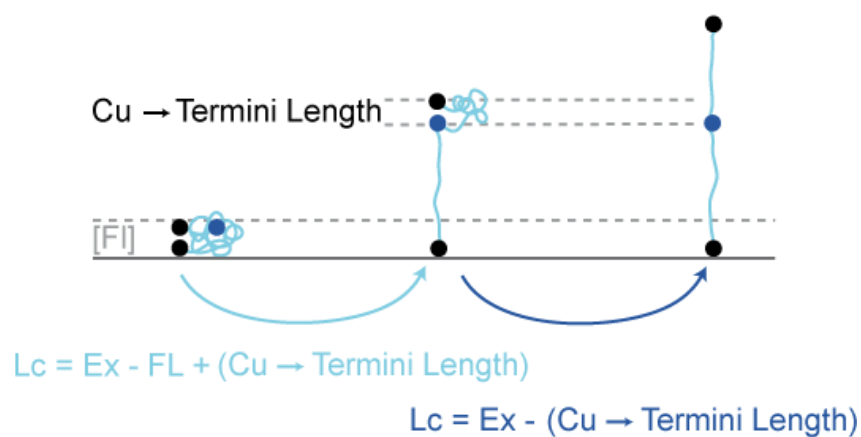


Figure 5. Schematic representation of the calculated increment in contour lengths for the unfolding of a single plastocyanin molecule. (A) When unfolding occurs in a two-state process, the calculated increment in contour length simply corresponds to the protein extension [Ex, given by number of amino acids * 0.38nm/aa - the folded length (FL) of the protein]. The folded length (Fl = 1.53 nm) is the distance between the N- and C- termini obtained from the crystal structure. (B) When an intermediate is observed, the equation shown in (A) must still hold true between the first and final event, however the folded length between the copper and the remaining folded protein must be accounted for. The light blue equation gives the increment in contour length for termini unfolding up to the copper, and the equation in dark blue gives the remaining length once the copper-bond has ruptured.

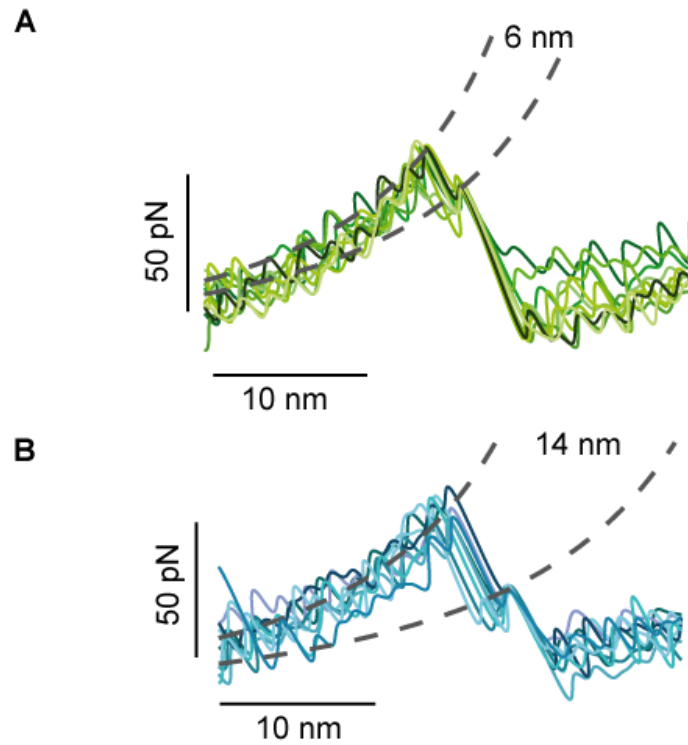


Figure 6. Plastocyanin displays two readily distinct mechanical intermediates. (A) The overlay of $n=10$ plastocyanin unfolding events containing a clear mechanical intermediate of ~ 6 nm, corresponding to the unfolding of the C-terminus up to the C89 or H92 – Cu bond. (B) Similarly, unfolding from the N-terminus up to the mechanically resilient H39 copper bond corresponds to an increment in contour length of ~ 14 nm (WLC fit, grey dashed lines, $n=9$).

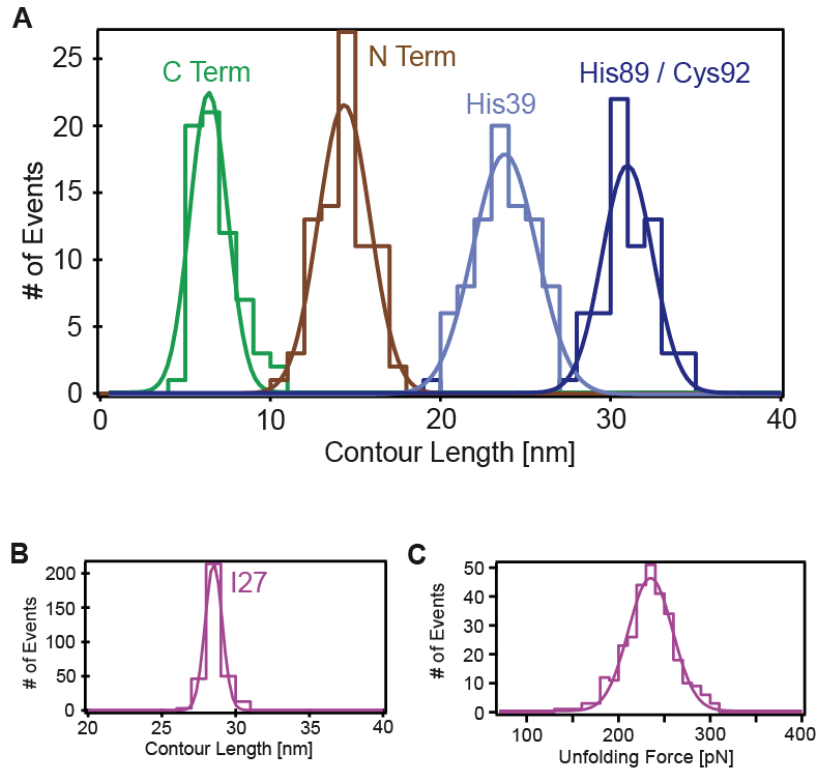


Figure 7. Increment in contour length hallmarking the intermediates encountered upon unfolding plastocyanin from the N- and C-termini. (A) Histogram of contour lengths corresponding to the protein unfolding from the N-terminus up to the H39 – Cu bond (brown, 13.8 ± 1.4 nm, $n=82$) and to the remaining extension once the bond has ruptured up to the total protein contour length (light blue, 23.3 ± 1.9 nm, $n=82$). Similarly, when the C-terminus unfolds up to either the C89 or H92 – Cu bond, an increment of 6nm is obtained (green, , 5.9 ± 1.1 nm $n=66$). Once this bond has been ruptured, the remaining 30.5 ± 1.5 nm is measured (dark blue, $n=66$). (B) Histogram of increment in contour length for the I27 marker protein (28.0 ± 0.6 nm, $n=326$) and (C) its corresponding unfolding force (230 ± 24 pN , $n=326$)

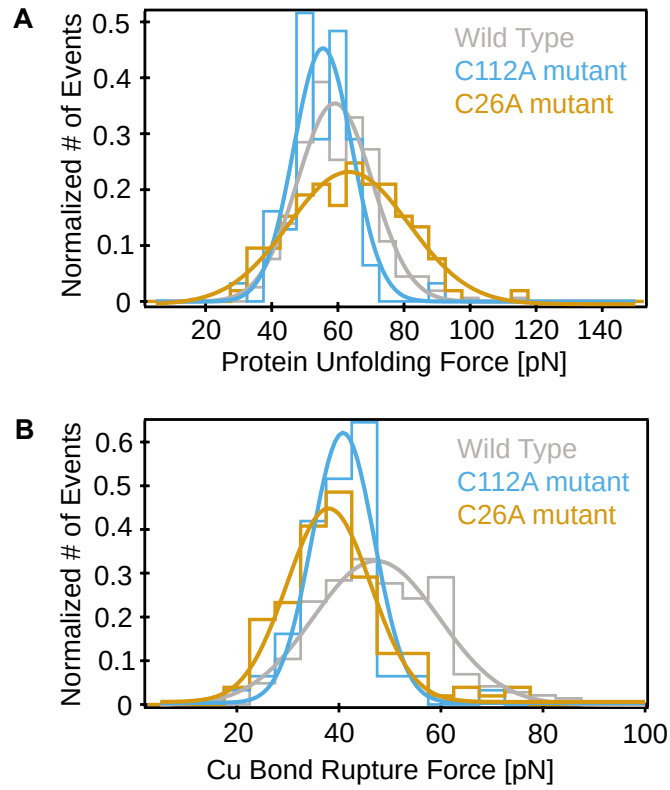


Figure 8. The unfolding forces are identical for the three $(\text{AzucI27})_4$, $(\text{AzucI12A-I27})_4$ and $(\text{AzucI26A-I27})_4$ polyproteins. Similarly, the measured forces required to disrupt the intermediate conformations upon breaking individual Cu-S or Cu-N bonds are almost identical for all three constructs. (A) Gaussian fit to the distribution of the forces required to unfold the wt-Azu protein (grey) yields a value of 54.1 ± 11 pN, $n = 320$. The values obtained for the unfolding of the AzucI12A (blue) and AzucI26A (yellow) polyproteins exhibit similar unfolding forces of 50.5 ± 8.8 pN, $n = 62$ and 58.3 ± 18 pN, $n = 105$, respectively. (B) Similarly, the force required to disrupt the mechanical intermediates is also almost identical in all cases (wt-Azu protein, grey, 42.3 ± 12 pN, $n = 289$; AzucI12A , blue, 35.9 ± 6 pN, $n = 62$; AzucI26A , yellow, 35.5 ± 8 pN, $n = 105$). Surprisingly, the rupture forces are almost indistinguishable in the case of the rupture of the Cu-S or Cu-N bond.

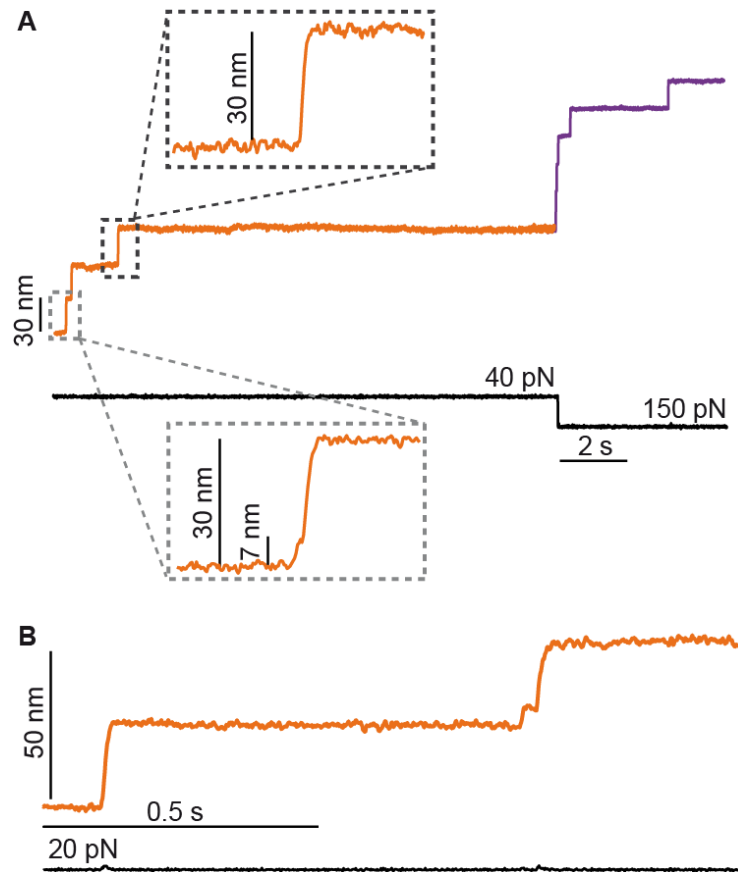


Figure 9. Force clamp experiments further demonstrate a complex unfolding pathway in the mechanical unfolding of azurin; while some events occur in a simple two-state scenario, other events exhibit well-defined mechanical intermediate conformations. (A) Unfolding trajectory of the (AZUC112A-I27)₄ polypeptide. A first pulse of 40 pN triggers the unfolding of the azurin monomers within the polypeptide chain (orange). After 15 seconds, a second pulse at higher forces of 150 pN triggers the unfolding of the I27 monomers (purple). Inset: while the first unfolding event occurs via a clear intermediate of ~7 nm, thus corresponding to disruption of the Cu-N_{His46} bond after the protein has unfolded from the N-terminus, the third event occurs in an all-or-none fashion, lacking any intermediate. (B) Pulling the protein at forces as low as 20 pN slows down the unfolding rate and enables the time-separation of both events. While the first event occurs in an all-or-none process, the second event exhibits a well-defined intermediate event of ~7 nm, corresponding to the ~9 nm event observed under constant-velocity conditions.

Azurin (PDB:1AZU) – Formed disulfide bond

Total number of amino acids : 105 (= 128 – 23aa trapped behind bond)

Folded length : 1.12nm

N-terminus → Cu length : 2.30 nm

C-terminus → Cu length : 2.76 nm

Disulfide bond length : 0.67 nm

Whole Protein (2 state unfolding)		
Expected Length	Measured Length	Measured Force
39.45 nm $(105 \times 0.38) - 1.12 + 0.67$	38.6 ± 1.0 nm $(n = 94)$	52.7 ± 12 pN
N Terminus unfolding → His46		
11.1 nm $(23 \times 0.38) - 1.12 + 0.67 + 2.76$		
His46 → Fully unfolded		
28.4nm $(82 \times 0.38) - 2.76$		
C Terminus unfolding → Cys112		
7.26 nm $(16 \times 0.38) - 1.12 + 2.30$		
Cys112 → Fully unfolded		
32.2 nm $(89 \times 0.38) - 2.30 + 0.67$		

Azurin – Without disulfide bond

Whole protein (2 state unfolding)		
Expected Length	Measured Length	Measured Force
47.5 nm (128*0.38) – 1.12	47.6 ± 1.2 nm (n = 77)	56.8 ± 15 pN
N Terminus → His46		
19.1 nm (46*0.38) – 1.12 + 2.76		
His46 → Fully unfolded		
28.4 nm (82*0.38) – 2.76		
C Terminus → Cys112		
7.3 nm (16*0.38) – 1.12 + 2.30		
Cys112 → Fully unfolded		
40.3 nm (112*0.38) – 2.30		

Table 1. Table of calculated increment in contour lengths for the a single azurin molecule for all unfolding scenarios both with and without a formed disulfide bond.

Plastocyanin (PDB:3BQV)

Total number of amino acids : 105

Folded length : 1.53nm

N-terminus → Cu length : 2.43 nm

C-terminus → Cu length : 2.94 nm

Whole Protein (2 state unfolding)		
Expected length	Measured Length	Measured Force
38.4 nm (105*0.38) – 1.53	36.8 ± 1.1 nm (<i>n</i> = 135)	75.8 ± 12 pN
N Terminus unfolding → His39		
16.23 nm (39*0.38) - 1.53 + 2.94	13.8 ± 1.5 nm (<i>n</i> = 82)	75.3 ± 12 pN
His39 → Fully unfolded		
22.1nm (66*0.38) – 2.94	23.3 ± 1.9 nm (<i>n</i> = 82)	46.0 ± 11 pN
C Terminus unfolding → Cys89		
7.0 nm (16*0.38) - 1.53 + 2.43	5.9 ± 1.1 nm (<i>n</i> = 66)	65.9 ± 13 pN
Cys89 → Fully unfolded		
31.4 nm (89*0.38) – 2.43	30.5 ± 1.5 nm (<i>n</i> = 66)	51.8 ± 13 pN

Table 2. Table of calculated increment in contour lengths for the a single plastocyanin molecule for all unfolding scenarios.

APPENDIX II

Supporting information for chapter 4: Protein S-sulfenylation is a fleeting molecular switch that regulated non-enzymatic oxidative folding.

Methods

Protein engineering

All protein engineering was performed by Ainhua Lezamiz (King's College London) with the following procedure:

The (I27_{E24C-K55C})₈, (I27_{G32C-A75C})₈, (I27)₈ and (I27_{C47A-C63A})₈ polyproteins were subcloned using the BamHI, BglII and KpnI restriction sites. The multidomain proteins and the I27_{E24C-K55C} monomer were cloned into the pQE80L (Qiagen) expression vector, and transformed into the BLR(DE3) *Escherichia coli* expression strain. The cells were grown in LB broth supplemented with 100 µg/mL ampicillin at 37°C. After reaching an OD₆₀₀ of ~0.6 the cultures were induced with Isopropyl β-D-1-thiogalactopyranoside (1 mM) and incubated overnight at 20°C. After harvesting the cells, disruption using a French Press was performed. The polyproteins from the lysate were purified by metal affinity chromatography on Talon resin (Clontech) followed by gel-filtration using a Superdex 200 10/300 GL column (GE Biosciences). Protein concentration in the samples was estimated using the Bradford protein assay.

Force spectroscopy

Single molecule force-clamp spectroscopy Atomic Force Microscopy (AFM) experiments were conducted at room temperature using both a home-made set-up³⁵⁴ and a commercial Luigs and Neumann force spectrometer¹⁶⁷. The sample was prepared by depositing 1-10 µL of protein in PBS solution (at a concentration of 1-10 mg ml⁻¹) onto a freshly evaporated gold coverslide. Each cantilever (Si₃N₄ Bruker MLCT-AUHW) was individually calibrated using the equipartition theorem, giving rise to a typical spring constant of ~12-17 pN nm⁻¹. Single proteins were picked up from the surface and pulled at a constant force. Individual polyproteins were fished by pushing the cantilever onto the surface exerting a contact force of 500-1500 pN so as to promote the non-

specific adhesion of the proteins on the cantilever surface. The piezoelectric actuator was then retracted to produce a set deflection (force), which was set constant throughout the experiment (~8-45 seconds) thanks to an external, active feedback mechanism while the extension was recorded. The force feedback was based on a proportional, integral and differential amplifier (PID) whose output was fed to the piezoelectric positioner. The feedback response is limited to ~3-5 ms. Thanks to the high-resolution piezoelectric actuator, our measurements of protein length have a peak-to-peak resolution of ~0.5 nm. Data of the force traces was filtered using a pole Bessel filter at 1 kHz. In all instances the protein was unfolded at 150 pN for 0.5s, followed by a high force pulse of 500 pN, except when the protein was extended for very short times (0.5s and 0.75s) when a pulse of 800pN was applied. Experiments were carried out in a sodium phosphate buffer solution, specifically, 50 mM sodium phosphate (Na_2HPO_4 and NaH_2PO_4), 150 mM NaCl. The pH was adjusted at pH =7.4 and pH=12.8. The final pH value in each solution was adjusted by adding the required amounts of NaOH solutions. Measuring solutions were filtered through a 0.2- μm membrane prior to each experiment. In the case of the dimedone experiments, 5,5-Dimethyl-1,3-cyclohexanedione (Sigma-Aldrich, 95%) was dissolved in pH =12.8 PBS buffer to give a final concentration of 30 mM. The pH of the dimedone solution was checked prior to each experiment. 3-Mercapto-1-propanol (Sigma-Aldrich, 95%) was freshly prepared prior to every experiment, dissolved in pH=7.4 PBS buffer to give a final concentration of 400 mM. Pre-reduced (I27_{E24C-K55C})₈ polyproteins were obtained by incubating the polyproteins in 10 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP, Sigma-Aldrich) for 20 minutes prior to AFM experiments. Hydrogen peroxide solutions (H_2O_2 , Sigma-Aldrich) were obtained by dissolving the 30% (v/v) solution in PBS to a final concentration of 300 μM H_2O_2 .

Data analysis

All data were recorded and analysed using custom software written in Igor Pro 6.0 (Wavemetrics). For all polyproteins, only recordings showing the signature of at least 5 unfolding events (15 nm steps) followed by 5 reduction events (10 nm steps) were analysed. We considered only recordings corresponding to the *test* pulse in which the protein was extended back to the same full unfolded length characterising the *initial* pulse, to ensure that the same protein was completely stretched in the two pulses. No traces that included unfolding events during the high force pulse were included in the analysis. The number of independent observations, N , for a particular condition was counted as the total number of protein domains (as observed in the *initial* pulse). Standard error of the mean (SEM) for the refolding fraction was estimated through the bootstrap method, where each recording was treated as an independent data point. SEM for the fit parameters was determined as standard error for the coefficient in the fit, accounting for the measurement errors of the individual data points. To assess the goodness of the fits in Fig. 4.5 and Fig. 4.6, we conducted reduced chi-squared, χ^2_{red} , analysis. Data in Figure 4b was compared to the Worm-like Chain model of polymer elasticity³⁴, using as fixed parameters $L_c = 28.5$ nm, $p = 1.2$ nm.

Mass spectrometry experiments

All protein engineering was performed by Steven Lynham (King's College London) with the following procedure:

Enzymatic digestion

The monomeric protein samples were prepared by diluting 20 μ l of protein (0.5 μ g/ μ l) in 20 μ l of either PBS pH 7.4, PBS pH12.8 or PBS pH12.8 + 30mM dithionite. The high pH samples had a calculated final pH = 12.5. In-solution digestion with trypsin of 5 μ g of total protein monomers was performed prior

to subsequent analysis by mass spectrometry. Trypsin digestion at a ratio of 1:20 (enzyme:substrate) was carried out for 2 hours at 37 °C

LC-MS/MS

Chromatographic separations were performed using an EASY NanoLC system (ThermoFisherScientific, UK). Peptides from a total protein amount of 2 µg on column were resolved by reversed phase chromatography on a 75 µm C18 column using a three step linear gradient of acetonitrile in 0.1% formic acid. The gradient was delivered to elute the peptides at a flow rate of 300 nL/min over 60 min. The eluate was ionised by electrospray ionisation using an Orbitrap Velos Pro (ThermoFisherScientific, UK) operating under Xcalibur v2.2. The instrument was programmed to acquire in automated data-dependent switching mode, selecting precursor ions based on their intensity for sequencing by collision-induced fragmentation using a Top20 CID method. The MS/MS analyses were conducted using collision energy profiles that were chosen based on the mass-to-charge ratio (m/z) and the charge state of the peptide. Details on the procedure followed for data analysis are provided in the Supplementary Information section.

Mass spectrometry database searching

Raw mass spectrometry data were processed into peak list files using Proteome Discoverer (ThermoScientific; v1.4). Processed raw data was searched using the Mascot search algorithm (www.matrixscience.com) against an in-house database containing the sequence of the mutant protein monomer. The data was also searched using the Error Tolerant function which by iterating through a comprehensive list of chemical and post-translational modifications, determines the mass of additions or losses from the residues in the peptide sequences without prior knowledge of the expected modifications. LC/MS/MS analysis has successfully identified 100% sequence coverage of the

mutant protein monomer with various forms of oxoacid modifications. The Mascot database searching algorithm applies a 95% probability CI in the MOWSE scoring that is an identification threshold. Mascot search output files were uploaded into the Scaffold spectral visualisation software (v.4.4.6; www.proteomesoftware.com) which allows statistical filtering of the data at the protein and peptide level for which these filters were applied to the data at the lowest stringency of 5% confidence interval (CI) for minimum protein, 0% CI for peptide values and minimum one peptide assignment. This was chosen to include all assigned peptides that may have fallen below the database applied 95% CI due to poor fragmentation matching and incorrect sequence assignment. All fragmentation spectra were verified manually to determine correct assignment of sequences and post –translation modifications.

Initial data searching concentrated on finding sulfenic acid modifications after the addition of the cyclic diketone compound, dimedone. Dimedone specifically traps this oxoacid form and binds covalently to the cysteine residue. This binding with the subsequent loss of two protons adds a mass of 138.1 Da to the residue which is detected on the peptide by fragmentation in MS/MS mode in the mass spectrometer.

Under conditions of high pH and without the presence of dimedone to specifically trap the sulfenic acid, there is possible evolution of the oxoacids to be present in their different forms. The protein monomer treated with high pH alone was searched against the in-house database which included the modifications sulfenic acid, sulfinic acid and sulfonic acid. Further in-depth searching of the database was also performed with the modifications aldehyde, dehydroalanine and sulphide. Database searching assigned 99% sequence coverage of the protein monomers and also assigned various oxoacid modifications on the cysteine residues of interest.

Further analysis was performed using a fresh sample of monomers in a high pH only environment and treated with enzymatic digestion as previously described. This sample was digested almost immediately following the addition of the high pH buffer. This was performed to determine if the previously identified oxoacid modifications were present at point of treatment, or if they had evolved over the length of time it had spent in the buffer prior to trypsin digestion. Almost complete sequence coverage was achieved once again following database searching of the mass spectrometer raw data. Oxoacid cysteine modifications sulfonic and sulfinic acid were identified at residue Cys55 as were the products of both α -elimination, aldehyde and β -elimination, persulfide and dehydroalanine.

Figures

Supplementary Figures

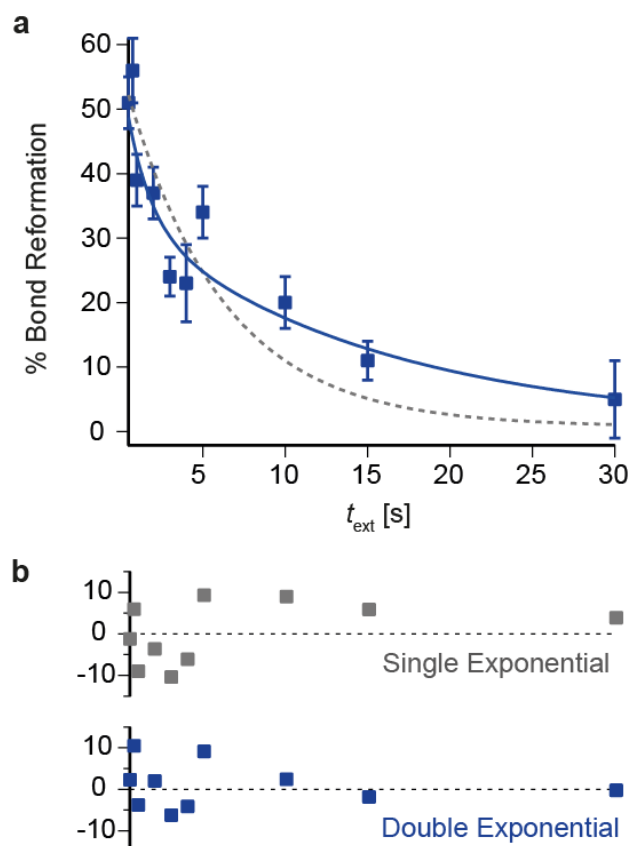


Figure 1. The dramatic decrease of the probability of disulfide reoxidation as a function of Δt_{ext} is better captured by a double exponential decay. (a) Fitting the time-course of disulfide bond reformation as a function of the time the reduced protein is kept stretched at a high force, t_{ext} , (Figure 4.6 in the main text) shows a time-course evolution that can be readily captured with a single exponential fit, yielding a decay rate of $k = 0.17 \text{ s}^{-1}$ ($\chi^2_{\text{red}} = 5.02$). However, a double exponential fit, reminiscent of at least two concomitant processes occurring at the same time with decay rates of $k_1 = 0.065 \text{ s}^{-1}$ and $k_2 = 0.695 \text{ s}^{-1}$, fits the data slightly better ($\chi^2_{\text{red}} = 2.78$) as demonstrated in (b) in the graphs showing the residuals corresponding to each distinct fit.

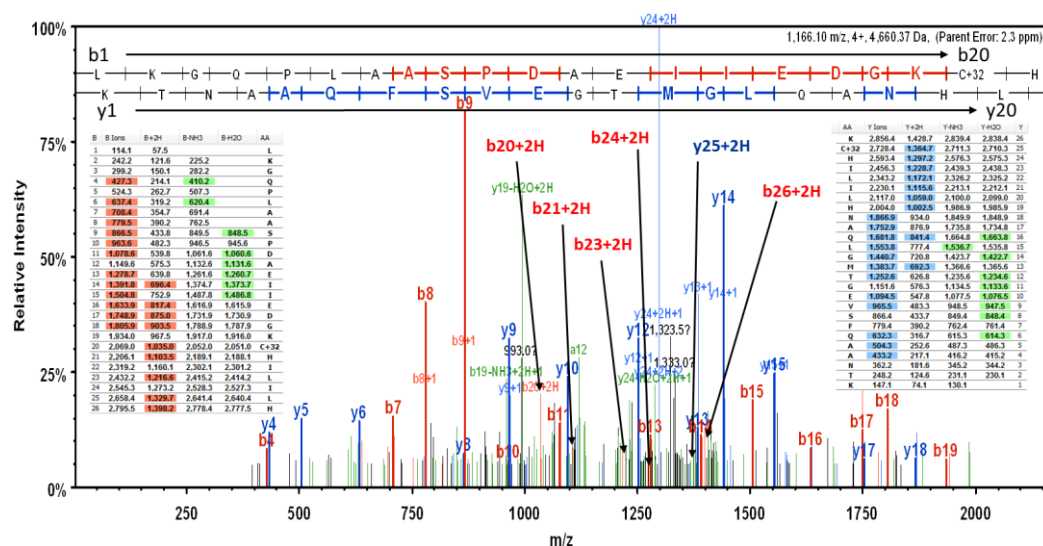


Figure 2. Evidence for sulfinic acid modification on a peptide with m/z 1166.10⁴⁺. Evidence for a sulfinic acid modification with the addition of 32 Da on the cysteine residue at Cys55. Detection of the modification is found on the doubly-charged b20-ion and correct assignment of residue masses after the modification at the doubly-charged b21, b23, b25 and b26-ions. Assignment of the doubly-charged y25-ion including the modification is further supporting information for the correct assignment.

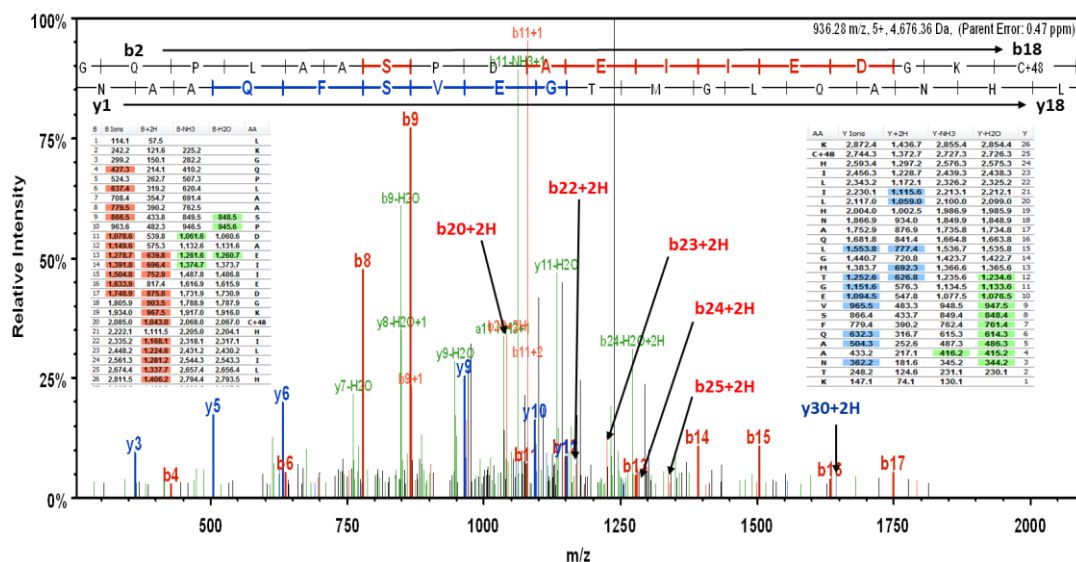


Figure 3. Evidence for a sulfonic acid modification on a peptide with m/z 936.28⁺. Correct peptide identification was provided by the assignment of the sequence data from the b and y-ions in the spectra. The oxidation modification of the cysteine residue, with the addition of 48 Da, was determined by assignment of the doubly-charged b20-ion and subsequent ions after the modification in the peptide sequence (doubly-charged b22 and b25-ions). This determines the modification of sulfonic acid (SO₃H) has been correctly assigned to the cysteine residue at Cys55.

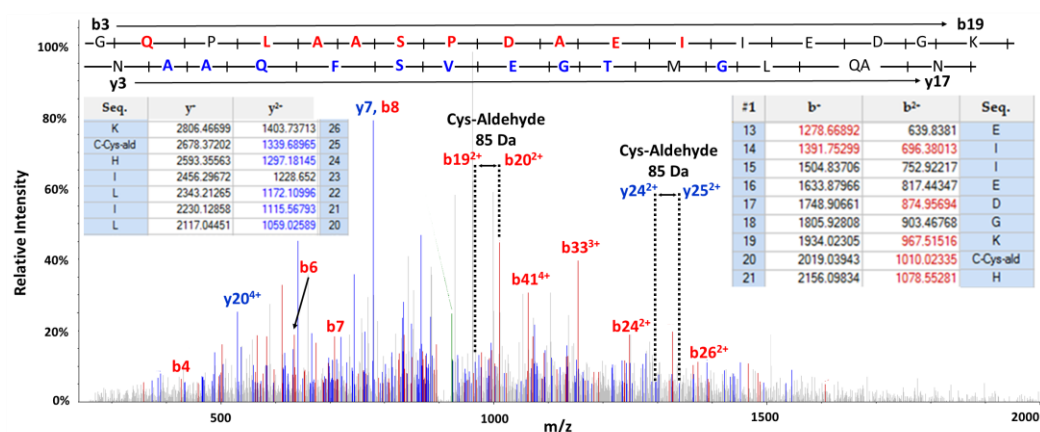


Figure 4. Confirmation of cysteine-aldehyde formation by alpha-elimination in reducing conditions. A peptide with m/z 923.08⁵⁺ was detected with a loss of 17.9928 Da from the cysteine residue in the sequence at the doubly-charged b20-ion and corresponding doubly-charged y25-ion. Further evidence was detected in the spectra for correct assignment of the modified cysteine on the y25³⁺, y25⁴⁺ and y25⁵⁺ fragment ions as well as the b20³⁺ fragment ion (not shown in the spectra to reduce confusing labelling).

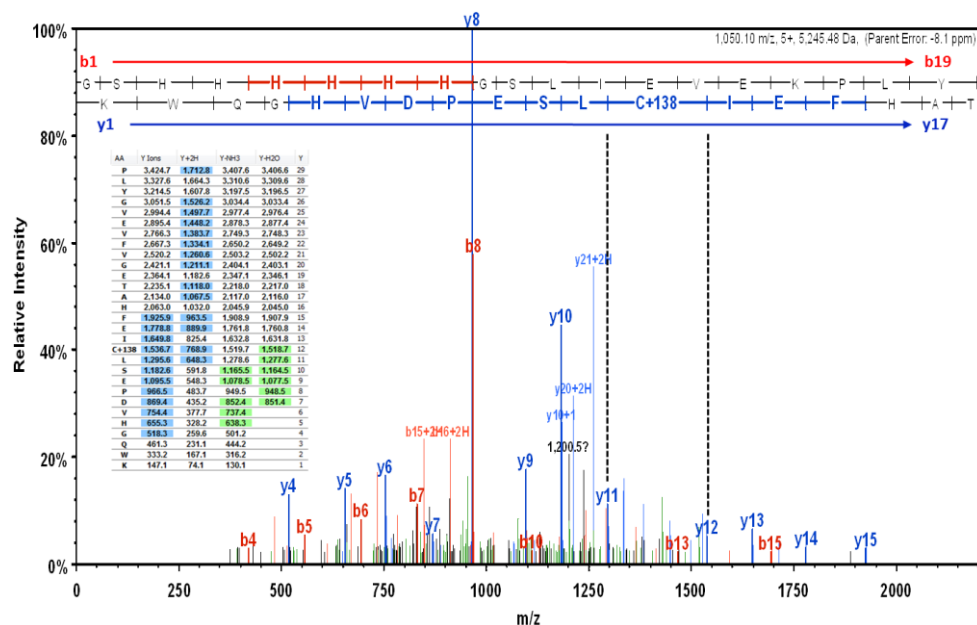


Figure 5. Evidence for a modified peptide following exposure to dimedone. A well matched sequence tag from the y5-y15-ions and from the b5-b8-ions successfully determines the correct assignment of the peptide to the protein of interest including the His tag. A mass shift of m/z 138.1 was detected on the cysteine residue at y12. Correct assignment of the subsequent residues following the modified cysteine gives strong evidence for the correct location of the modification.

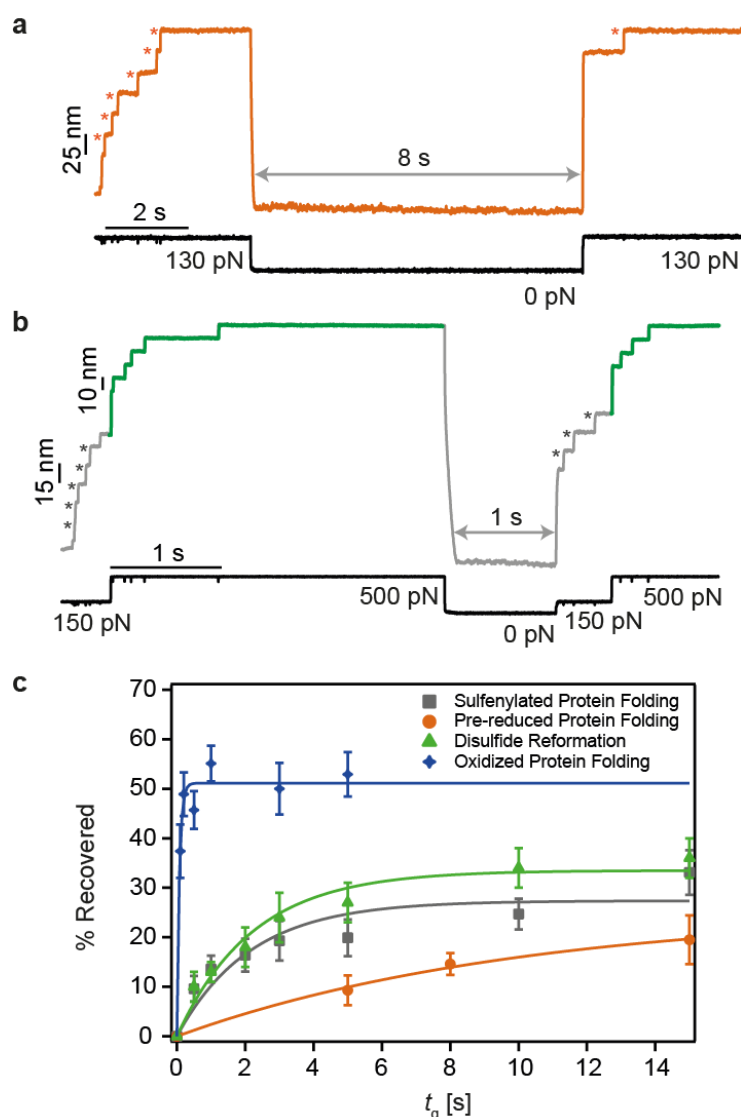
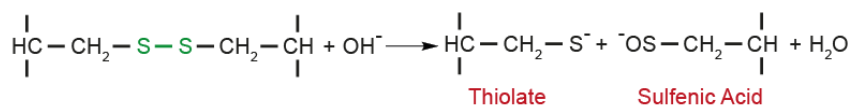


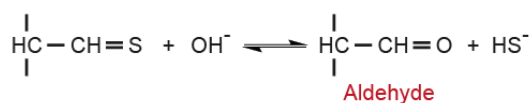
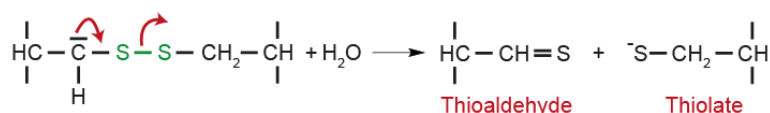
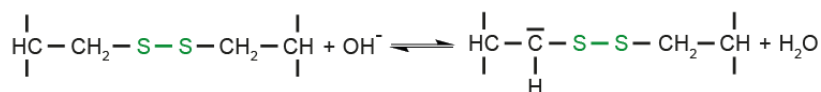
Figure 6. The sulfenic acid-mediated reformation of a disulfide bond increases the rate and yield of protein folding. (a) Incubating (I27_{E24C-K55C})₈ in 10mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) for 30 mins prior to AFM experiments results in the full rupture of all disulfide bonds, as fingerprinted by the ~25 nm steps of unfolding, which corresponds to the complete extension of the reduced protein, void of a disulfide bond. The % of refolding of the pre-reduced protein at pH =12.8 is surprisingly low, even at long t_q . (b) By contrast, after protein sulfenylation (created by the alkaline hydrolysis of the disulfide bond-containing (I27_{E24C-K55C})₈ protein) the refolding efficiency is drastically improved, as observed by the presence of 15 nm steps (grey asterisks) in the test pulse preceding the disulfide bond scission events (green). (c) Measuring the percentage of successful refolding events over a range of quench times t_q probes the time-dependent mechanical folding process. Fitting the data with a single exponential yields the associated rate of folding. While the folding kinetics of pre-reduced (I27_{E24C-K55C})₈ polyprotein at pH = 12.8 yields a folding rate of $k_{\text{fold}} = 0.1 \text{ s}^{-1}$ (orange), the folding rate of the same sulfenylated protein (grey) increases to up to $k_{\text{fold}} = 0.46 \text{ s}^{-1}$. Crucially, the folding yield is also significantly increased when folding is linked to disulfide bond reformation. For comparison, the kinetics of disulfide bond reformation (green, $k_{\text{ref}} = 0.45 \text{ s}^{-1}$) and oxidised protein (blue, $k_{\text{fold}} = 13 \text{ s}^{-1}$) are shown. Altogether, these results demonstrate that the reformation of the disulfide bond triggered by sulfenic acid increases both the rate and yield of protein refolding, thus acting as a rescue element that prevents the protein from irreversibly misfolding.

a. Disulfide scission under high alkaline conditions:

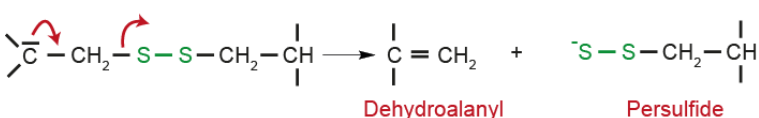
Hydrolysis:



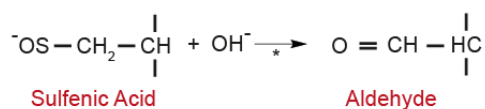
α-Elimination:



β-Elimination:



b. Sulfenic evolution to aldehyde in high alkaline conditions:



*Same mechanism as above α-elimination

Figure 7. The reactivity of a protein disulfide bond and a sulfenic acid moiety under high alkaline conditions. (a) In solution, three possible documented mechanisms can lead to disulfide bond scission in the presence of hydroxide anions (OH⁻); (i) The hydrolysis mechanism corresponds to the scission of the S-S bond through an S_N2 chemical reaction, giving rise to a sulfenic acid and a thiolate; (ii) In the α-elimination mechanism, the carbon in α position is deprotonated, giving rise to an unstable carbanion that can internally re-attack, giving rise to a thiolate and to a thioaldehyde that rapidly evolves to an aldehyde moiety; (iii) Alternatively, during β-elimination, the carbon in β position is deprotonated, resulting in a dehydroalanyl and a persulfide. Our MS results show evidences of the presence of sulfenic acid, aldehyde, dehydroalanyl and persulfide thus suggesting the presence of the three competing mechanisms. (b) Under high pH conditions, the sulfenic acid moiety produced during the hydrolysis of the disulfide bond can evolve into a stable aldehyde moiety via an α-elimination mechanism akin to that observed for the scission of the disulfide bond.

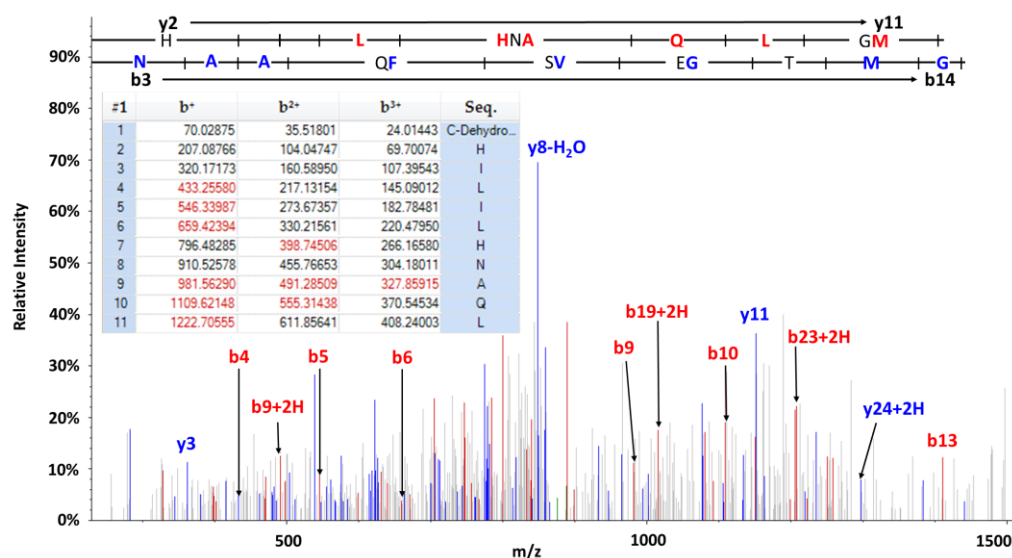


Figure 8. Weak evidence of a dehydroalanine formation from cysteine by beta-elimination in reducing conditions. A peptide with m/z 888.13³⁺ was detected with a loss of 33.9877 Da from the cysteine residue in the sequence at the singly-charged b4-b6-ions in the fragmentation spectra. The modified residue lies at the b1-ion and was not detected in this spectra along with the b2-ion. The presence of correctly matched b-ions after the loss from the cysteine is evidence to suggest that this modified peptide has been correctly matched in the database. This is weak evidence for the presence of a dehydroalanine from Cys55.

Supplementary Table 1: Identification of cysteine containing modified peptides following exposure to high pH buffer prior to digestion. Mass spectrometry analysis by database searching and manual verification.

Cys Modified	24	55	55	55	55	55
Signal	44130	31880	6809	40950	8260	83530
Intensity						49590
Charge State	5 ⁺	4 ⁺	4 ⁺	3 ⁺	5 ⁺	4 ⁺
Mr	5188.40	4468.16	4484.15	2662.38	4611.39	4661.34
<i>m/z</i>	1038.69	1118.05	1122.04	888.13	923.08	1166.09
Modification	Sulfonic acid*	Sulfinic acid*	Sulfonic acid	Dehydroalanine*	Aldehyde	Persulfide
Sequence	² GSHHHHHH GSLIEVEKPL YGVVEFVGE TAHFELDSEP DVHGQWK ⁴⁶	⁴⁹ GQPLAASP DAEIEDGKc HILLIHNAQL GMTGEVSFQ AANTK ⁹⁰	⁴⁹ GQPLAASP DAEIEDGKc HILLIHNAQL GMTGEVSFQ AANTK ⁹⁰	⁶⁶ cHILLIHNA QLGMTGEVS FQAANTK ⁹⁰	⁴⁷ LKGQPLAA SPDAEIEDG KcHILLIHNA QLGMTGEVS FQAANTK ⁹⁰	⁴⁷ LKGQPLAA SPDAEIEDG KcHILLIHNA QLGMTGEVS FQAANTK ⁹⁰
						⁴⁷ LKGQPLAA SPDAEIEDG KcHILLIHNA QLGMTGEVS FQAANTK ⁹⁰

55	
21760	
4 ⁺	
4676.33	
1170.09	
Sulfonic acid	
⁴⁷ LKGQPLAA SPDAEIIIDG KcHILLHNA QLGMTGEYS FQAANTK ⁹⁰	

*weak evidence

Supplementary Table 2: Cysteine containing peptides modified with Sulfenic acid-dimedone following database searching and manual verification.

Cys Modified	24	55	55	55
Signal Intensity	92030	12570	114800	13730
Charge State	5 ⁺	4 ⁺	4 ⁺	4 ⁺
Mr	5245.48	4766.47	4525.27	2833.43
<i>m/z</i>	1050.10	1192.62	1132.33	709.36
Modification	Dimedone	Dimedone	Dimedone	Dimedone
Sequence	² GSHHHHHHGSLEIV EKPL YGVEVFVGETA HFEIDLSEPDVHGQW K ⁴⁶	⁴⁷ LKGQPLAASPD AEII EDGKcHILLHN AQL GMTGEVSFQAANTK ⁹⁰	⁴⁹ CQPLAASPD AEIIED GKcHILLHN AQLGM TGEVSFQAANTK ⁹⁰	⁶⁶ cHILLHN AQLGMT GEVSFQAANTK ⁹⁰

APPENDIX III

Supporting information for chapter 5: Tailoring protein nanomechanics with
chemical reactivity.

Materials and methods

Protein engineering

Protein engineering was kindly performed by Palma Rico and Ainhua Lezamiz (King's College London).

The (I27_{E24C-K55C})₈ and (I27_{G32C-A75C})₈ polyproteins were subcloned using the BamHI, BglII and KpnI restriction sites. The multidomain polyproteins and the I27_{E24C-K55C} monomer were cloned into the pQE80L (Qiagen) expression vector, and transformed into the BLR(DE3) *Escherichia coli* expression strain (VWR International). The cells were grown in LB broth supplemented with 100 µg/mL ampicillin at 37°C. After reaching an OD₆₀₀ of ~0.6 the cultures were induced with Isopropyl β-D-1-thiogalactopyranoside (1 mM) and incubated overnight at 20°C. After harvesting the cells, disruption using a French Press was performed. The polyproteins from the lysate were purified by metal affinity chromatography on Talon resin (Clontech) using equilibration buffer (50 mM sodium phosphate pH 7.0, 300 mM NaCl), washing buffer (50 mM sodium phosphate pH 7.0, 300 mM NaCl, 20 mM Imidazole) and elution buffer (50 mM sodium phosphate pH 7.0, 300 mM NaCl, 250 mM Imidazole). This was followed by gel-filtration using a Superdex 200 10/300 GL column (GE Biosciences) using PBS buffer pH 7.3. Protein concentration in the samples was estimated using the Bradford protein assay.

Force clamp spectroscopy

Single molecule force-clamp spectroscopy Atomic Force Microscopy (AFM) experiments were conducted at room temperature using both a home-made set-up³⁵⁴ and a commercial Luigs and Neumann force spectrometer¹⁶⁷. The sample was prepared by depositing 1-10 µL of protein in PBS solution (at a concentration of 1-10 mg ml⁻¹) onto a freshly evaporated gold coverslide. Each cantilever (Si₃N₄ Bruker MLCT-AUHW) was individually calibrated using the

equipartition theorem, giving rise to a typical spring constant of $\sim 12\text{-}17$ pN nm⁻¹. Individual polypeptides were fished by pushing the cantilever onto the surface exerting a contact force of 500-1500 pN so as to promote the non-specific adhesion of the proteins on the cantilever surface. The piezoelectric actuator was then retracted to produce a set deflection (force), which was set constant throughout the experiment (up to ~ 80 seconds) thanks to an external, active feedback mechanism while the extension was recorded. The force feedback was based on a proportional, integral and differential amplifier (PID) whose output was fed to the piezoelectric positioner. The feedback response is limited to $\sim 2\text{-}5$ ms. Data of the force traces was filtered using a pole Bessel filter at 1 kHz. In the disulfide bond rupture kinetics experiments, the protein was first unfolded at 150 pN for 0.5s, and subsequently the force was raised up to 200 pN and left constant for a large, variable period of time to capture the full kinetics of disulfide reduction for each studied nucleophile. In the oxidative folding experiments, the protein was unfolded at 150 pN for 0.5s, followed by a high force pulse (varying between 200-300 pN depending on the use nucleophile) for 3 seconds. The quench pulse where the pulling force is completely removed was set in all cases to $t_q = 8$ s. Experiments were carried out by keeping the concentration of the deprotonated form of L-cysteine and N-aceyl-cysteine nucleophiles constant to 0.7 mM, whereas the deprotonated form of DL-homocysteine was kept at 0.55mM due to solubility issues, and the total concentration of each nucleophile in the measuring solution was likewise adjusted according to the pK_a of each compound. The different thiol nucleophiles [L-cysteine (Santa Cruz biotechnology, 98%), DL-homocysteine (Sigma-Aldrich, 95%), N-acteyl-cysteine (Sigma-Aldrich, 98%) and TCEP (Sigma-Aldrich)] were diluted into a sodium phosphate buffer solution, specifically, 50 mM sodium phosphate (Na_2HPO_4 and NaH_2PO_4), 150 mM NaCl. The final pH of in each solution was adjusted by adding the required amounts of HCl/NaOH solutions. Measuring solutions were filtered through

a 0.2- μ m membrane prior to each experiment. Each solution was freshly prepared prior to every experiment.

Data analysis

All data were recorded and analysed using custom software written in Igor Pro 6.0 (Wavemetrics). For the kinetics studies, only recordings showing the signature of at least 5 unfolding events (15 nm steps) followed by 5 reduction events (10 nm steps) were analysed. We followed the same criteria for the oxidative folding experiments, in which we considered only recordings where, in the test pulse, the protein was extended back to the same full unfolded length characterising the initial pulse. The number of independent observations, N , for a particular condition was counted as the total number of protein domains (as observed in the *initial* pulse). Standard error of the mean (SEM) for the refolding fraction was estimated through the bootstrap method, where each recording was treated as an independent data point. SEM for the fit parameters was determined as standard error for the coefficient in the fit, accounting for the measurement errors of the individual data points.

DFT calculations

DFT calculations were kindly performed by Guillaume Stirnemann.

Electronic structure calculations using DFT were performed using the Gaussian 09 software³⁵⁷. The B3LYP correlation function^{358, 359} was employed with the 6-311+G(d,p) basis set. All structures were optimized in implicit water solvent consisting of a dielectric medium using the PCM model³⁶⁰. Frequency calculations in the gas phase were performed on these optimized structures. The reported free-energies correspond to the sum of gas-phase energy at 0 K, correction for solvation, zero-point energy, and thermal correction at 298 K. The HOMO energy levels, the electrostatic potentials and the charges

(obtained using the Natural Bond Orbital analysis implemented in Gaussian) were determined from the optimized structures in implicit solvent.

To mimic protein-protein and mixed disulfide bond, we considered dimers of cysteines or its derivatives. At neutral pH, amino-acids would be present in a zwitterionic $\text{NH}_3^+ \dots \text{COO}^-$ form. However, as already observed for amino-acids (including cysteine^{361, 362}), the neutral tautomer $\text{NH}_2 \dots \text{COOH}$ was found to be more stable in all cases. Only the inclusion of explicit water molecules would provide the necessary stabilization of the zwitterionic state by the formation of strong hydrogen-bond with water³⁶³, but it is beyond the scope of the current work. The reported energies therefore refer to the $\text{NH}_2 \dots \text{COOH}$ forms of thiolates for cysteine and homocysteine. For NAC, the NH_2 group is replaced by an amide group, which is not protonated at neutral pH, so the $\text{NH} \dots \text{COO}^-$ form was considered.

Mass spectrometry experiments

Mass spectrometry experiments were kindly performed by Steven Lynham.

Enzymatic digestion

The monomeric I27_{E24C-K55C} protein samples were prepared by diluting 20 μl of protein (0.5 $\mu\text{g}/\mu\text{l}$) in 20 μl PBS pH 7.4 containing L-cysteine, homocysteine and N-acetyl cysteine at the same concentration as used in the single molecule experiments. In-solution digestion with trypsin of 5 μg of total protein monomers was performed prior to subsequent analysis by mass spectrometry. Trypsin digestion at a ratio of 1:20 (enzyme:substrate) was carried out for 2 hours at 37 °C.

Liquid chromatography-mass spectrometry

Chromatographic separations were performed using an EASY NanoLC system (ThermoFisherScientific, UK). Peptides from a total protein amount of

2 µg on column were resolved by reversed phase chromatography on a 75 µm C18 column using a three step linear gradient of acetonitrile in 0.1% formic acid. The gradient was delivered to elute the peptides at a flow rate of 300 nL/min over 60 min. The eluate was ionised by electrospray ionisation using an Orbitrap Velos Pro (ThermoFisherScientific, UK) operating under Xcalibur v2.2. The instrument was programmed to acquire in automated data-dependent switching mode, selecting precursor ions based on their intensity for sequencing by collision-induced fragmentation using a Top20 CID method and a Top10 high-energy collision dissociation (HCD) method. Both types of MS/MS collision dissociation analyses were conducted using collision energy profiles that were chosen based on the mass-to-charge ratio (m/z) and the charge state of the peptide.

UV spectrophotometric measurements for pK_a determination

UV spectrophotometric experiments were kindly performed by Marc Mora.

Absorbance measurements were performed with a JENWAY 6305 Spectrophotometer at room temperature. Absorbance measurements of L-cysteine and N-acetyl-L-cysteine at a final concentration of 250 µM in PBS buffer were performed at $\lambda = 240$ nm from acidic to alkaline pH values, by successive addition of small aliquots of sodium hydroxide. All absorbance measurements were performed at $\lambda = 240$ nm^{336, 337} at a final concentration of 250 µM with 1mM of tris(2-carboxyethyl)phosphine to prevent dimerization in PBS buffer. The basal absorbance of tris(2-carboxyethyl)phosphine at $\lambda = 240$ nm was subtracted from the total absorbance signal in the pK_a determination procedure. All samples were homogenized by strongly stirring immediately before each measurement and their pH value was measured with a Mettler Toledo pH-meter. The pK_a values were determined by fitting the Henderson-Hasselbach equation to the absorbance-pH data.

Mass spectrometry of l-cysteine and l-homocysteine

Mass spectrometry experiments were kindly performed by Marc Mora.

ESI-MS experiments were performed with a Thermo LTQ XL linear ion Trap mass spectrometer with electron transfer dissociation source. The electrospray voltage was kept at 4.00 KV, the capillary temperature was set at 350 °C and the capillary voltage was 19 V. A full MS (m/z 107-400) scan was acquired with a nominal mass resolution. Samples were delivered by means of flow injection. MS data was analyzed with Xcalibur version 2.0.7 software. L-homocysteine was freshly prepared right before the MS experiment at a pH of 9.2 and a final concentration of 63 μ M, and was confirmed by following the m/z molecular ion peak at 136 m/z . Dimerization to L-homocystine was induced by preparing a 63 μ M L-homocysteine solution at pH 9.2, and measured after 15h of sample preparation. L-homocystine was confirmed by following the m/z molecular ion peak at 269 m/z . Both species present in the same spectrum were obtained by bubbling the measuring solution with N₂ prior to MS determination.

Figures

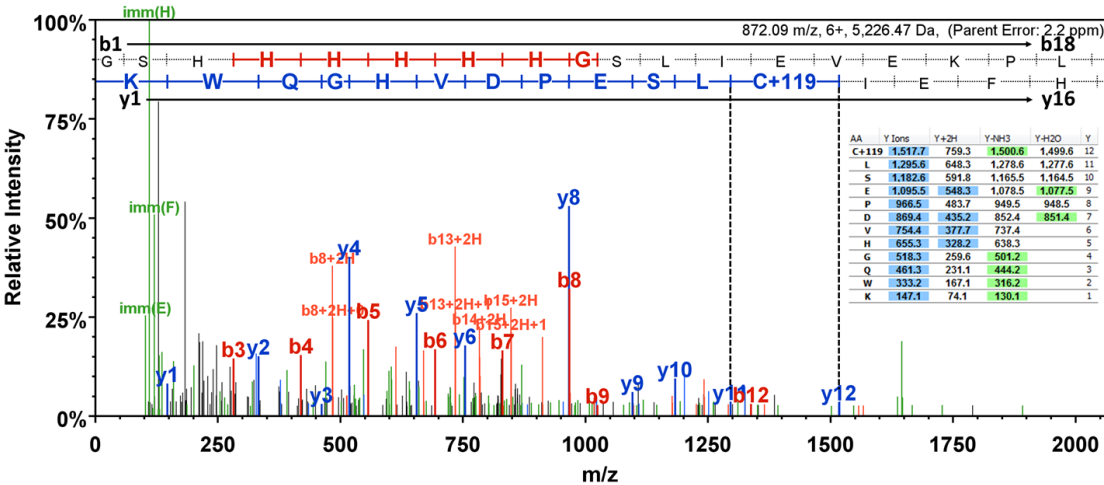


Figure 1. Evidence for a L-cysteine modification on a peptide with m/z 872.09⁶⁺. HCD fragmentation followed by database searching has assigned a strong sequence string of consecutive ions from b4-b9 and y1-y12, which accurately confirms the peptide identification. A modification of 119 Da corresponding to an additional L-cysteine was assigned to the y12-ion (Cys24).

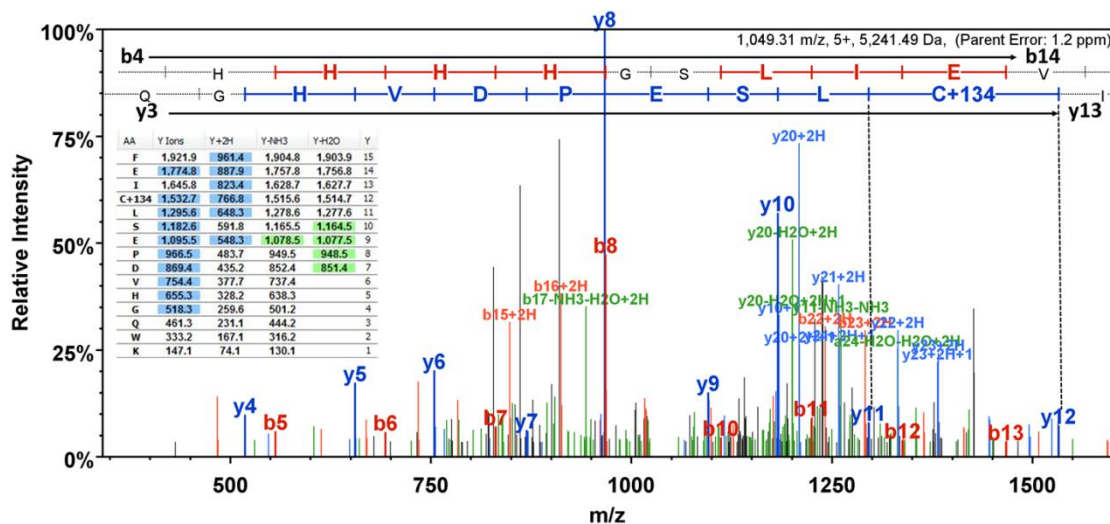


Figure 2. Evidence for a S-homocysteine modification on a peptide with m/z 1049.31⁵⁺. Evidence of a homocysteine modification on the cysteine-24 residue. A sequence tag of y-ions from y5-y12 and b-ions from b6-b8 and b11-b13 correctly identifies the peptide sequence of the protein at the N-terminus. An addition of 134 Da was detected on the Cys24 residue in the fragmentation spectra, which confirms homocysteine modification.

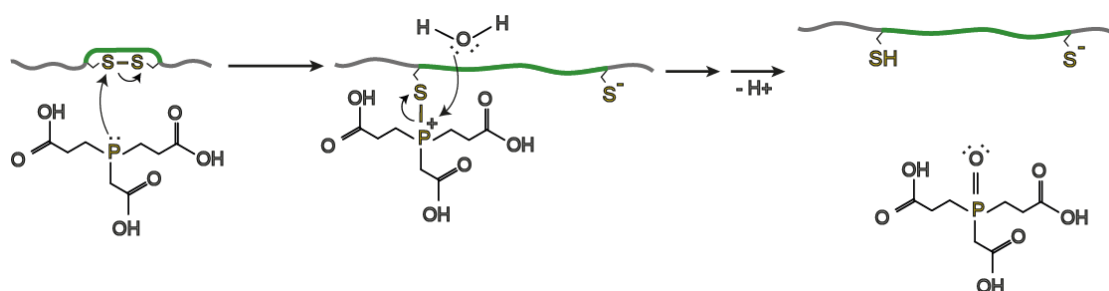


Figure 3. Scheme of the protein disulfide reactivity with TCEP. TCEP induces disulfide bond reduction through an S_N2 chemical reaction to create an unstable S-P thioalkoxyphosphonium cation that can be rapidly attacked by water molecules. The chemical attack results in a fully extended protein that harbours two reduced cysteines which, upon force withdrawal, can successfully refold into a reduced protein conformation devoid of the mechanically stiff disulfide bond.

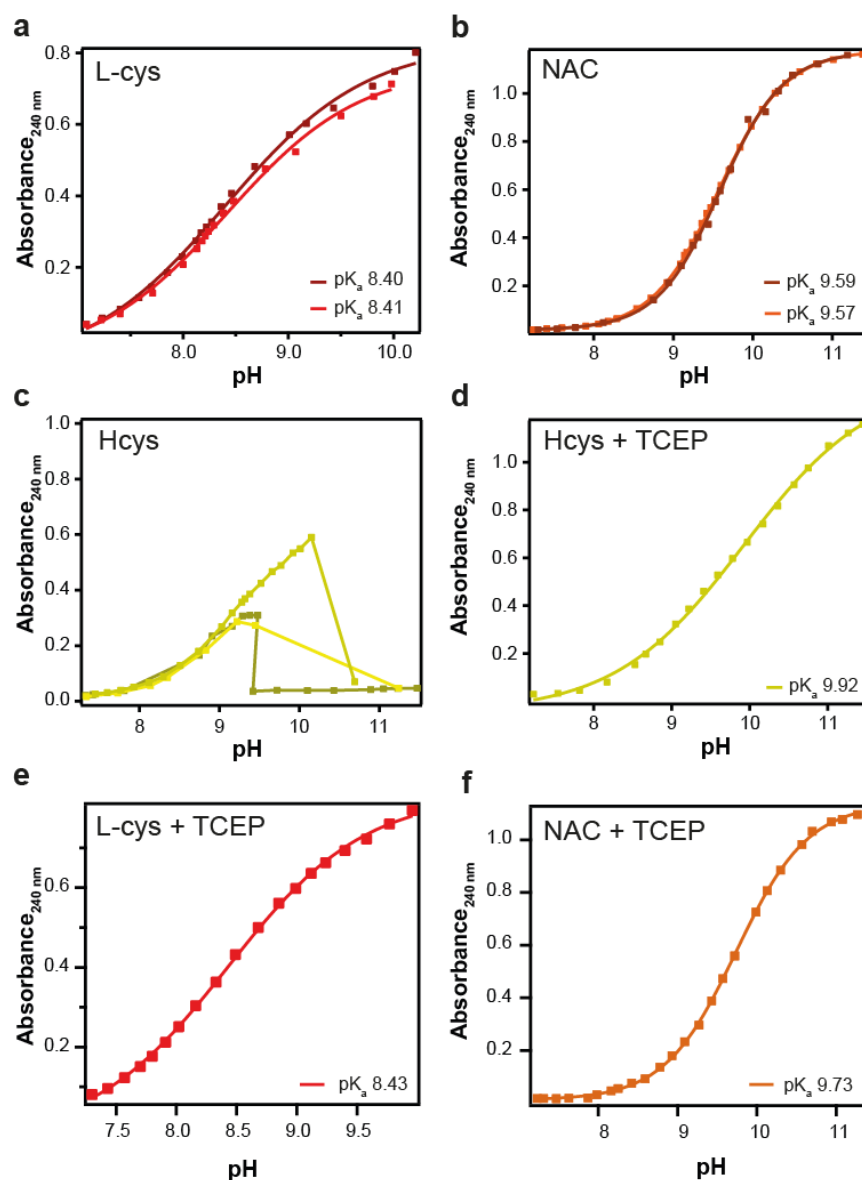


Figure 4. pK_a determination for the different low molecular weight thiols. The determination of the pK_a value for the different small thiols through absorbance measurements at $\lambda = 240$ nm proved highly reproducible within individual absorbance measurements as a function of pH. (a) A $pK_a = 8.40$ was measured for cysteine, (b) and a $pK_a = 9.6$ was obtained for NAC. (c) Homocysteine measurements showed a dramatic decrease in absorbance over time, especially at high pH values. Addition of 1 mM TCEP (highlighting the possible oxidation and homocysteine dimerization and precipitation) reversed the situation, allowing to measure a $pK_a = 9.9$. By contrast, adding 1 mM TCEP to the measuring solution did not affect the pK_a determination of (e) cysteine or (f) NAC, consistent with their monomeric state in solution.

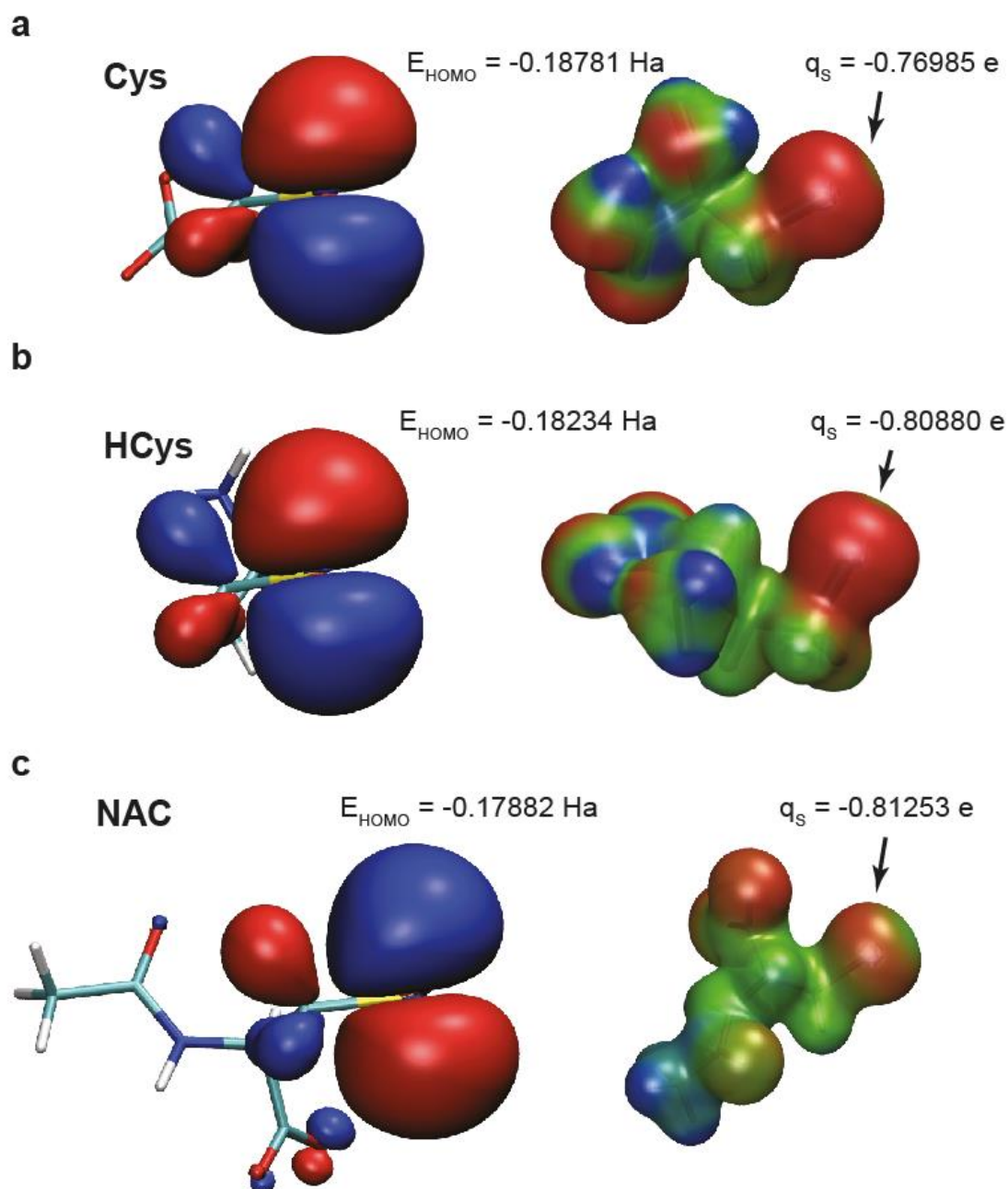


Figure 5. (a) HOMO calculations (left) and electrostatic potentials (from red: negative, to blue: positive) mapped on an electronic isodensity (0.02) surface (right), together with the charges the sulfur atom for (a) L-cysteine, (b) homocysteine and (c) NAC, highlighting the good nucleophilic power of homocysteine and NAC.

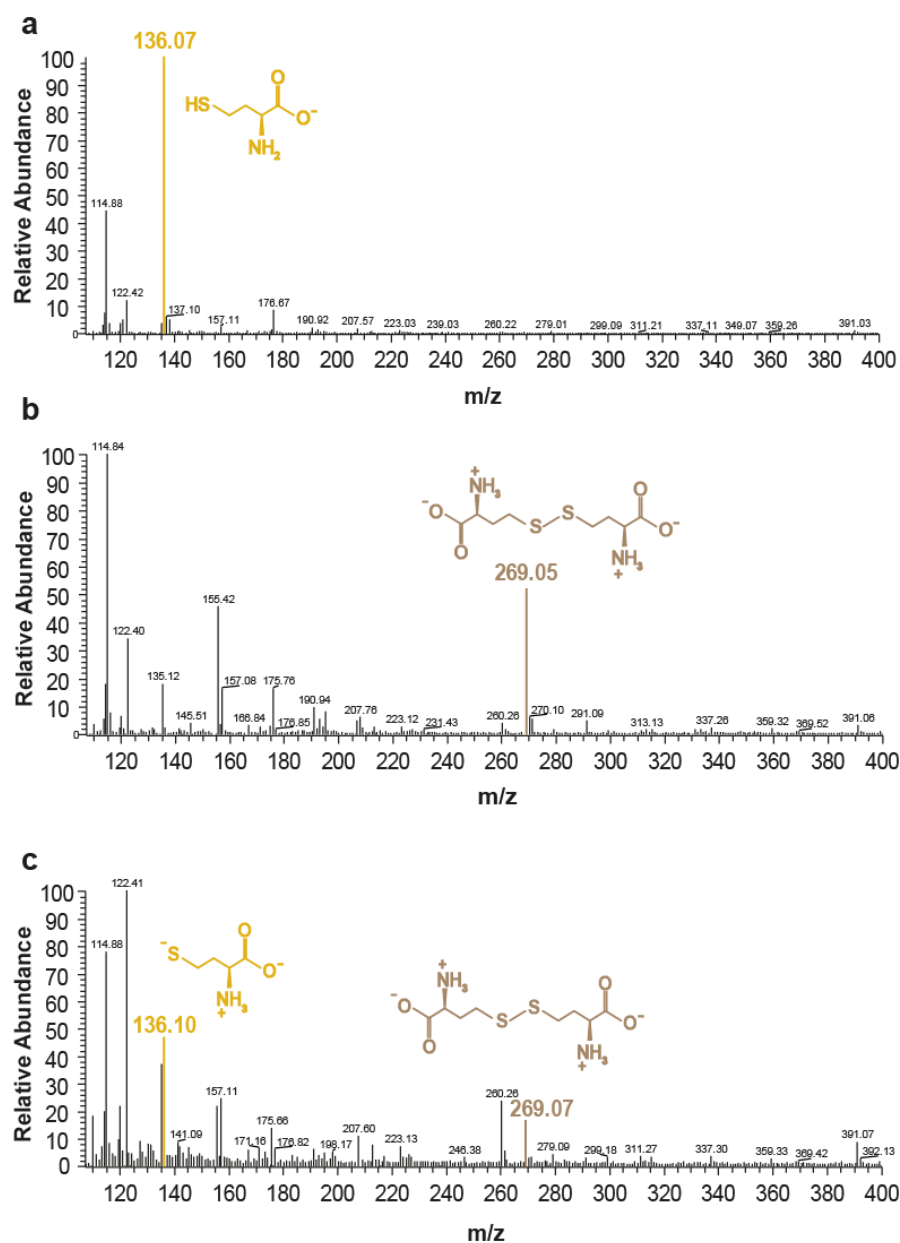


Figure 7. Solution homocysteine undergoes fast oxygen-mediated dimerization, demonstrating the high tendency to form stable disulfide bonds. (a) A 62.5 μM homocysteine solution at pH = 7 does not dimerize under ambient conditions for 15 hours. (b) By contrast, exposing the same solution for 3 hours at pH = 9 induced full dimerization to homocysteine (the homocysteine dimer). (c) Similar experiments in the presence of bubbling N_2 significantly delayed homocysteine appearance, thus suggesting a oxygen-induced homocysteine oxidation process. Crucially, homocysteine is highly insoluble. These experiments highlight the high tendency of homocysteine to create disulfide bonds.

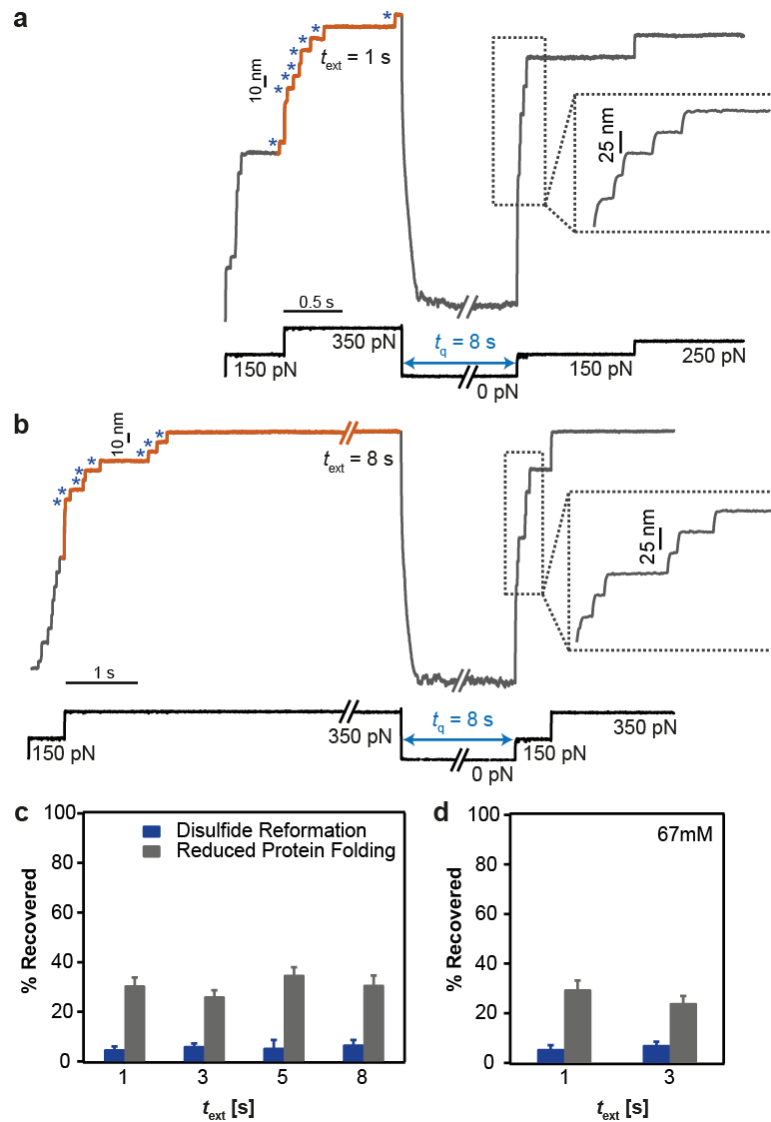


Figure 8. NAC-mediated evolution of oxidative and reduced folding over time. The time-evolution presence of 10 nm steps (signature of oxidative folding) and 25 nm steps (hallmarking successful refolding of the reduced protein) in the test pulse was tested by changing the time, t_{ext} , the reduced protein is left unfolded and stretched, allowing it to react with solution nucleophiles. Surprisingly, leaving the protein extended for a short (a) $t_{\text{ext}} = 1 \text{ s}$ and (b) long $t_{\text{ext}} = 8 \text{ s}$ does not significantly change the folding percentages (c). Similarly, changing the solution concentration of NAC ($\pm \text{s.d.}$) (d) does not have a significant effect on the folding fate of the protein. Hence, these experiments suggest that the reactivity of the unfolded and stretched protein occurs under thermodynamic—and not kinetic—control ($\pm \text{s.d.}$).

Table 1. Identification of peptide modifications by L-cysteine, Homocysteine and N-acetyl-cysteine. Mass spectrometry analysis by database searching and manual verification.

Sequence	m/z	Mr	Charge State	Intensity	Residue	Modification
² GSHHHHHH GSLIEVEKPL YGVEVFGVE TAHFEIcLSEP DVHGQWK ⁴⁶	872.09	5,226.47	6+	455100	Cys24	L-Cys
⁶⁶ cHILILHNA QLGMTGEVS FQAANTK ⁹⁰	939.13	2,814.36	3+	29730	Cys55	L-Cys
² GSHHHHHH GSLIEVEKPL YGVEVFGVE TAHFEIcLSEP DVHGQWK ⁴⁶	1049.31	5,241.49	5+	306700	Cys24	Hcys
⁶⁶ cHILILHNA QLGMTGEVS FQAANTK ⁹⁰	944.14	2,829.38	3+	144600	Cys55	Hcys
² GSHHHHHH GSLIEVEKPL YGVEVFGVE TAHFEIcLSEP DVHGQWK ⁴⁶	1049.3	5,241.48	5+	1118000	Cys24	Hcys
⁶⁶ cHILILHNA QLGMTGEVS FQAANTK ⁹⁰	566.88	2,829.38	5+	242200	Cys55	Hcys
⁶⁶ cHILILHNA QLGMTGEVS FQAANTK ⁹⁰	1,415.70	2,829.39	2+	504100	Cys55	Hcys
⁴⁷ LKGQPLAA SPDAEIIEDG KcHILILHNA QLGMTGEVS FQAANTKSA ANLK ⁹⁶	896.79	5,374.70	6+	33960	Cys55	NAC
² GSHHHHHH GSLIEVEKPL YGVEVFGVE TAHFEIcLSEP DVHGQWK ⁴⁶	879.26	5,269.50	6+	805900	Cys24	NAC

⁶⁶ cHILILHNA QLGMTGEVS FQAANTK ⁹⁰	715.35	2,857.37	4+	447700	Cys55	NAC
² GSHHHHHH GSLIEVEKPL YGVEVFGVE TAHFEIcLSEP DVHGQWK ⁴⁶	874.59	5,241.47	6+	530100	Cys24	Hcys pH 7.6
⁶⁶ cHILILHNA QLGMTGEVS FQAANTK ⁹⁰	708.35	2,829.39	4+	293400	Cys55	Hcys pH 7.6

APPENDIX IV

ARTICLE

Received 24 Dec 2014 | Accepted 24 Jun 2015 | Published 3 Aug 2015

DOI: 10.1038/ncomms8894

OPEN

The mechanochemistry of copper reports on the directionality of unfolding in model cupredoxin proteins

Amy E.M Beedle¹, Ainhua Lezamiz², Guillaume Stirnemann³ & Sergi Garcia-Manyes^{1,2}

Understanding the directionality and sequence of protein unfolding is crucial to elucidate the underlying folding free energy landscape. An extra layer of complexity is added in metalloproteins, where a metal cofactor participates in the correct, functional fold of the protein. However, the precise mechanisms by which organometallic interactions are dynamically broken and reformed on (un)folding are largely unknown. Here we use single molecule force spectroscopy AFM combined with protein engineering and MD simulations to study the individual unfolding pathways of the blue-copper proteins azurin and plastocyanin. Using the nanomechanical properties of the native copper centre as a structurally embedded molecular reporter, we demonstrate that both proteins unfold via two independent, competing pathways. Our results provide experimental evidence of a novel kinetic partitioning scenario whereby the protein can stochastically unfold through two distinct main transition states placed at the N and C termini that dictate the direction in which unfolding occurs.

¹Department of Physics, King's College London, London WC2R 2LS, UK. ²Randall Division of Cell and Molecular Biophysics, King's College London, London WC2R 2LS, UK. ³CNRS - Institut de Biologie Physico-Chimique - PSL Research University, Laboratoire de Biochimie Théorique, 75005 Paris, France. Correspondence and requests for materials should be addressed to S.G.-M. (email: sergi.garcia-manyes@kcl.ac.uk).

During folding, unfolded polypeptides progressively narrow down their conformational space search, often briefly visiting intermediate conformations reminiscent of localized energy minima^{1–3} until they eventually reach the functional natively folded structure^{4–8}. Understanding the molecular mechanisms underpinning each intermediate as well as identifying the sequential order statistics in which each conformation occurs along an individual (un)folding pathway, remains a challenge in the protein folding field².

Single molecule experiments are ideally suited to detect individual short-lived protein conformations, often occurring through statistically rare pathways that would be otherwise smeared out in the ensemble. In particular, force experiments excel at precisely mapping out individual mechanical (un)folding pathways by monitoring the protein extension along a well-defined one-dimension reaction coordinate; the protein's end-to-end length. These experiments have revealed signatures of complexity in the unfolding trajectories of a wide variety of proteins. While earlier works reported the presence of well-defined mechanical intermediates for a number of topologically distinct small proteins^{9–13}, more recent studies on larger multidomain proteins have experimentally demonstrated the concept of kinetic partitioning^{14–17}, where both the number and the position of these intermediates stochastically change after the main unfolding event, which remains the same in all trajectories^{18–24}. A further layer of complexity would be added if a protein also exhibited stochastic behaviour in the rupture of the main mechanical clamp, defining a multiplicity of pathways from the early stages of unfolding.

A key advantage of single molecule atomic force microscopy (AFM) mechanical experiments is the ability to directly measure the mechanical activation of individual chemical bonds. These measurements enable a direct correlation between the strength and nature of the bond and its mechanical stability. For example, while mechanical protein unfolding involves the disruption of a key set of hydrogen bonds, typically requiring forces ranging from ~20 to 200 pN (ref. 25), the mechanical stability of covalent bonds is significantly higher, up to ~3 nN (ref. 26). In particular, covalent-disulfide bonds cannot be broken on the application of force, unless a nucleophile is added to the solution²⁷. The repertoire of chemical interactions that are ruptured during the mechanical unfolding trajectories have recently been extended to include organometallic bonds^{22,28}. Metals play an important structural and biological role in a vast variety of proteins, spanning one-half of all known protein crystal structures in the Protein Data Bank. Hence, understanding the mechanochemistry of individual organometallic bonds and the kinetics of metal release from the protein core is essential when studying the mechanical (un)folding of metalloproteins. While pioneering recent studies have mainly focused on the mechanical properties of iron-binding proteins such as rubredoxin^{22,28,29}, the nanomechanics of other organometallic bonds such as copper, which is an essential metal in biological processes involved in a variety of redox reactions, remains vastly elusive.

Here we use a combination of protein engineering and single molecule force experiments to study the individual unfolding pathways of two model blue-copper proteins: azurin, involved in electron-transfer processes in the redox system of *Pseudomonas aeruginosa* and other bacteria³⁰, and plastocyanin^{31–33}, which functions as an electron carrier in the photosynthetic electron-transfer chain, shuttling electrons between the cytochrome b6f complex and Photosystem I in all higher plants as well as in a number of green algae and cyanobacteria such as *Phormidium laminosum*³⁴. Azurin is topologically rich, exhibiting a β -barrel structure organized in a Greek key motif, Fig. 1a,b (refs 35,36). Crucially, its structure harbours the three main families of

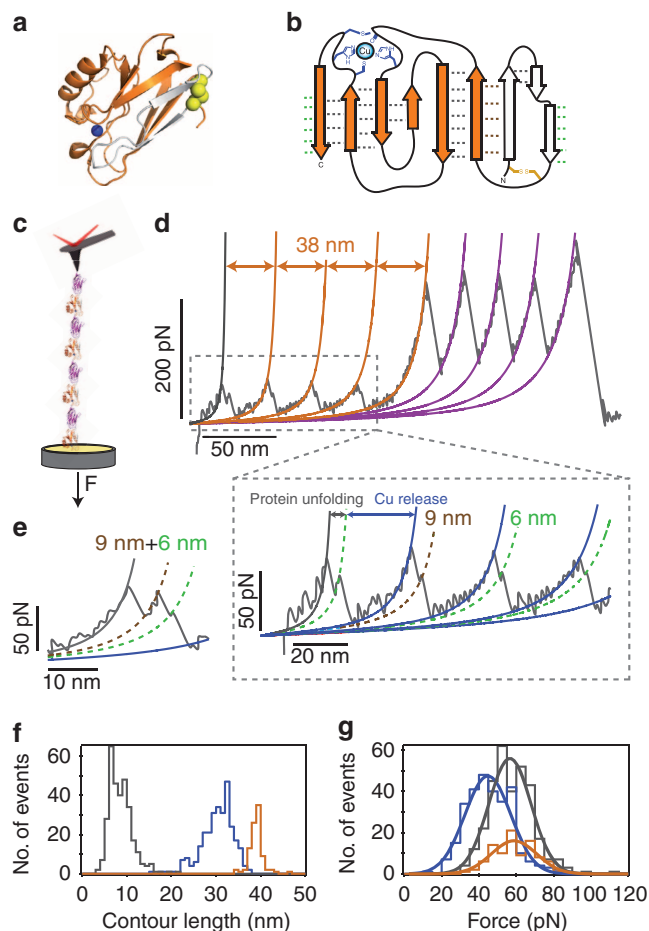


Figure 1 | Wt-azurin unfolds through two mechanically labile intermediates.

(a) Schematic structure of wt-azurin (PDB: 4AZU). (b) The residues trapped after the disulfide bond (yellow) are marked in white. The rest are represented in orange. (c) Scheme of the engineered (Azu-I27)₄ polyprotein being pulled in the AFM set-up. (d) Pulling on an individual (Azu-I27)₄ polyprotein results in an unfolding trajectory, where the unfolding of azurin (orange) occurs at ~55 pN, followed by the unfolding of the I27 domains, distinguished by an unfolding force of ~210 pN and an increment in contour length of $\Delta L = 28$ nm (purple). *Inset*: The unfolding of wt-azurin occurs via two well-defined mechanically stable intermediates occurring at ~6 nm (green dashed line) or ~9 nm (brown dashed line). (e) In a few trajectories ($n = 25$), a four-state unfolding pathway is measured, where both intermediates (~6 and ~9 nm) are captured within the same trajectory. (f,g) Worm-like chain fits to the data reveal that each complete azurin unfolding event occurs concomitant to an increment in contour length of $\Delta L = 38.6 \pm 0.9$ nm, $n = 93$ (f, orange histogram), requiring an unfolding force of 56.3 ± 12.1 pN, $n = 93$ (g, orange). When unfolding occurs via a mechanical intermediate, the main unfolding event (dashed lines), occurring at 54.1 ± 11.3 pN, $n = 323$ (g, grey), elicits a protein length of <10 nm (f, grey histogram) before it reaches a mechanically resistant intermediate. The mechanical disruption of such intermediate form occurs at forces of $\sim 42.3 \pm 12.3$ pN, $n = 289$ (g, blue), with a concomitant protein elongation of $\Delta L \sim 30$ nm (f, blue histogram).

chemical determinants that confer mechanical stability to proteins, namely, the presence of two sets of key hydrogen bonds at each of its termini, a penta-coordinated copper ion and a native covalent-disulfide bond. Such interesting structural diversity renders azurin particularly attractive from the mechanochemistry perspective. Using the native copper as an internal 'built-in' mechanical intermediate reporter, we demonstrate that azurin can unfold via two competing independent pathways.

We show that the protein can stochastically unfold along either the N–C or the C–N direction, suggesting that the barriers to unfolding from both termini are largely equivalent at the probed pulling velocity. Each individual unfolding pathway is unambiguously distinguished by the position of the subsequent rupture of the copper–ligand bond and further supported by directed mutagenesis and molecular dynamics (MD) simulations. Crucially, the mechanical stability of each individual organocopper bond is surprisingly low despite of their highly covalent character. These observations are expanded to the topologically similar plastocyanin protein, exhibiting analogous behaviour. Our results provide experimental evidence of a novel kinetic partitioning scenario, where the stochastic nature of the unfolding process originates already at the very early stages, as the protein chooses between two competing main unfolding transition states that precisely dictate the unravelling sense along the termini direction.

Results

The nanomechanics of azurin unfolding. Using single molecule spectroscopy AFM we stretched a single azurin polyprotein whereby each azurin monomer is flanked by two modules of the titin 27th Ig domain, which provide an internal molecular fingerprint³⁷ (Fig. 1c). A typical force-extension unfolding trajectory of the (Azu–I27)₄ polyprotein marks first the unfolding of the azurin proteins (orange), occurring at ~ 56 pN, followed by the unfolding of the I27 modules (purple), requiring ~ 210 pN to unfold (Fig. 1d) and exhibiting an increase in contour length of $\Delta L \sim 28$ nm (ref. 37). Fitting the worm-like chain model of polymer elasticity (WLC) to the data reveals that each azurin unfolding event entails a concomitant increase in contour length of $\Delta L \sim 38$ nm (orange fits). Assuming a contour length per residue of 0.38 nm (ref. 38), unfolding and extending each azurin monomer right from the termini would release $\Delta L \sim 48$ nm. However, a disulfide bond between the amino acids C3–C26 shortcuts the protein (Fig. 1a,b, white), and effectively shortens it by ~ 10 nm. Hence, the increase of $\Delta L \sim 38$ nm that we observed is perfectly compatible with the unfolding of the whole azurin protein from a position very close to the termini in the presence of the disulfide bond (orange region in Fig. 1a, Supplementary Fig. 1 and Supplementary Table 1).

Close inspection of each azurin unfolding event (Fig. 1d, inset) reveals, in $\sim 80\%$ of the occurrences, the presence of a clear mechanical intermediate (Supplementary Fig. 2). Remarkably, the position of this intermediate is not unique, since we measured (with similar probability) events occurring either ~ 6 nm (dashed green fit) or ~ 9 nm (dashed brown fit, Fig. 1d, inset) after the main unfolding event, occurring at ~ 54 pN (Fig. 1g, grey). Despite the close proximity in the measured length defining both intermediates, the grey histogram in Fig. 1f is bimodal and the presence of both intermediates is statistically significant, with $>99.9\%$ certainty (Hartigan's diptest, Supplementary Fig. 3). The mechanical disruption of such intermediate conformations occurs at forces of ~ 43 pN (Fig. 1g, blue), with a concomitant protein elongation of $\Delta L \sim 32$ or $\Delta L \sim 28$ nm, respectively (Fig. 1f, blue histogram). Remarkably, in 8% ($n=25$) of the total unfolding trajectories showing an intermediate, the protein unfolds via a four-state unfolding pathway, where both intermediates (~ 6 and ~ 9 nm) are readily captured (Fig. 1e). Crucially, in these cases the order of occurrence of both ~ 6 and ~ 9 nm events is stochastic. In the remaining $\sim 20\%$ of events, the protein unfolds through a simple two-state scenario, extending a total length of ~ 38 nm, requiring forces of ~ 56 pN (Fig. 1g, orange). To further validate the presence of the two distinct mechanical intermediates, we conducted experiments under force-clamp conditions,

whereby the protein extension is recorded as a function of time (Supplementary Fig. 4). These experiments confirmed the presence of two statistically distinct mechanical intermediates. Details on the detection of unfolding intermediates in constant velocity and force-clamp experiments are shown in Supplementary Figs 5 and 6, respectively.

According to the presence of the copper site in the tertiary structure of azurin, we speculate that the mechanical intermediates that we measured are associated with the mechanical rupture of the copper centre. In azurin (PDB: 4AZU), copper exhibits a characteristic trigonal bipyramidal geometry, whereby the metal interacts with five key residues; Met-121, His-117, Cys-112, His-46 and Gly-45 (Fig. 1b). Of these, structural and spectroscopic studies reveal that the three equatorial residues Cys-112, His-117 and His-46 exhibit the strongest interactions with the metal³¹. Taking this into consideration, the ~ 6 -nm intermediate would be compatible with an unfolding scenario where the protein unfolds from the C terminus starting by disrupting the β -sheets $\beta 2b$ – $\beta 8$ (Fig. 1b, green hydrogen bonds) up to the Cu–S_{Cys112} bond, hence releasing the amino acids contained between the C terminus (Lys-128) and Cys-112 (Supplementary Fig. 1 and Supplementary Table 1). However, the ~ 9 nm intermediate event cannot be accounted for by this unfolding sequence. By contrast, it could be readily explained assuming that the protein unfolds from the N terminus instead (Fig. 1b, brown hydrogen bonds). In that case, the position of the mechanical intermediate would match perfectly with the unfolding up to the Cu–N_{His46} bond. Indeed, extending the amino acids released on unfolding the protein from the N terminus up to His-46 (after deducting those residues that are trapped behind the disulfide bond formed between Cys-3 and Cys-26, white region in Fig. 1a,b) results in a length increment of ~ 11 nm (Supplementary Fig. 1 and Supplementary Table 1), consistent with the observed ~ 9 nm intermediate. Due to the weak character of the backbone carbonyl of Gly-45 (ref. 39), it is highly unlikely that the rupture of such interaction with the Cu centre can account for the mechanical intermediate observed in our experiments.

The distinct unfolding pathways revealed by mutagenesis. To verify the hypothesis that azurin unfolds via two competing pathways fingerprinted by the rupture of the Cu–S and Cu–N individual bonds, we first engineered a polyprotein containing the Cys112Ala mutant⁴⁰, thus removing the Cu–S_{Cys112} bond (Fig. 2a). Crucially, circular dichroism experiments certified the correct folding of the protein (Supplementary Fig. 7) and inductively coupled plasma–optical emission spectroscopy (ICP-OES) measurements revealed that, although only partially, Cu was still embedded within the azurin structure (Supplementary notes). Stretching the (Azu_{C112A}–I27)₄ polyprotein revealed the presence of only the ~ 9 nm intermediate (brown fits, Fig. 2b; and brown histogram, Fig. 2d). The absence of the ~ 6 nm intermediate in the Azu_{C112A} case firmly suggests that the ~ 6 nm intermediate observed in the unfolding trajectories of wildtype (wt)-Azu (Fig. 1d) is indeed triggered by the mechanical disruption of the Cu–S_{Cys112} bond.

To further investigate the origin of the ~ 9 nm intermediate, we studied the unfolding properties of the (Azu_{C26A}–I27)₄ polyprotein, lacking the native disulfide bond (Fig. 2c). The absence of the disulfide bond was further checked by the Ellman's assay^{41,42}, revealing an increase in the number of free sulphydryls in the polyprotein chain in the mutated protein as opposed to the wt-form (Supplementary notes). As expected, each azurin unfolding event showed a larger associated increment in contour

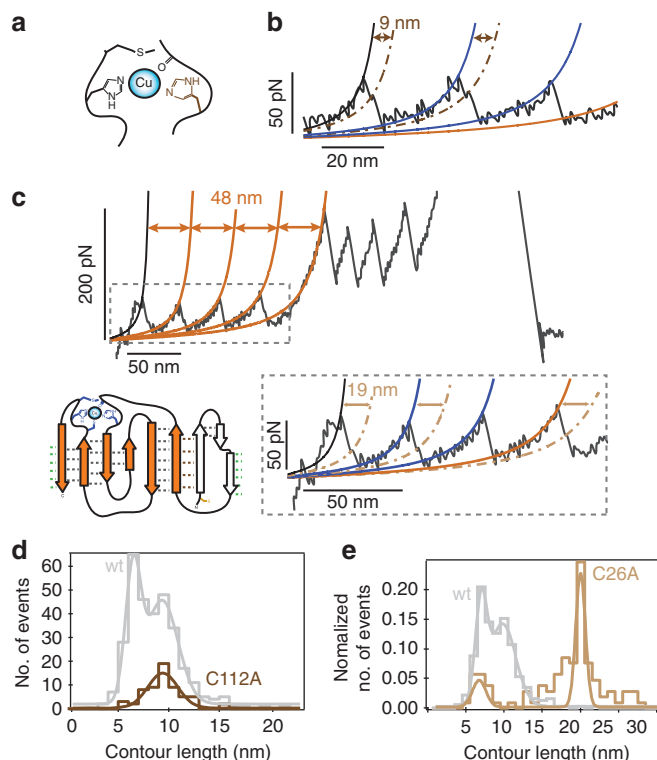


Figure 2 | The Azu_{C112A} and the Azu_{C26A} mutations confirm the existence of parallel unfolding pathways. (a) The Azu_{C112A} mutation prevents the formation of the Cu–S_{Cys112} bond. (b) On unfolding, the Azu_{C112A} protein extends by ~9 nm (dashed brown fit) before encountering a mechanical intermediate that corresponds to the disruption of the Cu–His₄₆ bond (dark blue fit). (c) The C26A mutation prevents the formation of the disulfide bond. In the absence of the disulfide bond, each azurin unfolding event occurs concomitant to an increase in contour length of $\Delta L = 47.6 \pm 1.2$ nm, $n = 77$ (orange fits). The mechanical intermediate is now mostly observed (80% of the occurrences) at $\Delta L = 19.0 \pm 0.5$ nm (dashed light brown fits), $n = 78$, consistent with the unfolding and extension of the total protein length from the N terminus up to His-46. In the remaining 20% of the cases in which the protein unfolds via a three-state pattern, the protein unfolds via the C terminus, as fingerprinted by the intermediate occurring at $\Delta L = 5.6 \pm 1.1$ nm, $n = 18$. (d) The histogram corresponding to the increment in contour length elicited after the main unfolding event measured for the Azu_{C112A} mutant reveals a main unfolding event occurring at $\Delta L = 9.4 \pm 1.5$ nm, $n = 62$ nm, which matches rather well the broader histogram of the wt-Azu data shown in Fig. 1f (background light grey bars), containing both intermediate events. (e) The normalized histogram corresponding to the unfolding of the Azu_{C26A} mutant shows well-defined intermediates, occurring at $\Delta L = 6.2 \pm 1.0$ nm, $n = 18$ and $\Delta L = 19.5 \pm 0.7$ nm, $n = 81$. Crucially, the N terminus pathway is highly favoured in the absence of the native disulfide bond.

length of $\Delta L \sim 48$ nm, corresponding to the full protein length (orange fits). Akin to the wt- case, the unfolding trajectories showed two well-defined intermediates; the first one occurring at ~6 nm and the second one now occurring at ~19 nm (dashed light brown fits). The distribution of events is shown in the contour length histogram of Fig. 2e (light brown). Hence, on removal of the disulfide bond, only the N terminus unfolding pathway was modified by adding to the length required to reach the His-46 the protein length previously trapped behind the disulfide bond resulting in the total length increase of ~19 nm, thus separating both populations of intermediate events. Interestingly, the relative abundance of both pathways for this disulfide bond-depleted construct is not ~50% as for the

wt- construct; rather, the ‘N-terminus’ pathway, eliciting the 19-nm intermediate, is observed in ~80% of the events featuring an intermediate, as observed in the normalized histogram shown in Fig. 2e.

The plastocyanin protein exhibits a similar unfolding scheme.

Azurin is a model protein for Type 1 (T1) copper centres. T1 copper proteins are distinct from other copper proteins because of their unique geometry and ligand sets³⁰. Despite the fact that the sequence identity between the T1 copper proteins is <20% (ref. 43), the overall structural fold of different T1 copper proteins is highly conserved³¹. Such common cupredoxin fold consists of eight β -strands arranged into a Greek β -barrel (Supplementary Fig. 8). Inspired by such large structural homology, we investigated whether the intriguing features observed in the mechanical unfolding pathway of azurin could be generalized to other structurally similar T1 model blue-copper proteins with the cupredoxin fold, such as plastocyanin (Fig. 3a). To this end, we studied the mechanical unfolding pathways of the (Plastocyanin-I27)₄ polypeptide (Fig. 3b). Complete unfolding of plastocyanin, occurring at ~75 pN (Fig. 3e), results in an increase of contour length of $\Delta L \sim 37$ nm (Fig. 3c, cyan fits and Supplementary Fig. 9), which is consistent with the extension of the full protein (Supplementary Table 2). Remarkably, in 60% of the unfolding events a mechanical intermediate was also observed (Fig. 3c, inset), which, as in the case of azurin, clustered in two main length populations at ~6 and ~14 nm (Supplementary Figs. 10–12 and Supplementary Table 2), giving rise to the—now well-separated—bimodal histogram shown in Fig. 3d. Akin to azurin, in plastocyanin the copper ion is bound to the sulfur of Cys-89 and to two nitrogens (from His-39 and His-92) in the trigonal plane, and to a methionyl sulfur (from Met-97) in the axial position at a relatively longer distance (Fig. 3f). Taking this structure into account, the ~6-nm intermediate can be readily explained in this case by the unfolding of the protein from the C terminus until Cys-89. In this case, after unfolding, the protein would be expected to extend 7.0 nm (Supplementary Fig. 9 and Supplementary Table 2). By contrast, unfolding the protein from the N terminus up to His-39 would extend 16.2 nm, thus consistent with the second intermediate that we observed. Crucially, mechanical disruption of the copper organometallic bonds requires forces of ~45 pN (Fig. 3g).

Molecular dynamics simulations. To gain further molecular insight into such enticing unfolding scheme, we conducted MD simulations of azurin and plastocyanin unfolding under a constant force³⁸. While present force fields are not yet adequate to reliably measure the rupture of organometallic bonds, which are constrained with harmonic potentials during the course of the simulation, the simulations provide a detailed molecular picture of the mechanisms underlying the first stages of protein unfolding, where a set of key hydrogen bonds are disrupted by the stretching force. Importantly, constriction of the organocopper bonds does not significantly influence the initial stages of protein unfolding (Supplementary Fig. 13). Individual azurin unfolding trajectories were generated by pulling on one terminal C $_{\alpha}$ (the other being fixed) under a constant force in the range of 250–350 pN (Fig. 4a) that guarantees that unfolding occurs on a timescale accessible to the simulation (~10 ns) as shown in Fig. 4b. In 90% of the cases ($n = 19$), the protein unfolded from the C terminus while in a single trajectory ($n = 1$), the protein was found to unfold from the N terminus instead (Fig. 4c and solid bars in Fig. 4d). In another individual trajectory the protein was observed to unfold from both ends simultaneously. When unfolding from the C terminus, the β 2b– β 8 strands unzip first,

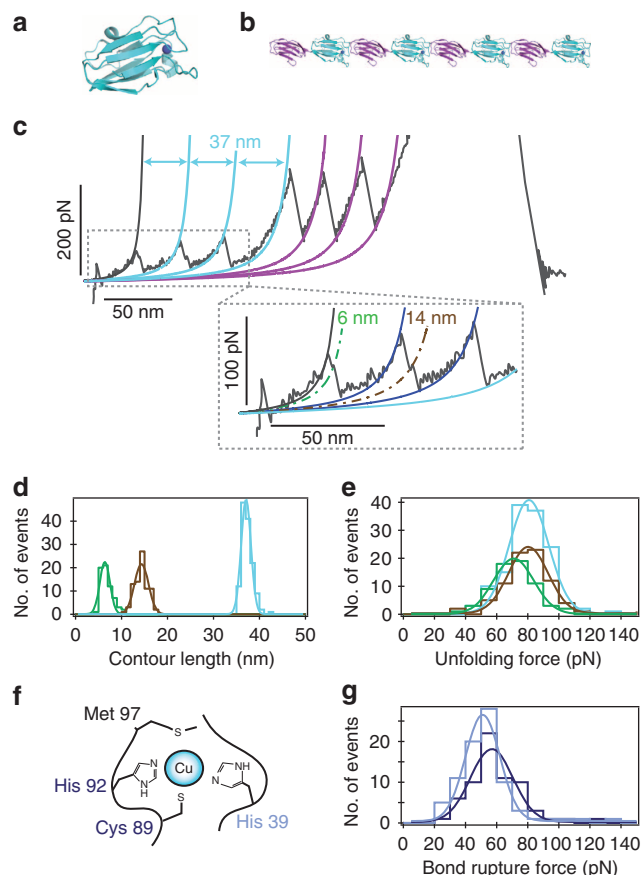


Figure 3 | Plastocyanin unfolds through parallel pathways fingerprinted by the rupture of individual Cu-S and Cu-N bonds. (a) Plastocyanin structure (PDB: 3BQV). (b) Scheme of the (Plastocyanin-I27)₄ polyprotein. (c) Individual constant velocity unfolding trajectory of a single (Plastocyanin-I27)₄ polyprotein, whereby the unfolding of each plastocyanin domain (cyan worm-like chain (WLC) fits) occurs before the unfolding of the I27 markers (purple WLC fits). *Inset*: Close inspection to each plastocyanin unfolding event reveals that the unfolding process occurs through two distinct mechanical intermediates, which are characterized (d) by an increase in contour length of either $\Delta L = 5.9 \pm 1.1$ nm, $n = 66$ (green) or $\Delta L = 13.8 \pm 1.5$ nm, $n = 82$ (brown). In the cases where no intermediate is observed, suggestive of an apoplastocyanin form, unfolding triggers the extension of the complete length of the protein, $\Delta L = 36.8 \pm 1.1$ nm, $n = 135$ (cyan). (e) Remarkably, the unfolding forces measured for each individual pathway (N terminus, C terminus and two-state unfolding) are very similar (75.3 ± 12 pN, $n = 82$, brown; 65.9 ± 13 pN, $n = 66$, green and 75.8 ± 12 pN, $n = 135$, cyan), respectively. In all cases, protein unfolding occurs at higher forces ~ 76 pN than the rupture of (f) the Cu-N_{His39} bond (46.0 ± 11 pN, $n = 82$, light blue) or the Cu-S_{Cys89} individual bond, occurring at 51.8 ± 13 pN, $n = 66$ (dark blue), as shown in (g).

followed by the disruption of $\beta 7$ – $\beta 8$. On the contrary, unfolding from the N terminus encompasses the rupture of the hydrogen bonds between three pairs of β -sheets, namely, $\beta 2b$ – $\beta 8$, $\beta 1$ – $\beta 3$ and $\beta 3$ – $\beta 6$. In sharp contrast with these results, pulling on the reduced protein (containing free cysteines, without the disulfide bond present) revealed unfolding trajectories whereby the protein mainly unfolded (80%, $n = 7$) from the N terminus (patterned bars in Fig. 4d), in close agreement with the experimental results (Fig. 2e).

Similarly, pulling on the wt-plastocyanin protein under a constant force of 250–350 pN revealed that, in 70% of the

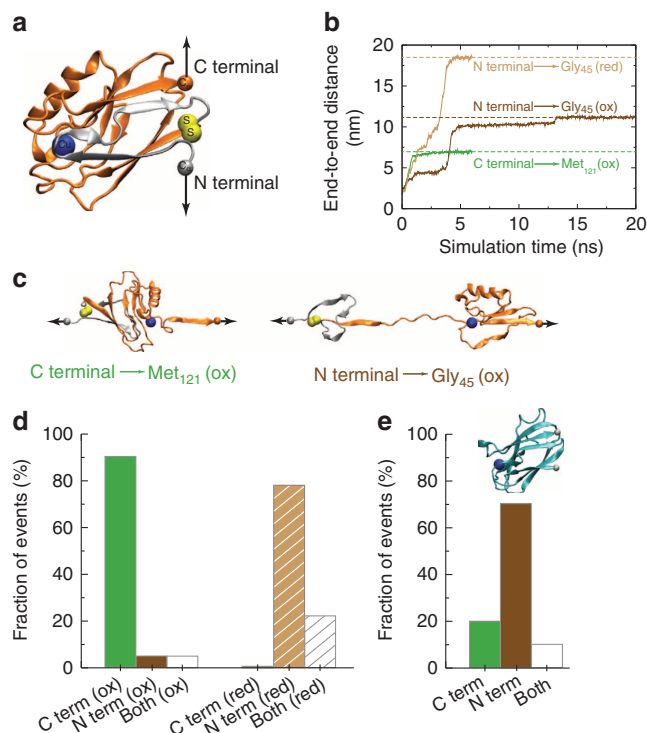


Figure 4 | MD simulations capture the distinct unfolding pathways.

(a) Azurin PDB structure (4AZU). Copper is explicitly shown as a blue sphere, the two sulfur atoms of the disulfide bond Cys3–Cys26 in yellow, the residues trapped after the disulfide bond in white and the rest, in orange. In a steered MD set-up, one of the termini is fixed in the laboratory frame and a constant force is applied to the other one. (b) Unfolding trajectories represented by the time evolution of the protein's end-to-end distance for three representative scenarios: wt-Azu unfolding from the N-terminus up to Gly-45 (brown), from the C-terminus up to Met-121 (green) and unfolding of wt-Azu lacking its disulfide bond from the N-terminus side up to Gly-45 (light brown). Each simulation was stopped once the end-to-end distance had reached the maximum possible extension at this force, the organometallic bonds being intact. (c) Representative structures following unfolding on the C-terminal side and up to Met-121 (left) and unfolding on the N-terminal side up to Gly-45 (right). (d) Percentage of the population of events falling in each scenario (green: C-terminal unfolding; brown: N-terminal; grey: both simultaneously), for both the oxidized form of wt-Azu (solid bars) and the reduced version lacking its disulfide bond (patterned bars). (e) Unfolding trajectories of wt-plastocyanin (3BQV). In $\sim 70\%$ of the trajectories ($n = 7$), plastocyanin unfolds through the N terminus (brown). In 20% of the occurrences ($n = 2$), the protein unfolds via the C terminus instead (green), and in the remaining trajectory the protein unfolds from both termini simultaneously (grey). Hence, from a qualitative viewpoint, these *in-silico* results support the experimental observations.

trajectories ($n = 7$), the protein unfolded from the N terminus, while in 20% of the trajectories ($n = 2$), the unfolding started at the C terminus instead (Fig. 4e). In an individual trajectory ($n = 1$), unfolding from both termini occurred concomitantly. Very similar ratios were observed for the case of apoplastocyanin (80/20/0%, respectively, $n = 10$), further validating the followed approach of constraining the organocopper bonds to study the initial stages of mechanical unfolding (Supplementary Fig. 13). Altogether, the *in-silico* unfolding trajectories demonstrate the unusual capability of both azurin and plastocyanin to mechanically unfold from both termini, lending strong support to the

experimental data. Precise information of each individual simulation run is detailed in the Supplementary Tables 3–6.

Discussion

Here we employed the mechanochemistry of copper to elucidate, at the single molecule level, the distinct mechanical unfolding pathways of the model T1 copper proteins azurin and plastocyanin. The active site of azurin contains a copper centre in a trigonally distorted tetrahedral geometry with a short Cu–S_{Cys112} bond (~ 2.1 Å) and two typical Cu–N_{His} bonds (~ 1.95 Å) in the equatorial plane³⁹. The axial position is occupied by a methionine at a relatively longer distance Cu–S_{Met121} (2.9 Å) (ref. 31). Finally, a weak-ionic interaction from an axial carbonyl oxygen (Gly-45) is also present. Similarly, in plastocyanin Cu(II) is coordinated by four atoms, the N^δ atoms of His-39 (~ 2.2 Å) and His-92 (~ 2.0 Å), the S^γ atom of Cys-89 (~ 2.1 Å), and the S^δ atom of Met-97 (~ 2.6 Å), forming a distorted tetrahedral geometry³⁴. Blue-copper proteins owe their name to a strong (~ 2.1 eV)⁴⁴ absorption around 600 nm associated with a S_{Cys112}, $p\pi \rightarrow Cu$ ligand to metal charge transfer transition (Supplementary Fig. 14). These particular spectral features are related to the highly covalent nature of the Cu–S bond^{45,46}, resulting from the strong interaction between the Cu 3d_{xy} acceptor orbital and the 3S $p\pi$ donor orbital^{31,47}. The high Cu–S stretching frequency of 400 cm^{-1} measured in Resonance Raman experiments supports the strong Cu–S_{Cys112} bond^{39,48}. S K-edge X-ray absorption spectroscopy results show that $\sim 38\%$ S character is mixed into the Cu 3d_{xy} orbital in the ground state³⁹. Moreover, the highly covalent copper-thiolate bond imparts a narrow four-line hyperfine splitting in the electron paramagnetic resonance spectra³¹. Such strong covalent character notwithstanding, the forces required to break the Cu–S_{Cys112} and the Cu–N_{His46} bonds (Cu–S_{Cys89} and Cu–N_{His39} in plastocyanin) are surprisingly low, of only ~ 45 pN (Supplementary Fig. 15). Interestingly, while the forces to unfold plastocyanin are slightly higher (~ 75 pN) than those required to unfold azurin (~ 55 pN), the mechanical stability of the organocopper bonds in both proteins is ~ 45 pN. These values are in sharp contrast with the high mechanical forces required to break homolytic covalent Si–Si bonds (2.1 nN) or heterolytic C–Si (2.0 nN) or Au–S (> 2.5 nN) bonds^{26,27,49}. The low mechanical forces required to disrupt the copper-ligand bonds in azurin and plastocyanin, together with the relatively low mechanical stability measured for the Fe–S (~ 211 pN; ref. 28) and Zn–S bonds (~ 170 pN; ref. 50) in rubredoxin, suggest that, from a mechanical perspective, the labile covalent organometallic bonds are significantly distinct from the mechanically resilient non-metallic covalent bonds such as Si–Si, S–S or Si–C bonds⁴⁹, despite their similar dissociation energies. These experimental findings call for computational efforts aiming at developing novel force fields that couple organometallic bonds with mechanical strain to help elucidation of the quantum-mechanical origin of the force-induced organometallic bond rupture.

The precise novel characterization of the nanomechanical properties of the copper-based organometallic bonds allowed us to uncover the complexity in the azurin and plastocyanin unfolding pathways when placed under mechanical force. In the case of azurin, we attribute the rupture of the unfolding intermediate to the breaking of the Cu–S_{Cys112} and Cu–N_{His46} bonds; however, no obvious intermediate was detected that was compatible with the rupture of the Cu–N_{His117} bond, which has been described using chemical denaturation techniques to be still formed even after azurin unfolding⁵¹. It is possible that due to the short distance (five residues) within Cys-112 and His-117 within the structure, the length resolution of our instrumentation makes

them indistinguishable, being all included in the data defining 6-nm intermediate. However, the observation that for the (Azur_{C112A}–I27)₄ mutant (Fig. 2d) the population of intermediate events at ~ 6 nm (or lower) was vanishingly small renders this option unlikely. Hence, it is plausible that on unfolding from the C terminus, the disruption of the $\beta 2b$ – $\beta 8$ sheets induces a conformational change in the protein loop connecting $\beta 7$ – $\beta 8$ strands that concomitantly distorts the active site structure thus perturbing the Cu–N_{His117} bond.

A remarkable finding in our experiments was the stochastic rupture of the main unfolding barrier in both studied blue-copper proteins, azurin and plastocyanin. Due to the particular β -barrel topology of the cupredoxin fold, two independent mechanical transition states, each presumably placed close to both of the termini, govern the mechanical stability of the proteins. Fortuitously, the height of the energy barrier to unfold both termini in each protein was found to be almost equivalent when a pulling velocity of 400 nm s^{-1} was employed, requiring a relatively low force of ~ 55 pN to unfold both termini in azurin (Supplementary Fig. 15), and ~ 75 pN to unfold both C and N termini in the case of plastocyanin (Fig. 3e). In fact, MD simulations on azurin clearly revealed that for both transition states, the rupture of the hydrogen bonds between the involved β -sheets occurs through an unzipping mechanism because the force is applied parallel to the hydrogen bonds axes (Fig. 4a). The precise dynamics of the unfolding sequence from both protein termini is detailed in the Supplementary Methods. These results put forward a rather unique global mechanical unfolding scenario for wt-azurin and wt-plastocyanin (Fig. 5). In 20% of the cases (40% in the case of plastocyanin), azurin unfolds in a single two-state event. In this case, we cannot pinpoint which mechanical transition state, either the N or C termini, breaks first. It is plausible that in some of these events, the resolution of our instrumentation was not able to detect the fast mechanical intermediates. However, by pulling in force clamp at forces as low as 20 pN (Supplementary Fig. 16), and hence readily expanding

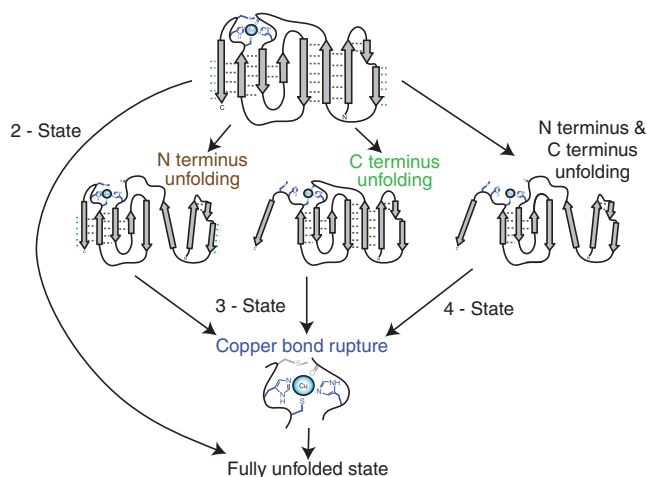


Figure 5 | A generic kinetic partition scheme for the complex mechanical unfolding of wt-azurin and wt-plastocyanin. On the application of mechanical force, azurin can unfold following multiple pathways. In 20% of the occurrences (40% in the case of plastocyanin), the protein unfolds according to an all-or-none process. By contrast, in 75% of the cases (50% for plastocyanin), azurin unfolds via one mechanically stable intermediate after unfolding from either the N terminus (brown) or the C terminus (green). In the remainder 5% (10%) of the cases, a four-state unfolding pathway involving the simultaneous unfolding of the N and C termini is observed.

the time-span of each trajectory, we could still capture both populations of events; those exhibiting a clear mechanical intermediate and those events where the unfolding was two-state, lacking any measurable intermediate. Indeed, ICP-OES measurements, which precisely measure the copper content in the protein sample, revealed that copper uptake was not total (Supplementary notes). We hence hypothesize that the observed two-state unfolding trajectories correspond to the unfolding of individual apo-azurin (or apoplastocyanin) modules. In 80% of the occurrences, azurin unfolds through a complex pathway involving at least one mechanical intermediate. Of these, in $\sim 45\%$ of the trajectories mechanical unfolding starts from the C terminus (green pathway), whereas in $\sim 45\%$ of the cases, unfolding proceeds instead through the N terminus (brown pathway). Finally, in the remainder $\sim 10\%$ of the occurrences, the protein unfolds first from one of the termini, then from the other. On a few occasions ($n = 16$) a unique 15 nm step corresponding to the concomitant unfolding of both termini is observed. Similarly, wt-plastocyanin follows a similar global unfolding scheme, with slightly different percentages corresponding to each unfolding pathway.

The observed similar probability of unfolding occurring from either the C- or the N terminus in wt-azurin is severely biased when the native disulfide bond is absent (Fig. 2e). In this case, unfolding mostly occurs ($\sim 80\%$) from the N terminus. MD simulations further confirm the experimental results by showing that, in 80% of the cases, the reduced wt-Azu lacking its disulfide bond (equivalent to the C26A mutant) unfolds indeed by breaking the N terminus β -sheets. Crucially, in our experiments we observed that wt-azurin can unfold from the N or C termini with almost equal probability. By contrast, our MD simulations show a drastic change in the pathway preference since 90% of the trajectories unfold from the C terminus. It is likely that such discrepancy can be explained by the slightly different distance to the transition state, Δx , associated with the unfolding from both C and N termini, giving rise to distinct force dependencies of unfolding. Since the force range probed in the *in-vitro* and *in-silico* experiments is significantly different, it is plausible that both approaches capture different relative unfolding probabilities at the distinct sampled pulling forces.

Altogether, the whole unfolding cascade is reminiscent of a complex kinetic partition scheme. In the recently reported examples for T4 Lysozyme¹⁸, Maltose Binding protein²¹ and Leucine Binding Protein²³ the main unfolding event was unique and always present, while mechanical intermediates further along the unfolding pathway were found to be stochastic and occurring in different positions. In the case of GFP, a bifurcation in the unfolding pathway was observed after the detachment of an N-terminal α -helix, fingerprinted by internal engineered disulfide bonds²⁴ or circular permutations⁵². By contrast, azurin and plastocyanin exhibit two distinct mechanical clamps placed very close to each terminus that exhibit a similar stability. Hence, on application of force, the protein stochastically crosses one of the two main barriers to unfolding. The directionality of the entire unfolding pathway is therefore determined by this early random event. Once the barrier is crossed, the protein unravels and gets held 'on pathway' at well-defined positions around the copper site. Taken together, the new unfolding scheme depicted for both blue-copper proteins follows a complex kinetic partition scheme in which the stochastic nature of the process is highlighted at the very beginning of the unfolding process with the rupture of the hydrogen bonds determining the main barrier for unfolding.

Our experiments have allowed us to pinpoint the unfolding directionality in the studied blue-copper proteins, occurring either from the C or N termini independently. The choice in the terminus that initiates unfolding might be of biological relevance,

as demonstrated by recent nanopore-based experiments, which have successfully recorded the distinct co-translocational unfolding dynamics depending on the terminus that is first presented to the nanopore mouth⁵³. Similarly, during degradation in the proteasome, proteins are mechanically unfolded before entering the narrow proteolytic chamber^{54,55}. The direction of entry in both processes depends on the force needed to unfold its two termini⁵⁶. When the N and C termini are mechanically non-equivalent, the protein will presumably tend to search for the weak link to initiate unfolding at the pore mouth⁵⁷. In this regard, proteins with dual mechanical transition states might provide an interesting case study for these applications, especially if one of the N or C termini can be tuned at will to act as the protein's Achilles' heel⁵⁸.

Altogether, our experiments demonstrate the power of single molecule techniques at revealing complex unfolding pathways in topologically rich proteins such as the case of metalloproteins, of high relevance in nature.

Methods

Protein engineering. The (Azu-I27)₄, (Azu_{C112A}-I27)₄, (Azu_{C26A}-I27)₄ and (Plasto-I27)₄ polyproteins were subcloned using the BamHI, BglII and KpnI restriction sites⁵⁹. Synthetic gene for plastocyanin was produced by Life Technologies (Paisley, UK). The multidomain proteins were cloned into the pQE80L (Qiagen) expression vector, and transformed into the BLR(DE3) *Escherichia coli* expression strain. All mutations were introduced by using the QuikChange Site-Directed Mutagenesis kit (Agilent). The cells were grown in LB broth supplemented with 100 $\mu\text{g ml}^{-1}$ ampicillin at 37 °C. After reaching an OD₆₀₀ of ~ 0.6 the cultures were induced with Isopropyl β -D-1-thiogalactopyranoside (1 mM) and incubated overnight at 20 °C (constructs containing Azurin) or 16 °C (Plastocyanin construct). For some constructs, CuSO₄ (0.5 mM) was supplemented to the cultures right before induction to increase copper intake. After harvesting the cells, disruption using a French Press was performed. The polyproteins from the lysate were purified by metal affinity chromatography on Talon resin (Clontech) followed by gel-filtration using a Superdex 200 10/300 GL column (GE Biosciences). Protein concentration in the samples was estimated using the Bradford protein assay.

Details on the biochemistry characterization assays regarding circular dichroism, Ellman's assays and ICP-OES measurements are detailed in the Supplementary Methods Section.

Force spectroscopy. Constant velocity AFM experiments were conducted at room temperature using both a home-made set-up⁶⁰ and a commercial Luigs and Neumann force spectrometer⁶¹. In all cases, the sample was prepared by depositing 1–10 μl of protein in PBS solution (at a concentration of 1–10 mg ml^{-1}) on a freshly evaporated gold coverslide. Each cantilever (Si₃N₄ Bruker MLCT-AUHW) was individually calibrated using the equipartition theorem, giving rise to a typical spring constant of ~ 12 –35 pN nm⁻¹. Single proteins were picked up from the surface and pulled at a constant velocity of 400 nm s⁻¹. Experiments were carried out in a sodium phosphate buffer solution, specifically, 50 mM sodium phosphate (Na₂HPO₄ and NaH₂PO₄), 150 mM NaCl, pH = 7.2.

Data analysis. All data were recorded and analysed using the custom software written in Igor Pro 6.0 (Wavemetrics). For all polyproteins, only recordings showing the signature of at least two events corresponding to the unfolding of the I27 fingerprints were analysed. Statistical tests were performed by applying Hartigan's diptest statistic for unimodality/multimodality^{62,63} and provided a test with simulation based *P* values based on a package compiled in R. Information about the data supporting this research and conditions of access can be found by emailing research.data@kcl.ac.uk.

MD simulations. Constant-force steered MD were performed on azurin starting from the PDB structure 4AZU solvated in a water box large enough in the direction of pulling to accommodate the partially unfolded protein ($6 \times 6 \times 18$ or $6 \times 6 \times 21$ nm). The NAMD 2.10 software was used to run the simulations on pure CPU or GPU/GPU nodes. Atomic parameters correspond to those in the CHARMM36 force-field⁶⁴, while atomic charges for copper II and atoms of the five ligands were taken from quantum calculations⁶⁵. We also constrained the organometallic bond distances as described elsewhere⁶⁵. Starting from a previously equilibrated protein whose CN axis was aligned along the longer edge of the box (direction of pulling), the C $_{\alpha}$ atom of the first residue was fixed while a constant force (250, 310 or 350 pN) was applied to the C $_{\alpha}$ atom of the last residue. In half of the trajectories, C $_{\alpha}$ from residue no. 1 was fixed and C $_{\alpha}$ from residue no. 128 was pulled; for the other half, C $_{\alpha}$ of residue no. 128 was fixed and C $_{\alpha}$ from residue no. 1 was pulled in the opposite direction. Twenty-one such trajectories, lasting between 6 and 20 ns (until full possible extension was reached), were generated for the

protein containing the disulfide bond (10 or 11 in each direction) in a ~ 18 -nm long box; nine trajectories total were generated in the absence of disulfide bond, in a slightly larger box (~ 21 nm). A very similar methodology was employed to simulate plastocyanin (3BQV, $n = 10$ trajectories—5 in each direction) and apoplastocyanin (2PCY, $n = 10$ trajectories—5 in each direction). As the active site of plastocyanin is very similar to that of azurin, we used the same set of atomic charges and bond potentials, albeit at slightly different equilibrium distances corresponding to that measured in the crystal structure.

References

- Onuchic, J. N. & Wolynes, P. G. Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70–75 (2004).
- Brockwell, D. J. & Radford, S. E. Intermediates: ubiquitous species on folding energy landscapes? *Curr. Opin. Struct. Biol.* **17**, 30–37 (2007).
- Bryngelson, J. D. & Wolynes, P. G. Intermediates and barrier crossing in a random energy-model (with applications to protein folding). *J. Phys. Chem.* **93**, 6902–6915 (1989).
- Bryngelson, J. D. & Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl Acad. Sci. USA* **84**, 7524–7528 (1987).
- Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19 (1997).
- Lazaridis, T. & Karplus, M. "New view" of protein folding reconciled with the old through multiple unfolding simulations. *Science* **278**, 1928–1931 (1997).
- Fersht, A. R. & Daggett, V. Protein folding and unfolding at atomic resolution. *Cell* **108**, 573–582 (2002).
- Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. Navigating the folding routes. *Science* **267**, 1619–1620 (1995).
- Marszalek, P. E. *et al.* Mechanical unfolding intermediates in titin modules. *Nature* **402**, 100–103 (1999).
- Bertz, M. & Rief, M. Mechanical unfoldons as building blocks of maltose-binding protein. *J. Mol. Biol.* **378**, 447–458 (2008).
- Schwaiger, I., Kardinal, A., Schleicher, M., Noegel, A. A. & Rief, M. A mechanical unfolding intermediate in an actin-crosslinking protein. *Nat. Struct. Mol. Biol.* **11**, 81–85 (2004).
- Schwaiger, I., Schleicher, M., Noegel, A. A. & Rief, M. The folding pathway of a fast-folding immunoglobulin domain revealed by single-molecule mechanical experiments. *EMBO Rep.* **6**, 46–51 (2005).
- Li, L. W., Huang, H. H. L., Badilla, C. L. & Fernandez, J. M. Mechanical unfolding intermediates observed by single-molecule force spectroscopy in a fibronectin type III module. *J. Mol. Biol.* **345**, 817–826 (2005).
- Thirumalai, D., Klimov, D. K. & Woodson, S. A. Kinetic partitioning mechanism as a unifying theme in the folding of biomolecules. *Theor. Chem. Acc.* **96**, 14–22 (1997).
- Kiefhaber, T. Kinetic traps in lysozyme folding. *Proc. Natl Acad. Sci. USA* **92**, 9029–9033 (1995).
- Wright, C. F., Lindorff-Larsen, K., Randles, L. G. & Clarke, J. Parallel protein-unfolding pathways revealed and mapped. *Nat. Struct. Biol.* **10**, 658–662 (2003).
- Zaidi, F. N., Nath, U. & Udgaonkar, J. B. Multiple intermediates and transition states during protein unfolding. *Nat. Struct. Biol.* **4**, 1016–1024 (1997).
- Peng, Q. & Li, H. Atomic force microscopy reveals parallel mechanical unfolding pathways of T4 lysozyme: evidence for a kinetic partitioning mechanism. *Proc. Natl Acad. Sci. USA* **105**, 1885–1890 (2008).
- He, C., Genchev, G. Z., Lu, H. & Li, H. Mechanically untying a protein slipknot: multiple pathways revealed by force spectroscopy and steered molecular dynamics simulations. *J. Am. Chem. Soc.* **134**, 10428–10435 (2012).
- Gao, X. *et al.* Single-molecule experiments reveal the flexibility of a Per-ARNT-Sim domain and the kinetic partitioning in the unfolding pathway under force. *Biophys. J.* **102**, 2149–2157 (2012).
- Aggarwal, V. *et al.* Ligand-modulated parallel mechanical unfolding pathways of maltose-binding proteins. *J. Biol. Chem.* **286**, 28056–28065 (2011).
- Zheng, P., Takayama, S. J., Mauk, A. G. & Li, H. B. Single Molecule Force Spectroscopy Reveals That Iron Is Released from the Active Site of Rubredoxin by a Stochastic Mechanism. *J. Am. Chem. Soc.* **135**, 7992–8000 (2013).
- Kotamarthi, H. C., Sharma, R., Narayan, S., Ray, S. & Aivarapu, S. R. K. Multiple unfolding pathways of leucine binding protein (LBP) probed by single-molecule force spectroscopy (SMFS). *J. Am. Chem. Soc.* **135**, 14768–14774 (2013).
- Mickler, M. *et al.* Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations. *Proc. Natl Acad. Sci. USA* **104**, 20268–20273 (2007).
- Galera-Prat, A., Gomez-Sicilia, A., Oberhauser, A. F., Cieplak, M. & Carrion-Vazquez, M. Understanding biology by stretching proteins: recent progress. *Curr. Opin. Struct. Biol.* **20**, 63–69 (2010).
- Beyer, M. K. & Clausen-Schaumann, H. Mechanochemistry: The mechanical activation of covalent bonds. *Chem. Rev.* **105**, 2921–2948 (2005).
- Garcia-Manyes, S., Liang, J., Szoszkiewicz, R., Kuo, T. L. & Fernandez, J. M. Force-activated reactivity switch in a bimolecular chemical reaction. *Nat. Chem.* **1**, 236–242 (2009).
- Zheng, P. & Li, H. B. Highly covalent ferric-thiolate bonds exhibit surprisingly low mechanical stability. *J. Am. Chem. Soc.* **133**, 6791–6798 (2011).
- Zheng, P., Takayama, S. I. J., Mauk, A. G. & Li, H. B. Hydrogen bond strength modulates the mechanical strength of ferric-thiolate bonds in rubredoxin. *J. Am. Chem. Soc.* **134**, 4124–4131 (2012).
- Solomon, E. I. *et al.* Copper active sites in biology. *Chem. Rev.* **114**, 3659–3853 (2014).
- Liu, J. *et al.* Metalloproteins containing cytochrome, iron-sulfur, or copper redox centers. *Chem. Rev.* **114**, 4366–4469 (2014).
- Redinbo, M. R., Yeates, T. O. & Merchant, S. Plastocyanin: structural and functional analysis. *J. Bioenerg. Biomembr.* **26**, 49–66 (1994).
- Gross, E. L. Plastocyanin: structure and function. *Photosynth. Res.* **37**, 103–116 (1993).
- Bond, C. S. *et al.* The structure of plastocyanin from the cyanobacterium *Phormidium laminosum*. *Acta. Crystallogr. D Biol. Crystallogr.* **55**, 414–421 (1999).
- Rizzuti, B., Daggett, V., Guzzi, R. & Sportelli, L. The early steps in the unfolding of azurin. *Biochemistry* **43**, 15604–15609 (2004).
- Sandberg, A., Leckner, J. & Karlsson, B. G. Apo-azurin folds via an intermediate that resembles the molten-globule. *Protein Sci.* **13**, 2628–2638 (2004).
- Carrion-Vazquez, M. *et al.* Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl Acad. Sci. USA* **96**, 3694–3699 (1999).
- Stirnemann, G., Giganti, D., Fernandez, J. M. & Berne, B. J. Elasticity, structure, and relaxation of extended proteins under force. *Proc. Natl Acad. Sci. USA* **110**, 3847–3852 (2013).
- Sarangi, R. *et al.* Spectroscopic and density functional theory studies of the blue-copper site in M121SeM and C112SeC azurin: Cu-Se versus Cu-S bonding. *J. Am. Chem. Soc.* **130**, 3866–3877 (2008).
- Sandberg, A., Leckner, J., Shi, Y., Schwarz, F. P. & Karlsson, B. G. Effects of metal ligation and oxygen on the reversibility of the thermal denaturation of *Pseudomonas aeruginosa* azurin. *Biochemistry* **41**, 1060–1069 (2002).
- Ellman, G. L. A colorimetric method for determining low concentrations of mercaptans. *Arch. Biochem. Biophys.* **74**, 443–450 (1958).
- Riener, C. K., Kada, G. & Gruber, H. J. Quick measurement of protein sulfhydryls with Ellman's reagent and with 4,4'-dithiodipyridine. *Anal. Bioanal. Chem.* **373**, 266–276 (2002).
- De Rienzo, F., Gabdoulline, R. R., Menziani, M. C. & Wade, R. C. Blue copper proteins: a comparative analysis of their molecular interaction properties. *Protein Sci.* **9**, 1439–1454 (2000).
- Do, H., Deeth, R. J. & Besley, N. A. Computational study of the structure and electronic circular dichroism spectroscopy of blue copper proteins. *J. Phys. Chem. B* **117**, 8105–8112 (2013).
- Solomon, E. I., Szilagyi, R. K., DeBeer George, S. & Basumallick, L. Electronic structures of metal sites in proteins and models: contributions to function in blue copper proteins. *Chem. Rev.* **104**, 419–458 (2004).
- Solomon, E. I. & Hadt, R. G. Recent advances in understanding blue copper proteins. *Coord. Chem. Rev.* **255**, 774–789 (2011).
- Solomon, E. I., Hare, J. W. & Gray, H. B. Spectroscopic studies and a structural model for blue copper centers in proteins. *Proc. Natl Acad. Sci. USA* **73**, 1389–1393 (1976).
- Han, J. *et al.* Resonance Raman spectra of plastocyanin and pseudoazurin: evidence for conserved cysteine ligand conformations in cupredoxins (blue copper proteins). *Biochemistry* **30**, 10904–10913 (1991).
- Grandbois, M., Beyer, M., Rief, M., Clausen-Schaumann, H. & Gaub, H. E. How strong is a covalent bond? *Science* **283**, 1727–1730 (1999).
- Zheng, P. & Li, H. B. Direct measurements of the mechanical stability of zinc-thiolate bonds in rubredoxin by single-molecule atomic force microscopy. *Biophys. J.* **101**, 1467–1473 (2011).
- Pozdnyakova, I., Guidry, J. & Wittung-Stafshede, P. Probing copper ligands in denatured *Pseudomonas aeruginosa* azurin: unfolding His117Gly and His46Gly mutants. *J. Biol. Inorg. Chem.* **6**, 182–188 (2001).
- Perez-Jimenez, R., Garcia-Manyes, S., Aivarapu, S. R. & Fernandez, J. M. Mechanical unfolding pathways of the enhanced yellow fluorescent protein revealed by single molecule force spectroscopy. *J. Biol. Chem.* **281**, 40010–40014 (2006).
- Rodriguez-Larrea, D. & Bayley, H. Protein co-translocational unfolding depends on the direction of pulling. *Nat. Commun.* **5**, 4841 (2014).
- Aubin-Tam, M. E., Olivares, A. O., Sauer, R. T., Baker, T. A. & Lang, M. J. Single-molecule protein unfolding and translocation by an ATP-fueled proteolytic machine. *Cell* **145**, 257–267 (2011).
- Maillard, R. A. *et al.* ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell* **145**, 459–469 (2011).
- Berko, D. *et al.* The direction of protein entry into the proteasome determines the variety of products and depends on the force needed to unfold its two termini. *Mol. Cell* **48**, 601–611 (2012).
- Rodriguez-Larrea, D. & Bayley, H. Multistep protein unfolding during nanopore translocation. *Nat. Nanotechnol.* **8**, 288–295 (2013).
- Matouschek, A. & Bustamante, C. Finding a protein's Achilles heel. *Nat. Struct. Biol.* **10**, 674–676 (2003).

59. Carrion-Vazquez, M. *et al.* Mechanical and chemical unfolding of a single protein: A comparison. *Proc Natl Acad. Sci. USA* **96**, 3694–3699 (1999).
60. Schlierf, M., Li, H. & Fernandez, J. M. The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proc. Natl Acad. Sci. USA* **101**, 7299–7304 (2004).
61. Popa, I., Kosuri, P., Alegre-Cebollada, J., Garcia-Manyes, S. & Fernandez, J. M. Force dependency of biochemical reactions measured by single-molecule force-clamp spectroscopy. *Nat. Prot.* **8**, 1261–1276 (2013).
62. Hartigan, P. M. Computation of the dip statistic to test for unimodality. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **34**, 320–325 (1985).
63. Hartigan, J. A. & Hartigan, P. M. The dip test of unimodality. *Ann. Stat.* **13**, 70–84 (1985).
64. Huang, J. & MacKerell, Jr A. D. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
65. van den Bosch, M. *et al.* Calculation of the redox potential of the protein azurin and some mutants. *Chembiochem* **6**, 738–746 (2005).

Acknowledgements

We thank GW Canters, Leiden Institute of Physics (The Netherlands) for kindly sharing the azurin plasmid. We thank Judit Perales-Calvo for help in cloning the plastocyanin polyprotein and Josep Relat-Goberna for help in software development. A.E.M.B. is funded by an EPSRC DTP fellowship. GS acknowledges support from GENCI for HPC resources (DARI project x2015077399). This work was supported by the Marie Curie CIG (293462) grant, BBSRC (J00992X/1) grant, Royal Society Research grant (RG120038) and EPSRC Fellowship (K00641X/1) to S.G.-M.

Author contributions

A.E.M.B and S.G.-M. designed research; A.E.M.B performed single molecule mechanical experiments; A.L. conducted molecular biology work and biochemistry experiments; A.E.M.B analysed experimental data; G.S. performed and analysed MD simulations and A.E.M.B and S.G.-M wrote the paper. All authors contributed in revising and editing the manuscript.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interest.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Beedle, A. E. M. *et al.* The mechanochemistry of copper reports on the directionality of unfolding in model cupredoxin proteins. *Nat. Commun.* 6:7894 doi: 10.1038/ncomms8894 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

APPENDIX V

ARTICLE

Received 11 Dec 2015 | Accepted 8 Jul 2016 | Published 22 Aug 2016

DOI: 10.1038/ncomms12490

OPEN

Protein S-sulfenylation is a fleeting molecular switch that regulates non-enzymatic oxidative folding

Amy E.M Beedle¹, Steven Lynham² & Sergi Garcia-Manyes¹

The post-translational modification S-sulfenylation functions as a key sensor of oxidative stress. Yet the dynamics of sulfenic acid in proteins remains largely elusive due to its fleeting nature. Here we use single-molecule force-clamp spectroscopy and mass spectrometry to directly capture the reactivity of an individual sulfenic acid embedded within the core of a single Ig domain of the titin protein. Our results demonstrate that sulfenic acid is a crucial short-lived intermediate that dictates the protein's fate in a conformation-dependent manner. When exposed to the solution, sulfenic acid rapidly undergoes further chemical modification, leading to irreversible protein misfolding; when cryptic in the protein's microenvironment, it readily condenses with a neighbouring thiol to create a protective disulfide bond, which assists the functional folding of the protein. This mechanism for non-enzymatic oxidative folding provides a plausible explanation for redox-modulated stiffness of proteins that are physiologically exposed to mechanical forces, such as cardiac titin.

¹Department of Physics and Randall Division of Cell and Molecular Biophysics, King's College London, London WC2R 2LS, UK. ²Centre of Excellence for Mass Spectrometry, King's College London, London SE5 8AF, UK. Correspondence and requests for materials should be addressed to S.G.-M. (email: sergi.garcia-manyes@kcl.ac.uk).

Protein folding defines a molecular self-assembly process that is finely tuned by a subtle interplay between enthalpic and entropic contributions^{1–3}. An extra layer of complexity is added when a native covalent disulfide bond needs to reform along the folding pathway^{4,5}. Other chemical alterations, such as post-translational modifications, pose further challenges to the successful completion of the folding process.

Cysteine is a principal target for post-translational modifications, especially those related to oxidative stress⁶. Due to the presence of empty *d*-orbitals ready for bonding, the highly polarizable sulfur atom can display oxidation states ranging from -2 to $+6$ (refs 7,8). Cysteine sulfenic acid (Cys-SOH) is the first oxidation step of cysteinyl thiols (Cys-SH)⁹ and the resulting post-translational modification, cysteine S-sulfenylation, has been identified as a redox sensor in an increasing number of proteins^{10,11}. However, the unusually high reactivity of -SOH—a consequence of its dual electrophilic and nucleophilic character^{7,10,12}—, together with its invisibility to ultraviolet-visible and fluorescent spectroscopic detection¹³, has rendered its characterization challenging. In general, stabilization of protein sulfenic acid is highly dependent on the protein microenvironment¹¹, following the general notion that solvent-exposed cysteines feature increased chemical lability¹⁴.

Identification of protein sulfenic acids has been mostly indirect, through chemical labelling with 5,5-dimethyl-1,3-cyclohexadione (dimedone)¹⁵, a mass-shifting molecular probe that specifically traps -SOH moieties and is easily identifiable with mass spectrometry (MS)^{16–18}. Despite this phenomenal progress¹², given the irreversible nature of the detection methods, which interfere with the intrinsically fast reactivity of Cys-SOH, direct identification of the dynamics of a particular protein -SOH and its implication on folding remains largely elusive.

Under oxidative conditions, Cys-SOH can be readily oxidized to more stable products, namely sulfinic (Cys-SO₂H) or sulfonic (Cys-SO₃H) acids⁶. The transition to these higher oxidation state species has been generally regarded as irreversible, and a marker for oxidative damage. Such deleterious pathway can be avoided if the electrophilic sulfur atom in -SOH reacts with an intramolecular neighbouring protein thiol nucleophile to form a stable disulfide bond^{19,20}. Such oxidant-mediated disulfide bond formation process, typically occurring within the reducing conditions of the cytosol on a transient increase in oxidative stress, differs from the enzyme-catalysed formation of structural disulfide bonds. This latter process takes place in the oxidizing milieu of the endoplasmic reticulum or the mitochondrial inter membrane space in eukaryotes or in the periplasmic space in bacteria^{19,21} via the formation of a mixed disulfide intermediate between the catalytic enzyme and the protein substrate^{22,23}. Despite its *in vivo* relevance, a comprehensive molecular description of the formation of disulfide bonds within the context of non-enzymatic oxidative folding is still missing.

Here we use a combination of single-molecule force-clamp spectroscopy, protein engineering techniques and MS to directly monitor the reactivity of an individual force-induced -SOH moiety occurring within a single immunoglobulin domain of the giant titin protein. Our results demonstrate that the fleeting -SOH intermediate, the life-time of which is largely conformation dependent, acts as a molecular switch that directly modulates protein function. In a short window of opportunity in the second timescale, the -SOH moiety is able to avoid irreversible protein misfolding—induced by cysteine hyperoxidation and/or aldehyde condensation—by readily forming a disulfide bond that guarantees the functional final folding of the protein. We hypothesize that this mechanism for non-enzymatic oxidative folding may be a common place for oxidation-induced post-translational modifications occurring on solvent-exposed cysteines, drastically

affecting the elastic properties of proteins that are exposed to mechanical forces.

Results

-SOH triggers disulfide bond formation. -SOH is typically induced by exposing a cysteine residue to high concentrations of hydrogen peroxide (H₂O₂)²⁴. However, its fast kinetics (10 – 10^7 M⁻¹ s⁻¹)^{25,26} and its further reactivity to higher oxidation state species often precludes the capture of this transient -SOH intermediate. An alternative means to create a protein -SOH in a much more controlled manner is based on the alkaline hydrolysis of an individual disulfide bond, resulting in the stoichiometric formation of a -SOH and a thiolate^{27–29} (Fig. 1a). Albeit admittedly less physiologically relevant, this strategy is also found in nature^{30,31}, as reported for phosphoglycerate kinase reductase^{32,33}.

Using an experimental single-molecule mechanical assay designed previously³⁴, an individual -SOH moiety can be readily created by the hydrolysis of a single protein disulfide bond under high-pH conditions (Fig. 1a). The atomistic detail of this process has been provided by highly revealing molecular quantum mechanics/molecular mechanics (QM/MM) simulations^{35,36}. We now set out to extend our experimental approach to probe whether each cleaved disulfide bond can be reformed in the absence of a catalytic enzyme, and how such putative reformation process affects the entire protein folding phenomena.

In our experiments, we used a single-molecule force-clamp atomic force (AFM) spectrometer (Fig. 1b) to stretch under high alkali conditions (pH = 12.8), a polyprotein containing eight identical domains of a mutated titin Ig27 domain containing two cysteines in positions 24 and 55, which readily form a structurally buried disulfide bond (Fig. 1a). We designed a five-pulse force protocol to measure the protein elongation over time, which enables independent capture of each individual cleavage and reformation event, with single bond resolution (Fig. 1c). Applying an initial pulse of 150 pN for 0.5 s to the (I27_{E24C-K55C})₈ polyprotein resulted first in the presence of fast elongation steps of ~15 nm (inset, grey), corresponding to the unfolding and extension of each individual protein domain up to the rigid disulfide bond, which became exposed to the solvent environment (Fig. 1c, grey). A second force pulse at a much higher force (500 pN) triggered the forced-induced scission of each individual disulfide bond via a S_N2 nucleophilic attack by the hydroxyl (OH⁻) anion. Each individual hydrolysis event is fingerprinted by a step of ~10 nm (inset, green), corresponding to the extension of the amino acids that were trapped behind the disulfide bridge (Fig. 1c, green). The two distinctive step-size populations in these first two high-force pulses (initial pulse) are clearly observed in the bimodal histogram shown in Fig. 1d. Notably, after the cleavage of all disulfide bonds at 500 pN, the fully unfolded and extended polyprotein is extended to almost its contour length and all of its backbone residues are fully exposed to the chemical environment of the measuring solution.

Shortly after the last individual rupture event, the stretching force was completely withdrawn for $t_q = 5$ s to trigger the folding of the polypeptide. Such force-quench approach allows dissection of the distinct phases involving mechanical folding from highly extended states^{37–40}. After the quenching pulse, a test pulse—composed again of a stretching sequence of 0.5 s at 150 pN followed by a pulse at a higher force of 500 pN—was applied to the protein to probe its folding status.

The elongation pattern of the protein during the test pulse enables discrimination among three different possible oxidative folding scenarios; (i) if the protein has not been able to fold

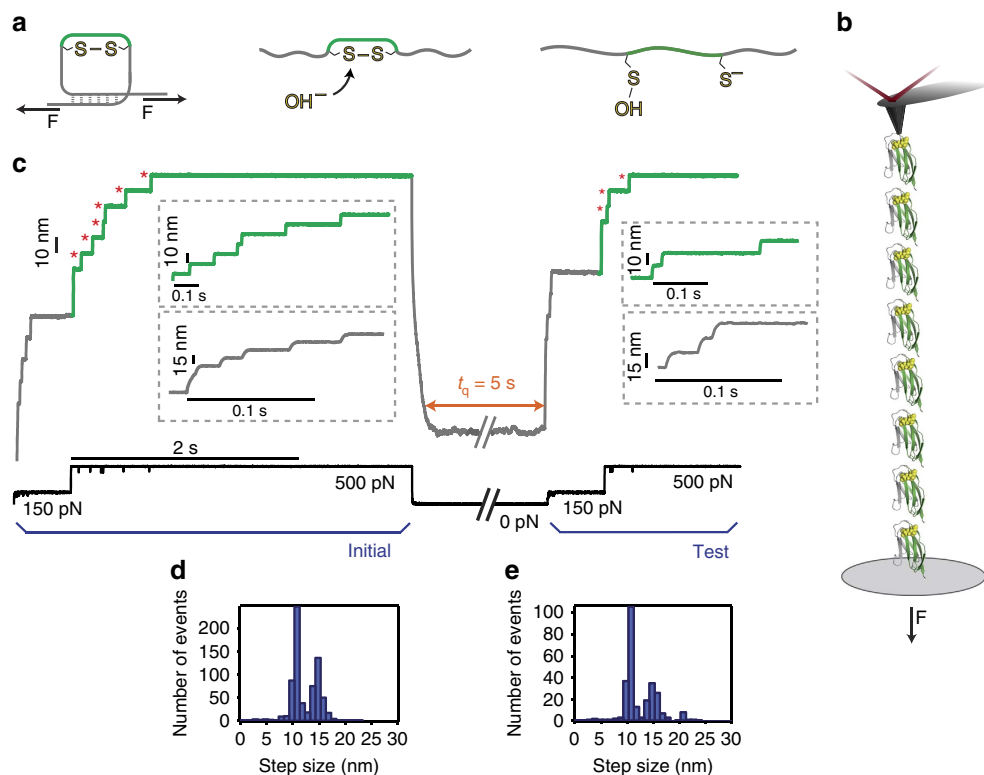


Figure 1 | Single-molecule force-clamp spectroscopy captures the non-enzymatic reformation of individual disulfide bonds. (a) The hydrolysis of the protein disulfide bond via a $\text{S}_{\text{N}}2$ attack by hydroxide anions results in the creation of a sulfenic acid and thiolate. (b) Schematics of the engineered $(\text{I27}_{\text{E24C-K55C}})_8$ polyprotein being stretched in the AFM set-up. (c) Pulling on an individual polyprotein at constant force using a specifically designed pulse-protocol captures oxidative protein folding. The initial pulse unfolds the protein and ruptures the disulfide bond; the first pulse $F = 150$ pN unfolds and extends all the modules of the $(\text{I27}_{\text{E24C-K55C}})_8$ polyprotein to the mechanical clamp created by the disulfide bonds, which exposes them to the solution (grey line). Each unfolding event is marked by a step increase of ~ 15 nm (inset, grey line). A second stronger force pulse ($F = 500$ pN) triggers disulfide rupture (green line), fingerprinted by equally spaced ~ 10 nm steps (inset, green line) that correspond to the extension of the peptide chain released on the scission of each disulfide bond (red asterisks). Releasing the pulling force ($F = 0$ pN) triggers the protein to collapse and fold for a time quench $t_q = 5$ s. The subsequent test pulse probes the folding status of the protein. Mirroring the initial pulse, a force pulse of $F = 150$ pN unfolds and extends the refolded protein up to the newly created disulfide bond (inset, grey line) whereas the last $F = 500$ pN pulse is able to rupture it again (inset, green line). The step-size histograms corresponding to the (d) initial and (e) test pulse highlight in both cases the presence of ~ 10 nm and ~ 15 nm steps.

within the quench time, it will extend featuring the absence of well-defined steps^{37,39}. Instead, (ii) if the protein is able to refold—thus regaining mechanical stability—but the disulfide bond has not been reformed, the protein will extend featuring steps of 25 nm, corresponding to the unfolding and stretching of the fully folded protein in the absence of the mechanically rigid, covalent disulfide bond³⁸. Finally, (iii) if the protein has been able to both successfully refold and reform the disulfide bond, the test pulse will be then composed of the unfolding ~ 15 nm steps followed by the ~ 10 nm steps corresponding to the (re)-rupture of each successfully reformed disulfide bond. In summary, a 25 nm step provides the signature of a properly folded domain with no disulfide, whereas a 15 nm step (followed by a 10 nm step) indicates a properly folded domain with a formed disulfide.

The latter scenario was indeed observed in our folding trajectories (Fig. 1c). The histogram corresponding to the distribution of step sizes populating the test pulse hallmarks the presence of both ~ 15 nm and ~ 10 nm steps (Fig. 1e). Surprisingly, not a single event corresponding to the ~ 25 nm steps was observed. Interestingly, a small population ($\sim 4\%$) of ~ 21 nm events was also present in the high-force pulse, corresponding to the scission of a non-native disulfide bond formed between two contiguous domains within the polyprotein chain⁴⁰ (Supplementary Fig. 1). As a control experiment, we verified that the oxidized protein was able to independently refold

under high-pH conditions, as shown in Supplementary Fig. 2. Altogether, our experiments qualitatively show that an individual disulfide bond is able to reform in the absence of catalytic enzymes.

We next examined whether such oxidative folding scenario could also be reproduced when the specific location of the disulfide bond within the structure of the I27 protein was changed. We repeated the experiments reported in Fig. 1 on a distinct polyprotein construct, $(\text{I27}_{\text{G32C-A75C}})_8$, in which the cysteine mutations were engineered in different positions of the I27 sequence. The resulting trajectories (Supplementary Fig. 3) qualitatively certified the successful reformation of the disulfide bond along the folding pathway.

The kinetics of disulfide bond reformation. To quantify the kinetics of disulfide bond reformation, we repeated the experiments shown in Fig. 1 and varied the quench time (t_q) within the range spanning $t_q = 0.5$ –15 s (Fig. 2). As observed in the individual trajectories shown in Fig. 2a,b, the longer the protein was allowed to refold in the absence of force, the higher the percentage of disulfide bonds that reformed (hallmarked by the appearance of ~ 10 nm steps in the test pulse). As shown in Fig. 2c, the per cent of disulfide bond reformation increased exponentially with the quench duration, suggesting that the

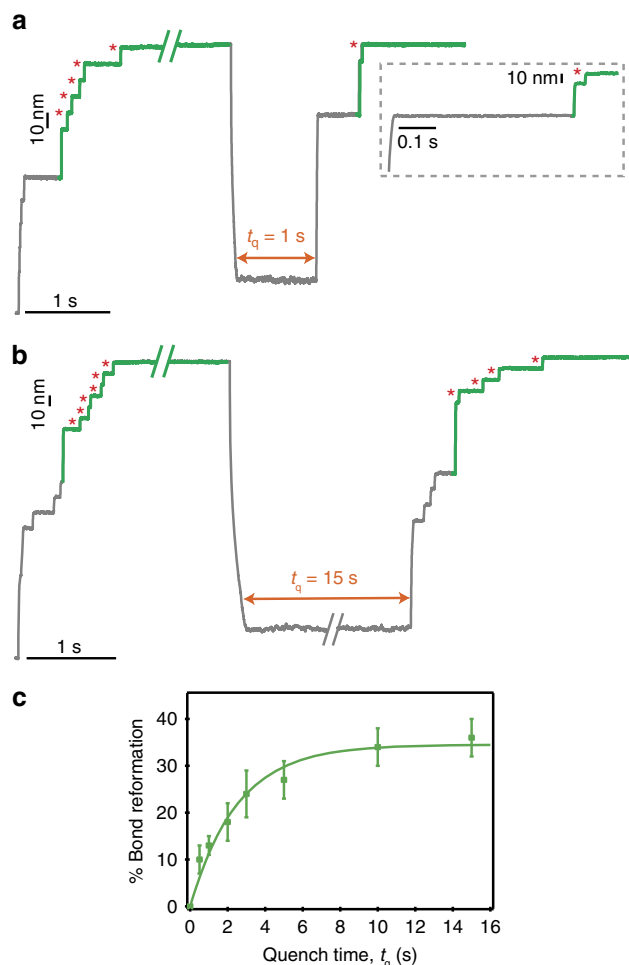


Figure 2 | Kinetics of disulfide bond reformation. (a,b) The percentage of disulfide reformation, fingerprinted by the presence of ~ 10 nm steps (red asterisks), increases with t_q . **a** illustrates a particular trajectory showing that the reformation of the disulfide bond (marked by the 10 nm step) occurred in the absence of protein refolding, since the 10 nm step, corresponding to the rupture of the disulfide bond, is not preceded by a 15 nm step, hallmark of protein refolding. This suggests that, in this particular trajectory, we interrupted the oxidative folding process at an instant whereby the reformation of the disulfide bond had occurred but the protein did not have yet time to fold. **(c)** The percentage of disulfide bond reformation increases exponentially with t_q with an associated rate constant of $k_{\text{ref}} = 0.45 \text{ s}^{-1}$ (solid green line, $\chi^2_{\text{red}} = 0.513$).

reformation of the protein disulfide bond is a two-state process with an associated characteristic time of $k_{\text{ref}} = 0.45 \text{ s}^{-1}$. This is consistent with a scenario where the nucleophilic thiol would attack the electrophilic -SOH centre via a concerted S_N2 reaction, occurring through a well-defined transition state^{41,42}, to reform the disulfide bond. Noteworthy, the measured reformation rate is expected to be pH dependent, since the nucleophilic capacity of the free thiolate to attack the electrophilic -SOH will be largely dependent on the thiolate being completely deprotonated. However, given the expected $pK_a \sim 8$ for each free cysteine, we do not anticipate these rates to be significantly affected in our high-pH experiments, although this might have an effect at lower pH values. Remarkably, the measured rate of disulfide reformation for the (I27_{G32C-A75C})₈ construct ($k_{\text{ref}} = 0.42 \text{ s}^{-1}$) was found to be almost identical to that measured for the (I27_{E24C-K55C})₈ polypeptide, thus demonstrating that the disulfide bond reformation kinetics is not significantly altered by changing

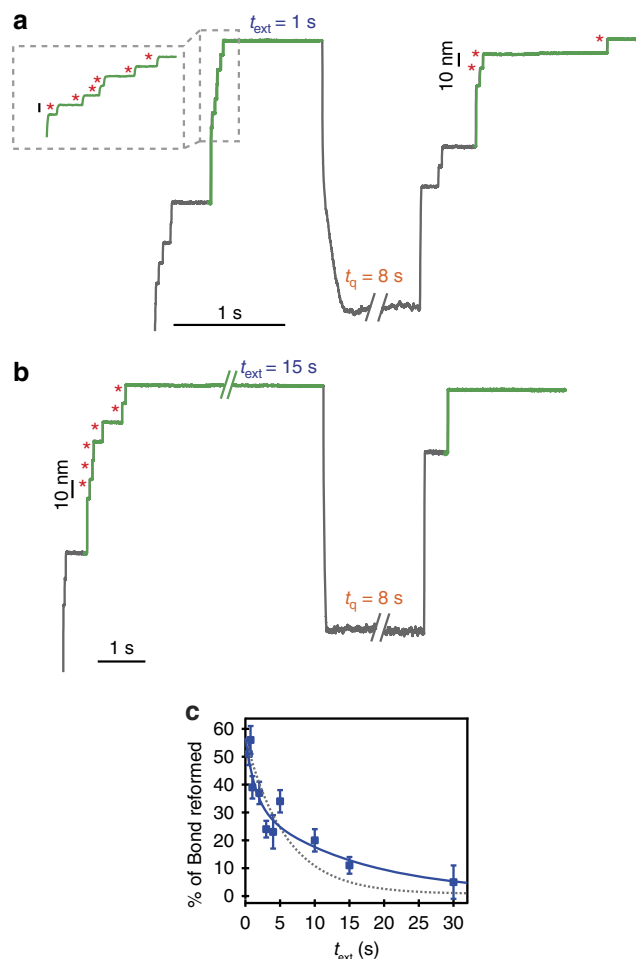


Figure 3 | S-sulfenylation of cryptic cysteines inhibits oxidative folding.

(a,b) Increasing the time, t_{ext} , the protein is reduced and stretched at high force ($F = 500$ pN) results in a drastic decrease of successful oxidative folding (marked by the decrease of both ~ 10 nm steps, red asterisks, and ~ 15 nm steps in the test pulse). **(c)** While the decrease in the % of disulfide bond reformation as a function of t_{ext} can be fitted with a single exponential (dotted grey line) with a rate constant $k = 0.17 \text{ s}^{-1}$ ($\chi^2_{\text{red}} = 5.02$), the trend is best described ($\chi^2_{\text{red}} = 2.78$) by a double exponential fit (solid blue line), with associated rate constants $k_1 = 0.065 \text{ s}^{-1}$ and $k_2 = 0.695 \text{ s}^{-1}$, reminiscent of two competing processes occurring concomitantly.

the position of the disulfide bond within the protein structure (Supplementary Fig. 3).

Reactivity of -SOH depends on solvent accessibility. While generally short, the life-time of -SOH has been described to crucially depend on the degree of solvent exposure^{20,43}. Force-clamp AFM allows to explore new protein conformations that are not usually sampled in classical ensemble experiments conducted in the bulk⁴⁴. In particular, the force-triggered low-entropy, highly extended unfolded state fully exposes the backbone residues to the solvent environment⁴⁵. To interrogate the conformation-dependent reactivity of -SOH, we systematically changed the time the fully reduced and mechanically stretched protein was exposed to the solvent, t_{ext} , thereby increasing the probability to be chemically modified, while keeping the quench time constant ($t_q = 8$ s, Fig. 3a,b). Leaving the protein stretched for $t_{\text{ext}} = 0.5$ s resulted in a $\sim 55\%$ disulfide bond reformation. Substantially increasing the extended time up to $t_{\text{ext}} = 30$ s

considerably decreased the reformation probability down to $\sim 5\%$. The absence of disulfide bond reformation inevitably resulted in protein misfolding, since we never observed the presence of a 15 nm step without the companion 10 nm step in the test pulse. The evolution of the per cent of disulfide bond reformation as a function of t_{ext} showed a marked, fast decrease (Fig. 3c). While fitting a single exponential to the data qualitatively described the time-course of the process with an associated decaying rate of $k = 0.17 \text{ s}^{-1}$ ($\chi^2_{\text{red}} = 5.02$), a double exponential ($k_1 = 0.065 \text{ s}^{-1}$ and $k_2 = 0.695 \text{ s}^{-1}$) slightly increased the goodness of the fit ($\chi^2_{\text{red}} = 2.78$, Supplementary Fig. 4), reminiscent of two independent reaction pathways occurring concomitantly.

To directly quantify the relationship between the degree of solvent exposure and -SOH reactivity, we systematically varied the stretching force after triggering the formation of -SOH. Using a six-pulse force protocol (Fig. 4), the initial two pulses unfolded the protein (150 pN, 0.3 s) and triggered the rupture of the disulfide bond within a short time (500 pN, 2 s). The force was then varied within a range spanning 0–500 pN during 4 s (grey box) before being withdrawn (0 pN, $t_q = 8 \text{ s}$) to trigger collapse³⁹ and oxidative folding. As before, the folding success was probed by two high force test pulses (150 pN followed by 500 pN). The folding trajectories shown in Fig. 4a qualitatively show that the number of 10 nm steps observed in the test pulse (red asterisks) significantly decreased as the stretching force was increased. This tendency was quantitatively verified in Fig. 4b, displaying the percentage of disulfide bond reformation as a function of the stretching force. Crucially, the observed trend can be nicely reproduced by the worm-like chain (WLC) model of polymer elasticity (using a persistence length $P = 1.2 \text{ nm}$)⁴⁵, which relates the protein's end-to-end length with the pulling force. At each defined force, the WLC model defines an equilibrium end-to-end conformation that is characterized by a different degree of solvent exposure of the protein residues⁴⁶. Altogether, these experiments demonstrate that the reactivity of -SOH—intimately related to the oxidative folding fate of the protein—is largely determined by the protein's conformation.

To investigate the molecular origin of the chemical modification that prevented the reformation of the disulfide bond at high pulling forces, we conducted MS measurements on the monomeric I27_{E24C-K55C} protein under high-pH conditions. MS spectra revealed the presence of both sulfinic and sulfonic acid, thus suggesting the successive irreversible oxidation of -SOH (Supplementary Figs 5 and 6 and Supplementary Table 1). Notably, the presence of aldehyde (with a concomitant loss of H_2S) was also detected with good evidence (Supplementary Fig. 7 and Supplementary Table 1). Indeed, aldehyde formation has been long reported as a sub-product of the decomposition of -SOH under high alkali conditions^{28,47,48}. While both the protein conformation and the timescales sampled in the MS and single-molecule mechanics experimental approaches are certainly different, the MS data provide excellent complementary information that illustrates the most plausible chemical processes responsible for the misfolding events captured in the nanomechanical experiments, greatly affecting the folding and elastic properties of the individual titin polypeptides.

Dimedone binding confirms the presence of -SOH. To provide a final, unambiguous proof that -SOH is actively involved in the molecular mechanism driving the reformation of disulfide bonds, we directly tested the effect of dimedone on the outcome of our oxidative folding assay. Dimedone has been vastly demonstrated to effectively react with -SOH (but not with thiols) with high specificity (Fig. 5a)^{7,15}. Addition of 30 mM of dimedone to the

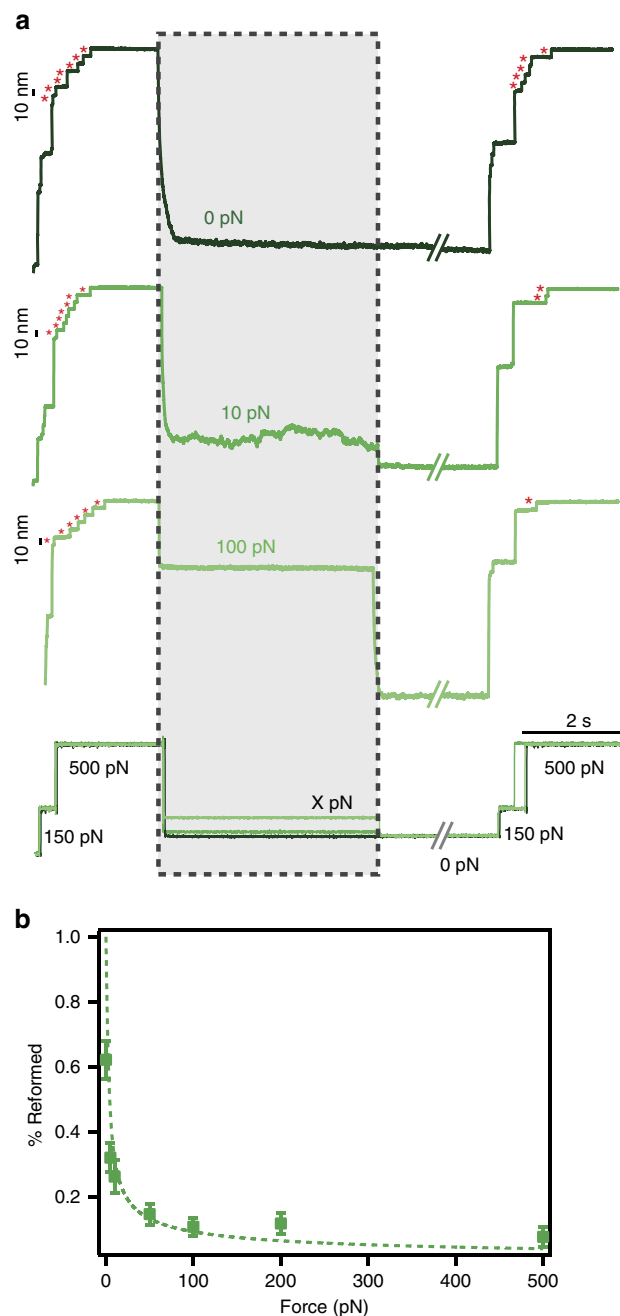


Figure 4 | The chemical evolution of sulfenic acid is force-dependent.

The degree of solvent exposure is tuned by the pulling force, which determines the (I27_{E24C-K55C})₈ protein end-to-end length. **(a)** A six-pulse force protocol probes the effect of force and solvent exposure on the outcome of oxidative folding. The number of 10 nm steps (red asterisks) significantly decreases with increasing the force applied to the protein after the sulfenic acid is formed (grey box), varying from 0 pN (dark green) to 100 pN (light green). **(b)** The percentage of disulfide bond reformation shows a drastic decrease with an increase in the stretching force, and the evolution can be captured by the worm-like chain (WLC) model of polymer elasticity ($P = 1.2 \text{ nm}$). This trend is indicative of the distinct reactivity of sulfenic acid as a function of its exposure to the solvent. At low forces the local protein environment shields the sulfenic acid, allowing it to condense into a reformed disulfide bond that ensures successful oxidative folding. By contrast, at higher forces the sulfenic acid is fully exposed to the solution and undergoes further chemical modification, ultimately leading to protein misfolding, hallmarked by the inability of the protein to recover its native mechanical stability.

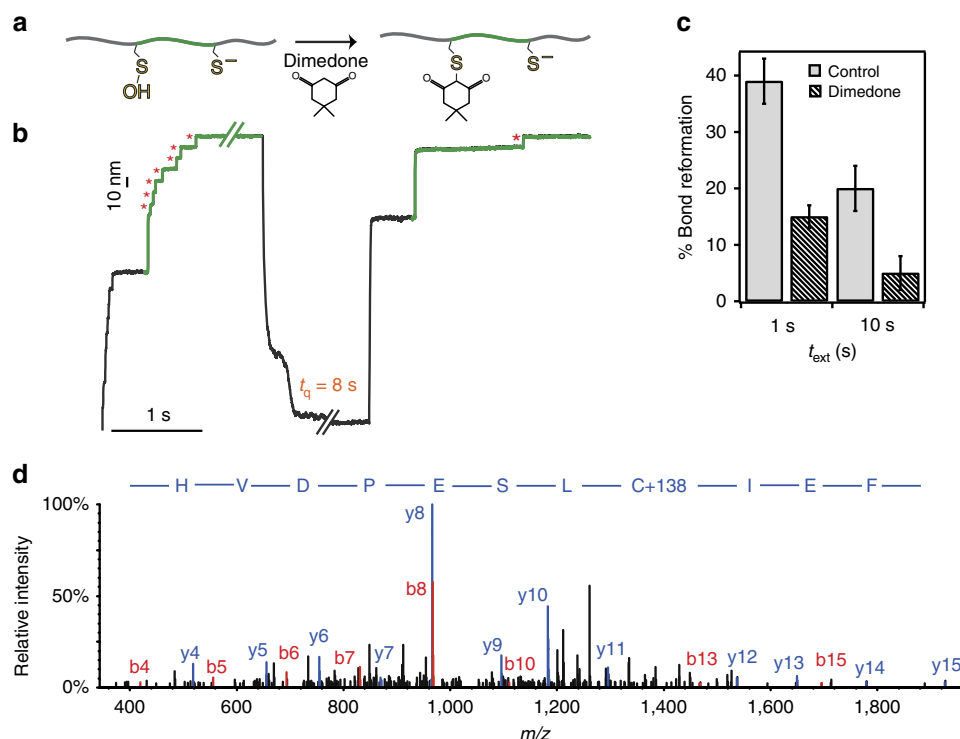


Figure 5 | Dimedone selectively targets sulfenic acid and blocks oxidative folding. (a) The newly formed sulfenic acid moiety can be selectively trapped by dimerone. (b) When an individual $(I27_{E24C-K55C})_8$ polyprotein is reduced and stretched in the presence of a 30 mM solution of dimerone, the subsequent test pulse shows a dramatic decrease in the presence of both ~ 10 nm (red asterisks) and ~ 15 nm steps, implying that oxidative folding is hindered. (c) The percentage of disulfide bond reformation (and hence of successful folding) decreases with t_{ext} , the time the protein remains extended and stretched at a high force. For $t_{\text{ext}} = 1$ s, the % reformation decreases down to $15.6 \pm 2.3\%$ in the presence of dimerone. On increasing $t_{\text{ext}} = 10$ s, the percentage of reformation is further reduced ($6.0 \pm 3.3\%$). (d) The MS/MS fragmentation spectra of an $I27_{E24C-K55C}$ monomer under high-pH conditions provided unambiguous evidence of the presence of the dimerone, hallmarked by a cysteine mass shift of $\Delta m/z = +138$ Da.

measuring solution markedly blocked both the reformation of the disulfide bond and the folding of the protein (Fig. 5b), decreasing the bond reformation efficiency down to $\sim 15\%$ (Fig. 5c). This drastic decrease in the reformation (and subsequent refolding) efficiency was further exacerbated (down to a residual $\sim 6\%$) when the protein was left fully extended at high forces for a longer period of time, $t_{\text{ext}} = 10$ s, promoting dimerone binding (Fig. 5c). By contrast, dimerone did not prevent the protein from successfully refolding when the disulfide bond remained oxidized (Supplementary Fig. 8). To indisputably verify the presence of the covalent S-dimerone adduct, we conducted MS measurements on the monomeric $I27_{E24C-K55C}$ protein under high-pH conditions. The resulting MS/MS fragmentation spectra identified, for both 24 and 55 cysteines, a shift in $\Delta m/z = +138$ Da (Fig. 5d, Supplementary Fig. 9 and Supplementary Table 2), which confirmed the presence of dimerone.

Discussion

Oxidative stress defines elevated intracellular levels of reactive oxygen species (ROS)—mostly including superoxide anion, hydrogen peroxide and hydroxyl radicals—that cause damage to DNA, lipids and proteins⁴⁹. While ROS were originally known for oxidizing various cellular compartments and promoting aging and a broad range of pathologies, more recent evidence shows that ROS also acts as signalling molecules that regulate basic cellular processes including growth, differentiation and cell migration^{26,50}. It is increasingly clear that a major mechanism of redox signalling is the dynamic regulation of protein function by the chemiselective oxidation of cysteine residues. The first

product of oxidation is Cys-SOH, exhibiting an oxidation state of (0). Despite its ephemeral nature, -SOH has been trapped in a number of crystallized proteins^{12,51}. These carry a wide range of functionalities¹², encompassing enzyme catalysis (such as protein tyrosine phosphatases⁵², kinases⁵³ and cysteine proteases⁵⁴), transcription factors (including OxyR⁵⁵, Yap1 (ref. 56) and NF- κ B⁵⁷) and ion channels (such as ORAI1 (ref. 58) and Kv1.5 (ref. 59)). This progress notwithstanding these structures provide a static snapshot, and the fast dynamics of such important post-translational modification often evaded characterization.

The relatively small number of crystallized proteins contrasts with the increasingly larger number of proteins where -SOH has been identified *in vivo*. Indeed, the dimerone-based cell-permeable chemical probe DAZ-2, capable of detecting SOH-modified proteins directly in living cells⁵⁰, has allowed for a comprehensive analysis of the cellular ‘sulfenome’¹⁸, revealing the presence of ($>1,000$) S-sulfonylation sites on more than 700 proteins in intact cells, highly prevalent in different cancer cell lines^{14,60}. Importantly, these newly identified protein targets largely expand the diversity of functionalities of the already identified sulfonylated proteins. Therefore, unveiling the molecular mechanisms defining -SOH reactivity is an essential requirement for the understanding of redox regulation in cells.

Our single-molecule approach allows independent measurement of the protein folding and disulfide bond formation processes, enabling us to directly visualize the molecular events defining the complex kinetics of the oxidative folding mechanism. The crucial role of disulfide reformation in protein refolding was further probed in experiments on the pre-reduced $(I27_{E24C-K55C})_8$ polyprotein at pH = 12.8. While the oxidized protein was able to

refold under high-pH conditions (Supplementary Fig. 10), the folding efficiency of the pre-reduced protein was significantly decreased down to $\sim 20\%$, exhibiting a seemingly slow folding kinetics, $k_{\text{fold}} = 0.10 \text{ s}^{-1}$ (Supplementary Fig. 10). This result contrasts with the much higher folding efficiency of the sulfenylated protein, which folds on the reformation of the disulfide bond ($k_{\text{ref}} = 0.45 \text{ s}^{-1}$), ultimately setting the associated folding rate ($k_{\text{fold}} = 0.46 \text{ s}^{-1}$). These experiments demonstrate that while the protein is mostly unable to successfully refold on its own under high pH, the reformation of the disulfide bond dramatically increases its folding efficiency. In this sense, -SOH emerges as a key molecular player able to rescue the protein from misfolding by forming a protective disulfide bond that allows the protein to successfully refold. This kinetic scheme seems to differ from the enzyme-mediated oxidative folding findings catalysed by PDI⁴⁰ and DsbA⁶¹, whereby the same I27 protein mutant was able to refold on its own while keeping the catalytic enzyme attached through a mixed disulfide conformation before the disulfide bond was successfully reformed at a later stage, the overall kinetics being largely controlled by the off-rate dynamics of the catalytic enzyme.

We postulate that the oxidative folding mechanism promoted by -SOH condensation with a neighbouring intramolecular thiolate is mechanistically different from the process where the reformation of the disulfide bond occurs through a mixed disulfide intermediate, which characterizes the enzyme-catalysed oxidative folding process²². To test this hypothesis, we repeated the force-quench experiments at pH = 7.2 using the small thiol 3-mercapto-1-propanol as the attacking nucleophile⁴¹ (Supplementary Fig. 11). In this case, the rupture of each individual disulfide bond resulted in the formation of a thiolate and the newly formed mixed disulfide. Unlike the case where -SOH could readily reform the cleaved disulfide, thus triggering the protein to correctly fold, when 3-mercapto-1-propanol was used as a nucleophile, the number of reformation events was vanishingly small. By contrast, these experiments highlighted (albeit with a relatively low abundance) the presence of 25 nm steps in the test pulse. Altogether, our experiments demonstrate the unique capabilities of -SOH to form a disulfide bond employing a chemical mechanism that is fundamentally different from that occurring during enzyme-mediated protein folding^{22,40}. In this novel -SOH-mediated scenario, the newly formed disulfide reduces the conformational space of the protein, and subsequently triggers almost ‘instantaneous’ folding of the remaining protein in the timescale of our experiments. Thus, the kinetics of disulfide bond formation is the rate-limiting step that defines the overall kinetics of the non-enzymatic oxidative folding process. It is certainly remarkable that enzyme-free oxidative folding occurs under high pH conditions at a similar rate ($k_{\text{ref}} = 0.45 \text{ s}^{-1}$) than that catalysed by a dedicated enzyme such as PDI ($k_{\text{ref}} = 0.35 \text{ s}^{-1}$) and DsbA ($k_{\text{ref}} = 0.70 \text{ s}^{-1}$)^{40,61}.

Our experiments demonstrate that the probability of reforming the disulfide bond increases exponentially with time (Fig. 2). This implies that, when the protein lies within the non-folded collapsed conformation after the force-quench, the highly reactive -SOH moiety is sterically stabilized within the microenvironment of the collapsed protein, most likely through becoming cryptic to the solvent. This situation contrasts with the fast evolution of the -SOH in the fully extended state (Fig. 3). At the present stage it is difficult to unambiguously pinpoint the exact origin of the fast reactivity undergone by the solvent-exposed -SOH. Due to the relatively short timescale of our single-molecule experiments, it is unlikely that the progression to higher oxidation state species is favoured, and hence we hypothesize that the aldehyde formation mechanism might be prevalent^{28,47,48}, which would account for the main pathway described by the single exponential fit. Given

the slightly higher accuracy of the double exponential fit, we postulate a possible competition between the high oxidation pathway (promoted by the dissolved oxygen in solution) and the aldehyde formation mechanism, exhibiting similar rate constants. Noteworthy, we cannot rule out the possibility that a fraction of the aldehyde population found in our MS experiments is a result of the direct alkali-mediated cleavage of the disulfide bond through an α -elimination mechanism, which competes with a β -elimination and the hydroxyl-mediated $\text{S}_{\text{N}}2$ nucleophilic attack^{27–29,62} (Supplementary Fig. 12). Indeed, our MS spectra show evidence of the three competing pathways (Supplementary Figs 7, 13 and 14 and Supplementary Table 1). However, it is very likely that, in the presence of force, the hydrolysis ($\text{S}_{\text{N}}2$) pathway is largely favoured over the elimination route, as suggested by recent quantum mechanical simulations under force⁶³, in which case the presence of aldehyde is likely to stem mostly from the evolution of -SOH. Further verification of the plausible competing reaction pathways under force, probably guided by *ab initio* simulations, might provide a comprehensive understanding of this complex reactivity landscape. In any case, while the high-pH conditions tested in our experiments probably trigger a faster -SOH reactivity towards the aldehyde product than one would perhaps expect under more physiologically relevant conditions, our experiments readily demonstrate that the reactivity of -SOH is largely conformation dependent, crucially contributing to the folding fate of the protein. We speculate that this mechanism might hold true in situations where -SOH is triggered by other biologically occurring oxidative species such as H_2O_2 . To test this hypothesis, we conducted folding experiments on the wt-I27 immunoglobulin module (containing two cryptic and native cysteines in positions 47 and 63 that do not form a disulfide bond) in the presence of $300 \mu\text{M}$ H_2O_2 at neutral pH (Supplementary Fig. 15). Stretching and unfolding the (wt-I27)₈ polypeptide exposed the cryptic cysteines to the oxidative conditions. Quenching the force and re-unfolding back again resulted in (i) the misfolding of the protein, fingerprinted by the absence of recovery of mechanical stability, and (ii) the inability of the protein to completely re-extend up to its unfolded contour length, suggestive of the presence of a stiff, non-native (probably interdomain) disulfide bond. Controls were performed with the double mutant (I27_{C47A–C63A})₈ polypeptide (devoid of both cysteines) showing full refolding and complete protein extensibility. These experiments, conducted in the presence of the less controllable yet physiologically more relevant H_2O_2 oxidative agent, strongly suggest that, when solvent exposed on mechanical unfolding⁶⁴, cryptic cysteines undergo post-translational modifications that markedly block protein folding, while lending further support to the SOH-mediated mechanism of disulfide bond formation that we investigated using the high-pH experimental strategy.

Taken together, our nanomechanical experiments put forward a rather unique kinetic scheme (Fig. 6) that captures the intricate conformation-dependent reactivity of the transitory -SOH intermediate, directly linked with the folding fate of the protein. In the case of titin, a protein that is physiologically exposed to mechanical force, such SOH-regulated folding scenario might have important consequences for the functional protein elasticity.

Titin is a molecular ruler that determines the passive elasticity of muscle, and is key to ensure, for example, cardiac function⁶⁵. Indeed, alterations in the stiffness of cardiac titin are related to a number of myopathies, such as diastolic dysfunction⁶⁶. The I-band region of titin is formed by a compliant, unstructured region (PEVK and N2B) intercalated between folded immunoglobulin domains, which are mechanically resistant⁶⁷. To adapt to the mechanical perturbations induced by the heartbeats, titin needs to reversibly modulate its length by

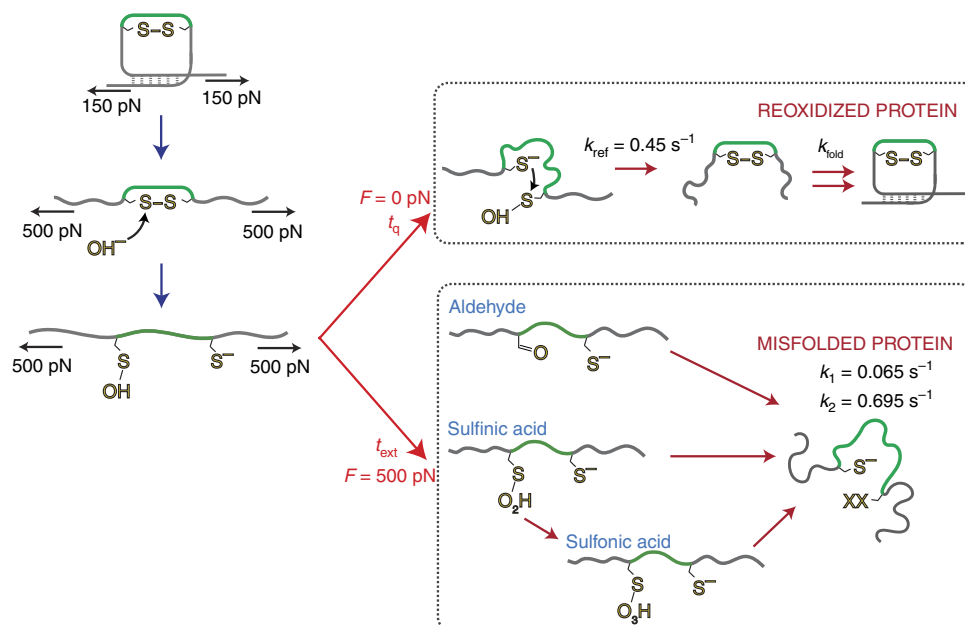


Figure 6 | Kinetic scheme for the force-triggered, conformation-dependent reactivity of sulfenic acid and its implications for oxidative protein folding.

On the application of mechanical force, the hydroxide anions rupture a protein disulfide bond, creating a sulfenic acid and a thiolate. In the absence of force ($F = 0$ pN) during the quench time (t_q) the sulfenic acid moiety remains cryptic in the protein's collapsed state. Under these conditions, the longer t_q , the higher the probability the thiolate anion can attack the sulfenic acid moiety (in the absence of dimedone), thus creating a disulfide bond with an associated rate of $k_{ref} = 0.45 \text{ s}^{-1}$. The reformation of the disulfide bond triggers the 'instantaneous' refolding of the protein (essentially set by the disulfide bond reformation rate). By contrast, when the protein is reduced and stretched at high force ($F = 500$ pN), sulfenic acid is fully exposed to the solvent and undergoes further modification either to the formation of sulfinic acid (which may be further oxidized to sulfonic acid) or to aldehyde. Such further reactivity of sulfenic acid results in irreversible misfolding of the reduced (I27_{E24C-K55C})₈ polyprotein, with a consequent decrease in its stiffness properties.

entropic extension/recoil of the compliant region and unfolding/refolding (at least of the mechanically weak) immunoglobulin domains⁶⁴. On immunoglobulin domain unfolding, the stiffness of the protein is greatly reduced, and it is increased back again on refolding. Recent pioneering evidence shows that S-glutathionylation of cryptic cysteines blocks titin refolding⁶⁴. Our results are likely to add S-sulenylation to the toolbox of post-translational modifications that regulate titin elasticity; whereas cysteine hyperoxidation irreversibly suppresses Ig27 refolding, thus severely decreasing its stiffness, the reformation of a protective disulfide bond has the opposing effect. In this sense, -SOH would act as 'mechanical switch', helping titin adapt its elasticity in response to the fluctuating oxidative stress conditions. We hence speculate that -SOH might work as an active redox sensor that directly regulates protein elasticity.

Methods

Protein engineering. The (I27_{E24C-K55C})₈, (I27_{G32C-A75C})₈, (I27)₈ and (I27_{C47A-C63A})₈ polyproteins were subcloned using the *Bam*HI, *Bgl*II and *Kpn*I restriction sites. The multidomain proteins and the I27_{E24C-K55C} monomer were cloned into the pQE80L (Qiagen) expression vector, and transformed into the BLR(DE3) *Escherichia coli* expression strain. The cells were grown in LB broth supplemented with $100 \mu\text{g ml}^{-1}$ ampicillin at 37°C . After reaching an OD_{600} of ~ 0.6 , the cultures were induced with Isopropyl β -D-1-thiogalactopyranoside (1 mM) and incubated overnight at 20°C . After collecting the cells, disruption using a French Press was performed. The polyproteins from the lysate were purified by metal affinity chromatography on Talon resin (Clontech) followed by gel filtration using a Superdex 200 10/300 GL column (GE Biosciences). Protein concentration in the samples was estimated using the Bradford protein assay.

Force spectroscopy. Single-molecule force-clamp spectroscopy AFM experiments were conducted at room temperature using both a home-made set-up⁶⁸ and a commercial Luigs and Neumann force spectrometer⁶⁹. The sample was prepared by depositing 1–10 μl of protein in PBS solution (at a concentration of $1\text{--}10 \text{ mg ml}^{-1}$) onto a freshly evaporated gold coverslide. Each cantilever (Si₃N₄ Bruker MLCT-AUHW) was individually calibrated using the equipartition

theorem, giving rise to a typical spring constant of $\sim 12\text{--}17 \text{ pN nm}^{-1}$. Single proteins were picked up from the surface and pulled at a constant force. Individual polyproteins were fished by pushing the cantilever onto the surface exerting a contact force of 500–1,500 pN so as to promote the non-specific adhesion of the proteins on the cantilever surface. The piezoelectric actuator was then retracted to produce a set deflection (force), which was set constant throughout the experiment ($\sim 8\text{--}45 \text{ s}$) thanks to an external, active feedback mechanism while the extension was recorded. The force feedback was based on a proportional, integral and differential amplifier whose output was fed to the piezoelectric positioner. The feedback response is limited to $\sim 3\text{--}5 \text{ ms}$. Thanks to the high-resolution piezoelectric actuator, our measurements of protein length have a peak-to-peak resolution of $\sim 0.5 \text{ nm}$. Data of the force traces was filtered using a pole Bessel filter at 1 kHz. In all instances the protein was unfolded at 150 pN for 0.5 s, followed by a high-force pulse of 500 pN, except when the protein was extended for very short times (0.5 and 0.75 s) when a pulse of 800 pN was applied. Experiments were carried out in a sodium phosphate buffer solution, specifically, 50 mM sodium phosphate (Na_2HPO_4 and NaH_2PO_4), 150 mM NaCl. The pH was adjusted at pH = 7.4 and pH = 12.8. The final pH value in each solution was adjusted by adding the required amounts of NaOH solutions. Measuring solutions were filtered through a $0.2\text{-}\mu\text{m}$ membrane before each experiment. In the case of the dimedone experiments, 5,5-Dimethyl-1,3-cyclohexanedione (Sigma-Aldrich, 95%) was dissolved in pH = 12.8 PBS buffer to give a final concentration of 30 mM. The pH of the dimedone solution was checked before each experiment. 3-Mercapto-1-propanol (Sigma-Aldrich, 95%) was freshly prepared before every experiment, dissolved in pH = 7.4 PBS buffer to give a final concentration of 400 mM. Pre-reduced (I27_{E24C-K55C})₈ polyproteins were obtained by incubating the polyproteins in 10 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP, Sigma-Aldrich) for 20 min before AFM experiments. Hydrogen peroxide solutions (H_2O_2 , Sigma-Aldrich) were obtained by dissolving the 30% (v/v) solution in PBS to a final concentration of $300 \mu\text{M H}_2\text{O}_2$.

Data analysis. All data were recorded and analysed using custom software written in Igor Pro 6.0 (Wavemetrics). For all polyproteins, only recordings showing the signature of at least 5 unfolding events (15 nm steps) followed by 5 reduction events (10 nm steps) were analysed. We considered only recordings corresponding to the test pulse in which the protein was extended back to the same full unfolded length characterizing the initial pulse, to ensure that the same protein was completely stretched in the two pulses. No traces that included unfolding events during the high force pulse were included in the analysis. The number of

independent observations, N , for a particular condition was counted as the total number of protein domains (as observed in the initial pulse). S.e.m. for the refolding fraction was estimated through the bootstrap method, where each recording was treated as an independent data point. S.e.m. for the fit parameters was determined as standard error for the coefficient in the fit, accounting for the measurement errors of the individual data points. To assess the goodness of the fits in Fig. 2 and Fig. 3, we conducted reduced χ^2 , χ^2_{red} analysis. Data in Fig. 4b was compared with the WLC model of polymer elasticity⁷⁰, using as fixed parameters $L_c = 28.5$ nm, $P = 1.2$ nm.

MS experiments. *Enzymatic digestion.* The monomeric protein samples were prepared by diluting 20 μ l of protein ($0.5 \mu\text{g} \mu\text{l}^{-1}$) in 20 μ l of either PBS pH 7.4, PBS pH12.8 or PBS pH12.8 + 30 mM dimedone. The high-pH samples had a calculated final pH = 12.5. In-solution digestion with trypsin of 5 μ g of total protein monomers was performed before subsequent analysis by MS. Trypsin digestion at a ratio of 1:20 (enzyme:substrate) was carried out for 2 h at 37 °C.

LC-MS/MS. Chromatographic separations were performed using an EASY NanoLC system (ThermoFisherScientific, UK). Peptides from a total protein amount of 2 μ g on column were resolved by reversed phase chromatography on a 75 μ m C18 column using a three step linear gradient of acetonitrile in 0.1% formic acid. The gradient was delivered to elute the peptides at a flow rate of 300 nl min⁻¹ over 60 min. The eluate was ionized by electrospray ionization using an Orbitrap Velos Pro (ThermoFisherScientific, UK) operating under Xcalibur v2.2. The instrument was programmed to acquire in automated data-dependent switching mode, selecting precursor ions based on their intensity for sequencing by collision-induced fragmentation using a Top20 CID method. The MS/MS analyses were conducted using collision energy profiles that were chosen based on the mass-to-charge ratio (m/z) and the charge state of the peptide.

MS database searching. Raw MS data were processed into peak list files using Proteome Discoverer (ThermoScientific; v1.4). Processed raw data was searched using the Mascot search algorithm (www.matrixscience.com) against an in-house database containing the sequence of the mutant protein monomer. The data was also searched using the Error Tolerant function which by iterating through a comprehensive list of chemical and post-translational modifications, determines the mass of additions or losses from the residues in the peptide sequences without prior knowledge of the expected modifications. LC/MS/MS analysis has successfully identified 100% sequence coverage of the mutant protein monomer with various forms of oxoacid modifications. The Mascot database searching algorithm applies a 95% probability CI in the MOWSE scoring that is an identification threshold. Mascot search output files were uploaded into the Scaffold spectral visualization software (v.4.4.6; www.proteomesoftware.com) which allows statistical filtering of the data at the protein and peptide level for which these filters were applied to the data at a the lowest stringency of 5% confidence interval (CI) for minimum protein, 0% CI for peptide values and minimum one peptide assignment. This was chosen to include all assigned peptides that may have fallen below the database applied 95% CI due to poor fragmentation matching and incorrect sequence assignment. All fragmentation spectra were verified manually to determine correct assignment of sequences and post-translation modifications.

Initial data searching concentrated on finding -SOH modifications after the addition of the cyclic diketone compound, dimedone. Dimedone specifically traps this oxoacid form and binds covalently to the cysteine residue. This binding with the subsequent loss of two protons adds a mass of 138.1 Da to the residue, which is detected on the peptide by fragmentation in MS/MS mode in the mass spectrometer.

Under conditions of high pH and without the presence of dimedone to specifically trap the -SOH, there is possible evolution of the oxoacids to be present in their different forms. The protein monomer treated with high pH alone was searched against the in-house database, which included the modifications -SOH, sulfinic acid and sulfonic acid. Further in-depth searching of the database was also performed with the modifications aldehyde, dehydroalanine and sulfide. Database searching assigned 99% sequence coverage of the protein monomers and also assigned various oxoacid modifications on the cysteine residues of interest.

Further analysis was performed using a fresh sample of monomers in a high pH only environment and treated with enzymatic digestion as previously described. This sample was digested almost immediately following the addition of the high-pH buffer. This was performed to determine whether the previously identified oxoacid modifications were present at point of treatment, or if they had evolved over the length of time it had spent in the buffer before trypsin digestion. Almost complete sequence coverage was achieved once again following database searching of the mass spectrometer raw data. Oxoacid cysteine modifications sulfinic and sulfonic acid were identified at residue Cys55 as were the products of both α -elimination, aldehyde and β -elimination, persulfide and dehydroalanine.

Data availability. The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD004514. Additional information and the data supporting this research, including the single-molecule nanomechanics experiments, can be obtained from the corresponding author on request.

References

- Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167–195 (1995).
- Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. Navigating the folding routes. *Science* **267**, 1619–1620 (1995).
- Karplus, M. & Weaver, D. L. Protein-folding dynamics. *Nature* **260**, 404–406 (1976).
- Anfinsen, C. B. & Haber, E. Studies on the reduction and re-formation of protein disulfide bonds. *J. Biol. Chem.* **236**, 1361–1363 (1961).
- Anfinsen, C. B., Haber, E., Sela, M. & White, Jr. F. H. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl Acad. Sci. USA* **47**, 1309–1314 (1961).
- Paulsen, C. E. & Carroll, K. S. Cysteine-Mediated Redox Signaling: Chemistry, Biology, and Tools for Discovery. *Chem. Rev.* **113**, 4633–4679 (2013).
- Gupta, V. & Carroll, K. S. Sulfenic acid chemistry, detection and cellular lifetime. *Biochim. Biophys. Acta Gen. Subjects* **1840**, 847–875 (2014).
- Yang, J., Carroll, K. S. & Liebler, D. C. The expanding landscape of the thiol redox proteome. *Mol. Cell. Proteomics* **15**, 1–11 (2016).
- Allison, W. S. Formation and reactions of sulfenic acids in proteins. *Accounts Chem. Res.* **9**, 293–299 (1976).
- Roos, G. & Messens, J. Protein sulfenic acid formation: From cellular damage to redox regulation. *Free Radic. Biol. Med.* **51**, 314–326 (2011).
- Poole, L. B. & Nelson, K. J. Discovering mechanisms of signaling-mediated cysteine oxidation. *Curr. Opin. Chem. Biol.* **12**, 18–24 (2008).
- Lo Conte, M. & Carroll, K. S. The redox biochemistry of protein sulfonylation and sulfinylation. *J. Biol. Chem.* **288**, 26480–26488 (2013).
- Kettenhofen, N. J. & Wood, M. J. Formation, Reactivity, and Detection of Protein Sulfenic Acids. *Chem. Res. Toxicol.* **23**, 1633–1646 (2010).
- Yang, J., Gupta, V., Carroll, K. S. & Liebler, D. C. Site-specific mapping and quantification of protein S-sulphenylation in cells. *Nat. Commun.* **5**, 4776 (2014).
- Benitez, L. V. & Allison, W. S. The inactivation of the acyl phosphatase activity catalyzed by the sulfenic acid form of glyceraldehyde 3-phosphate dehydrogenase by dimedone and olefins. *J. Biol. Chem.* **249**, 6234–6243 (1974).
- Furdui, C. M. & Poole, L. B. Chemical approaches to detect and analyze protein sulfenic acids. *Mass. Spectrom. Rev.* **33**, 126–146 (2014).
- Witze, E. S., Old, W. M., Resing, K. A. & Ahn, N. G. Mapping protein post-translational modifications with mass spectrometry. *Nat. Methods* **4**, 798–806 (2007).
- Yang, J. *et al.* Global, *in situ*, site-specific analysis of protein S-sulfonylation. *Nat. Protoc.* **10**, 1022–1037 (2015).
- Cremers, C. M. & Jakob, U. Oxidant sensing by reversible disulfide bond formation. *J. Biol. Chem.* **288**, 26489–26496 (2013).
- Rehder, D. S. & Borges, C. R. Cysteine sulfenic Acid as an Intermediate in Disulfide Bond Formation and Nonenzymatic Protein Folding. *Biochemistry* **49**, 7748–7755 (2010).
- Sevier, C. S. & Kaiser, C. A. Formation and transfer of disulphide bonds in living cells. *Nat. Rev. Mol. Cell. Biol.* **3**, 836–847 (2002).
- Nagy, P. Kinetics and mechanisms of thiol-disulfide exchange covering direct substitution and thiol oxidation-mediated pathways. *Antioxid. Redox. Signal.* **18**, 1623–1641 (2013).
- Depuydt, M., Messens, J. & Collet, J. F. How proteins form disulfide bonds. *Antioxid. Redox. Signal.* **15**, 49–66 (2011).
- Dickinson, B. C. & Chang, C. J. Chemistry and biology of reactive oxygen species in signaling or stress responses. *Nat. Chem. Biol.* **7**, 504–511 (2011).
- Reddie, K. G. & Carroll, K. S. Expanding the functional diversity of proteins through cysteine oxidation. *Curr. Opin. Chem. Biol.* **12**, 746–754 (2008).
- D'Autreaux, B. & Toledano, M. B. ROS as signalling molecules: mechanisms that generate specificity in ROS homeostasis. *Nat. Rev. Mol. Cell. Biol.* **8**, 813–824 (2007).
- Danehy, J. P. & Hunter, W. E. Alkaline decomposition of organic disulfides.2. Alternative pathways as determined by structure. *J. Org. Chem.* **32**, 2047–2204 (1967).
- Pal, B. C., Uziel, M., Doherty, D. G. & Cohn, W. E. Isolation and characterization of a pyrimidine sulfenic acid via scission of sulfur-sulfur bond in methyl analog of bis(4-thiouridine) disulfide. *J. Am. Chem. Soc.* **91**, 3634–363 (1969).
- Florence, T. M. Degradation of protein disulfide bonds in dilute alkali. *Biochem. J.* **189**, 507–520 (1980).
- Stathakis, P. *et al.* Angiotensin formation involves disulfide bond reduction and proteolysis in kringle 5 of plasmin. *J. Biol. Chem.* **274**, 8910–8916 (1999).
- Hogg, P. J. Disulfide bonds as switches for protein function. *Trends. Biochem. Sci.* **28**, 210–214 (2003).

32. Lay, A. J. *et al.* Phosphoglycerate kinase acts in tumour angiogenesis as a disulphide reductase. *Nature* **408**, 869–873 (2000).
33. Lay, A. J., Jiang, X. M., Daly, E., Sun, L. & Hogg, P. J. Plasmin reduction by phosphoglycerate kinase is a thiol-independent process. *J. Biol. Chem.* **277**, 9062–9068 (2002).
34. Garcia-Manyes, S., Liang, J., Szoszkiewicz, R., Kuo, T. L. & Fernandez, J. M. Force-activated reactivity switch in a bimolecular chemical reaction. *Nat. Chem.* **1**, 236–242 (2009).
35. Dopieralski, P. *et al.* The Janus-faced role of external forces in mechanochemical disulfide bond cleavage. *Nat. Chem.* **5**, 685–691 (2013).
36. Hofbauer, F. & Frank, I. CPMD simulation of a bimolecular chemical reaction: nucleophilic attack of a disulfide bond under mechanical stress. *Chemistry* **18**, 16332–16338 (2012).
37. Fernandez, J. M. & Li, H. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science* **303**, 1674–1678 (2004).
38. Garcia-Manyes, S., Brujic, J., Badilla, C. L. & Fernandez, J. M. Force-clamp spectroscopy of single-protein monomers reveals the individual unfolding and folding pathways of I27 and ubiquitin. *Biophys. J.* **93**, 2436–2446 (2007).
39. Garcia-Manyes, S., Dougan, L., Badilla, C. L., Brujic, J. & Fernandez, J. M. Direct observation of an ensemble of stable collapsed states in the mechanical folding of ubiquitin. *Proc. Natl Acad. Sci. USA* **106**, 10534–10539 (2009).
40. Kosuri, P. *et al.* Protein folding drives disulfide formation. *Cell* **151**, 794–806 (2012).
41. Garcia-Manyes, S., Kuo, T. L. & Fernandez, J. M. Contrasting the individual reactive pathways in protein unfolding and disulfide bond reduction observed within a single protein. *J. Am. Chem. Soc.* **133**, 3104–3113 (2011).
42. Vollhardt, P. N. S. *Organic Chemistry: Structure and Function* (Freeman and Company, 2003).
43. Poole, L. B., Karplus, P. A. & Claiborne, A. Protein sulfenic acids in redox signaling. *Annu. Rev. Pharmacol.* **44**, 325–347 (2004).
44. Stirnemann, G., Kang, S. G., Zhou, R. & Berne, B. J. How force unfolding differs from chemical denaturation. *Proc. Natl Acad. Sci. USA* **111**, 3413–3418 (2014).
45. Walther, K. A. *et al.* Signatures of hydrophobic collapse in extended proteins captured with force spectroscopy. *Proc. Natl Acad. Sci. USA* **104**, 7916–7921 (2007).
46. Stirnemann, G., Giganti, D., Fernandez, J. M. & Berne, B. J. Elasticity, structure, and relaxation of extended proteins under force. *Proc. Natl Acad. Sci. USA* **110**, 3847–3852 (2013).
47. Crowder, J. A. & Harris, M. Mechanism of the sulfur lability in the alkali degradation of wool protein. *J. Res. Natl Bur. Stand.* **16**, 475–480 (1936).
48. Toennies, G. Oxidation of cysteine in non-aqueous media—The ‘sulfenic acid’ as the primary oxidation product. *J. Biol. Chem.* **122**, 27–47 (1937).
49. Schieber, M. & Chandel, N. S. ROS Function in redox signaling and oxidative stress. *Curr. Biol.* **24**, R453–R462 (2014).
50. Leonard, S. E., Reddie, K. G. & Carroll, K. S. Mining the thiol proteome for sulfenic acid modifications reveals new targets for oxidation in cells. *ACS Chem. Biol.* **4**, 783–799 (2009).
51. Salsbury, F. R., Knutson, S. T., Poole, L. B. & Fetrow, J. S. Functional site profiling and electrostatic analysis of cysteines modifiable to cysteine sulfenic acid. *Protein Sci.* **17**, 299–312 (2008).
52. Lee, S. R. *et al.* Reversible inactivation of the tumor suppressor PTEN by H₂O₂. *J. Biol. Chem.* **277**, 20336–20342 (2002).
53. Giannoni, E., Buricchi, F., Raugei, G., Ramponi, G. & Chiarugi, P. Intracellular reactive oxygen species activate Src tyrosine kinase during cell adhesion and anchorage-dependent cell growth. *Mol. Cell. Biol.* **25**, 6391–6403 (2005).
54. Cotto-Rios, X. M., Bekes, M., Chapman, J., Ueberheide, B. & Huang, T. T. Deubiquitinases as a signaling target of oxidative stress. *Cell Rep.* **2**, 1475–1484 (2012).
55. Lee, C. *et al.* Redox regulation of OxyR requires specific disulfide bond formation involving a rapid kinetic reaction path. *Nat. Struct. Mol. Biol.* **11**, 1179–1185 (2004).
56. Paulsen, C. E. & Carroll, K. S. Chemical dissection of an essential redox switch in yeast. *Chem. Biol.* **16**, 217–225 (2009).
57. Reynaert, N. L. *et al.* Dynamic redox control of NF- κ B through glutaredoxin-regulated S-glutathionylation of inhibitory κ B kinase beta. *Proc. Natl Acad. Sci. USA* **103**, 13086–13091 (2006).
58. Bogeski, I. *et al.* Differential redox regulation of ORAI ion channels: a mechanism to tune cellular calcium signaling. *Sci. Signal.* **3**, ra24 (2010).
59. Svoboda, L. K. *et al.* Redox-sensitive sulfenic acid modification regulates surface expression of the cardiovascular voltage-gated potassium channel Kv1.5. *Circ. Res.* **111**, 842–853 (2012).
60. Seo, Y. H. & Carroll, K. S. Profiling protein thiol oxidation in tumor cells using sulfenic acid-specific antibodies. *Proc. Natl Acad. Sci. USA* **106**, 16163–16168 (2009).
61. Kahn, T. B., Fernandez, J. M. & Perez-Jimenez, R. Monitoring oxidative folding of a single protein catalyzed by the disulfide oxidoreductase DsbA. *J. Biol. Chem.* **290**, 14518–14527 (2015).
62. Parker, A. J. & Kharasch, N. The scission of the sulfur-sulfur bond. *Chem. Rev.* **583**–628 (1959).
63. Dopieralski, P., Ribas-Arino, J., Anjukandi, P., Krupicka, M. & Marx, D. Force-induced reversal of beta-eliminations: stressed disulfide bonds in alkaline solution. *Angew. Chem. Int. Ed. Engl.* **55**, 1304–1308 (2016).
64. Alegre-Cebollada, J. *et al.* S-Glutathionylation of cryptic cysteines enhances titin elasticity by blocking protein folding. *Cell* **156**, 1235–1246 (2014).
65. Gautel, M. The sarcomeric cytoskeleton: who picks up the strain? *Curr. Opin. Cell. Biol.* **23**, 39–46 (2011).
66. Chung, C. S. *et al.* Shortening of the elastic tandem immunoglobulin segment of titin leads to diastolic dysfunction. *Circulation* **128**, 19–28 (2013).
67. Li, H. *et al.* Reverse engineering of the giant muscle protein titin. *Nature* **418**, 998–1002 (2002).
68. Schlierf, M., Li, H. & Fernandez, J. M. The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proc. Natl Acad. Sci. USA* **101**, 7299–7304 (2004).
69. Popa, I., Kosuri, P., Alegre-Cebollada, J., Garcia-Manyes, S. & Fernandez, J. M. Force dependency of biochemical reactions measured by single-molecule force-clamp spectroscopy. *Nat. Protoc.* **8**, 1261–1276 (2013).
70. Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. Entropic Elasticity of Lambda-Phage DNA. *Science* **265**, 1599–1600 (1994).

Acknowledgements

We thank Ainhoa Lezamiz for protein purification and Roger Morris for critical reading of the manuscript. We thank Aritz Garcia for help in acquiring preliminary data and Julio Fernández and Hagan Bayley for helpful discussions. A.E.M.B. is funded by an EPSRC DTP fellowship. This work was supported by the Marie Curie CIG (293462) grant, BBSRC (J00992X/1) grant, Royal Society Research grant (RG120038), BHF grant (PG/13/50/30426) and EPSRC Fellowship (K00641X/1), all to S.G.-M.

Author contributions

S.G.-M. conceived research. A.E.M.B. and S.G.-M. designed experiments. A.E.M.B. conducted single-molecule experiments and analysed data. S.L. performed mass spectrometry experiments and analysed data. A.E.M.B. and S.G.-M. wrote the paper.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Beedle, A.E.M. *et al.* Protein S-sulfonylation is a fleeting molecular switch that regulates non-enzymatic oxidative folding. *Nat. Commun.* **7**:12490 doi: 10.1038/ncomms12490 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016

APPENDIX VI

ARTICLE

Received 11 Dec 2016 | Accepted 13 Apr 2017 | Published 6 Jun 2017

DOI: 10.1038/ncomms15658

OPEN

Tailoring protein nanomechanics with chemical reactivity

Amy E.M. Beedle¹, Marc Mora¹, Steven Lynham², Guillaume Stirnemann³ & Sergi Garcia-Manyes¹

The nanomechanical properties of elastomeric proteins determine the elasticity of a variety of tissues. A widespread natural tactic to regulate protein extensibility lies in the presence of covalent disulfide bonds, which significantly enhance protein stiffness. The prevalent *in vivo* strategy to form disulfide bonds requires the presence of dedicated enzymes. Here we propose an alternative chemical route to promote non-enzymatic oxidative protein folding via disulfide isomerization based on naturally occurring small molecules. Using single-molecule force-clamp spectroscopy, supported by DFT calculations and mass spectrometry measurements, we demonstrate that subtle changes in the chemical structure of a transient mixed-disulfide intermediate adduct between a protein cysteine and an attacking low molecular-weight thiol have a dramatic effect on the protein's mechanical stability. This approach provides a general tool to rationalize the dynamics of S-thiolation and its role in modulating protein nanomechanics, offering molecular insights on how chemical reactivity regulates protein elasticity.

¹Department of Physics and Randall Division of Cell and Molecular Biophysics, King's College London, WC2R 2LS London, UK. ²Centre of Excellence for Mass Spectrometry, King's College London, SE5 8AF London, UK. ³CNRS Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, Univ. Paris Denis Diderot, Sorbonne Paris Cité, PSL Research University, 13 rue Pierre et Marie Curie, 75005 Paris, France. Correspondence and requests for materials should be addressed to S.G.-M. (email: sergi.garcia-manyes@kcl.ac.uk).

Modular proteins represent a natural strategy to achieve highly elastic materials with enhanced mechanical properties¹. For example, the giant protein titin is formed by a large number of stiff immunoglobulin (Ig) and Fibronectin-like (Fn) domains intercalated between non-structured, extensible sequences (PEVK, N2B) that are mechanically compliant^{2–4}. Combined, the distinct mechanical properties of both structurally diverse components regulate the large-scale passive elasticity of muscle under physiological conditions². At the local scale, modulation of the nanomechanics of a single protein is mainly achieved through four specific molecular mechanisms; first and foremost (i), forced unfolding converts the mechanically rigid native state into an elastic unfolded conformation devoid of mechanical stability⁵. (ii) Secondly, a simple strategy to alter the mechanical properties of the protein's native state is through site-directed mutagenesis in well-defined positions within the force-bearing ('mechanical clamp') structural motif⁶. Alternatively, (iii) ligand binding can drastically increase the mechanical stability of a folded protein^{7–9}, or prevent an unfolded polypeptide from recovering its native mechanical stability by blocking successful refolding¹⁰. Both strategies often result in an all-or-none switch between both extreme mechanical stabilities¹¹. (iv) A far less explored mechanism to modulate protein mechanics is achieved via post-translational modifications, through subtle yet crucial changes in chemical reactivity within the protein core^{12,13}. The most relevant protein modification with a mechanical impact is arguably the formation of disulfide bonds, which are crucial modulators of protein extensibility. Disulfide bonds establish a rigid, molecular shortcut that impairs the complete force-induced unfolding of a variety of proteins including titin¹⁴, and the extracellular cell adhesion molecules (CAM) protein superfamily¹⁵. In these cases, modulation of protein elasticity is generally dynamic and occurs under redox control¹⁶; while oxidizing conditions guarantee the presence of the covalently rigid disulfide bond, reducing conditions induce its rupture, thereby triggering full protein unfolding under force. Hence, the ability of the protein to form a disulfide bond during its folding route, a process known as oxidative folding¹⁷, emerges as a crucial functional determinant of protein mechanics.

In vivo, the formation of disulfide bonds occurs through two independent pathways¹⁸. Within the 'redox route', a cysteine thiol can be oxidized via one or two electrons to a higher oxidation state species. Most typically, the free thiol is first oxidized to sulfenic acid, which then condenses with a neighbour protein thiol to form a protective disulfide bond that prevents irreversible protein damage due to cysteine overoxidation^{13,19}. The second and most prevalent means of disulfide bond formation is catalysed by a superfamily of dedicated oxidoreductases—including thioredoxin, glutaredoxin, the bacterial Dsb and the eukaryotic protein disulfide isomerase enzymes—that share a common active site containing two highly conserved catalytic cysteines^{20,21}. Their mechanism of action employs a thiol-exchange nucleophilic reaction, creating a transient yet obligatory mixed-disulfide intermediate between a substrate cysteine and the *N*-terminal cysteine of the enzyme's active site²². The determinant role of the mixed disulfide conformation to orchestrate oxidative folding has deserved great attention^{23,24}. While the static conformational picture has been accessed through a variety of protein structures^{25–29}, the dynamics of this crucial fleeting mixed disulfide intermediate has only now begun to be understood^{30,31}.

Inspired by the functioning of oxidoreductases enzymes, we now aim to acquire a deep understanding of the naturally employed chemical tactics to rationalize the molecular mechanisms underpinning other closely related protein S-modifications.

We will use this *ad hoc* acquired knowledge to design new experimental strategies that rationally exploit the chemical properties of naturally occurring small molecules to modulate protein elasticity. A main advantage of this experimental approach is that it is not restricted to enzymatic activity, thus having the potential to be scaled up to large quantities, it has a predictive character and it offers exquisite control over the chemical selectivity of the small molecules as it avoids the inextricable conformational changes involved during protein-protein interactions that define enzymatic catalysis.

The most obvious candidates for the small chemical modulators are low-weight molecular thiols (LMW-SH)³², especially encompassing cysteine (Cys), cysteinylglycine (CysGly), homocysteine (Hcys) and glutathione (GSH). Present both intracellularly and also in the human plasma, these small thiols are often regarded as biomarkers for oxidative stress³³, and their presence as mixed-disulfide with proteins increases with age^{34,35}. While GSH reactivity has been the focus of extensive research^{36,37}, the reactivity of Cys and Hcys and their related protein post-translational modifications, namely S-cysteinylation and S-homocysteinylation, have comparatively received much less attention. In particular, high levels of Hcys have been associated to reduced muscle function and integrity³⁸. In fact, hyperhomocysteinemia is considered as an independent risk factor for cardiovascular disease^{39–42}, and has been linked to pathological alterations in mechanically functional proteins such as the impaired fibrillin–fibronectin assembly⁴³. Despite important efforts, a direct molecular link between the presence of S-homocysteinylation and the pathological mechanical effect is not completely understood.

Here we use single-molecule force-clamp spectroscopy, supported by mass spectrometry (MS) and density functional theory (DFT) calculations, to provide mechanistic insights into the direct link between the chemical reactivity of biologically relevant LMW thiols and their effect on mechanical protein folding. Our results demonstrate that the life-time of the ephemeral mixed disulfide intermediate structure is an essential modulator of the nanomechanics of cardiac titin, switching between two successfully refolded conformations with markedly different stiffness properties.

Results

S-cysteinylation promotes non-enzymatic oxidative folding.

We applied a constant force to a polyprotein composed of eight identical domains of a titin Ig27 mutant (I27_{E24C–K55C})₈ that harbours a buried disulfide bond between amino acids 24 and 55 (Fig. 1a). A specifically designed five-pulse force protocol measures the protein elongation over time in the presence of 13.3 mM of L-cysteine (Fig. 1b), enabling the independent capture of each individual disulfide bond rupture and reformation event, with single bond resolution (Fig. 1c)¹³. Applying a short (500 ms) pulse of 150 pN to the (I27_{E24C–K55C})₈ polyprotein elicits steps of ~15 nm, corresponding to the unfolding of the protein up to the position of the rigid disulfide bond (grey), which becomes exposed to the solution. A second pulse at a higher force (300 pN) triggers a ~10 nm stepwise increase in the protein length (red), corresponding to the force-catalysed rupture of each individual disulfide bond by the attacking L-cysteine nucleophile. The initial pulse, composed of ~10 nm and ~15 nm steps (Fig. 1d), results in a fully stretched and reduced protein form exhibiting a free cysteine thiolate and a mixed disulfide conformation between the native cysteine and the attacking L-cysteine (Fig. 1b).

Removal of the pulling force for $t_q = 8$ s (a period of time that is long enough to ensure quantitative protein refolding) triggered the collapse and folding of the protein. A second test pulse that

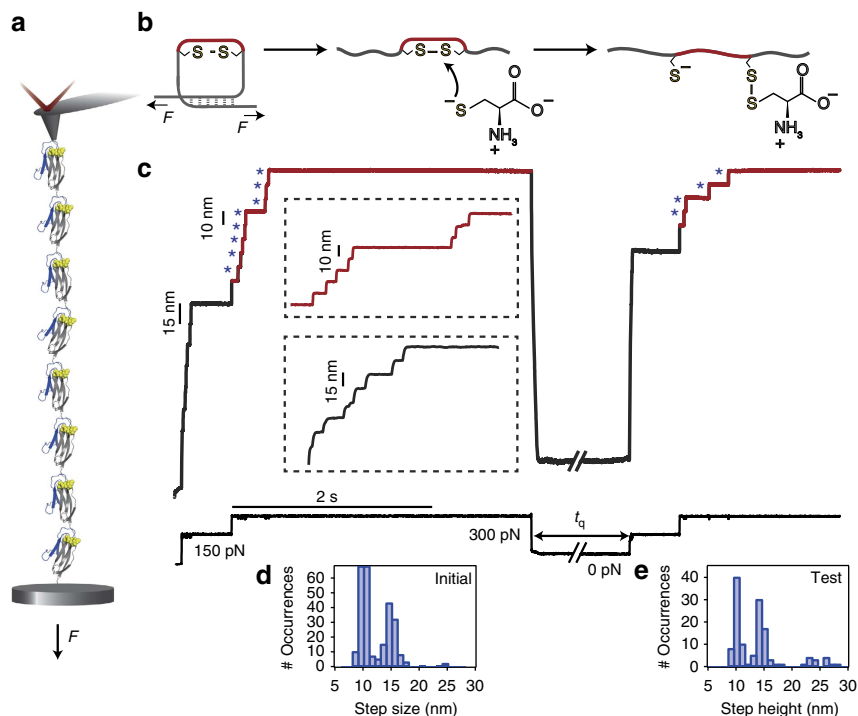


Figure 1 | L-cysteine mediates the non-enzymatic reformation of disulfide bonds. (a) A polyprotein construct formed of eight identical repeats of a mutated I27th Ig domain of titin containing a buried disulfide bond ($I27_{E24C-K55C}$)₈ is tethered between a gold substrate and an AFM cantilever tip, which applied a constant stretching force throughout the experiment. (b) Upon the application of force, the protein unfolds (grey) and exposes the disulfide bond to a L-cysteine solution (13.3 mM, pH = 7.2), which is able to attack the disulfide bond through a S_N2 nucleophilic attack, marked by the extension of the amino acids that were trapped behind the disulfide bond (red). The force-induced cleavage of the protein disulfide results in a stretched protein conformation containing a reduced protein thiol and a mixed disulfide bond between the attacking L-cysteine and a native cysteine. (c) Experimental folding trajectory of the ($I27_{E24C-K55C}$)₈ revealed by a five-pulse force quench protocol. The initial pulse unfolds the protein and ruptures the disulfide bond; the first pulse $F = 150$ pN unfolds and extends all the modules of the ($I27_{E24C-K55C}$)₈ in steps of ~ 15 nm (grey inset), exposing the disulfide bonds to the L-cysteine solution. The second pulse at $F = 300$ pN catalyses the rupture of each disulfide bond (blue asterisks), giving rise to a staircase-like increase of protein length occurring in ~ 10 nm steps (red inset). Quenching the pulling force ($F = 0$ pN) triggers the protein to collapse and fold for a time quench $t_q = 8$ s. The subsequent test pulse probes the folding status of the protein. Mirroring the initial pulse, a force pulse of $F = 150$ pN unfolds and extends the refolded protein up to the newly created disulfide bond (grey line) whereas the last $F = 300$ pN pulse is able to rupture it again (red line). The step-size histograms corresponding to the (d) initial and (e) test pulse highlight in both cases the presence of ~ 10 nm and ~ 15 nm steps.

mirrors the initial pulse probes the folding success of the folding reaction for a given t_q . Similar to the initial pulse, the elongation of the protein in the test pulse occurs in steps of ~ 15 nm (unfolding) followed by the sequential reduction of each individual disulfide bond (~ 10 nm steps, Fig. 1e). This sequence of events fingerprints the success of non-enzymatic oxidative folding process. In particular, the presence of ~ 10 nm steps in the test pulse unambiguously hallmarks the non-enzymatic reformation of individual disulfide bonds, surprisingly occurring in the absence of any oxidising agent, and its intricate link to the successful refolding of the protein.

S-homocysteinylation blocks oxidative folding. To probe the generality of the small thiol-mediated non-enzymatic oxidative folding process, we next probed the outcome of oxidative folding in the presence of the chemically analogous homocysteine nucleophile (Fig. 2a), only differing from L-cysteine in the presence of an extra methylene group next to the nucleophilic thiolate (Fig. 2b). Surprisingly, in this case the test pulse was mostly populated by steps of ~ 25 nm, and the proportion of ~ 10 nm steps, hallmark of disulfide reformation, was vanishingly small (Fig. 2a). The presence of ~ 25 nm in the 150 pN test pulse stands for a mechanically weaker protein that has properly folded but has not reformed its internal stiff disulfide bond. The percentage of reduced, folded protein (~ 25 nm steps) is dramatically

higher ($\sim 40\%$) than in the case of L-cysteine ($\sim 3\%$), whereas the situation is completely reversed in the case of disulfide bond reformation, which is favoured with L-cysteine ($\sim 30\%$) and significantly reduced ($\sim 10\%$) when Hcys is the attacking nucleophile (Fig. 2c,d). Altogether, these results conclude that the attacking nucleophile has a profound effect on the nanomechanics of the refolded protein; while L-cysteine promotes a stiff, refolded protein with a properly reformed disulfide bond, S-homocysteinylation blocks oxidative folding but eventually allows the protein to fold into a more compliant native state conformation void of the reformed disulfide bond. The dramatic contrast in the effect that both analogous post-translational modifications have on the oxidative folding fate of the protein and on its nanomechanical properties is certainly surprising given the closely related chemical structures of both low-molecular-weight thiols.

The mixed disulfide plays a key role in disulfide formation.

It is very plausible that the mechanism underpinning the non-enzymatic reformation of a protein disulfide bond observed for L-cysteine and also—to a much lower extent—for homocysteine, involves the reactivity of the mixed disulfide intermediate. MS measurements on the $I27_{E24C-K55C}$ monomer detected with high evidence the post-translational modification of both 24 and 55 cysteines by both L-cysteine and homocysteine (Supplementary

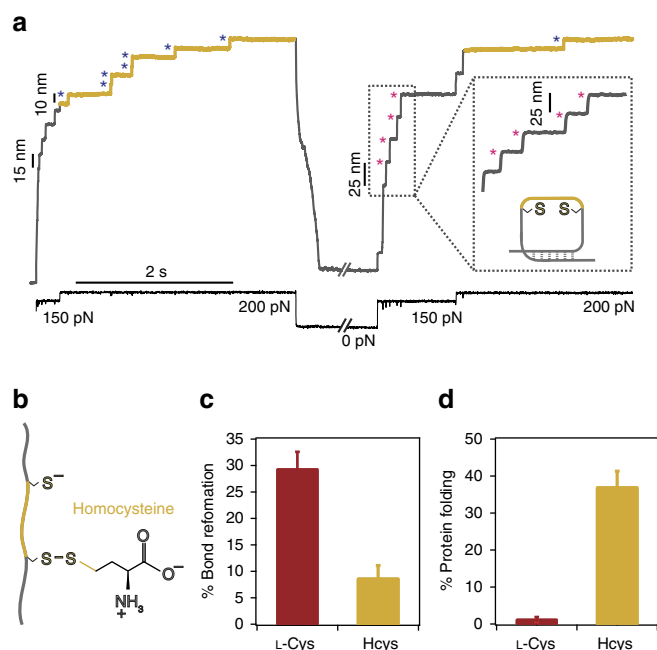


Figure 2 | DL-homocysteine prevents successful non-enzymatic oxidative folding. (a) The oxidative folding trajectory in the presence of homocysteine (111 mM, pH = 7.6) exhibits a test pulse that is composed of steps of ~25 nm (marking the refolding of a reduced protein form, pink asterisks) followed by a vanishingly small population of ~10 nm steps, fingerprinting disulfide bond reformation (blue asterisks). (b) The asymmetric mixed disulfide between a native cysteine and the attacking homocysteine, which only differs from L-cysteine in the presence of a methylene group (yellow) next to the nucleophilic thiol plays a crucial role in the outcome of the folding reaction. (c) While the percentage of disulfide bond reformation in the presence of homocysteine (yellow) is negligible compared to that of L-cysteine (red), (d) the situation is reversed for the population of folded and reduced proteins (~25 nm steps), which is largely predominant for homocysteine (yellow) when compared to the residual fraction measured for L-cysteine (red) (error bar $\pm$ s.d.).

Figs 1,2 and Supplementary Table 1), which unambiguously confirmed the presence of the mixed disulfide. Noteworthy, given the different degree of solvent exposure of the protein disulfide bond obtained in MS and mechanical measurements, the yield of protein modification is not comparable in both experimental approaches. To prove the necessary participation of the mixed disulfide adduct in the dynamics of disulfide bond reoxidation, we conducted analogous experiments on an equivalent (I27_{G32C-A75C})₈ polyprotein using the phosphorus-based TCEP nucleophile (we chose the I27_{G32C-A75C} mutant, harbouring the buried disulfide bond in a different position within the protein structure, because TCEP can partially reduce the disulfide bond of the (I27_{E24C-K55C})₈ polyprotein in the absence of force⁴⁴). In the case of the P-based nucleophile TCEP, the force-induced S_N2 disulfide cleavage⁴⁵ obviously does not yield a stable mixed disulfide, but rather a highly reactive mixed S-P thioalkoxyphosphonium cation intermediate⁴⁶ (Supplementary Fig. 3), which can be easily displaced by water molecules to render a completely unfolded protein with two reduced thiolates (Fig. 3a). Analysis of the test pulse in our refolding trajectories (Fig. 3b) reveals the absolute absence of ~10 nm steps, indicating that no disulfide bond is present in the otherwise successfully refolded protein (confirmed by the presence of ~25 nm steps, Fig. 3c). These experiments demonstrated the crucial implication of the mixed disulfide species in the thiol-mediated mechanism of oxidative folding.

Homocysteine exhibits an anomalous high pK_a. To elucidate the molecular origin of the markedly different behaviour observed for L-cysteine and homocysteine, we investigated how the physico-chemical properties of both small thiols could determine their reactivity within the context of a nucleophilic substitution (S_N2) reaction. It is well established that a charged thiolate will be a better nucleophile than a protonated thiol⁴⁷. Hence, there is a general direct connection between the goodness of a particular nucleophile/leaving group and its acid-base properties. Although basicity is a thermodynamic property and nucleophilicity is a kinetic phenomenon, a good correlation between both properties has been observed, generally following the rule of thumb stating that good leaving groups are the conjugate bases of strong acids⁴⁸. While L-cysteine is reported to exhibit a pK_a = 8.15 (ref. 49) – 8.3 (ref. 50), there is a surprising disagreement in the literature regarding the pK_a value for homocysteine, for which two largely disparate values (pK_a = 8.7 (ref. 51) and pK_a = 10 (refs 50,52,53)) are reported. To resolve such a dichotomy, we independently measured the pK_a of both compounds by monitoring the free thiol absorbance at λ = 240 nm (refs 54,55) as a function of pH for both L-cysteine and homocysteine compounds (Fig. 4a). While an expected pK_a = 8.4 was measured for L-cysteine, the pK_a of homocysteine was found to be significantly higher (pK_a = 9.9) in our experimental conditions, supporting previous findings⁵⁰. Crucially, our homocysteine measurements required reducing conditions to attenuate its fast dimerization, especially under basic conditions (Supplementary Fig. 4). The dissimilar pK_a value found for both isolated compounds in solution is likely to account for their different reactivity when selectively bound to a protein via a protein post-translational modification.

DFT calculations provide a thermodynamic perspective.

To obtain a more precise understanding of the thermodynamically allowed reactivity between the different thiol species involved in our experiments (namely (i) the initial disulfide bond between two structurally close cysteines, (ii) the newly mixed disulfide bond species; (iii) the resulting free cysteine thiol after disulfide bond cleavage and (iv) the free small thiol nucleophile species in solution), we performed DFT calculations on model thiol and disulfide species using the expanded 6-311 + G(d,p) basis set and the B3LYP functional, which has been used in the context of thiol-disulfide exchange⁵⁶ (Table 1). Noteworthy, in this simplified approach any free-energy contributions arising from a protein conformational change upon formation of a native or mixed disulfide are neglected (only the free-energy associated with an isolated disulfide bond rupture is computed).

The attack of the Cys_{prot}-Cys mixed disulfide bond by a solution L-cysteine is, from a free-energy perspective, an equilibrium process with no net energy contribution ($\Delta G = 0$ kcal mol⁻¹ reaction 1, Table 1), whereas the rupture of the mixed non-equivalent disulfide between Cys_{prot}-Hcys by free Hcys in solution (reaction 2, Table 1) is favourable ($\Delta G = -4$ kcal mol⁻¹). By contrast, the attack of the same mixed disulfide (Cys_{prot}-Hcys) by the freed protein cysteine (reaction 3, Table 1) is not favourable ($\Delta G = +2.6$ kcal mol⁻¹). Hence, these results can rationalize a few key non-expected experimental observations, namely that: (i) the symmetric mixed disulfide formed by two equivalent cysteines can be attacked by the freed protein thiolate when in close physical proximity to reform a disulfide bond (Fig. 4b); by contrast, (ii) the asymmetric Cys_{prot}-Hcys is not quantitatively attacked by the free thiol protein cysteine to reform the previously cleaved disulfide bond but (iii) can be attacked by free homocysteine molecules in solution, thus resolving the mixed disulfide and leaving the protein with two reduced thiolates able to successfully refold in the absence of a formed disulfide bond giving rise to the ~25 nm steps.

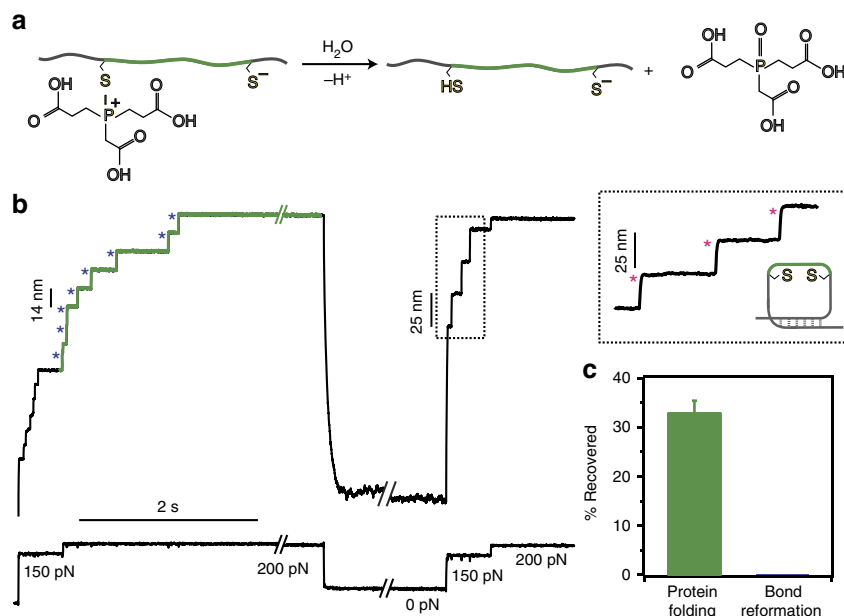


Figure 3 | TCEP does not create a mixed disulfide and therefore the protein refolds into a conformation void of the disulfide bond. (a) When a phosphorous nucleophile such as TCEP is used to induce the force-activated cleavage of the disulfide bond on a structurally analogous (I27_{G32C-A75C})₈ polyprotein (4 mM, pH 7.2), a highly unstable mixed S-P thioalkoxyphosphonium cation intermediate is formed, which can be rapidly resolved by water solvolysis. TCEP-mediated disulfide bond rupture results in a stretched protein harbouring two reduced cysteines. (b) The test pulse in the folding trajectories is composed of steps of ~25 nm (pink asterisks), fingerprinting a natively folded protein devoid of the rigid disulfide bond. (c) Crucially, the absence of a mixed disulfide precludes the reformation of the native disulfide bond in the context of non-enzymatic oxidative folding (error bar $\pm$ s.d.).

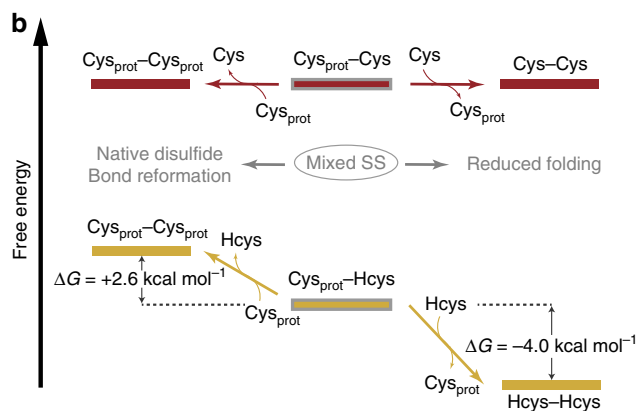
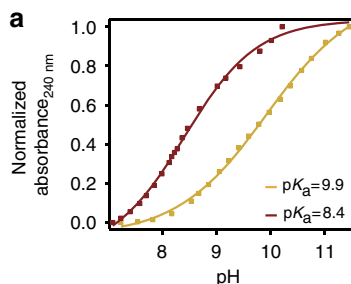


Figure 4 | The high pK_a of homocysteine leads to a thermodynamically stable asymmetric mixed-disulfide. (a) Free homocysteine exhibits a surprisingly higher $pK_a = 9.9$ (red) than the homologous L-cysteine ($pK_a = 8.4$, yellow) revealed by absorbance measurement at $\lambda = 240$ nm in solution. (b) DFT calculations predict the high stability of the Cys-Hcys mixed disulfide, which cannot be attacked by a neighbouring protein free thiol ($\Delta G = +2.6$ kcal mol⁻¹) to reform the native disulfide.

Table 1 | DFT free-energy calculations corresponding to the disulfide-exchange reactions induced by the studied low-molecular-weight thiols.

Reaction	ΔG (kcal mol ⁻¹)
1 Cys-Cys + Cys ⁻ → Cys-Cys + Cys ⁻	0.0
2 Cys-Hcys + Hcys ⁻ → Hcys-Hcys + Cys ⁻	-4.0
3 Cys-Hcys + Cys ⁻ → Cys-Cys + Hcys ⁻	+2.6
4 Cys-NAC ⁻ + Cys ⁻ → Cys-Cys + NAC ²⁻	+7.0
5 Cys-NAC ⁻ + NAC ²⁻ → NAC-NAC ²⁻ + Cys ⁻	-7.7

The distinct nucleophilic power of L-cys and Hcys. While free-energy DFT calculations agree with our (thermodynamic) acid-base measurements, full understanding of the S_N2 nucleophilic substitution reaction would also require kinetic considerations. At the fundamental level, thiol-disulfide exchange is a classical bimolecular nucleophilic exchange S_N2 reaction that occurs through a linear (S-S-S) Walden transition state complex^{47,56,57}. Crucially, mechanical force accelerates disulfide cleavage by lowering the kinetic barrier^{45,58,59}. To compare the nucleophilic power of both small thiols, we measured the rate of disulfide reduction at the same stretching force ($F = 200$ pN) for two solutions containing the same concentration of active, deprotonated thiolate species at the same pH (Fig. 5a). Our results demonstrate that the rate of disulfide reduction is clearly higher for homocysteine ($r = 1.11$ s⁻¹) than for cysteine ($r = 0.08$ s⁻¹), thus highlighting the better nucleophilic nature of homocysteine (Fig. 5b). From the computational perspective, a very crude estimate of the relative nucleophilic character of both cysteine and homocysteine compounds can be obtained by comparing the energies of the HOMO orbitals of the isolated small compound. Our calculations revealed (Supplementary Fig. 5) that the HOMO energy for Hcys is higher than that of L-Cys, suggesting an enhanced reactivity toward electrophiles.

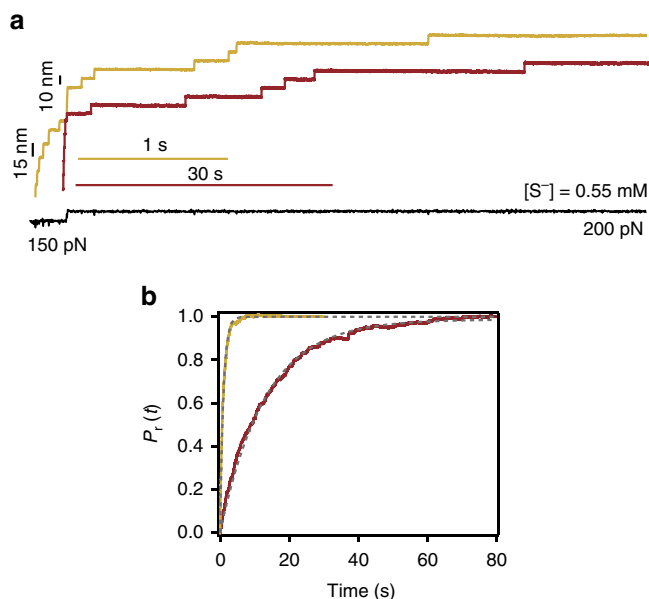


Figure 5 | The kinetics of disulfide reduction is largely accelerated by homocysteine due to its high nucleophilic power. (a) A two-pulse force protocol measures the kinetics of disulfide reduction when the protein is exposed to the same concentration of active deprotonated species of L-cysteine (red) and homocysteine (yellow). The time-course of disulfide bond reduction is dramatically longer for L-cysteine as illustrated by the drastically different scale bars (b) The rate of disulfide bond reduction at $F = 200$ pN is ~ 14 -fold faster for homocysteine ($r = 1.11 \text{ s}^{-1}$) than for L-Cys ($r = 0.08 \text{ s}^{-1}$) under the same experimental conditions, demonstrating that homocysteine is a better nucleophile than L-cysteine.

Similarly, the effective electrostatic negative charge localized on the attacking sulfur atom in homocysteine is more negative ($-0.81e$) than in L-cysteine ($-0.77e$). Altogether, these results suggest that Hcys is a better nucleophile (and thus, a worse leaving group) than L-Cys (Supplementary Fig. 5), qualitatively agreeing with the experimental results and the DFT free-energy calculations (Table 1).

Noteworthy, precise quantification of both thermodynamic and kinetic parameters defining the system would require taking into account the subtle yet important effects of the larger and complex protein environment, and also the non-trivial, catalytic effect that the pulling force has on the reaction kinetics^{58,59}. Further free-energy calculations at the QM/MM level with explicit solvent, hopefully coupled with a force perturbation^{60–62} (which goes beyond the scope of the current work), will help elucidate the finer parameters (for example, solvation, bond orientation⁶²) regarding the mechanistic dynamics of the process. In any case, our simple estimates regarding thermodynamic and kinetic considerations of the isolated reactants provide a good general qualitative description of the experimental results, describing a localized chemical reaction occurring in the core of a much complex protein environment.

The experimental approach predicts the reactivity of NAC. To further test if these considerations are sufficient predictors of disulfide reactivity and thus, of the final outcome of the oxidative folding reaction, we conducted similar nanomechanical experiments using *N*-acetylcysteine (NAC) as a nucleophile (Fig. 6a,b). NAC is a drug derivative of cysteine with an acetyl group attached to its nitrogen (Fig. 6a), which has been used in clinical practice for decades as a mucolytic agent and also employed in the

treatment of numerous disorders including paracetamol intoxication and vascular and cardiac injury⁶³. The uniqueness in the action of NAC stems from its particular physicochemical properties, which render it an optimal reducing agent of protein disulfide bonds⁶⁴. Measurement of the pK_a for NAC through our absorbance experiments yielded a value $pK_a = 9.6$ (Fig. 6c). Similarly, DFT calculations predicted that, from a thermodynamics perspective, the reaction whereby a free protein cysteine attacks a NAC–Cys_{prot} mixed disulfide bond is energetically impaired ($\Delta G = +7 \text{ kcal mol}^{-1}$, reaction 4, Table 1). By contrast, the NAC–Cys_{prot} mixed disulfide can be attacked by free solution NAC molecules ($\Delta G = -7.7 \text{ kcal mol}^{-1}$, reaction 5, Table 1). Moreover, HOMO and electrostatic calculations (Supplementary Fig. 5) highlight the high nucleophilic power of NAC with respect to L-cysteine. Hence, from the physicochemical viewpoint, NAC is closely related to Hcys. In effect, our single-molecule folding trajectories in the presence of NAC (Fig. 6b) certify that while the protein can fold into its reduced form (Fig. 6d), its ability to effectively reform the previously reduced disulfide bond is vanishingly small (showing an even smaller percentage than homocysteine, Fig. 6e). The presence of the NAC–protein mixed disulfide was further certified through MS analysis (Supplementary Fig. 6).

Discussion

Detection of chemical intermediates is the Holy Grail in the elucidation of reaction mechanisms. However, due to their ephemeral nature, their capture is frequently challenging. An extra layer of complexity is added when the chemical reactions occur within the core of an individual protein—giving rise to the generally called post-translational modifications—mainly because the reactivity of the newly created intermediate adduct is intrinsically coupled with the conformational motions of the protein. Hence, direct observation of the chemical evolution of a protein-based chemical reaction and its crucial effect on the folding fate of a protein remains largely elusive.

Due to the highly polarizable sulfur atom, the amino acid cysteine is a particular target for numerous chemical modifications in the cellular milieu⁶⁵, of which the most commonly known is the formation of disulfide bonds. Such thiol-disulfide exchange reactions have crucial roles in biology, both stabilizing structural protein motifs and also as essential regulators of the dynamic functional properties of many enzymes, including most oxidoreductases¹⁸. Albeit perhaps less studied, another important natural example of disulfide bond formation involves small plasma thiols. Indeed, an increase in plasma disulfides, including low-molecular-mass protein-mixed disulfides, has been postulated as a biomarker of oxidative stress and aging. In particular, direct correlations were found between the disulfide forms of Cys and Hcys and age^{33–35}. Several pathologies, including cognitive impairment, osteoporosis, eye's disorders, complications in pregnancy, reduced skeletal muscle function and integrity and cardiovascular disease have been directly correlated with high homocysteine levels⁵⁰. Hence, it is enticing to speculate about the presence of a set of distinct chemical properties of Hcys that underlie its ability to dimerize and to form stable disulfide bonds with protein cysteine residues that impair or alter normal protein function. Several proteins with distinct biological function such as albumin⁴⁹, fibronectin, fibrillin, tropoelastin⁴³, transthyretin, alpha-crystallin⁶⁶ and metallothionein have been identified as natural Hcys targets³⁰. While a few pioneering studies postulated a molecular-targeting hypothesis that link the S-homocysteinylolation post-translational modification with protein function^{50,52}, direct experimental evidence remained elusive.

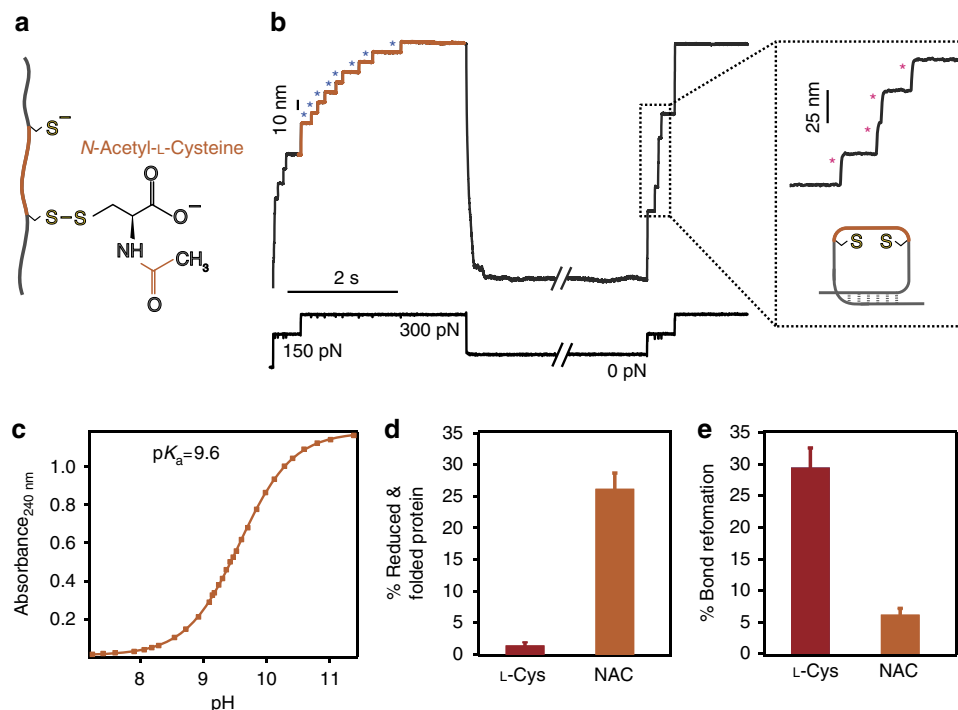


Figure 6 | N-acetylcysteine creates a stable protein mixed-disulfide that largely blocks oxidative folding. (a) The force-induced cleavage of a protein disulfide bond results in an asymmetrical mixed disulfide intermediate of a protein cysteine with NAC, a cysteine derivative with an acetyl group (orange) attached to its nitrogen. (b) The individual single-molecule trajectories track the success of the oxidative folding assay. The initial pulse results in the unfolding of the (I27_{E24C-K55C})₈ polypeptide (15 nm steps, grey) followed by the force-induced cleavage of the solvent-exposed disulfide bonds, marked by the 10 nm step-wise increase of the protein length (blue asterisks). By contrast, the test pulse is mostly composed of 25 nm steps (pink asterisks), which hallmark that the protein has successfully refolded into a reduced conformation that is void of the native disulfide bond. (c) pK_a determination through absorbance measurements at $\lambda = 240$ nm yields a pK_a = 9.6. These results demonstrate that NAC is a weak acid, and hence its conjugated base will be a poor leaving group, in agreement with the high stability of the NAC-mediated mixed disulfide. (d) While the % of successful folding to a reduced conformation is largely prevalent in the presence of NAC, (e) the % of disulfide bond reformation is vanishingly small for NAC as compared to the structurally akin L-cysteine (error bar $\pm$ s.d.).

Our single-molecule approach enables us to independently monitor the protein folding and disulfide bond formation events, hence allowing direct observation of both the molecular mechanisms underlying oxidative folding and the effect that closely related post-translational modifications have on the protein's mechanical function. Here we present a chemical strategy to promote successful non-enzymatic oxidative protein folding that is based on the high reactivity of a symmetric Cys_{prot}-Cys mixed disulfide obtained through protein S-cysteinylation. In this scenario, being in energetic equilibrium with a neighbouring protein thiol, the shuffling between three equivalent cysteines when in close spatial proximity allows disulfide isomerization⁶⁷, eventually leading to the spontaneous reformation of the native disulfide bond. By contrast, homocysteine forms a highly stable asymmetrical disulfide bond with a protein cysteine that can hardly be re-attacked by the previously freed neighbour thiol. We hypothesize that the high propensity of Hcys to form stable mixed disulfides with protein cysteines can be a generic mechanism that underlies the lack of protein functionality upon protein S-homocysteinylation. Similarly, the high affinity of NAC (and in general, all the tested small thiols) to create disulfide bonds might explain its important role in protein protection under oxidative stress⁶⁸, avoiding irreversible thiol overoxidation to sulfinic and sulfonic acid.

A particular characteristic of our experiments is that chemical modifications occur on cryptic cysteines, which only become solvent exposed after mechanical unfolding of the protein. In this

sense, the outcome of the folding process is not only limited by the chemical reactivity of the protein, but also by protein conformation, which is precisely modulated by mechanical force. Hence, our findings suggest a combined mechanochemical scheme that reveals how the distinct physicochemical properties and reactivity of closely related small thiols dictates the folding fate of the protein, giving rise to two successfully refolded forms, yet with very distinct mechanical properties (Fig. 7a). In our approach, the native disulfide bond is first reduced by the attacking nucleophile thiol upon application of force. While the redox potential is suggested to be slightly modulated by the pulling force⁶², force-activated reduction of a protein disulfide bond is mostly under kinetic control⁵⁸, resulting in a protein stretched conformation containing a completely solvent-exposed mixed disulfide which is *a priori* susceptible to be attacked by the thiol nucleophiles present in solution.

Given that in the presence of TCEP (where no mixed-disulfide bond can occur) the protein successfully refolds (certified by the presence of 25 nm steps in the test pulse), we speculate that, in the presence of L-Cys, Hcys and NAC, the observed 25 nm steps (especially obvious with the experiments with Hcys and NAC) correspond to a scenario whereby free thiolates in solution are able to attack the mixed disulfide conformation (namely Cys-Cys_{prot}, Hcys-Cys_{prot} and NAC-Cys_{prot}), therefore resulting in a protein containing two reduced thiolates that are able to properly refold, void of the disulfide bond (Fig. 7b). Indeed, DFT calculations show that the attack of solution free Hcys to the solvent-exposed, mixed asymmetrical Hcys-Cys mixed disulfide

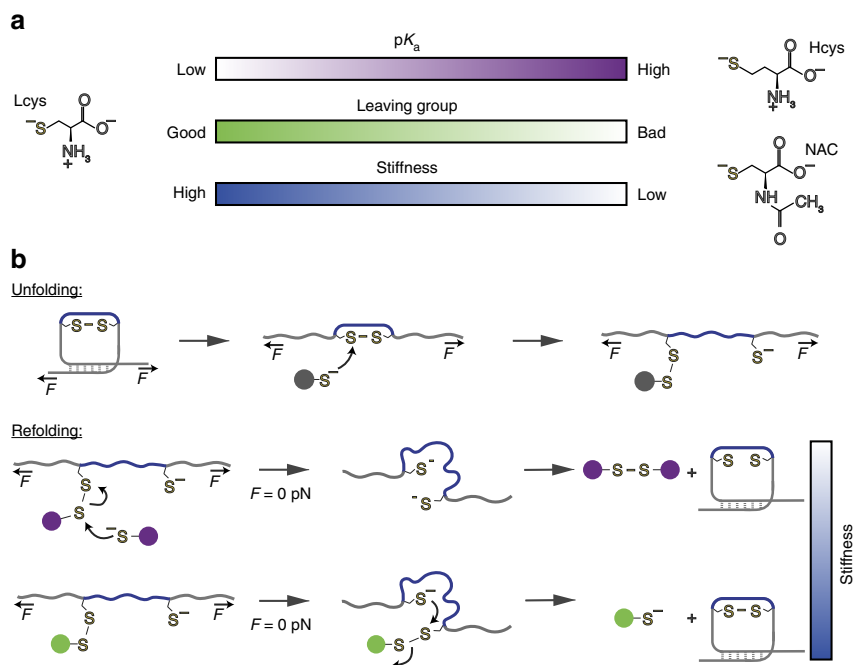


Figure 7 | Protein reactivity induced by S-thiolation regulates protein nanomechanics. (a) The physicochemical properties of the different small thiols govern the life-time of the mixed disulfide, which largely controls the outcome of the oxidative folding reaction and determines protein nanomechanics. (b) Schematics of the reactivity of the S-thiolated protein under force. On the application of mechanical force, the different small thiolates rupture a protein disulfide bond, creating a mixed disulfide and a reduced protein thiolate. In such extended and reduced protein conformation, the protein thiol that is bound to a strong nucleophile acts as the leaving group when attacked by free thiols in solution (purple spheres, for example, homocysteine and NAC) resulting in two reduced protein cysteines. In the absence of force ($F = 0$ pN), the reduced protein can recover a more compliant folded form that is devoid of the native disulfide bond. By contrast, when a softer nucleophile is used (green sphere, for example, L-cysteine), the highly reactive symmetric mixed disulfide is in equilibrium with the free protein cysteine thiol. In this scenario, disulfide bond isomerization occurs, resulting in a stiff folded protein characterized by the reformation of the native disulfide bond.

is thermodynamically very favoured ($\Delta G = -4.0$ kcal mol⁻¹, reaction 2, Table 1), and it is very likely that this equilibrium is further displaced due to the high insolubility of the Hcys-Hcys homodimer (additional small molecule MS experiments reveal that monomeric Hcys in solution undergoes oxygen-mediated oxidation to its dimeric, insoluble form⁴² within the timescale of ~ 1 h, Supplementary Fig. 7). Similarly, the attack of NAC free molecules in solution to the NAC-Cys_{prot} heterodimer is even more favourable from the thermodynamic perspective ($\Delta G = -7.7$ kcal mol⁻¹, reaction 5, Table 1). These free-energy considerations are in very good agreement with our experimental observations, showing that while the percentage of reduced folding is vanishingly small for L-Cys (Fig. 2d), it is instead predominant for both Hcys (Fig. 2d) and also for NAC (Fig. 6d). Further experiments where the time the reduced protein is left extended and exposed to NAC molecules in the solvent (Supplementary Fig. 8) confirmed that the percentage of reduced folding (requiring a solution NAC molecule attacking the asymmetric disulfide bond) remained invariable over time, suggesting that the overall reactivity process seems to occur under thermodynamic control, probably reflecting the simplicity of the reaction and the lack of conformational and steric hindrance.

Supported by recent findings using a cysteine-defective thioredoxin mutant³⁰ and reduced glutathione¹², we speculate that the presence of the mixed disulfide blocks protein folding (giving rise to a compliant protein form), and that it needs to be resolved, either through the attack of a neighbouring protein cysteine to reform the native disulfide bond after protein collapse or via reaction with thiols free in solution. However, we cannot completely rule out at this stage the possibility that the protein is able to refold around its mixed-disulfide and regain mechanical stability.

Upon withdrawal of the applied force, the protein collapses into a molten globule conformation⁶⁹ where reactive cysteines are physically close and sterically buried from the protein environment. In the case of L-cysteine, disulfide bond isomerization⁶⁷ allows the reformation of the native disulfide bond, giving rise to a rigid protein form (Fig. 7b). By contrast, in the case of Hcys and NAC where the mixed disulfide bond has been resolved by reacting with free solution nucleophiles while the protein is stretched under force, quenching the force results in a refolded and reduced protein form that displays a mechanical stability that is intermediate between the mechanically stiff oxidized protein containing the disulfide bond, and the mechanically compliant misfolded protein that harbours a stable, non-resolved mixed disulfide species (Fig. 7b).

The measurements presented here are conducted on an Ig domain of cardiac titin, a protein that carries an important mechanical role *in vivo*. A key requirement of elastomeric proteins, always subject to multiple stretching-relaxation cycles, is their reversible functionality and reliability over time⁷⁰. Our approach provides an extra level of understanding on the molecular mechanisms of disulfide bond reformation under force, and its direct link to mechanical folding. Crucially, our measurements reflect the largely distinct role of the closely related S-cysteinylation and S-homocysteinylation on protein elasticity, and add to the emerging research field that highlights the importance of force-induced exposure of cryptic sites to regulate protein nanomechanics, both via post-translational modifications, such as S-glutathionylation¹² and S-sulfenylation¹³, or through the unravelling of metal binding sites⁷¹. More generally, our proof-of-principle approach demonstrates the use of highly localized chemical reactivity—which can be potentially

expanded to other distinct chemical mechanisms—to rationally tailor protein elasticity, and offer a new approach to engineer reversible redox-responsive protein-based biomaterials with adaptable mechanical properties.

Methods

Protein engineering. The (I27_{E24C-K55C})₈ and (I27_{G32C-A75C})₈ polyproteins were constructed using the BamHI, BglII and KpnI restriction sites (polyprotein sequences shown in Supplementary Table 2). The engineered polyproteins and the I27_{E24C-K55C} monomer (synthetic gene, Thermo Fisher Scientific) were then cloned into the pQE80L (Qiagen) expression vector, and transformed into the BLR(DE3) *Escherichia coli* expression strain (VWR International). Cells were grown at 37 °C and supplemented with 100 µg ml⁻¹ ampicillin. After an OD₆₀₀ of ~0.6 was reached, cells were induced with Isopropyl β-D-1-thiogalactopyranoside (1 mM) and incubated overnight at 20 °C. Cells were disrupted using a French Press. The polyproteins obtained after lysis were purified by metal affinity chromatography on Talon resin (Clontech) using equilibration buffer (50 mM sodium phosphate pH 7.0, 300 mM NaCl), washing buffer (50 mM sodium phosphate pH 7.0, 300 mM NaCl, 20 mM Imidazole) and elution buffer (50 mM sodium phosphate pH 7.0, 300 mM NaCl, 250 mM Imidazole). This was followed by gel-filtration using a Superdex 200 10/300 GL column (GE Biosciences) using PBS buffer pH 7.3. We measured the protein concentration using the Bradford assay.

Force-clamp spectroscopy. To perform the single-molecule force-clamp spectroscopy atomic force microscopy experiments we used both a home-made set-up⁷² and a commercial Luigs and Neumann force spectrometer⁷³. We deposited 1–10 µl of protein in PBS solution (at a concentration of 1–10 mg ml⁻¹) onto a freshly evaporated gold coverslide. Prior to each experiment we independently calibrated the cantilever (Si₃N₄ Bruker MLCT-AUHW) by means of the equipartition theorem, exhibiting a typical spring constant of ~12–17 pN nm⁻¹. Individual polyprotein molecules were fished by pressing the cantilever onto the gold surface at forces spanning 500–1,500 pN, thus enhancing non-specific physisorption. The deflection of the cantilever (force) was then kept constant throughout the experiment (up to ~80 s) thanks to a proportional, integral and differential amplifier (PID) feedback system, achieving a time response to ~2–5 ms. The data was filtered at 1 kHz with a pole Bessel filter. In the disulfide bond rupture kinetics experiments, the protein was first unfolded at 150 pN for 0.5 s, and subsequently the force was raised up to 200 pN and left constant for a large, variable period of time to capture the full kinetics of disulfide reduction for each studied nucleophile. In the oxidative folding experiments, the protein was unfolded at 150 pN for 0.5 s, followed by a high force pulse (varying between 200 and 300 pN depending on the use nucleophile) for 3 s. The quench pulse where the pulling force is completely removed was set in all cases to $t_q = 8$ s. Experiments were carried out by keeping the concentration of the deprotonated form of L-cysteine and N-acetyl-cysteine nucleophiles constant to 0.7 mM, whereas the deprotonated form of D,L-homocysteine was kept at 0.55 mM due to solubility issues, and the total concentration of each nucleophile in the measuring solution was likewise adjusted according to the pK_a of each compound. The different thiol nucleophiles (L-cysteine (Santa Cruz biotechnology, 98%), D,L-homocysteine (Sigma-Aldrich, 95%), N-acetyl-cysteine (Sigma-Aldrich, 98%) and TCEP (Sigma-Aldrich)) were diluted into a sodium phosphate buffer solution, specifically, 50 mM sodium phosphate (Na₂HPO₄ and NaH₂PO₄), 150 mM NaCl. The final pH of in each solution was adjusted by adding the required amounts of HCl/NaOH solutions. Measuring solutions were filtered through a 0.2-µm membrane prior to each experiment. Each solution was freshly prepared prior to every experiment.

Data analysis. Both the acquisition and analysis of the experimental data was performed using customized software written in Igor Pro 6.0 (Wavemetrics). For the kinetics studies, we only analysed trajectories exhibiting at least five unfolding events (15 nm steps) followed by five reduction events (10 nm steps). We followed the same criteria for the oxidative folding experiments, and we considered only trajectories whereby the test pulse reached the full unfolded length characterizing the initial pulse. We used the bootstrap method, to estimate the standard error of the mean (s.e.m.) for the refolding fraction.

DFT calculations. Electronic structure calculations using DFT were performed using the Gaussian 09 software⁷⁴. The B3LYP correlation function^{75,76} was employed with the 6-311 + G(d,p) basis set. All structures were optimized in implicit water solvent consisting of a dielectric medium using the PCM model⁷⁷. Frequency calculations in the gas phase were performed on these optimized structures. The reported free energies correspond to the sum of gas-phase energy at 0 K, correction for solvation, zero-point energy, and thermal correction at 298 K. The HOMO energy levels, the electrostatic potentials and the charges (obtained using the Natural Bond Orbital analysis implemented in Gaussian) were determined from the optimized structures in implicit solvent.

To mimic protein–protein and mixed disulfide bond, we considered dimers of cysteines or its derivatives. At neutral pH, amino acids would be present in a zwitterionic NH₃⁺...COO⁻ form. However, as already observed for amino acids

(including cysteine^{78,79}), the neutral tautomer NH₂...COOH was found to be more stable in most cases. Only the inclusion of explicit water molecules would provide the necessary stabilization of the zwitterionic state by the formation of strong hydrogen bond with water⁸⁰, but it is beyond the scope of the current work. The reported energies therefore refer to the NH₂...COOH forms of thiolates for cysteine and homocysteine. For NAC, the NH₂ group is replaced by an amide group, which is not protonated at neutral pH, so the NH...COO⁻ form was considered.

Mass spectrometry experiments. Enzymatic digestion. The I27_{E24C-K55C} monomer samples were prepared by diluting 20 µl of protein (0.5 µg µl⁻¹) in 20 µl PBS pH 7.4 containing L-cysteine, homocysteine and N-acetyl cysteine at the same concentration as used in the single molecule experiments. Prior to the MS analysis, samples were exposed for 2 h at 37 °C to trypsin digestion at a ratio of 1:20 (enzyme:substrate).

Liquid chromatography–mass spectrometry. An EASY NanoLC system (ThermoFisherScientific, UK) was used to conduct the chromatographic separations. We used a reversed phase chromatography on a 75 µm C18 column to resolve the peptides from a total protein amount on column of 2 µg. A three-step linear gradient of acetonitrile in 0.1% formic acid was employed to elute the peptides at a flow rate of 300 nl min⁻¹ over 60 min. The eluate was ionized by electrospray ionization using an Orbitrap Velos Pro (ThermoFisherScientific, UK) operating under Xcalibur v2.2. Precursor ions were selected according to their intensity using the collision-induced fragmentation employing a Top20 CID method and a Top10 high-energy collision dissociation method. The MS/MS collision dissociation analyses were performed by using the collision energy profiles, which were chosen according to both the mass-to-charge ratio (m/z) and the charge state of the peptide.

Ultraviolet spectrophotometric measurements for pK_a determination.

Absorbance measurements were performed with a JENWAY 6305 Spectrophotometer at room temperature. Absorbance measurements of L-cysteine and N-acetyl-L-cysteine at a final concentration of 250 µM in PBS buffer were performed at $\lambda = 240$ nm from acidic to alkaline pH values, by successive addition of small aliquots of sodium hydroxide. All absorbance measurements were performed at $\lambda = 240$ nm (refs 54,55) at a final concentration of 250 µM with 1 mM of tris(2-carboxyethyl)phosphine to prevent dimerization in PBS buffer. The basal absorbance of tris(2-carboxyethyl)phosphine at $\lambda = 240$ nm was subtracted from the total absorbance signal in the pK_a determination procedure. All samples were homogenized by strongly stirring immediately before each measurement and their pH value was measured with a Mettler Toledo pH-metre. The pK_a values were determined by fitting the Henderson–Hasselbach equation to the absorbance–pH data.

Mass spectrometry of L-cysteine and L-homocysteine. ESI-MS experiments were performed with a Thermo LTQ XL linear ion Trap mass spectrometer with electron transfer dissociation source. The electrospray voltage was kept at 4.00 kV, the capillary temperature was set at 350 °C and the capillary voltage was 19 V. A full MS (m/z 107–400) scan was acquired with a nominal mass resolution. Samples were delivered by means of flow injection. MS data was analysed with Xcalibur version 2.0.7 software. L-homocysteine was freshly prepared right before the MS experiment at a pH of 9.2 and a final concentration of 63 µM, and was confirmed by following the m/z molecular ion peak at 136 m/z . Dimerization to L-homocysteine was induced by preparing a 63 µM L-homocysteine solution at pH 9.2, and measured after 15 h of sample preparation. L-homocysteine was confirmed by following the m/z molecular ion peak at 269 m/z . Both species present in the same spectrum were obtained by bubbling the measuring solution with N₂ prior to MS determination.

Data availability. The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD005981. Additional information and the data supporting this research, including the single-molecule nanomechanics experiments, can be obtained from the corresponding author upon reasonable request.

References

- Smith, B. L. *et al.* Molecular mechanistic origin of the toughness of natural adhesives, fibres and composites. *Nature* **399**, 761–763 (1999).
- Li, H. B. *et al.* Reverse engineering of the giant muscle protein titin. *Nature* **418**, 998–1002 (2002).
- Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. & Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109–1112 (1997).
- Labeit, S. & Kolmerer, B. Titins - giant proteins in charge of muscle ultrastructure and elasticity. *Science* **270**, 293–296 (1995).
- Stirnemann, G., Giganti, D., Fernandez, J. M. & Berne, B. J. Elasticity, structure, and relaxation of extended proteins under force. *Proc. Natl Acad. Sci. USA* **110**, 3847–3852 (2013).

6. Li, H., Carrion-Vazquez, M., Oberhauser, A. F., Marszalek, P. E. & Fernandez, J. M. Point mutations alter the mechanical stability of immunoglobulin modules. *Nat. Struct. Biol.* **7**, 1117–1120 (2000).
7. Hu, X. & Li, H. Force spectroscopy studies on protein-ligand interactions: a single protein mechanics perspective. *FEBS Lett.* **588**, 3613–3620 (2014).
8. Perales-Calvo, J., Lezamiz, A. & Garcia-Manyes, S. The mechanochemistry of a structural zinc finger. *J. Phys. Chem. Lett.* **6**, 3335–3340 (2015).
9. Vogel, V. Mechanotransduction involving multimodular proteins: converting force into biochemical signals. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 459–488 (2006).
10. Bechtluft, P. *et al.* Direct observation of chaperone-induced changes in a protein folding pathway. *Science* **318**, 1458–1461 (2007).
11. Ainarapu, S. R., Li, L., Badilla, C. L. & Fernandez, J. M. Ligand binding modulates the mechanical stability of dihydrofolate reductase. *Biophys. J.* **89**, 3337–3344 (2005).
12. Alegre-Cebollada, J. *et al.* S-glutathionylation of cryptic cysteines enhances titin elasticity by blocking protein folding. *Cell* **156**, 1235–1246 (2014).
13. Beedle, A. E., Lynham, S. & Garcia-Manyes, S. Protein S-sulenylation is a fleeting molecular switch that regulates non-enzymatic oxidative folding. *Nat. Commun.* **7**, 12490 (2016).
14. Mayans, O., Wuerges, J., Canela, S., Gautel, M. & Wilmanns, M. Structural evidence for a possible role of reversible disulphide bridge formation in the elasticity of the muscle protein titin. *Structure* **9**, 331–340 (2001).
15. Carl, P., Kwok, C. H., Manderson, G., Speicher, D. W. & Discher, D. E. Forced unfolding modulated by disulfide bonds in the Ig domains of a cell adhesion molecule. *Proc. Natl Acad. Sci. USA* **98**, 1565–1570 (2001).
16. Wu, C., Belenda, C., Leroux, J. C. & Gauthier, M. A. Interplay of chemical microenvironment and redox environment on thiol-disulfide exchange kinetics. *Chemistry* **17**, 10064–10070 (2011).
17. Anfinsen, C. B. & Haber, E. Studies on the reduction and re-formation of protein disulfide bonds. *J. Biol. Chem.* **236**, 1361–1363 (1961).
18. Nagy, P. Kinetics and mechanisms of thiol-disulfide exchange covering direct substitution and thiol oxidation-mediated pathways. *Antioxid. Redox Signal.* **18**, 1623–1641 (2013).
19. Gupta, V. & Carroll, K. S. Sulfenic acid chemistry, detection and cellular lifetime. *Biochim. Biophys. Acta* **1840**, 847–875 (2014).
20. Holmgren, A. Thioredoxin. *Annu. Rev. Biochem.* **54**, 237–271 (1985).
21. Wilkinson, B. & Gilbert, H. F. Protein disulfide isomerase. *Biochim. Biophys. Acta* **1699**, 35–44 (2004).
22. Sevier, C. S. & Kaiser, C. A. Formation and transfer of disulphide bonds in living cells. *Nat. Rev. Mol. Cell Biol.* **3**, 836–847 (2002).
23. Walker, K. W. & Gilbert, H. F. Scanning and escape during protein-disulfide isomerase-assisted protein folding. *J. Biol. Chem.* **272**, 8845–8848 (1997).
24. Walker, K. W., Lyles, M. M. & Gilbert, H. F. Catalysis of oxidative protein folding by mutants of protein disulfide isomerase with a single active-site cysteine. *Biochemistry* **35**, 1972–1980 (1996).
25. Paxman, J. J. *et al.* The structure of the bacterial oxidoreductase enzyme DsbA in complex with a peptide reveals a basis for substrate specificity in the catalytic cycle of DsbA enzymes. *J. Biol. Chem.* **284**, 17835–17845 (2009).
26. Qin, J., Clore, G. M., Kennedy, W. M., Huth, J. R. & Gronenborn, A. M. Solution structure of human thioredoxin in a mixed disulfide intermediate complex with its target peptide from the transcription factor NF kappa B. *Structure* **3**, 289–297 (1995).
27. Di Jeso, B. *et al.* Mixed-disulfide folding intermediates between thyroglobulin and endoplasmic reticulum resident oxidoreductases ERp57 and protein disulfide isomerase. *Mol. Cell. Biol.* **25**, 9793–9805 (2005).
28. Kadokura, H., Tian, H., Zander, T., Bardwell, J. C. & Beckwith, J. Snapshots of DsbA in action: detection of proteins in the process of oxidative folding. *Science* **303**, 534–537 (2004).
29. Molinari, M. & Helenius, A. Glycoproteins form mixed disulphides with oxidoreductases during folding in living cells. *Nature* **402**, 90–93 (1999).
30. Kosuri, P. *et al.* Protein folding drives disulfide formation. *Cell* **151**, 794–806 (2012).
31. Kahn, T. B., Fernandez, J. M. & Perez-Jimenez, R. Monitoring oxidative folding of a single protein catalyzed by the disulfide oxidoreductase DsbA. *J. Biol. Chem.* **290**, 14518–14527 (2015).
32. Wang, M., Zhao, Q. & Liu, W. The versatile low-molecular-weight thiols: beyond cell protection. *Bioessays* **37**, 1262–1267 (2015).
33. Cumming, R. C. *et al.* Protein disulfide bond formation in the cytoplasm during oxidative stress. *J. Biol. Chem.* **279**, 21749–21758 (2004).
34. Giustarini, D., Dalle-Donne, I., Lorenzini, S., Milzani, A. & Rossi, R. Age-related influence on thiol, disulfide, and protein-mixed disulfide levels in human plasma. *J. Gerontol. A Biol. Sci. Med. Sci.* **61**, 1030–1038 (2006).
35. Rossi, R., Giustarini, D., Milzani, A. & Dalle-Donne, I. Cysteinylation and homocysteinylation of plasma protein thiols during ageing of healthy human beings. *J. Cell. Mol. Med.* **13**, 3131–3140 (2009).
36. Ghezzi, P. Regulation of protein function by glutathionylation. *Free Radic. Res.* **39**, 573–580 (2005).
37. Dalle-Donne, I., Rossi, R., Colombo, G., Giustarini, D. & Milzani, A. Protein S-glutathionylation: a regulatory device from bacteria to humans. *Trends Biochem. Sci.* **34**, 85–96 (2009).
38. Veeranki, S. & Tyagi, S. C. Defective homocysteine metabolism: potential implications for skeletal muscle malfunction. *Int. J. Mol. Sci.* **14**, 15074–15091 (2013).
39. Hankey, G. J. & Eikelboom, J. W. Homocysteine and vascular disease. *Indian Heart J.* **52**, S18–S26 (2000).
40. Eikelboom, J. W., Lonn, E., Genest, Jr J., Hankey, G. & Yusuf, S. Homocyst(e)ine and cardiovascular disease: a critical review of the epidemiologic evidence. *Ann. Intern. Med.* **131**, 363–375 (1999).
41. Mudd, S. H. *et al.* Homocysteine and its disulfide derivatives - a suggested consensus terminology. *Arterioscler. Thromb. Vasc. Biol.* **20**, 1704–1706 (2000).
42. Ragone, R. Homocysteine solubility and vascular disease. *FASEB J.* **16**, 401–404 (2002).
43. Hubmacher, D., Cirulis, J. T., Miao, M., Keeley, F. W. & Reinhardt, D. P. Functional consequences of homocysteinylation of the elastic fiber proteins fibrillin-1 and tropoelastin. *J. Biol. Chem.* **285**, 1188–1198 (2010).
44. Ainarapu, S. R., Wiita, A. P., Huang, H. H. & Fernandez, J. M. A single-molecule assay to directly identify solvent-accessible disulfide bonds and probe their effect on protein folding. *J. Am. Chem. Soc.* **130**, 436–437 (2008).
45. Koti Ainarapu, S. R., Wiita, A. P., Dougan, L., Uggerud, E. & Fernandez, J. M. Single-molecule force spectroscopy measurements of bond elongation during a bimolecular reaction. *J. Am. Chem. Soc.* **130**, 6479–6487 (2008).
46. Dmitrenko, O., Thorpe, C. & Bach, R. D. Mechanism of SN2 disulfide bond cleavage by phosphorus nucleophiles. Implications for biochemical disulfide reducing agents. *J. Org. Chem.* **72**, 8298–8307 (2007).
47. Whitesides, G. M., Houk, J. & Patterson, M. A. K. Activation parameters for thiolate disulfide interchange reactions in aqueous-solution. *J. Org. Chem.* **48**, 112–115 (1983).
48. Peter Vollhardt, N. S. *Organic Chemistry*. 7th edn (W.H. Freeman and Company, 2003).
49. Sengupta, S. *et al.* Albumin thiolate anion is an intermediate in the formation of albumin-S-S-homocysteine. *J. Biol. Chem.* **276**, 30111–30117 (2001).
50. Glushchenko, A. V. & Jacobsen, D. W. Molecular targeting of proteins by L-homocysteine: mechanistic implications for vascular disease. *Antioxid. Redox Signal.* **9**, 1883–1898 (2007).
51. Friedman, M., Cavins, J. F. & Wall, J. S. Relative nucleophilic reactivities of amino groups and mercaptide ions in addition reactions with alpha beta-unsaturated compounds. *J. Am. Chem. Soc.* **87**, 3672 (1965).
52. Hubmacher, D., Sabatier, L., Annis, D. S., Mosher, D. F. & Reinhardt, D. P. Homocysteine modifies structural and functional properties of fibronectin and interferes with the fibronectin-fibrillin-1 interaction. *Biochemistry* **50**, 5322–5332 (2011).
53. Benesch, R. E. & Benesch, R. The acid strength of the -Sh group in cysteine and related compounds. *J. Am. Chem. Soc.* **77**, 5877–5881 (1955).
54. Noda, L. H., Kuby, S. A. & Lardy, H. A. Properties of thioesters - kinetics of hydrolysis in dilute aqueous media. *J. Am. Chem. Soc.* **75**, 913–917 (1953).
55. Polgar, L. Spectrophotometric determination of mercaptide ion, an activated form of Sh-group in thiol enzymes. *FEBS Lett.* **38**, 187–190 (1974).
56. Bach, R. D., Dmitrenko, O. & Thorpe, C. Mechanism of thiolate-disulfide interchange reactions in biochemistry. *J. Org. Chem.* **73**, 12–21 (2008).
57. Garcia-Manyes, S., Kuo, T. L. & Fernandez, J. M. Contrasting the individual reactive pathways in protein unfolding and disulfide bond reduction observed within a single protein. *J. Am. Chem. Soc.* **133**, 3104–3113 (2011).
58. Wiita, A. P., Ainarapu, S. R., Huang, H. H. & Fernandez, J. M. Force-dependent chemical kinetics of disulfide bond reduction observed with single-molecule techniques. *Proc. Natl Acad. Sci. USA* **103**, 7222–7227 (2006).
59. Garcia-Manyes, S., Liang, J., Szoszkiewicz, R., Kuo, T. L. & Fernandez, J. M. Force-activated reactivity switch in a bimolecular chemical reaction. *Nat. Chem.* **1**, 236–242 (2009).
60. Dopieralski, P. *et al.* The Janus-faced role of external forces in mechanochemical disulfide bond cleavage. *Nat. Chem.* **5**, 685–691 (2013).
61. Hofbauer, F. & Frank, I. CPMD simulation of a bimolecular chemical reaction: nucleophilic attack of a disulfide bond under mechanical stress. *Chemistry* **18**, 16332–16338 (2012).
62. Baldus, I. B. & Grater, F. Mechanical force can fine-tune redox potentials of disulfide bonds. *Biophys. J.* **102**, 622–629 (2012).
63. Zafarullah, M., Li, W. Q., Sylvester, J. & Ahmad, M. Molecular mechanisms of N-acetylcysteine actions. *Cell. Mol. Life Sci.* **60**, 6–20 (2003).
64. Samuni, Y., Goldstein, S., Dean, O. M. & Berk, M. The chemistry and biological activities of N-acetylcysteine. *Biochim. Biophys. Acta* **1830**, 4117–4129 (2013).

65. Reddie, K. G. & Carroll, K. S. Expanding the functional diversity of proteins through cysteine oxidation. *Curr. Opin. Chem. Biol.* **12**, 746–754 (2008).
66. Khazaei, S., Yousefi, R. & Alavian-Mehr, M. M. Aggregation and fibrillation of eye lens crystallins by homocysteinylation; implication in the eye pathological disorders. *Protein J.* **31**, 717–727 (2012).
67. Alegre-Cebollada, J., Kosuri, P., Rivas-Pardo, J. A. & Fernandez, J. M. Direct observation of disulfide isomerization in a single protein. *Nat. Chem.* **3**, 882–887 (2011).
68. Hochgrafe, F. *et al.* S-cysteinylation is a general mechanism for thiol protection of *Bacillus subtilis* proteins after oxidative stress. *J. Biol. Chem.* **282**, 25981–25985 (2007).
69. Garcia-Manyes, S., Dougan, L., Badilla, C. L., Brujic, J. & Fernandez, J. M. Direct observation of an ensemble of stable collapsed states in the mechanical folding of ubiquitin. *Proc. Natl Acad. Sci. USA* **106**, 10534–10539 (2009).
70. Tatham, A. S. & Shewry, P. R. Elastomeric proteins: biological roles, structures and mechanisms. *Trends Biochem. Sci.* **25**, 567–571 (2000).
71. 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.* **6**, 7569 (2015).
72. Schlierf, M., Li, H. & Fernandez, J. M. The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proc. Natl Acad. Sci. USA* **101**, 7299–7304 (2004).
73. Popa, I., Kosuri, P., Alegre-Cebollada, J., Garcia-Manyes, S. & Fernandez, J. M. Force dependency of biochemical reactions measured by single-molecule force-clamp spectroscopy. *Nat. Protoc.* **8**, 1261–1276 (2013).
74. Frisch, M. J. *et al.* *Gaussian 09* (Gaussian Inc., 2009).
75. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A Gen. Phys.* **38**, 3098–3100 (1988).
76. Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B Condens. Matter* **37**, 785–789 (1988).
77. Miertus, S., Scrocco, E. & Tomasi, J. Electrostatic interaction of a solute with a continuum - a direct utilization of *ab initio* molecular potentials for the prevision of solvent effects. *Chem. Phys.* **55**, 117–129 (1981).
78. Pecul, M. Theoretical simulation of the ROA spectra of neutral cysteine and serine. *Chem. Phys. Lett.* **427**, 166–176 (2006).
79. Steudel, R. & Steudel, Y. Derivatives of cysteine related to the thiosulfate metabolism of sulfur bacteria by the multi-enzyme complex 'Sox'-studied by B3LYP-PCM and G3X(MP2) calculations. *Phys. Chem. Chem. Phys.* **12**, 630–644 (2010).
80. Tajkhorshid, E., Jalkanen, K. J. & Suhai, S. Structure and vibrational spectra of the zwitterion L-alanine in the presence of explicit water molecules: a density functional analysis. *J. Phys. Chem. B* **102**, 5899–5913 (1998).

Acknowledgements

We thank Ainhoa Lezamiz and Dr Palma Rico for protein purification. A.E.M.B. is funded by an EPSRC DTP fellowship. M.M. is funded by a Fight for Sight PhD studentship. G.S. acknowledges support from CNRS through a PICS allocation (PICS07571). This work was supported by the Marie Curie CIG (293462) grant, BBSRC (J00992X/1) grant, Royal Society Research grant (RG120038), BHF grant (PG/13/50/30426) and EPSRC Fellowship (K00641X/1), all to S.G.-M.

Author contributions

S.G.-M. conceived research. A.E.M.B. and S.G.-M. designed experiments. A.E.M.B. conducted single-molecule mechanical experiments and analysed data. M.M. performed absorbance experiments and small compound mass spectrometry experiments. S.L. performed protein-based mass spectrometry experiments and analysed data.; G.S. performed and analysed DFT calculations. A.E.M.B. and S.G.-M. wrote the paper. All authors contributed in revising and editing the manuscript.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Beedle, A. E. M. *et al.* Tailoring protein nanomechanics with chemical reactivity. *Nat. Commun.* **8**, 15658 doi: 10.1038/ncomms15658 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017

1. Ribas-Arino J, Marx D. Covalent Mechanochemistry: Theoretical Concepts and Computational Tools with Applications to Molecular Nanomechanics. *Chemical reviews* **112**, 5412-5487 (2012).
2. Beyer MK, Clausen-Schaumann H. Mechanochemistry: The mechanical activation of covalent bonds. *Chemical reviews* **105**, 2921-2948 (2005).
3. Hickenboth CR, Moore JS, White SR, Sottos NR, Baudry J, Wilson SR. Biasing reaction pathways with mechanical force. *Nature* **446**, 423-427 (2007).
4. Wang JP, *et al.* Inducing and quantifying forbidden reactivity with single-molecule polymer mechanochemistry. *Nature chemistry* **7**, 323-327 (2015).
5. Schuler B, *et al.* Reversible Bergman cyclization by atomic manipulation. *Nature chemistry* **8**, 220-224 (2016).
6. Ladenthin JN, *et al.* Force-induced tautomerization in a single molecule. *Nature chemistry* **8**, 935-940 (2016).
7. Radiom M, Kong P, Maroni P, Schafer M, Kilbinger AFM, Borkovec M. Mechanically induced cis-to-trans isomerization of carbon-carbon double bonds using atomic force microscopy. *Phys Chem Chem Phys* **18**, 31202-31210 (2016).
8. Brujic J, Hermans RI, Walther KA, Fernandez JM. Single-molecule force spectroscopy reveals signatures of glassy dynamics in the energy landscape of ubiquitin. *Nat Phys* **2**, 282-286 (2006).
9. Rief M, Grubmuller H. Force spectroscopy of single biomolecules. *Chemphyschem* **3**, 255-261 (2002).
10. Marszalek PE, Dufrene YF. Stretching single polysaccharides and proteins using atomic force microscopy. *Chemical Society reviews* **41**, 3523-3534 (2012).
11. Rief M, Oesterhelt F, Heymann B, Gaub HE. Single molecule force spectroscopy on polysaccharides by atomic force microscopy. *Science* **275**, 1295-1297 (1997).

12. Liphardt J, Onoa B, Smith SB, Tinoco I, Bustamante C. Reversible unfolding of single RNA molecules by mechanical force. *Science* **292**, 733-737 (2001).
13. Li H, Carrion-Vazquez M, Oberhauser AF, Marszalek PE, Fernandez JM. Point mutations alter the mechanical stability of immunoglobulin modules. *Nature structural biology* **7**, 1117-1120 (2000).
14. Linke WA. Sense and stretchability: The role of titin and titin-associated proteins in myocardial stress-sensing and mechanical dysfunction. *Cardiovasc Res* **77**, 637-648 (2008).
15. Grutzner A, Garcia-Manyes S, Kotter S, Badilla CL, Fernandez JM, Linke WA. Modulation of Titin-Based Stiffness by Disulfide Bonding in the Cardiac Titin N2-B Unique Sequence. *Biophysical journal* **97**, 825-834 (2009).
16. Marshall BT, Long M, Piper JW, Yago T, McEver RP, Zhu C. Direct observation of catch bonds involving cell-adhesion molecules. *Nature* **423**, 190-193 (2003).
17. Yakovenko O, *et al.* FimH forms catch bonds that are enhanced by mechanical force due to allosteric regulation. *Journal of Biological Chemistry* **283**, 11596-11605 (2008).
18. Thomas WE, Vogel V, Sokurenko E. Biophysics of catch bonds. *Annual review of biophysics* **37**, 399-416 (2008).
19. Echelman DJ, Alegre-Cebollada J, Badilla CL, Chang CY, Ton-That H, Fernandez JM. CnaA domains in bacterial pili are efficient dissipaters of large mechanical shocks. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 2490-2495 (2016).
20. Alegre-Cebollada J, Badilla CL, Fernandez JM. Isopeptide Bonds Block the Mechanical Extension of Pili in Pathogenic *Streptococcus pyogenes*. *Journal of Biological Chemistry* **285**, 11235-11242 (2010).
21. Kamsma D, Sitters G, Thalhammer G, Ritsch-Marte M, Peterman EJG, Wuite GJL. Acoustic Force Spectroscopy. *Biophysical journal* **108**, 355a-355a (2015).
22. Yang D, Ward A, Halvorsen K, Wong WP. Multiplexed single-molecule force spectroscopy using a centrifuge. *Nature communications* **7**, (2016).

23. Rodriguez-Larrea D, Bayley H. Multistep protein unfolding during nanopore translocation. *Nature nanotechnology* **8**, 288-295 (2013).
24. Freedman KJ, Haq SR, Edel JB, Jemth P, Kim MJ. Single molecule unfolding and stretching of protein domains inside a solid-state nanopore by electric field. *Scientific reports* **3**, (2013).
25. Nickels PC, *et al.* Molecular force spectroscopy with a DNA origami-based nanoscopic force clamp. *Science* **354**, 305-307 (2016).
26. Neuman KC, Nagy A. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nature methods* **5**, 491-505 (2008).
27. Hughes ML, Dougan L. The physics of pulling polypeptides: a review of single molecule force spectroscopy using the AFM to study protein unfolding. *Rep Prog Phys* **79**, (2016).
28. Popa I, Kosuri P, Alegre-Cebollada J, Garcia-Manyes S, Fernandez JM. Force dependency of biochemical reactions measured by single-molecule force-clamp spectroscopy. *Nature protocols* **8**, 1261-1276 (2013).
29. Sotomayor M, Schulten K. Single-molecule experiments in vitro and in silico. *Science* **316**, 1144-1148 (2007).
30. Rief M, Gautel M, Oesterhelt F, Fernandez JM, Gaub HE. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109-1112 (1997).
31. Kellermayer MSZ, Smith SB, Granzier HL, Bustamante C. Folding-unfolding transitions in single titin molecules characterized with laser tweezers. *Science* **276**, 1112-1116 (1997).
32. Oberhauser AF, Hansma PK, Carrion-Vazquez M, Fernandez JM. Stepwise unfolding of titin under force-clamp atomic force microscopy. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 468-472 (2001).

33. Dietz H, *et al.* Cysteine engineering of polyproteins for single-molecule force spectroscopy. *Nature protocols* **1**, 80-84 (2006).
34. Bustamante C, Marko JF, Siggia ED, Smith S. Entropic Elasticity of Lambda-Phage DNA. *Science* **265**, 1599-1600 (1994).
35. Stigler J, Ziegler F, Gieseke A, Gebhardt JCM, Rief M. The Complex Folding Network of Single Calmodulin Molecules. *Science* **334**, 512-516 (2011).
36. Hoffmann T, Dougan L. Single molecule force spectroscopy using polyproteins. *Chemical Society reviews* **41**, 4781-4796 (2012).
37. Zheng P, Arantes GM, Field MJ, Li H. Force-induced chemical reactions on the metal centre in a single metalloprotein molecule. *Nat Commun* **6**, 7569 (2015).
38. Perales-Calvo J, Lezamiz A, Garcia-Manyes S. The Mechanochemistry of a Structural Zinc Finger. *The journal of physical chemistry letters* **6**, 3335-3340 (2015).
39. Zheng P, Takayama SIJ, Mauk AG, Li HB. Hydrogen Bond Strength Modulates the Mechanical Strength of Ferric-Thiolate Bonds in Rubredoxin. *Journal of the American Chemical Society* **134**, 4124-4131 (2012).
40. Elms PJ, Chodera JD, Bustamante CJ, Marqusee S. Limitations of Constant-Force-Feedback Experiments. *Biophysical journal* **103**, 1490-1499 (2012).
41. Greenleaf WJ, Woodside MT, Abbondanzieri EA, Block SM. Passive all-optical force clamp for high-resolution laser trapping. *Phys Rev Lett* **95**, (2005).
42. Popa I, *et al.* A HaloTag Anchored Ruler for Week-Long Studies of Protein Dynamics. *Journal of the American Chemical Society* **138**, 10546-10553 (2016).
43. Alegre-Cebollada J, *et al.* S-Glutathionylation of Cryptic Cysteines Enhances Titin Elasticity by Blocking Protein Folding. *Cell* **156**, 1235-1246 (2014).

44. Beedle AE, Lynham S, Garcia-Manyes S. Protein S-sulfenylation is a fleeting molecular switch that regulates non-enzymatic oxidative folding. *Nature communications* **7**, 12490 (2016).
45. Schonfelder J, Perez-Jimenez R, Munoz V. A simple two-state protein unfolds mechanically via multiple heterogeneous pathways at single-molecule resolution. *Nature communications* **7**, 11777 (2016).
46. Yao MX, *et al.* Force-dependent conformational switch of alpha-catenin controls vinculin binding. *Nature communications* **5**, (2014).
47. Schlierf M, Li HB, Fernandez JM. The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 7299-7304 (2004).
48. Wiita AP, Ainavarapu RK, Huang HH, Fernandez JM. Force-dependent chemical kinetics of disulfide bond reduction observed with single-molecule techniques. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 7222-7227 (2006).
49. Berkovich R, Garcia-Manyes S, Klafter J, Urbakh M, Fernandez JM. Hopping around an entropic barrier created by force. *Biochem Bioph Res Co* **403**, 133-137 (2010).
50. Valle-Orero J, Eckels EC, Stirnemann G, Popa I, Berkovich R, Fernandez JM. The elastic free energy of a tandem modular protein under force. *Biochem Bioph Res Co* **460**, 434-438 (2015).
51. Dudko OK, Hummer G, Szabo A. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 15755-15760 (2008).
52. Popa I, Fernandez JM, Garcia-Manyes S. Direct Quantification of the Attempt Frequency Determining the Mechanical Unfolding of Ubiquitin Protein. *Journal of Biological Chemistry* **286**, 31072-31079 (2011).
53. Rivas-Pardo JA, Eckels EC, Popa I, Kosuri P, Linke WA, Fernandez JM. Work Done by Titin Protein Folding Assists Muscle Contraction. *Cell Rep* **14**, 1339-1347 (2016).

54. Valle-Orero J, *et al.* Mechanical Deformation Accelerates Protein Ageing. *Angewandte Chemie*, (2017).
55. Lv SS, Bu TJ, Kayser J, Bausch A, Li HB. Towards constructing extracellular matrix-mimetic hydrogels: An elastic hydrogel constructed from tandem modular proteins containing tenascin FnIII domains. *Acta Biomater* **9**, 6481-6491 (2013).
56. Li HB, Cao Y. Protein Mechanics: From Single Molecules to Functional Biomaterials. *Accounts Chem Res* **43**, 1331-1341 (2010).
57. Cao Y, Li HB. Engineering tandem modular protein based reversible hydrogels. *Chem Commun*, 4144-4146 (2008).
58. Fang J, Li HB. A Facile Way to Tune Mechanical Properties of Artificial Elastomeric Proteins-Based Hydrogels. *Langmuir : the ACS journal of surfaces and colloids* **28**, 8260-8265 (2012).
59. Kong N, Peng Q, Li HB. Rationally Designed Dynamic Protein Hydrogels with Reversibly Tunable Mechanical Properties. *Adv Funct Mater* **24**, 7310-7317 (2014).
60. Li HB, *et al.* Reverse engineering of the giant muscle protein titin. *Nature* **418**, 998-1002 (2002).
61. Lv S, Dudek DM, Cao Y, Balamurali MM, Gosline J, Li HB. Designed biomaterials to mimic the mechanical properties of muscles. *Nature* **465**, 69-73 (2010).
62. Chung J, Kushner AM, Weisman AC, Guan ZB. Direct correlation of single-molecule properties with bulk mechanical performance for the biomimetic design of polymers. *Nat Mater* **13**, 1055-1062 (2014).
63. Vogel V, Sheetz M. Local force and geometry sensing regulate cell functions. *Nat Rev Mol Cell Bio* **7**, 265-275 (2006).
64. Cost AL, Ringer P, Chrostek-Grashoff A, Grashoff C. How to Measure Molecular Forces in Cells: A Guide to Evaluating Genetically-Encoded FRET-Based Tension Sensors. *Cell Mol Bioeng* **8**, 96-105 (2015).

65. Morimatsu M, Mekhdjian AH, Adhikari AS, Dunn AR. Molecular Tension Sensors Report Forces Generated by Single Integrin Molecules in Living Cells. *Nano letters* **13**, 3985-3989 (2013).
66. Meng FJ, Sachs F. Orientation-based FRET sensor for real-time imaging of cellular forces. *J Cell Sci* **125**, 743-750 (2012).
67. Grashoff C, *et al.* Measuring mechanical tension across vinculin reveals regulation of focal adhesion dynamics. *Nature* **466**, 263-U143 (2010).
68. Fisher TE, Oberhauser AF, Carrion-Vazquez M, Marszalek PE, Fernandez JM. The study of protein mechanics with the atomic force microscope. *Trends in biochemical sciences* **24**, 379-384 (1999).
69. Linke WA, *et al.* PEVK domain of titin: an entropic spring with actin-binding properties. *Journal of structural biology* **137**, 194-205 (2002).
70. Li H, Oberhauser AF, Redick SD, Carrion-Vazquez M, Erickson HP, Fernandez JM. Multiple conformations of PEVK proteins detected by single-molecule techniques. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 10682-10686 (2001).
71. Rief M, Gautel M, Oesterhelt F, Fernandez JM, Gaub HE. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109-1112 (1997).
72. Carrion-Vazquez M, Oberhauser AF, Fisher TE, Marszalek PE, Li H, Fernandez JM. Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Progress in biophysics and molecular biology* **74**, 63-91 (2000).
73. Matouschek A, Bustamante C. Finding a protein's Achilles heel. *Nature structural biology* **10**, 674-676 (2003).
74. Lu H, Schulten K. The key event in force-induced unfolding of Titin's immunoglobulin domains. *Biophysical journal* **79**, 51-65 (2000).

75. Klimov DK, Thirumalai D. Native topology determines force-induced unfolding pathways in globular proteins. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 7254-7259 (2000).
76. Fisher TE, Marszalek PE, Fernandez JM. Stretching single molecules into novel conformations using the atomic force microscope. *Nature structural biology* **7**, 719-724 (2000).
77. Zheng P, Cao Y, Bu T, Straus SK, Li H. Single molecule force spectroscopy reveals that electrostatic interactions affect the mechanical stability of proteins. *Biophysical journal* **100**, 1534-1541 (2011).
78. He C, Li H. Staphylokinase Displays Surprisingly Low Mechanical Stability. *Langmuir : the ACS journal of surfaces and colloids* **33**, 1077-1083 (2017).
79. Carrion-Vazquez M, Li H, Lu H, Marszalek PE, Oberhauser AF, Fernandez JM. The mechanical stability of ubiquitin is linkage dependent. *Nature structural biology* **10**, 738-743 (2003).
80. Jagannathan B, Elms PJ, Bustamante C, Marqusee S. Direct observation of a force-induced switch in the anisotropic mechanical unfolding pathway of a protein. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 17820-17825 (2012).
81. Brockwell DJ, *et al.* Pulling geometry defines the mechanical resistance of a beta-sheet protein. *Nature structural biology* **10**, 731-737 (2003).
82. Lee W, *et al.* Mechanical anisotropy of ankyrin repeats. *Biophysical journal* **102**, 1118-1126 (2012).
83. Dietz H, Berkemeier F, Bertz M, Rief M. Anisotropic deformation response of single protein molecules. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 12724-12728 (2006).
84. Dietz H, Rief M. Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 16192-16197 (2004).
85. Kotamarthi HC, Sharma R, Narayan S, Ray S, Ainavarapu SR. Multiple unfolding pathways of leucine binding protein (LBP) probed by single-

molecule force spectroscopy (SMFS). *Journal of the American Chemical Society* **135**, 14768-14774 (2013).

86. Peng Q, Li H. Atomic force microscopy reveals parallel mechanical unfolding pathways of T4 lysozyme: evidence for a kinetic partitioning mechanism. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 1885-1890 (2008).
87. Guinn EJ, Jagannathan B, Marqusee S. Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein. *Nature communications* **6**, 6861 (2015).
88. Aggarwal V, Kulothungan SR, Balamurali MM, Saranya SR, Varadarajan R, Ainavarapu SR. Ligand-modulated parallel mechanical unfolding pathways of maltose-binding proteins. *J Biol Chem* **286**, 28056-28065 (2011).
89. Hu X, Li H. Force spectroscopy studies on protein-ligand interactions: a single protein mechanics perspective. *FEBS letters* **588**, 3613-3620 (2014).
90. Cao Y, Yoo T, Zhuang S, Li H. Protein-protein interaction regulates proteins' mechanical stability. *Journal of molecular biology* **378**, 1132-1141 (2008).
91. Bertz M, Rief M. Ligand binding mechanics of maltose binding protein. *Journal of molecular biology* **393**, 1097-1105 (2009).
92. Wang CC, Tsong TY, Hsu YH, Marszalek PE. Inhibitor binding increases the mechanical stability of staphylococcal nuclease. *Biophysical journal* **100**, 1094-1099 (2011).
93. Rivas-Pardo JA, Alegre-Cebollada J, Ramirez-Sarmiento CA, Fernandez JM, Guixé V. Identifying sequential substrate binding at the single-molecule level by enzyme mechanical stabilization. *ACS nano* **9**, 3996-4005 (2015).
94. Ainavarapu SR, Li L, Badilla CL, Fernandez JM. Ligand binding modulates the mechanical stability of dihydrofolate reductase. *Biophysical journal* **89**, 3337-3344 (2005).
95. Junker JP, Hell K, Schlierf M, Neupert W, Rief M. Influence of substrate binding on the mechanical stability of mouse dihydrofolate reductase. *Biophysical journal* **89**, L46-48 (2005).

96. Li H. Engineering proteins with tailored nanomechanical properties: A single molecule approach. *Org Biomol Chem* **5**, 3399-3406 (2007).
97. Cao Y, Li H. Engineered elastomeric proteins with dual elasticity can be controlled by a molecular regulator. *Nature nanotechnology* **3**, 512-516 (2008).
98. Li HB. 'Mechanical Engineering' of Elastomeric Proteins: Toward Designing New Protein Building Blocks for Biomaterials. *Adv Funct Mater* **18**, 2643-2657 (2008).
99. Grandbois M, Beyer M, Rief M, Clausen-Schaumann H, Gaub HE. How strong is a covalent bond? *Science* **283**, 1727-1730 (1999).
100. Schoeler C, *et al.* Ultrastable cellulosome-adhesion complex tightens under load. *Nature communications* **5**, 5635 (2014).
101. Otten M, *et al.* From genes to protein mechanics on a chip. *Nature methods* **11**, 1127-1130 (2014).
102. Baumann F, Bauer MS, Milles LF, Alexandrovich A, Gaub HE, Pippig DA. Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. *Nature nanotechnology* **11**, 89-94 (2016).
103. Vera AM, Carrion-Vazquez M. Direct Identification of Protein-Protein Interactions by Single-Molecule Force Spectroscopy. *Angewandte Chemie* **55**, 13970-13973 (2016).
104. Cao Y, Yoo T, Li H. Single molecule force spectroscopy reveals engineered metal chelation is a general approach to enhance mechanical stability of proteins. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 11152-11157 (2008).
105. Zheng P, Li HB. Highly Covalent Ferric-Thiolate Bonds Exhibit Surprisingly Low Mechanical Stability. *Journal of the American Chemical Society* **133**, 6791-6798 (2011).
106. Zheng P, Li HB. Direct Measurements of the Mechanical Stability of Zinc-Thiolate Bonds in Rubredoxin by Single-Molecule Atomic Force Microscopy. *Biophysical journal* **101**, 1467-1473 (2011).

107. Perales-Calvo J, Lezamiz A, Garcia-Manyes S. The Mechanochemistry of a Structural Zinc Finger. *Journal of Physical Chemistry Letters* **6**, 3335-3340 (2015).
108. Wei W, *et al.* Structural Insights and the Surprisingly Low Mechanical Stability of the Au-S Bond in the Gold-Specific Protein GolB. *Journal of the American Chemical Society* **137**, 15358-15361 (2015).
109. Zheng P, Arantes GM, Field MJ, Li HB. Force-induced chemical reactions on the metal centre in a single metalloprotein molecule. *Nature communications* **6**, (2015).
110. Zheng P, Takayama SJ, Mauk AG, Li HB. Single Molecule Force Spectroscopy Reveals That Iron Is Released from the Active Site of Rubredoxin by a Stochastic Mechanism. *Journal of the American Chemical Society* **135**, 7992-8000 (2013).
111. Li QC, Barret DG, Messersmith PB, Holten-Andersen N. Controlling Hydrogel Mechanics via Bio-Inspired Polymer-Nanoparticle Bond Dynamics. *ACS nano* **10**, 1317-1324 (2016).
112. Grindy SC, *et al.* Control of hierarchical polymer mechanics with bioinspired metal-coordination dynamics. *Nat Mater* **14**, 1210-1216 (2015).
113. Erickson HP. Reversible Unfolding of Fibronectin Type-Iii and Immunoglobulin Domains Provides the Structural Basis for Stretch and Elasticity of Titin and Fibronectin. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 10114-10118 (1994).
114. Schwaiger I, Sattler C, Hostetter DR, Rief M. The myosin coiled-coil is a truly elastic protein structure. *Nat Mater* **1**, 232-235 (2002).
115. Levinthal C. Are There Pathways for Protein Folding. *J Chim Phys Pcb* **65**, 44-+ (1968).
116. Bujalowski PJ, Oberhauser AF. Tracking unfolding and refolding reactions of single proteins using atomic force microscopy methods. *Methods* **60**, 151-160 (2013).

117. Schonfelder J, De Sancho D, Perez-Jimenez R. The Power of Force: Insights into the Protein Folding Process Using Single-Molecule Force Spectroscopy. *Journal of molecular biology* **428**, 4245-4257 (2016).
118. Garcia-Manyes S, Brujic J, Badilla CL, Fernandez JM. Force-clamp spectroscopy of single-protein monomers reveals the individual unfolding and folding pathways of I27 and ubiquitin. *Biophysical journal* **93**, 2436-2446 (2007).
119. Garcia-Manyes S, Dougan L, Fernandez JM. Osmolyte-induced separation of the mechanical folding phases of ubiquitin. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 10540-10545 (2009).
120. Berkovich R, Garcia-Manyes S, Urbakh M, Klafter J, Fernandez JM. Collapse Dynamics of Single Proteins Extended by Force. *Biophysical journal* **98**, 2692-2701 (2010).
121. Garcia-Manyes S, Dougan L, Badilla CL, Brujic J, Fernandez JM. Direct observation of an ensemble of stable collapsed states in the mechanical folding of ubiquitin. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 10534-10539 (2009).
122. Schlierf M, Rief M. Surprising Simplicity in the Single-Molecule Folding Mechanics of Proteins. *Angew Chem Int Edit* **48**, 820-822 (2009).
123. Valle-Orero J, Tapia-Rojo R, Eckels EC, Rivas-Pardo JA, Popa I, Fernandez JM. Proteins Breaking Bad: A Free Energy Perspective. *Journal of Physical Chemistry Letters* **8**, 3642-3647 (2017).
124. Chen H, *et al.* Dynamics of Equilibrium Folding and Unfolding Transitions of Titin Immunoglobulin Domain under Constant Forces. *Journal of the American Chemical Society* **137**, 3540-3546 (2015).
125. Stigler J, Rief M. Ligand-Induced Changes of the Apparent Transition-State Position in Mechanical Protein Unfolding. *Biophysical journal* **109**, 365-372 (2015).
126. Rognoni L, Most T, Zoldak G, Rief M. Force-dependent isomerization kinetics of a highly conserved proline switch modulates the mechanosensing region of filamin. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 5568-5573 (2014).

127. Neupane K, Manuel AP, Woodside MT. Protein folding trajectories can be described quantitatively by one-dimensional diffusion over measured energy landscapes. *Nat Phys* **12**, 700-+ (2016).
128. Neupane K, Foster DAN, Dee DR, Yu H, Wang F, Woodside MT. Direct observation of transition paths during the folding of proteins and nucleic acids. *Science* **352**, 239-242 (2016).
129. Javadi Y, Fernandez JM, Perez-Jimenez R. Protein Folding Under Mechanical Forces: A Physiological View. *Physiology* **28**, 9-17 (2013).
130. Junker JP, Ziegler F, Rief M. Ligand-Dependent Equilibrium Fluctuations of Single Calmodulin Molecules. *Science* **323**, 633-637 (2009).
131. Dobson CM. Protein folding and misfolding. *Nature* **426**, 884-890 (2003).
132. Soto C. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat Rev Neurosci* **4**, 49-60 (2003).
133. Oberhauser AF, Marszalek PE, Carrion-Vazquez M, Fernandez JM. Single protein misfolding events captured by atomic force microscopy. *Nature structural biology* **6**, 1025-1028 (1999).
134. Mashaghi A, Mashaghi S, Tans SJ. Misfolding of Luciferase at the Single-Molecule Level. *Angew Chem Int Edit* **53**, 10390-10393 (2014).
135. Yu H, Dee DR, Liu X, Brigley AM, Sosova I, Woodside MT. Protein misfolding occurs by slow diffusion across multiple barriers in a rough energy landscape. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 8308-8313 (2015).
136. Hoffmann A, Neupane K, Woodside MT. Single-molecule assays for investigating protein misfolding and aggregation. *Phys Chem Chem Phys* **15**, 7934-7948 (2013).
137. Yu H, *et al.* Direct observation of multiple misfolding pathways in a single prion protein molecule. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 5283-5288 (2012).

138. Scholl ZN, Li Q, Yang WT, Marszalek PE. Single-molecule Force Spectroscopy Reveals the Calcium Dependence of the Alternative Conformations in the Native State of a γ -Crystallin Protein. *Journal of Biological Chemistry* **291**, 18263-18275 (2016).
139. Garcia-Manyes S, *et al.* Single-molecule Force Spectroscopy Predicts a Misfolded, Domain-swapped Conformation in human D-Crystallin Protein. *Journal of Biological Chemistry* **291**, 4226-4235 (2016).
140. Dobson CM, Sali A, Karplus M. Protein folding: A perspective from theory and experiment. *Angew Chem Int Edit* **37**, 868-893 (1998).
141. Luca VC, *et al.* SIGNAL TRANSDUCTION Notch-Jagged complex structure implicates a catch bond in tuning ligand sensitivity. *Science* **355**, 1320+ (2017).
142. Evans E, Leung A, Heinrich V, Zhu C. Mechanical switching and coupling between two dissociation pathways in a P-selectin adhesion bond. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 11281-11286 (2004).
143. Wang X, Ha T. Defining single molecular forces required to activate integrin and notch signaling. *Science* **340**, 991-994 (2013).
144. Buckley CD, *et al.* Cell adhesion. The minimal cadherin-catenin complex binds to actin filaments under force. *Science* **346**, 1254211 (2014).
145. Puchner EM, *et al.* Mechanoenzymatics of titin kinase. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 13385-13390 (2008).
146. Rognoni L, Stigler J, Pelz B, Ylanne J, Rief M. Dynamic force sensing of filamin revealed in single-molecule experiments. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 19679-19684 (2012).
147. Kim J, Hudson NE, Springer TA. Force-induced on-rate switching and modulation by mutations in gain-of-function von Willebrand diseases. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 4648-4653 (2015).

148. Zhang X, Halvorsen K, Zhang CZ, Wong WP, Springer TA. Mechanoenzymatic cleavage of the ultralarge vascular protein von Willebrand factor. *Science* **324**, 1330-1334 (2009).
149. Yao M, *et al.* The mechanical response of talin. *Nature communications* **7**, 11966 (2016).
150. del Rio A, Perez-Jimenez R, Liu RC, Roca-Cusachs P, Fernandez JM, Sheetz MP. Stretching Single Talin Rod Molecules Activates Vinculin Binding. *Science* **323**, 638-641 (2009).
151. Yan J, Yao MX, Goult BT, Sheetz MP. Talin Dependent Mechanosensitivity of Cell Focal Adhesions. *Cell Mol Bioeng* **8**, 151-159 (2015).
152. Yao M, Goult BT, Chen H, Cong P, Sheetz MP, Yan J. Mechanical activation of vinculin binding to talin locks talin in an unfolded conformation. *Scientific reports* **4**, 4610 (2014).
153. Aquilano K, Baldelli S, Ciriolo MR. Glutathione: new roles in redox signaling for an old antioxidant. *Front Pharmacol* **5**, (2014).
154. Bianco P, Martonfalvi Z, Naftz K, Koszegi D, Kellermayer M. Titin Domains Progressively Unfolded by Force Are Homogenously Distributed along the Molecule. *Biophysical journal* **109**, 340-345 (2015).
155. Krammer A, Craig D, Thomas WE, Schulten K, Vogel V. A structural model for force regulated integrin binding to fibronectin's RGD-synergy site. *Matrix Biol* **21**, 139-147 (2002).
156. Kubow KE, *et al.* Mechanical forces regulate the interactions of fibronectin and collagen I in extracellular matrix. *Nature communications* **6**, (2015).
157. Burridge K, Guilluy C. Focal adhesions, stress fibers and mechanical tension. *Exp Cell Res* **343**, 14-20 (2016).
158. Fass D. Disulfide bonding in protein biophysics. *Annual review of biophysics* **41**, 63-79 (2012).

159. Raina S, Missiakas D. Making and breaking disulfide bonds. *Annual review of microbiology* **51**, 179-202 (1997).
160. Linke K, Jakob U. Not every disulfide lasts forever: disulfide bond formation as a redox switch. *Antioxidants & redox signaling* **5**, 425-434 (2003).
161. Zewail AH. Femtochemistry: Atomic-Scale Dynamics of the Chemical Bond Using Ultrafast Lasers (Nobel Lecture) Copyright((c)) The Nobel Foundation 2000. We thank the Nobel Foundation, Stockholm, for permission to print this lecture. *Angewandte Chemie* **39**, 2586-2631 (2000).
162. Wiita AP, Ainaravapu SR, Huang HH, Fernandez JM. Force-dependent chemical kinetics of disulfide bond reduction observed with single-molecule techniques. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 7222-7227 (2006).
163. Ainaravapu SR, Wiita AP, Huang HH, Fernandez JM. A single-molecule assay to directly identify solvent-accessible disulfide bonds and probe their effect on protein folding. *J Am Chem Soc* **130**, 436-437 (2008).
164. Liang J, Fernandez JM. Mechanochemistry: one bond at a time. *ACS nano* **3**, 1628-1645 (2009).
165. Manteca A, Alonso-Caballero A, Fertin M, Poly S, De Sancho D, Perez-Jimenez R. The Influence of Disulfide Bonds on the Mechanical Stability of Proteins is Context Dependent. *The Journal of biological chemistry*, (2017).
166. Garcia-Manyes S, Liang J, Szoszkiewicz R, Kuo TL, Fernandez JM. Force-activated reactivity switch in a bimolecular chemical reaction. *Nature chemistry* **1**, 236-242 (2009).
167. Popa I, Kosuri P, Alegre-Cebollada J, Garcia-Manyes S, Fernandez JM. Force dependency of biochemical reactions measured by single-molecule force-clamp spectroscopy. *Nature protocols* **8**, 1261-1276 (2013).
168. Ainaravapu SR, *et al.* Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. *Biophysical journal* **92**, 225-233 (2007).

169. Garcia-Manyes S, Kuo TL, Fernandez JM. Contrasting the individual reactive pathways in protein unfolding and disulfide bond reduction observed within a single protein. *Journal of the American Chemical Society* **133**, 3104-3113 (2011).
170. Koti Ainavarapu SR, Wiita AP, Dougan L, Uggerud E, Fernandez JM. Single-molecule force spectroscopy measurements of bond elongation during a bimolecular reaction. *Journal of the American Chemical Society* **130**, 6479-6487 (2008).
171. Fernandes PA, Ramos MJ. Theoretical insights into the mechanism for thiol/disulfide exchange. *Chem-Eur J* **10**, 257-266 (2004).
172. Dopieralski P, Ribas-Arino J, Anjukandi P, Krupicka M, Kiss J, Marx D. The Janus-faced role of external forces in mechanochemical disulfide bond cleavage. *Nature chemistry* **5**, 685-691 (2013).
173. Dopieralski P, Ribas-Arino J, Anjukandi P, Krupicka M, Marx D. Force-Induced Reversal of beta-Eliminations: Stressed Disulfide Bonds in Alkaline Solution. *Angew Chem Int Ed Engl* **55**, 1304-1308 (2016).
174. Dopieralski P, Ribas-Arino J, Anjukandi P, Krupicka M, Marx D. Unexpected mechanochemical complexity in the mechanistic scenarios of disulfide bond reduction in alkaline solution. *Nature chemistry* **9**, 164-170 (2017).
175. Li W, Grater F. Atomistic evidence of how force dynamically regulates thiol/disulfide exchange. *Journal of the American Chemical Society* **132**, 16790-16795 (2010).
176. Holmgren A. Thioredoxin structure and mechanism: conformational changes on oxidation of the active-site sulfhydryls to a disulfide. *Structure* **3**, 239-243 (1995).
177. Holmgren A. Thioredoxin. *Annual review of biochemistry* **54**, 237-271 (1985).
178. Wiita AP, *et al.* Probing the chemistry of thioredoxin catalysis with force. *Nature* **450**, 124-127 (2007).

179. Kahn TB, Fernandez JM, Perez-Jimenez R. Monitoring Oxidative Folding of a Single Protein Catalyzed by the Disulfide Oxidoreductase DsbA. *The Journal of biological chemistry* **290**, 14518-14527 (2015).
180. Perez-Jimenez R, *et al.* Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. *Nature structural & molecular biology* **18**, 592-596 (2011).
181. Perez-Jimenez R, *et al.* Diversity of chemical mechanisms in thioredoxin catalysis revealed by single-molecule force spectroscopy. *Nature structural & molecular biology* **16**, 890-896 (2009).
182. Grutzner A, Garcia-Manyes S, Kotter S, Badilla CL, Fernandez JM, Linke WA. Modulation of titin-based stiffness by disulfide bonding in the cardiac titin N2-B unique sequence. *Biophysical journal* **97**, 825-834 (2009).
183. Carl P, Kwok CH, Manderson G, Speicher DW, Discher DE. Forced unfolding modulated by disulfide bonds in the Ig domains of a cell adhesion molecule. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 1565-1570 (2001).
184. Wilkinson B, Gilbert HF. Protein disulfide isomerase. *Biochimica et biophysica acta* **1699**, 35-44 (2004).
185. Bulleid NJ, Ellgaard L. Multiple ways to make disulfides. *Trends in biochemical sciences* **36**, 485-492 (2011).
186. Van Laer K, Hamilton CJ, Messens J. Low-molecular-weight thiols in thiol-disulfide exchange. *Antioxidants & redox signaling* **18**, 1642-1653 (2013).
187. Alegre-Cebollada J, Kosuri P, Rivas-Pardo JA, Fernandez JM. Direct observation of disulfide isomerization in a single protein. *Nature chemistry* **3**, 882-887 (2011).
188. Kosuri P, *et al.* Protein folding drives disulfide formation. *Cell* **151**, 794-806 (2012).
189. Onuchic JN, Wolynes PG. Theory of protein folding. *Current Opinion in Structural Biology* **14**, 70-75 (2004).

190. Brockwell DJ, Radford SE. Intermediates: ubiquitous species on folding energy landscapes? *Curr Opin Struct Biol* **17**, 30-37 (2007).
191. Bryngelson JD, Wolynes PG. Intermediates and Barrier Crossing in a Random Energy-Model (with Applications to Protein Folding). *Journal of Physical Chemistry* **93**, 6902-6915 (1989).
192. Bryngelson JD, Wolynes PG. Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci U S A* **84**, 7524-7528 (1987).
193. Dill KA, Chan HS. From Levinthal to pathways to funnels. *Nature Structural Biology* **4**, 10-19 (1997).
194. Lazaridis T, Karplus M. "New view" of protein folding reconciled with the old through multiple unfolding simulations. *Science* **278**, 1928-1931 (1997).
195. Fersht AR, Daggett V. Protein folding and unfolding at atomic resolution. *Cell* **108**, 573-582 (2002).
196. Wolynes PG, Onuchic JN, Thirumalai D. Navigating the Folding Routes. *Science* **267**, 1619-1620 (1995).
197. Marszalek PE, *et al.* Mechanical unfolding intermediates in titin modules. *Nature* **402**, 100-103 (1999).
198. Bertz M, Rief M. Mechanical unfoldons as building blocks of maltose-binding protein. *Journal of Molecular Biology* **378**, 447-458 (2008).
199. Schwaiger I, Kardinal A, Schleicher M, Noegel AA, Rief M. A mechanical unfolding intermediate in an actin-crosslinking protein. *Nature Structural & Molecular Biology* **11**, 81-85 (2004).
200. Schwaiger I, Schleicher M, Noegel AA, Rief M. The folding pathway of a fast-folding immunoglobulin domain revealed by single-molecule mechanical experiments. *EMBO Rep* **6**, 46-51 (2005).

201. Li LW, Huang HHL, Badilla CL, Fernandez JM. Mechanical unfolding intermediates observed by single-molecule force spectroscopy in a fibronectin type III module. *Journal of Molecular Biology* **345**, 817-826 (2005).
202. Thirumalai D, Klimov DK, Woodson SA. Kinetic partitioning mechanism as a unifying theme in the folding of biomolecules. *Theoretical Chemistry Accounts* **96**, 14-22 (1997).
203. Kiefhaber T. Kinetic Traps in Lysozyme Folding. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 9029-9033 (1995).
204. Wright CF, Lindorff-Larsen K, Randles LG, Clarke J. Parallel protein-unfolding pathways revealed and mapped. *Nature Structural Biology* **10**, 658-662 (2003).
205. Zaidi FN, Nath U, Udgaonkar JB. Multiple intermediates and transition states during protein unfolding. *Nature Structural Biology* **4**, 1016-1024 (1997).
206. He C, Genchev GZ, Lu H, Li H. Mechanically untying a protein slipknot: multiple pathways revealed by force spectroscopy and steered molecular dynamics simulations. *J Am Chem Soc* **134**, 10428-10435 (2012).
207. Gao X, *et al.* Single-molecule experiments reveal the flexibility of a Per-ARNT-Sim domain and the kinetic partitioning in the unfolding pathway under force. *Biophys J* **102**, 2149-2157 (2012).
208. Kotamarthi HC, Sharma R, Narayan S, Ray S, Ainavarapu SRK. Multiple Unfolding Pathways of Leucine Binding Protein (LBP) Probed by Single-Molecule Force Spectroscopy (SMFS). *Journal of the American Chemical Society* **135**, 14768-14774 (2013).
209. Mickler M, Dima RI, Dietz H, Hyeon C, Thirumalai D, Rief M. Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 20268-20273 (2007).
210. Galera-Prat A, Gomez-Sicilia A, Oberhauser AF, Cieplak M, Carrion-Vazquez M. Understanding biology by stretching proteins: recent progress. *Curr Opin Struct Biol* **20**, 63-69 (2010).

211. Garcia-Manyes S, Liang J, Szoszkiewicz R, Kuo TL, Fernandez JM. Force-activated reactivity switch in a bimolecular chemical reaction. *Nature Chemistry* **1**, 236-242 (2009).
212. Solomon EI, *et al.* Copper active sites in biology. *Chem Rev* **114**, 3659-3853 (2014).
213. Liu J, *et al.* Metalloproteins containing cytochrome, iron-sulfur, or copper redox centers. *Chem Rev* **114**, 4366-4469 (2014).
214. Redinbo MR, Yeates TO, Merchant S. Plastocyanin: structural and functional analysis. *J Bioenerg Biomembr* **26**, 49-66 (1994).
215. Gross EL. Plastocyanin: Structure and function. *Photosynth Res* **37**, 103-116 (1993).
216. Bond CS, *et al.* The structure of plastocyanin from the cyanobacterium *Phormidium laminosum*. *Acta Crystallogr D Biol Crystallogr* **55**, 414-421 (1999).
217. Rizzuti B, Daggett V, Guzzi R, Sportelli L. The early steps in the unfolding of azurin. *Biochemistry* **43**, 15604-15609 (2004).
218. Sandberg A, Leckner J, Karlsson BG. Apo-azurin folds via an intermediate that resembles the molten-globule. *Protein Sci* **13**, 2628-2638 (2004).
219. Carrion-Vazquez M, *et al.* Mechanical and chemical unfolding of a single protein: a comparison. *Proc Natl Acad Sci U S A* **96**, 3694-3699 (1999).
220. Stirnemann G, Giganti D, Fernandez JM, Berne BJ. Elasticity, structure, and relaxation of extended proteins under force. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 3847-3852 (2013).
221. Sarangi R, *et al.* Spectroscopic and density functional theory studies of the blue-copper site in M121SeM and C112SeC azurin: Cu-Se versus Cu-S bonding. *J Am Chem Soc* **130**, 3866-3877 (2008).

222. Sandberg A, Leckner J, Shi Y, Schwarz FP, Karlsson BG. Effects of metal ligation and oxygen on the reversibility of the thermal denaturation of *Pseudomonas aeruginosa* azurin. *Biochemistry* **41**, 1060-1069 (2002).
223. Ellman GL. A colorimetric method for determining low concentrations of mercaptans. *Arch Biochem Biophys* **74**, 443-450 (1958).
224. Riener CK, Kada G, Gruber HJ. Quick measurement of protein sulfhydryls with Ellman's reagent and with 4,4'-dithiodipyridine. *Anal Bioanal Chem* **373**, 266-276 (2002).
225. De Rienzo F, Gabdoulline RR, Menziani MC, Wade RC. Blue copper proteins: A comparative analysis of their molecular interaction properties. *Protein Science* **9**, 1439-1454 (2000).
226. Do H, Deeth RJ, Besley NA. Computational Study of the Structure and Electronic Circular Dichroism Spectroscopy of Blue Copper Proteins. *Journal of Physical Chemistry B* **117**, 8105-8112 (2013).
227. Solomon EI, Szilagyi RK, DeBeer George S, Basumallick L. Electronic structures of metal sites in proteins and models: contributions to function in blue copper proteins. *Chem Rev* **104**, 419-458 (2004).
228. Solomon EI, Hadt RG. Recent advances in understanding blue copper proteins. *Coordination Chemistry Reviews* **255**, 774-789 (2011).
229. Solomon EI, Hare JW, Gray HB. Spectroscopic studies and a structural model for blue copper centers in proteins. *Proc Natl Acad Sci U S A* **73**, 1389-1393 (1976).
230. Han J, *et al.* Resonance Raman spectra of plastocyanin and pseudoazurin: evidence for conserved cysteine ligand conformations in cupredoxins (blue copper proteins). *Biochemistry* **30**, 10904-10913 (1991).
231. Pozdnyakova I, Guidry J, Wittung-Stafshede P. Probing copper ligands in denatured *Pseudomonas aeruginosa* azurin: unfolding His117Gly and His46Gly mutants. *Journal of Biological Inorganic Chemistry* **6**, 182-188 (2001).
232. Perez-Jimenez R, Garcia-Manyes S, Aivarapu SR, Fernandez JM. Mechanical unfolding pathways of the enhanced yellow fluorescent protein

- revealed by single molecule force spectroscopy. *J Biol Chem* **281**, 40010-40014 (2006).
233. Rodriguez-Larrea D, Bayley H. Protein co-translocational unfolding depends on the direction of pulling. *Nat Commun* **5**, 4841 (2014).
 234. Aubin-Tam ME, Olivares AO, Sauer RT, Baker TA, Lang MJ. Single-molecule protein unfolding and translocation by an ATP-fueled proteolytic machine. *Cell* **145**, 257-267 (2011).
 235. Maillard RA, *et al.* ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell* **145**, 459-469 (2011).
 236. Berko D, *et al.* The direction of protein entry into the proteasome determines the variety of products and depends on the force needed to unfold its two termini. *Mol Cell* **48**, 601-611 (2012).
 237. Rodriguez-Larrea D, Bayley H. Multistep protein unfolding during nanopore translocation. *Nat Nanotechnol* **8**, 288-295 (2013).
 238. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167-195 (1995).
 239. Wolynes PG, Onuchic JN, Thirumalai D. Navigating the folding routes. *Science* **267**, 1619-1620 (1995).
 240. Karplus M, Weaver DL. Protein-folding dynamics. *Nature* **260**, 404-406 (1976).
 241. Anfinsen CB, Haber E. Studies on the reduction and re-formation of protein disulfide bonds. *J Biol Chem* **236**, 1361-1363 (1961).
 242. Anfinsen CB, Haber E, Sela M, White FH, Jr. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A* **47**, 1309-1314 (1961).
 243. Paulsen CE, Carroll KS. Cysteine-Mediated Redox Signaling: Chemistry, Biology, and Tools for Discovery. *Chemical Reviews* **113**, 4633-4679 (2013).

244. Gupta V, Carroll KS. Sulfenic acid chemistry, detection and cellular lifetime. *Biochimica Et Biophysica Acta-General Subjects* **1840**, 847-875 (2014).
245. Yang J, Carroll KS, Liebler DC. The Expanding Landscape of the Thiol Redox Proteome. *Mol Cell Proteomics* **15**, 1-11 (2016).
246. Allison WS. Formation and Reactions of Sulfenic Acids in Proteins. *Accounts of Chemical Research* **9**, 293-299 (1976).
247. Roos G, Messens J. Protein sulfenic acid formation: From cellular damage to redox regulation. *Free Radical Biology and Medicine* **51**, 314-326 (2011).
248. Poole LB, Nelson KJ. Discovering mechanisms of signaling-mediated cysteine oxidation. *Curr Opin Chem Biol* **12**, 18-24 (2008).
249. Lo Conte M, Carroll KS. The redox biochemistry of protein sulfenylation and sulfinylation. *J Biol Chem* **288**, 26480-26488 (2013).
250. Kettenhofen NJ, Wood MJ. Formation, Reactivity, and Detection of Protein Sulfenic Acids. *Chemical Research in Toxicology* **23**, 1633-1646 (2010).
251. Yang J, Gupta V, Carroll KS, Liebler DC. Site-specific mapping and quantification of protein S-sulphenylation in cells. *Nature Communications* **5**, (2014).
252. Benitez LV, Allison WS. The inactivation of the acyl phosphatase activity catalyzed by the sulfenic acid form of glyceraldehyde 3-phosphate dehydrogenase by dimedone and olefins. *J Biol Chem* **249**, 6234-6243 (1974).
253. Furdui CM, Poole LB. Chemical approaches to detect and analyze protein sulfenic acids. *Mass Spectrom Rev* **33**, 126-146 (2014).
254. Witze ES, Old WM, Resing KA, Ahn NG. Mapping protein post-translational modifications with mass spectrometry. *Nat Methods* **4**, 798-806 (2007).
255. Yang J, Gupta V, Tallman KA, Porter NA, Carroll KS, Liebler DC. Global, in situ, site-specific analysis of protein S-sulfenylation. *Nature Protocols* **10**, 1022-1037 (2015).

256. Cremers CM, Jakob U. Oxidant sensing by reversible disulfide bond formation. *J Biol Chem* **288**, 26489-26496 (2013).
257. Rehder DS, Borges CR. Cysteine sulfenic Acid as an Intermediate in Disulfide Bond Formation and Nonenzymatic Protein Folding. *Biochemistry* **49**, 7748-7755 (2010).
258. Sevier CS, Kaiser CA. Formation and transfer of disulphide bonds in living cells. *Nature Reviews Molecular Cell Biology* **3**, 836-847 (2002).
259. Nagy P. Kinetics and mechanisms of thiol-disulfide exchange covering direct substitution and thiol oxidation-mediated pathways. *Antioxid Redox Signal* **18**, 1623-1641 (2013).
260. Depuydt M, Messens J, Collet JF. How Proteins Form Disulfide Bonds. *Antioxidants & Redox Signaling* **15**, 49-66 (2011).
261. Dickinson BC, Chang CJ. Chemistry and biology of reactive oxygen species in signaling or stress responses. *Nature Chemical Biology* **7**, 504-511 (2011).
262. Reddie KG, Carroll KS. Expanding the functional diversity of proteins through cysteine oxidation. *Current Opinion in Chemical Biology* **12**, 746-754 (2008).
263. D'Autreaux B, Toledano MB. ROS as signalling molecules: mechanisms that generate specificity in ROS homeostasis. *Nat Rev Mol Cell Biol* **8**, 813-824 (2007).
264. Danehy JP, Hunter WE. Alkaline Decomposition of Organic Disulfides .2. Alternative Pathways as Determined by Structure. *Journal of Organic Chemistry* **32**, 2047-& (1967).
265. Pal BC, Uziel M, Doherty DG, Cohn WE. Isolation and Characterization of a Pyrimidine Sulfenic Acid Via Scission of Sulfur-Sulfur Bond in Methyl Analog of Bis(4-Thiouridine) Disulfide. *Journal of the American Chemical Society* **91**, 3634-& (1969).
266. Florence TM. Degradation of Protein Disulfide Bonds in Dilute Alkali. *Biochemical Journal* **189**, 507-520 (1980).

267. Stathakis P, Lay AJ, Fitzgerald M, Schlieker C, Matthias LJ, Hogg PJ. Angiostatin formation involves disulfide bond reduction and proteolysis in kringle 5 of plasmin. *Journal of Biological Chemistry* **274**, 8910-8916 (1999).
268. Hogg PJ. Disulfide bonds as switches for protein function. *Trends in Biochemical Sciences* **28**, 210-214 (2003).
269. Lay AJ, *et al.* Phosphoglycerate kinase acts in tumour angiogenesis as a disulphide reductase. *Nature* **408**, 869-873 (2000).
270. Lay AJ, Jiang XM, Daly E, Sun L, Hogg PJ. Plasmin reduction by phosphoglycerate kinase is a thiol-independent process. *Journal of Biological Chemistry* **277**, 9062-9068 (2002).
271. Hofbauer F, Frank I. CPMD simulation of a bimolecular chemical reaction: nucleophilic attack of a disulfide bond under mechanical stress. *Chemistry* **18**, 16332-16338 (2012).
272. Fernandez JM, Li H. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science* **303**, 1674-1678 (2004).
273. Garcia-Manyes S, Kuo TL, Fernandez JM. Contrasting the Individual Reactive Pathways in Protein Unfolding and Disulfide Bond Reduction Observed within a Single Protein. *Journal of the American Chemical Society* **133**, 3104-3113 (2011).
274. Vollhardt P, N S. *Organic Chemistry: Structure and Function*. Freeman and Company (2003).
275. Poole LB, Karplus PA, Claiborne A. Protein sulfenic acids in redox signaling. *Annual Review of Pharmacology and Toxicology* **44**, 325-347 (2004).
276. Stirnemann G, Kang SG, Zhou R, Berne BJ. How force unfolding differs from chemical denaturation. *Proc Natl Acad Sci U S A* **111**, 3413-3418 (2014).
277. Walther KA, Grater F, Dougan L, Badilla CL, Berne BJ, Fernandez JM. Signatures of hydrophobic collapse in extended proteins captured with force spectroscopy. *Proc Natl Acad Sci U S A* **104**, 7916-7921 (2007).

278. Stirnemann G, Giganti D, Fernandez JM, Berne BJ. Elasticity, structure, and relaxation of extended proteins under force. *Proc Natl Acad Sci U S A* **110**, 3847-3852 (2013).
279. Crowder JA, Harris M. Mechanism of the sulfur lability in the alkali degradation of wool protein. *Journal of Research of the National Bureau of Standards* **16**, 475-480 (1936).
280. Toennies G. Oxidation of cysteine in non-aqueous media - The "sulfenic acid" as the primary oxidation product. *Journal of Biological Chemistry* **122**, 27-47 (1937).
281. Schieber M, Chandel NS. ROS Function in Redox Signaling and Oxidative Stress. *Current Biology* **24**, R453-R462 (2014).
282. Leonard SE, Reddie KG, Carroll KS. Mining the thiol proteome for sulfenic acid modifications reveals new targets for oxidation in cells. *ACS Chem Biol* **4**, 783-799 (2009).
283. Salsbury FR, Knutson ST, Poole LB, Fetrow JS. Functional site profiling and electrostatic analysis of cysteines modifiable to cysteine sulfenic acid. *Protein Science* **17**, 299-312 (2008).
284. Lee SR, Yang KS, Kwon J, Lee C, Jeong W, Rhee SG. Reversible inactivation of the tumor suppressor PTEN by H₂O₂. *J Biol Chem* **277**, 20336-20342 (2002).
285. Giannoni E, Buricchi F, Raugei G, Ramponi G, Chiarugi P. Intracellular reactive oxygen species activate Src tyrosine kinase during cell adhesion and anchorage-dependent cell growth. *Mol Cell Biol* **25**, 6391-6403 (2005).
286. Cotto-Rios XM, Bekes M, Chapman J, Ueberheide B, Huang TT. Deubiquitinases as a signaling target of oxidative stress. *Cell Rep* **2**, 1475-1484 (2012).
287. Lee C, *et al.* Redox regulation of OxyR requires specific disulfide bond formation involving a rapid kinetic reaction path. *Nat Struct Mol Biol* **11**, 1179-1185 (2004).

288. Paulsen CE, Carroll KS. Chemical dissection of an essential redox switch in yeast. *Chem Biol* **16**, 217-225 (2009).
289. Reynaert NL, *et al.* Dynamic redox control of NF-kappaB through glutaredoxin-regulated S-glutathionylation of inhibitory kappaB kinase beta. *Proc Natl Acad Sci U S A* **103**, 13086-13091 (2006).
290. Bogeski I, *et al.* Differential redox regulation of ORAI ion channels: a mechanism to tune cellular calcium signaling. *Sci Signal* **3**, ra24 (2010).
291. Svoboda LK, *et al.* Redox-sensitive sulfenic acid modification regulates surface expression of the cardiovascular voltage-gated potassium channel Kv1.5. *Circ Res* **111**, 842-853 (2012).
292. Seo YH, Carroll KS. Profiling protein thiol oxidation in tumor cells using sulfenic acid-specific antibodies. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 16163-16168 (2009).
293. Kahn TB, Fernandez JM, Perez-Jimenez R. Monitoring Oxidative Folding of a Single Protein Catalyzed by the Disulfide Oxidoreductase DsbA. *Journal of Biological Chemistry* **290**, 14518-14527 (2015).
294. Parker AJ, Kharasch N. The scission of the sulfur-sulfur bond. *Chemical Reviews*, 583-628 (1959).
295. Gautel M. The sarcomeric cytoskeleton: who picks up the strain? *Curr Opin Cell Biol* **23**, 39-46 (2011).
296. Chung CS, *et al.* Shortening of the Elastic Tandem Immunoglobulin Segment of Titin Leads to Diastolic Dysfunction. *Circulation* **128**, 19-28 (2013).
297. Li H, *et al.* Reverse engineering of the giant muscle protein titin. *Nature* **418**, 998-1002 (2002).
298. Smith BL, *et al.* Molecular mechanistic origin of the toughness of natural adhesives, fibres and composites. *Nature* **399**, 761-763 (1999).
299. Labeit S, Kolmerer B. Titins - Giant Proteins in Charge of Muscle Ultrastructure and Elasticity. *Science* **270**, 293-296 (1995).

300. Cao Y, Li HB. Engineered elastomeric proteins with dual elasticity can be controlled by a molecular regulator. *Nature Nanotechnology* **3**, 512-516 (2008).
301. Vogel V. Mechanotransduction involving multimodular proteins: Converting force into biochemical signals. *Annual Review of Biophysics and Biomolecular Structure* **35**, 459-488 (2006).
302. Bechtluft P, *et al.* Direct observation of chaperone-induced changes in a protein folding pathway. *Science* **318**, 1458-1461 (2007).
303. Alegre-Cebollada J, *et al.* S-glutathionylation of cryptic cysteines enhances titin elasticity by blocking protein folding. *Cell* **156**, 1235-1246 (2014).
304. Mayans O, Wuerges J, Canela S, Gautel M, Wilmanns M. Structural evidence for a possible role of reversible disulphide bridge formation in the elasticity of the muscle protein titin. *Structure* **9**, 331-340 (2001).
305. Keten S, Chou CC, van Duin AC, Buehler MJ. Tunable nanomechanics of protein disulfide bonds in redox microenvironments. *J Mech Behav Biomed Mater* **5**, 32-40 (2012).
306. Wu C, Belenda C, Leroux JC, Gauthier MA. Interplay of chemical microenvironment and redox environment on thiol-disulfide exchange kinetics. *Chemistry* **17**, 10064-10070 (2011).
307. Nagy P. Kinetics and Mechanisms of Thiol-Disulfide Exchange Covering Direct Substitution and Thiol Oxidation-Mediated Pathways. *Antioxidants & Redox Signaling* **18**, 1623-1641 (2013).
308. Sevier CS, Kaiser CA. Formation and transfer of disulphide bonds in living cells. *Nat Rev Mol Cell Biol* **3**, 836-847 (2002).
309. Walker KW, Gilbert HF. Scanning and escape during protein-disulfide isomerase-assisted protein folding. *J Biol Chem* **272**, 8845-8848 (1997).
310. Walker KW, Lyles MM, Gilbert HF. Catalysis of oxidative protein folding by mutants of protein disulfide isomerase with a single active-site cysteine. *Biochemistry* **35**, 1972-1980 (1996).

311. Paxman JJ, *et al.* The structure of the bacterial oxidoreductase enzyme DsbA in complex with a peptide reveals a basis for substrate specificity in the catalytic cycle of DsbA enzymes. *J Biol Chem* **284**, 17835-17845 (2009).
312. Qin J, Clore GM, Kennedy WM, Huth JR, Gronenborn AM. Solution structure of human thioredoxin in a mixed disulfide intermediate complex with its target peptide from the transcription factor NF kappa B. *Structure* **3**, 289-297 (1995).
313. Di Jeso B, *et al.* Mixed-disulfide folding intermediates between thyroglobulin and endoplasmic reticulum resident oxidoreductases ERp57 and protein disulfide isomerase. *Mol Cell Biol* **25**, 9793-9805 (2005).
314. Kadokura H, Tian H, Zander T, Bardwell JC, Beckwith J. Snapshots of DsbA in action: detection of proteins in the process of oxidative folding. *Science* **303**, 534-537 (2004).
315. Molinari M, Helenius A. Glycoproteins form mixed disulphides with oxidoreductases during folding in living cells. *Nature* **402**, 90-93 (1999).
316. Wang M, Zhao Q, Liu W. The versatile low-molecular-weight thiols: Beyond cell protection. *Bioessays* **37**, 1262-1267 (2015).
317. Cumming RC, Andon NL, Haynes PA, Park M, Fischer WH, Schubert D. Protein disulfide bond formation in the cytoplasm during oxidative stress. *Journal of Biological Chemistry* **279**, 21749-21758 (2004).
318. Giustarini D, Dalle-Donne I, Lorenzini S, Milzani A, Rossi R. Age-related influence on thiol, disulfide, and protein-mixed disulfide levels in human plasma. *J Gerontol A Biol Sci Med Sci* **61**, 1030-1038 (2006).
319. Rossi R, Giustarini D, Milzani A, Dalle-Donne I. Cysteinylation and homocysteinylation of plasma protein thiols during ageing of healthy human beings. *J Cell Mol Med* **13**, 3131-3140 (2009).
320. Ghezzi P. Regulation of protein function by glutathionylation. *Free Radic Res* **39**, 573-580 (2005).

321. Dalle-Donne I, Rossi R, Colombo G, Giustarini D, Milzani A. Protein S-glutathionylation: a regulatory device from bacteria to humans. *Trends Biochem Sci* **34**, 85-96 (2009).
322. Veeranki S, Tyagi SC. Defective Homocysteine Metabolism: Potential Implications for Skeletal Muscle Malfunction. *International Journal of Molecular Sciences* **14**, 15074-15091 (2013).
323. Hankey GJ, Eikelboom JW. Homocysteine and vascular disease. *Indian Heart J* **52**, S18-26 (2000).
324. Eikelboom JW, Lonn E, Genest J, Jr., Hankey G, Yusuf S. Homocyst(e)ine and cardiovascular disease: a critical review of the epidemiologic evidence. *Ann Intern Med* **131**, 363-375 (1999).
325. Mudd SH, *et al.* Homocysteine and its disulfide derivatives - A suggested consensus terminology. *Arteriosclerosis Thrombosis and Vascular Biology* **20**, 1704-1706 (2000).
326. Ragone R. Homocystine solubility and vascular disease. *FASEB J* **16**, 401-404 (2002).
327. Hubmacher D, Cirulis JT, Miao M, Keeley FW, Reinhardt DP. Functional Consequences of Homocysteinylation of the Elastic Fiber Proteins Fibrillin-1 and Tropoelastin. *Journal of Biological Chemistry* **285**, 1188-1198 (2010).
328. Dmitrenko O, Thorpe C, Bach RD. Mechanism of SN2 disulfide bond cleavage by phosphorus nucleophiles. Implications for biochemical disulfide reducing agents. *J Org Chem* **72**, 8298-8307 (2007).
329. Whitesides GM, Houk J, Patterson MAK. Activation Parameters for Thiolate Disulfide Interchange Reactions in Aqueous-Solution. *Journal of Organic Chemistry* **48**, 112-115 (1983).
330. Peter Vollhardt NS. *Organic Chemistry*, 7th Edition edn. W. H. Freeman and Company (2003).
331. Sengupta S, *et al.* Albumin thiolate anion is an intermediate in the formation of albumin-S-S-homocysteine. *Journal of Biological Chemistry* **276**, 30111-30117 (2001).

332. Glushchenko AV, Jacobsen DW. Molecular targeting of proteins by L-homocysteine: Mechanistic implications for vascular disease. *Antioxidants & Redox Signaling* **9**, 1883-1898 (2007).
333. Friedman M, Cavins JF, Wall JS. Relative Nucleophilic Reactivities of Amino Groups and Mercaptide Ions in Addition Reactions with Alpha Beta-Unsaturated Compounds. *Journal of the American Chemical Society* **87**, 3672-& (1965).
334. Hubmacher D, Sabatier L, Annis DS, Mosher DF, Reinhardt DP. Homocysteine modifies structural and functional properties of fibronectin and interferes with the fibronectin-fibrillin-1 interaction. *Biochemistry* **50**, 5322-5332 (2011).
335. Benesch RE, Benesch R. The Acid Strength of the -Sh Group in Cysteine and Related Compounds. *Journal of the American Chemical Society* **77**, 5877-5881 (1955).
336. Noda LH, Kuby SA, Lardy HA. Properties of Thiolesters - Kinetics of Hydrolysis in Dilute Aqueous Media. *Journal of the American Chemical Society* **75**, 913-917 (1953).
337. Polgar L. Spectrophotometric Determination of Mercaptide Ion, an Activated Form of Sh-Group in Thiol Enzymes. *Febs Letters* **38**, 187-190 (1974).
338. Bach RD, Dmitrenko O, Thorpe C. Mechanism of thiolate-disulfide interchange reactions in biochemistry. *J Org Chem* **73**, 12-21 (2008).
339. Baldus IB, Grater F. Mechanical force can fine-tune redox potentials of disulfide bonds. *Biophys J* **102**, 622-629 (2012).
340. Zafarullah M, Li WQ, Sylvester J, Ahmad M. Molecular mechanisms of N-acetylcysteine actions. *Cell Mol Life Sci* **60**, 6-20 (2003).
341. Samuni Y, Goldstein S, Dean OM, Berk M. The chemistry and biological activities of N-acetylcysteine. *Biochim Biophys Acta* **1830**, 4117-4129 (2013).
342. Reddie KG, Carroll KS. Expanding the functional diversity of proteins through cysteine oxidation. *Curr Opin Chem Biol* **12**, 746-754 (2008).

343. Khazaei S, Yousefi R, Alavian-Mehr MM. Aggregation and fibrillation of eye lens crystallins by homocysteinylation; implication in the eye pathological disorders. *Protein J* **31**, 717-727 (2012).
344. Hochgrafe F, Mostertz J, Pother DC, Becher D, Helmann JD, Hecker M. S-cysteinylation is a general mechanism for thiol protection of *Bacillus subtilis* proteins after oxidative stress. *J Biol Chem* **282**, 25981-25985 (2007).
345. Garcia-Manyes S, Dougan L, Badilla CL, Brujic J, Fernandez JM. Direct observation of an ensemble of stable collapsed states in the mechanical folding of ubiquitin. *Proc Natl Acad Sci U S A* **106**, 10534-10539 (2009).
346. Tatham AS, Shewry PR. Elastomeric proteins: biological roles, structures and mechanisms. *Trends in Biochemical Sciences* **25**, 567-571 (2000).
347. Echelman DJ, Lee AQ, Fernandez JM. Mechanical forces regulate the reactivity of a thioester bond in a bacterial adhesin. *Journal of Biological Chemistry* **292**, 8988-8997 (2017).
348. Chen Z, *et al.* Mechanochemical unzipping of insulating poly(ladderene) to semiconducting polyacetylene. *Science* **357**, 475-479 (2017).
349. Lei H, Guo YB, Hu XD, Hu CG, Hu XT, Li HB. Reversible Unfolding and Folding of the Metalloprotein Ferredoxin Revealed by Single-Molecule Atomic Force Microscopy. *Journal of the American Chemical Society* **139**, 1538-1544 (2017).
350. Zheng P, Wang YY, Li HB. Reversible Unfolding-Refolding of Rubredoxin: A Single-Molecule Force Spectroscopy Study. *Angew Chem Int Edit* **53**, 14060-14063 (2014).
351. Fernandez JM, Li HB. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science* **303**, 1674-1678 (2004).
352. Beedle AEM, Mora M, Lynham S, Stirnemann G, Garcia-Manyes S. Tailoring protein nanomechanics with chemical reactivity. *Nature communications* **8**, 15658 (2017).

353. Beedle AEM, Lynham S, Garcia-Manyes S. Protein S-sulfenylation is a fleeting molecular switch that regulates non-enzymatic oxidative folding. *Nature communications* **7**, (2016).
354. Schlierf M, Li H, Fernandez JM. The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 7299-7304 (2004).
355. Huang J, MacKerell AD, Jr. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* **34**, 2135-2145 (2013).
356. van den Bosch M, *et al.* Calculation of the redox potential of the protein azurin and some mutants. *Chembiochem* **6**, 738-746 (2005).
357. Frisch MJ, *et al.* Gaussian 09. (ed[^](eds). Gaussian, Inc. (2009).
358. Becke AD. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys Rev A Gen Phys* **38**, 3098-3100 (1988).
359. Lee C, Yang W, Parr RG. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B Condens Matter* **37**, 785-789 (1988).
360. Miertus S, Scrocco E, Tomasi J. Electrostatic Interaction of a Solute with a Continuum - a Direct Utilization of Abinitio Molecular Potentials for the Prevision of Solvent Effects. *Chemical Physics* **55**, 117-129 (1981).
361. Pecul M. Theoretical simulation of the ROA spectra of neutral cysteine and serine. *Chemical Physics Letters* **427**, 166-176 (2006).
362. Steudel R, Steudel Y. Derivatives of cysteine related to the thiosulfate metabolism of sulfur bacteria by the multi-enzyme complex "Sox"-studied by B3LYP-PCM and G3X(MP2) calculations. *Physical Chemistry Chemical Physics* **12**, 630-644 (2010).
363. Tajkhorshid E, Jalkanen KJ, Suhai S. Structure and vibrational spectra of the zwitterion L-alanine in the presence of explicit water molecules: A density functional analysis. *Journal of Physical Chemistry B* **102**, 5899-5913 (1998).

