

BIOLOGY &
BIOLOGICAL CHEMISTRY

ACS IN
FOCUS

Mechanical Unfolding Response of Proteins

Ionel Popa &
Ronen Berkovich



ACS Publications

ACS In Focus

Mechanical Unfolding Response of Proteins

Ionel Popa

Ronen Berkovich

PREFACE

Proteins are busy workers, keeping our cells alive and responsible for the performance of most of their functions. Biochemical assays have been known to explore proteins since men utilized fire and boiled the first egg. However, there is a secret life that proteins have in vivo, of which we know very little, which involves localized mechanical interactions, that cause disruptions in the folded structure of a protein under a transient pulling vector from a neighboring molecule, and large assemblies of many proteins acting in a tag-of-war. These mechanical interactions have eluded scientists for a long time, since they could not be studied with typical biochemical assays.

In this primer, we invite you to join us on a journey where we will look at proteins from a new perspective, that of mechanical interactions. Mechanical unfolding of proteins is accompanied by the extension of the peptide chain by an application of force. This extension enables unique signaling cascades and intricate modalities to efficiently store, release or dissipate energy. In the first Section we will look at structural elements forming the 3-dimensional (3D) structure of proteins with a mechanical emphasis, followed by an improved description of the protein folding process, which tracks the journey that a peptide takes from synthesis by the ribosome to the final native structure. In the following two Sections we track the mechanical response of proteins from a thermodynamic and kinetic perspective, involving the use 2D and 3D energy landscape concepts. In the fourth Section we discuss current methods and experimental techniques used to study the mechanical unfolding response of proteins. In the final Section we present specific signaling and regulation examples, where mechanical unfolding of proteins and their refolding under a force vector plays an important role in the human body.

When designing and writing this primer, we were led by the editor's enlightening and insightful comment – write as if we could address a younger version of ourselves. Accordingly, by writing to our younger selves, we intended that this primer could serve as a starting point for early-stage young scientists (graduate students), or other any other person new to the field of mechanical forces effects on protein performance. Beyond the introduction to the physics behind protein behavior for those coming from more biological background, this primer can be useful for application of concepts from thermodynamics and statistical mechanics of polymers to understand protein compliance to external forces.

To conclude, in this primer we were motivated to assemble a (relatively) easy to read and yet comprehensive overview that gathers fundamental concepts in protein mechanics, while attempting to balance between fine details on one hand, and breadth on the other. We encourage the reader to be extremely critic when performing measurements, and particularly when analyzing it by using the proper models to interpret the data, since many artifacts (as discussed in this primer) can manifest themselves. Naturally we encourage the young readers of this primer to

expand their reading based on the references of each Section, and even beyond that. This primer is also appropriate to readers who did not take any advanced biophysics or biology courses, but are curious to learn how the engines of our cells work inside our bodies.

AUTHORS



Ionel Popa graduated with a Bachelor of Chemical Engineering in from Gh. Asachi Technical University, in Iasi, Romania, and obtained his doctoral degree in Chemistry and Biochemistry from University of Geneva, in Switzerland, while working with Prof. Michal Borkovec. He then joined the Department of Biological Sciences at Columbia University, first as a postdoc and later as an associate research scientist. At Columbia he worked with Prof. Julio M. Fernandez, where they developed a novel approach based on magnetic tweezers and covalent attachment that can study the mechanical unfolding of a single protein molecule for many hours or days. Since 2015, he joined the Physics Department at the University of Wisconsin-Milwaukee, where he currently holds the title of Associate Professor. His laboratory is studying how the mechano-biology of proteins relate to how our muscles work, how cancer develops and spreads, and how bacteria adhere and interact with antibodies.



Ronen Berkovich is an Associate Professor in the Department of Chemical Engineering at Ben-Gurion University of the Negev (BGU). He received her B.Sc. in Chemical Engineering from Technion in 2002, and after pursuing several years in the industry as an engineer, he obtained his Ph.D. from the Department of Chemical Physics at Tel-Aviv University in 2010 in Chemical Physics with Prof. Joseph (Yossi) Klafter and Prof. Michael Urbakh. He completed postdoctoral training at Columbia University at the Department of Biological Sciences with Prof. Julio M. Fernandez, prior to joining BGU as a faculty member in October 2013. Beyond single molecule biophysics, his research group is studying non-linear phenomena related to soft matter mechanics, nanotribology, and particle resuspension.

DEDICATIONS/ACKNOWLEDGMENTS

We dedicate this primer to Prof. Julio M. Fernandez from Columbia University, who was the postdoctoral mentor to both authors. Prof. Fernandez is an honest and passionate scientist, a true intellectual, and one of the founding fathers of the field of single protein force spectroscopy. The scientific journey outlined in this primer began in his lab, where the authors met and continued collaborating after they started their independent careers. In particular we would like to recall the

stimulating and inspiring scientific ‘good discussions’ over coffee at the Hungarian Pastry Shop in Amsterdam Ave, which were later moved to JoeCoffee in the Northwest Corner Building in Columbia University. Thank you, Julio.

Additionally, R.B. would like to thank and acknowledge his Ph.D. advisors, Prof. Yossi Klafter and Prof. Michael Urbakh, for their involvement and contribution to the initial part of the theoretical aspects of this work, which influenced its continuation.

I.P. also would like to thank his Ph.D. advisor, Prof. Michal Borkovec from the University of Geneva, for inspiring the art-of-science. I.P. also acknowledges National Science Foundation (MCB-1846143), who awarded him the Career award that enabled this quest for understanding how protein unfolding plays a functional role in vivo.

Both authors acknowledge the financial support by US-Israel Binational Science Foundation (BSF, grant number 2020042).

Finally, we thank the American Chemical Society for giving us this opportunity and for the efforts and resources they have put into this project. Their highly professional team made this endeavor interesting (and enjoyable!) didactical voyage. Whole heartly thanks Sara Tenney, Mary Catherine Hager, Tricia Louvar, and the skilled art and graphical directors Benedicte Rossi and Carrie Elleman.

Ionel Popa and Ronen Berkovich,

March 2023.

Chapter 1 – Protein folding and structure

1.1 INTRODUCTION

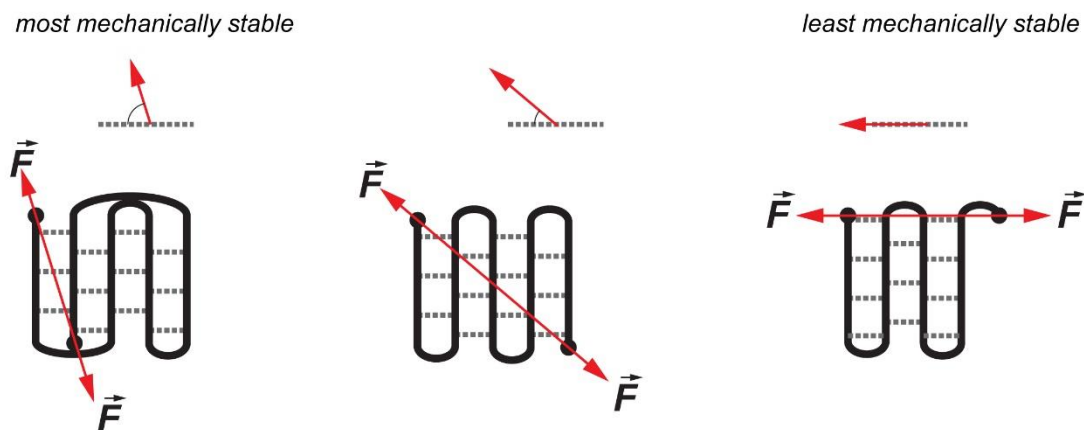
‘Force equals mass times acceleration’ is one of the first physics laws that students learn when starting to study how objects interact with each other. A force is characterized by a magnitude and a direction, hence its vectorial nature, and acts on objects as a pull or push, to change their velocity or acceleration. But how mechanical forces act on biological molecules such as proteins is yet a poorly understood but an equally fascinating topic that will be the focus of this digital primer. Proteins are complex biomolecules that accomplish most of the cellular functions in nature, from structural support to enzymatic activity. They are made of a sequence of twenty amino acids, known as the primary structure, that drive the formation of well-defined 3-dimensional (3D) structures. It is these 3D structures that give proteins their function. These 3D structures, also known as the native structures, are composed of two main elements, alpha helices and beta strands, connected with flexible loops and tight turns, and are held together by hydrogen bonds and, sometimes, disulfide bridges. The alpha (*a*-) helices and beta (*b*-) strands, which represent the secondary structure of a protein, form during the *protein folding* process. Protein folding, which describes the journey from a polypeptide chain of a certain sequence to a specific 3D *native* structure, is still one of the unsolved quests in science. This process involves several steps, starting with the entropic collapse of the polypeptide chain, followed by the hydrophobic collapse, formation of the molten globule and finally the acquisition of the native functional structure. This chapter will further discuss the main structural components forming the protein structure and the steps composing the folding process. The structure will be presented based on its organization level (primary, secondary, tertiary and quaternary), while the protein folding process will be described from a time and energy perspective (as polypeptide synthesis, entropic collapse, hydrophobic collapse, leading to the native state).

While protein folding is an active area of research, its inverse process – protein unfolding – is equally captivating. Unfolding is a yet to be recognized signaling mechanism, where the folded and unfolded states of protein domains represent the ones and zeros of a molecular computational unit. Apart from the fast kinetics of the unfolding/refolding reaction compared to protein turn-over, ‘flipping’ the unfolding switch allows for a sudden change in extension and for exposure of a hidden binding site ¹. There are three main ways to unfold a protein, all requiring a specific type of energy. Chemical unfolding involves the addition of a chemical denaturant to disrupt the molecular interactions holding the tertiary structure, and this chemical needs typically to be added in higher molar concentration. **Thermal unfolding** relies on heat to drive the breaking of the hydrogen bonds forming the tertiary structure of a protein, taking place at temperatures well above the physiological temperature of warm-blooded organisms, of 37 °C. Mechanical unfolding relies on the use of pulling forces, that shear or unzip the hydrogen bonds stabilizing a protein and this process typically takes place when the protein experiences forces of several piconewtons (pN) (**Figure 1.1**). The type of energy being used to trigger unfolding – chemical, thermal or mechanical – matters, and proteins can be, for example, highly stable to thermal denaturation and weakly stable to mechanical forces. This difference between thermal and mechanical stability also explains why there are organisms living in hot springs, such as those in Yellowstone National

Park. Their proteins are thermally very stable, withstanding temperatures of over 80 C. However, if the proteins from these organisms would be equally mechanically stable, it would be impossible for them to turn over their proteins. As discussed in Chapter 5, the proteasome uses mechanical unfolding to denature and then degrades proteins.

Unlike mechanical unfolding, thermal and chemical unfolding are global processes, where most of the hydrogen bonds holding the protein structure together break almost simultaneously. The origin of the difference between thermal-chemical and mechanical unfolding goes back to the previously mentioned characteristic of force being a vector. Due to its vectorial nature, mechanical unfolding is a localized event, triggered by the breaking of a few hydrogen bonds. These hydrogen bonds inside the protein structure are known as the *mechanical clamp* of a protein and their breaking, followed by the immediate diffusion of water molecules taking over the newly exposed sites, represents the transition state of the mechanical unfolding process. Hence, mechanical unfolding is a highly localized event. Furthermore, how the hydrogen bonds forming the mechanical clamp are oriented with respect to the force vector matters (**Figure 1.1**). A protein with hydrogen bonds perpendicular to the force vector will display a higher mechanical stability than a protein with the same number of bonds parallel to the force. This behavior has been proven experimentally by attaching the tethers used to apply force in various points of a folded protein². The same protein would then show vastly different mechanical stability, dependent on the orientation of the force vector.

Figure 1.1. Effect of the pulling geometry on the mechanical stability of a protein.



The geometry of the tethering ends in respect to the orientation of the hydrogen bonds maintaining the tertiary structure of a protein plays a critical role in determining the mechanical stability of a protein. The closer the angle between hydrogen bonds and direction of the applied force is to 90 degrees, the more stable the protein is, while a protein that has most of its angles between the force and hydrogen bonds close to 0 degrees will be mechanically weak. Schematics represent all-beta proteins in several pulling geometries. Dotted grey lines symbolize hydrogen bonds, the black dots - the ends of the protein, the red line – the shortest distance determining the direction

of the pulling force. Insets top show the angle between the force vector and the hydrogen bonds holding the structure together.

Unlike chemical and thermal unfolding, which rarely take place *in vivo*, mechanical unfolding is ubiquitous in nature. A protein would at least once be mechanically unfolded during its lifetime, during its degradation by the proteosome³. But many proteins have evolved to operate under mechanical force, and their repeated unfolding and refolding under a changing force vector is an important mechanism to transform mechanical energy into biochemical signaling and vice-versa¹. For this process to take place effectively, the proteins that operate under a force vector *in vivo* are typically segregated into many domains, and might also have purely elastic unstructured regions. This arrangement resembles that of beads-on-a-string. These multidomain proteins (or [polyproteins](#)) can respond to a mechanical perturbation by unfolding some of their domains, followed by the entropic extension of the now unstructured peptide chain, under the changing force vector. Through unfolding, the protein chain adds contour length to the ensemble, hence driving the decrease of the overall force experienced by the polyprotein molecule. The differences between single-domain and multi-domain proteins, as well as between chemical, thermal and mechanical stabilities of proteins, will be discussed in the last part of this chapter.

1.2. PROTEIN STRUCTURE

We begin our journey towards discussing mechanical unfolding of proteins with a short introduction on the structure of proteins. Unlike most books and reviews describing protein structure, our focus here will be on how various structure elements can become force-bearing. It is important to note that the human genome contains ~20,000 genes, and a typical cell has about 42 million protein molecules⁴. Among these protein molecules, some are present in very few copies (fewer than 10 molecules per cell), while others are very abundant (over 10,000 copies per cell). But most proteins are present in the 1,000-10,000 range. Furthermore, many of the proteins are being secreted by cells either in a soluble or insoluble form. Some secreted soluble proteins, such as hormones, allow cells to communicate with each other in a delocalized manner, across tissues and organs. Some secreted insoluble proteins form what is known as the extracellular matrix (ECM) and are specific for a tissue or organ, both in terms of composition and mechanical properties. The ECM is continuously reshaped by the surrounding cells and is actively used for local communications between cells through mechanical interactions⁵. The protein structure plays a key role in determining function, and here we will focus on mechanical aspects related to various structural organization of proteins.

1.2.1. PRIMARY STRUCTURE OF PROTEINS

The primary structure of proteins is formed by a sequence of up to 20 amino acids. Amino acids have a carboxylic (-COOH) and an amine (-NH₂) chemically reactive group, linked together through an alpha-carbon (C^α) (**Figure 1.2**). In physiological pH conditions, amino acids are [zwitterions](#). Due to the presence of these chemically reactive groups, [amino acids self-ligation](#) results in extremely long polypeptide chains. This ligation process is carried out by the ribosome, as further discussed in **Section 1.3**. Apart from the bonds forming directly the polypeptide chain, some proteins also have disulfide bonds, which form succeeding the folding process ⁶. These disulfide bonds will further constitute load-bearing bonds when a protein is experiencing mechanical unfolding. A summary of the lengths, energies and breaking forces of the bonds involved in the mechanical response of proteins is given in **Table 1.1**. The forces or mechanical energies required to break covalent bonds are typically too high to take place *in vivo*, and these bonds are broken chemically rather than mechanically. Such a process of mechanically breaking covalent bonds would be a homolytic process that produces free radicals and uncontrolled cross-linking between biological molecules⁷, as opposed to chemical cleavage, which typically is a heterolytic process. Nature does not like to play dice, as such radical-based processes would result in many reaction products, such as reactive oxygen species thought to damage cellular DNA. An exception to this rule is represented by the dislocation of bacteria tethered via single proteins, which can happen during transient mechanical events such as coughing ⁸.

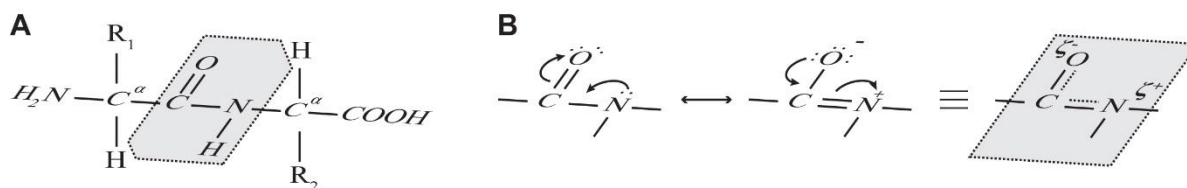
Table 1.1. Average parameters for covalent bonds found in protein structure.

Bond	Bond breaking force [nN]	Bond length [Å]	Bond energy [kT]
C – C	4.1	1.54	139
C – N	4.1	1.43	117
C – O	4.25	1.43	141
C – S	2.0	1.51	105
S – S	2.25	2.05	87

Compiled from refs. 9, 10

Apart from the extremely high bond energies holding a polypeptide chain together, a further reinforcement comes from the double carbon-oxygen bond that forms the carboxyl group. In the isopeptide bond, the electrons that form the carboxyl groups are delocalized over the O-C-N (**Figure 1.2**). This delocalization, also known as *amido–imido tautomerization*, has two effects: (i) it further strengthens the C-N bond inside the polypeptide backbone; (ii) it fixes this bond to a single plane, by restricting its rotation.

Figure 1.2. The steric of the peptide bond.



Partial delocalization of electrons between O-C-N prevents the peptide bond from rotating and increases its strength.

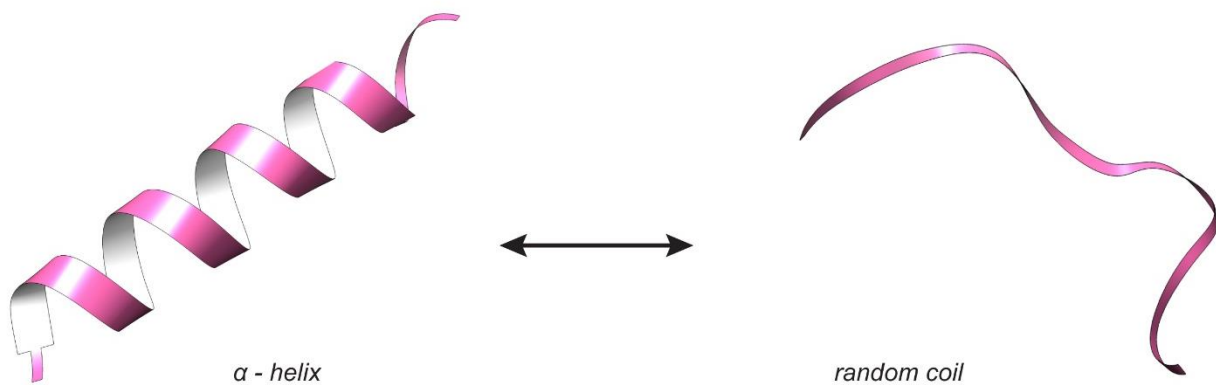
In conclusion, while the primary structure is unique for each protein, the backbone is essentially identical. Since the covalent bonds forming the polypeptide chain are very stable, both mechanically and thermally, they do not yield to the forces or temperatures typically encountered *in vivo*.

1.2.2. SECONDARY STRUCTURE OF PROTEINS

In 1951, Linus Pauling and Robert Corey were the first to report the structure of the main two components forming the secondary structure of proteins: the α -helices and the β -sheets¹¹. These elements, first found in keratin and silk proteins, were later shown to be part of all structured proteins. The reason for their common occurrence in proteins lies in the fact that the hydrogen bonds that drive the formation of α -helices and β -sheets appear between the N-H and C=O bonds that are part of the polypeptide chain, rather than between the side chains. The main differences in how the two secondary structure elements respond to a force vector is related to their assembly length and orientation to the force vector.

To form an α -helix, a polypeptide chain only needs to wind around onto itself (**Figure 1.3**). This twisting gives the α -helix a well-defined structure, resembling that of a spring, with a complete turn roughly every 3.6 amino acids and hydrogen bonds longitudinal to its orientation. The presence of the amino acids side chains can have subtle effects on the overall shape of the helix, which can show various kinks or additional twists. The denaturing of an α -helix into an unstructured polypeptide chain is known as the *helix-to-coil transition*. As hydrogen bonds forming a helix are parallel to the force, they will typically be mechanically weak.

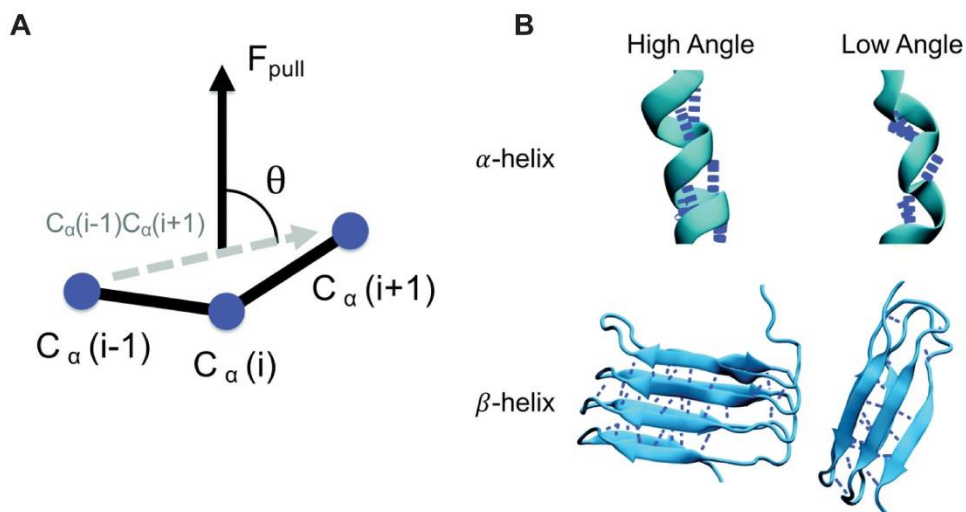
Figure 1.3. Helix-to-coil transition of a polypeptide chain.



Representations of the α -helix to random coil transition.

The stability of an α -helix taken outside the folded structure depends both on the orientation of the hydrogen bonds to the force vector, as well as its dimensions. Helices with hydrogen bonds oriented along the pulling axes (high angle between the hydrogen bonds and the pulling vector in **Figure 1.4**) will respond differently to force than those having additional twists (low angle in **Figure 1.4**)¹². As the load increases, an α -helix will display three distinct responses: initially a molecular rearrangement will orient the helix to the force vector without bond breakage, followed by the breaking of hydrogen bonds holding the helix structure resulting in a linear increase in elongation, and terminated by the stretching of the unstructured polypeptide backbone lacking any internal hydrogen bonds¹³. Furthermore, the stability of an α -helix also depends on its length¹⁴. For example, the α -helix from the protein **talin** (domain R6) having ~25 amino acids was shown to withstand 7 pN of force¹⁵, while another α -helix from the same protein but from a different domain (domain R3) was measured to form at forces all the way up to 20 pN¹⁶. However, in this last case, the coil-to-globule transition was catalyzed by the presence of another ligand protein, vinculin, which binds to two adjacent helices, that have together almost twice the number of amino acids (58 amino acids). For more details on this system, see **Chapter 5.3**.

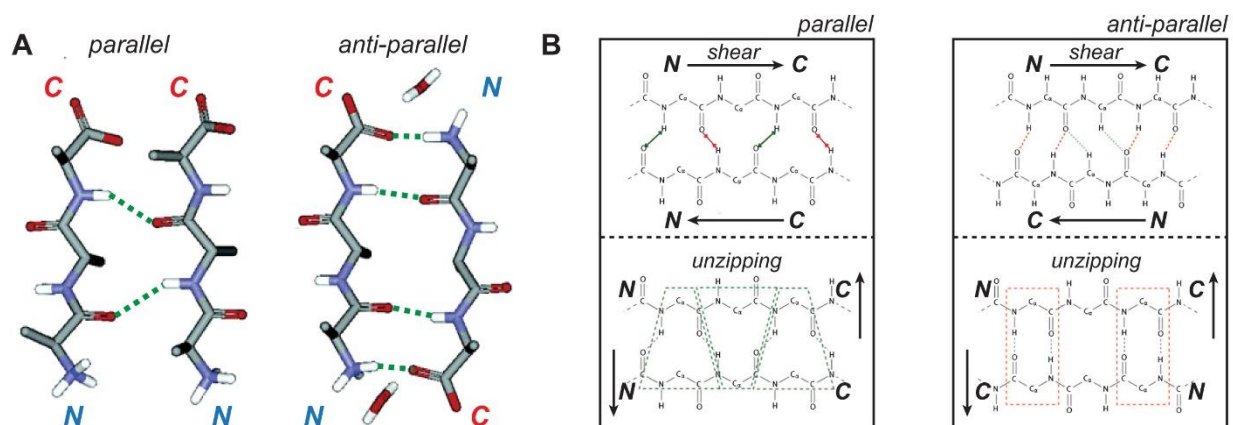
Figure 1.4. Mechanical stability of secondary structure elements, as affected by a pulling angle.



A) Diagram showing the pulling angle with respect to the alpha carbons. B) Renderings of α -helices (*top*) and β -strands (*bottom*) under a high (*left*) or low angle (*right*). C) Change in the angle between hydrogen bonds and force vector measured before the breaking of hydrogen bonds for keratin (*top* – having a single coiled coil structure) and biofilm-forming protein CsgA (*bottom* – having repeating beta strands subunits). Adapted from ref. 12 with permissions.

Unlike α -helices, which have sequential longitudinal hydrogen bonds along the amino acid chain, β -strands form between different regions of the protein, which are spatially brought about by transverse hydrogen bonds. Most of the β -sheets have more than two structural elements (strands) and will be further discussed in the following section, when talking about the tertiary structure of proteins. When two regions of a protein come spatially together to form two β -strands, they can be facing each other in a parallel or anti-parallel orientation (**Figure 1.5**). From a thermodynamic perspective, the antiparallel β -strands are considered more stable than the parallel ones, as in this configuration the hydrogen bonds are closer together. However, from a mechanical perspective, the parallel strands are more resistant to shear, as under force, hydrogen bonds will stretch and compress in succession¹⁷ (**Figure 1.5**). By contrast, the anti-parallel β -strands in a shear configuration will have all hydrogen bonds stretched at the same time. On the other hand, if the force vector is transverse to the β -strands, the hydrogen bonds will break similarly to opening of a zipper. In this case, as the anti-parallel orientation has more isolated hydrogen bonds, it will be mechanically less stable. Importantly, the energy needed to break β -strands longitudinally (shear) is significantly larger than the one needed to break them transversally (zipper). Hence, it is not surprising that the most mechanically stable proteins characterized thus far have two parallel beta strands formed from the two ends of their sequence, holding their tertiary structure firmly in place.

Figure 1.5. Configurations of β -strands.



A) Stick representation of two β -strands in parallel (left) and anti-parallel (right) configuration, with hydrogen bonds marked with dotted green lines. B) Effect of mechanical forces on the two β -strand conformations (parallel – left and anti-parallel – right) when applied in a longitudinal (shear – top) orientation and transversal (unzipping/peeling – bottom) geometry. Parallel β -strands in shear mode are the most mechanically stable, as hydrogen bonds elongate and compress during force loading (green and red arrows, respectively). Adapted from refs. 18 and 17 with permission.

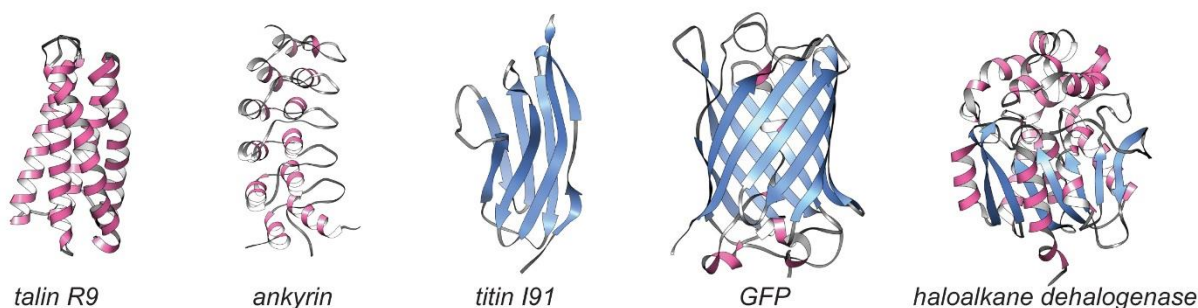
1.2.3. TERTIARY STRUCTURE OF PROTEINS

The tertiary structure of proteins is composed of α -helices, β -sheets and combinations of the two elements. Proteins made entirely from α -helices (all- α -proteins) are encountered in naturally forming load-bearing networks, such as the actin filaments, which are part of the cytoskeleton, motor protein myosin, membrane-supporting spectrin or mechano-transducing proteins such as talin and vinculin. A special type of all- α -fold is seen in ankyrins, which mediate the attachment of membrane proteins to cytoskeleton (**Figure 1.6**). The ankyrin fold has pairs of α -helices having hydrogen bonds transversal to the N-to-C coordinate, which makes them mechanically stable, resulting in a unique mechanical response to force¹⁹. Several proteins have only β -strands, which are connected into β -sheets. The **immunoglobulin fold** (Ig-domains) is found for example in most muscle **titin** domains, as well as in antibodies, extracellular protein fibronectin, in silk and amyloid fibrils. A special all- β fold is known as β -barrel, where the β -strands twist and coil to form a closed structure (**Figure 1.6**). This structure is seen in pore-forming membrane proteins (such as porin), as well as in the most popular class of fluorescent proteins (green fluorescent protein – GFP). Another interesting all- β fold is the β -solenoid (**Figure 1.4B bottom**). In this case, a superhelix is formed from antiparallel β -strands. β -solenoids appear in adhesins and other proteins used by bacteria during attachment¹². Many proteins have both α -helices and β -sheets in their fold. The last representation in **Figure 1.6** depicts the structure of haloalkane dehalogenase, an enzyme that catalyzes the replacement of a 1-haloalkane with a primary alcohol group. HaloTag, a mutant of this enzyme which lacks the group needed for the last hydrolysis step, was shown to have

completely different unfolding responses when tethered between the centered enzymatic site and the N or the C-terminus ²⁰.

Before we conclude the subsection on the tertiary structure of proteins, it is worth reiterating that many of the proteins that operate under force *in vivo* are composed of several domains, resembling the beads-on-a-string arrangement. For example, talin has four head, 13 tail and one dimerization domain ⁵. Titin on the other hand is an alternately spliced protein, having different lengths depending on the muscle type, but has over 100 IgG-like domains and several unstructured regions, as well as two terminal dimerization domains, Z1Z2 ²¹. As explained in **Section 1.4**, these domains linked together should be seen as a single protein and produce some unique properties. Due to this arrangement as beads-on-a-string, most mechano-transducing proteins have a filament-like shape, rather than a globular shape.

Figure 1.6. Tertiary structure of proteins



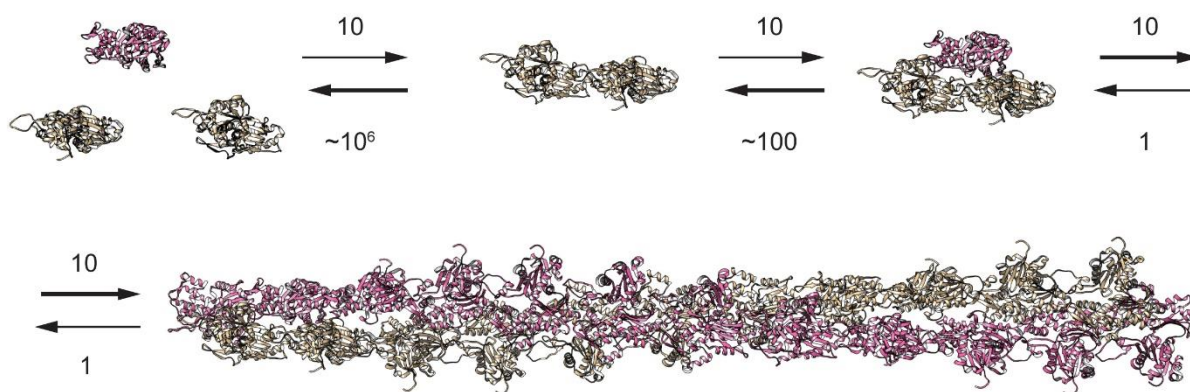
From left to right: all- α domain R9 of talin, a protein involved in focal adhesion formations (pdb: 2kbb); all- α with special configuration ankyrin domains (pdb: 4b93); all- β muscle titin IgG-like domain **I91** (pdb:1tit); all- β barrel domain GFP (pdb: 1gfl); α - β haloalkane dehalogenase enzyme (pdb: 3g9x). Structures rendered using Chimera.

1.2.4. QUATERNARY STRUCTURE OF PROTEINS

Proteins operating under force *in vivo* are often the molecular trusses holding the shape of a cell or are part of the surrounding extracellular matrix. Hence, it is no surprise that these proteins are found linked together in either **homoomers and heteromers**, and even forming extended supramolecular assemblies. For example, to form focal adhesions, which help link cells to their extracellular matrix, more than 50 different types of proteins are needed during assembly ⁵. The connections forming the quaternary structure of load-bearing proteins determine both their stability and assembly/disassembly dynamics. Often the intra-domain interactions are relatively weak, with binding constants having values of micromolar (mM) or hundreds of nanomolar (nM), and the cooperativity between these interactions plays an important role. However, in some cases, these interactions are strong and require enzymes or large force to form and break. Such

is the case of actin, which is an important component of the cellular cytoskeleton and part of the sarcomere, the contractile unit of muscle cells ²². Actin is present both in a monomeric and polyprotein form. The monomeric form is known as globular (g-) actin, while the polymeric form as filamentous (f-) actin. F-actin is made of two helical strands wound around each other and can dynamically grow or shrink by recruiting or eliminating actin monomer (**Figure 1.7**). The polymerization of actin is driven by ATP, and accessory proteins such as profilin and cofilin continuously regulate its assembly and disassembly. The continuous reshaping of actin network is responsible for the crawling locomotion of cells, as well as maintaining the structure of the cell and providing a trafficking network ²².

Figure 1.7. Assembly and quaternary structure of actin.



Actin monomers spontaneously polymerize to form filamentous (f-) actin. Small oligomers have higher dissociation than association constants, while filaments are more stable and with higher association constants (values above/below arrows). Structures rendered using Chimera from pdb: 5mvy.

1.3. THE MAIN STAGES OF PROTEIN FOLDING

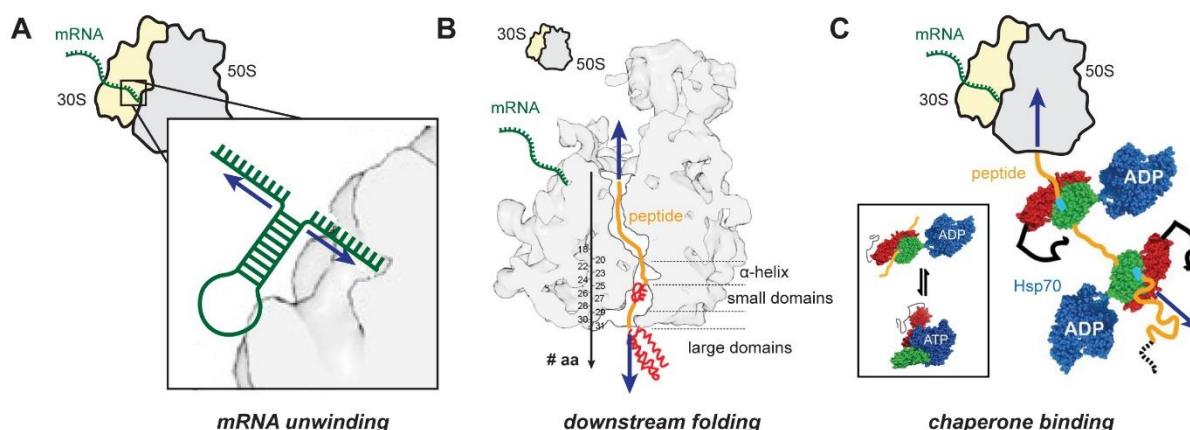
Protein folding is a fascinating process that involves acquisition of a well-defined 3D structure from a given amino acid sequence. Cyrus Levinthal pointed out an apparent paradox, which states that to randomly sample all the possible states, an average protein-forming polypeptide chain (~ 100 amino acids) would require an incredibly large amount of time (even longer than the age of the universe), in stark contrast to the actual measured folding time of proteins *in vivo*, which typically ranges from milliseconds to several seconds. The fact that the protein folding process is somehow guided and follows some yet-to-be-discovered rules was further confirmed with the introduction of the AlphaFold, a machine learning algorithm introduced by Google in 2018. Without understanding the inner workings of the protein folding process, AlphaFold can accurately predict protein structure from amino acid sequences, after being trained with over 170,000 known structures, from public repositories ²³. Hence, this machine learning success further supports the

idea that some general rules exist to reduce the conformational search and to optimize the path taken by a polypeptide chain to become a structured protein. Below we will focus on the main steps of protein folding and try to understand how chemical and mechanical elements interplay in the process. We will argue that the theoretical conformation space is in effect reduced by orders of magnitude by entropic forces, which are also responsible for how much a polypeptide chain extends under a force vector.

1.3.1. THE RIBOSOME AND HOW THE POLYPEPTIDE CHAIN IS ASSEMBLED

The ribosome is the molecular factory of the cell and assembles transfer (t)RNA-bound amino acids together into a long polypeptide chain. During translation, one side of the ribosome complex reads the nucleotide sequence along a messenger (m)RNA molecule, three nucleotides (one codon) at a time. On another side, the ribosome is synchronously assembling a polypeptide chain from a possible of 20 amino acids, based on the matching codons. Protein folding effectively starts as the primary structure is assembled and extruded from the exit tunnel of the ribosome. There are several key aspects that we will focus our discussion on in this subsection, where mechanical forces are involved: (i) pull forces that the ribosome needs to exert to break secondary structure elements of mRNA, (ii) **steric** exclusion forces generated by the nascent polypeptide chain along the exit tunnel and (iii) the **chaperones** that assist the ribosome during co-translational folding (**Figure 1.8**).

Figure 1.8. Forces related to the assembly of the primary structure of proteins.



A) Schematics showing the interaction of the ribosome with mRNA. Before the ribosome can read the mRNA codon sequence, it needs to unwind secondary structures, which otherwise would be too large to enter. B) Downstream folding of multidomain proteins generates a steric extrusion force on the nascent polypeptide. Black vertical arrow shows number of amino acids (aa) inside the exit tunnel. C) Chaperones can bind and transform chemical energy into mechanical force to

maintain a polypeptide unfolded. The force vector is shown in all panels with blue arrows. Figure adapted from refs. 24 and 25.

The first-time force is required during protein synthesis is when the ribosome unzips the mRNA molecule, which encodes the amino acid sequence of a given protein. Unlike DNA, RNA is single-stranded and can form elements of secondary structure, such as hairpins and pseudoknots (**Figure 1.8**). These hairpins are too large to enter the ribosome and are not accessible for readout, slowing down the translocation of mRNA. From a geometric perspective, mRNA hairpins can be in a zipper conformation. Ribosomes unwind mRNA hairpins by force, using a forward rotation of a head subunit, while coupled to a translocation factor ²⁶. The slowing down of mRNA translocation due to hairpin unwinding can play a role in pacing the cotranslational folding of proteins. Interestingly, while the force needed for mRNA unzipping is proportional to the number of hydrogen bonds, the mechanically stronger hairpins have a similar effect as weaker ones, acting more like an allosteric switch between a fast and a slow translation mechanism ²⁶.

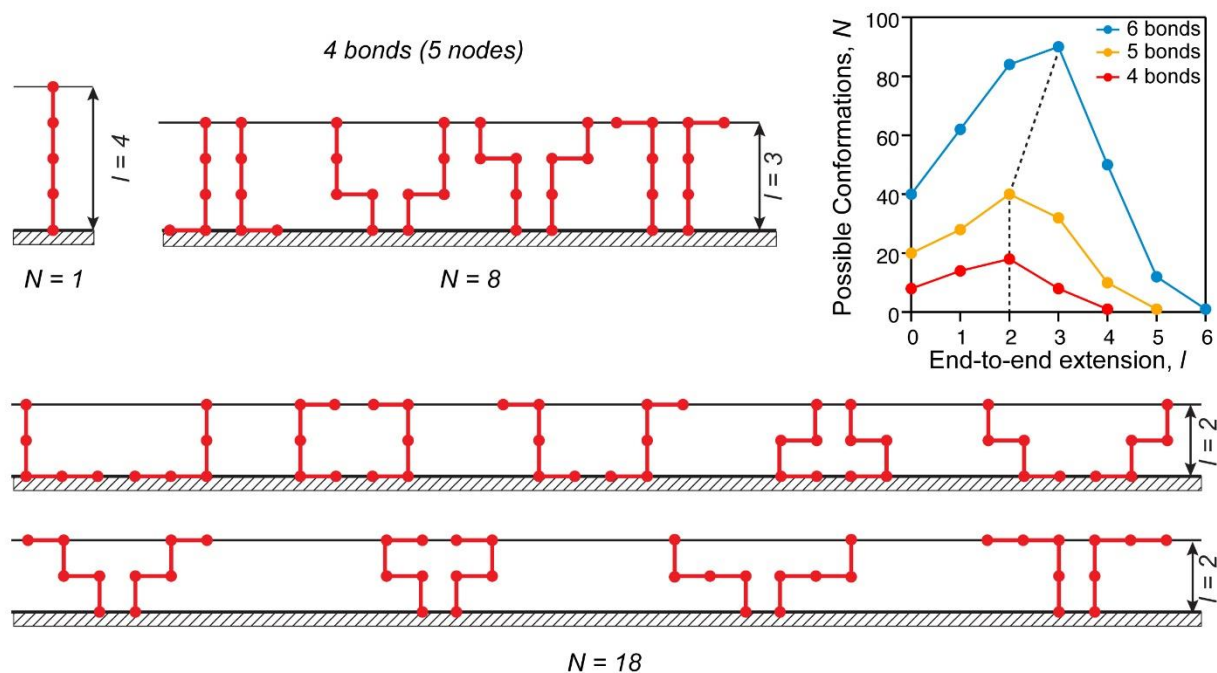
The exit tunnel of the ribosome extruding the nascent polypeptide chain has a diameter of up to 1.4 nm and a length of ~8 nm ²⁷. This tunnel is too narrow to accommodate most folded proteins, but secondary structure elements (typically α -helices, which have diameters of ~0.92 nm) and small protein folds have been reported to form inside it (**Figure 1.8**) ²⁴. Due to this inability of proteins to fold inside the exit tunnel of the ribosome, a steric exclusion force is generated, which plays an important role in pulling the nascent chain during domain folding, especially for multidomain proteins. The pull force can be generated by an adjacent domain already translated, which can fold faster than the translation speed. While still being assembled, the polypeptide chain experiences for the first time a pulling force! This force was measured to be at least 10 pN on an emerging peptide of 15-22 amino acids, generated by the downstream folding of a 93 amino acids-protein ²⁸. Hence the rate of expression of a polypeptide chain forming a multidomain protein will spike as the preceding domains outside the [ribosome exit tunnel](#) acquire its structure.

Chaperones are helper molecules that ensure the correct folding of proteins, translocation across plasma membrane and even disassembly of protein aggregates. They typically operate close to the exit tunnel of the ribosome, by either preventing a protein to fold until more of its sequence is produced, or by sterically confining a polypeptide chain inside a barrel-like assembly, to ensure what is known as frustrated folding. One of the chaperone proteins operating near the translocation pore of the ribosome, the heat shock protein (Hsp)70, was shown to relieve translation stalling on the ribosome by exerting a pulling force ²⁹ (**Figure 1.8**). Hsp70 requires chemical energy in the form of ATP to bind to a nascent polypeptide and generate a pull force through loss of entropy. During binding of the relatively large chaperone molecule to the polypeptide chain, the extruded volume of the complex effectively limits the conformation search space of the nascent polypeptide, reducing the number of accessible conformations. This entropic force due to increased excluded volume was found to exceed thermal fluctuations for emerging polypeptides smaller than 30 amino acids and was estimated to be up to 20 pN (equivalent to energies up to 5 kT) ²⁹.

1.3.2. THE ENTROPIC COLLAPSE REDUCES THE CONFORMATIONAL SPACE OF A POLYPEPTIDE CHAIN

The currently unsolved question of how proteins can acquire a unique well-defined structure given the immense conformational search space of a polypeptide has been puzzling scientists for decades. Among driving forces, entropy is the main contributor for reducing the number of available conformations and decreasing the folding time from taking longer than the age of the universe to several milliseconds to seconds. Surprisingly, most books and reviews show the picture of the rugged folding landscape of protein folding only after the entropic collapse happened, and after folding already proceeded over 99% of the way. Entropic collapse can seem like an esoteric force, as it quantifies the amount of free energy that is **not** available for conversion into mechanical work, rather than formation of physical bonds. To demystify it, we consider a chain of four units that can only move on orthogonal coordinates in 2D space (up, down, left, right), and has a fixed end (as would be the case of a polypeptide chain tethered to the ribosome). We also consider that two segments cannot occupy the same space (excluded volume approximation) (**Figure 1.9**). Given these assumptions, our four-units chain can acquire a single conformation when fully stretched, eight conformations when the end-to-end distance is three units, and so on.

Figure 1.9. Stick model describing how entropy drives polymer chain collapse



A hypothetical polymer composed of 5 monomers (4 bonds) and only constrained to vertical and horizontal coordinates, has the natural tendency to collapse, in order to maximize its entropy, S . Bottom surface is assumed as a physical barrier and anchoring point (equivalent to the ribosome surface). Driven by the change in entropy, the change in free energy ΔG will vary (approximated

from equations (1.1) and (1.2) from 0 kT (since for $l = 4$ we have $\ln(1) = 0$) to $-2 kT$ (for $l = 3$, $\ln(8) \sim 2$), and further to $-3 kT$ (since $\ln(16) \sim 3$ for $l = 2$). (Top right inset): Change in the number of possible conformations with number of bonds; dotted line marks the end-to-end extension l with the highest entropy.

As with any process in nature, protein folding must follow the laws of thermodynamics and requires a negative change in free energy, ΔG , to proceed:

$$\Delta G = \Delta H - T\Delta S < 0 \quad (1.1)$$

where ΔH is the change in enthalpy, ΔS is the change in entropy, and T is the process temperature. The entropy is defined as:

$$S = k_B \ln(N) \quad (1.2)$$

with k_B the Boltzmann constant and N is the number of equivalent states. In **Figure 1.9** the entropy of an extended chain is zero ($\ln(1) = 0$), the three units extended state is $k_B \ln(8)$, while the two units extended state has $k_B \ln(16)$. Note that a chain extending two-units has a smaller entropy than one extending three units, due to surface effects, as it is costly for a chain to collapse too close to a surface.

The likely reason entropic effects are typically ignored when describing protein is that at no point in time is the polypeptide chain fully stretched after exiting the ribosome channel or being released by a chaperone. For our four-units peptide chain an energy of only $\sim 3k_B T$ would be required to completely stretch it from a minimum energy state. However, as the number of bonds increases, the energy required to fully stretch a polypeptide chain will quickly exceed that needed to hold together the covalent bonds forming the backbone. Hence, the entropic energy should be seen more like a potential energy that continuously drives the collapse of the polypeptide chain as it is being synthesized. This potential energy can however be partially sampled using force spectroscopy techniques, which can mechanically unfold single proteins and drive the polypeptide chain in an extended denatured state ³⁰. More details on this extended denatured state will be presented in the following chapters.

When the polypeptide chain is relatively extended, the enthalpy ΔH will not play a significant role, apart from the formation of a few α -helices along the peptide chain. Entropic collapse is important not only for folding, but also for how proteins function under force *in vivo*. It is the reason why, at physiological forces, a protein can refold against a force vector and instantaneously contract tens of nanometers, well above the range of hydrogen bonds and hydrophobic effects. It is also the reason why an unfolded polypeptide chain extends to a given end-to-end distance at a given pulling force. In **Chapter 2** we will describe how entropy can create a barrier and separate an extended and collapsed state tens of nanometers from each other, while in **Chapter 5** we will

look at how the entropic collapse of a single domain of the muscle protein titin can deliver similar amounts of work as that produced during the stroke of myosin motors walking on actin filaments.

1.3.3. THE HYDROPHOBIC COLLAPSE AND THE MOLTEN GLOBULE STRUCTURES

As the name implies, the hydrophobic collapse is driven by interactions of amino acid residues with water molecules. About half of the natural amino acids have hydrophobic residues and have the tendency of clustering together while being excluded by the solvent, as they favor interactions among themselves to interactions with water molecules. This segregation of hydrophobic and hydrophilic residues due to their interaction with water is an enthalpy driven process. Obviously, a segregation of amino acids based on their interaction with water and themselves comes at the expense of entropy. The free energy ΔG (**equation 1.1**) decreases during the hydrophobic collapse, because the ΔH term decreases faster against the increase in the entropic term $-T\Delta S$.

As the polypeptide chain collapses, the amino acid residues also come close enough to allow for the dynamic formation of hydrogen bonds. This part of the hydrophobic collapse gives rise to a structure known as the *molten globule*³¹. As its name suggests, the molten globule is a dynamic structure, where some hydrogen bonds continuously form and break. Furthermore, water molecules leave the folding core, resulting in further collapse. There is no significant difference between the gyration radius of the molten globule state and that of the native state (typically less than 1 nm). However, the molten globule states do not have fully formed secondary structure elements (especially β -strands that form between different regions of the polypeptide chain), but retain many of the contacts between amino acids over time³². Hence, molten globule structures, while having similar end-to-end distance as the native state, are significantly less stable.

1.3.4. THE NATIVE STATE OF PROTEINS

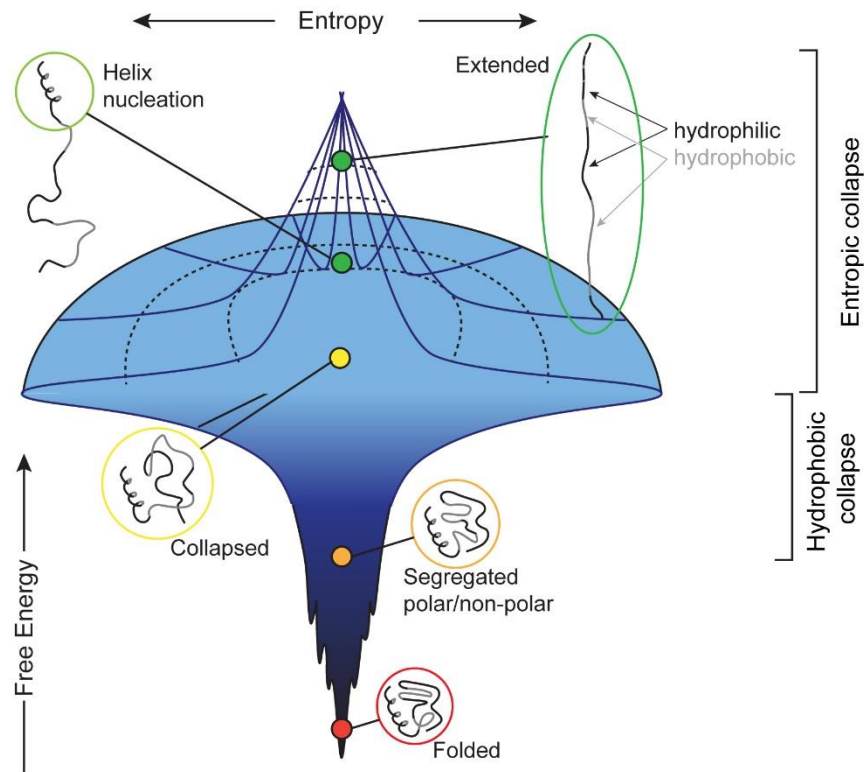
As the number of conformations decreases beyond the molten globule, a protein acquires a well-defined number of α -helices and β -strands, separated by unstructured regions. These structures are held together by hydrogen bonds, that can dynamically break and form without significant changes in the number and orientation of the secondary structure elements. Some proteins can fold in more than one native structure, and these proteins are known as fold-switching proteins³³. For some proteins, the decrease in the enthalpy is not sufficient to compensate for the change in entropy ($-T\Delta S$) to reach a final native structure. These proteins are known as intrinsically disordered proteins, and they can acquire a tertiary structure (or structures) only after the interaction with a ligand or with the membrane³⁴.

1.3.5. TIMING AND ENERGY LANDSCAPE REPRESENTATION OF THE FOLDING PROCESS

The formation of α -helices can be even faster than the entropic collapse, as it is driven by hydrogen bonds forming between adjacent amino acids. Similarly, to a spring recoiling in place, the formation of α -helices takes place on the nanoseconds time scale ³⁵. The formation of β -sheets on the other hand requires some major rearrangements, as parts of the polypeptide chain far away in sequence must come together in space. Hence β -sheets will form after the entropic collapse, between the molten globule and the native states. The entropic collapse of a polypeptide chain tethered from a single end takes place on the microsecond (ms) timescale, as the diffusion dynamics and dihedral rearmament of a polymer chain are relatively fast processes ³⁶. The entropic collapse time will depend on the chain length, and a different way of estimating it is by using its diffusion coefficient, which was determined using both molecular dynamics simulations and experiments to be $D \sim 2 \times 10^8 \text{ nm}^2/\text{s}$ ³⁶. Hence the entropic collapse (in the absence of a tethering probe) takes place on a nanosecond timescale ³⁶. As both helix nucleation and entropic collapse are faster than the synthesis rate of the ribosome, the two processes will follow the rate-limiting transcription. However, these processes are extremely relevant when talking about the folding energy, as well as when describing proteins unfolding and refolding under force. The first steps that have a significantly long time, and can compete with translation, are the hydrophobic collapse and formation of the molten globule structures. These steps were predicted theoretically and measured experimentally to take place on the millisecond timescale ^{35, 37}.

From an energy perspective, the free energy changes of the folding process are first driven by chain entropy, and then enthalpic interactions take over. The protein folding process, from the (potentially) extended polypeptide chain to the native 3D structure, can be summarized through its **energy landscape** (**Figure 1.10**). As the free energy decreases (y -axis), the entropy of the molecule first increases (x -axis), driving its transition from an extended to an entropically collapsed state. Concurrent to this process, α -helix formation occurs ³⁵. As the protein undergoes the hydrophobic collapse the entropy starts to decrease, while enthalpy becomes the driving force. As folding proceeds through the segregation of hydrophilic/hydrophobic amino acids and acquisition of the native state, the entropy further decreases, mirroring the fewer number of available states. However, this entropy decreases starting from the hydrophobic collapse ($\Delta S < 0$) is compensated at every step by a steeper decrease in enthalpy (ΔH), producing a net negative change in the free energy ΔG (see eq. 1.1). Each protein will have a unique structure and a well-defined number of interactions preserving its native state, and these interactions will be reflected by the enthalpic term, which is important at the last stages of the folding process.

Figure 1.10. Energy landscape of protein folding and its main steps driving protein folding energy landscape with significant steps.



The transparent lines at the top represent the potential energy coming from the entropic collapse of a polypeptide chain. The entropic collapse is followed by a hydrophobic collapse and formation of the molten globule structures, accompanied by the segregation of the hydrophilic and hydrophobic amino acids while interacting with the water environment. The representation of the lower funnel part of the landscape was adapted from ref. 35.

1.4. PROTEINS AND POLYPROTEINS

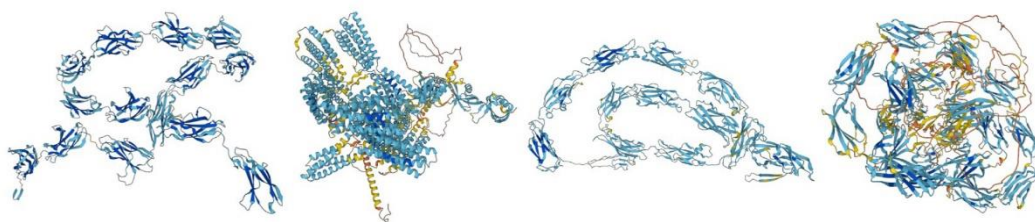
As previously mentioned, polyproteins are multidomain proteins covalently connected in series by extensible unstructured polypeptide segments. Polyproteins have a beads-on-a-string configuration. A prominent example can be met in the giant muscle protein titin that traverses across the half-sarcomere (the basic contractile unit along the muscle fiber), which is comprised of hundreds of domains of fibronectin type III, globular immunoglobulins and flexible [PEVK](#) sections ³⁸. Such a serial arrangement can be advantageous from several structural and functional perspectives. First, it is less susceptible to complications during its folding as it folds into small consecutive domains (unlike large complex proteins) ³⁹, and it also induces pulling forces during the protein synthesis at the ribosome surface ²⁸. Second, when operating under mechanical load, polyproteins made of different domains can adjust the tension distribution through hierarchical unfolding and extension, starting with the less stable domains ⁴⁰. A polyprotein made from identical domains can form correlations along its consecutive domains, thus stiffening its backbone ^{41, 42} and stabilizing ^{43, 44} its integrity.

Natural polyproteins can also be found in viruses and in non-viral organisms ⁴⁵. In viruses they are implicated in proteome structuring, such as in human immunodeficiency virus (HIV) ⁴⁶ and in SARS coronavirus ⁴⁷. Non-viral polyproteins can be found in tissues ⁴⁸, in the cell-extracellular matrix interface ⁴⁹, and also sometimes, when referred to as tandemly repetitive polyproteins, they are produced as large precursor proteins ⁵⁰.

Specific to mechano-related physiological processes, the unique conformation of polyproteins is exploited in their biological settings. In response to mechanical load, polyproteins can regulate their elasticity through transient conformational changes of their constituent protein domains and linkers ⁵¹. This means that when subjected to externally applied forces, unstructured polypeptide segments and linkers can elastically extend, followed by possible unfolding of protein domains to form additional linker extensions. Upon the removal of the load, the unfolded proteins can recoil, collapse, and eventually refold. Through this mechanism, the polyproteins can regain their mechanical stability and consequently reduce the unfolded polymeric linker length. These conformational changes eventually lead to elasticity changes, sometimes necessary for the exposure of cryptic sites that are protected by the protein/domain folded conformation ⁵². More importantly, this unfolding and refolding process can be repeated many times during the lifecycle of protein domains, during which additional contour length is added and subtracted, leading to faster elasticity changes than what would be otherwise alternatively possible through protein synthesis or denaturation.

In addition to the previously mentioned giant multi-domain titin polyprotein that regulates tension and energy storage during muscle contraction ^{53, 54, 55}, other examples of natural elastomeric polyproteins can be found in talin ^{56, 57} and [cadherin](#) ^{58, 59}, which are involved in mechano-transduction, and the extracellular matrix polyprotein fibronectin ^{60, 61} that participates in cellular adhesion (illustrated in **Figure 1.11** and discussed in **Chapter 5**).

Figure 1.11. Structure models of polyproteins.



From left to right: segment of muscle titin (pdb: AF_AFQ8WZ42F97); talin (pdb: AF_AFQ9Y4G6F1); cadherin (pdb: AF_AFQ9H251F2); fibronectin (pdb: AF_AFP02751F1).

In addition to natural polyproteins, a new class of synthetic polyproteins emerged in recent decades using single-chain engineering approaches. These engineered polyproteins have applications in biotechnology and therapeutics ^{62, 63, 64}, and are extensively used for basic research ^{1, 65, 66, 67, 68, 69, 70}. These artificial polyproteins, which are made of tandem protein repeats, are synthesized by cloning only the sequence coding for certain domains into an expression plasmid.

They can be comprised from the same domain or different variations of domains, where each domain has its own mechanical properties. These engineered protein constructs, in which the same domain is repeated several times (typically 8-12 times), constitute also a novel molecular tool to study the mechanical response of proteins *in vitro*. Engineered polyproteins not only provide a unique molecular fingerprint as opposed to a monomeric protein, but also allow for the study of subtle effects, such as how the chain entropy arising from the growing unfolded polypeptide chain affects the stability of the remaining folded protein domains.

1.5. CHEMICAL, THERMAL AND MECHANICAL STABILITIES

Protein stability refers to the ability of the protein to persist in its native folded conformation. In their native state, globular proteins consist of a hydrophobic core that forms in their center, and surrounded by a hydrophilic exterior, which is stabilized by the aqueous environment ⁷¹. In their natural surroundings, proteins are exposed to a fluctuating environment, in which temperature, chemical surrounding and forces interweave. Small-scale fluctuations affect the native state, which is somewhat dynamic in nature, yet these small instabilities are not sufficient to modify it ⁷². Large fluctuations, however, can (and sometimes do), intervene with the integrity of the protein's conformation, as proteins can unfold spontaneously for short periods of time. In some cases, these perturbations are an integral part of the protein functionality, while in other situations, they pose a threat to it in situations of inflammation and disease.

Hydrogen bonding is one of the important interactions within proteins that hold their tertiary structure. Hydrogen bonds form between a hydrogen and an electronegative oxygen or nitrogen atoms ⁷³. Hydrogen bonds hierarchically form interactions between structural elements within the protein, and with its surrounding molecules. The location and magnitude of the hydrogen bonds are fundamentally associated with the protein structure, stability, and function ⁷⁴. Hydrogen bond networks can regulate the thermodynamic and mechanical properties of the protein, through fluctuations in which local bonds dissociate and re-associate ⁷⁵.

Solvent-protein interactions essentially govern the thermal and mechanical stability, conformation, association, and kinetics of proteins. They are essential in a vast array of physicochemical and biological phenomena spreading from human physiology to biotechnology and ecology applications. The solutions in which proteins are immersed constitute environments that satisfy specific chemical and interfacial conditions required for their optimal functionality. Protein stability relies on weak non-covalent interactions, where structural changes, such as folding/unfolding, are a consequence of modifications involving the previously mentioned hydrogen bonds, van der Waals and electrostatic interactions. These interactions are affected to large extent by the protein environment, which involves the solvent (composition), and net forces applied by temperature changes and mechanical forces. These interactions are also prevalent in the protein environment and can have stabilizing effects on both folded and unfolded conformations.

Chemical (solvent interactions), thermal and mechanical (force) perturbations are the main three mechanisms for disrupting and intervening with the structure of the protein and they affect

the stability of a protein through different mechanisms. Solvent molecules can chemically interact with specific amino acids on the surface or along the backbone of the protein and have a stabilizing/destabilizing effect. Raising the temperature increases the internal energy of the protein, and consequently hydrogen bonds within it become unstable, and hydrophobic/hydrophilic interactions can be modulated. Force acts on the protein, and hierarchically disrupts its structural elements. We now briefly focus on each one of these mechanisms.

1.5.1. Chemical denaturation

Structure and function of proteins are greatly affected by their solvent environment ⁷⁶. The quality of the solvent is a general property accounting for overall effects that eventually induces steric (conformational) modifications of the protein. Protein solubility depends on the sequence and number of the exposed residues, which determines the distribution of hydrophobic and hydrophilic areas along its surface and core. Higher content of surface-exposed charged and polar residues can increase the protein solubility through the interaction with the water molecules and ions in the surrounding media, whereas a higher number of surface hydrophobic residues can decrease the protein solubility ⁷⁷.

Specific ion effects are typically complex in nature as they result from intermolecular interactions involving ions, solvent molecules and the protein solute. Protein conformations, depending on properties such as their shape and charge density, vary not only from their interactions from ion to ion, but can also be affected by changes in the local concentration of the same ion ⁷⁸. These effects are prevalent under a variety of settings, such as in neurodegenerative diseases (protein fibrillation, aggregation and amyloid formation) ⁷⁹, in pharmaceuticals and biotechnology ⁸⁰, in the food industry ⁸¹, and in the design of protein-based materials ⁸².

Specific ion effects on protein stability were first ranked by the *Hofmeister* in his famous series based on an ion's ability to stabilize the conformation of a protein (precipitation or [salting-out](#)) or to interrupt it by unfolding ([salting-in](#)) ^{83, 84}. When the stability of a folded protein increases with increasing salt concentration, the effect is referred to as salting-in, while when it is decreasing, it is known as salting-out. In diluted aqueous electrolyte solutions, the process is viewed from the perspective of water structuring rather than protein stabilization, and can be either [kosmotropic](#); (hard and strongly hydrated) or [chaotropic](#) (soft and weakly hydrated). Hydration theory ⁸⁵ defines kosmotropes (water order-makers) as hard and strongly hydrated ions that are either small or that have high charge density (stabilize proteins). By ordering the water molecules around them, they effectively deplete the water molecules surrounding the protein, making the solution a poor solvent for the protein, and thus increasing the stability of the protein's folded (or collapsed) state. Chaotropes (water disorder-makers), on the other hand, are weakly hydrated soft ions that, by making the solution a better solvent, induce unfolding to increase the surface area of the solute in contact with the solvent. This effect is achieved by disordering the water molecules in their vicinity, thus facilitating the hydrations of the proteins in the solution ⁸⁶. An alternative proposes that these effects are a result of the direct interactions between the ions and the specific structure

and composition of each protein (or other biomolecule) in terms of the chemical properties of each residue ⁸⁷ or that they take place via the first hydration layer ⁸⁸.

Apart from folding and refolding dynamics, solvent quality dramatically manifests itself through the coil-to-globule phase transition (collapse). This transition is considered as a second order phase transition ⁸⁹, in which the chain collapses from an expanded coiled state into a compact globular conformation due to changes in solvent quality ⁹⁰. The coil-to-globule phase transition is considered to be essential for protein folding ⁹¹. For expanding the definitions of solvent quality and conformational states, the reader is referred to **Chapter 2**.

Overall, the presence and species of solvent and osmolyte molecules can stabilize the native structure of the protein or drive it towards unfolding. Their concentration and species tune the biological performance of the protein in its biological environment and can also impair it in situations of disease when they considerably deviate from their normal characteristic values.

1.5.2. Temperature and thermal stability

Proteins perform through conformational changes that involve energies in order of magnitude of thermal energy. From the initial stages of protein synthesis and folding from the ribosome, a polypeptide chain is subjected to the surrounding temperatures. At physiological temperatures, proteins maintain their functional dynamics and activity through their structural flexibility, enabled by an interplay between the internal weak interactions (van der Waals and hydrogen bonds, salt bridges, dipoles and electrostatic effects, etc.), and with their external environment. Some proteins have naturally evolved to function under high temperatures in thermophilic and hyper-thermophilic conditions. The heat stability of such proteins emerges from enhancement of all forms of their weak interactions that eventually stabilize their structural conformations ⁹².

From an energetic perspective, the transitions between folding and unfolding, as driven by temperature, can be simplistically described as a two-state process over an energy landscape, in which the folded and unfolded states are separated by an **activation barrier** ⁹³. Heat is a natural denaturant that affects the bonding and interactions within the protein and with its surroundings. The unfolding barrier was intensively studied with respect to temperature effect on its thermodynamic properties ^{94, 95}.

The **Gibbs free energy** (G) of the transition is comprised from enthalpy (H) and entropy (S) terms through the relation that reflects a balance given in Eq. (1) ($\Delta G = \Delta H - T\Delta S$, where Δ denotes the differences between the folded and unfolded states, and T is the temperature). This is the free energy needed to perform the work required to disrupt the folded conformation of the protein at a given temperature and pressure, and it provides a measure for the thermodynamic stability of the protein ⁹⁶. From this relation, it can be seen that increasing the temperature will decrease the overall value of the change in free energy for the transition, up to the point where thermal fluctuations will drive the transition to the unfolded state.

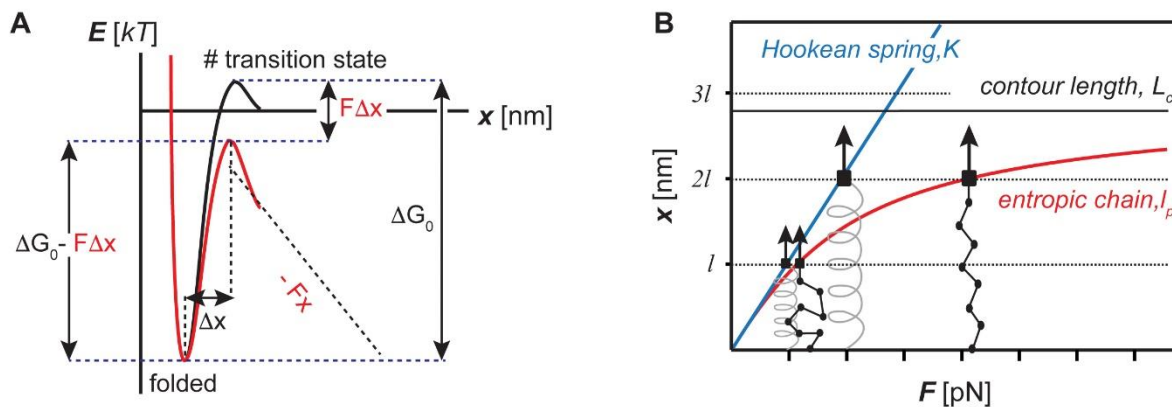
The activation enthalpy (ΔH) of the protein is determined by van der Waals and electrostatic interactions, and reflects the energetic interference in these bonds (that preserve the protein

structure). The enthalpy change is relatively low (within the order of fractions to several units of kT), however the overall ΔG is on the orders of tens to about a hundred of kT under ambient conditions⁹⁵. Entropy is mostly dominated by the configuration of the protein and is related to the number of possible conformations that are available to the polypeptide chain. At high temperatures, the activation enthalpy changes (on one hand, weak interactions decrease, however hydrophobic effects get stronger^{94, 97}), and the number of accessible conformations grows considerably, thus increasing the activation entropy ($-T\Delta S$). It should be noted that the validity of the two-state model approximation to describe thermal denaturation is debated⁹⁸, and will be further addressed in **Chapter 2**.

1.5.3. Force and mechanical perturbations

Force is a natural agent that is physiologically exploited by organisms and biomolecular processes to modify the structure of proteins, and mechanical unfolding can play a functional role⁹⁹. If temperature speeds up the diffusion along the energy landscape separating the folded and unfolded states of proteins, force reshapes the energy landscape altogether. In this sense, force can be considered as a denaturant that interferes with the protein's activation barrier by separating the folded and unfolded states, and consequently affects the probabilities of being in specific conformations (**Figure 1.12A**). Force also changes the spatial separation on the energy landscape between the minima of the folded and refolded states. Similar to a spring that will extend longer under a higher force, a polypeptide will have a larger end-to-end distance with increasing force. However, unlike a [Hookean spring](#), a protein will only extend linearly at relatively low forces and will asymptotically tend to its contour length as the force increases (**Figure 1.12B**). Proteins exposed to forces in physiological systems typically experience typical forces of 0 – 30 pN and undergo conformational changes across lengths of tens of nanometers (nm). Exposure to mechanical forces over time has been applied to a selection process, and through this process, mechanically stable proteins and proteins performing under load have evolved.

Figure 1.12. Mechanical aspects of force application to a protein.



A. Simplistic one-dimensional representation of the free energy of a protein with an unperturbed energy barrier with magnitude ΔG_0 and the distance to the transition state Δx (black line). Upon the application of a force F , the free energy is modified (red line). The application of force reduces the height of the energy barrier by extent of $-F\Delta x$, and the probability of leaving the folded state (unfolding) increases. B. Elastic description of the polymeric behavior of the unfolded protein with contour length L_c , as Hookean spring (blue) characterized with a spring constant K , and as entropic chain with a persistence length l_p .

Different elements and motifs within the protein respond to force in different compliances. Beta structures (β -sheets) are modules that can endure forces even of several hundred pN, while alpha structures (α -helices) are considerably weaker, and unfold at a few tens of piconewtons^{100, 101, 102}. β -sheets gain their mechanical stability from a motif called mechanical-clamp¹⁰³. In its most stable conformation, the mechanical clamp has a primary set of hydrogen bonds arrayed between two parallel sheets, which simultaneously endure shear stresses (see also **Figure 1.5**).

When a mechanical force is applied to a protein, the molecule will align with respect to the vectorial direction of the force. Once the amplitude of the force breaks the mechanical clamp and the protein unfolds, the internal tension along the backbone will decrease, and the now unfolded polypeptide will continue to be elastically stretched. The direction in which the force is applied on the protein also plays an important role, since it affects the distribution of the tension along the protein, particularly when mechanical clamps are involved. In the latter case, the unzipping mechanism will require less force than shearing, hence, the effect of force on the mechanical resilience is non-isotropic^{104, 105}.

One has to consider the effect that the unfolding perturbation has on a protein or to the system the protein is part of. While temperature is relatively isotopically distributed along an ensemble of proteins (or along systems in which proteins and polyproteins are incorporated), chemical agents, such as ion or other desaturating molecules, depend on crowding and consequently on diffusion rates, while mechanical forces, depend on the direction of its propagation and localization effects (acting on specific non-covalent force-withstanding subregions). The folded conformation of the protein is well defined, however its unfolded state is more ambiguous¹⁰⁶: chemical and temperature induced unfolding causes small changes in the overall mean globular size of the protein, which is characterized by a change of few nanometers in its dimensions⁸⁹. In mechanical unfolding, in which the force is applied vectorially, the globular folded protein can extend up to several tens of nanometers along its end-to-end distance, covering up to 80% of its contour length¹⁰⁷. In all, comparison between different unfolding experiments with different denaturing approaches (chemical, thermal, and force) can be misleading, as each perturbation method samples different ensembles of structural-activation energy barriers^{106, 108}, stretched along different [reaction coordinates](#) on a multidimensional energy hyper surface⁹⁷. This means that a protein can have high thermal stability and low mechanical stability at the same time. This issue will be discussed in detail in **Chapter 2**.

That's a Wrap

- Various levels of structural organization in proteins are governed by specific mechanical forces and orientation geometry.
- Protein folding is inextricably related to chain entropy and can be separated from a time and energy perspective as the assembly of the primary structure and its extrusion through the exit channel of the ribosome, entropic collapse, hydrophobic collapse and formation of molten globule, followed by acquisition of the native structure.
- Many proteins operating under force have a rod-like structure and are composed of many domains, similarly to a beads-on-a-string arrangement.
- Unlike thermal and chemical denaturants, mechanical force affects the stability of proteins in a localized way and pulling geometry plays a significant role.

Read These Next

- Sinan Keten and Markus J. Buehler, Geometric Confinement Governs the Rupture Strength of H-bond Assemblies at a Critical Length Scale, Nano Lett. 2008, 8, 2, 743–748.
- Milo M. Lin, Dmitry Shorokhov, and Ahmed H. Zewail, Dominance of misfolded intermediates in the dynamics of α -helix folding, P Natl Acad Sci USA 2014, 111 (40), 14424-14429.
- Bruce Alberts, Molecular Biology of the Cell, Garland Publishing, 1989.

Chapter 2 – Energy landscapes as the protein’s blueprint

2.1 WHAT IS AN ENERGY LANDSCAPE?

Protein functionality is directly related to its folded three-dimensional structure, therefore, understanding the physical aspects of the forces that shape the formation of that native conformation of proteins is of high importance. In particular, the complex folding mechanism was shaped by evolutionary pressure to provide the functional gain of the protein. Transitioning between folded and various unfolded states is accompanied by conformational changes in the protein structural and energetical states. Unlike the defined native state for stable proteins, unfolded states occupy a vast number of conformational possibilities, ranging from partially (intermediates) unfolded states, through [molten-globules](#) to completely unstructured random coiled conformations^{109, 110, 111, 112}.

The folded native state is typically compact and thermodynamically stable in the natural surroundings of the protein, which means that the free energy of this conformation is minimized¹¹³. Acquiring the folded structure involves an “oriented-random” search across the conformational space available to the unfolded protein chain until the proper interactions between specific amino acids come into contact^{113, 114, 115}. This search is done by transient sampling of localized conformational regions from all the accessible states, and eventually leads to the optimal structural and energetic properties that define the native state.

Since conformational space is related with the free energy of the system, a free energy landscape can map all possible states, and therefore conceptually is viewed as a “blueprint” of the protein. The amino-acid composition of the protein determines the contour and possible states and barriers of the free energy landscape, to direct fast and effective folding. This specific sequence naturally determined by the genetics of the organism was evolutionary optimized to minimize energetic competition between various interactions of the different functional amino acid residues¹¹⁶. Accordingly, the thermodynamic depiction of the possible conformational states that the protein can hold can then set the basis for understanding the kinetic mechanisms of the transport between these states^{117, 118}.

Free energy is defined as the amount of available energy to perform work in a [thermodynamic system](#). It is also associated with dynamic [equilibrium](#). A dynamic equilibrium can be described in a reversible reaction in a closed system, in which reactants can turn into products, and products can turn back to become reactants. This reversible equilibrium process is characterized by an equilibrium, or rate-coefficient (see **section 2.2**), that accounts for the ratio of the fluxes between the products and the reactants at equilibrium. The reactants/products can be chemical substances (represented through their concentrations), or populations, defined by some states. The dynamic equilibrium can be changed by performing some work on the system, which will make one of the species favorable, and the system will reach a new dynamical equilibrium, with a new characteristic reaction rate. These species, referred to as reactants/products, can be considered as the folded/unfolded states in our context (or any population of specific structural state, such as

collapsed, misfolded, molten globule, etc.). The rate-coefficient characterizes the transitioning between such states.

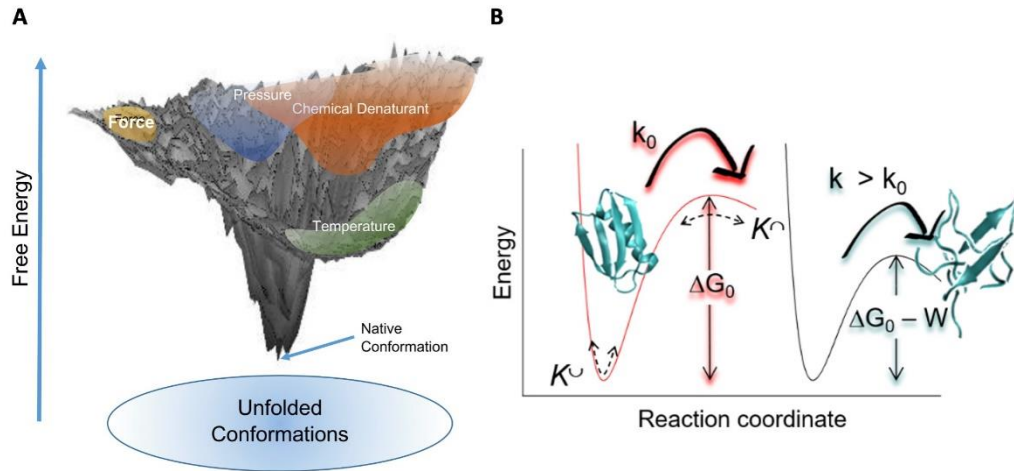
The Gibbs free energy, G , describing the free energy for systems at constant temperature and pressure conditions, is given as a thermal balance between their entropic and enthalpic components (as mentioned in **Chapter 1.5.2**). This classical view is expanded through the incorporation of ideas from polymer physics and statistical thermodynamics¹¹⁹. In the following section we briefly survey basic notions and concepts on free energy, starting from simplistic (yet highly useful!) one-dimensional descriptions of the free energy, which will be followed by incorporation of fundamental aspects of polymer physics contributions, and concluding with the multidimensional expansion of the free energy hypersurface.

2.2 TWO-STATE MODEL OF PROTEIN FOLDING/UNFOLDING

Many small globular proteins fold/unfold by transitioning between their native (folded) state, and an unfolded state in a two-state fashion^{120, 121}. Thermodynamically, the native state is represented by a minimum in the free energy landscape of the protein^{122, 123}. This minimum defines the folded conformation as a stable state that encompasses all the interactions that maintain it. Under the assumption that an activation barrier, ΔE_0 , separates this state from other fully or partially unfolded states, the probability that a protein will undergo spontaneous unfolding is very low (although it exists due to thermal fluctuations).

To actively induce unfolding, it is necessary to overcome this energy barrier using thermal, chemical, or mechanical perturbations (**Figure 2.1A**). The perturbations that overcome the energy barrier correspond to various molecular interactions, and eventually are a product of the effective interactions (bond energies) that hold the folded conformation intact. This can be described using the standard model in biochemistry in which the overall effective interactions scales with the free energy of the bond strength, i.e., $\Delta E_0 = \Delta G_0$. These interactions affect the free energy landscape by performing chemical/mechanical work, W , which reduces the activation barrier separating the native folded state from other unfolded state. These changes, along with the resulting increase in the unfolding rate, are illustrated in **Figure 2.1B**.

Figure 2.1. Schematic representations of protein folded and unfolded states on a free energy scheme.



A. Illustration of the conformational space at different perturbations, where different unfolded ensembles are designated by the nature of the perturbation on different locations (sometimes overlapping) in the energy landscape . Adapted from ref. 124} with permission. B. Unfolding under a perturbation using a single-well free-energy landscape across a reaction coordinate comprised from a single-well, representing the folded state and an activation barrier. The intrinsic unperturbed free-energy landscape (left, red) is characterized by an activation energy barrier, ΔG_0 , and some rate-coefficient k_0 ; The activation energy barrier changes with the application of external perturbation (right, black) through the reduction of the energy barrier by $\Delta G_0 - W$, accompanied by an increase with the unfolding rate-coefficient, k .

Performing external work thus lowers this barrier, and consequently reduces the probability of populating the folded state (and increasing the unfolding rate). Under [quasi-adiabatic approximation](#), a first-order rate equation can be used to describe the probability of remaining in the folded state by its survival probability, $S(t)$:

$$\frac{dS(t)}{dt} = -k(t)S(t) \quad (2.1)$$

where $k(t)$ is the escape (or unfolding) rate. According to this formulation, the probability of refolding (or alternatively, refolding rate) is neglected. The survival probability of remaining folded is given by

$$S(t) = e^{-\int k(t')dt'} \quad (2.2)$$

between t_0 and a given time t (t' is the integration variable). The time-dependent unfolding probability, $p(t)$, which is the probability to unfold at time t , or alternatively, the unfolding time

distribution, is related to the survival probability via $S(t) = 1 - \int p(t') dt'$. Assuming that under the application of a given amount of perturbation, the reduced barrier, $\Delta G(W) = \Delta G_0 - W$, yields unfolding time distribution that is characterized with a single characteristic unfolding rate, k , which leads to

$$p(t) = k e^{-kt} \quad (2.3)$$

The unfolding rate itself can be conveniently described within the framework of the two-state model as a diffusional process across a one-dimensional energy barrier using the [Kramers theory](#) under high-barrier approximation^{97, 125, 126}:

$$k = k_0 e^{-\frac{\Delta G}{k_B T}} \quad ; k_0 = \frac{D \sqrt{K^U K^N}}{2\pi k_B T} \quad (2.4)$$

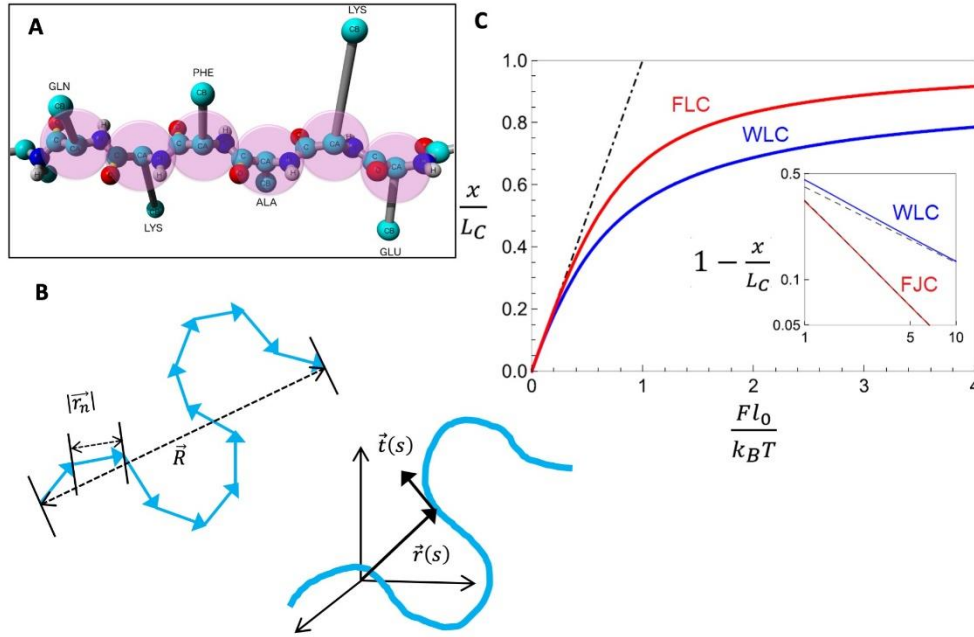
where k_0 is the attempt frequency for crossing the unperturbed barrier, k_B is *Boltzmann's* constant, and T is the absolute temperature. If ΔG varies with time, for instance under a time dependent force (see **Chapter 3.2**) then $k = k(t)$ as well. D is the diffusion coefficient, accounting for the dynamics across the energy landscape, K^E and K^G are the curvatures of the energy landscape at the minimum of the (folded state) well, and at the maximum of the activation barrier respectively, and are given by the second derivatives of the energy at these given positions (illustrated in **Figure 2.1B**).

The one-dimensional two-state model predicts an exponential dependency or [Poisson distribution](#) of the unfolding probability with time, which makes it appealing for the interpretation of experimental folding/unfolding measurements, by capturing minimal routes^{127, 128, 129}.

2.3 PROTEINS AND POLYMERS

Proteins are initially synthesized at the ribosome as polypeptides that fold into their native state, from which they acquire their functional form¹³⁰. The polypeptide is a hetero-polymeric chain comprised of a defined sequence of amino acids, which defines the primary structure of the protein. This means that in their foundation, proteins are polymeric materials, whose monomeric units are amino acids (**Figure 2.2A**).

Figure 2.2. Polymeric description of polypeptide chain.



A. Illustration of a polypeptide chain, where each monomeric amino acid is marked with a pink circle. Taken from ref. 131 with permission. B. Schematic representation of discrete (left) and continuous (right) WLC polymer models. C. Force extension scheme of the FJC and WLC models. The scaling length $l_0 = \frac{1}{3}b$ for FJC and $\frac{2}{3}l_p$ for WLC were chosen to create a unit initial slope (dot-dashed line). The inset shows on a log-log plot the deviation between the models. Taken from ref. 132 with permission.

Biopolymers display substantial elasticity, particularly when they are incorporated in biological systems whose operation depends on this property, as in cellular signaling and in tissue elasticity. The statistical approach to polymer physics is based on the parallel view of a polymeric chain to **random walk** schemes. This framework enables the formulation of the most simple polymer models. Implementing simple diffusional concepts, as we will soon see, polymer chain models can describe fundamental chain properties, such as their end-to-end distances (and their distributions), from their physical properties^{133, 134, 135}. According to this approach, we will consider a polymer chain as made of a sequence of rigid segments, connected by unrestricted hinges.

Flexible and semi-flexible polymers have often been characterized by models such as the Freely Jointed Chain (FJC) and/or the Worm-Like Chain (WLC) models. Generally, a polymer can be modeled by a set of N consecutive $\vec{r}_n$ segments with length $|\vec{r}_n| = b_0$ (**Figure 2.2B**, left). The size of such chain can be described by several length parameters: The nominal length is defined by the contour length, $L_C = \sum_{n=1}^N b_0 = N b_0$, its end-to-end length $\vec{R} = \vec{R}_N - \vec{R}_1 = \sum_{n=1}^N \vec{r}_i$, and its characteristic ideal random coil size is given by the root mean square displacement of the chain, i.e., $R_0 = \langle \vec{R}^2 \rangle^{1/2} = N^{1/2} b_0$.

The ideal chain (FJC) modeling assumes that there are no correlations along the chain, and every macromolecular conformation has equal probability. The final volume of the monomers is not considered, and also polymer-solvent interactions are excluded. In principle, the mean square displacement of the chain can be calculated as

$$\langle \vec{R}^2 \rangle = \langle (\sum_{n=1}^N \vec{r}_n) (\sum_{m=1}^N \vec{r}_m) \rangle = \sum_{n=1}^N \sum_{m=1}^N \langle \vec{r}_n \cdot \vec{r}_m \rangle \quad (2.5)$$

where n and m represent various locations along the contour of the chain ($n > m$), which makes $\langle \vec{r}_n \cdot \vec{r}_m \rangle$ an autocorrelation function that sums over all the contributions arising from the position of the segments. One can consider it as a relation, or “memory” along the segments: is $\vec{r}_n$ affected by $\vec{r}_m$. From symmetry properties, **equation 2.5** can be rewritten:

$$\langle \vec{R}^2 \rangle = \sum_{n=1}^N \langle \vec{r}_n^2 \rangle + 2 \sum_{n>m}^N \langle \vec{r}_n \cdot \vec{r}_m \rangle = N b_0^2 + 2 \sum_{n>m}^N \langle \vec{r}_n \cdot \vec{r}_m \rangle \quad (2.6)$$

For $n \neq m$, $\langle \vec{r}_n \cdot \vec{r}_m \rangle = \langle \vec{r}_n \rangle \langle \vec{r}_m \rangle = 0 \cdot 0 = 0$, hence there is no correlation along the chain vectors, which means that $\langle \vec{R}^2 \rangle = N b_0^2 = R_0^2$ (R_0^2 scales with N , in accord to a Gaussian decay, as in Brownian motion).

Before going to the continuous (WLC) description, let us consider a situation in which there is a constant angle q between every consecutive segment, i.e. $\langle \vec{r}_n \rangle |_{\vec{r}_m, \vec{r}_{m+1}, \dots, \vec{r}_{n-1} = \text{fixed}} = \vec{r}_{n-1} \cos(\theta)$, yet each segment can rotate freely around its axis. This model is called the Freely-Rotating-Chain (FRC). This creates a correlation along the chain, which can be written as:

$$\langle \vec{r}_n \cdot \vec{r}_m \rangle = b_0^2 \cos(\theta)^{|n-m|} = b_0^2 e^{-\frac{|n-m| b_0}{l_p}} \quad (2.7)$$

This means that the correlation decays exponentially along the chain with respect to a new length scale, l_p , termed the persistence length, and is defined from **equation 2.7**:

$$l_p = \frac{b_0}{-\ln[\cos(\theta)]} \quad (2.8)$$

It provides a measure to the bending stiffness of the chain, by defining a length at which the correlation progresses along the chain. For “soft” biomolecules, such as polypeptides, l_p is about the size of the monomeric unit (amino acids, i.e., $l_p \sim 0.4 \text{ nm}$ ¹³⁶), while for “stiff” biomolecules, such as DNA strands, it can reach their contour length, meaning tens of nanometers¹³⁷.

To formulate the continuous (WLC) model, the segments' length and bending have to approach their limits ($b_0 \rightarrow 0$, and $L_C \gg b_0$). Considering a position s , and the corresponding tangent vector $\vec{t}(s)$ along the chain (**Figure 2.2B**, right), **equation 2.7** can be written as:

$$\langle \vec{t}(s) \cdot \vec{t}(s') \rangle = b_0^2 e^{-\frac{|s-s'|b_0}{l_p}} \quad (2.9)$$

From this expression the length of the chain can be calculated using, **equation 2.6**¹³⁸:

$$\langle \vec{R}^2 \rangle_{WLC} = \int_0^{L_C} \left[\int_0^{L_C} \langle \vec{t}(s) \cdot \vec{t}(s') \rangle ds' \right] ds = 2l_p L_C - 2l_p^2 \left(1 - e^{-\frac{L_C}{l_p}} \right) \quad (2.10)$$

This expression has a wide spectrum of elastic behavior, ranging from the ideal chain limit ($L_C \gg l_p$): $\langle \vec{R}^2 \rangle \sim 2l_p L_C \sim N \tilde{b}_0^2$ (where $\tilde{b}_0^2 \equiv 2b_0 l_p$), to the rod limit ($L_C \leq l_p$): $\langle \vec{R}^2 \rangle \sim L_C^2$.

We now revisit the two elastic models, the FJC and the WLC in the situations of high extensions, under the application of external forces. We start with the FJC model. As long as the polymeric chain is very long with respect to the scale of the short-range interactions (correlations), it can be conceptually divided into N "FJC sub-chains" with length b , termed the Kuhn length:

$$b \sim \frac{\langle R^2 \rangle}{L_C} \sim 2l_p \quad (2.11)$$

This means that in the FJC description, the Kuhn length provides a measure for the chain stiffness, as the persistence length in the WLC model. From the calculation of the free energy of the FJC, its end-to-end position under the application of external force F can be derived¹³⁹:

$$\langle R \rangle_F = x_{FJC} = bN \left[\coth \left(\frac{Fb}{k_B T} \right) - \frac{1}{Fb/k_B T} \right] \quad (2.12)$$

The expression in the square brackets is called the Langevin function: $\mathcal{L}(\beta) = \coth(\beta) - 1/\beta$.

For the WLC there is no simple explicit analytical solution for the chain elongation, however it was approximated with an interpolation formula¹⁴⁰, which is frequently used:

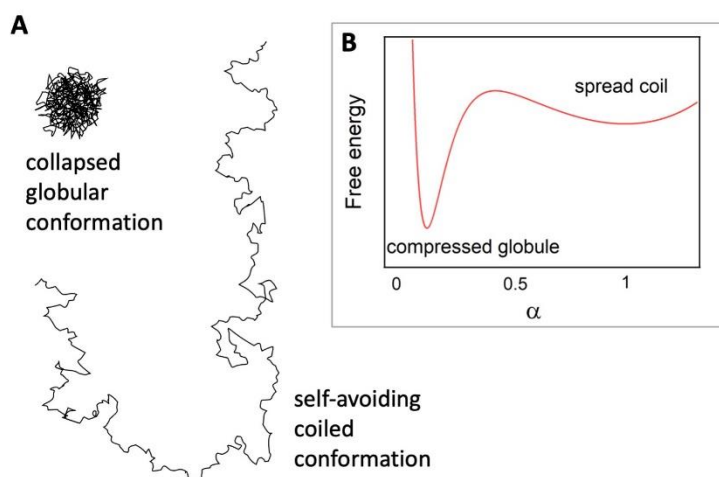
$$F_{WLC} \sim \frac{k_B T}{l_p} \left[\frac{1}{4} \left(1 - \frac{x}{L_C} \right)^{-2} - \frac{1}{4} + \frac{x}{L_C} \right] \quad (2.13)$$

In both **equations 2.12** and **2.13**, x denotes the end-to-end length of the chain. The differences between these models are illustrated in **Figure 2.2C**.

More elaborate polymer models that extend the FJC and WLC have been developed since the basic models were not sufficient to describe high stretching behaviors observed experimentally^{141, 142, 143}.

Polymer conformation strongly depends on its surrounding media (solvent) and on the temperature. Repulsive interactions between monomeric units of the polymer can lead to a coiled conformation. In an ideal solvent, entropy drives the chain collapse. When attractive interactions prevail, the chain assumes a compact globular conformation. These interactions are used within the definition of the solvent quality as good or bad solvent, where good solvent promotes the extended coil conformation, and bad solvent stabilizes the compact globule conformation. If the monomeric interactions become sufficiently strong, the polymer undergoes a phase transition (equivalent to condensation) in which the expanded coil transitions into a dense globular blob. This transition is the coil-to-globule transition (**Figure 2.3A**).

Figure 2.3. Polymeric coil-to-globule transition.



A. Coiled and globule polymeric conformations. B. Free energy landscape of the coil-to-globule phase transition as a function of the swelling parameter, α .

To understand monomeric interactions, we need to define the solvent quality terminology, and particularly θ -conditions. Generally, low temperatures induce attraction, while higher temperatures lead to repulsive interactions between the segments of the polymer. These interactions are usefully characterized with the [Virial theory](#). According to this theory, the energy, U , can be expanded as a series of powers in the segment density, r : $U = Vk_B T(Br^2 + Cr^3 + \dots)$ ¹⁴⁴. V is the volume of the polymer, B and C are the second and third virial coefficients that depend

on the temperature and the interaction energy between the segments. The second virial coefficient, B , is proportional to the square of the density, which relates it to the binary pair collision probability, and C in a similar manner is associated with three-body interactions.

At low temperatures, B (and the internal energy of the polymer coil, U) is negative, which means that attractive interactions become dominant. These conditions describe bad solvent, in which we observe precipitation of the polymer globules. At high temperatures, B (and U) is positive, promoting repulsive interactions along the segments of the polymer chain, and consequently stabilizing its coiled conformation. Here the solvent is referred to as good solvent, and the coiled polymer chains are better dissolved. This means that the definition of the solvent quality is therefore affected by the temperature.

The transition between the two states occurs when $B = 0$. At this point, the temperature is called the θ -temperature (or the θ -point), and the polymer becomes ideal. At this temperature, repulsive interactions due to excluded volume cancel attractive interactions due to van der Waals attractive effects. Accordingly, good solvent is defined at $T > \theta$, and bad solvent at $T < \theta$. It should be noted that the temperature is not the sole component, and other parameters, such as the solvent composition (ions, sugars etc.) and their concentrations also affect the coil-to-globule transition⁸³. Additionally, the application of mechanical forces provokes a similar coil-to-globule transition along a single chain, between a collapsed and extended states^{145, 146}.

The free energy of the coil-to-globule transition is typically described as a function of the swelling coefficient, $\alpha = R/R_0$, when $\alpha > 1$ alludes to a coiled (swollen) conformation, and $\alpha < 1$ to a compact dense globule. The free energy is comprised of entropic contributions, related to the conformation of the chain, and from the energy of the monomeric interactions along the chain (that invoke the use of the second virial coefficient). **Figure 2.3B** shows the free energy of the coil-to-globule transition as a function of the swelling parameter, α .

Proteins are in fact biopolymers that for most cases reside in globular conformation in their natural surroundings. And as will be discussed in detail in **Chapter 3**, the coil to globule phase transition is an integrative part in the folding process from high extensions of the unfolded protein chain. Moreover, the polymeric nature of the folded and unfolded states manifests itself in an extremely large number of possible conformational states that manifest themselves in the complex dynamics that proteins exhibit. Consequently, this has a dramatic effect on the energy landscape of the proteins and makes it more complicated than the basic two-state model described above.

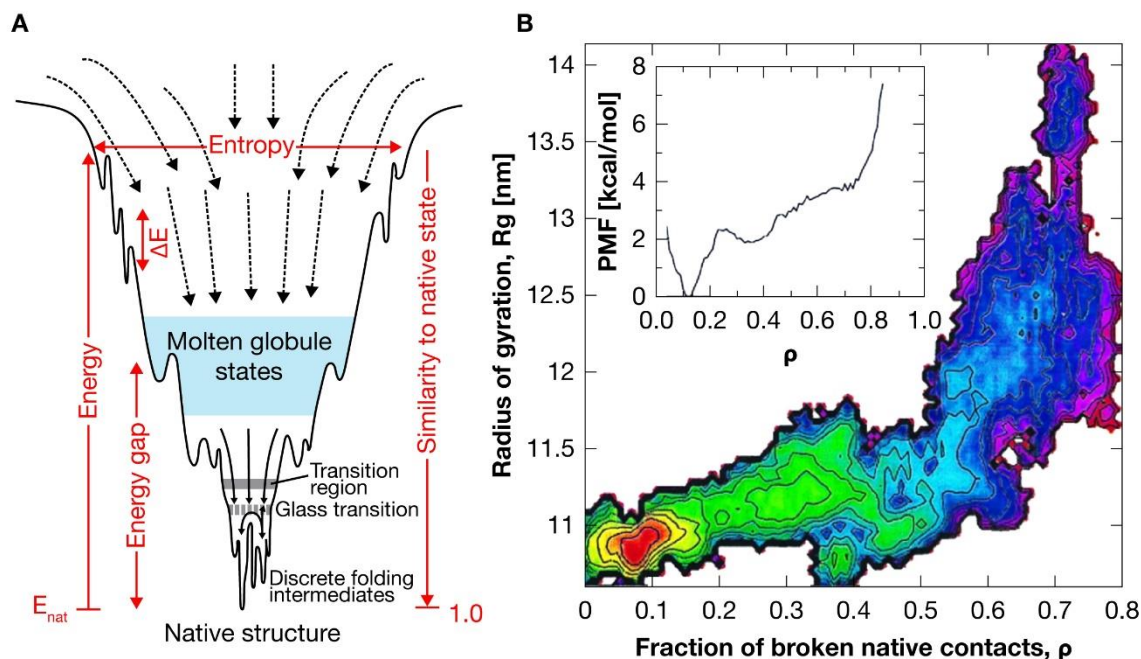
2.4 MULTI-DIMENSIONALITY

Folding of globular proteins can be viewed as a thermodynamic process, where the native structure typically corresponds to the lowest Gibbs energy minimum in the conformational space of the polypeptide chain. The complexity of this process gave rise to new theoretic approaches that go beyond the simple two-state model. According to these theories, protein folding is energetically driven by a downhill descent along a multidimensional free energy landscape to its

most stable (low energy) native state, interweaving entropy and enthalpy from both thermodynamic and kinetic aspects^{97, 116, 147}.

While the folded (native) state of the protein is characterized by low minima in the energy landscape, the generally termed unfolded state, the unfolding is actually represented by more than a single state, with many possible structural conformations¹⁴⁸. These states can be represented through the multidimensional free energy hypersurface of the protein^{97, 118, 147, 149} that emerges from numerous degrees of freedom of the protein conformations (associated with its amino acid sequence, temperature, and solvent interactions). This energetic-conformational space provides the possibility for various folding/unfolding pathways. Therefore, a funneled¹¹⁶ (or hypergutter¹⁵⁰) orientation of the free energy landscape assures minimal frustration, since it quickly directs the folding process down to the folded conformation at the basin of the energy landscape (**Figure 2.4A**). The presence of local maxima and minima along the energetic hypersurface form roughness that eventually reduce the conformational space to lower dimensions¹¹⁴.

Figure 2.4. Multidimensionality of the energy surface.



A. Schematics of the free energy funnel directing the downhill folding with respect to its various energetic and conformational states (taken from ref. 151). B. 2D projection of the free energy of GB1 protein on two reaction coordinates: radius of gyration (R_g) and fraction of broken native contacts (ρ), calculated from MD simulations in explicit solvent. The inset shows the 1D PMF (projection) over the ρ coordinate. Adapted from ref. 152} with permission.

Several interactions are involved in the folding/unfolding processes, such as the number of **native contacts** that form/rupture during conformational transitions, presence of water molecules that are being excluded or that enter the hydrophobic core of the folded protein, and the possible number of conformational states (reflected through the size of the folded/unfolded protein). These and other interactions that were not listed can serve as reaction coordinates for the energy landscape. The projection of the energy hypersurface of the protein over any of these reaction coordinates can be described by its **potential of mean force (PMF)** profile^{97, 115, 150}, which is sometimes referred to as the free energy landscape, although these terms have different meanings¹⁵³.

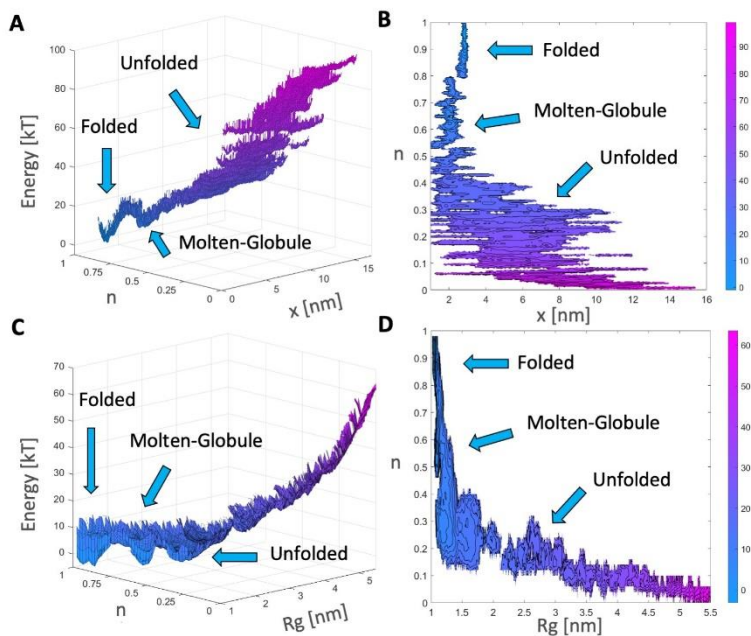
In this section, we use the example of the globular GB1 protein as a model system to demonstrate the aspects of various degrees of freedom at different reaction coordinates of the free energy hypersurface. This B1 segment of Streptococcal protein G is a small (56 residues) protein, whose native conformation consists of an α -helix and four-stranded β -sheet¹⁵⁴ and has been the subject of folding studies with respect to its free energy landscape^{152, 155, 156, 157, 158, 159}.

Figure 2.4B shows the calculated 2D free energy projection over two specific reaction coordinates: The **radius of gyration** of the protein, R_g , and, the fraction of broken native contacts, ρ , (that estimates how close the protein is to its native state) all calculated from **molecular dynamics (MD) simulations** that were performed in explicit solvent¹⁵². It can be seen from this 2D energy surface that folding begins with initial collapse, where R_g decreases with no substantial growth of native contacts: $R_g \sim 1.2$ nm while $\sim 45\%$ of contacts have formed, then it changes very little as the number of native contact increases (number of native contacts gets to $\sim 90\%$ contacts, while R_g decreases by only ~ 0.15 nm), during the folding to the native state. This behavior forms an inverse downhill L-shaped energy surface, from high energy (purple and dark blue colored) to the lowest conformation of the native state (yellow and red colored). The Figure 2.3B inset shows the 1D PMF projection on the ρ coordinate, which displays a gradually descending energy profile with a distinct folded, partially unfolded/intermediate/molten-globule, and unfolded states (at $\rho = \sim 0.1$, ~ 0.35 , and ~ 0.7 respectively).

As mentioned, the MD simulations from which the energy surface and profile were calculated (**Figure 2.4B**) were performed in explicit solvent conditions. The meticulous attention to the presence of solvent molecules is of great importance since water plays an important role in the folding and unfolding process. The differences between good and bad solvents become apparent here (see **Section 2.3**). During folding, hydrophobic interactions between specific residues increase, and water molecules are excluded from the emerging hydrophobic core of the protein. At the same time, hydrophobic amino acid residues are exposed to the surroundings in the exterior surface of the folded protein (see also the discussion on hydrophobic collapse in **Chapter 1.3.3**). These internal hydrophobic and external hydrophilic effects stabilize the conformation of the native state. The balance between these interactions during the folding process gives rise to the formation of metastable collapsed/molten-globule states^{111, 112, 160}, which follow the polypeptide chain collapse¹⁰⁹ and the coil-to-globule transition¹¹⁰. The molten-globule state is formed between the native (folded) and the completely unfolded states and is responsible for the (here inversed) L-shape of the energy surface. In this case, the folding of GB1 would be a three-state process over the ρ reaction coordinate^{151, 152}.

In the example above, the free energy hypersurface was represented by two specific coordinates. We now examine the representation of the free energy landscape of the same GB1 protein over different reaction coordinates, and how their combination affects the general morphology. The following energy surfaces were also obtained from MD simulations in explicit solvent¹⁶¹. **Figure 2.5A** show the energy surface of GB1 over two reaction coordinates, n , representing the normalized number of native contacts (unlike ρ in **Figure 2.4**, that was defined as the fraction of broken native contacts), and x , the end-to-end length of the chain. **Figure 2.5B** shows the PMF surface over n , and R_g . **Figures 2.5C** and **2.4D** show the 2D projections of the 3D energy surfaces shown in **Figures 2.4A** and **2.5B**, respectively. Notice that due to the different definition of the number of contacts (n instead of ρ), the projected PMF displays L-shape (opposing to the inversed L-shape shown in **Figure 2.4C**). This means that regardless of the choice of the reaction coordinates here, the three-state architecture is preserved. Additionally, note that in this representation there is a difference only in a single reaction coordinate (x , and R_g), yet the general shape of the energy landscapes changes.

Figure 2.5. Free energy surfaces of protein GB1 from MD simulations in explicit solvent.



A. Energy surface as a function of the normalized number of native contacts, n , and the end-to-end length, x , coordinates. B. Energy surface as a function of n , and the radius of gyration, R_g , coordinates. C. 2D projection of the energy surface shown in A. D. 2D projection of the energy surface shown in C. Adapted from ref. 161} with permission.

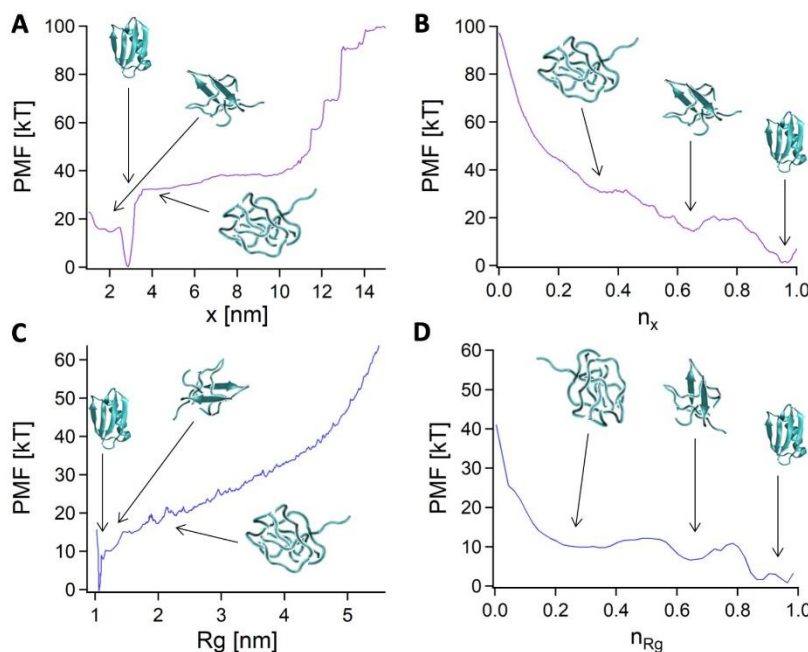
For a better observation of the detail of the PMFs along the different reaction coordinates and their interactions, one can observe the projections of the PMF on each coordinate. The reduction

in dimensionality occurs as other conformations along the z -coordinate are being averaged out in the following procedure, for example for an arbitrary coordinate y :

$$PMF(y) = -k_B T \ln \left[\int e^{-PMF(y,z)/k_B T} dz \right] \quad (2.14)$$

Here y and z stand for the pair of coordinates n , x and Rg . **Figures 2.6A** and **2.6B** show the minimal path PMF profiles along the x and n coordinates from the energy surface shown in **Figure 2.5A**. **Figures 2.6C** and **2.6D** show the minimal path PMF profiles along the Rg and n coordinates from the energy surface shown in **Figure 2.5B**. Although in both representations of the energy surfaces the same n coordinate is calculated, the energy surface projection on it combined with a different reaction coordinate, x or Rg , alters the morphology along the n coordinate. To this end we denote them in **Figures 2.6B** and **2.6D** as n_x and n_{Rg} with respect to the energy surface they are obtained from.

Figure 2.6. Minimal path profiles from unperturbed energy surfaces in Figure 2.5.



1D PMF projections onto the x (A) and n_x (B) coordinates (from the x - n energy surface), and 1D PMF projections onto the Rg (C) and n_{Rg} (D) coordinates (from the Rg - n energy surface). Adapted from ref. 161} with permission.

The 1D representation of the various minimal path PMFs demonstrates the rearrangement of the energy surface when projected on two different conformational spaces (on the x - n and Rg - n planes) and helps visualize how each parameter changes during protein folding. While each length coordinate displays a single defined folded state, the two contact coordinates show the three thermodynamic states that were displayed above. These three states are more pronounced along the arrangement of n_{Rg} than on n_x . Yet these three states are distinguishable, and as such provide an efficient reaction coordinate (as r in **Figure 2.4B**). Due to the funneled L-shaped form of the energy surfaces, the three states observed on n collapse within a narrow range on the x and Rg coordinates. Consequently, these states, particularly the folded and the partially folded/unfolded molten-globule, are indistinguishable and appear as a single energy minimum on the x/Rg minimal 1D PMF projection.

These representations exemplify that in order to adequately understand the folding/unfolding pathways it is necessary to consider the reaction coordinate that captures it. For instance, as we will discuss in **Chapter 3**, application of force to a protein is conveniently monitored with the end-to-end length coordinate using single molecule techniques. This means that a 1D PMF(x) is probed, and the mechanical work performed through the application of force tilts this PMF profile and creates a new unfolded state, leading to the formation of a two-state PMF. Therefore, acknowledging the possibility that several thermodynamic states can be “buried” along the probed reaction coordinate can prevent misinterpretation of the data^{161, 162, 163, 164}. In such a situation, complementing information (and measurements) can be helpful.

To summarize, the funneled free energy hypersurface provides a highly useful framework in the study of processes associated with different states of proteins and transitioning between them, such as folding/unfolding, misfolding, aggregation and biomolecular recognition, and disordered proteins^{130, 165, 166, 167, 168, 169}. Although the free energy hypersurface provides a conceptual approach that offers insights into the thermodynamics and kinetics of protein functional mechanisms, the use of the two-state is a useful tool to obtain direct information on specific transitioning between conformational states.

That's a Wrap

- A protein has a well-defined folded state, but many unfolded states.
- A mechanically unfolded protein extends in the direction of the force vector, and its extension depends mainly on its number of amino acids and the magnitude of the force.
- Polymer elasticity models can be used to describe the extension of unfolded proteins under a force vector.
- The energy landscape of a protein can be a good representation for the probability of different states and the transition between them.
- The energy landscape of a protein under force shows on the pulling coordinate a distinct minimum between an unfolded-extended state and a folding state, but cannot distinguish between native and molten globule states, due to their similar size.

- When projected on the number of contacts coordinate, the energy landscape however shows distinct minima between folded, molten globule and unfolded states.

Read These Next

Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D., Navigating the folding routes. *Science* **1995**, 267 (5204), 1619-1620. Zwanzig, R., Two-state models of protein folding kinetics. *P Natl Acad Sci USA* **1997**, 94 (1), 148-150.

Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G., Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins*. **1995**, 21 (3), 167-195.

Doi, M.; Edwards, S. F.; Edwards, S. F., *The theory of polymer dynamics*. Clarendon Press: 1988; Vol. 73, p 408.

Khokhlov, A. R. G., A. Y.; Pande, V. S., *Statistical Physics of Macromolecules*. American Institute of Physics Melville, NY: 1994; p XXVIII, 350.

Berkovich, R.; Garcia-Manyes, S.; Urbakh, M.; Klafter, J.; Fernandez, J. M., Collapse Dynamics of Single Proteins Extended by Force. *Biophys.J.* **2010**, 98 (11), 2692-2701.

Chapter 3 – Protein functioning under force

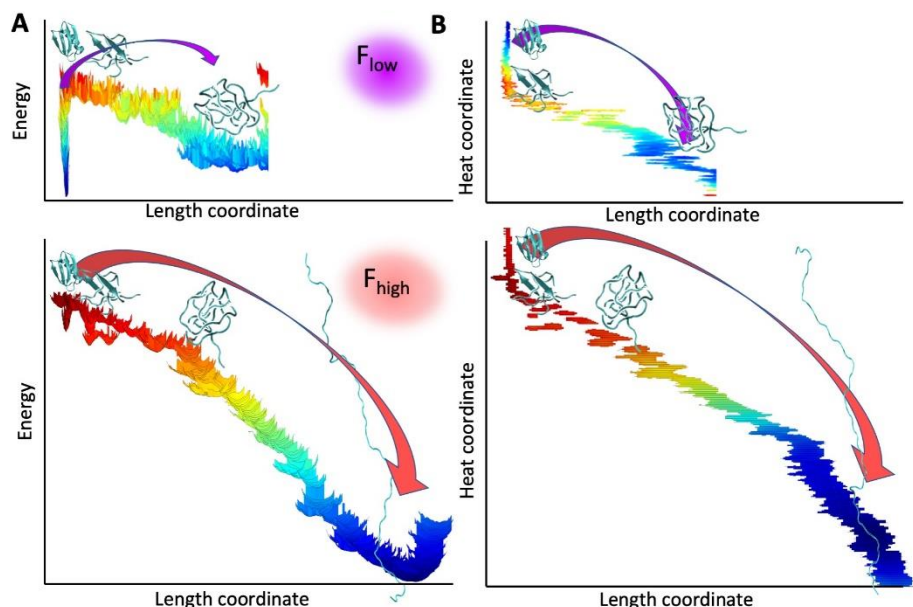
3.1 INTRODUCTION: PROTEINS AND MECHANICAL FORCES

During their physiological function, some proteins are exposed to mechanical forces, with specific examples discussed in **Chapter 5**. These forces may induce conformational changes of the protein structure^{65, 170, 171}. While some physiological functions require that the proteins will maintain their native folded state, other processes rely on the unfolding and extension of the peptide chain under force. For example, a protein has to unfold in order to be translocated across a membrane during import into mitochondria, chloroplasts or the endoplasmic reticulum¹⁷². Additionally, unfolding and refolding under force was shown to be a mechanism by which titin protein stores and delivers energy during muscle contraction^{48, 173}.

Conventional studies of protein folding/unfolding under chemical or thermal denaturing conditions are performed in bulk (ensemble) experiments. However, these ensemble-based approaches average over many proteins with various conformations and are limited by their ability to capture the conformational heterogeneity of the individual protein molecules. Consequently, it is difficult to distinguish amid different folding/unfolding pathways. Single molecule techniques, on the other hand, enable studying one molecule at a time by exploring its energetically metastable behavior, and probing both static and dynamic heterogeneity in the intermediate states among individual proteins^{174, 175, 176}. In particular, single molecule data can follow time evolving trajectories of protein folding and unfolding, as well as chemical reactions that are hard (or impossible) to synchronize at the ensemble level¹⁷⁷. The replacement of bulk ensemble averages with time [probability density functions \(PDFs\)](#) can expand the amount of information available regarding the construction of propagators, kinetic schemes, potentials and so on¹⁷⁸. The use of [single molecule force spectroscopy \(SMFS\)](#) techniques is also appealing, since the kinetics of the protein and its thermodynamic properties can be monitored on a single, well-defined, reaction coordinate – its end-to-end length. These techniques are presented and discussed in **Chapter 4**.

This chapter discusses the implications of protein performance under mechanical forces. Force as an external stimulus can induce transitions between conformational states such as the native (folded) state, intermediate (partially unfolded) states, disordered (unfolded coiled or globular) states^{32, 37, 89, 175, 179, 180}. In the previous chapter we discussed the multidimensional nature of the unperturbed energy landscape of simple globular protein, and its projection over different reaction coordinates. Here we delve into the specific effect forces have on the protein as it unfolds and refolds. **Figure 3.1** demonstrates the effect of mechanical load on the free energy landscape of GB1 protein, calculated from MD simulations¹⁶¹. When a protein experiences mechanical force, interactions that stabilize its native conformation are overcome, leading to unfolding. This process can be represented by crossing an energy barrier, which is reduced by the work applied by the applied load with respect to the surrounding thermal environment. Here it is important to take into account that not only the amplitude of the force affects the unfolding/refolding processes, but also the way in which the force is applied. The impact of the applied force also depends on the rate in which it is employed^{181, 182, 183}, the polymeric properties of unstructured components^{42, 184, 185, 186}, and the direction along which the force is applied^{2, 187}.

Figure 3.1. The effect of force on the energy landscape of protein GB1.



Trend of protein conformations along its free energy landscape under the application of low (upper panels) and high (lower panels) forces represented on one (A) and two (B) dimensions. The colors indicate increasing energy, ranging from red (high) to blue (low). Notice how the force expands the energy landscape, as more unfolded extended states are explored as the protein stretches (adapted from ref. 161).

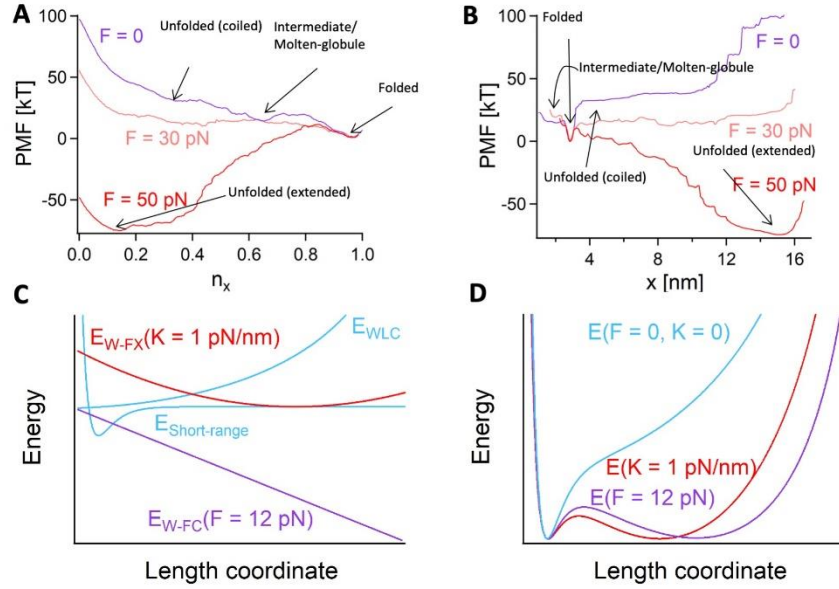
Figure 3.1A shows how the energy landscape is being distorted along the extension reaction coordinate (here, its end-to-end length) under low and high forces (upper and lower panels, respectively), while **Figure 3.1B** shows the same energy landscapes, as energy surfaces along the length and heat (represented here by the normalized number of native contacts) coordinates. As can be seen, when the magnitude of the force increases, the energy landscape expands as the unfolded proteins explore more conformations. Eventually the polymeric nature of the unfolded protein will enable its extension based on the extent of the applied force – while high forces maintain highly extended state, low forces enable sampling coiled and intermediate configurations. Therefore, mechanical unfolding under high loads is immediate, (even before stretching upon the applied load begins) and short-lived, which means that it is not always observable on the length coordinate.

3.2 UNFOLDING UNDER FORCE

As generally introduced in the previous chapters, thermodynamics and kinetics provide a valuable means to understand protein folding/unfolding on the molecular level. The one-dimensional (1D) reduction of the protein's energy surface provides a highly useful framework to

describe and study the dynamic transitions between the conformational states of the protein under force^{123, 127, 129, 146, 161, 171, 180, 188, 189, 190, 191}. We examine the projection of the 2D energy landscapes of GB1 over their 1D minimal path potentials of mean force (PMFs)^{68, 161}. **Figures 3.2A** and **3.2B** show the 1D PMFs (from **Figure 3.1**) along the n_x (fraction of native contacts) and x (end-to-end distance) coordinates, respectively, for an unperturbed protein, and for a protein unfolding under a constant force of 30, and 50 pN. In the absence of force (purple), the free energy of the native state is the lowest. The application of an external force (red shades) lowers the barriers to the unfolded states and increases the probability for transitioning between other conformational states. While only a single barrier is observed on the x coordinate, the 1D projection onto the n_x coordinate indicates the existence of the three states: folded, intermediate/molten-globule and unfolded (coiled under zero/low loads, and extended at high loads) respectively.

Figure 3.2. Illustration of the effect of force on the 1D projection of the energy landscape of a protein.



Projection on the minimal path PMFs over the fraction of native contacts, n_x (A) and over the end-to-end length, x (B) at different loads (Adapted from ref. 161). C. Schematics of the short-range interactions and WLC polymeric conformation components of the energy contour projection (light blue) under the mechanical work performed at constant force, E_{W-FC} (purple) and under force extension, E_{W-FX} (red). D. Model of the unperturbed protein free energy along its extension coordinate (light blue), extended under a direct force application (purple), and under a time varying elastic potential (red) (Parameters taken from refs. 146 and 163).

When external load is applied to the protein, its native contacts break as it unfolds through intermediate states into a disordered polypeptide chain, which continues to stretch^{136, 181}. The

application of force (in various manners and to various extents) affects the activation barriers that separate the states, and driving the conformational changes via different pathways, which are sometimes not apparent on the 1D projection of the protein energy surface^{68, 161}. This can be observed in **Figures 3.1** and **3.2**, where the first barrier from the native state explicitly appears on the n_x coordinate, while the two minima this barrier separates overlap, and are hard to distinguish along the single x coordinate. It can also be seen how the experienced force not only expands the PMFs, but also performs mechanical work ($\int F dx$) that tilts the barriers and conformational states (shifting the local maxima and minima along the contour of the PMFs). This effect, in turn, shifts the probabilities of occupying the various (compact and stretched) unfolded states that sample equilibrium conformations under the given temporal conditions.

The adjacency of the folded states on the x projected PMF that overlap within less than ~2 nm can be modeled as short-range interactions (with [Lenard-Jones](#) or [Morse potentials](#), for example) followed by the high energy conformational space due to the polymeric nature of the unfolded chain. While the coiled conformations of the unfolded protein are best modeled by random coiled, or Gaussian chain, the stretched conformations can be described using the Freely-Jointed-Chain (FJC), or the Worm-Like-Chain (WLC) models, as described in **Section 2.3**. The modeling of the unperturbed PMF components over the x -coordinate, given by Eq. (3.1) are shown in **Figure 3.2C**, and their joint PMF in **Figure 3.2D** in light blue. This potential is produced by combining the Morse potential for the short-range interactions, and the WLC potential for the polymeric extensibility of the unfolded protein:

$$E_0(x) = E_{Short-range}(x) + E_{WLC}(x) \quad (3.1)$$

$$= U_0 \left\{ \left[1 - e^{-2\frac{b}{R_c}(x-R_c)} \right]^2 - 1 \right\} + \frac{k_B T}{l_p} \left\{ \frac{L_c}{4} \left[\left(1 - \frac{x}{L_c} \right)^{-1} - 1 \right] - \frac{x}{4} + \frac{x^2}{2L_c} \right\}$$

Here, U_0 , R_c , and b are the parameters of the Morse potential that respectively define the amplitude of the potential (height of the energy barrier), the position of the potential minimum (size of the folded protein), and its spread (curvature). Beyond the well, as illustrated in **Figure 3.2C**, the short-range interaction vanishes. The WLC potential is characterized with respect to the thermal energy, given by $k_B T$, where k_B is Boltzmann's constant, and T is the absolute temperature. The WLC parameters are l_p , the persistence length, and L_c , the contour length (see chapter 2.3). The overall unperturbed potential given by Eq. (3.1) in **Figure 3.2D** (light blue) resembles the unperturbed PMF in **Figure 3.2B** (purple).

The perturbations are presented in **Figure 3.2C** using two form of mechanical work: [constant force \(FC\)](#), purple, and [force-extension \(FX\)](#), red. The constant force potential is given by $E_{W-FC} = Fx$, where F is the applied load, and x is the end-to-end distance coordinate of the molecule. Its combination with the protein unperturbed potential is plotted in purple in **Figure 3.2D**. As can be seen, the application of mechanical work tilts the potential, and forms a new state at high extension, as the unfolded protein chain is stretched. This state will be further deepened and extended as the applied force will increase. It should be noted that the extent of the effect of this force will depend on its proportion to the other potential parameters. At high forces, if the

application of a direct force is immediate (if the pulling velocity is very fast), the perturbation on the free energy will be sharp, and the unfolding of the protein will accordingly be dominated by the highest barrier, which will govern the unfolding rate). Here the force amplitude, F , representing the structural interference agent, is the perturbation parameter.

For moderate force transmission that could range from low to high values, in which the load is applied at constant velocity, v , the force extension potential can be approximated as an elastic spring, $E_{W-FX}(x,t) = \frac{1}{2}K(x - x_0)^2$, where K is the stiffness (spring constant) of the protein and the interface that conveys the load, and $x_0 = vt$ is the position in which the stretching is arrested, where t is time.^{146, 163} The PMF at $K = 1$ pN/nm, and $x_0 = 21$ nm is plotted in red in **Figure 3.2D**. Here the stiffness, K , and the velocity, v (or alternatively, x_0), are the perturbation parameters. Notice that unlike the FC potential, which is only a function of its position, the FX potential varies with both position and time.

In unperturbed systems, unfolding and refolding occur due to thermal fluctuations. Unfolding is a fast process, however force application in natural environments is transmitted by surfaces/molecules to which the unfolding proteins are tethered to, which slows down the forced unfolding¹⁹². Therefore, the force experienced by the protein is considered as stationary over the time scale of the fastest possible thermal excitations. The corresponding rates thus correspond to an [Arrhenius-like dependency](#)^{193, 194}, which is related to the height of the activation potential barrier between the native and the transition state, ΔG_0 . As will be shown below, it also depends on Δx , the distance between the minimum of the folded state to the maximum of the unfolding barrier, or the distance to the transition state (see **Figure 3.3**).

Figure 3.3. The effect of force on the energy landscape of a protein

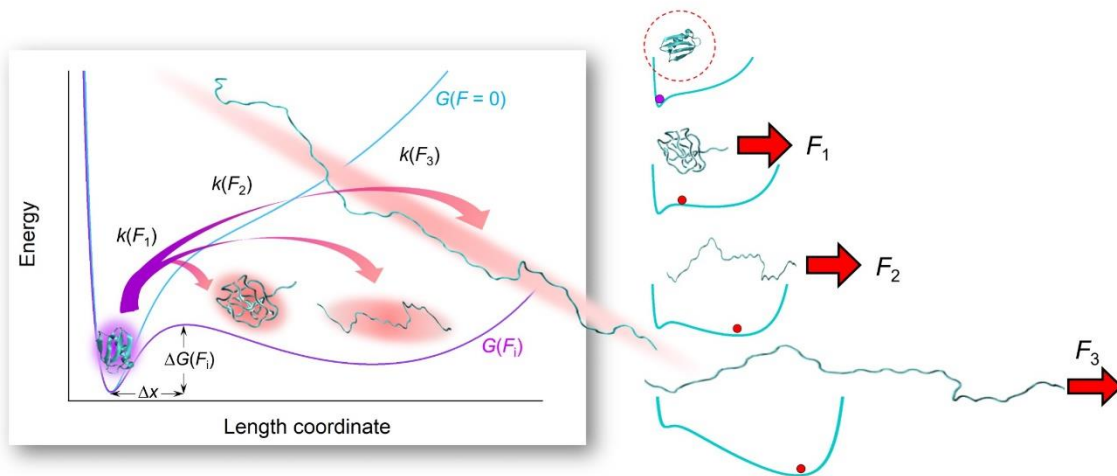


Illustration of unfolding in two-state kinetics under the load. Increasing forces ($F_1 < F_2 < F_3$) results in different unfolded-stretched conformations. Consequently, the force dependent unfolding barrier, $\Delta G(F_i)$, separated by a length Δx from the native state, decreases as the applied force increases. This affects the unfolding rate $k(F_i)$ that increases as well with the increase of the force.

Elevated forces stabilize highly extended conformations, as the mechanical work deepens this state.

Incorporating the two-state model of protein unfolding within the framework of reaction rate theory^{182, 188, 195} provides a convenient approach to quantify the unfolding kinetics. Reaction rate theory states that the probability of reactant to react decays exponentially with respect to a reaction rate, which depends on temperature, and the thermodynamics of the system. Here the reactant is the folded state, and the reaction rate in this context is the rate of unfolding. Therefore, using this approach, we can quantify the force dependent unfolding probabilities, and from their consequent unfolding rates, we can calculate the moments of the unfolding probability, i.e., the mean unfolding force, and its variance. The latter, as will be shown below, can be related to the thermodynamic properties of the protein, such as ΔG_0 , Δx , and k_0 (the unfolding off-rate attempt frequency).

Under the simplifying assumption that unfolding takes place as a two-state process, the dynamics of the survival probability, $S(t)$, the probability to persist at the folded state under the influence of some applied load before time t , can be treated with a first order kinetic equation, $\dot{S} = -k[F(t)]S$. This equation was given as Eq. (2.1) in **Section 2.2**, with the difference that here $k[F(t)]$, the rate of unfolding, is a function of the time dependent force at which the transition occurs. This equation implies that unfolding occurs in a two-state manner, in which refolding rate can be neglected. Following the solution of this equation, given by Eq. (2.2), $S[F(t)]$ is given as a time-dependent single exponential:

$$S(t) = \exp[-k(F)t] \quad (3.2)$$

where $F = F(t)$. This means that the survival probability depends on the force (time) dependent unfolding rate. The [cumulative distribution function \(CDF\)](#) of unfolding force, $P(F)$, is associated to the unfolding force probability density, $p(F)$, and to the survival probability via:

$$P(F) = 1 - S(F) = \int_0^F p(F')dF' \quad (3.3)$$

where $p(F)dF$ provides the probability to unfold during the force interval between F and $F + dF$ (corresponding to the probability to unfold within the time interval t' and $t' + dt'$).

Taking the time derivative of Eq. (2), the following relation is obtained: $-\dot{S}dt = p(F)dF$, leading to $p(F) = -dS(F)/dF$. As previously mentioned, force can be transduced in two main channels: as a constant force with an amplitude F , and at some force rate, $\dot{F} = dF/dt$. If the force rate application is constant, then $\dot{F} = r$, and if it applied at a constant velocity, v , then $\dot{F} = K \cdot v$, with K being the elastic stiffness of the system.

Based on the force rate definition ($\dot{F} = dF/dt$), the time-dependent expressions above can be transformed into force. The survival probability in Eq. (3.2) will now become:

$$S(F) = e^{-\frac{1}{F} \int_0^F k(F') dF'} \quad (3.4)$$

Then, the unfolding probability can be calculated by taking the minus force derivative of the survival probability:

$$p(F) = \frac{k(F)}{F} e^{-\frac{1}{F} \int_0^F k(F') dF'} \quad (3.5)$$

Once the probability density function, $p(F)$, is known, the mean force (first moment of $p(F)$), can be evaluated via:

$$\langle F \rangle = \int_0^F F' p(F') dF' \quad (3.6)$$

The variance can be calculated as $\langle F^2 \rangle - \langle F \rangle^2$. However, to accomplish this, the relation between $k(F)$ with ΔG_0 , Δx , and k_0 has to be resolved.

Mechanical unfolding of a protein involves the rupture of its native bonds (and weak bonds involved in higher structures), as well as the extension of the released unfolded chain. In 1996, R. Zwanzig observed that the folding of some proteins appears to be a two-state kinetic process. By investigating ensembles of unfolded states arriving to thermodynamic equilibrium, he showed the reduction of multiple-state rate equations to a two-state equation¹²⁰. The effect of force on the lifetime of a bond, in terms of crossing a potential barrier separating two states, was first addressed by George Bell in 1978¹⁹⁶, and revisited in more detail by Evans & Ritchie^{193, 194} to describe the force dependent rate for a two-state process:

$$k[F(t)] = k_0 e^{-\frac{\Delta G[F(t)]}{k_B T}} \quad (3.7)$$

To solve equation (3.7), the profile of the free energy, $\Delta G(F)$, has to be determined. A simple phenomenological linear relation was suggested by Bell for cell adhesion¹⁹⁶, according to which $\Delta G(F) = \Delta G_0 - F\Delta x$. However, this expression has a limited range of validity, as it does not consider changes in Δx . As such, this approximation is valid for the regime of small forces. The dependency of the energy barrier can be expanded to a more general form¹⁹⁵:

$$\Delta G = \Delta G_0 \left(1 - \frac{F\Delta x}{\Delta G_0}\right)^{\nu} \quad (3.8)$$

where ν is an additional fitting parameter which represents the curvature of energy barrier. At $\nu = 1$ the linear phenomenological solution is obtained, and $\nu = 1/2$ and $\nu = 2/3$ correspond to a harmonic cusp¹⁸³ and linear cubic^{182, 195} approximations, respectively. Therefore, for a thermally activated high damping dynamics, the rate of crossing the transition barrier is given by¹⁹⁵:

$$k(F) = k_0 \left(1 - \frac{F\Delta x}{\Delta G_0}\right)^{\nu-1} \exp \left[-\frac{\Delta G_0}{k_B T} \left(1 - \frac{F\Delta x}{\Delta G_0}\right)^{\nu} \right] \quad (3.9)$$

Using the expression for the energy profile, given by Eq. (3.8), and integrating eq. (3.6) from zero to F , the relation between the mean rupture forces and the loading rate¹⁸⁸ is:

$$\langle F \rangle = \frac{\Delta G_0}{\nu \Delta x} \left\{ 1 - \left\{ \frac{k_B T}{\Delta G_0} \ln \left[\frac{k_0 k_B T e^{\Delta G_0 + \gamma}}{\left(\frac{dF}{dt}\right) \Delta x} \right] \right\}^{\nu} \right\} \quad (3.10)$$

Here $\gamma = 0.577\dots$ is the Euler-Mascheroni constant. By fitting this expression to a set of measured mean unfolding forces measured at various loading rates, it is possible to extract not only ΔG_0 , Δx , but also k_0 and ν ^{171, 188}.

This approach was further expanded, and culminated with a more comprehensive expression that introduces the parameter m , which corresponds to the shape of the transition landscape¹⁹⁷:

$$k(F) = k_0 \left(1 - \frac{\mu F \Delta x}{\Delta G_0}\right)^{2 - \frac{1}{\mu}} e^{\frac{\Delta G_0}{k_B T} \left[1 - \left(1 - \frac{\mu F \Delta x}{\Delta G_0}\right)^{1/\mu} \right]} \quad (3.11)$$

At $\mu = 1/2$ the shape of the potential barriers represented as a harmonic-cusp, while at $\mu = 2/3$ it is approximated as linear-cubic. At $\mu = 1$, the free energy profile assumes the Bell's limit, in a fully ductile regime which decreases exponentially the transition state distance.

3.3 REFOLDING AGAINST FORCE

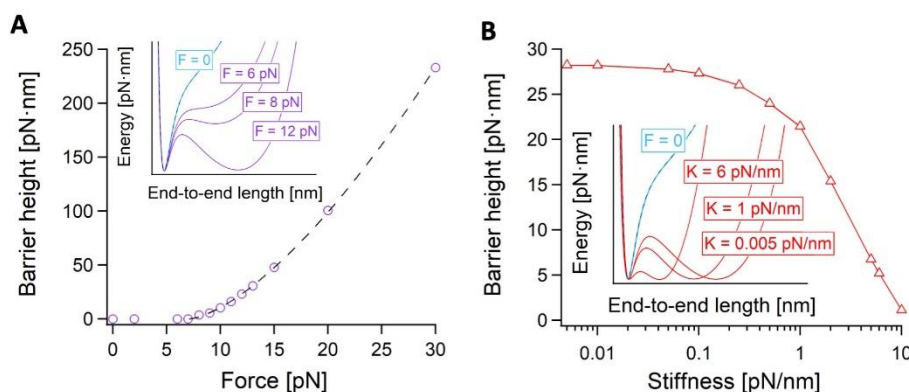
The response of a protein to force is different from its behavior under changing temperatures^{106, 108}, where solvent alterations mediate the structural perturbation, and eventually different transitional barriers are sampled. Therefore, when interpreting kinetic data obtained from protein

unfolding measurements, one should seek to understand which barrier is being probed, and if the analysis does not involve mixing of more than one barrier. Therefore, we have to consider the process of refolding against force.

When the external force is reduced or removed, the unfolded extended chain contracts (collapses) and eventually refolds^{146, 161, 198, 199, 200}. As discussed in the previous section, unfolding from the native state can be well described in a two-state manner, where only the initial conditions (reflected through the survival probability) are considered when crossing the activation energy barrier. The unfolded state however, and its nature, are not considered, because the applied force exponentially increases the unfolding probabilities (according to the two-state model), such that refolding rates are considered as very small with respect to the unfolding rates.

Refolding from a highly extended conformation, by reduction or removal of the force, displays a more convoluted behavior^{68, 146, 161, 198, 199, 201}. Generally, based on the 1D description on the x coordinate given by Eq. (3.1), and discussed in Section 3.2, once the applied force is reduced (or completely eliminated), the unfolded chain entropically recoils into an unstructured coiled conformation (see **Chapter 2.3**). According to this description, if some lower force is applied by a mechanical work $\int F dx$, or $\frac{1}{2}K(x - x_0)^2$ while being arrested at some extension x_0 , then a local minimum is formed in the PMF, which is separated by a barrier^{146, 163}. This barrier, ΔU , illustrated in **Figure 3.4A**, is related to the polymeric properties of the unfolded coiled chain (i.e. the extent of elongation), and the applied load. It spreads across a distance $(x_0 - Rc) > \Delta x$, and it reflects the distance between the unstructured coiled and the globular conformations of the unfolded protein chain.

Figure 3.4. Refolding against force: effect of the unfolded protein from an extended state.



A. Height of the collapse (entropic) energy barrier, ΔU , at different applied forces F (purple circles). The dashed line is given by the relation $\Delta U \sim (F - F_c)^{3/2}$, fitted to the data. Inset: The application of force forms a barrier, ΔU , associated with the polymeric collapse of the unfolded protein chain. (Adapted from ref. 146). B. Height of the collapse barrier with the curvature (spring constant) of the force transducer, K . Inset: Variation in the stiffness of the system affecting the properties of ΔU (adapted from ref. 163).

With the reduction of the force, this barrier decreases as $\Delta U \sim (F - F_c)^{3/2}$ (dashed curve in **Figure 3.4A**), where F_c is the critical force at which the barrier disappears (the maximum becomes a saddle point)¹⁴⁶. Crossing of this barrier marks the coil-to-globule transition¹⁹⁸, after which folding to the native state can be achieved^{37, 199}. If the force is ramped down at a constant v , then ΔU is susceptible to K , the stiffness of the system (see also **Section 3.5**). **Figure 3.4B** shows how the 1D projection of the energy landscape over the x coordinate is reduced with the stiffness of the system, when being held at a constant position $x_0 = vt$. For this comparison, it is interesting to observe that for the used parameters, as K gets smaller, it asymptotically reaches the equivalent constant force barrier (here at ~ 12 pN)¹⁶³.

Under the application of high forces, the protein is driven to occupy highly extended states in its energy landscape. Once the protein crosses ΔU , to a collapsed coiled state, bonds begin to form on a reaction coordinate that manifest in very small changes in x (see **Chapter 2.4**). If the applied force is sufficiently low (or completely removed), then the formation of intermediate, and eventually native bonds involves crossing barriers, that arrange in a downhill configuration (**Figure 3.2A**). This naturally depends on the ratio between these barriers, where low energy barriers will lead to short-lived intermediate states, and higher to longer ones. Overlay of the projected barriers on the bond/heat coordinate on the x coordinate can obscure monitoring of the actual folding process^{37, 146, 198}. However, recent technological advances enable scientists to measure and distinguish between several unfolding/refolding pathways with high accuracy^{173, 199, 202}.

3.4 REGULATING MECHANICAL STRESSES TO PROTEINS UNFOLDING/REFOLDING

Regardless of the complexity of protein structure, the two-state approximation proves a powerful tool to capture the unfolding kinetics of protein folding under force. For instance, it was shown using SMFS (see **Section 4.2**) techniques that the force required to unfold a protein domain increases with the applied extension rate of application (v)^{181, 203}. This behavior is described by the two-state model within the framework of reaction rate theory^{182, 188, 195}, and became a common practice in data analysis. Yet, in this context, thermal activation to cross an energetic barrier also incorporates the contribution of the disordered polymeric chain components, which reflected through the chain compliance (or inverted stiffness)¹⁹⁴. Generally, with the increase of velocity, undulations across the chain are smoothed as they have less time to relax, and consequently more coordinates are included. This effect becomes even more pronounced in sequential unfolding in polyproteins^{43, 183, 194}. This aspect will be discussed in this section.

Polyproteins are unique macromolecules since they are comprised from proteins arrayed in tandem. They respond to mechanical loads through partial unfolding and extension of their unfolded domains. In hetero-polyproteins (made from tethered different proteins), unfolding is expected to follow the mechanical stability ranking (reflected through their energy barriers) of the proteins along the chain, where the weakest protein will unfold first, and the most stable will unfold last²⁰⁴. Accordingly, for homo-polyprotein (made for the same domains in tandem), where all

proteins are considered similar (thus having comparable energetics), one would expect that the unfolding process will be random (along the vectorial direction of the force application, every protein experiences the same load). However, this is not the case since correlations are formed during the sequential unfolding of proteins along the polyprotein chain ^{42, 43, 44, 205, 206, 207, 208}.

If each domain that unfolds within the polyprotein is assumed to occur in a two-state manner, where each unfolding event is considered individually, then there is no relation to the history of the system. This means that previous unfolding events have no effect on current unfolding events. As such, sequential unfolding times within a polyprotein are inherently assumed to be **independent and identically distributed (iid)**, thus expected to follow an exponential decay for the cumulative unfolding time distribution, in which the dwell times between unfolding events are a function of the applied force, $\Delta t(F)$:

$$P(F) = 1 - e^{-\frac{\Delta t}{\tau}} \quad (3.12)$$

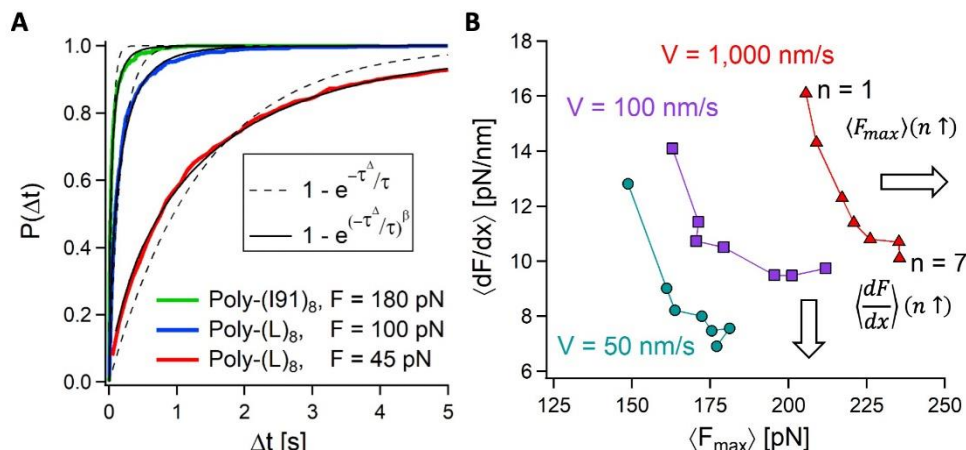
This expression follows Eq. 3.3, where the characteristic time is the mean unfolding time, which is given by the unfolding rate, $t = \langle \Delta t \rangle = 1/k$. According to this assumption, the unfolding kinetics of a polyprotein made of N proteins is equivalent to the unfolding kinetics of such N individual proteins. However, SMFS experiments provide evidence that forced unfolding-times proteins within a polyprotein do not follow the exponential distribution ^{44, 200, 205, 206, 208, 209}, but are better defined with a stretched exponential (or “**Weibull distribution**”):

$$P(F) = 1 - e^{-\left(\frac{\Delta t}{\tau}\right)^\beta} \quad (3.13)$$

While in Eq. (3.12) t represents the mean unfolding time as the characteristic time, in Eq. (3.13) it provides an estimation of the timescale of the unfolding process that stretches across several decades with respect to the exponential constant, β ($\beta = 1$ regains the Poisson distribution, and $\beta = 2$ describes a **Gaussian distribution**).

Figure 3.5A plots the unfolding empirical CDFs of two polyproteins under different constant forces, the highly mechanically stable poly-(I91)₈ (which is an all-beta domain from muscle titin) and the less stable poly-(protein L)₈ (which is an alpha-beta bacterial protein). The high force measurements (100 pN and above) were performed with Atomic Force Microscopy (AFM) ⁴⁴, and the lower (45 pN and below) with Magnetic Tweezers (MT) ²⁰⁸. These methodologies will be described in **Chapter 4**. The CDFs were fitted with Eqs. 3.12 and 3.13 (dashed and solid lines, respectively), demonstrating that the measured data indeed deviates from single exponentiality.

Figure 3.5. Mechanical unfolding of polyproteins.



A. Unfolding time empirical CDFs calculated from constant force (FC) measurements of poly-(I91)₈ at 180 pN (green), and poly-(L)₈ at 100 pN (blue), and 45 pN (red). The CDFs are fitted with a single exponential, Eq. 3.12 dashed line, and stretched exponential, Eq. 3.13 solid line (adapted from refs. 44 and 208). B. Mean chain stiffness, $\langle dF/dx \rangle$ plotted against the maximal unfolding force, $\langle F_{max} \rangle$, calculated from measurements at constant velocity (FX) of poly-(I91)₈ for sequential unfolding events, n , at pulling velocities of 50 (turquoise circles), 100 (purple squares), and 1,000 (red triangles) nm/s (adapted from ref. 42).

When proteins unfold their unstructured segments operate as linkers. The mechanical properties of these polymeric linkers affect the tension propagation²¹⁰ towards the structured parts of the protein. The introduction of an external tension was shown to introduce correlations between the sequential events during the unfolding along the polyprotein^{44, 206, 208}. This effect is apparent in FX measurements, in which the force is applied at a constant velocity, v , and is manifested through a hierarchy in the unfolding probabilities^{42, 43, 184, 207}. With every unfolding event along the polyprotein chain, the polymeric component grows, and consequently lowers the stiffness of the chain. The unfolding forces and stiffnesses can be directly measured in FX experiments, by collecting the maximal force values (F_{max}) at each unfolding force peak, and by taking the slope (dF/dx) before each unfolding event, respectively.

Figure 3.5B shows the mean stiffnesses that were measured during sequential unfolding of poly-(I91)₈ against their corresponding mean unfolding forces at three different pulling velocities⁴². While the unfolding forces increase with v (see section 3.2), the stiffness, in general, appears to also increase with velocity. This means that in addition to elastic effects, viscoelastic effects are also present^{42, 134, 184}. More interestingly, it can be seen that, for every velocity, while the mean unfolding force increases with domain number ($n = 1: 7$ here), the chain elasticity is reduced.

Here one has to consider the local velocity of the chain along its x -coordinate, $\dot{x}$ (the time derivative of the measured evolution of the end-to-end length of the molecules). Although the measurements are performed at constant v , the local coordinate changes in time. It slows down with the relaxation of the chain after each unfolding event, and accelerates as the unfolded chain

further elongates. At low $\dot{x}$, the chain has enough time to randomize, which means that the mean force it experiences is near the equilibrium elastic limit. As the chain elongates to high stretch, $\dot{x}$ increases and thermal randomization narrows down, and viscous effects become dominant. This results in the increase in tension propagation along the chain²¹⁰. Hence, as $\dot{x}$ grows, the chain stiffens as the force it conveys also grows in addition to the stiffness of the pulling apparatus (at a given v)¹⁹⁴. With every unfolding event, the polymeric component of the polypeptide grows, and becomes more compliant (less stiff) at the force required to unfold the next protein. This change in compliance requires additional mechanical work, and consequently the force required to unfolding increase with n .

After examining the CDFS, we now look at the manifestation of the observed deviation from exponentiality in the PDFs. If the system would comply with Poisson behavior (iid), then its PDF will be described with an exponential decay:

$$p(F) = 1 - \frac{1}{\tau} e^{-\frac{\Delta t}{\tau}} \quad (3.14)$$

The introduction of an external tension introduces correlations between the sequential events during the unfolding along the polyprotein. This observation confronts the iid assumption. The PDFs of systems that show such asymptotic behavior of the empirical distributions can also be well described by an algebraic decay (power law) of the form $p(\Delta t) \sim (t/\Delta t)^{1+\alpha}$, where α is a disorder parameter, ranging between 0 and 1, and t is a scaling factor. For such a dependency, the rate is given by $k \sim (\alpha/t^\alpha)\Delta t^{\alpha-1}$, and at $\alpha = 1$ the same $k = 1/t$ dependency as is obtained for the exponential decay. To reflect the physical boundaries of the system, we multiply the power law with an exponential. This form is called **truncated power law (TPL)**, and it is used to describe temporal distribution for anomalous transport^{211, 212}:

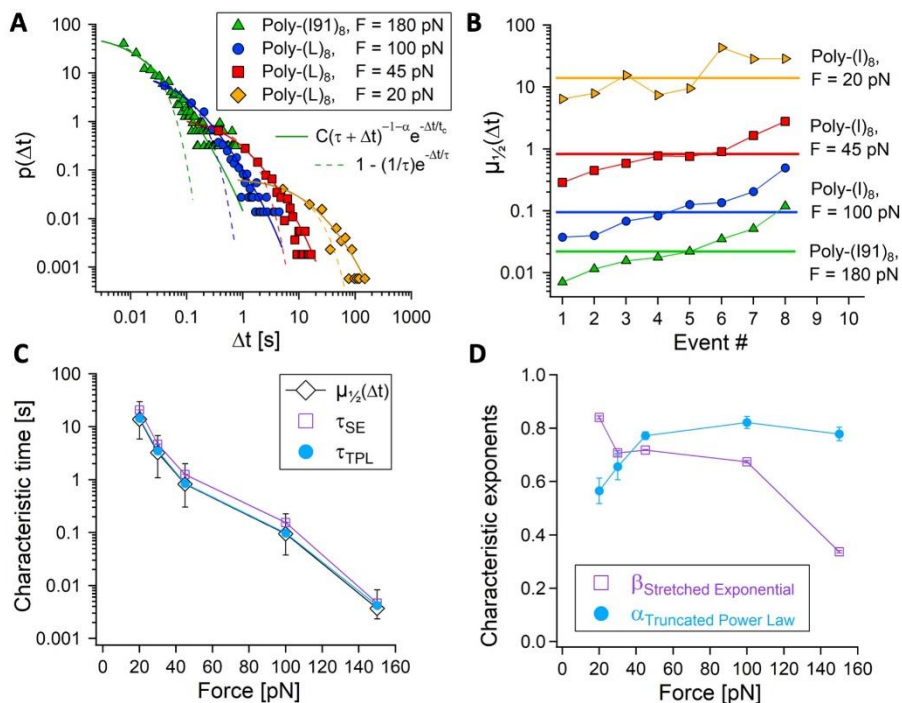
$$p(\Delta t) = C(\tau + \Delta t)^{-1-\alpha} e^{-\Delta t/t_c} \quad (3.15)$$

The longest observation unfolding time during the experiment duration time, t_c , sets the cutoff of system, and bounds the values of Δt , C is a normalization factor²¹². This expression (TPL) is used to characterize systems that display subdiffusive transport behavior. It is justified by the coexistence of extremely short and long dwell times during the transitioning from one energy barrier to a another as unfolding progresses along the polyprotein.

Figure 3.6A shows that the TPL (3.15) fits better the empirical distributions of the unfolding time of poly-(I91)₈, and poly-(protein L)₈ PDFs at several forces compared to the exponential decay (Eq. 3.14)^{44, 208}. This provides further verification that experimental data that display nonexponentially behavior be assessed and analyzed using different models (with sound physical meaning) than the conventional single exponential, which are convenient and appealing to use. With respect to the individual unfolding events along the chain, the medians calculated from the data show in **Figure 3.6B** distinct hierarchy with the event number (as unfolding progresses). The

choice of characterizing the unfolding times with the medians rather than the mean resulted from the asymptotic (nonexponential) behavior of the distributions. In such cases the median is more statistically resistant to the large fluctuations in the outlays of the distributions.

Figure 3.6. Unfolding dwell time analysis showing hierarchy.



A. Empirical protein unfolding time PDFs calculated from constant force (FC) measurements of poly-(I91)₈ at 180 pN (green triangles), and poly-(L)₈ at 100 (blue circles), 45 (red squares) and 20 (orange diamonds) pN. The PDFs are fitted with a single exponential, Eq. 3.14 dashed lines, and TPL, Eq. 3.15 solid lines. B. Medians calculated from the PDFs shown in A, using the same color coding for all the unfolding times (straight thick lines), and for the individual unfolding events (colored symbols). C. Overall characteristic unfolding times (for all n at each force), for poly-(L)₈ given by their medians (empty diamonds) and fitted values from their CDFs using the stretched exponential (purple circles), and PDFs using the TPL (light blue triangles). D. The fitted characteristic exponents, corresponding to the fits in C (adapted from refs. 44 and 208).

In **Figure 3.6C** the unfolding time medians are compared to the characteristic times obtained from the stretched exponential and TPL fits to the CDFs and PDFs, respectively (for poly-(protein L)₈). The TPL fitted t shows considerably better agreement (relative error smaller than 10%) with the empirical medians compared to the stretched exponential fitted t (relative errors of about 50%)²⁰⁸. **Figure 3.6D** shows the overall characteristic exponents from the fitting in **Figure 3.6C**. First, it can be seen that both α and β are smaller than 1. This means that the observed dynamics is

indeed of subdiffusive transport ²¹³. Yet, they display opposing trends in which α decreases with the applied force, and β increases. Further statistical examination of these trends showed that the correlations increase with the applied force, which may support the behavior of the TPL modeling.

The energy barriers of the individual unfolding events can be estimated from the characteristic unfolding times $\Delta E(n, F) = k_B T \ln[t(n, F)/A]$, where $A = k_0 \exp(\Delta G_0/k_B T)$ is the attempt frequency at $F = 0$. Based on the trend in Figure 3.6B, a hierarchy in the unfolding barriers with the progression of the unfolding process along the polypeptide molecule is obtained ²⁰⁸. This can provide a physical ground for the observed deviation from exponentiality, and for the adequacy of describing it with anomalous subdiffusive transport models that well describe the exhibited PDFs ²¹³. The increase in t with n means that the unfolding rate decreases as unfolding progresses, due to the deepening of the effective activation barriers. A process in which the system “gets stuck” for longer and longer time is referred to as aging ²¹⁴. The increase of the sequential unfolding effective barriers can be related to the elongating chain with each unfolding event that intervenes with the unfolding probabilities, as the chain stiffness decreases ^{42, 186}.

That’s a Wrap

- Both the magnitude of the force, as well as its orientation (points of contact), influence the mechanical unfolding kinetics of a protein.
- Due to its non-equilibrium nature, the faster the force is applied (loading rate), the more stable a protein will be,
- Refolding of a protein from a highly extended conformation is governed by the entropic collapse of the polypeptide chain.
- Mechanical unfolding of polyprotein is influenced by the entropy of each unfolded domain.
- Sequential unfolding in polyproteins results in nonexponential distribution due to correlations that form along the chain.

Read These Next

Garg, A., Escape field distribution for escape from a metastable potential well subject to a steadily increasing bias field. *Phys. Rev. B* **1995**, 51 (21), 15592-15595.

Bustamante, C.; Chemla, Y. R.; Forde, N. R.; Izhaky, D., Mechanical processes in biochemistry. *Annu. Rev. Biochem.* **2004**, 73, 705-748.

Hughes, M. L.; Dougan, L., The physics of pulling polyproteins: a review of single molecule force spectroscopy using the AFM to study protein unfolding. *Rep. Prog. Phys* **2016**, 79 (7).

Dudko, O. K.; Hummer, G.; Szabo, A., Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.* **2006**, 96 (10).

Cossio, P.; Hummer, G.; Szabo, A., On artifacts in single-molecule force spectroscopy. *P Natl Acad Sci USA* **2015**, 112 (46), 14248-14253.

Popa, L.; Berkovich, R., Mechanobiology: protein refolding under force. *Emerg. Top. Life Sci.* **2018**, 2 (5), 687-699.

Bura, E.; Klimov, K.; Barsegov, V., Analyzing forced unfolding of protein tandems by ordered variates, 1: Independent unfolding times. *Biophys.J.* **2007**, 93 (4), 1100-1115.

Chetrit, E.; Sharma, S.; Maayan, U.; Pelah, M. G.; Klausner, Z.; Popa, I.; Berkovich, R., Nonexponential kinetics captured in sequential unfolding of polyproteins over a range of loads. *Curr. Opin. Res. Biol.* **2022**, 4, 106-117.

Chapter 4. Methods to study the mechanical unfolding of proteins

4.1. Introduction

Mechanical unfolding of proteins requires exposure to a force vector, which has both a direction and a magnitude. As such, a protein is tethered between two points along its structure, typically between its N- and C- termini. The response of a tethered protein is inextricably influenced by the probes used to apply force (bead, cantilever, surface) and by its surrounding environment. Ideally these probes would be much smaller than the molecule being investigated, to minimally perturb them (imagine a laser beam bouncing off a moving ball to measure the ball's position), but in reality, these probes largely exceed the size of the molecules and impact the measured dynamics (see discussion at the end of this chapter). Here we will focus our discussion on methods used to study the unfolding response of proteins outside their cellular environment, on force sensors used at the interface between cells and substrates, on tissue-like approaches and on molecular simulations. We will also discuss how measuring devices and conditions impact the measured behavior. Each measuring technique will have two main elements: one used to apply mechanical force and one used to measure the effect of the mechanical force on proteins. Apart from these two elements, an important consideration has to be given to attaching proteins to the force generating/measuring probes, which we discuss first.

4.2. In vitro force spectroscopy methods

In vitro methods rely on using purified proteins taken outside their native environment and measured in a solution with similar pH and ionic strength as inside a cell. As one needs to apply a well-defined mechanical perturbation, most in vitro methods operate at the single molecule level and are termed *single molecule force spectroscopy (SMFS)* approaches. Below we describe three of the most used SMFS approaches: atomic force microscopy (AFM), optical tweezers (OT) and magnetic tweezers (MT). Other SMFS methods that have been used to some extent to study the mechanical response of biomolecules are surface force apparatus²¹⁵, sheer-flow^{216, 217}, acoustic force spectroscopy²¹⁸ and centrifugal force spectroscopy²¹⁹. We refer the reader to the cited references to find out more about these techniques. Apart from the different methodologies of applying force and measuring its effect on a biomolecule, a critical aspect of all SMFS techniques is related to how proteins are attached to the tethering probes. Hence, we begin our discussion with various molecular attachment approaches. We will then go over how the three main SMFS techniques operate and finish this subsection by talking about bulk approaches to apply force.

4.2.1. Attachment chemistries for single molecule force spectroscopy methods

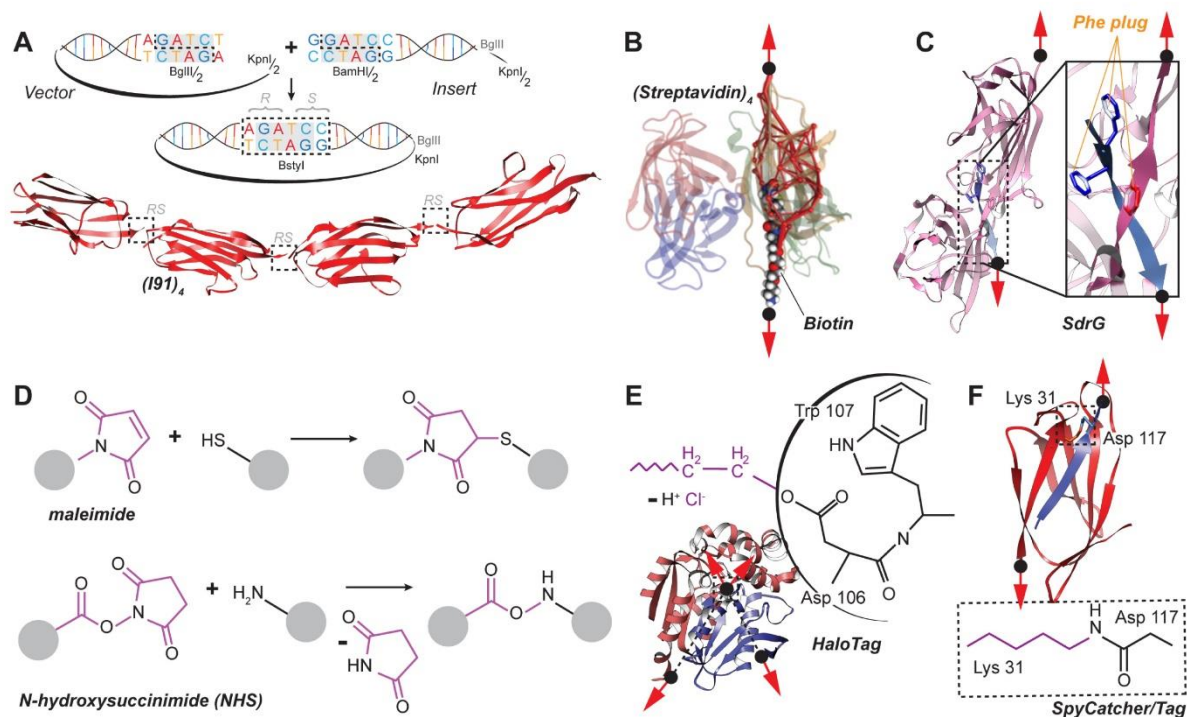
The main strategies to attach and tether molecules under force rely on using physical interactions, specific non-covalent chemical interactions and targeted covalent attachments. Each vary in the level of complexity, attachment strength and specificity, as further discussed below.

4.2.1.1. *Non-specific physical interactions*

The force that can be applied to a single molecule is as high as the weakest link between that molecule and the tethering probes. Many early approaches relied on *non-specific physical interactions* for tethering proteins. These physical interactions may come from van der Waals attraction, or electrostatic and hydrophobic connections. While these interactions are intrinsically weak, their overall effect can be significant if many of them act in parallel⁷³. Because in SMFS measurements the tethering probes are approached and retracted 'blindly' in the hope of tethering a molecule. These methods are often compared anecdotally to fishing. So how do you catch fish without a bait on your hook? One way is to saturate the waters with fish and increase your chances of a random encounter. A similar approach has been taken for AFM experiments, where a surface is typically completely covered with the protein of interest. In the specific case of AFM, to ensure that only a single molecule is tethered, a cantilever with a relatively sharp tip is used (typical curvature radius smaller than 10 nm, comparable to the size of a protein). This small radius of the cantilever improves the chances of attaching to a single molecule. However, by simply using a monomeric protein, there is no direct way of knowing when one pulls on two molecules perfectly in parallel. Furthermore, there is no direct way of knowing if the molecule being measured is actually the protein added to the fluid chamber.

It is a well-known fact that biological samples tend to be rather 'dirty' and some proteins would degrade or aggregate. Hence statistical analysis is needed to remove the multi-tethers and 'junk' molecules²²⁰, but blunted tips and contaminated sample can produce inaccurate (or even wrong) results. An elegant approach to circumvent this uncertainty when using physical interactions was introduced by Fernandez and collaborators and relies on molecular fingerprints¹³⁶. In their approach, the scientists took the gene of the protein of interest and repeated it several times (**Figure 4.1A**). The researchers took advantage of the unique sequences of two known digestion sites BamHI (GGATCC) and BglII (AGATCT), which are both asymmetrically cleaved after the first base and can form in the subsequent ligation step a new site, BstYI (AGATCC). The other ends of the vector/insert can be ligated back using a second site, such as KpnI. A new nondigested BglII site is added with the insert and can be used to open the vector again and add more repeating units. Through subsequent BamHI/BglII digestions (which do not cleave the newly formed BstYI sites), followed by ligation, one can easily repeat the same protein sequence many times (**Figure 4.1A**). Competent *E. coli* cells are typically used for the polyprotein expression. Because expression approach becomes inefficient for proteins over 100 kDa, the protein of interest is typically repeated for a maximum of 8-12 times. When pulled, these polyproteins produce a unique molecular fingerprint, coming from the repeated unfolding of each of the protein domains, which takes place at slightly different times (examples of such traces will be shown in Chapter 5).

Figure 4.1. Various approaches used for tethering single proteins under force.



A) Nonspecific adsorptions require the use of polyproteins. (*top*): A method to repeat the sequence encoding a certain protein relies on the use of a special feature of two **restriction sites**, BamHI and BglII, which have a similar middle sequence (grey square), that allows them to form a new site, BstY. (*bottom*) Ribbon representation of a titin I91 tetramer (pdb: 1tit), which was one of the first constructs to be produced as a polyprotein ¹³⁶. B) Structure of a complex between Streptavidin-Biotin (also known as AviTag) showing the most stable attachment state (of the possible four. Adapted with permission from ref. 221. C) SdrG forms one of the strongest non-covalent bonds to a peptide ligand (marked in orange) through a “screwlike” mechanism through a Phenylalanine (Phe) screw. Rendered from pdb: 1r17. D) Popular chemical crosslinking strategies used to attach proteins to substrates: (*top*): addition reaction between a thiol group (part of a cysteine amino acid) and maleimide; (*bottom*): reaction between amines (present as part of terminal, lysine or arginine amino acids) and a N-hydroxysuccinimide (NHS) group. E) HaloTag technology relies on a mutant Haloalkane Dehalogenase, which forms an ester bond with a chloroalkane ligand. Depending on the application of the force between the catalytic center of HaloTag and the remainder of the protein, different extensions and stabilities are obtained ²⁰. F) SpyCatcher/Tag covalent attachment uses a split protein from a fibronectin binding bacterial protein, which naturally forms an internal isopeptide bond ²²². While having slower binding kinetics than HaloTag, SpyCatcher/Tag attachment has the advantage of using peptide tags (rather than small chemicals), which can be natively expressed by cells.

4.2.1.2. Specific non-covalent interactions

Strong non-covalent interactions have been successfully used to tether proteins in SMFS approaches. They typically rely on known stable protein-protein and protein-ligand interactions. Among the most used non-covalent interactions are antibody-antigen, Biotin-Streptavidin and split proteins, most of them having nM [dissociation constants](#). Antibody-antigen attachment is used extensively in OT measurements, where Digoxigenin (DIG) group is added to a DNA linker to bind to anti-DIG functionalized polystyrene beads ²²³. Biotin (also known as vitamin H or B7 or AviTag) binds to a homo-tetrameric Streptavidin, a protein derived from a bacterium. While chemically, Streptavidin-Biotin is one of the most stable non-covalent bonds in nature, mechanically it breaks at <100 pN in under 2-3 min, but some bonds were shown to survive longer (hours or even days) ²²⁴. Recently it was discovered that how the force is transmitted through the Biotin-Streptavidin complex impacts the strength of this interaction, with the complex that bears the force directly in-between the surface and biomolecule producing the most stable tether ²²⁵ (**Figure 4.1B**). Other important non-covalent interactions used in force spectroscopy were engineered using secreted bacteria proteins. Adhesin SD-repeat protein G (SdrG), which is secreted by a pathogen targeting a short peptide, can withstand over 2 nN of force ²²⁶. The target peptide anchors inside the SdrG protein in a screwlike manner, providing the exceptional stability through a [catch-bond](#) mechanism (**Figure 4.1C**). Similarly, the cohesin–dockerin system was shown to withstand nN forces using a molecular locking mechanism ²²⁷.

4.2.1.3. *Specific covalent interactions*

Due to their intrinsic strength and specificity, covalent bonds provide ideal attachment anchors. However, they are relatively hard to implement. One way of realizing covalent bonds is through short linkers, such as using an amino acid with a reactive group placed at the end of a protein construct. Among the 20 amino acids, cysteines provide a good opportunity for covalent attachment as they enable covalent S-C, S-S or S-Au bonds using specific chemistries (such as an addition reaction between a thiol and a maleimide group, which produces a S-C bond) (**Figure 4.1D**). Amine-terminated amino acids can also be linked through their reaction with N-hydroxysuccinimide (NHS) (**Figure 4.1D**). Histidine amino acids, which are also extensively used in protein purification, were also transformed into potential covalent attachment tethers via a phosphorylation strategy ²²⁸. A [Sortase](#) driven approach has also managed to produce peptide bonds between terminal N-G-L and G-L amino acids ²²⁹, but there are challenges related to catalytical efficiency of the enzyme ²³⁰. [Click-chemistry](#), which relies on an azide-alkyne cycloaddition reaction ²³¹, is also extensively used in bioconjugation, and its developers earned the 2022 Nobel Prize in Chemistry.

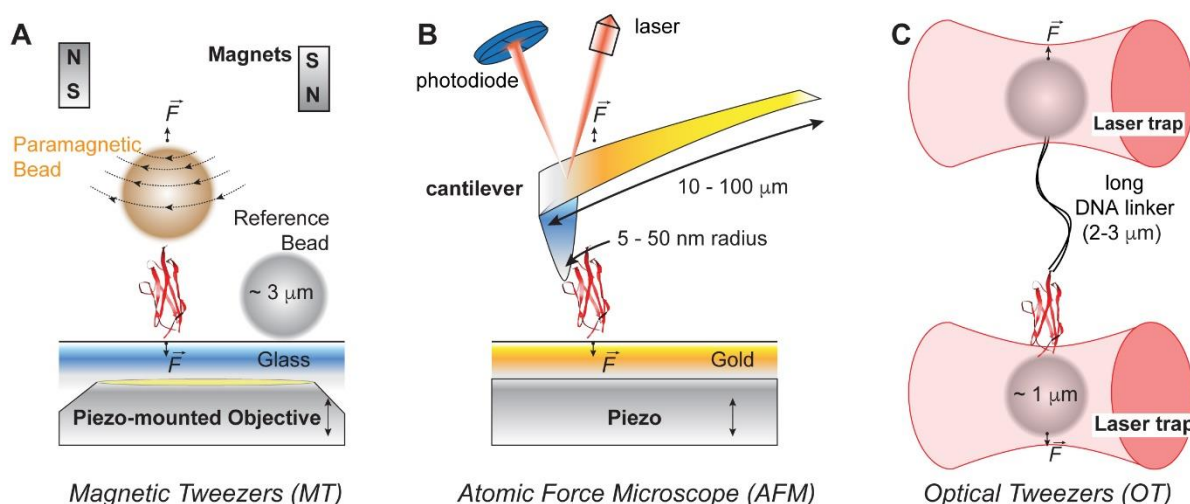
A more specific approach to producing covalent tethers relies on using active protein domains, engineered through molecular biology approaches, to be part of the protein construct of interest. These protein domains can form covalent bonds with small linkers or tags. HaloTag technology is currently the most effective covalent attachment approach for proteins, and relies on a mutant Haloalkane Dehalogenase enzyme, which forms an ester bond with a chloroalkane ligand (**Figure 4.1E**). HaloTag lacks a critical amino acid present at the catalytic site of the wildtype Dehalogenase, which is otherwise responsible with releasing the substrate through the

hydrolysis of the newly formed ester bond. Interestingly, HaloTag will respond differently when placed at the N or C-terminus of a protein, as in each case force is applied to a different half of the protein, connecting its catalytic site to the rest of the domains ²⁰. The downside of HaloTag is that its ligand is a small chemical, which cannot be expressed in cells. Circumventing this shortcoming, the SpyTag-SpyCatcher technology relies on a split protein approach, taken from a bacterium protein domain that forms an internal isopeptide bond (**Figure 4.1E**) ²²². As discussed in Chapter 5, this isopeptide bond, which naturally forms between two amino acids with opposite charges located inside the protein structure, prevents the mechanical unfolding of these domain. In the split protein, the smaller section (SpyTag) docks to the larger domain (SpyCatcher) and forms in time this internal isopeptide bond ²²². While the SpyCatcher maturation is slower than HaloTag (~30 min vs <1 s), it has the advantage that it uses a peptide tag. Having a peptide-based tag is especially important for labeling experiments, where both the tag and catcher can be engineered as linked to other proteins of interest and can be expressed together inside the same cell.

4.2.2. Single Molecule Force Spectroscopy methods

The most popular SMFS methods used to measure the mechanical response of proteins are Magnetic Tweezers (MT), Atomic Force Microscopy (AFM), and Optical Tweezers (OT) (**Figure 4.2**). Each method has a way to apply force and a way to measure the molecular extension, as discussed below.

Figure 4.2. Single molecule force spectroscopy techniques used to measure the mechanical response of proteins.



A) Magnetic Tweezers (MT) tethers a biomolecule between a glass surface and a paramagnetic bead; force is applied using either a pair of permanent magnets or an electromagnet; B) Atomic Force Microscope (AFM) tethers biomolecules between a gold or silica surface and a sharp tip at the end of a cantilever; to apply force, the surface is moved vertically with the help of a piezo-

actuator, and to measure the developing force, the deflection of the cantilever is gauged by bouncing a laser beam from its end onto a quadrant photodiode. C) Optical Tweezers (OT) tethers biomolecules between two trapped beads, with a strong **trap** used to apply force and a weak trap used to measure the experienced force.

4.2.2.1. Magnetic Tweezers (MT)

To apply pN forces to a tethered biomolecule, magnetic tweezers (MT) uses at one end a **superparamagnetic bead** of 2-3 μm in diameter, while the opposite end the molecule is attached to a glass coverslip (**Figure 4.2 A**). As there is no precise way to measure the instantaneous position of the glass slide, and hence the non-moving end of the tether, a second non-magnetic (reference) bead is needed, which is glued to the surface. Hence MT does not measure absolute extensions, but changes in molecular length from a low magnetic force (typically ~ 1 pN) to a high magnetic force (up to ~ 100 pN). To monitor the position of the magnetic and reference beads, MT takes advantage of the interference pattern forming around the two beads, which is highly sensitive to focal changes. In the beginning of an experiment, the objective is moved with the help of a **piezo actuator** in predefined steps (10-20 nm) and a stack library is saved. During measurement, each interference pattern of the two beads is correlated against its respective stack library to determine the absolute position of each bead, and the difference between these two positions produces the extension of the molecule. During the experiment, the reference bead plays another important role – it allows to monitor and correct for focal drift by moving the objective-piezo assembly to maintain this bead at its initial position. It is this active drift correction mechanism that allows for measurements that can last hours, or even days ²²⁴.

MT is best suited for applying physiological-like forces (which are typically <10 pN per molecule ²³²) over extended times, of hours, comparable to the protein turn-over time in vivo. As the moving end relies on a levitating bead which is only physically attached to the measuring setup through the molecule, MT has a low spring constant of the trap, which gives it good accuracy at low forces. Since the applied force changes with the gradient of the magnetic field, which varies over a mm scale, any unfolding events that a protein domain might have moves the bead over a few tens of nm. This orders-of-magnitude difference in distance will ensure that the experienced force stays constant as the molecule extends (this feature is also defined as a passive force clamp). However, MT relies on live image processing, which limits its acquisition speed to ~ 1 kHz, and makes it best suited to measuring molecular events that take place on the second-time scale. Also MT typically uses short tethers (tens-to-hundreds nm long) close to a surface, where the viscosity is different, which complicates the calibration procedure.

4.2.2.2. Atomic force microscopy (AFM)

The atomic force microscope (AFM) tethers biomolecules between a piezo-mounted surface (such as a gold-coated surface, **Figure 4.2B**) and a cantilevered tip. To apply force, the surface is moved vertically away from the cantilever tip with the help of a piezo actuator. To measure the

molecular tension, the deflection of the cantilever is determined with the help of a laser beam, which bounces from the back of the cantilever's free end onto a [quadrant photodiode](#). While the force is given by the deflection of the cantilever, the extension is inferred from the movement of the piezo. Typically the piezo is moved away from the cantilever with constant velocity, but in some instrumentations an active feedback mechanism can continuously adjust for the piezo position to maintain a constant setpoint force ²³³.

While the tip of the cantilever has 5-50 nm radial curvature, increasing its chances of single tethers, the cantilever itself is 10-200 mm long and at least a few mm wide, as it needs to be comparable to the focused laser spot used to monitor its deflection. The cantilever is also thin, to allow for a low spring constant (and hence high deflection/force sensitivity) and carving out part of the cantilever can further improve its response ²³⁴. To improve their laser reflection, cantilevers are typically coated with a top gold reflective layer. The precise localization of the cantilever and accurate positioning of the surface allows AFM to excel and measure molecular extensions with high bandwidth (~10 kHz) and with sub-nm resolution (given by the precision of the piezo actuator). However, it is the cantilever's dimensions that limit the functionality of the AFM to forces over 10 pN, due to the relatively large thermal motion of its moving end. Furthermore, the gold coating, which is continuously heated by the laser beam and cooled by the surrounding environment, transforms the cantilever into a very sensitive [thermocouple](#), which limits the measurement time to a few seconds, before the thermal drift makes the measured force imprecise. As AFM can apply forces of up to a few nN with excellent time resolution, it is best suited to measure strong binding interactions, highly-mechanically stable proteins and breakage of covalent bonds ²³³. The AFM is also a very popular surface imaging technique, and was pioneered by Gerd Binnig, who was awarded the Nobel Prize in Physics in 1986.

4.2.2.3. Optical Tweezers (OT)

Optical tweezers (OT) rely on the net force generated by a high-power infrared (IR) laser beam to attract objects of comparable size toward its focal point ²³⁵. This property of focused light, known as optical trapping, is used to control the position of micron-sized polystyrene beads (~1 mm in diameter), and for its discovery, Alan Ashkin was awarded the Nobel Prize for Physics in 2018. In its most popular form, OT uses a dual-trap arrangement, where biomolecules are tethered between two trapped beads (**Figure 4.2C**). In this arrangement, one stiff trap is used to apply force, while the other softer trap measures molecular extension. As the trap used to apply force requires a stiffer spring constant, and hence a higher laser power, a long (micron-sized) DNA linker is used to space the biomolecule of interest away from laser beam and minimize photothermal damage. The molecular displacement is determined from the movement of the bead from the focal point of the soft trap, using a quadrant photodiode. When operated in a narrow trap displacement, OT can achieve passive force-clamp behavior ²³⁶.

Similarly to MT, OT uses micron-sized beads, which limits the thermal effects of the probes and allows for precise measurements in the force range of most physiological processes (< 10 pN per molecule). As with the AFM, the detection system on which it relies is a quadrant photodiode, which can produce acquisition frequencies of ~10 kHz, improving position localization. However, the use of intense laser beams needed for trapping results in photochemical ([singlet oxygen](#)) and photothermal damage²³⁷, which limits the length of OT traces

to a few minutes. Furthermore, the DNA linkers used to space the biomolecule from the trapped bead can undergo an overstretching transition of 65 pN. Hence OT are ideal to measure biological transitions taking place on the millisecond time scale and at low (<65 pN) forces.

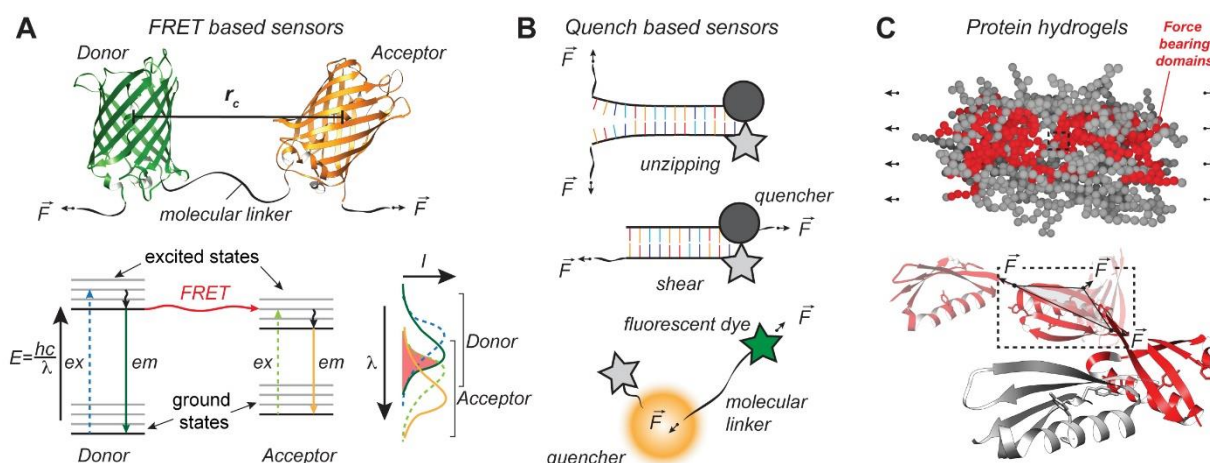
4.3. In vivo and tissue-like approaches to study protein unfolding

The main approaches used in vivo and in biomaterials to apply force and measure the mechanical response rely on molecular tension probes, protein-based materials and molecular dynamics simulations, as discussed next.

4.3.1. Molecular tension probes

Molecular tension probes use a fluorescence-based response to measure the applied force at the interface between a cell and a substrate. They are typically engineered as force switches, which turn on and off when a certain force is reached and are built using peptide/protein or DNA linkers. The peptide/protein tension probes use a pair of either fluorescent proteins (such as those from the GFP family) or fluorescent dyes (such as Cyanine dyes). The measurement relies on Förster resonance energy transfer (FRET) between these two fluorescent probes (**Figure 4.3A**). FRET is an energy transfer process which can take place between two **chromophores**, when one of them (known as donor) has an emission spectrum that partially overlaps with the excitation spectrum of the second probe (known as acceptor; in **Figure 4.3A** bottom right the emission spectrum of a donor chromophore is shown with continuous green and the excitation of the acceptor with dotted light green, while the overlapping region is marked in red). FRET is highly sensitive to the separation and relative orientation of the donor and acceptor molecules (the transfer **FRET efficiency** scales with the Förster separation r_c to the power of 6 and typically takes place when donor and acceptor chromophores are less than 10 nm apart, through a non-radiative dipole-dipole coupling). By using peptides or structured linkers (such as α -helix) of given lengths, one can correlate pulling forces with the decrease in FRET, since under tension the separation between the fluorescent probes changes sharply with force (and follows polymer elasticity laws – see also discussion in **Chapter 2**). Protein based FRET probes can be expressed in cells and used to measure forces between intracellular structures ²³⁸. FRET probes require relatively short linkers, as it has a maximum sensitivity at Förster separations of 5-6 nm, limiting their force range ²³⁹.

Figure 4.3. In vivo and tissue-like measurements.



A) FRET-based tension sensors. (*top*): representation of a FRET tension sensor that uses a donor-acceptor protein pair separated by an unstructured polypeptide linker; (*bottom left*): energy diagram illustrating the ground (bottom) and excited (top) electronic states (black lines) and their vibrational levels (grey lines) of a Donor (left) and Acceptor (right) chromophores.; (*bottom right*): corresponding excitation (*ex* – dotted lines) and emission (*em* – simple lines) spectra relating how intensity (*I*) changes with wavelength (*λ*). The red area marks the overlap between the emission spectrum of the donor and excitation spectrum of the acceptor, where FRET can occur. B) Quench-based sensors; (*top*) sensor based on unzipping of a DNA fragment, which releases the quencher from the chromophore above a force threshold, turning the substrate fluorescent; (*middle*): a similar approach using the shear of DNA strands; (*bottom*): a gold-bead acting as a quencher for a chromophore attached through a polymeric linker; when force moves the chromophore away from the bead, the chromophore starts to fluoresce (green vs grey star). C) Tissue-like approaches based on protein hydrogels; (*top*): coarse-grained representation of a protein hydrogel showing the load bearing network (red) and inactive protein domains (grey); taken from ref. 240 with permission; (*bottom*): representation of protein domains inside a biomaterial tethered between their N- and C- termini, as well as at specific cross-linking sites; grey triangle shows the force distribution on a protein domain acting as a network node.

Another approach to molecular sensors uses a quencher to keep a fluorescent protein/dye in the ground state (**Figure 4.3B**). The quenchers can be modified nonfluorescent acceptor dyes, which capture the donor energy through FRET²⁴¹, or noble metal surfaces/beads (such as gold), which quench fluorescence through surface energy transfer²⁴². The quench-based sensors have the advantage that they do not show a fluorescence signal unless the quencher and chromophore are spaced apart by force. To control the force threshold, the quenched-based sensors can rely on double stranded DNA fragments, which have the fluorophore and quencher at one side, and the ends used to attach the fragment to the surface and to a liable group on the opposite sides (for unzipping mode) or apart (for shear mode – see **Figure 4.3B**). The force threshold of the force sensor can be controlled by changing the size and composition of the DNA linker²⁴³.

Molecular tension probes allow scientists to sample physical connections between a cell and a substrate or inside the cell, and have been instrumental in measuring the formation and maturation of cellular focal adhesions. Molecular tension probes do not provide direct knowledge of the molecular structures or number of tethers producing a given force, and have the downfalls related to using fluorescent probes, such as bleaching. However, they represent an elegant approach to a hands-off approach to apply mechanical forces and measure their effects in vivo. Other sources provide more in-depth discussion of molecular tension probes ²⁴⁴.

4.3.2. Protein hydrogels and tissue-like materials

Rheometry-based approaches were developed decades ago to study the response of materials to stress and determine their elastic and plastic deformations, as well as their yield point. Similar approaches have been successfully implemented to study muscle fibers and muscle sarcomeres, where stripped down tissues are attached between an actuator and a force sensor ^{245, 246}. Using rheometry-inspired approaches, we and others have developed a biomaterial-based approach to study the mechanical response of proteins ^{247, 248}. This approach is relatively straightforward as it relies on covalently cross-linking proteins at exposed tyrosine sites via a light-triggered chemical reaction (**Figure 4.3C**). Once a protein solution is turned into a biomaterial, it can be physically manipulated and attached between an actuator and a force sensor, similarly to standard rheometry approaches. As protein-based biomaterials contain over 80% water ²⁴⁸, they are also called hydrogels. This high-water content is important, as it allows protein domains to stay folded and behave as they would be in solution. When a protein has fewer than three cross-linking sites (which is the minimum number required to form a network) or has low solubility (and precipitates before a gelation concentration is reached), one can use the same polyprotein approach developed for AFM to produce hydrogels (**Figure 4.1A**).

When measuring the mechanical response of protein-based materials, a setup that can operate under force-clamp conditions is desirable, such that the stress stays constant throughout the experiment. Such a setup uses an active feedback mechanism which continuously adjusts the strain of the biomaterial to match a setpoint stress, as protein domains inside the material unfold and extend or contract and refold as the force is cycled between a high and a low value ²⁴⁸. While synthesizing protein-based materials and measuring their force-response is relatively straightforward, and much simpler to implement than the other techniques discussed thus far, scaling the macroscopic response of protein-biomaterials to the nanoscopic unfolding of protein domains is relatively complicated, as it requires coarse-grained modeling approaches that can separate the network effects from the unfolding responses ²⁴⁹. As mentioned in **Chapter 1**, the mechanical response of a protein depends on the tethering geometry and proteins inside hydrogels will have several tethering orientations (**Figure 4.3C**). However, using polyproteins can result in the majority of the protein domains experiencing a N-to-C force ²⁴⁹, but this does not guarantee that the macroscopic response will not be significantly influenced by the other tethering geometries, which might produce weaker states. A further complication is also coming from the fact that not all domains forming the protein-biomaterial contribute to the force-response, as they might not be part of the load-bearing network (**Figure 4.3C** red vs grey domains) ²⁴⁰. In spite of

all these shortcomings, these approaches utilizing biomaterials made from pure proteins allow for the study of the mechanical response of proteins in crowded environments, resembling that of cells and tissues, and constitute straightforward way to study mechanical changes induced by solvent or binding partners, as one only has to change the solution inside the measuring chamber where the hydrogel rests. Furthermore, a single stress-relaxation or high stress-low stress curve produces the averaged unfolding and refolding response of billions of proteins at once, a feat that no single molecule technique can reach.

4.4. In silico methods - steered molecular dynamics (MD) simulations

Molecular dynamics (MD) simulations rely on force fields to reproduce the movement of atoms inside a protein structure, and assume controlled thermodynamic conditions and Newtonian dynamics^{250, 251}. A MD simulation starts by placing a protein structure inside a virtual box filled with solvent molecules, and the system is first left to equilibrate. The protein structure used was typically obtained experimentally, through crystallography or NMR, and contains information on the most likely locations of all atoms inside a folded native state of a protein. Following the equilibration step, one end of the molecule is fixed and the other is pulled using a potential well (and hence the name “steered”) and the molecular force is calculated from the displacement of the pulled atom from the center of the well.

Steered MD simulations can provide remarkable information on how hydrogen bonds form and break and how various secondary structure elements move in respect with each other while a folded protein turns into a polypeptide chain, under a force vector. Unlike the measuring techniques used to study the mechanical response of proteins, MD simulations use point-like probes, and the measured response is not affected by the dynamics and viscous drag of the tethering objects. They can also directly map the load bearing parts of a protein domains and locate the mechanical clamp or load-opposing mechanism of a protein under force. Due to the computational costs of solving the equation of motion numerically in ns timesteps, MD simulations are however limited to time scales of only a few ms, and hence need to use forces (or loading rates) ~10-100x higher than the ones applied experimentally, or to utilize coarse-grained approximations. Furthermore, the force fields rely on various assumptions (parameters) and simulations done for the same protein but using different software packages or graining resolution can produce slightly different results^{252, 253}. Additionally, different barriers might be sampled under different force loading rates (e.g. streptavidin-biotin interaction described in **Section 4.1**) and these higher loading rates might bias certain unfolding pathways²⁵⁴.

4.5. Concluding remarks on limitations of the measuring techniques to study mechanical response of proteins

In an ideal system, the probe used to sample a process does not affect the sampled process itself. If one shines a light beam to measure the position and velocity of a flying bird, the beam

will influence neither. If one throws a baseball to measure the same parameters, the position might be accurate, but the velocity and moving trajectory of that bird will change after the ball hits it. As mentioned above, all the SMFS methods use macroscopic probes to tether nanoscopic proteins, and there are at least three orders of magnitude difference in size between the tethered protein and of the tethering probes (nm vs mm). This situation is even more exacerbated when using molecular probes or macroscopic approaches, as a cell or the tethering clusters, respectively, would be much larger than the tethered protein.

If we consider the case of a protein unfolding under force, from an energy landscape perspective there are a folded and an unfolded minima, separated by a barrier (see also **Chapters 2 and 3**). On the energy landscape, the location of the unfolded in respect to the folded minimum depends on the final conformation of the unfolded peptide chain. Hence, how much the separation between the two ends of the tether molecule will change during unfolding will depend solely on the number of amino acids forming the structure of the protein and the experienced force (at high enough force, where secondary structure elements are no longer stable; these extensions are well characterized by polymer physics theories – see also **Chapters 2 and 3**). Similarly, as the distance from folded to transition state of most proteins (with a mechanical clamp between α -strands) is within the size of a water molecule ²⁵⁵, the tethering probe will have a limited effect on the measured unfolding kinetics (the path along the end-to-end reaction coordinate is ~ 0.25 nm, within the range of molecular fluctuations, allowing the protein to transverse the barrier even if the probe does not immediately follow suit). However, how fast the polypeptide chain extends along the force direction after crossing the barrier and diffusing toward its entropic minimum, which is several nanometers apart, requires the diffusion of the tethering object along with it (bead or cantilever for SMFS). Because the kinetic behavior of proteins is heavily influenced by the tethering probes, an energy landscape constructed from SMFS measurements will have accurate locations for the folded and unfolded minima, but will produce inaccurate measurements of intermediate states and of the curvature of the barrier separating the folded and unfolded states. Furthermore, a slower diffusion coefficient than that of a freely floating molecule will be needed to sample this energy landscape, reflected by the retarded diffusion of the bead/cantilever ^{164, 192}. If one is solely interested in the physics of the unfolding process, various strategies have been derived to produce an energy landscape that is deconvoluted from the movement of the probes ^{256, 257, 258}. However, it would be hard to validate such strategies, given the fact that the size difference between probes and molecules is so huge. One such validation could come with improved computation from MD simulations. Currently these simulations are done at much higher forces and for much shorter times than SMFS experiments.

So at first glance, it might seem that the scientists wanting to study the mechanical response of proteins are doomed, with the experimental techniques having probes too large (and hence too slow) and simulations being performed in non-physiological time and force conditions. However, one has to wonder how relevant a measurement that removes the effect of the tethers is for our understanding of how proteins function in vivo under force. In vivo, proteins are operating in packed environments and tethered between large molecular or cellular assemblies, and their mechanical response is inextricably linked to the dynamics of their surroundings and tethers. Hence, not only are all the tools discussed in this chapter invaluable in determining how force might influence binding interactions, expose buried reaction sites, change the elasticity of a tissue

due to unfolding, or determine relative stability between different proteins or different geometries, but in our opinion these tools represent a better window into the behavior of proteins in vivo. These features play an important role in making force a key driver for transducing mechanical signals into chemical signals and vice versa. In the final chapter we will follow some biological systems, which will further clarify these concepts.

That's a Wrap

- There are three main single molecule force spectroscopy techniques to study the mechanical unfolding of proteins: magnetic tweezers (MT), atomic force microscopy (AFM) and optical tweezers (OT).
- Attachment chemistries and protein engineering play a critical role in the success of measurements involving tethering of single molecules.
- MT is ideally suited for low forces (<100 pN) and transitions occurring on (10 ms – 1h); AFM is best for higher forces (20pN- 10nN) and fast transitions (1ms – 10s); OT operates best at low forces (<65 pN) to measure fast transitions (1ms – 10s).
- In vivo methods based on fluorescent molecular probes, and tissue-like approaches, based on protein-materials, can describe the mechanical response of proteins in crowded environments.
- Molecular dynamics (MD) simulations produce mechanistic views of the mechanical unfolding and protein binding, but are currently limited to short times (microseconds).

Read These Next

Popa, I., Rivas-Pardo, J. A., Eckels, E. C., Echelman, D. J., Badilla, C. L., Valle-Orero, J. & Fernandez, J. M. A HaloTag Anchored Ruler for Week-Long Studies of Protein Dynamics. *Journal of the American Chemical Society* **138**, 10546-10553, (2016).

Bustamante, C. J., Chemla, Y. R., Liu, S. & Wang, M. D. Optical tweezers in single-molecule biophysics. *Nature Reviews Methods Primers* **1**, 25, (2021).

Bender, R. & Salaita, K. *Molecular Force Sensors*. ACS Focus (American Chemical Society, 2021).

Chapter 5. How cells and tissues use mechanical unfolding and refolding as a gain-of-function

5.1. Key characteristics related to mechanical unfolding and refolding of proteins in vivo

Mechanotransducing proteins are typically part of large protein assemblies and have their ends tethered and capable of moving in response to a force vector. To represent a gain-of-function, the mechanical unfolding and refolding of proteins needs to convert mechanical work into a biochemical signal or in some kind of energy transfer. The main known gain-of-function mechanisms can be summarized as described from the perspective of

1. Exposure of buried reactive site
2. Energy storage and release
3. A quantized response
4. Fine-tuned unfolding response through ligand binding ¹.

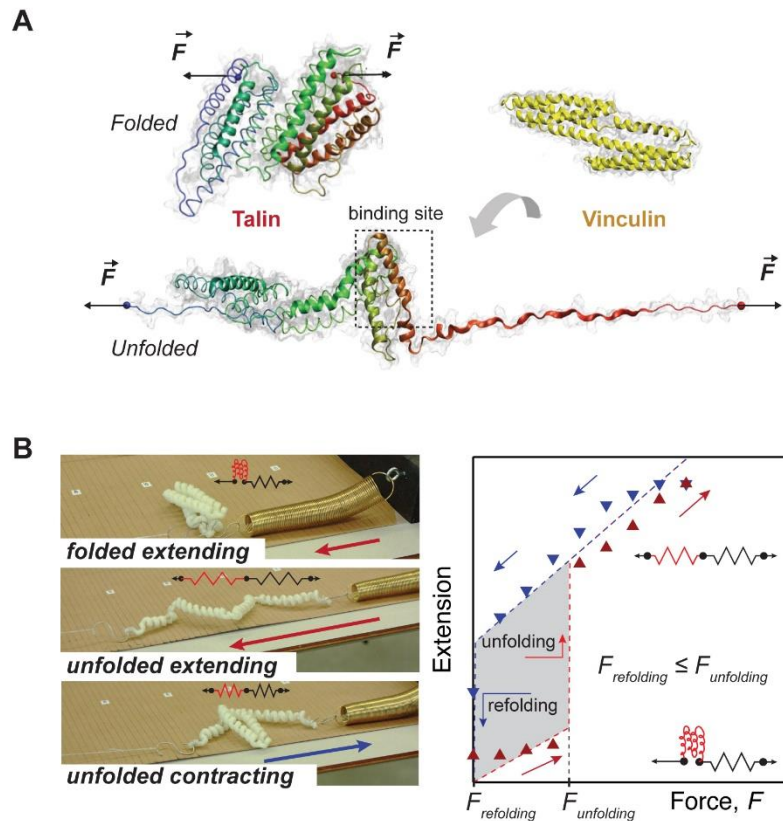
All these characteristics accompanying mechanical unfolding of proteins are then put in the context of selected physiological systems, in the second part of this Chapter.

5.1.1. Exposure of buried reactive site

Extension of the peptide chain following unfolding can expose to the environment reactive sites buried inside the folded structure, which in turn can result in either (i) recruitment of other proteins or (ii) posttranslational modifications. (i) When force-induced binding takes place, the buried site acts as a substrate for a ligand that is typically already present in the proximity of the tethered protein (**Figure 5.1A**, this mechanism will be discussed in greater detail in **Chapter 5.3**). Hence, through force-induced binding to otherwise buried binding sites, unfolding can trigger recruitment of other proteins in what resembles a feed-back mechanism that allows cells to match the strength of their connections. (ii) When unfolding exposes a reactive amino acid or disulfide group to the environment, a chemical reaction can take place. For example, a protein domain having a buried disulfide bond can unfold and extend up to that disulfide bond, as the force applied (typically in the pN range) is not sufficient to mechanically break S-S bonds (which break in the nN range) ^{259, 260}. However, the presence of reducing molecules (such as glutathione) can chemically cleave disulfide bonds, effectively increasing the contour length (and hence extensibility) of protein domains. Several post-translational modifications can also take place when liable amino acids are exposed to the medium. For example, it was shown that reaction of thiols can lead to their reaction with glutathione ²⁶¹. Due to steric effects, the presence of a chemical group attached at a cysteine that is typically buried deep inside the folded structure, results in either a reduced folding rate and mechanical stability, or prevents refolding all-together

²⁶¹. Hence through chemical modifications of a buried site, a protein domain can produce a change in the contour length, and hence final extension at a given force.

Figure 5.1. Mechanical unfolding as a gain-of-function in vivo.



A) Representation of the folded structure (top) and an unfolded state (bottom) of talin rod domains (colored molecule) in the presence of vinculin (yellow molecule), which binds only to the unfolded and extended conformation of talin. Taken from ref. 262 with permission. B) Model showing the extension and contraction under force of a protein in series with an unstructured element (represented by a spring). At low forces (below $F_{unfolding}$) the extensibility is given almost entirely by the extension of the spring element, while above $F_{unfolding}$ the unfolded peptide chain transforms into a second spring element. Figure adapted from ref. 263.

5.1.2. Energy storage and release

When fast adaptation of elasticity is required, such as is the case for muscles or attached bacteria (described in detail in the second part of this chapter), unfolding can provide an immediate energy dissipation mechanism and produce a change in extension followed by a decrease in experienced force through the release of hidden length (as the contour length of an

unfolded peptide is several times larger than that of a folded domain)²⁶⁴. As described in previous chapters, unfolding entails breaking of the hydrogen bonds holding the tertiary structure together, and, when under force, it is also accompanied by the entropic extension of the peptide chain. This extension can be tens of nanometers between the two ends of an unfolded protein, and under force the chain can be too stretched to allow for the amino acids to come together to form hydrogen bonds, which are short ranged. For amino acids to come together, the polypeptide chain needs to first collapse, which takes place at relatively low forces (typically well below 10 pN, **Figure 5.1B**). For this reason, for many proteins that are mechanically stable and unfold at tens or hundreds of pN of force, the refolding force is significantly lower than the unfolding force (e.g. I91 of titin unfolds at ~200 pN²⁶⁵ and refolds at ~6 pN¹⁷³). In this case, in an extension versus force representation, a significant hysteresis is seen between the path taken when the force is increased and that taken when the force is decreased (illustrated also by the grey area in **Figure 5.1B**). This hysteresis represents an energy dissipation mechanism. This dissipation mechanism plays key roles in avoiding damage due to force transients, as domains act as semi-sacrificial dumpers by adding length to a polypeptide tether and reducing the overall tension on the molecule. When proteins have stabilities in the <10 pN force range, the unfolding and refolding forces may be similar, and both transitions can be seen at the same force. In this case, the folding and unfolding transitions take place without loss of energy and refolding of a protein against a force vector can play another important role, that of performing mechanical work. (We will discuss this aspect in detail when talking about muscle contraction).^{262, 263}

5.1.3. A quantized response

The typical architecture of globular proteins operating under force is that of being segregated into multiple domains (polyproteins), resembling beads-on-a-string. *In vivo*, each domain has a slightly different sequence and mechanical stability. This architecture (i) helps avoid misfolded states, as domains can sequentially fold while they are being expressed by the ribosome; (ii) allows the protein to act as a molecular battery, as folding represents a mechanism of storing energy and compacting the polypeptide chain (as discussed at the point above, a folded structure is more compact than a collapsed polymer); and (iii) can produce a quantized response. This quantized response can be visualized similarly to the ones and zeros of a computer algorithm, with one being the folded state and zero the unfolded state of each domain⁵. The quantized response, where domains - which typically have different stabilities - unfold in response to force buildup and trigger different signaling events, allows for a fine-tuned molecular behavior based on the changing biomechanical environment.

5.1.4 Fine-tuned unfolding response through ligand binding

Binding of ligands to protein domains can act as a force rheostat, by altering their mechanical stability¹. Typically, binding of ligands and small ions induces an increase of mechanical stability of a domain, as it caps the binding site which otherwise has exposed hydrophobic amino acids

that decrease its stability. Similarly, when binding takes place to an unfolded state (as discussed at point 1 for talin), refolding probability decreases¹⁶. Hence the molecular program (as discussed at point 3 just above) can be fine-tuned without the need to re-express new proteins, but rather through binding to folded or to unfolded states.

Note that, as also discussed in previous chapters on diffusion on the energy landscape under force, and explained in **Chapters 3 and 4**, mechanical unfolding and refolding are non-equilibrium probabilistic events driven by thermal motion, that depend not only on the value of the experienced force, but also on how fast the force is applied. A protein will unfold at a higher force if the force is increased faster, as detailed in **Chapter 3.2**. Similarly, a very stable protein can unfold at a constant low (or even zero) force, but unfolding might take a longer time than its turnover. When we describe here an unfolding force without associating it with a time or loading rate dependency, we mean that the force was increased linearly with time at a rate relevant for physiological processes, or that unfolding took place on a physiologically-relevant time scale.

5.2. Muscular contraction and protein refolding-induced function

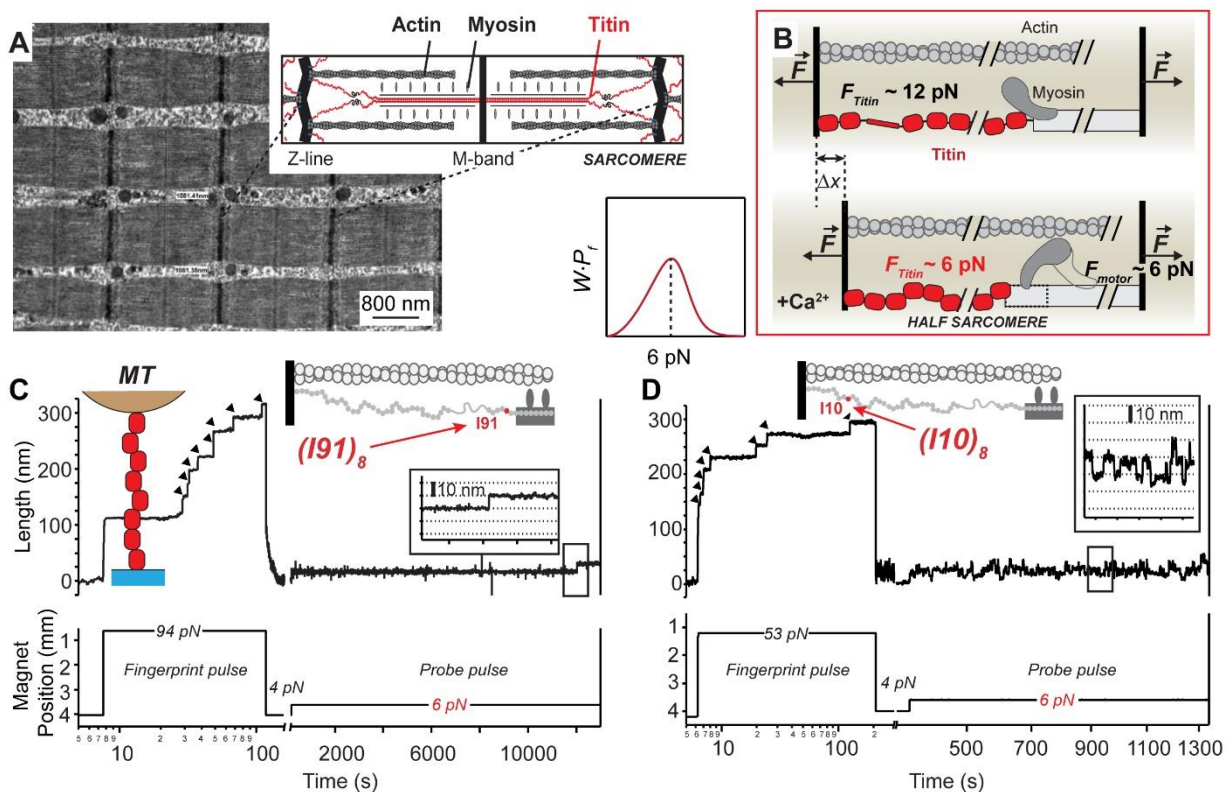
Muscles are linked to bones via tendons and are responsible for body locomotion, posture and blood circulation. They are divided in three categories: skeletal, cardiac, and smooth. While having different cellular arrangements, all types of muscles rely on transducing electrical signals from motoneurons into contractile motion. Among the three groups, the skeletal muscles allow for voluntary control. These muscles have a well-aligned structure along the direction on which they operate (**Figure 5.2A**). Sarcomeres are the smallest functional units of muscles, and are composed of three main filaments: actin, myosin and titin. Actin forms fibrous assemblies both in muscles and in cells, which have double helix-like structures made from the interwinding of two protein strains. Myosin motors can perform power strokes to engage and move on actin filaments, enabling the shortening of sarcomeres and driving contraction. The power stroke requires chemical energy in the form of ATP and is controlled through the diffusion of Ca^{2+} ions (**Figure 5.2A and B**). Spanning along each half of the sarcomere is titin, the largest protein in the human body. Titin is formed of two unstructured regions, resembling molecular springs, and over 100 Ig-like domains. Each muscle has a different titin *isoform*, which varies in length and number of Ig-like domains, from ~2.2 mm in cardiac muscle to ~3 mm in *soleus* muscle^{21, 266, 267}. Titin's main role was thought to be that of a molecular spring, and until recently it was unknown if any of its Ig-like domains can unfold at physiological forces. The reason for this unknown came from the fact that one of the most used proteins to study mechanical unfolding by the force spectroscopy community is titin I91, which is a very stable domain (it unfolds in force extension at ~200 pN²⁶⁵, well above the physiological range, and it takes ~3.5 hours for 1/8 domains to unfold at 6 pN¹⁷³ – see **Figure 5.2C**). Hence, from the perspective of gain-of-function discussed in the beginning of this chapter, the large difference between the unfolding and refolding forces makes I91 an ideal domain for energy dissipation, that would only unfold rarely, potentially avoiding catastrophic failure.

The recent discovery that Ig-like domains from the proximal region of titin can have equilibrium unfolding-refolding transitions at physiological forces (<10 pN per molecule), without

energy loss through heat, has changed the paradigm on how titin operates in the human body (**Figure 5.2D**)¹⁷³. From this perspective, unfolding of titin domains takes place during muscle contraction, while myosin motors are not engaged and titin is experiencing the entire tension (**Figure 5.2B top**). At this stage, the force-per-titin was estimated to be ~ 12 pN²¹, which is too high to allow for refolding. As myosin motors engage actin, the force-per-titin decreases, as a single myosin motor generates 4-8 pN¹⁷³. The remaining force range is between 4-8 pN allows for titin domains to refold. In this range, the folding of one titin domain under force can deliver even more **contractile energy** than the power stroke of a myosin motor²¹ (**Figure 5.2B bottom**). While the contractile energy will be the highest at the upper force range (as the decrease in length is higher), its probability of refolding is also the lowest. Hence a better way to estimate the range where titin refolding has the largest impact is through the product between folding probability P_f and contractile energy W delivered by refolding. Through this approach, it was estimated that titin refolding has the largest impact on muscle contraction at ~ 6 pN¹⁷³ (**Figure 5.2B Inset**).

Titin domains, regularly unfolding and refolding under physiological working force range, opens new avenues for understanding muscle regulation. For example, by changing the number of domains unfolding and refolding during contraction, muscles can quickly adapt their sarcomere length and energy delivery, when one needs to go (for e.g.) from a standing to a running posture. Also, buried titin sites that can be exposed through mechanical unfolding and lead to a long-term change in the overall length of the sarcomere through posttranslational modifications that would impair refolding²⁶⁸.

Figure 5.2. Titin refolding delivers energy during muscle contraction.

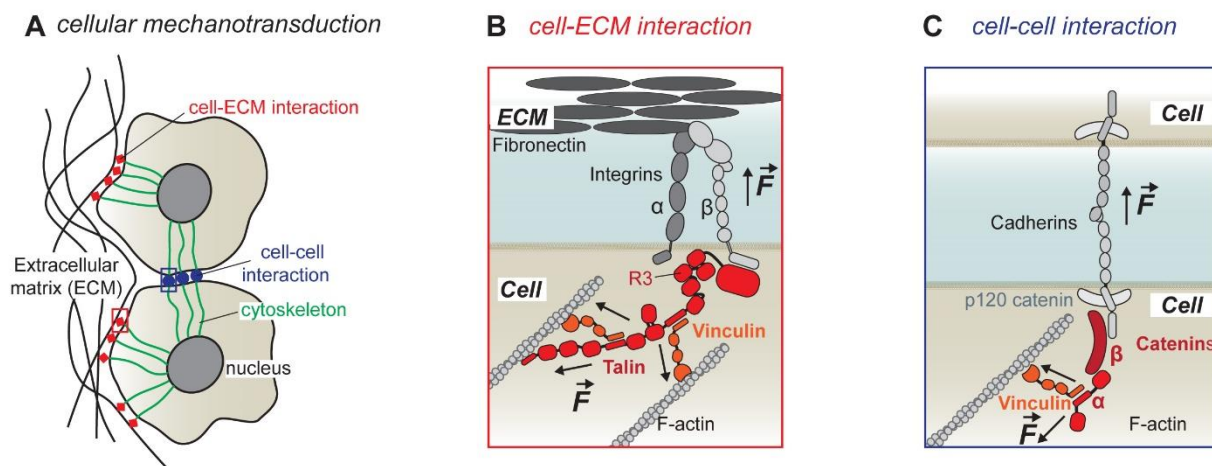


A) Transmission electron microscopy image of a striated muscle biopsy, taken from Rayat et. al, Austin J Pathol Lab Med. 2016; 3(1): 1016. Inset: Schematics of a muscle sarcomere, showing the main filaments; taken from ref. 1; B) Schematics of how titin unfolding and refolding operates during muscle contraction, adapted from ref. 1. When myosin motors do not engage actin, titin experiences the largest force (~ 12 pN)^{21, 173}. As myosin motors perform power stroke movements, the force-per-titin reduces and at ~ 6 pN it delivers the largest contractile energy through refolding. Inset bottom left: diagram of the product between contractile energy W and folding probability P_f as a function of force for a single titin domain.¹⁷³ C) and D) Magnetic tweezers traces showing first the unfolding of an octameric construct made from titin domains I91 and I10, respectively. Insets show schematics of the measuring system, the relative position of the two domains in the sarcomere and zoomed-in views of unfolding events.

5.3. Cellular mechano-transduction regulates dynamics of the cellular cytoskeleton

Cells receive and send chemical and mechanical cues, which allows them to regulate their function, to divide or to enter [apoptosis](#). Chemical signaling typically has a more global effect (for example, secretion of the adrenaline hormone acts on nearly all cells), while mechanical communication is highly localized to a cell's neighbors, which can be other cells or their extracellular matrix (ECM) environment (**Figure 5.3A**).

Figure 5.3. Mechanotransduction regulates how cells function.



A) Schematics of cell-extracellular matrix (ECM) and cell-cell interactions, which are coupled through the cellular membrane all the way to the nucleus through the cytoskeleton; B) Schematics illustrating how mechanical unfolding of talin reinforces the connection between the actin filaments and the integrin-ECM complex through recruitment of vinculin to the unfolded states of talin rod domains; adapted from ref. 1; C) Schematics of the mechanical response of catenins unfolding and binding vinculin to drive cell-cell connections via cadherins; adapted from ref. 1.

The ECM consists of secreted proteins and is reshaped by and guides the growth of the hosted cells ²⁶⁹. There are over 50 different proteins actively assembling and disassembling the connections between a cell and its ECM ²⁷⁰. The large majority of both extracellular and intracellular proteins involved in **mechanotransduction** are segregated into multiple domains ⁵. Cells have developed mechanisms to sense the stiffness, pattern and movements of the ECM and to respond accordingly ⁴⁹. An important mechanism used by these proteins to transform mechanical forces into chemical cues involves exposure of cryptic sites ²⁷¹. For example, extracellular fibronectin, which has over 50 repeating subunits and self-assembles into fibrils, can respond to a pull force and expose sites formed of three amino acids: arginine-glycine-aspartic acid (also known as RGD, from the letter alphabet of these amino acids) ^{272, 273}. Exposure of RGD sites can trigger recruitment of integrins, which are a transmembrane family of proteins formed from an α and a β subunit ²⁷⁴ (**Figure 5.3B**). Integrin activation produces one of the largest molecular movements, as the cytosolic sides of the α and β subunits move apart ~ 7.5 nm ²⁷⁵. The exposed β subunit of integrin activates talin, a multidomain protein which then attaches to an opposing end to actin filaments. Talin has 13-rod (R) domains which contain 11 buried sites that can bind vinculin when exposed through mechanical unfolding ^{276, 277}. Talin acts as a molecular computer ⁵, which integrates the increases in force of the cell with its ECM by unfolding its R-domains and binding vinculin, to either reinforce current actin connections or recruit new actin filaments (**Figure 5.3B**). Like in a tag-of-war, the extracellular forces are matched internally through dynamic re-arrangement of the cellular cytoskeleton.

When cells interact with other cells, they form an extracellular homo-dimeric complex via cadherins, a linkage driven by the presence of Ca^{2+} ions ⁵⁹. The intracellular side of cadherin binds to actin filaments through the α/β catenin complex, through a linkage mediated by p120-catenin (**Figure 5.3C**). As force develops, α -catenin can too unfold and expose a binding site for vinculin, which can add one more reinforcement to this connection ^{278, 279}.

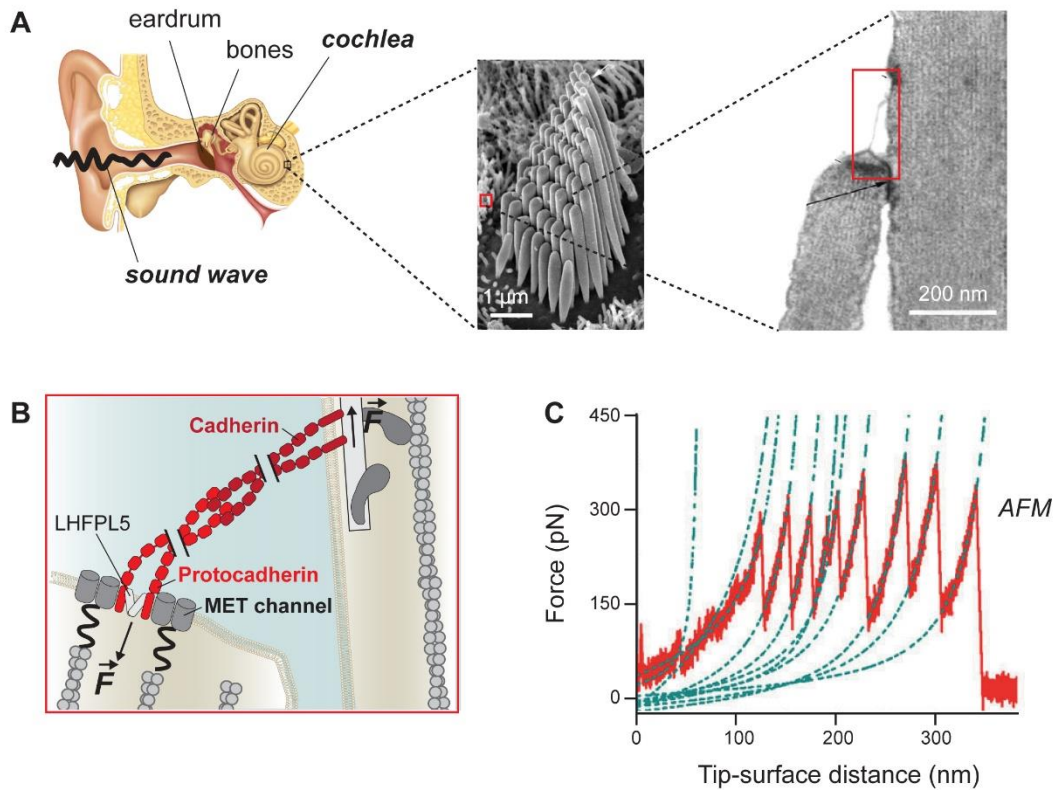
As described above, the same protein, vinculin, acts as a network reinforcer for both cell-ECM and cell-cell connections ²²¹. While up to eleven vinculin connections can be formed by talin during cell-ECM interactions, a single vinculin is recruited at the catenin site for cell-cell interactions. Force plays a dual role for vinculin recruitment. For example, it was shown that at low forces vinculin binding can induce a coil-2-globule transition in the α -helix of talin containing the binding site (as discussed in **Chapter 1.2.2**), while higher forces can even lead to expulsion of vinculin and breaking of the complex ¹⁶. The mechanical stability of talin domains is also regulated through binding. For example, one of the most stable connections of talin to one of its partners, DLC-1, keeps talin R8 domain in a folded conformation, preventing both binding of vinculin, as well as recruitment of talin to the plasma membrane ²⁸⁰.

5.4. How cells in our ears transform pressure waves into perceived sound

Sound travels as pressure waves through the auditory canal of the ear, then reaches the eardrum, which produces vibrations through three tiny, connected bones (**Figure 5.4A**). These

bones amplify the eardrum vibrations before they travel to the cochlea. The cochlea has a snail-like shaped structure with ~2.75 turns around its axis and is decorated with hair cell bundles connected to the hearing nerves and immersed in fluid. These hair cell bundles, called stereocilia, vibrate at specific sound frequencies, depending on their position inside the cochlea (progressively lower pitched vibrations are detected as sound travels through the cochlea). Their movement turns vibration waves into electrical signals by opening ion channels that generate an ionic flux ²⁸¹. Essential for sound transmission and sound amplitude regulation is a filamentous link which is formed by two homodimer proteins, cadherin 23 and protocadherin 15 (**Figure 5.4B**). As discussed in **Chapter 5.3**, cadherins are multidomain proteins used to establish Ca^{2+} -dependent cell-cell connections. The two proteins form a dual molecular handshake ^{282, 283} and are responsible for maintaining the tension between the bundle cells. Structure analysis suggests a stiff coupled force transmission ²⁸⁴, while the length of the tip was measured to vary in situ ²⁸⁵. Under tension, waves make the cells vibrate leading to opening of ion channels MET/LHFPL5, which generate electrical currents²⁸¹. So how can both, a force coupler state and a length-changing state, co-exist within the same molecular assembly? The reader can probably already make an educated guess: it must be related to the folded states of the domains forming the cadherin-protocadherin connection. As explained in the beginning of this chapter, folded domains will produce a stiff connection, while unfolded domains will dissipate mechanical energy while changing the connection length. When a transient loud sound excites the bundle, domain unfolding can adsorb the mechanical perturbation through unfolding, while refolding will restore the molecular tension. But what if our ears are continuously exposed to loud sounds? Then the molecular link will allow some sliding between the two cells, which will decrease their sensitivity. So going to a loud concert makes us less sensitive to sound! However, the sensitivity is regained over time, as myosin motors move on actin filaments to restore the resting tension between stereocilia (**Figure 5.4B**). This process allows us the next day after a loud concert to hear just fine! Hence, protein refolding and myosin-actin interaction allows for certain plasticity for sound perception ²⁸⁶.

Figure 5.4. Mechanical regulation in hearing.



A) (*left*): Schematics of the inner ear vestibule, which contains receptors for balance, and cochlea, which contains the hearing nerves; (*middle*): Scanning electron microscopy (SEM) image of hair cells bundles from cochlea, arranged in a staircase-like pattern, with each cell increasing in height across the bundle; image taken from ref. 287 with permission; (*right*): Transmission electron microscopy (TEM) image of the tip link complex, formed by a protein connection maintain tension between two auditory hair cells; image taken from ref. 288 with permission; B) Representation of the cadherin-protocadherin complex, regulating the tension between hair-cells. C) AFM traces of the cadherin (Cdh23) and protocadherin (Pcdh15), showing a saw-tooth like pattern in the measured force as the complex extends and unfolds its component domains, measured at a loading rate of 10^5 pN/s. Taken from ref. 289,

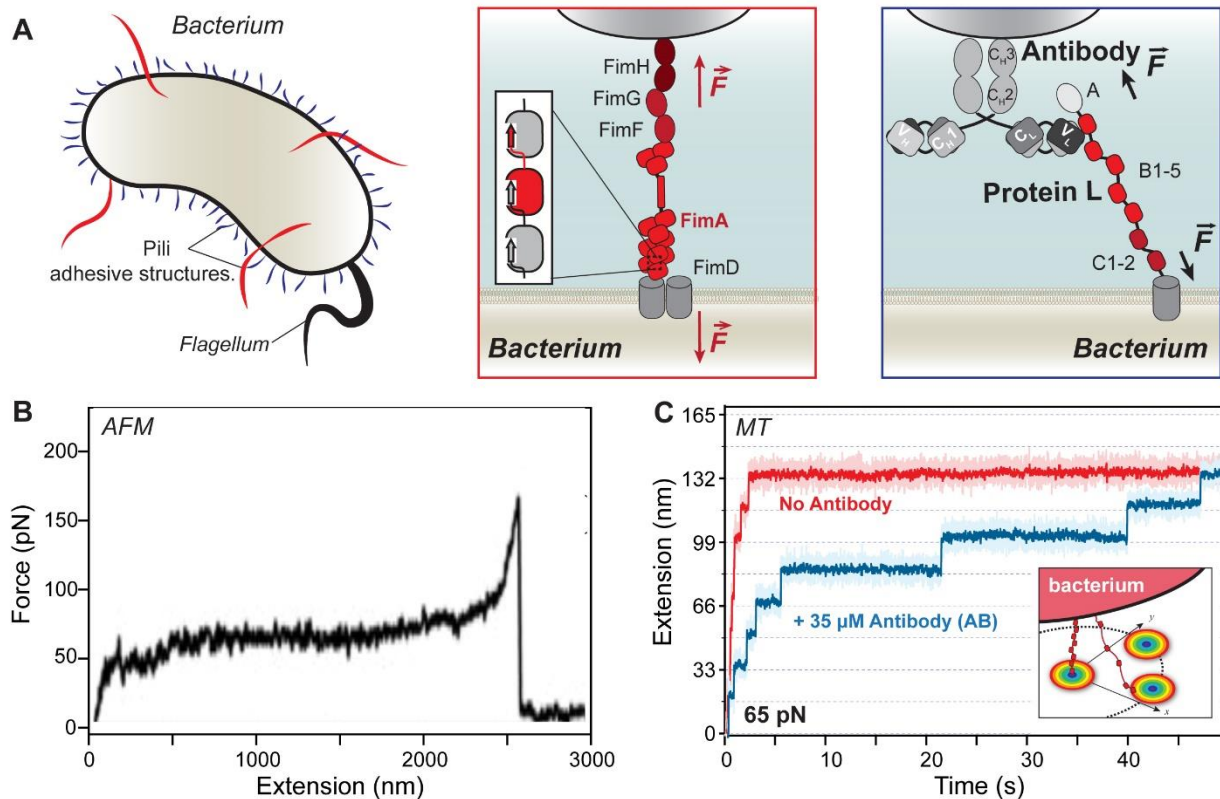
Two recent single molecule studies measured the unfolding under force of the cadherin 23 and protocadherin 15 domains in force-extension mode (**Figure 5.4 C**)^{289,289, 290}. In one study done with AFM, the average unfolding force of the domains forming the connecting proteins was found to be between 300-400 pN²⁸⁹, while in a second study done with OT, the force was 20-50 pN²⁹⁰. The discrepancy is only apparent as explained in the first part of this chapter, how fast the force is increased (loading rate) influences the measured unfolding force; the AFM study used loading rates of $\sim 10^5$ pN/s, while the OT employed loading rates of $\sim 10^2$ pN/s. The loading rate in vivo will obviously be determined by the coupling frequency of each cell bundle, while the force will be given by the amplitude of the incoming sound wave. Hence, this increase in mechanical

stability of proteins with force loading rate can itself represent a regulation mechanism for the elasticity of the cadherin-protocadherin connection.

5.5. Force-regulated attachment of bacterial adhesion

Bacteria form dynamic connections with their host tissues, as they need to navigate the high shear forces of the mucus, or withstand transient forces as those developed during air or fluid flow (such as those generated by coughing or urination)²⁹¹. Typically, bacteria have a flagellum, used for swimming and locomotion, and secrete a variety of **pili** and other adhesive molecules, used for colonization (**Figure 5.5A**). The mechanisms used by bacteria to attach to their target are extremely diverse. Here we will only explore two systems.

Figure 5.5. The mechanics of bacterial adhesion.



A) (left): Schematics of bacterium; (middle) Schematics of type I pilus. Adapted from ref. 1. (right): Schematics of secreted antibody binding multidomain protein L, which can attach to its target at the light chain region through domains B1-5; adapted from refs. 1, 292; B) AFM measurement of the system from panel B, measuring the unraveling of a type I pilus as a force plateau at 20-60 pN while having >1mm long extension; this plateau allows bacteria to dissipate energy and avoid dislocation²⁹³; Inset: reconstructed image of bacterium binding antibodies – taken from ref. 294

with permissions. C) MT traces of eight domains of protein L in the absence and presence of antibody ligands. Inset: mechanism describing how the unfolding response can be used by the bacterium to optimize its antibody cluster search. Adapted from ref. 292 with permission.

A system known as type I pili is found throughout a family of an *E. coli* bacteria known to cause bladder infections. The secreted pili play a key role in initiating infection and maintaining the bacteria attached to its host cells to avoid clearance during urination²⁹⁵. Each type I pilus has a flexible fibrillar tip made from three protein domains (FimF, FimG and FimH), which is joined to a rigid rod-like structure composed of thousands of repeating units of the same protein domain, FimA^{296, 297} (**Figure 5.5B**). The FimA rod is 6.9 nm thick and 1-2 mm long²⁹⁸. All the Fim domains, including the FimA rod, are assembled through a β -strand complementation mechanism, which is realized by having the terminal amino acids of one protein domain inserted as a β -strand in the structure of the following domain^{299, 300} (**Figure 5.5B inset**). The FimH tip has a two-domain structure and can bind mannose receptors, found on the surface of the host cells³⁰¹. During a transient force, the quaternary structure of the rod domain is the first to yield and displays an elongation plateau at forces between 30-60 pN of up to 700% their initial length^{293, 302} (**Figure 5.5C**). Through such a large extension, the bacterium can dissipate a large amount of mechanical energy to maintain its contact with the host cells¹. Interestingly, if higher forces were to be developed, the Fim proteins can themselves unfold in a well-defined sequential manner²⁹⁷.

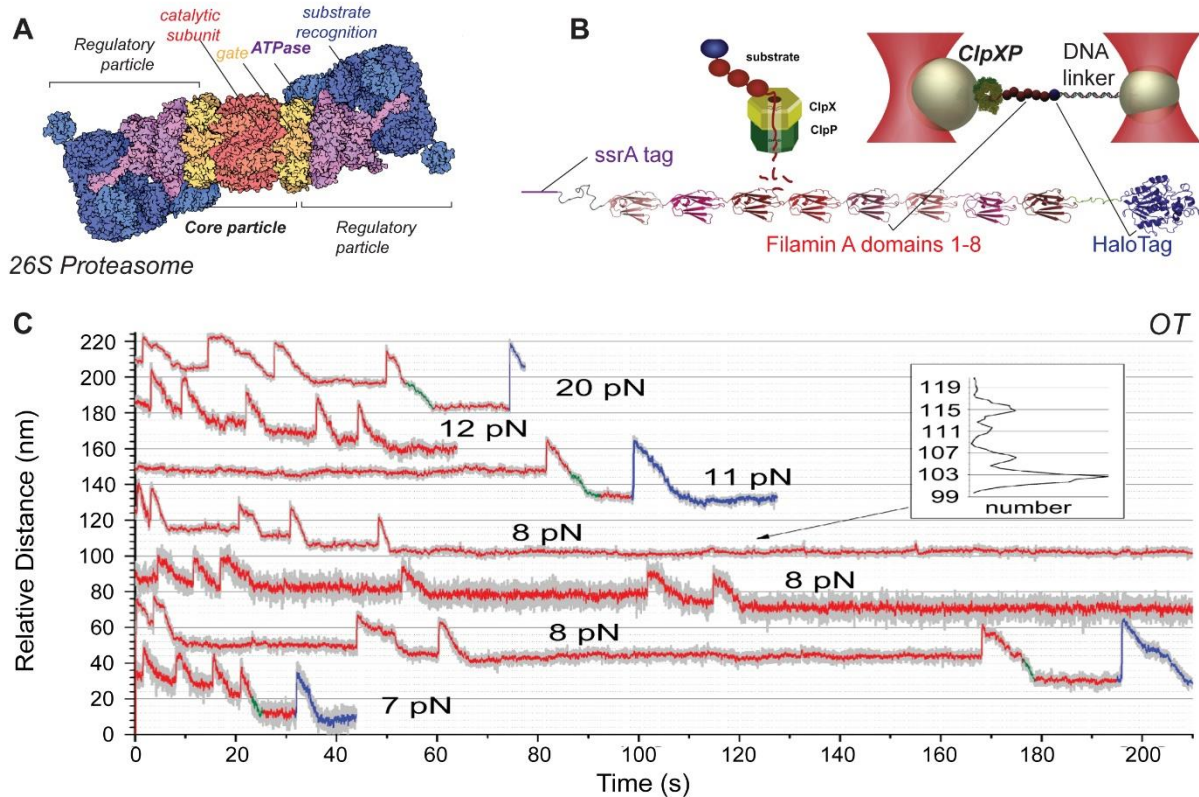
A second system that we will discuss here is of a class of bacteria that can secrete pili that bind antibodies – known as antibody binding proteins (**Figure 5.5D**)^{303, 304}. Antibodies are an important part of the immune response, and bind to pathogens to tag them for clearance. Antibodies are either on the surface of a certain type of white blood cells (called B cells) or in soluble form. They have a Y-shape and recognize a distinct pathogen molecule, called antigen. By secreting antibody binding proteins, some bacteria can protect themselves from being tagged by antibodies (**Figure 5.5D**). In this case, the antibody which ‘hunts’ for pathogens becomes the hunted. One such example is *Finegoldia magna*, a bacterium which colonizes skin, oral cavities, implants and mucous membranes^{305, 306}. This bacterium secretes a polyprotein L, which has several domains that bind to antibodies (titled B-domains)³⁰³ (**Figure 5.5D**). While the B domains of protein L have a very simple a-b structure, each can bind two antibodies³⁰³. One of the binding sites of protein L has a high affinity ($K_d \sim 150$ nM), while the other has a lower binding strength ($K_d \sim 25$ mM)^{303, 307}. Using single molecule MT, it was shown that protein L has a ~ 2 fold increase in mechanical stability when binding to antibodies²⁹² (**Figure 5.5E**). Interestingly, the mechanically stable state was induced by the binding of the second antibody, at the lower affinity interface. This discovery points to a mechanism where the bacterium uses its high affinity interface for attachment and low affinity site for tuning its search radius²⁹². When attached to an antibody cluster, the B domains of protein L would stay folded, allowing the bacterium to focus its pili toward that cluster (**Figure 5.5E inset**). When binding to a region with low antibody concentration, only the high-affinity site would engage, while allowing protein L to unfold and extend the search area of the bacterium under flow²⁹². Through unfolding, the search radius can be increased by ~ 15 nm per domain²⁹².

5.6. Degradation of proteins requires mechanical unfolding of the tertiary structure

In **Chapter 1**, we discussed various types of unfolding mechanisms (mechanical, thermal, chemical) and we emphasized that a protein will have a different stability for each type of perturbation. This difference can explain the existence of organisms living in hot springs, in temperatures close to that of boiling water: their proteins must be stable enough to operate in high temperature, but unstable enough to be degraded and turned-over. Such organisms would not exist if the mechanical and thermal stability of their proteins would be the same, but their proteins must have high thermal and regular mechanical stability. This example points to the fact that the unfolding of proteins preceding degradation by the proteasome is mechanical, as it we discuss next.

The proteasome is the molecular assembly responsible for the unfolding of structured proteins followed by the chemical breaking of their peptide bonds³⁰⁸. It plays an important role in regulating protein concentration inside a cell and in removing non-functional proteins (such as misfolded structures, that are not viable). It is formed of a core particle and one or two regulatory units^{309, 310}(**Figure 5.6A**). The regulatory units recognize and bind proteins tagged for degradation, remove the ubiquitin degradation tag and use six connected ATPase motor proteins (hexameric ring) to unfold and linearly translocate proteins into the degradation chamber. The core particle has a gate that controls the access to the catalytical chamber, and three proteolytic sites inside the catalytic chamber, that cleave the polypeptide chain into smaller fragments^{309, 310}.

Figure 5.6. The mechanical unfolding and chemical degradation of proteins by the proteasome.



A) Structure of the yeast proteasome solved by hybrid methods ³⁰⁹, taken from <https://pdb101.rcsb.org/motm/166>. B) (left): schematics of the ClpXP-filamin system; (right): schematics of the experiment; (bottom): the polyprotein construct used, which contains a recognition tag, eight filamin domains and a HaloTag. C) Single molecule OT traces monitoring the change in extension between a bead with an immobilized ClpXP system and a polyprotein made of eight repeats of filamin A and a HaloTag, attached to the second bead through a DNA linker. Inset: spacing of extension plateaus. Adapted from ref. 3 with permissions.

The analogous proteolytic system in bacteria, ClpXP, has a central proteolytic unit (made from two homo-heptameric rings of ClpP), flanked between one or two hexameric rings at the ends, ClpX ³¹¹ (**Figure 5.6B inset** shows the schematics of one ClpX attached to ClpP). The ClpX acts as an ATPase motor, which unfolds protein domains and controls their translocation in the proteolytic ClpP chamber. The unfolding kinetics and translocation velocity of ClpXP were measured using OT ³. The ClpX ATPase motors continuously pull the attached protein domain into the proteolytic chamber through a small pore of ~3 nm in diameter ³¹². As the pore is too narrow to fit folded protein domains, unfolding must occur first. Following unfolding, the peptide chain immediately extends due to the small force vector (7 to 20 pN) applied during the experiment (**Figure 5.6B**). This extension is measured as a step increase in the overall extension. Following unfolding, the ATPase translocated the peptide into the proteolytical chamber, which was degraded and released as small peptide fragments. Note that at forces higher than 33 pN the ATPase was found to stall, which is a good estimate for the equivalent peak pulling force ³ and

that there is ~4 nm difference between the end-to-end distance before and after each unfolding-translocation event (Inset **Figure 5.6C**), which is the size of a folded protein domain. Without any force, the translocation velocity was ~32 amino acids per second ³. Once a protein domain was unfolded and degraded, a linear plateau was measured, until the ATPase motor managed to unfold the following domain (**Figure 5.6B**). Important here, the plateaus before and after the unfolding and translocation are the size of a folded protein domain (~5 nm).

5.7. Concluding remarks

Looking through the magnifying lens of force-spectroscopy measurements, everything in biology may now seem mechanical. Many articles and textbooks, on the other hand, focus entirely on biochemical reactions, where different proteins interact with each other based on their concentrations and affinities, and form complicated signaling cascades. Here we tried to emphasize novel concepts that have so far been ignored or overlooked by mainstream books. These concepts have at the center the mechanical connections formed between proteins inside a cell or crossing the membrane. Unlike chemical reactions, mechanical connections are highly localized. Hence the mechanical environment for a cell is represented by its neighboring cells and surrounding extracellular matrix; similarly, the mechanical environment for a protein is represented by its neighboring molecules, which are typically at the termini of that protein. To have an adequate response to force, many of the proteins (including those discussed here) have evolved in multidomains (polyproteins), an architecture resembling that of beads-on-a-string. This architecture is not only important for folding, as domains can fold independently of each other while still being produced by the ribosome (and avoid misfolding), but it is important for their force response. Typically these domains have similar (but not identical) structures, with a similar number of amino acids and comparable folding forces. However, they tend to have slightly different mechanical stabilities. As discussed above, this slightly different mechanical stability produces a segregated response, with the weaker domains unfolding first. However, all domains inside a given protein tend to refold at the same force ¹⁷³.

One way of looking at polyproteins under force is through their folded or unfolded states, which can be seen as the ones and zeros of a mechanical computation unit. If chemical signaling is controlled through the expression and turn-over of various protein components along a signaling cascade, the proteins operating under force are too large to be controlled through turn-over (e.g. it takes more than one hour for muscle cells to make one titin molecule and many more to degrade it ³¹³). When mechanical signals play a physiological role, the experienced force acts as a rheostat dial. Force continuously turns the protein domains from folded to unfolded and vice versa, as a computer code turns CPU transistors into highs and lows to perform a computation. As both their synthesis and turn-over are longer than the response elicited through mechanical signaling, proteins under force use the mechanical unfolding and refolding of their domains as an adaptation to immediately perform their signaling roles. This role can be further fine-tuned chemically. We discussed here how binding to a folded domain typically increases its stability, while binding to a site on an unfolded domain might prevent its refolding. Furthermore, unfolding exposes amino

acids that can be posttranslationally modified, temporarily or permanently affecting the mechanical response of that domain to force.

As our understanding evolves on how mechanical forces act as a signaling pathway *in vivo*, we expect to see a better integration between the biochemical and mechanical analysis when trying to understand how the human body works. These advancements have been made possible by the development of novel innovative techniques to study proteins, the importance of which was underlined throughout this digital primer by mentioning the many Nobel prizes awarded to their developers. The study of how cells sense and respond to mechanical signals (also known as mechanobiology) is currently an emerging discipline. We predict that as more scientists become interested in this field of science, we will improve our understanding not only on processes that have force as the central driving perturbation, such as cell development or muscle responses, but also on more intricate signaling cascades and longer plasticity-inducing pathways.

That's a Wrap

- Mechanical unfolding has some unique features for protein activity under force, such as (1) exposure of previously buried binding sites to the solution, (2) addition/subtraction of contour length with immediate change in molecular elasticity, (3) storage and release of energy coming from the fact that the average unfolding force of a protein domain is higher or equal than its average folding force, and (4) can be fine-tuned through binding of ligands, that typically increase the mechanical stability of folded proteins.
- Titin refolding can deliver an energy boost to muscle contraction in the 4-8 pN molecular force range.
- Talin and catenins unfold and extend to recruit vinculin protein and branch the connections that they mediate between cell-ECM and cell-cell, respectively.
- Unfolding of cadherin-protocadherin domains forming the molecular connections between hearing hair cells modulates adaptation to transient or long-lasting sound wave through the force loading-dependent unfolding response.
- Bacteria regulate their attachment length and response to transient forces through mechanical unfolding of secreted protein domains.
- Protein degradation requires mechanical unfolding by an ATPase molecular motor at the entry of the proteosome.

Read These Next

Sharma, S.; Subramani, S.; Popa, I., Does protein unfolding play a functional role *in vivo*? *Febs Journal* **2021**, 288 (6), 1742-1758.

Eckels, E. C.; Tapia-Rojas, R.; Rivas-Pardo, J. A.; Fernandez, J. M., The Work of Titin Protein Folding as a Major Driver in Muscle Contraction. *Annu Rev Physiol* **2018**, 80, 327-351.

Haining, A. W. M.; Lieberthal, T. J.; Hernandez, A. D., Talin: a mechanosensitive molecule in health and disease. *Faseb J* **2016**, *30* (6), 2073-2085.

Persat, A.; Nadell, C. D.; Kim, M. K.; Ingremeau, F.; Siryaporn, A.; Drescher, K.; Wingreen, N. S.; Bassler, B. L.; Gitai, Z.; Stone, H. A., The Mechanical World of Bacteria. *Cell* **2015**, *161* (5), 988-997.

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* **2011**, *145* (2), 257-267.

Glossary

Activation barrier: The minimum required energy that is required for a compound to undergo chemical or structural modification. It is represented as a potential/energy barrier that separates two minima (stable thermodynamic states) in the potential energy/energy landscape.

Amino acid self-ligation: Amino acids have both amine (-NH_2) and carboxyl groups (-COOH) which can react with each other to form a peptide bond (-CO-NH-) and eliminate a water molecule; this reaction can produce polymeric chains ($\text{-NH-C(HR)-CO-})_n$, where R can be one of the 20 natural amino acids; a typical protein has ~ 100 amino acids, and the polymeric chain is also known as a polypeptide, due to the peptide bond linking the amino acids together.

Apoptosis: Is the death of cells which occurs as a normal and controlled regulation mechanism. It is needed to maintain an equilibrium cell population, and, for this reason, it is often called “programmed” cell death. On average, a healthy individual loses ~ 60 billion cells daily, which amounts to less than 0.2% of the total cells. When the programmed cell death is triggered via mechanical cues (e.g., detachment from the extracellular matrix), it is called anoikis. Anoikis prevents inappropriate attachment and migration of cells, which can lead to abnormal growth and cancer.

Arrhenius-like dependency: The Arrhenius equation exponentially relates the rate coefficient to the activation energy. In transition state theory, the Eyring formulation of Arrhenius law describes the activation energy with the Gibbs free energy.

Boltzmann constant: Is a proportionality constant that connects the thermodynamic temperature of a molecular system with its kinetic energy; it is also central to statistical mechanics, as it is used to connect microstates of equivalent energy with the macroscopic entropy of the system; due to its importance and to the fact that Boltzmann spent most of his years defending, the formula for entropy (Eq. 1.2) is inscribed on the tombstone of Boltzmann’s grave.

Cadherin: A family of large transmembrane proteins (proteins that span across the cell membrane), which is involved in cellular adhesion and their function depends on calcium ions.

Catch-bond: Is a non-covalent bond that increases stability with applied force, typically due to the formation of new hydrogen bonds under force. There is a force range where catch-bonds have maximum stability and were first discovered for rolling leukocytes (immune white blood cells). The leukocytes use this mechanism to avoid clogging of capillaries in normal flow, while binding in high flow, such as that generated by a wound.

Chaotropic: Defined by hydration theory, chaotropes (water disorder-makers), are weakly hydrated soft ions that by making the solution a better solvent, induce unfolding to increase the surface area of the solute in contact with the solvent. This effect is achieved by disordering the water molecules in their vicinity, thus facilitating the hydrations of the proteins in the solution.

Chaperone (protein): Member of a group of proteins which assists other proteins during folding and reduce protein misfolded states (acquisition of a non-native functional structures); some chaperones are expressed by cells during increases in heat, when the body temperature rises to fight an infection; in this case, the expressed chaperones (which are known as heat-shock proteins) mitigate the effect of thermal unfolding of proteins, which otherwise would lead to formation of toxic aggregates.

Chromophore: Refers to a molecule capable of emitting a fluorescence signal in the visible light spectrum when excited with a certain wavelength; as the emitted energy is lower than the energy used to excite a chromophore, the emission wavelength is larger than the excitation wavelength.

Click-chemistry: Refers to a series of covalent bioconjugation reactions which typically involve a molecule linked to an azide group (which has three double bonded nitrogen atoms $-N=N+=N-$) and another molecule linked to an alkyne group (which has a triple bond between two carbon atoms, $-C \equiv CH$) and can be catalyzed by copper monovalent ions (Cu^+).

Constant force (FC): Mode of force application to a molecule in single molecule force spectroscopy. Also known as "force clamp", in which a constant value of force is applied to a molecule, which is maintained by either a feedback loop or by a constant magnetic field.

Constant velocity (FX): Or, Force Extension, is a mode of force application to a molecule in single molecule force spectroscopy. The force is applied by pulling on the molecule at constant velocity, where its length changes nonlinearly with respect to the increase in the applied load.

Contractile energy: An energy that can be estimated as the product between the force and contraction length.

Cumulative distribution function (CDF): Describes the probability of a random variable X , when it equals, or smaller than some real value a , i.e., $P(a) = \Pr(X \leq a)$. On one limit the CDF has an asymptote at 0, and at 1 at the other limit. In between these values it grows monotonically, meaning that for $a \leq b$, $P(a) \leq P(b)$. The CDF of a continuous variable can be expressed as the integral of its probability density function, as shown in Eq. (3.3).

dB: A measure for sound volume representing the intensity on a logarithmic scale. Humans can hear sounds up to 130 dB (the pain threshold) at frequencies between 20 Hz and 20 kHz. A sound level of 55 dB is that of a normal conversation, 85 dB continuous exposure is the limit of hearing damage; damage can occur when the links between cadherin 23 and protocadherin 15 are broken.

Dissociation constants: Are a measure for the equilibria between an associated complex and its components in free form. The strength of an interaction is defined through either the association constant to form a complex, but more often through the dissociation constant to break a complex. The two constants are inversely proportional. An interaction is considered as strong in biology if its dissociation constant is in the nanomolar (nM) range).

Energy landscape: Is an elegant way for scientists to quantify how likely a state can be populated in a reaction and figure out the path that a process might take; similarly to a ball that, due to gravity, will travel downhill in a geographical landscape, a molecule will want to reach a minimum every on an energy landscape; a stable state is marked as a valley, while a transition state as a hill; the energy decreases with the log of number of states; hence for a state to decrease two folds in energy it needs to occur ~10 times more often.

Equilibrium: Refers to a situation in which material/energy fluxes are flowing in opposite direction with equal rates, such that no net change in the system is observed.

FRET efficiency: Is defined as the proportion of energy released by the donor that is transferred to the acceptor chromophore; the FRET efficiency is highly dependent on the spacing between donor and acceptor and is affected by the relative dipole orientation of the two chromophores and the spectral overlap between the emission of the donor and excitation of the acceptor.

Gaussian distribution: Also known as "normal distribution", is a symmetric distribution about the mean (which is equal to the median, and mode) that is represented by the peak of the distribution. The width of this distribution is defined by the standard deviation. Its empirical rule states that 68.2% of the observations it describes will be within a range of one standard deviation around the mean, 95.4% within two standard deviations, and 99.7% within three standard deviations.

Gibbs free energy: Also referred to as "free enthalpy", is an important thermodynamic property. It is a thermodynamic potential that is used to quantify the maximal amount of work in a thermodynamic closed system under constant pressure and temperature.

Homoemers and heteromers: Refers to molecular assemblies made from the same or different proteins, respectively.

Hookean spring: Refers to a spring that obeys Hook's law: $F = k \times x$, where F is force, x is extension and k is spring constant; as the polypeptide chain is made of covalent bonds, which are inelastic, a polypeptide chain will only behave as a Hookean spring at very low forces (< 10 pN), where the force straightens the backbone.

Hydrogel: Is a crosslinked hydrophilic p material that contains more water by weight than dry mass but does not dissolve in water.

I91: I91 is a titin domain in the distal region; it is sometimes referred to as I27 due to a misrepresentation of the titin gene.

Immunoglobulin fold: A 2-layer sandwich of 7-9 antiparallel β -strands arranged in two β -sheets. Some proteins are considered all- β even when they have some isolated α -helices at the periphery of their structure.

Independent and identically distributed (iid): This term means that collection of random variables sampled from different events do not share dependency in each other and are identically distributed, which means that they share the same mother distribution.

Isoform: A protein variant originating from the same gene or gene family. Titin isoforms are produced through alternative splicing, as during the transcription process the mRNA is assembled to encode for a different number of Ig-like domains and slightly different unstructured regions, specific for each muscle. The cardiac muscle has two titin isoforms.

Kosmotropic: Defined by hydration theory, kosmotropes (water order-makers) as hard and strongly hydrated ions that are either small or that have high charge density. Their presence stabilizes the native state of proteins. By ordering the water molecules around them, they effectively deplete the water molecules surrounding the protein, making the solution a poor solvent for the protein, and thus increasing the stability of the protein's folded (or collapsed) state.

Kramers theory (or Kramers model): Introduced by the Dutch physicist Hendrik Anthony "Hans" Kramers in 1940 to describe the crossing of an energy barrier with respect for friction and thermal agitation effects along a reaction coordinate. It enables the determination of the barrier crossing rates from fluxes.

Lenard-Jones or Morse potentials: Originally these are two mathematical representations of pair interaction potentials between two non-bonded and uncharged atoms/molecules (van der Waals interactions). Their usefulness as intermolecular potentials expanded their use, as in the case presented here, where they can conveniently be used to define a potential of folded conformation along a distance reaction coordinate.

Loading rate: Represents a measure of how fast the force increases with time. It is estimated from the product between the retraction velocity and the spring constant of the probe. Most simulations and many measurements are done under constant loading rate (or constant velocity) rather than constant force conditions. The two approaches produce opposite results: at constant force the higher the force, the shorter the unfolding time; at constant velocity, the faster the loading rate, the higher the unfolding force a protein has.

Mannose: Is a sugar molecule important in cellular metabolism; it is attached to proteins through the posttranslational modification of several amino acids and tunes the structural and functional roles of membrane proteins.

Mechanical clamp: Is a region in the protein structure that opposes the greatest resistance to a force vector and typically breaking of this region results in the immediate unfolding of the protein; can be seen as the equivalent of a load bearing wall holding a structure.

Mechanotransduction: Describes the ability of a system to sense, integrate and convert a mechanical input into a biochemical output.

Molecular dynamics (MD) simulations: A numerical approach to simulate the dynamical evolution of a large system comprised of atoms and molecules through the application of force fields and local interactions between them.

Molten-globule: A set of states where many hydrogen bonds and structure elements are significantly more stable than in the previous phases of folding note that there is still debate if all the water molecules leave the molten globule state and some authors distinguish between a dry and solvent containing molten globule states.

Native contacts: The contacts that form between the residues of amino acids that stabilize the structure of the protein native state.

Perturbation parameter: A means/agent by which a perturbation is introduced to the system.

PEVK: This unstructured peptide region of titin is called as such, because it contains in a proportion larger than 70% one of the four amino acids: Proline (P), Glutamic Acid (E), Valine (V) and Lysine (K).

Piezo actuator: A ceramic based device which can move with sub-nm resolution in response to an applied voltage. Piezoelectric materials can produce energy due to the development of electric charge generated by the movement of electrons upon application of stress. As the effect is reversible, a strong electrical field can change the dimensions of piezoelectric materials.

Pili (pilus for singular): Is also referred to as fimbria. As a consequence, proteins in type I pilus discussed in this chapter are named FimA, FimD, etc. Note that some researchers reserve the term pilus for the larger appendages that are used for bacterial conjugation.

pN: 1 piconewton (pN) represents 10⁻¹² Newtons. At single molecule level the forces are in the pN range, while the extensions are in the nanometer (nm) range.

Poisson distribution: Named after the French mathematician Siméon Denis Poisson, this distribution describes the probability of random events that within a temporal or spatial interval if these events have a constant mean rate and are independent of each other.

Polyproteins: Are chains of covalently conjoined smaller proteins that occur in nature as versatile means to organize the proteome; we use this term here as a class of multidomain proteins that do not have long-lasting interdomain interactions.

Potential of mean force (PMF): Represents the projection of the free energy surface along a specific reaction coordinate.

Probability density function (PDF): Is a function that describes the density of a variable X in all its probability space. The probability to find X within some interval ($a \leq X \leq b$) equals the integral over the PDF in this interval.

Quadrant photodiode: A device that reports on the intensity of a laser beam as measured over its quadrants. The deflection of a cantilever is measured from the difference between top and bottom halves, while the torsion from the difference between the left and right halves. A perfectly centered beam produces an output of zero volts.

Quasi-adiabatic approximation: An adiabatic process is defined as a process in which no heat is accepted or rejected by the system at any time. In our context, the quasi-adiabatic (sometimes referred to as “low-loss”) approximation implies that the unfolding rate is slower than the characteristic relaxation rate.

Radius of gyration: A measure of averaged size, which is calculated by the root mean square distance from the edges of the surface volume to its center of mass, without changing its moment of rotational inertia.

Random walk: A random process describing a trajectory made of a sequence of steps with random direction, whose size and dwell time are samples from probabilistic distributions. If the step size has a Gaussian distribution shape, and the dwell-times have Poisson distribution shape, then this random walk describes Brownian motion, or a Markov process (where each step is independent of its previous one).

Reaction coordinate: A coordinate that signifies the progress of the process along it is being modified. It can be the end-to-end length of a molecule, radius of gyration, the number of hydrogen bonds, bond angle, etc.

Restriction sites: Are 6–8 base pairs of DNA that are recognized and cleaved by specific enzymes; typically expression vectors with a gene of interest (called plasmids), which are pieces of circular DNA having 2-6 kbp, are used for protein overexpression; for scientists to add a gene of interest into an expression vector, they use such restriction enzymes, that open the two strands asymmetrically, producing what is known as ‘sticky ends’ (dotted line in top Figure 4.1A), followed by a ligation procedure, where complementary strands are glued together.

Rheometry: Refers to a class of techniques which measure the amount of deformation of a material of a fluid under an applied force or stress.

Ribosome exit tunnel: The terminal structure of ribosome, from where the newly produced peptide chain emerges in solution has some flexibility and the diameter is not uniform, nor well-defined; the channel changes depending on the shape and size of the emerging sequence; it can accommodate between 30 to 35 amino acids, but this number is reduced to 15 to 22 amino acids in the presence of a steric extrusion force which stretches the chain.

Salting-in and salting-out: These terms refer to how increasing the amount of salt in a solution affects the stability of a protein; any change in salt concentration affects the ionic strength of the

solution, which in turn affects any electrostatic interaction that stabilized the folded state, such as hydrogen bonds and salt bridges (which are oppositely charged amino acids groups in close spatial proximity inside the protein structure).

Single molecule force spectroscopy (SMFS): A methodology involving several techniques to study the dynamical behavior of individual molecules under the application of external forces.

Singlet oxygen: The lowest excited state of the dioxygen molecule, with a lifetime in solution of $\sim 3 \mu\text{sec}$; as the most popular and frequently used laser wavelengths for optical traps are in the near-IR (to minimize thermal effects); these wavelengths are however within the absorption bands of ground state molecular oxygen; apart from singlet oxygen, two photons can be adsorbed for a short time, behaving as one photon with double energy (especially when using pulsed laser beams), leading to photochemical effects; finally, while the laser wavelength and low power are typically chosen to minimize thermal effects (in the near-IR the sample is typically heating with $\sim 1 \text{ K}/100 \text{ mW}$ laser power), the high rate of change of temperature can too have toxic effects; all these effects are discussed in detail in the reference cited in this paragraph (Blázquez-Castro A, Micromachines, 2019).

Soleus: The longest muscle in the human body, located on the back of the lower leg; it activates during standing, walking, or jumping. It runs from the knee to the heel.

Sortase: A class of transpeptidase enzymes that covalently attach an array of proteins to the surface of Gram-positive bacteria.

Steric: Refers to the spatial atomic arrangement within a molecule. Also known as the “steric effect”. Conformational and reactivity of molecules and ions are affected by their steric and electronic distribution and structure.

Superparamagnetic bead: A bead does not have hysteresis (or memory effects) in their magnetization, meaning that the bead will behave the same if the magnetic field decreases or increases to a given value. To achieve such a property, manufacturers imbed magnetic nanocrystals ($\sim 20 \text{ nm}$ in size) in a polystyrene core. The crystals can then rotate in place and orient to force without rotating the bead.

Talin: A long protein (about 60 nm), comprised from a globular head domain and a flexible rod domain, all containing protein domain in tandem. It is of high importance in mechanotransduction, and is involved in focal adhesion.

Tautomeric conformations: The two equilibrium conformations; for O-C-N bond, the molecule has either a double C=O or C=N and leading to delocalization of electrons are also known as tautomeric conformations.

Thermal unfolding: Thermal denaturation is the process that happens to the proteins inside an egg during boiling, which transforms a protein solution into a solid nutritional food.

Thermocouple: Refers to a sensor that measures temperature by having two materials glued together; typically it relies on the electromotive force across two points of an electrically conducting material when there is a temperature difference between them; when a strip or cantilever is made from two layers of materials with different thermal expansion coefficients, the change in temperature will lead to one layer to expand more than the other and to the overall cantilever to bend toward the material with the smallest thermal expansion coefficient.

Thermodynamic system: A defined system of matter which is separated from its surroundings.

Titin: Named after the pre-Olympian gods from the Greek mythology, the muscle protein titin is the largest protein known so far. It is about 1 mm in length and comprised of 244 folded proteins that are connected by unstructured peptides. It operates as a molecular spring that adjusts muscle elasticity.

Trap (in the context of magnetic and optical tweezers): Represents the magnetic field or the focused laser beam, respectively, that is used to move a bead tethered to a molecule and to apply mechanical force.

Truncated power law (TPL): The truncation introduces a "cutoff" to the power law distribution by multiplying it with an exponential. Since the exponent of the power law can display ill-behaved moments due to their possible long (heavy) tails, the truncation sets a limit to this asymptotic behavior. In physical systems, this cutoff will reflect the finite size, or boundary limit of the system.

Virial theory: The virial expansion extends the basic equation of state for non-ideal gases, and is used also for fluids, due to the presence of long- and short-term interactions. This is an equation of state that relates the pressure with the density and temperature. In the presence of ionic solutions, one can express the osmotic pressure as a series of powers in the concentration (density).

"Weibull" distribution: Named after the Swedish mathematician Waloddi Weibull, this distribution can describe a wide range of distribution shapes, but in particular it is flexible in modeling skewed

data (differently from a normal distribution). It is characterized by a shape parameter, b (sometimes k). if $b = 1$, then the Weibull distribution becomes identical to the exponential distribution, and if $b = 2$ it is equivalent to a Gaussian distribution.

Zwitterions: In moderate pH the amino acids are dipolar ions with both NH_3^+ and CO_2^- groups having charged states.

References

1. Sharma, S.; Subramani, S.; Popa, I., Does protein unfolding play a functional role in vivo? *FEBS J.* **2021**, *288* (6), 1742-1758.
2. Dietz, H.; Berkemeier, F.; Bertz, M.; Rief, M., Anisotropic deformation response of single protein molecules. *Proc. Natl. Acad. Sci. USA* **2006**, *103* (34), 12724-12728.
3. 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* **2011**, *145* (2), 257-267.
4. Ho, B.; Baryshnikova, A.; Brown, G. W., Unification of Protein Abundance Datasets Yields a Quantitative *Saccharomyces cerevisiae* Proteome. *Cell Syst.* **2018**, *6* (2), 192-205.e3.
5. Popa, I.; Gutzman, J. H., The extracellular matrix–myosin pathway in mechanotransduction: from molecule to tissue. *Emerg. Top. Life Sci.* **2018**, *2* (5), 727-737.
6. Kosuri, P.; Alegre-Cebollada, J.; Feng, J.; Kaplan, A.; Ingles-Prieto, A.; Badilla, C. L.; Stockwell, B. R.; Sanchez-Ruiz, J. M.; Holmgren, A.; Fernandez, J. M., Protein Folding Drives Disulfide Formation. *Cell* **2012**, *151* (4), 794-806.
7. Zapp, C.; Obarska-Kosinska, A.; Rennekamp, B.; Kurth, M.; Hudson, D. M.; Mercadante, D.; Barayeu, U.; Dick, T. P.; Denysenkov, V.; Prisner, T.; Bennati, M.; Daday, C.; Kappl, R.; Gräter, F., Mechanoradicals in tensed tendon collagen as a source of oxidative stress. *Nat. Commun.* **2020**, *11* (1).
8. Bernardi, R. C.; Durner, E.; Schoeler, C.; Malinowska, K. H.; Carvalho, B. G.; Bayer, E. A.; Luthey-Schulten, Z.; Gaub, H. E.; Nash, M. A., Mechanisms of Nanonewton Mechanostability in a Protein Complex Revealed by Molecular Dynamics Simulations and Single-Molecule Force Spectroscopy. *J. Am. Chem. Soc.* **2019**, *141* (37), 14752-14763.
9. Grandbois, M.; Beyer, M.; Rief, M.; Clausen-Schaumann, H.; Gaub, H. E., How strong is a covalent bond? *Science* **1999**, *283* (5408), 1727-1730.
10. 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. *Nat. Chem.* **2017**, *9* (2), 164-170.
11. Eisenberg, D., The discovery of the alpha-helix and beta-sheet, the principal structural features of proteins. *Proc. Natl. Acad. Sci. USA* **2003**, *100* (20), 11207-11210.
12. DeBenedictis, E. P.; Keten, S., Mechanical unfolding of alpha- and beta-helical protein motifs. *Soft Matter* **2019**, *15* (6), 1243-1252.
13. Zhmurov, A.; Kononova, O.; Litvinov, R. I.; Dima, R. I.; Barsegov, V.; Weisel, J. W., Mechanical Transition from α -Helical Coiled Coils to beta-Sheets in Fibrin(ogen). *J. Am. Chem. Soc.* **2012**, *134* (50), 20396-20402.

14. Neumaier, S.; Reiner, A.; Buttner, M.; Fierz, B.; Kiefhaber, T., Testing the diffusing boundary model for the helix-coil transition in peptides. *Proc. Natl. Acad. Sci. USA* **2013**, *110* (32), 12905-12910.
15. Wang, Y. N.; Yao, M. X.; Baker, K. B.; Gough, R. E.; Le, S.; Goult, B. T.; Yan, J., Force-Dependent Interactions between Talin and Full-Length Vinculin. *J. Am. Chem. Soc.* **2021**, *143* (36), 14726-14737.
16. Tapia-Rojo, R.; Alonso-Caballero, A.; Fernandez, J., Direct observation of a coil-to-helix contraction triggered by vinculin binding to talin. *Sci. Adv.* **2020**, *6* (21).
17. Haworth, N. L.; Wouters, M. A., Between-strand disulfides: forbidden disulfides linking adjacent beta-strands. *Rsc Adv.* **2013**, *3* (46), 24680-24705.
18. Asakura, T.; Okonogi, M.; Nakazawa, Y.; Yamauchi, K., Structural analysis of alanine tripeptide with antiparallel and parallel beta-sheet structures in relation to the analysis of mixed beta-sheet structures in Samia cynthia ricini silk protein fiber using solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **2006**, *128* (18), 6231-6238.
19. Lee, G.; Abdi, K.; Jiang, Y.; Michaely, P.; Bennett, V.; Marszalek, P. E., Nanospring behaviour of ankyrin repeats. *Nature* **2006**, *440* (7081), 246-249.
20. Popa, I.; Berkovich, R.; Alegre-Cebollada, J.; Badilla, C. L.; Rivas-Pardo, J. A.; Taniguchi, Y.; Kawakami, M.; Fernandez, J. M., Nanomechanics of HaloTag Tethers. *J. Am. Chem. Soc.* **2013**, *135* (34), 12762-12771.
21. Eckels, E. C.; Tapia-Rojo, R.; Rivas-Pardo, J. A.; Fernandez, J. M., The Work of Titin Protein Folding as a Major Driver in Muscle Contraction. *Annu. Rev. Physiol.* **2018**, *80*, 327-351.
22. Pollard, T. D.; Cooper, J. A., Actin, a Central Player in Cell Shape and Movement. *Science* **2009**, *326* (5957), 1208-1212.
23. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D., Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583-+.
24. Marino, J.; von Heijne, G.; Beckmann, R., Small protein domains fold inside the ribosome exit tunnel. *Febs. Lett.* **2016**, *590* (5), 655-660.
25. Clerico, E. M.; Tiliusky, J. M.; Meng, W. L.; Gierasch, L. M., How Hsp70 Molecular Machines Interact with Their Substrates to Mediate Diverse Physiological Functions. *J. Mol. Biol.* **2015**, *427* (7), 1575-1588.
26. Desai, V. P.; Frank, F.; Lee, A.; Righini, M.; Lancaster, L.; Noller, H. F.; Tinoco, I.; Bustamante, C., Co-temporal Force and Fluorescence Measurements Reveal a

Ribosomal Gear Shift Mechanism of Translation Regulation by Structured mRNAs. *Mol. Cell* **2019**, 75 (5), 1007-+.

27. Voss, N. R.; Gerstein, M.; Steitz, T. A.; Moore, P. B., The geometry of the ribosomal polypeptide exit tunnel. *J. Mol. Biol.* **2006**, 360 (4), 893-906.
28. Goldman, D. H.; Kaiser, C. M.; Milin, A.; Righini, M.; Tinoco, I.; Bustamante, C., Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo. *Science* **2015**, 348 (6233), 457-460.
29. De los Rios, P.; Ben-Zvi, A.; Slutsky, O.; Azem, A.; Goloubinoff, P., Hsp70 chaperones accelerate protein translocation and the unfolding of stable protein aggregates by entropic pulling. *Proc. Natl. Acad. Sci. USA* **2006**, 103 (16), 6166-6171.
30. Fisher, T. E.; Marszalek, P. E.; Fernandez, J. M., Stretching single molecules into novel conformations using the atomic force microscope. *Nat. Struct. Biol.* **2000**, 7 (9), 719-724.
31. Ohgushi, M.; Wada, A., 'Molten-globule state': a compact form of globular proteins with mobile side-chains. *Febs. Lett.* **1983**, 164 (1), 21-24.
32. Pande, V. S.; Rokhsar, D. S., Is the molten globule a third phase of proteins? *Proc. Natl. Acad. Sci. USA* **1998**, 95 (4), 1490-1494.
33. Bryan, P. N.; Orban, J., Proteins that switch folds. *Curr. Opin. Struct. Biol.* **2010**, 20 (4), 482-488.
34. Oldfield, C. J.; Dunker, A. K., Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu Rev Biochem* **2014**, 83, 553-584.
35. Lin, M. M.; Shorokhov, D.; Zewail, A. H., Dominance of misfolded intermediates in the dynamics of alpha-helix folding. *Proc. Natl. Acad. Sci. USA* **2014**, 111 (40), 14424-14429.
36. Stirnemann, G.; Giganti, D.; Fernandez, J. M.; Berne, B. J., Elasticity, structure, and relaxation of extended proteins under force. *Proc. Natl. Acad. Sci. USA* **2013**, 110 (10), 3847-3852.
37. 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* **2009**, 106 (26), 10534-10539.
38. Granzier, H.; Labeit, S., Structure-function relations of the giant elastic protein titin in striated and smooth muscle cells. *Muscle Nerve* **2007**, 36 (6), 740-755.
39. Han, J. H.; Batey, S.; Nickson, A. A.; Teichmann, S. A.; Clarke, J., The folding and evolution of multidomain proteins. *Nat. Rev. Mol. Cell Biol.* **2007**, 8 (4), 319-330.
40. Li, H. B.; Oberhauser, A. F.; Fowler, S. B.; Clarke, J.; Fernandez, J. M., Atomic force microscopy reveals the mechanical design of a modular protein. *Proc. Natl. Acad. Sci. USA* **2000**, 97 (12), 6527-6531.
41. Khatri, B. S.; Byrne, K.; Kawakami, M.; Brockwell, D. J.; Smith, D. A.; Radford, S. E.; McLeish, T. C. B., Internal friction of single polypeptide chains at high stretch. *Faraday Discuss.* **2008**, 139, 35-51.

42. Elias-Mordechai, M.; Chetrit, E.; Berkovich, R., Interplay between Viscoelasticity and Force Rate Affects Sequential Unfolding in Polyproteins Pulled at Constant Velocity. *Macromolecules* **2020**, *53* (8), 3021-3029.
43. Zinober, R. C.; Brockwell, D. J.; Beddard, G. S.; Blake, A. W.; Olmsted, P. D.; Radford, S. E.; Smith, D. A., Mechanically unfolding proteins: The effect of unfolding history and the supramolecular scaffold. *Protein Sci.* **2002**, *11* (12), 2759-2765.
44. Chetrit, E.; Meroz, Y.; Klausner, Z.; Berkovich, R., Correlations within polyprotein forced unfolding dwell-times introduce sequential dependency. *J. Struct. Biol.* **2020**, *210* (3).
45. Crepin, T.; Swale, C.; Monod, A.; Garzoni, F.; Chaillet, M.; Berger, I., Polyproteins in structural biology. *Curr. Opin. Struct. Biol.* **2015**, *32*, 139-146.
46. Bell, N. M.; Lever, A. M. L., HIV Gag polyprotein: processing and early viral particle assembly. *Trends Microbiol.* **2013**, *21* (3), 136-144.
47. Hilgenfeld, R., From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. *FEBS J.* **2014**, *281* (18), 4085-4096.
48. Kruger, M.; Linke, W. A., The Giant Protein Titin: A Regulatory Node That Integrates Myocyte Signaling Pathways. *J. Biol. Chem.* **2011**, *286* (12), 9905-9912.
49. Vogel, V.; Sheetz, M., Local force and geometry sensing regulate cell functions. *Nat. Rev. Mol. Cell Biol.* **2006**, *7* (4), 265-275.
50. Kennedy, M. W., The polyprotein lipid binding proteins of nematodes. *Biochim. Biophys. Acta* **2000**, *1476* (2), 149-164.
51. Erickson, H. P., Reversible unfolding of fibronectin type III and immunoglobulin domains provides the structural basis for stretch and elasticity of titin and fibronectin. *Proc. Natl. Acad. Sci. USA* **1994**, *91* (21), 10114-10118.
52. Garcia-Manyes, S.; Beedle, A. E. M., Steering chemical reactions with force. *Nat. Rev. Chem.* **2017**, *1* (11).
53. LeWinter, M. M.; Granzier, H., Cardiac Titin A Multifunctional Giant. *Circulation* **2010**, *121* (19), 2137-2145.
54. Freundt, J. K.; Linke, W. A., Titin as a force-generating muscle protein under regulatory control. *J. Appl. Physiol.* **2019**, *126* (5), 1474-1482.
55. Linke, W. A., Stretching the story of titin and muscle function. *Journal of Biomechanics* **2023**, *152*, 111553.
56. Critchley, D. R., Genetic, biochemical and structural approaches to talin function. *Biochem. Soc. Trans.* **2005**, *33*, 1308-1312.
57. Goult, B. T.; Yan, J.; Schwartz, M. A., Talin as a mechanosensitive signaling hub. *J. Cell Biol.* **2018**, *217* (11), 3776-3784.
58. Chothia, C.; Jones, E. Y., The molecular structure of cell adhesion molecules. *Annu. Rev. Biochem.* **1997**, *66*, 823-862.

59. Leckband, D. E.; de Rooij, J., Cadherin Adhesion and Mechanotransduction. In *Annual Review of Cell and Developmental Biology*, Vol 30, Schekman, R.; Lehmann, R., Eds. 2014; Vol. 30, pp 291-315.
60. Hynes, R. O., The dynamic dialogue between cells and matrices: Implications of fibronectin's elasticity. *Proc. Natl. Acad. Sci. USA* **1999**, 96 (6), 2588-2590.
61. Singh, P.; Carraher, C.; Schwarzbauer, J. E., Assembly of Fibronectin Extracellular Matrix. In *Annual Review of Cell and Developmental Biology*, Vol 26, Schekman, R.; Goldstein, L.; Lehmann, R., Eds. 2010; Vol. 26, pp 397-419.
62. Desmet, J.; Verstraete, K.; Bloch, Y.; Lorent, E.; Wen, Y. R.; Devreese, B.; Vandenbroucke, K.; Loverix, S.; Hettmann, T.; Deroo, S.; Somers, K.; Henderikx, P.; Lasters, I.; Savvides, S. N., Structural basis of IL-23 antagonism by an Alphabody protein scaffold. *Nat. Commun.* **2014**, 5.
63. Luke, G. A.; Ryan, M. D., "Therapeutic applications of the 'NPGP' family of viral 2As". *Rev. Med. Virol.* **2018**, 28 (6).
64. Zhao, Y.; Zhu, Y.; Liu, X.; Jin, Z. M.; Duan, Y. K.; Zhang, Q.; Wu, C. Y.; Feng, L.; Du, X. Y.; Zhao, J. Y.; Shao, M. L.; Zhang, B.; Yang, X. N.; Wu, L. J.; Ji, X. Y.; Guddat, L. W.; Yang, K. L.; Rao, Z. H.; Yang, H. T., Structural basis for replicase polyprotein cleavage and substrate specificity of main protease from SARS-CoV-2. *Proc. Natl. Acad. Sci. USA* **2022**, 119 (16).
65. Hughes, M. L.; Dougan, L., The physics of pulling polyproteins: a review of single molecule force spectroscopy using the AFM to study protein unfolding. *Rep. Prog. Phys* **2016**, 79 (7).
66. Scholl, Z. N.; Josephs, E. A.; Marszalek, P. E., Modular, Nondegenerate Polyprotein Scaffolds for Atomic Force Spectroscopy. *Biomacromolecules* **2016**, 17 (7), 2502-2505.
67. Schonfelder, J.; De Sancho, D.; Perez-Jimenez, R., The Power of Force: Insights into the Protein Folding Process Using Single-Molecule Force Spectroscopy. *J. Mol. Biol.* **2016**, 428 (21), 4245-4257.
68. Popa, L.; Berkovich, R., Mechanobiology: protein refolding under force. *Emerg. Top. Life Sci.* **2018**, 2 (5), 687-699.
69. Mora, M.; Stannard, A.; Garcia-Manyes, S., The nanomechanics of individual proteins. *Chem. Soc. Rev.* **2020**, 49 (19), 6816-6832.
70. Alegre-Cebollada, J., Protein nanomechanics in biological context. *Biophys. Rev.* **2021**, 13 (4), 435-454.
71. Yancey, P. H.; Clark, M. E.; Hand, S. C.; Bowlus, R. D.; Somero, G. N., Living with water-stress – evolution of osmolyte systems. *Science* **1982**, 217 (4566), 1214-1222.
72. DuBay, K. H.; Bowman, G. R.; Geissler, P. L., Fluctuations within Folded Proteins: Implications for Thermodynamic and Allosteric Regulation. *Acc. Chem. Res.* **2015**, 48 (4), 1098-1105.

73. Israelachvili, J. N., *Intermolecular and Surface Forces*, 3rd Edition. 2011; p 1-674.
74. Rose, G. D.; Wolfenden, R., Hydrogen-bonding, hydrophobicity, packing, and protein-folding. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, 22, 381-415.
75. Livesay, D. R.; Huynh, D. H.; Dallakyan, S.; Jacobs, D. J., Hydrogen bond networks determine emergent mechanical and thermodynamic properties across a protein family. *Chem. Cent. J.* **2008**, 2.
76. Fenimore, P. W.; Frauenfelder, H.; McMahon, B. H.; Parak, F. G., Slaving: Solvent fluctuations dominate protein dynamics and functions. *Proc. Natl. Acad. Sci. USA* **2002**, 99 (25), 16047-16051.
77. Prabhu, N.; Sharp, K., Protein-solvent interactions. *Chem. Rev.* **2006**, 106 (5), 1616-1623.
78. Kunz, W., Specific ion effects in colloidal and biological systems. *Curr. Opin. Colloid Interface Sci.* **2010**, 15 (1-2), 34-39.
79. Chiti, F.; Dobson, C. M., Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. In *Annual Review of Biochemistry*, Vol 86, Kornberg, R. D., Ed. 2017; Vol. 86, pp 27-68.
80. Manning, M. C.; Chou, D. K.; Murphy, B. M.; Payne, R. W.; Katayama, D. S., Stability of Protein Pharmaceuticals: An Update. *Pharm. Res.* **2010**, 27 (4), 544-575.
81. Loveday, S. M., Food Proteins: Technological, Nutritional, and Sustainability Attributes of Traditional and Emerging Proteins. In *Annual Review of Food Science and Technology*, Vol 10, Doyle, M. P.; McClements, D. J., Eds. 2019; Vol. 10, pp 311-339.
82. Olsen, B. D., Engineering Materials from Proteins. *AIChE J.* **2013**, 59 (10), 3558-3568.
83. Lo Nostro, P.; Ninham, B. W., Hofmeister Phenomena: An Update on Ion Specificity in Biology. *Chem. Rev.* **2012**, 112 (4), 2286-2322.
84. Jungwirth, P.; Cremer, P. S., Beyond Hofmeister. *Nat. Chem.* **2014**, 6 (4), 261-263.
85. Collins, K. D.; Neilson, G. W.; Enderby, J. E., Ions in water: Characterizing the forces that control chemical processes and biological structure. *Biophys. Chem.* **2007**, 128 (2-3), 95-104.
86. Cacace, M. G.; Landau, E. M.; Ramsden, J. J., The Hofmeister series: salt and solvent effects on interfacial phenomena. *Q. Rev. Biophys.* **1997**, 30 (3), 241-277.
87. Vrbka, L.; Jungwirth, P.; Bauduin, P.; Touraud, D.; Kunz, W., Specific ion effects at protein surfaces: A molecular dynamics study of bovine pancreatic trypsin inhibitor and horseradish peroxidase in selected salt solutions. *J. Phys. Chem. B* **2006**, 110 (13), 7036-7043.
88. Zhang, Y. J.; Cremer, P. S., Interactions between macromolecules and ions: the Hofmeister series. *Curr. Opin. Chem. Biol.* **2006**, 10 (6), 658-663.
89. Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N.; Dothager, R. S.; Seifert, S.; Thiagarajan, P.; Sosnick, T. R.; Hasan, M. Z.; Pande, V. S.; Ruczinski,

- I.; Doniach, S.; Plaxco, K. W., Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. USA* **2004**, *101* (34), 12491-12496.
90. Sherman, E.; Haran, G., Coil-globule transition in the denatured state of a small protein. *Proc. Natl. Acad. Sci. USA* **2006**, *103* (31), 11539-11543.
 91. Doniach, S.; Garel, T.; Orland, H., Phase diagram of a semiflexible polymer chain in a theta solvent: Application to protein folding. *J. Chem. Phys.* **1996**, *105* (4), 1601-1608.
 92. Feller, G., Protein folding at extreme temperatures: Current issues. *Semin. Cell Dev. Biol.* **2018**, *84*, 129-137.
 93. Jackson, S. E., How do small single-domain proteins fold? *Fold Des.* **1998**, *3* (4), R81-R91.
 94. Baldwin, R. L., Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl. Acad. Sci. USA* **1986**, *83* (21), 8069-8072.
 95. Karplus, M., Aspects of protein reaction dynamics: Deviations from simple behavior. *J. Phys. Chem. B* **2000**, *104* (1), 11-27.
 96. Privalov, P. L., Stability of Proteins Small Globular Proteins. In *Advances in Protein Chemistry*, Anfinsen, C. B.; Edsall, J. T.; Richards, F. M., Eds. Academic Press: 1979; Vol. 33, pp 167-241.
 97. Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G., Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins.* **1995**, *21* (3), 167-195.
 98. Liu, F.; Du, D. G.; Fuller, A. A.; Davoren, J. E.; Wipf, P.; Kelly, J. W.; Gruebele, M., An experimental survey of the transition between two-state and downhill protein folding scenarios. *Proc. Natl. Acad. Sci. USA* **2008**, *105* (7), 2369-2374.
 99. Javadi, Y.; Fernandez, J. M.; Perez-Jimenez, R., Protein Folding Under Mechanical Forces: A Physiological View. *Physiology* **2013**, *28* (1), 9-17.
 100. Rief, M.; Pascual, J.; Saraste, M.; Gaub, H. E., Single molecule force spectroscopy of spectrin repeats: Low unfolding forces in helix bundles. *J. Mol. Biol.* **1999**, *286* (2), 553-561.
 101. Lenne, P. F.; Raae, A. J.; Altmann, S. M.; Saraste, M.; Horber, J. K. H., States and transitions during forced unfolding of a single spectrin repeat. *Febs. Lett.* **2000**, *476* (3), 124-128.
 102. Altmann, S. M.; Grunberg, R. G.; Lenne, P. F.; Ylanne, J.; Raae, A.; Herbert, K.; Saraste, M.; Nilges, M.; Horber, J. K. H., Pathways and intermediates in forced unfolding of spectrin repeats. *Structure* **2002**, *10* (8), 1085-1096.
 103. Lu, H.; Schulten, K., The key event in force-induced unfolding of titin's immunoglobulin domains. *Biophys. J.* **2000**, *79* (1), 51-65.
 104. Carrion-Vazquez, M.; Li, H. B.; Lu, H.; Marszalek, P. E.; Oberhauser, A. F.; Fernandez, J. M., The mechanical stability of ubiquitin is linkage dependent. *Nat. Struct. Biol.* **2003**, *10* (9), 738-743.

105. Brockwell, D. J.; Paci, E.; Zinober, R. C.; Beddard, G. S.; Olmsted, P. D.; Smith, D. A.; Perham, R. N.; Radford, S. E., Pulling geometry defines the mechanical resistance of a beta-sheet protein. *Nat. Struct. Biol.* **2003**, *10* (9), 731-737.
106. Stirnemann, G.; Kang, S. G.; Zhou, R. H.; Berne, B. J., How force unfolding differs from chemical denaturation. *Proc. Natl. Acad. Sci. USA* **2014**, *111* (9), 3413-3418.
107. Schlierf, M.; Li, H. B.; Fernandez, J. M., The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proc. Natl. Acad. Sci. USA* **2004**, *101* (19), 7299-7304.
108. Tapia-Rojo, R.; Mazo, J. J.; Falo, F., Thermal versus mechanical unfolding in a model protein. *J. Chem. Phys.* **2019**, *151* (18).
109. Fersht, A. R., Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* **1997**, *7* (1), 3-9.
110. Zhou, Y. Q.; Karplus, M., Interpreting the folding kinetics of helical proteins. *Nature* **1999**, *401* (6751), 400-403.
111. Bychkova, V. E.; Semisotnov, G. V.; Balobanov, V. A.; Finkelstein, A. V., The Molten Globule Concept: 45 Years Later. *Biochem. (Mosc.)* **2018**, *83*, S33-S47.
112. Judy, E.; Kishore, N., A look back at the molten globule state of proteins: thermodynamic aspects. *Biophys. Rev.* **2019**, *11* (3), 365-375.
113. Dobson, C. M.; Sali, A.; Karplus, M., Protein folding: A perspective from theory and experiment. *Angew. Chem. Int. Ed.* **1998**, *37* (7), 868-893.
114. Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D., Navigating the folding routes. *Science* **1995**, *267* (5204), 1619-1620.
115. Dill, K. A.; Chan, H. S., From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **1997**, *4* (1), 10-19.
116. Onuchic, J. N.; LutheySchulten, Z.; Wolynes, P. G., Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545-600.
117. Tsai, C. J.; Ma, B. Y.; Sham, Y. Y.; Kumar, S.; Nussinov, R., Structured disorder and conformational selection. *Proteins* **2001**, *44* (4), 418-427.
118. Henzler-Wildman, K.; Kern, D., Dynamic personalities of proteins. *Nature* **2007**, *450* (7172), 964-972.
119. Baldwin, R. L., Protein folding: matching speed and stability. *Nature* **1994**, *369* (6477), 183-184.
120. Zwanzig, R., Two-state models of protein folding kinetics. *Proc. Natl. Acad. Sci. USA* **1997**, *94* (1), 148-150.
121. Fersht, A., *Structure and Mechanism in Protein Science*.
122. Finkelstein, A. V.; Ptitsyn, O. B., *Protein Physics. A Course of Lectures*. Second Edition ed.; Academic Press: Amsterdam, 2016; p 508.

123. Onuchic, J. N.; Wolynes, P. G., Theory of protein folding. *Curr. Opin. Struct. Biol.* **2004**, *14* (1), 70-75.
124. Lapidus, L., Protein unfolding mechanisms and their effects on folding experiments [version 1; peer review: 2 approved]. *F1000Research* **2017**, *6* (1723).
125. Kramers, H. A., Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940**, *7* (4), 284-304.
126. Hanggi, P.; Talkner, P.; Borkovec, M., Reaction-Rate Theory - 50 Years after Kramers. *Rev. Mod. Phys.* **1990**, *62* (2), 251-341.
127. Guo, Z. Y.; Thirumalai, D., Kinetics of protein folding: nucleation mechanism, time scales, and pathways. *Biopolymers* **1995**, *36* (1), 83-102.
128. Socci, N. D.; Onuchic, J. N.; Wolynes, P. G., Diffusive dynamics of the reaction coordinate for protein folding funnels. *J. Chem. Phys.* **1996**, *104* (15), 5860-5868.
129. Best, R. B.; Hummer, G., Diffusion models of protein folding. *Phys. Chem. Chem. Phys.* **2011**, *13* (38), 16902-16911.
130. Dill, K. A.; MacCallum, J. L., The Protein-Folding Problem, 50 Years On. *Science* **2012**, *338* (6110), 1042-1046.
131. Kukic, P.; Kannan, A.; Dijkstra, M. J. J.; Abeln, S.; Camilloni, C.; Vendruscolo, M., Mapping the Protein Fold Universe Using the CamTube Force Field in Molecular Dynamics Simulations. *PLoS Comput. Biol.* **2015**, *11* (10).
132. Meyer, A. C.; Öz, Y.; Gundlach, N.; Karbach, M.; Lu, P.; Müller, G., Molecular chains under tension: Thermal and mechanical activation of statistically interacting extension and contraction particles. *Phys. Rev. E* **2020**, *101* (2), 022504.
133. de Gennes, P. G., *Scaling Concepts in Polymer Physics*. First edition ed.; Cornell University Press: 1980; p 319.
134. Doi, M.; Edwards, S. F.; Edwards, S. F., *The theory of polymer dynamics*. Clarendon Press: 1988; Vol. 73, p 408.
135. Khokhlov, A. R. G., A. Y.; Pande, V. S., *Statistical Physics of Macromolecules*. American Institute of Physics Melville, NY: 1994; p XXVIII, 350.
136. Carrion-Vazquez, M.; Oberhauser, A. F.; Fowler, S. B.; Marszalek, P. E.; Broedel, S. E.; Clarke, J.; Fernandez, J. M., Mechanical and chemical unfolding of a single protein: A comparison. *Proc. Natl. Acad. Sci. USA* **1999**, *96* (7), 3694-3699.
137. Bustamante, C.; Marko, J. F.; Siggia, E. D.; Smith, S., Entropic elasticity of lambda-phage DNA. *Science* **1994**, *265* (5178), 1599-1600.
138. Flory, P. J., *Statistical Mechanics of Chain Molecules*. Interscience Publishers: 1969; p 432.
139. Fixman, M.; Kovac, J., Polymer conformational statistics. III. Modified Gaussian models of stiff chains. *J. Chem. Phys.* **1973**, *58* (4), 1564-1568.

140. Marko, J. F.; Siggia, E. D., Stretching DNA. *Macromolecules* **1995**, *28* (26), 8759-8770.
141. Fiasconaro, A.; Falo, F., Analytical results of the extensible freely jointed chain model. *Phys. A: Stat. Mech. Appl.* **2019**, 532.
142. Liao, X. Y.; Purohit, P. K.; Gopinath, A., Extensions of the worm-like-chain model to tethered active filaments under tension. *J. Chem. Phys.* **2020**, *153* (19).
143. Bao, Y.; Luo, Z. L.; Cui, S. X., Environment-dependent single-chain mechanics of synthetic polymers and biomacromolecules by atomic force microscopy-based single-molecule force spectroscopy and the implications for advanced polymer materials. *Chem. Soc. Rev.* **2020**, *49* (9), 2799-2827.
144. Grosberg, A. I. U.; Khokhlov, A. R.; de Gennes, P. G., *Giant Molecules: Here, There, and Everywhere*. World Scientific: 2011; p 322.
145. Morrison, G.; Hyeon, C.; Toan, N. M.; Ha, B. Y.; Thirumalai, D., Stretching homopolymers. *Macromolecules* **2007**, *40* (20), 7343-7353.
146. Berkovich, R.; Garcia-Manyes, S.; Urbakh, M.; Klafter, J.; Fernandez, J. M., Collapse Dynamics of Single Proteins Extended by Force. *Biophys. J.* **2010**, *98* (11), 2692-2701.
147. Bryngelson, J. D.; Wolynes, P. G., Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem.* **1989**, *93* (19), 6902-6915.
148. Smith, L. J.; Fiebig, K. M.; Schwalbe, H.; Dobson, C. M., The concept of a random coil - Residual structure in peptides and denatured proteins. *Fold Des.* **1996**, *1* (5), R95-R106.
149. Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G., The energy landscapes and motions of proteins. *Science* **1991**, *254* (5038), 1598-1603.
150. McLeish, T. C. B., Protein folding in high-dimensional spaces: Hypergutters and the role of nonnative interactions. *Biophys. J.* **2005**, *88* (1), 172-183.
151. Brooks, C. L.; Gruebele, M.; Onuchic, J. N.; Wolynes, P. G., Chemical physics of protein folding. *Proc. Natl. Acad. Sci. USA* **1998**, *95* (19), 11037-11038.
152. Sheinerman, F. B.; Brooks, C. L., Molecular picture of folding of a small α/β protein. *Proc. Natl. Acad. Sci. USA* **1998**, *95* (4), 1562-1567.
153. Kawai, S.; Komatsuzaki, T., Effect of timescale on energy landscape: Distinction between free-energy landscape and potential of mean force. *Phys. Rev. E* **2013**, *87* (3).
154. Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M., A Novel, Highly Stable Fold of the Immunoglobulin Binding Domain of Streptococcal Protein G. *Science* **1991**, *253* (5020), 657-661.
155. Sheinerman, F. B.; Brooks, C. L., Calculations on folding of segment B1 of streptococcal protein G. *J. Mol. Biol.* **1998**, *278* (2), 439-456.
156. Zhou, R. H.; Berne, B. J.; Germain, R., The free energy landscape for beta hairpin folding in explicit water. *Proc. Natl. Acad. Sci. USA* **2001**, *98* (26), 14931-14936.

157. Shimada, J.; Shakhnovich, E. I., The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. *Proc. Natl. Acad. Sci. USA* **2002**, 99 (17), 11175-11180.
158. Dudko, O. K.; Graham, T. G. W.; Best, R. B., Locating the Barrier for Folding of Single Molecules under an External Force. *Phys. Rev. Lett.* **2011**, 107 (20).
159. Morrone, A.; Giri, R.; Toofanny, Rudesh D.; Travaglini-Allocatelli, C.; Brunori, M.; Daggett, V.; Gianni, S., GB1 Is Not a Two-State Folder: Identification and Characterization of an On-Pathway Intermediate. *Biophys. J.* **2011**, 101 (8), 2053-2060.
160. Ptitsyn, O. B., Molten globule and protein folding. In *Advances in Protein Chemistry, Vol 47*, Anfinsen, C. B.; Edsall, J. T.; Richards, F. M.; Eisenberg, D. S., Eds. 1995; Vol. 47, pp 83-229.
161. Berkovich, R.; Mondal, J.; Paster, I.; Berne, B. J., Simulated Force Quench Dynamics Shows GB1 Protein Is Not a Two State Folder. *J. Phys. Chem. B* **2017**, 121 (20), 5162-5173.
162. Berezhkovskii, A. M.; Szabo, A.; Greives, N.; Zhou, H. X., Multidimensional reaction rate theory with anisotropic diffusion. *J. Chem. Phys.* **2014**, 141 (20).
163. Berkovich, R.; Garcia-Manyes, S.; Klafter, J.; Urbakh, M.; Fernandez, J. M., Hopping around an entropic barrier created by force. *Biochem. Biophys. Res. Commun.* **2010**, 403 (1), 133-137.
164. Cossio, P.; Hummer, G.; Szabo, A., On artifacts in single-molecule force spectroscopy. *Proc. Natl. Acad. Sci. USA* **2015**, 112 (46), 14248-14253.
165. Oliveberg, M.; Wolynes, P. G., The experimental survey of protein-folding energy landscapes. *Q. Rev. Biophys.* **2005**, 38 (3), 245-288.
166. Hartl, F. U.; Hayer-Hartl, M., Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **2009**, 16 (6), 574-581.
167. Schug, A.; Onuchic, J. N., From protein folding to protein function and biomolecular binding by energy landscape theory. *Curr Opin Pharmacol.* **2010**, 10 (6), 709-714.
168. Nussinov, R.; Wolynes, P. G., A second molecular biology revolution? The energy landscapes of biomolecular function. *Phys. Chem. Chem. Phys.* **2014**, 16 (14), 6321-6322.
169. Chong, S. H.; Im, H.; Ham, S., Explicit Characterization of the Free Energy Landscape of pKID-KIX Coupled Folding and Binding. *ACS Cent. Sci.* **2019**, 5 (8), 1342-1351.
170. Bustamante, C.; Chemla, Y. R.; Forde, N. R.; Izhaky, D., Mechanical processes in biochemistry. *Annu. Rev. Biochem.* **2004**, 73, 705-748.
171. Borgia, A.; Williams, P. M.; Clarke, J., Single-molecule studies of protein folding. *Annu. Rev. Biochem.* **2008**, 77, 101-125.
172. Huang, S. H.; Ratliff, K. S.; Matouschek, A., Protein unfolding by the mitochondrial membrane potential. *Nat. Struct. Biol.* **2002**, 9 (4), 301-307.

173. Rivas-Pardo, J. A.; Eckels, E. C.; Popa, I.; Kosuri, P.; Linke, W. A.; Fernandez, J. M., Work Done by Titin Protein Folding Assists Muscle Contraction. *Cell Rep.* **2016**, *14* (6), 1339-1347.
174. Ritort, F., Single-molecule experiments in biological physics: methods and applications. *J. Phys. Condens. Matter.* **2006**, *18* (32), R531-R583.
175. Elms, P. J.; Chodera, J. D.; Bustamante, C.; Marqusee, S., The molten globule state is unusually deformable under mechanical force. *Proc. Natl. Acad. Sci. USA* **2012**, *109* (10), 3796-3801.
176. Leake, M. C., The physics of life: one molecule at a time. *Philos. Trans. R. Soc. B: Biol. Sci.* **2013**, *368* (1611).
177. Deniz, A. A.; Mukhopadhyay, S.; Lemke, E. A., Single-molecule biophysics: at the interface of biology, physics and chemistry. *J. R. Soc. Interface* **2008**, *5* (18), 15-45.
178. Flomenbom, O.; Silbey, R. J., Utilizing the information content in two-state trajectories. *Proc. Natl. Acad. Sci. USA* **2006**, *103* (29), 10907-10910.
179. Schuler, B.; Eaton, W. A., Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* **2008**, *18* (1), 16-26.
180. Zhang, Y. J.; Dudko, O. K., A transformation for the mechanical fingerprints of complex biomolecular interactions. *Proc. Natl. Acad. Sci. USA* **2013**, *110* (41), 16432-16437.
181. Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E., Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **1997**, *276* (5315), 1109-1112.
182. Dudko, O. K.; Filippov, A. E.; Klafter, J.; Urbakh, M., Beyond the conventional description of dynamic force spectroscopy of adhesion bonds. *Proc. Natl. Acad. Sci. USA* **2003**, *100* (20), 11378-11381.
183. Hummer, G.; Szabo, A., Kinetics from nonequilibrium single-molecule pulling experiments. *Biophys. J.* **2003**, *85* (1), 5-15.
184. Kawakami, M.; Byrne, K.; Brockwell, D. J.; Radford, S. E.; Smith, D. A., Viscoelastic study of the mechanical unfolding of a protein by AFM. *Biophys. J.* **2006**, *91* (2), L16-L18.
185. Maitra, A.; Arya, G., Influence of pulling handles and device stiffness in single-molecule force spectroscopy. *Phys. Chem. Chem. Phys.* **2011**, *13* (5), 1836-1842.
186. Shoham, A.; Givli, S., Unfolding compactly folded molecular domains: Overall stiffness modifies the force-barrier relation. *Chem. Phys. Lett.* **2020**, *758*.
187. Best, R. B.; Paci, E.; Hummer, G.; Dudko, O. K., Pulling direction as a reaction coordinate for the mechanical unfolding of single molecules. *J. Phys. Chem. B* **2008**, *112* (19), 5968-5976.
188. Dudko, O. K.; Hummer, G.; Szabo, A., Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.* **2006**, *96* (10).
189. Zoldak, G.; Rief, M., Force as a single molecule probe of multidimensional protein energy landscapes. *Curr. Opin. Struct. Biol.* **2013**, *23* (1), 48-57.

190. Thirumalai, D.; Hyeon, C., RNA and protein folding: Common themes and variations. *Biochemistry* **2005**, *44* (13), 4957-4970.
191. Neupane, K.; Manuel, A. P.; Woodside, M. T., Protein folding trajectories can be described quantitatively by one-dimensional diffusion over measured energy landscapes. *Nat. Phys.* **2016**, *12* (7), 700-+.
192. Berkovich, R.; Hermans, R. I.; Popa, I.; Stirnemann, G.; Garcia-Manyes, S.; Berne, B. J.; Fernandez, J. M., Rate limit of protein elastic response is tether dependent. *Proc. Natl. Acad. Sci. USA* **2012**, *109* (36), 14416-14421.
193. Evans, E.; Ritchie, K., Dynamic strength of molecular adhesion bonds. *Biophys. J.* **1997**, *72* (4), 1541-1555.
194. Evans, E.; Ritchie, K., Strength of a weak bond connecting flexible polymer chains. *Biophys. J.* **1999**, *76* (5), 2439-2447.
195. Garg, A., Escape field distribution for escape from a metastable potential well subject to a steadily increasing bias field. *Phys. Rev. B* **1995**, *51* (21), 15592-15595.
196. Bell, G. I., Models for the Specific Adhesion of Cells to Cells. *Science* **1978**, *200* (4342), 618-627.
197. Cossio, P.; Hummer, G.; Szabo, A., Kinetic Ductility and Force-Spike Resistance of Proteins from Single-Molecule Force Spectroscopy. *Biophys. J.* **2016**, *111* (4), 832-840.
198. Fernandez, J. M.; Li, H. B., Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science* **2004**, *303* (5664), 1674-1678.
199. Chen, H.; Yuan, G. H.; Winardhi, R. S.; Yao, M. X.; Popa, I.; Fernandez, J. M.; Yan, J., Dynamics of Equilibrium Folding and Unfolding Transitions of Titin Immunoglobulin Domain under Constant Forces. *J. Am. Chem. Soc.* **2015**, *137* (10), 3540-3546.
200. 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.* **2007**, *93* (7), 2436-2446.
201. Hyeon, C. B.; Morrison, G.; Pincus, D. L.; Thirumalai, D., Refolding dynamics of stretched biopolymers upon force quench. *Proc. Natl. Acad. Sci. USA* **2009**, *106* (48), 20288-20293.
202. Tapia-Rojo, R.; Mora, M.; Board, S.; Walker, J.; Boujemaa-Paterski, R.; Medalia, O.; Garcia-Manyes, S., Enhanced statistical sampling reveals microscopic complexity in the talin mechanosensor folding energy landscape. *Nat. Phys.*
203. Kellermayer, M. S. Z.; Smith, S. B.; Granzier, H. L.; Bustamante, C., Folding-Unfolding Transitions in Single Titin Molecules Characterized with Laser Tweezers. *Science* **1997**, *276* (5315), 1112-1116.
204. Hoffmann, T.; Dougan, L., Single molecule force spectroscopy using polyproteins. *Chem. Soc. Rev.* **2012**, *41* (14), 4781-4796.

205. Brujic, J.; Hermans, R. I.; Walther, K. A.; Fernandez, J. M., Single-molecule force spectroscopy reveals signatures of glassy dynamics in the energy landscape of ubiquitin. *Nat. Phys.* **2006**, 2 (4), 282-286.
206. Bura, E.; Klimov, K.; Barsegov, V., Analyzing forced unfolding of protein tandems by ordered variates, 1: Independent unfolding times. *Biophys. J.* **2007**, 93 (4), 1100-1115.
207. Sumbul, F.; Marchesi, A.; Rico, F., History, rare, and multiple events of mechanical unfolding of repeat proteins. *J. Chem. Phys.* **2018**, 148 (12).
208. Chetrit, E.; Sharma, S.; Maayan, U.; Pelah, M. G.; Klausner, Z.; Popa, I.; Berkovich, R., Nonexponential kinetics captured in sequential unfolding of polyproteins over a range of loads. *Curr. Opin. Res. Biol.* **2022**, 4, 106-117.
209. Lannon, H.; Vanden-Eijnden, E.; Brujic, J., Force-Clamp Analysis Techniques Give Highest Rank to Stretched Exponential Unfolding Kinetics in Ubiquitin. *Biophys. J.* **2012**, 103 (10), 2215-2222.
210. Seifert, U.; Wintz, W.; Nelson, P., Straightening of Thermal Fluctuations in Semiflexible Polymers by Applied Tension. *Phys. Rev. Lett.* **1996**, 77 (27), 5389-5392.
211. Burov, S.; Jeon, J. H.; Metzler, R.; Barkai, E., Single particle tracking in systems showing anomalous diffusion: the role of weak ergodicity breaking. *Phys. Chem. Chem. Phys.* **2011**, 13 (5), 1800-1812.
212. Dentz, M.; Cortis, A.; Scher, H.; Berkowitz, B., Time behavior of solute transport in heterogeneous media: transition from anomalous to normal transport. *Adv. Water Resour.* **2004**, 27 (2), 155-173.
213. Metzler, R.; Klafter, J., The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.* **2000**, 339 (1), 1-77.
214. Barkai, E., Aging in subdiffusion generated by a deterministic dynamical system. *Phys. Rev. Lett.* **2003**, 90 (10).
215. Israelachvili, J.; Min, Y.; Akbulut, M.; Alig, A.; Carver, G.; Greene, W.; Kristiansen, K.; Meyer, E.; Pesika, N.; Rosenberg, K.; Zeng, H., Recent advances in the surface forces apparatus (SFA) technique. *Rep. Prog. Phys.* **2010**, 73 (3).
216. Perkins, T. T.; Smith, D. E.; Chu, S., Single polymer dynamics in an elongational flow. *Science* **1997**, 276 (5321), 2016-2021.
217. Akbari, E.; Shahhosseini, M.; Robbins, A.; Poirier, M. G.; Song, J. W.; Castro, C. E., Low cost and massively parallel force spectroscopy with fluid loading on a chip. *Nat. Commun.* **2022**, 13 (1), 6800.
218. Ozcelik, A.; Rufo, J.; Guo, F.; Gu, Y. Y.; Li, P.; Lata, J.; Huang, T. J., Acoustic tweezers for the life sciences. *Nat. Methods* **2018**, 15 (12), 1021-1028.
219. Yang, D.; Ward, A.; Halvorsen, K.; Wong, W. P., Multiplexed single-molecule force spectroscopy using a centrifuge. *Nat. Commun.* **2016**, 7.

220. Johnson, K. C.; Thomas, W. E., How Do We Know when Single-Molecule Force Spectroscopy Really Tests Single Bonds? *Biophys. J.* **2018**, *114* (9), 2032-2039.
221. Kluger, C.; Braun, L.; Sedlak, S. M.; Pippig, D. A.; Bauer, M. S.; Miller, K.; Milles, L. F.; Gaub, H. E.; Vogel, V., Different Vinculin Binding Sites Use the Same Mechanism to Regulate Directional Force Transduction. *Biophys. J.* **2020**, *118* (6), 1344-1356.
222. Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M., Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. USA* **2012**, *109* (12), E690-E697.
223. Jadhav, V. S.; Bruggemann, D.; Wruck, F.; Hegner, M., Single-molecule mechanics of protein-labelled DNA handles. *Beilstein J Nanotech* **2016**, *7*, 138-148.
224. Popa, I.; Rivas-Pardo, J. A.; Eckels, E. C.; Echelman, D. J.; Badilla, C. L.; Valle-Orero, J.; Fernandez, J. M., A HaloTag Anchored Ruler for Week-Long Studies of Protein Dynamics. *J. Am. Chem. Soc.* **2016**, *138* (33), 10546-10553.
225. Sedlak, S. M.; Schendel, L. C.; Gaub, H. E.; Bernardi, R. C., Streptavidin/biotin: Tethering geometry defines unbinding mechanics. *Sci. Adv.* **2020**, *6* (13), eaay5999.
226. Milles, L. F.; Schulten, K.; Gaub, H. E.; Bernardi, R. C., Molecular mechanism of extreme mechanostability in a pathogen adhesin. *Science* **2018**, *359* (6383), 1527-1532.
227. Schoeler, C.; Malinowska, K. H.; Bernardi, R. C.; Milles, L. F.; Jobst, M. A.; Durner, E.; Ott, W.; Fried, D. B.; Bayer, E. A.; Schulten, K.; Gaub, H. E.; Nash, M. A., Ultrastable cellulosome-adhesion complex tightens under load. *Nat. Commun.* **2014**, *5*.
228. Lei, H.; Zhang, J. S.; Li, Y.; Wang, X.; Qin, M.; Wang, W.; Cao, Y., Histidine-Specific Bioconjugation for Single-Molecule Force Spectroscopy. *Acs Nano* **2022**, *16* (9), 15440-15449.
229. Deng, Y. B.; Wu, T.; Wang, M. D.; Shi, S. C.; Yuan, G. D.; Li, X.; Chong, H. C.; Wu, B.; Zheng, P., Enzymatic biosynthesis and immobilization of polyprotein verified at the single-molecule level. *Nat. Commun.* **2019**, *10*.
230. Morgan, H. E.; Turnbull, W. B.; Webb, M. E., Challenges in the use of sortase and other peptide ligases for site-specific protein modification. *Chem. Soc. Rev.* **2022**, *51* (10), 4121-4145.
231. Moses, J. E.; Moorhouse, A. D., The growing applications of click chemistry. *Chem. Soc. Rev.* **2007**, *36* (8), 1249-1262.
232. Rospars, J. P.; Meyer-Vernet, N., Force per cross-sectional area from molecules to muscles: a general property of biological motors. *Roy. Soc. Open Sci.* **2016**, *3* (7).
233. 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.* **2013**, *8* (7), 1261-1276.
234. Edwards, D. T.; Faulk, J. K.; Sanders, A. W.; Bull, M. S.; Walder, R.; LeBlanc, M.-A.; Sousa, M. C.; Perkins, T. T., Optimizing 1- μ s-Resolution Single-Molecule Force

- Spectroscopy on a Commercial Atomic Force Microscope. *Nano Lett.* **2015**, *15* (10), 7091-7098.
235. Bustamante, C. J.; Chemla, Y. R.; Liu, S.; Wang, M. D., Optical tweezers in single-molecule biophysics. *Nat. Rev. Methods Primers* **2021**, *1* (1), 25.
 236. Greenleaf, W. J.; Woodside, M. T.; Abbondanzieri, E. A.; Block, S. M., Passive all-optical force clamp for high-resolution laser trapping. *Phys. Rev. Lett.* **2005**, *95* (20).
 237. Blazquez-Castro, A., Optical Tweezers: Phototoxicity and Thermal Stress in Cells and Biomolecules. *Micromachines* **2019**, *10* (8).
 238. Grashoff, C.; Hoffman, B. D.; Brenner, M. D.; Zhou, R. B.; Parsons, M.; Yang, M. T.; McLean, M. A.; Sligar, S. G.; Chen, C. S.; Ha, T.; Schwartz, M. A., Measuring mechanical tension across vinculin reveals regulation of focal adhesion dynamics. *Nature* **2010**, *466* (7303), 263-U143.
 239. Cost, A.-L.; 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.* **2015**, *8* (1), 96-105.
 240. Nowitzke, J.; Popa, I., What Is the Force-per-Molecule Inside a Biomaterial Having Randomly Oriented Units? *J. Phys. Chem. Lett.* **2022**, *13* (31), 7139-7146.
 241. Jurchenko, C.; Chang, Y.; Narui, Y.; Zhang, Y.; Salaita, K. S., Integrin-Generated Forces Lead to Streptavidin-Biotin Unbinding in Cellular Adhesions. *Biophys. J.* **2014**, *106* (7), 1436-1446.
 242. Liu, Y.; Yehl, K.; Narui, Y.; Salaita, K., Tension Sensing Nanoparticles for Mechano-Imaging at the Living/Nonliving Interface. *J. Am. Chem. Soc.* **2013**, *135* (14), 5320-5323.
 243. Wang, X. F.; Ha, T., Defining Single Molecular Forces Required to Activate Integrin and Notch Signaling. *Science* **2013**, *340* (6135), 991-994.
 244. Bender, R.; Salaita, K., *Molecular Force Sensors*. American Chemical Society: 2021; p -1.
 245. Linke, W. A.; Popov, V. I.; Pollack, G. H., Passive and active tension in single cardiac myofibrils. *Biophys. J.* **1994**, *67* (2), 782-792.
 246. Korte, F. S.; McDonald, K. S., Sarcomere length dependence of rat skinned cardiac myocyte mechanical properties: dependence on myosin heavy chain. *J. Physiol.* **2007**, *581* (2), 725-739.
 247. Lv, S.; Dudek, D. M.; Cao, Y.; Balamurali, M. M.; Gosline, J.; Li, H. B., Designed biomaterials to mimic the mechanical properties of muscles. *Nature* **2010**, *465* (7294), 69-73.
 248. Khoury, L. R.; Nowitzke, J.; Shmilovich, K.; Popa, I., Study of Biomechanical Properties of Protein-Based Hydrogels Using Force-Clamp Rheometry. *Macromolecules* **2018**, *51* (4), 1441-1452.
 249. Shmilovich, K.; Popa, I., Modeling Protein-Based Hydrogels under Force. *Phys. Rev. Lett.* **2018**, *121* (16).

250. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30* (10), 1545-1614.
251. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19-25.
252. Martin-Garcia, F.; Papaleo, E.; Gomez-Puertas, P.; Boomsma, W.; Lindorff-Larsen, K., Comparing Molecular Dynamics Force Fields in the Essential Subspace. *Plos One* **2015**, *10* (3).
253. Mykuliak, V. V.; Sikora, M.; Booth, J. J.; Cieplak, M.; Shalashilin, D. V.; Hytonen, V. P., Mechanical Unfolding of Proteins-A Comparative Nonequilibrium Molecular Dynamics Study. *Biophys. J.* **2020**, *119* (5), 939-949.
254. Stirnemann, G., Recent Advances and Emerging Challenges in the Molecular Modeling of Mechanobiological Processes. *J. Phys. Chem. B* **2022**, *126* (7), 1365-1374.
255. Hoffmann, T.; Tych, K. M.; Hughes, M. L.; Brockwell, D. J.; Dougan, L., Towards design principles for determining the mechanical stability of proteins. *Phys. Chem. Chem. Phys.* **2013**, *15* (38), 15767-15780.
256. Hummer, G.; Szabo, A., Free energy profiles from single-molecule pulling experiments. *Proc. Natl. Acad. Sci. USA* **2010**, *107* (50), 21441-21446.
257. Gebhardt, J. C. M.; Bornschlogla, T.; Rief, M., Full distance-resolved folding energy landscape of one single protein molecule. *Proc. Natl. Acad. Sci. USA* **2010**, *107* (5), 2013-2018.
258. Woodside, M. T.; Block, S. M., Reconstructing Folding Energy Landscapes by Single-Molecule Force Spectroscopy. *Annu Rev Biophys* **2014**, *43*, 19-39.
259. Wiita, A. P.; Perez-Jimenez, R.; Walther, K. A.; Grater, F.; Berne, B. J.; Holmgren, A.; Sanchez-Ruiz, J. M.; Fernandez, J. M., Probing the chemistry of thioredoxin catalysis with force. *Nature* **2007**, *450* (7166), 124-+.
260. Eckels, E. C.; Haldar, S.; Tapia-Rojo, R.; Rivas-Pardo, J. A.; Fernandez, J. M., The Mechanical Power of Titin Folding. *Cell Rep.* **2019**, *27* (6), 1836-+.
261. Alegre-Cebollada, J.; Kosuri, P.; Giganti, D.; Eckels, E.; Rivas-Pardo, J. A.; Hamdani, N.; Warren, C. M.; Solaro, R. J.; Linke, W. A.; Fernandez, J. M., S-Glutathionylation of Cryptic Cysteines Enhances Titin Elasticity by Blocking Protein Folding. *Cell* **2014**, *156* (6), 1235-1246.

262. del Rio, A.; Perez-Jimenez, R.; Liu, R. C.; Roca-Cusachs, P.; Fernandez, J. M.; Sheetz, M. P., Stretching Single Talin Rod Molecules Activates Vinculin Binding. *Science* **2009**, 323 (5914), 638-641.
263. Popa, I.; Saitis, F., Using Magnets and Flexible 3D-Printed Structures to Illustrate Protein (Un)foldings. *J. Chem. Educ.* **2022**.
264. Berkovich, R.; Fernandez, V. I.; Stirnemann, G.; Valle-Orero, J.; Fernandez, J. M., Segmentation and the Entropic Elasticity of Modular Proteins. *J. Phys. Chem. Lett.* **2018**, 9 (16), 4707-4713.
265. Li, H. B.; Linke, W. A.; Oberhauser, A. F.; Carrion-Vazquez, M.; Kerkvilliet, J. G.; Lu, H.; Marszalek, P. E.; Fernandez, J. M., Reverse engineering of the giant muscle protein titin. *Nature* **2002**, 418 (6901), 998-1002.
266. Granzier, H. L.; Labeit, S., The giant protein titin - A major player in myocardial mechanics, signaling, and disease. *Circ. Res.* **2004**, 94 (3), 284-295.
267. Linke, W. A., Sense and stretchability: The role of titin and titin-associated proteins in myocardial stress-sensing and mechanical dysfunction. *Cardiovasc. Res.* **2008**, 77 (4), 637-648.
268. Giganti, D.; Yan, K.; Badilla, C. L.; Fernandez, J. M.; Alegre-Cebollada, J., Disulfide isomerization reactions in titin immunoglobulin domains enable a mode of protein elasticity. *Nat. Commun.* **2018**, 9.
269. Frantz, C.; Stewart, K. M.; Weaver, V. M., The extracellular matrix at a glance. *J. Cell. Sci.* **2010**, 123 (24), 4195-4200.
270. Zamir, E.; Geiger, B., Molecular complexity and dynamics of cell-matrix adhesions. *J. Cell. Sci.* **2001**, 114 (20), 3583-3590.
271. Vogel, V., Mechanotransduction involving multimodular proteins: Converting force into biochemical signals. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, 35, 459-488.
272. Mao, Y.; Schwarzbauer, J. E., Fibronectin fibrillogenesis, a cell-mediated matrix assembly process. *Matrix Biol.* **2005**, 24 (6), 389-399.
273. Humphrey, J. D.; Dufresne, E. R.; Schwartz, M. A., Mechanotransduction and extracellular matrix homeostasis. *Nat. Rev. Mol. Cell Biol.* **2014**, 15 (12), 802-812.
274. Shattil, S. J.; Kim, C.; Ginsberg, M. H., The final steps of integrin activation: the end game. *Nat. Rev. Mol. Cell Biol.* **2010**, 11 (4), 288-300.
275. Zhu, J. Q.; Zhu, J. H.; Springer, T. A., Complete integrin headpiece opening in eight steps. *J. Cell Biol.* **2013**, 201 (7), 1053-1068.
276. Haining, A. W. M.; Lieberthal, T. J.; Hernandez, A. D., Talin: a mechanosensitive molecule in health and disease. *Faseb J* **2016**, 30 (6), 2073-2085.
277. Yao, M. X.; Goult, B. T.; Klapholz, B.; Hu, X.; Toseland, C. P.; Guo, Y. J.; Cong, P. W.; Sheetz, M. P.; Yan, J., The mechanical response of talin. *Nat. Commun.* **2016**, 7.

278. Yao, M. X.; Qiu, W.; Liu, R. C.; Efremov, A. K.; Cong, P. W.; Seddiki, R.; Payre, M.; Lim, C. T.; Ladoux, B.; Mege, R. M.; Yan, J., Force-dependent conformational switch of alpha-catenin controls vinculin binding. *Nat. Commun.* **2014**, *5*.
279. Seddiki, R.; Narayana, G. H. N. S.; Strale, P. O.; Balcioglu, H. E.; Peyret, G.; Yao, M. X.; Le, A. P.; Lim, C. T.; Yan, J.; Ladoux, B.; Mege, R. M., Force-dependent binding of vinculin to alpha-catenin regulates cell-cell contact stability and collective cell behavior. *Mol. Biol. Cell.* **2018**, *29* (4), 380-388.
280. Dahal, N.; Sharma, S.; Phan, B.; Eis, A.; Popa, I., Mechanical regulation of talin through binding and history-dependent unfolding. *Sci. Adv.* **2022**, *8* (28).
281. Ge, J. P.; Elferich, J.; Goehring, A.; Zhao, H. Y.; Schuck, P.; Gouaux, E., Structure of mouse protocadherin 15 of the stereocilia tip link in complex with LHFPL5. *Elife* **2018**, *7*.
282. Sotomayor, M.; Weihofen, W. A.; Gaudet, R.; Corey, D. P., Structure of a force-conveying cadherin bond essential for inner-ear mechanotransduction. *Nature* **2012**, *492* (7427), 128-+.
283. Choudhary, D.; Narui, Y.; Neel, B. L.; Wimalasena, L. N.; Klanseck, C. F.; De-la-Torre, P.; Chen, C. H.; Araya-Secchi, R.; Tamilselvan, E.; Sotomayor, M., Structural determinants of protocadherin-15 mechanics and function in hearing and balance perception. *Proc. Natl. Acad. Sci. USA* **2020**, *117* (40), 24837-24848.
284. Kachar, B.; Parakkal, M.; Kurc, M.; Zhao, Y.; Gillespie, P. G., High-resolution structure of hair-cell tip links. *Proc. Natl. Acad. Sci. USA* **2000**, *97* (24), 13336-13341.
285. Hackney, C. M.; Furness, D. N., The composition and role of cross links in mechanoelectrical transduction in vertebrate sensory hair cells. *J. Cell. Sci.* **2013**, *126* (8), 1721-1731.
286. Kliuchko, M.; Heinonen-Guzejev, M.; Monacis, L.; Gold, B. P.; Heikkila, K. V.; Spinosa, V.; Tervaniemi, M.; Brattico, E., The association of noise sensitivity with music listening, training, and aptitude. *Noise Health* **2015**, *17* (78), 350-357.
287. Fettiplace, R.; Kim, K. X., The Physiology of Mechanoelectrical Transduction Channels in Hearing. *Physiol. Rev.* **2014**, *94* (3), 951-986.
288. Hackney, C. M.; Fettiplace, R.; Furness, D. N., The Functional-Morphology of Stereociliary Bundles on Turtle Cochlear Hair-Cells. *Hearing Res.* **1993**, *69* (1-2), 163-175.
289. Arora, N.; Hazra, J. P.; Rakshit, S., Anisotropy in mechanical unfolding of protein upon partner-assisted pulling and handle-assisted pulling. *Commun. Biol.* **2021**, *4* (1).
290. Bartsch, T. F.; Hengel, F. E.; Oswald, A.; Dionne, G.; Chipendo, I. V.; Mangat, S. S.; El Shatanofy, M.; Shapiro, L.; Muller, U.; Hudspeth, A. J., Elasticity of individual protocadherin 15 molecules implicates tip links as the gating springs for hearing. *Proc. Natl. Acad. Sci. USA* **2019**, *116* (22), 11048-11056.
291. Persat, A.; Nadell, C. D.; Kim, M. K.; Ingremau, F.; Siryaporn, A.; Drescher, K.; Wingreen, N. S.; Bassler, B. L.; Gitai, Z.; Stone, H. A., The Mechanical World of Bacteria. *Cell* **2015**, *161* (5), 988-997.

292. Dahal, N.; Nowitzke, J.; Eis, A.; Popa, I., Binding-Induced Stabilization Measured on the Same Molecular Protein Substrate Using Single-Molecule Magnetic Tweezers and Heterocovalent Attachments. *J. Phys. Chem. B* **2020**, *124* (16), 3283-3290.
293. Miller, E.; Garcia, T.; Hultgren, S.; Oberhauser, A. F., The mechanical properties of E-coli type 1 pili measured by atomic force microscopy techniques. *Biophys. J.* **2006**, *91* (10), 3848-3856.
294. Nordenfelt, P.; Waldemarson, S.; Linder, A.; Morgelin, M.; Karlsson, C.; Malmstrom, J.; Bjorck, L., Antibody orientation at bacterial surfaces is related to invasive infection. *J Exp Med* **2012**, *209* (13), 2367-2381.
295. Johnson, J. R., Virulence Factors in Escherichia-Coli Urinary-Tract Infection. *Clin. Microbiol. Rev.* **1991**, *4* (1), 80-128.
296. Hahn, E.; Wild, P.; Hermanns, U.; Sebbel, P.; Glockshuber, R.; Haner, M.; Taschner, N.; Burkhard, P.; Aebi, U.; Muller, S. A., Exploring the 3D molecular architecture of Escherichia coli type 1 pili. *J. Mol Biol.* **2002**, *323* (5), 845-857.
297. Alonso-Caballero, A.; Schonfelder, J.; Poly, S.; Corsetti, F.; De Sancho, D.; Artacho, E.; Perez-Jimenez, R., Mechanical architecture and folding of E. coli type 1 pilus domains. *Nat. Commun.* **2018**, *9*.
298. Proft, T.; Baker, E. N., Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease. *Cell. Mol. Life Sci.* **2009**, *66* (4), 613-635.
299. Remaut, H.; Waksman, G., Protein-protein interaction through beta-strand addition. *Trends Biochem Sci* **2006**, *31* (8), 436-444.
300. Zyla, D. S.; Prota, A. E.; Capitani, G.; Glockshuber, R., Alternative folding to a monomer or homopolymer is a common feature of the type 1 pilus subunit FimA from enteroinvasive bacteria. *J. Biol. Chem.* **2019**, *294* (27), 10553-10563.
301. Baorto, D. M.; Gao, Z. M.; Malaviya, R.; Dustin, M. L.; vanderMerwe, A.; Lublin, D. M.; Abraham, S. N., Survival of FimH-expressing enterobacteria in macrophages relies on glycolipid traffic. *Nature* **1997**, *389* (6651), 636-639.
302. Andersson, M.; Uhlin, B. E.; Fallman, E., The biomechanical properties of E-coli Pili for urinary tract attachment reflect the host environment. *Biophys. J.* **2007**, *93* (9), 3008-3014.
303. Graille, M.; Stura, E. A.; Housden, N. G.; Beckingham, J. A.; Bottomley, S. P.; Beale, D.; Taussig, M. J.; Sutton, B. J.; Gore, M. G.; Charbonnier, J. B., Complex between Peptostreptococcus magnus protein L and a human antibody reveals structural convergence in the interaction modes of Fab binding proteins. *Structure* **2001**, *9* (8), 679-687.
304. Grover, R. K.; Zhu, X. Y.; Nieusma, T.; Jones, T.; Boero, I.; MacLeod, A. S.; Mark, A.; Niessen, S.; Kim, H. J.; Kong, L.; Assad-Garcia, N.; Kwon, K.; Chesi, M.; Smider, V. V.; Salomon, D. R.; Jelinek, D. F.; Kyle, R. A.; Pyles, R. B.; Glass, J. I.; Ward, A. B.; Wilson, I. A.; Lerner, R. A., A Structurally Distinct Human Mycoplasma Protein that Generically Blocks Antigen-Antibody Union. *Science* **2014**, *343* (6171), 656-661.

305. Boyanova, L.; Markovska, R.; Mitov, I., Virulence arsenal of the most pathogenic species among the Gram-positive anaerobic cocci, *Finegoldia magna*. *Anaerobe* **2016**, *42*, 145-151.
306. Neumann, A.; Bjorck, L.; Frick, I. M., *Finegoldia magna*, an Anaerobic Gram-Positive Bacterium of the Normal Human Microbiota, Induces Inflammation by Activating Neutrophils. *Front. Microbiol.* **2020**, *11*.
307. Beckingham, J. A.; Housden, N. G.; Muir, N. M.; Bottomley, S. P.; Gore, M. G., Studies on a single immunoglobulin-binding domain of protein L from *Peptostreptococcus magnus*: the role of tyrosine-53 in the reaction with human IgG. *Biochem. J.* **2001**, *353*, 395-401.
308. Rousseau, A.; Bertolotti, A., Regulation of proteasome assembly and activity in health and disease. *Nat. Rev. Mol. Cell Biol.* **2018**, *19* (11), 697-712.
309. Beck, F.; Unverdorben, P.; Bohn, S.; Schweitzer, A.; Pfeifer, G.; Sakata, E.; Nickell, S.; Plitzko, J. M.; Villa, E.; Baumeister, W.; Forster, F., Near-atomic resolution structural model of the yeast 26S proteasome. *Proc. Natl. Acad. Sci. USA* **2012**, *109* (37), 14870-14875.
310. Mao, Y., Structure, Dynamics and Function of the 26S Proteasome. In *Macromolecular Protein Complexes III: Structure and Function*, Harris, J. R.; Marles-Wright, J., Eds. Springer International Publishing: Cham, 2021; pp 1-151.
311. Stahlhut, S. G.; Alqarzaee, A. A.; Jensen, C.; Fisker, N. S.; Pereira, A. R.; Pinho, M. G.; Thomas, V. C.; Frees, D., The ClpXP protease is dispensable for degradation of unfolded proteins in *Staphylococcus aureus*. *Sci. Rep.* **2017**, *7*.
312. Fei, X.; Bell, T. A.; Jenni, S.; Stinson, B. M.; Baker, T. A.; Harrison, S. C.; Sauer, R. T., Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate. *Elife* **2020**, *9*.
313. Rudolph, F.; Huttemeister, J.; Lopes, K. D.; Juttner, R.; Yu, L.; Bergmann, N.; Friedrich, D.; Preibisch, S.; Wagner, E.; Lehnart, S. E.; Gregorio, C. C.; Gotthardt, M., Resolving titin's lifecycle and the spatial organization of protein turnover in mouse cardiomyocytes. *Proc. Natl. Acad. Sci. USA* **2019**, *116* (50), 25126-25136.