

Alma Mater Studiorum - Università di Bologna

Dottorato di Ricerca in:

Biologia dei Sistemi Molecolari e Cellulari – Ciclo XX (2008)
Settore scientifico-disciplinare di afferenza: CHIM/06 - CHIMICA
ORGANICA

Each One Teaches One: Characterizing
Active Forms of Proteins By Single
Molecule Force Spectroscopy

Presentata da: Massimo Sandal

Supervisore:

Prof. Bruno Samorì

Coordinatore:

Prof. Vincenzo Scarlato

Esame finale anno 2008

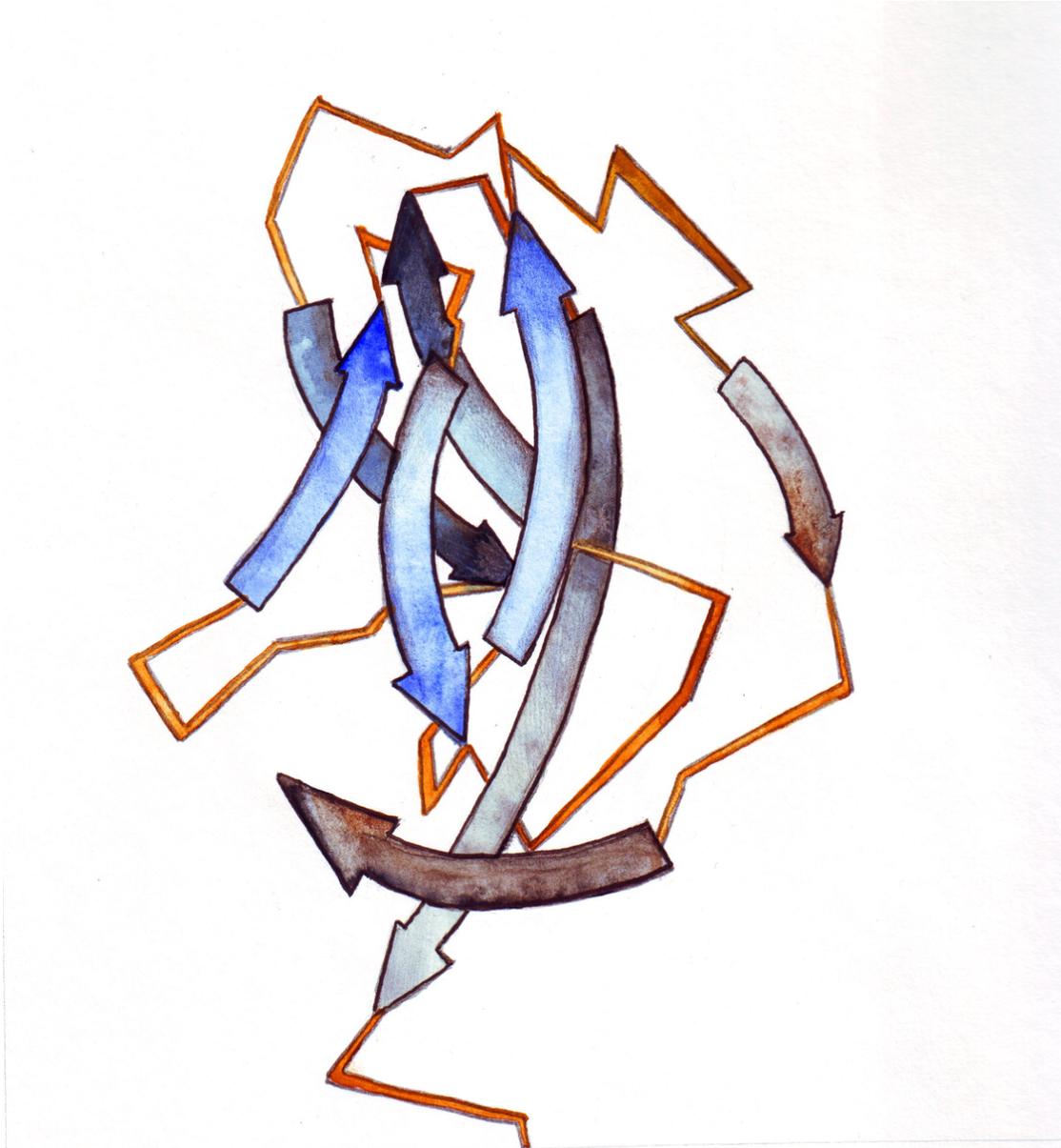


Figure 1: Ambra Galassi. I27 module of human titin. Pencil on paper, 2008.

Contents

1	Preface	3
1.1	Published works	4
1.2	Coauthors	5
I	Introduction: The structural diversity of a single protein	7
2	The structural diversity of a single protein	9
2.1	The roots of protein activity	9
2.2	The local folding problem	11
3	Far from classic equilibrium: mechanical forces at the protein level	15
3.1	Protein mechanobiology	16
3.1.1	The mechanochemical crosstalk between extracellular matrix and adherent cells	16
3.1.2	The transduction of mechanical signals by protein chains	17
3.1.3	Examples of mechanical regulation of proteins	19
3.2	Sacrificial bonds and protein adhesion	20
3.3	Force and intermolecular interactions	21
4	Hidden in a conformational equilibrium: intrinsically unstructured proteins and amyloids	25
4.1	Intrinsically unstructured proteins	25
4.1.1	Discovery of intrinsically unstructured proteins	26

CONTENTS

4.1.2	Intrinsic unfolding and function	27
4.1.3	Prediction of intrinsic disorder	28
4.1.4	The structural states of IUPs	29
4.2	Amyloid fibrils	31
4.2.1	Amyloid fibril structure	31
4.2.2	Conformational similarities and diversities of amyloids . . .	34
4.3	Mechanism of amyloid fibril formation	34
4.3.1	General features	35
4.3.2	Oligomers and protofibrils	35
4.3.3	The seeded and conformational models	38
4.4	Conformational disorder and amyloid aggregation	41
4.5	Single molecule techniques for intrinsically unstructured proteins and aggregation	43
II	Single molecule force spectroscopy of proteins	45
5	Single molecule force spectroscopy	47
5.1	Single molecule techniques	47
5.2	Atomic force microscopy	49
5.2.1	AFM as an imaging device	49
5.2.2	AFM as a dynamometer	51
5.2.3	AFM calibration	51
5.3	Constant speed force spectroscopy	53
5.3.1	Features of a force curve	53
6	SMFS investigation of multimodular proteins	57
6.1	Protein force spectroscopy operation	57
6.2	Why multimodular?	58
6.3	The mechanical unfolding of a multimodular protein	61
6.3.1	Understanding the sawtooth pattern	61
6.3.2	Polymer elasticity and the WLC model	62
6.3.3	Deviations from the WLC model	64
6.3.4	Basic information extracted from protein SMFS data	64
6.4	The polyprotein strategy for SMFS protein analysis	67

CONTENTS

6.4.1	Building polyproteins	68
6.4.2	Polyprotein design and heteromeric polyproteins	69
6.4.3	Open problems in polyprotein design	70
6.5	Molecular determinants of the mechanical stability of proteins . . .	73
6.5.1	Secondary and tertiary structure	73
6.5.2	Pulling geometry	74
6.5.3	Disulfide bonds	76
6.6	The physics of protein mechanical unfolding	76
6.6.1	Equilibrium and non-equilibrium unfolding	76
6.6.2	Description of single barrier mechanical unfolding	78
6.7	Mapping the mechanical unfolding pathway	81
6.7.1	Dynamic SMFS	81
6.7.2	Protein engineering analysis	83
6.7.3	Molecular dynamics simulations of mechanical unfolding . .	84
6.8	SMFS data analysis	85
6.8.1	Recognizing aspecific interactions	85
6.8.2	Recognizing single molecule events	86
6.8.3	Recognizing objects	87
6.8.4	Recognizing patterns	88
6.8.5	Data analysis automation	89
6.9	Force clamp SMFS of multimodular proteins	89
6.10	Limits and perspectives	92
6.10.1	SMFS as a structural technique	92
6.10.2	Increasing data throughput	92

III Hierarchical mechanochemical switches in angiotatin **95**

7	Background and theoretical elaboration	99
7.1	Disulfide bonds as molecular switches	99
7.1.1	Disulfide bond cleavage as a mechanism for protein function control	100
7.2	Coupling the redox equilibrium of a disulfide bond with an external mechanical force	102

CONTENTS

7.2.1	Mechanical effects of disulfide bonds	102
7.2.2	The redox modulation of the protein mechanics can switch new protein functionalities	104
7.2.3	Candidate proteins	105
7.2.4	Mechanoredox controls in the extracellular matrix	106
7.3	Angiostatin	106
7.3.1	Kringle structure	106
7.3.2	Angiostatin production	107
7.3.3	Angiostatin proposed mechanisms of action	108
7.3.4	Kringle structures activities	109
7.3.5	Kringle activities and disulfide bonds	110
7.4	A mechanoredox hypothesis for angiostatin	110
7.4.1	Mechanical forces acting on angiostatin	110
7.4.2	Forced unfolding accessibility of kringle domains	111
8	Rationale	113
9	Materials and methods	115
9.1	Force spectroscopy experiments with dithiotreitol	115
9.2	Force spectroscopy experiments with thioredoxin	116
9.3	Force spectroscopy raw data analysis	116
9.4	Monte Carlo simulations	117
9.5	Steered molecular dynamics simulations.	118
9.6	Protein-protein docking simulation.	118
10	Results	121
10.1	Steered molecular-dynamics simulations of the angiostatin mechan- ical unfolding under reducing conditions	121
10.2	Stretching single angiostatin molecules under reducing conditions by SMFS.	123
10.3	Stretching angiostatin in a condition that mimics the redox envi- ronment on the surface of a tumor cell.	130
10.4	Mechanochemical tuning of the binding of angiostatin with cell sur- face ATPase.	132

CONTENTS

10.5	Mechanochemical tuning of the exposure of the K4 fragments with the highest antimigratory activity.	134
11	Discussion and perspectives	137
11.1	The hierarchy of redox and mechanical regulation	139
IV	Probing the conformational diversity of the intrinsically unstructured protein α-synuclein at the single molecule level	141
12	Background	145
12.1	The α -synuclein protein	145
12.1.1	α -synuclein sequence	145
12.1.2	α -synuclein physiological function	146
12.1.3	α -synuclein in pathology	146
12.2	The problem of α -synuclein conformational equilibrium	147
12.2.1	Monomeric conformers	148
12.2.2	Polymeric conformers	149
12.3	The problem of α -synuclein aggregation	150
12.3.1	The partially folded intermediate hypothesis	151
12.3.2	The extended conformation hypothesis	151
13	Rationale	153
14	Materials and methods	155
14.1	Polyprotein design and expression	155
14.2	Circular dichroism experiments	158
14.3	Fluorescence experiments	158
14.4	Buffer elemental analysis	160
14.5	Surface preparation	160
14.6	Force spectroscopy experiments	160
14.7	Data analysis	161

15 Results	163
15.1 α -synuclein single molecules display conformational diversity on the AFM timescale	163
15.2 Conformers with the mechanical properties of a random coil	163
15.3 Conformers with the mechanical properties of a chain containing a β -like structured segment	165
15.3.1 Assignment of seven-peaked force curves to the pulling of a single 3S3 molecule	166
15.3.2 Features of the β -like α -synuclein segment	167
15.4 Conformers displaying mechanically weak interactions	168
15.4.1 Assignment of the mechanically weak interactions to the pulling of the α -synuclein module	168
15.4.2 Features of the mechanically weak interactions	169
15.5 Shift of conformational equilibrium due to sequence and buffer alterations	169
15.5.1 1 μ M Cu^{2+} buffer	169
15.5.2 A30P mutant	171
15.5.3 500 mM Tris	171
15.6 Fluorescence spectroscopy shows evidence of I27- α -synuclein interactions	171
15.7 Circular dichroism of the 3S3 construct	172
15.8 The distribution of mechanically weak interaction lengths is differently affected by variations in the environment or protein sequence.	172
16 Discussion	177
16.1 Conformational equilibria of α -synuclein as probed by AFM-based CS-SMFS	177
16.2 The population β -like conformers is correlated with the α -synuclein propensity to pathological aggregation	178
16.3 Are β -like structures really β ?	179
16.3.1 Experimental evidence	179
16.3.2 Evidence from the literature	180
16.4 The nature of mechanically weak interactions	181
16.4.1 An electrostatic model of I27- α -synuclein interaction	181

CONTENTS

16.4.2	The influence of sequence and chemical conditions on I27- α -synuclein interactions	183
16.4.3	Other possible structural assignments	184
16.4.4	Multimodular constructs for the measure of interprotein interactions?	185
16.5	Effect of the I27 handlers	186
17	Conclusion and perspectives	189
17.1	SMFS for the intrinsically unfolded proteins	189
17.2	AFM-based SMFS as a structural technique	190
17.3	Testing therapies against proteopathies by mean of SMFS	191
V	Hooke: an open source software platform for SMFS data analysis	193
18	Background	197
19	Design and implementation	199
19.1	An open approach to SMFS data analysis	199
19.1.1	A platform for SMFS data analysis	199
19.1.2	The Open Source philosophy	200
19.1.3	Platform independence	201
19.1.4	Independence from commercial software	201
19.2	The <i>Hooke</i> implementation	202
19.2.1	Choice of the programming environment	202
19.2.2	The <i>Hooke</i> architecture	204
19.3	Example plugins and features	206
19.3.1	Peak recognition and filtering	206
19.3.2	Mass analysis and selection of force curve data	208
VI	Final remarks	211
20	Sources	213

CONTENTS

21 Thank you	215
21.1 Family	215
21.2 Laboratory	216
21.3 Significant others	218
21.4 Software	219
Bibliography	220

CONTENTS

Do you want to see what human eyes have never seen? Look at the moon. Do you want to hear what ears have never heard? Listen to the bird's cry. Do you want to touch what hands have never touched? Touch the earth.

Jorge Luis Borges

The diversity of the phenomena of nature is so great, and the treasures hidden in the heavens so rich, precisely in order that the human mind shall never be lacking in fresh nourishment

Johannes Kepler

Confusion is next and next after that is the truth

Sonic Youth , "Confusion Is Next"

Tra frammenti di tecniche

sotto prodigi incerti

un affanno continuo

radio accese

tremo

per un non so

che si trova

a volte

a caso

CCCP, "And The Radio Plays"

CONTENTS

Chapter 1

Preface

This Ph.D. candidate thesis collects the research work I conducted under the supervision of Prof. Bruno Samorí in 2005, 2006 and 2007. Some parts of this work included in the Part III have been begun by myself during my undergraduate thesis in the same laboratory and then completed during the initial part of my Ph.D. thesis: the whole results have been included for the sake of understanding and completeness.

During my graduate studies I worked on two very different protein systems. The theoretical *trait d'union* between these studies, at the biological level, is the acknowledgement that protein biophysical and structural studies must, in many cases, take into account the dynamical states of protein conformational equilibria and of local physico-chemical conditions where the system studied actually performs its function. This is introduced in the introductory part in Chapter 2. Two different examples of this are presented: the structural significance deriving from the action of mechanical forces *in vivo* (Chapter 3) and the complexity of conformational equilibria in intrinsically unstructured proteins and amyloid formation (Chapter 4).

My experimental work investigated both these examples by using in both cases the single molecule force spectroscopy technique (described in Chapter 5 and Chapter 6). The work conducted on angiostatin focused on the characterization of the relationships between the mechanochemical properties and the mechanism of action of the angiostatin protein, and most importantly their intertwining with the further layer of complexity due to disulfide redox equilibria (Part III). These stud-

ies were accompanied concurrently by the elaboration of a theoretical model for a novel signalling pathway that may be relevant in the extracellular space, detailed in Chapter 7.2.

The work conducted on α -synuclein (Part IV) instead brought a whole new twist to the single molecule force spectroscopy methodology, applying it as a structural technique to elucidate the conformational equilibria present in intrinsically unstructured proteins. These equilibria are of utmost interest from a biophysical point of view, but most importantly because of their direct relationship with amyloid aggregation and, consequently, the aetiology of relevant pathologies like Parkinson's disease. The work characterized, for the first time, conformational equilibria in an intrinsically unstructured protein at the single molecule level and, again for the first time, identified a monomeric folded conformation that is correlated with conditions leading to aggregation in α -synuclein and, ultimately, Parkinson's disease.

Also, during the research work, I found myself in the need of a general-purpose data analysis application for single molecule force spectroscopy data analysis that could solve some common logistic and data analysis problems that are common in this technique. I developed an application that addresses some of these problems, herein presented (Part V), and that aims to be publicly released soon.

1.1 Published works

Both experimental works have been published on high-impact, peer-reviewed international journals:

Fabio Grandi*, Massimo Sandal*, Giovanni Guarguaglini, Emidio Capriotti, Rita Casadio, Bruno Samorí

Hierarchical mechanochemical switches in angiostatin

ChemBioChem 7 (11): 1774-1782 Nov 2006

(* = equally contributed)

Massimo Sandal*, Francesco Valle*, Isabella Tessari, Stefano Mammi, Elisabetta Bergantino, Francesco Musiani, Marco Brucale, Luigi Bubacco, Bruno Samorí

Conformational Equilibria in Monomeric α -Synuclein at the Single-Molecule Level

CHAPTER 1: PREFACE

PLoS Biology Vol. 6, No. 1, e6 Jan 2008

(*=equally contributed)

Also, during the course of the graduate studies, I contributed two reviews containing original theoretical elaboration:

Massimo Sandal, Fabio Grandi , Bruno Samorí

Single molecule force spectroscopy discovers mechanochemical switches in biology: The case of the disulfide bond

Polymer 47 (7): 2571-2579 Mar 22 2006

Francesco Valle, Massimo Sandal, Bruno Samorí

The interplay between chemistry and mechanics in the transduction of a mechanical signal into a biochemical function

Physics of Life Reviews 4 (3):157-188 Sep 01 2007

1.2 Coauthors

Apart from my supervisor Prof. Bruno Samorí, the work described in this thesis has been done in collaboration with (in random order): Francesco Valle, Marco Brucalc, Francesco Musiani, Fabio Grandi, Giovanni Guarguaglini, Fabrizio Benedetti, Emidio Capriotti, Prof. Rita Casadio (University of Bologna), Isabella Tessari, Luigi Bubacco (University of Padua).

CHAPTER 1: PREFACE

Part I

Introduction: The structural diversity of a single protein

Chapter 2

The structural diversity of a single protein

A set is a Many that allows itself to be thought of as a One.

Georg Cantor

2.1 The roots of protein activity

The acquisition of the first protein detailed three-dimensional structure by John Kendrew and coworkers in 1958[1] has been at the same time one of the most important achievements and disappointments of the history of biology. The biochemists, spoiled by the striking geometrical and functional elegance of the DNA structure, were appalled by the irregular, crooked structure of the globular proteins. As Max Perutz asked: *Could the search for ultimate truth really have revealed so hideous and visceral looking an object?*[2].

More than forty years later, the reasons of such a difference are clear. Apart from genetic information storage, that is done mostly by DNA, and topological insulation, that is due mostly to lipid bilayers, nearly every other structural and functional aspect of Earth's life biochemistry is dominated by proteins and supramolecular protein assemblies, and proteins are essential for controlling and synthesizing other biomolecules -including protein themselves. While DNA structure is better (at first approximation) independent from its sequence, since its main objective is to contain information regardless of the information itself, each pro-

tein sequence must correspond to a structure suited to its function. The structural complexity of proteins today is of no surprise to us -on the contrary, molecular biologists today are stricken by the fact that only a handful of secondary and tertiary structure motifs can explain the basic features of most proteins[3], thus partially vindicating the question of Max Perutz.

So, molecular biology has long recognized, *at the level of the proteome*, that proteins are (and, indeed, must be) enormously diverse and flexible things. The potential for this diversity however also recognized at the level of the single protein molecule. Virtually all biological processes that involve motion have their origin in protein dynamics. Dynamics also plays an important role in many proteins of which the primary function is not mobility itself. The ability to change conformation is also essential for the function of many transport proteins, proteins involved in signal transduction, proteins in the immune system, and numerous enzymes¹. In many enzymes, conformational changes serve to enclose the substrate, thereby preventing its release from the protein and ideally positioning it for the protein to perform its function, as in lysozyme. Immunoglobulins are highly flexible in order to be able to deal with a large range of ligands. In general, substrate binding and reversible covalent modifications like phosphorylation all act by modifying protein structure.

Usually these are subtle, local conformational changes. The working structure is therefore assumed to be the so-called *native structure* or slight modifications of it. Most importantly, a large majority of these modifications studied so far share a common property: they are relatively stable, long-lived conformers that can be often isolated and calmly structurally characterized by crystallization or NMR. However, there is no deep reason for thinking that all of these modifications must satisfy our experimental requirements. It is entirely possible therefore that our structural biology suffers a selection effect: our description of molecular biology comes from “native” structures because they are the ones readily available to study.

Molecular motors are a more dramatic example of the multiplicity of molecular structures. Muscle contraction is based on the combined action of actin and myosin. While actin behaves as a more or less static “track” , myosin is a motor protein -that is, an object that exploits chemical energy to perform significant motion. Other examples are the molecular motors kinesin, the F1-ATPase and the

bacterial flagellum. These two latter supramolecular assemblies are, more than motors, *machines* comprehending fully formed rotors and stators. Such objects necessarily involve global conformational changes, with motions of significant amplitude for large parts of a protein.

The role of conformational equilibria is only now ending to be underappreciated. By asserting that a protein assumes basically always the same structure in solution, we neglect the marginal stability¹ and the complexity of the protein folding energy landscape. Such oversimplification cannot explain phenomena like amyloid aggregation, prion strains and intrinsically unfolded proteins, that only in the recent decades have begun to be fully appreciated first in their pathological, and then biophysical, importance. It is entirely possible that the activity of a protein depends more on short-lived and/or poorly-populated conformations, than on the stablest equilibrium conformation. Regulation of the conformational equilibria in solution is another mean that evolution may have used² to control protein activity.

A picture is thus emerging of protein biological activity as a property arising not from a single structure, but from a dynamic *set* of multiple structures. Most of these structures are not easily accessible by mean of crystallography or NMR, but will be functional nonetheless. Elucidation of these structural ensembles is a daunting but essential task that will be only fully resolved by novel structural techniques, especially ones working at the single molecule level.

2.2 The local folding problem

The definition of native structure itself is somewhat confused: it is usually defined to be the conformation of the protein as in the living system. However detailed structures of proteins are rarely, if ever, probed in living systems. It is by no means sure that structures probed in relatively dilute solutions or in highly packed crystals are always equivalent to the ones assumed by the protein in the living cell -yet

¹The free energy $\Delta G_{unfolding} = \Delta G_{denatured} - \Delta G_{native}$ usually stays between 5 and 20 kcal/mol of protein. This is less than 0.1 kT per residue (k = Boltzmann constant and T = temperature)[4][5]

²If biology teaches us something, is that when evolution *may* do something, usually *did it*, no matter if as improbable as authors of Ph.D. thesis.

they are usually assumed to be native.

The physical and chemical environment of molecules in the living cell are in fact extremely different from those in the test tube or in crystals. There are several variables that only now begin to be recognized as essential to dissect the molecular machinery of life. Macromolecular crowding in solution is among the most obvious and long recognized ones, yet its influence has also long been neglected[6]. Recent studies are showing dramatically the complex influence of macromolecular crowding on folding and activity of proteins[7][8][9][10][11][12][13][14]. It must be emphasized that most of these experiments, by necessity, investigate crowding by using “generic” crowding agents like dextran or Ficoll, thus are still somewhat far from the exact characteristics of the immensely complex biomolecular mixture in the cell environment.

Disulfide redox equilibria are another example. While it is commonplace to think of the cytoplasm environment as reducing and of the extracellular one as oxidizing, this is only a safe approximation. In fact, active disulfide redox equilibria exist locally at the external surface of living cells and are known to regulate a wide variety of biological functions. In turn, these redox equilibria can have structural effects by themselves or affect the structural outcome of other physicochemical variables[15].

Perhaps the most striking example is however coming from the recent struggling field of protein mechanochemistry. Being life a thing of motion, it is intrinsically a thing of mechanical forces. Only in the latest years the action of these forces have begun to be systematically investigated. It has then been shown that the function of many proteins is fully appreciated not in a relaxed, but in a mechanically stretched state. Mechanical force partially unfolds proteins *in vivo*[16] and exposes recognition sites[17]. It is clear that in this case local, dynamic conditions induce structures that are simply not possible to explore with conventional crystallographic or spectroscopic methods.

In general, it is evident that, biologically, the folding problem is also a problem of *local* folding. The structure that will be biologically active is the one belonging to the chemical, physical, mechanical microenvironment(s) the protein will be actually working within. Characterization of those microenvironments, of the intertwining of physical and chemical conditions and of how these affect the protein energy landscapes is one of the next essential steps for the molecular understanding

CHAPTER 2: THE STRUCTURAL DIVERSITY OF A SINGLE PROTEIN

of Earth's life.

CHAPTER 2: THE STRUCTURAL DIVERSITY OF A SINGLE PROTEIN

Chapter 3

Far from classic equilibrium: mechanical forces at the protein level

Where force is necessary, one should make use of it boldly, resolutely, and right to the end.

Leon Trotsky

Classical biochemistry looked at the cell as a chemical reactor where diffusions and random collisions drove the reactions of a multitude of chemical species. This simple paradigm was challenged in the last 50 years by the acknowledgement that biochemistry is intimately connected with mechanical forces, that can transport directionally the involved chemical species.

The essential role of mechanical forces in the life cycle of cells is today being extensively recognized. Mechanical forces affect gene expression in vascular cells [18][19], muscle cells[20] and extracellular matrix [21][22][23]. Mechanical stress has also been proved to affect the intracellular nuclear structure[24]. Blood flow causes mechanical stresses that in turn are involved in the building and remodelling of arterial structure[25][26].

As the following sections show, the functional significance of these forces in the metazoan cell is mostly protein responsibility. In particular, many proteins are functionally significant in contexts dominated by mechanical forces. Their

structure-function relationship must therefore be studied as emerging from the set of force-induced conformers, and not by their folding at rest.

3.1 Protein mechanobiology

3.1.1 The mechanochemical crosstalk between extracellular matrix and adherent cells

The physical and chemical scaffold for cell adhesion, growth and differentiation in the Metazoa is the extracellular matrix (ECM). This scaffold is dynamic, being subjected to constant remodeling and action of local forces brought by cells. A close loop is thus formed: the remodeling leads to changes in the elasticity of the ECM substrate itself and shifts the application of mechanical tension on the adherent cells[27].

Intracellularly, actin microfilaments give tension on the cell surface, that is then transmitted to their surface membrane and then on the ECM[28][29]. The transducers of these mechanical signals are transmembrane receptors, like integrins, usually mediators of cell adhesion[30].

By measuring the forces exerted by a single focal adhesion, it has been possible to estimate the average stress density exerted by cells as being about $5 \text{ nN}/\mu\text{m}^2$. In these adhesion sites, each molecule is estimated to resist a force of no less than 1 pN [31][32]. Lower stress density values ($0.6 \text{ nN}/\mu\text{m}^2$) for endothelial cells have been reported[33]. The mechanical cell behaviour is however cooperative: a larger stress for single cell is measured for cells belonging to a subconfluent epithelium than for an isolated cell[33]. The adhesion complexes, mediated mainly by actin and integrin, are both mechanical and biochemical links between the ECM and the cytoskeleton. As such, they function as mechanosensors that react to changes in elasticity of their surroundings and to the intervening network of forces. Mechanochemical models for cell regulation are thus beginning to be proposed[27][30][34].

3.1.2 The transduction of mechanical signals by protein chains

The transduction of mechanical signals involves different molecules in both the cell and the ECM. Actin filaments and microtubules can be reshaped by the tension on the cytoskeleton. In turn, this can indirectly stretch proteins connected to it. ECM molecules might experience forces directly applied by adherent cells or due to the relative movements of the ECM components[28]. These forces may deform ECM proteins and therefore alter their biological functionality[35]. Other candidates for mechanochemical transduction can also be found within transmembrane elements. Integrins can deform under mechanical loading and change their ability to form focal adhesion complexes with the cytoskeleton[36]. Stretching forces applied on ion channels can alter the ion transport through the cell membrane[37]. All of the above considerations lead to the idea that a key role in the cells mechanochemical transduction process is played by the deformation of biomolecules in response to mechanical stress.

Recent experimental and computational data suggest that mechanical forces regulate the functional states of some proteins by stretching them into metastable intermediate conformations. Three different means has been evidenced[35][38], by which a stretching force can alter a multimodular protein conformation and therefore its functional state: exposing cryptic sites, deforming the conformation of a binding site, changing the relative distance of synergic binding sites:

Exposure of cryptic sites Force applied on a protein module can stretch it until a binding site, usually buried in the fold, becomes exposed and able to bind a receptor. The effect of the force is therefore to turn on a biochemical switch.

Deforming the conformation of a binding site Mechanical stress can modify the geometry of a binding site as to make binding favoured (or unfavoured).

Shifting synergic binding sites. When two binding sites that bind the same receptor or that need to be bound in concert are present, their position can be shifted by the action of mechanical force, thus altering synergic effects.

CHAPTER 3: FAR FROM CLASSIC EQUILIBRIUM: MECHANICAL FORCES AT THE PROTEIN LEVEL

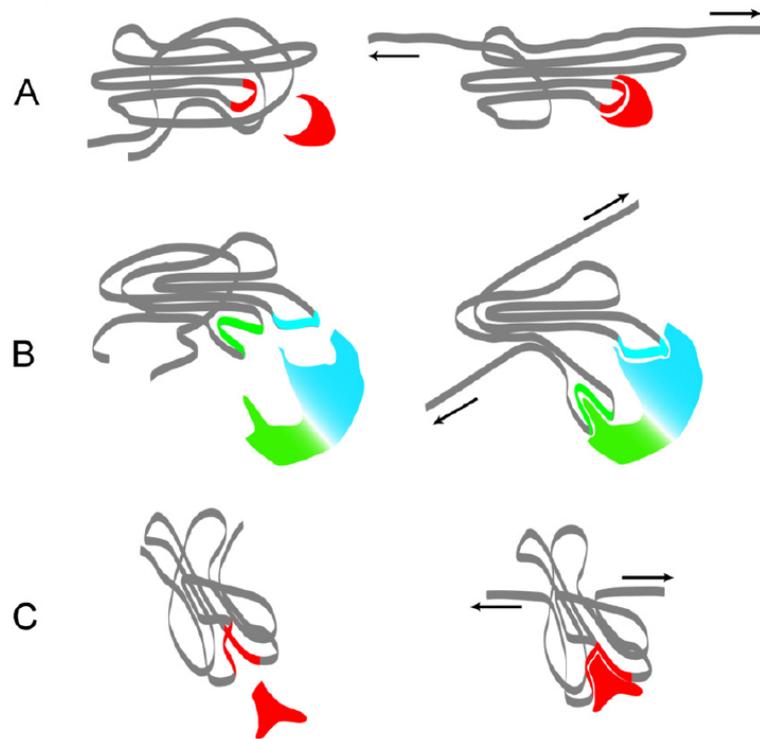


Figure 3.1: The conversion of a mechanical tensions into biochemical signals by force induced conformational transitions. A) By exposure of cryptic sites: a mechanical force can stretch a protein into a conformation in which a binding site, previously buried in the folded native structure (left), is exposed and thus enabled to bind a receptor (right). B) By changing the distance of binding sites: two sites that are not at a correct distance to bind cooperatively to a receptor can be brought by a force to a relative distance that makes their concerted binding possible. C) By changing the shape of the binding site: the force can modify the geometry of a potential binding site in order to make its binding to a ligand possible.

3.1.3 Examples of mechanical regulation of proteins

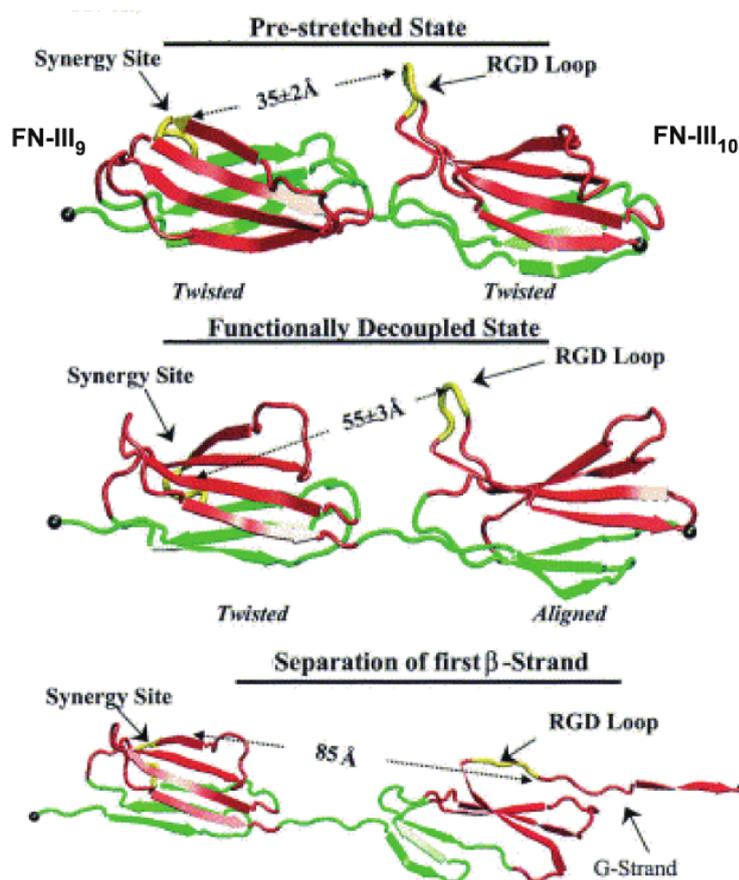


Figure 3.2: Fibronectin function is mechanically regulated in modules FNIII9-FNIII10. The pre-stretched state exposes the RGD-loop at a distance of 35 \AA from the synergy site. A partially stretched intermediate state (“functionally decoupled” state) is reached, where the distance with the FNIII9 site is enlarged and the synergistic effect is abated. Further stretching of the structure leads to the rupture of a β -sheet and to deformation and reduced exposure of the RGD loop, thus abolishing integrin binding[39]

Fibronectin (Fn), a cell adhesion protein abundant in many mammalian tissues, is the classical example of protein mechanical regulation. The sequence contains more than 50 characteristic β -sheet modular repeats. It is thought that fibronectin is one of the key players in the initial encoding of mechanical forces into biochemical signals[35], by binding integrins at focal adhesions. Fibronectin has the ability to

physiologically form supramolecular fibrils, whose formation is induced by the mechanical deformation of FNIII modules into a distorted conformation. Crucial sites required for fibrillogenesis, inaccessible in the native folded state[40], are thus exposed.

Also, mechanical forces can modulate the Fn-integrin binding affinity[35][41]. In particular, the RGD integrin-binding sequence of FNIII10 modules must have a distance of at least 11 Å from fibronectin surface to achieve optimal binding. In the native state the presence of a β -sheet holds RGD in the correct conformation. This β -sheet breaks under a mechanical force, thus reducing the exposure of the RGD loop (Figure 3.2). The RGD loop binding affinity is also regulated by a synergic site in the neighbour domain FNIII9, and this is also subject to mechanical regulation. Mechanical stretching of Fn in fact increases the distance from the 32 Å of the native structure (that allows them to work cooperatively) to 55 Å (Figure 3.2)[42].

Force can expose cryptic sites in proteins that unleash enzymatic activity. Fibronectin, again, is an example (see Figure 3.3). The FnI12 module has a partially cryptic protein disulfide isomerase activity[43]. The collagen-binding domain has cryptic metalloprotease activity[44]. Fibronectinase, a proteinase that specifically acts on actin and myosin, is cryptically hidden in the N-terminal heparin/fibrin binding domain. Fibronectinase seems to play a role in muscular dystrophy pathogenesis[45].

3.2 Sacrificial bonds and protein adhesion

Some mechanical proteins, like titin or a variety of cell surface receptors, can have evolved to function as “shock absorbers” . This peculiar mechanical role was first recognized by Paul Hansma et al. [46]. Their attention was focused on the incredible strength and toughness of certain natural materials [46] [47] [48] [49], such as spider dragline silk, whose breakage energy per unit weight is two orders of magnitude greater than high tensile steel[50] [51], or abalone shell, with its very high fracture resistance, or the mysterious molecular origin of the toughness of the bones[52]. When one stretches away the two ends of this type of protein before the breaking point of the polypeptide chain is reached, a domain unfolds, and the energy stored in the protein is dissipated as heat. Then, the constantly applied force puts the protein again under an increasing stress, until the next domain

breaks and so on. These domains were termed “sacrificial” because they are actually sacrificed in order to release the force before it can break the polypeptide chain. Only when no folded sacrificial domains are left, only after that point, will the chain finally break under the external force. By the same principle, in the case of muscles, multi-modular proteins not only transmit stresses but also protect the muscle structure from damage whenever extreme forces are acting on it. In particular, the unfolding of the proximal Ig domains of titin may serve as a buffer to protect cardiac sarcomeres from being damaged at forces exceeding the physiological range[53] (see Figure 3.4).

3.3 Force and intermolecular interactions

Also, intermolecular interactions can be regulated by force. The behavior of a mechanically activated switch must therefore be described in terms of the relation between force, lifetime and affinity, i.e. in terms of the dynamics of formation and dissociation of the interactions upon which it relies. This dynamics is controlled by the mechanical deformation of the energy landscape of the same interactions. Correspondingly, energy landscapes found in different natural processes that involve mechanical forces have been optimized by evolution. In general, energy barriers characterized by short barrier distance x_b can sustain high forces but small deformations and thus display a “brittle” character; conversely, barriers possessing large x_b are “elastic” or compliant, being capable of sustaining low forces but larger deformations[54]. Such differences have been found in the energy landscapes that drive the control of the rolling adhesion of leukocytes in the blood vessels *vs* the shape of the energy landscape of the interactions mediated by fibronectin that sustain the infection of staphylococci to the endothelial cells lining the vessels. These differences are in agreement with the macroscopic behaviour and biological meaning of these systems.

The physics underlying the dissociation of molecular interactions (intra- or intermolecular) by mechanical force, and the AFM force spectroscopy technique best suited to understand it, are discussed in 5 and 6.

CHAPTER 3: FAR FROM CLASSIC EQUILIBRIUM: MECHANICAL FORCES AT THE PROTEIN LEVEL

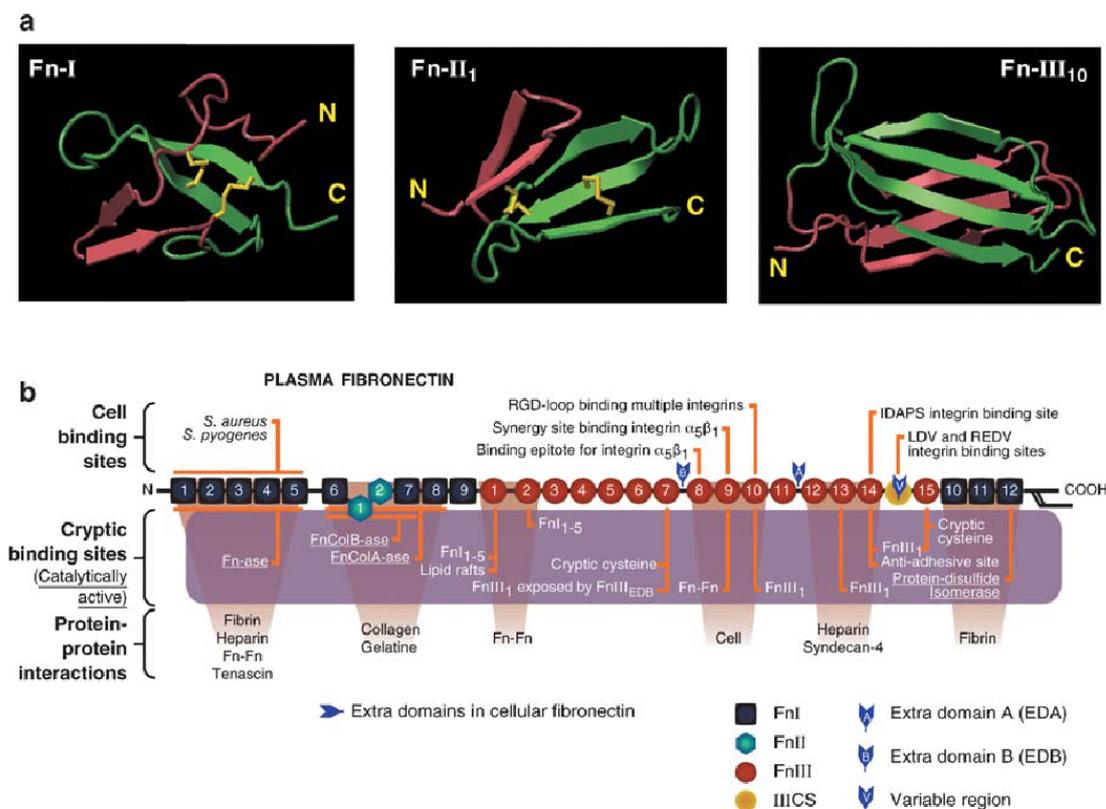


Figure 3.3: Diagram showing molecular recognition and cryptic sites of fibronectin. Upper panels: Fibronectins are dimeric molecules composed of more than 50 repeats of type FnI, FnII, and FnIII, whereby cellular fibronectin may contain additional alternatively spliced modules as indicated. All the repeats are composed of β -sheet motifs, and representative crystal structures are given in the upper panel for FnI, FnIII₁, and FnIII₁₀. Fibronectins contain a large number of molecular recognition and cryptic sites (lower panel), including the cell binding site RGD; the synergy site PHSRN, which is recognized by $\alpha_5\beta_1$ integrins; and the sequence IDAPS in the HepII region of fibronectin that supports $\alpha_4\beta_1$ -dependent cell adhesion. The IDAPS motif implicated in integrin $\alpha_4\beta_1$ binding is at the FnIII₁₃₋₁₄ junction. The cryptic sites include various Fn self-assembly sites whose exposure is needed to induce fibronectin fibrillogenesis, a cryptic fragment from FnIII₁ that localizes to lipid rafts and stimulates cell growth and contractility, and a binding site for tenascin. Other cryptic sites with enzymatic activity include FnCol-ase, a metalloprotease in the collagen binding domain of plasma fibronectin capable of digesting gelatin, helical type II and type IV collagen, α - and β -casein, and insulin β -chain; as well as a proteinase (Fn-ase) specific to fibronectin, actin, and myosin; and a disulfide isomerase. Figure and caption from [17]

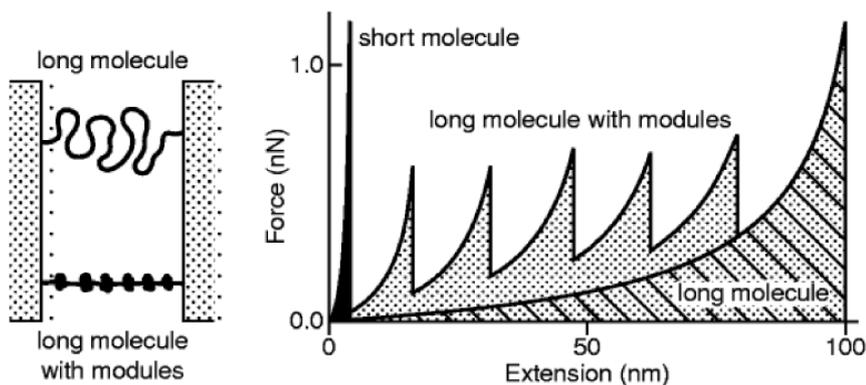


Figure 3.4: Biopolymer resistance to an applied traction force. The figure shows the force versus extension curve for three different kind of molecules: a short one, a long one, and a multi-domain one. The short molecule resists pulling up to a high force before it breaks at small extensions. The energy required to break this short molecule (area under the curve) is small. A long molecule behaves like an entropic spring and yields to the pulling force up to large extensions. The energy to break the long molecule is larger than that for the small molecule, but the forces at low extensions are small. Finally, the long molecule compacted into domains resists pulling already at small extensions. Before the molecule's backbone can break at high forces, modules unfold at intermediate forces. Repeated unfolding of molecules allows stretching up to large extensions and requires significant energy. From [46]

CHAPTER 3: FAR FROM CLASSIC EQUILIBRIUM: MECHANICAL FORCES AT THE
PROTEIN LEVEL

Chapter 4

Hidden in a conformational equilibrium: intrinsically unstructured proteins and amyloids

Many patterns of Nature are so irregular and fragmented [...] The existence of these patterns challenges us to study these forms that Euclid leaves aside as being “formless” to investigate the morphology of the “amorphous” .

Benoît Mandelbrot

4.1 Intrinsically unstructured proteins

The classical structure-function paradigm of protein biology has been recently been jeopardized by the discovery of intrinsically unstructured proteins (IUPs). It is not often acknowledged that unstructured segments > 50 residues are common in functional proteins[55], to the point that almost 30% of the eukaryotic proteome could encode for partially or completely disordered proteins. Far from being “molecular freaks” , disordered proteins evolved to be and function this way, as shown by their prevalence and evolutionary conservation. The genome fraction encoding IUPs increases with the complexity of organisms[55][56][57].

In fact, a whole “new view” of protein structure-function relationship is emerg-

ing, where conformational diversity, far from being a unusual quirk of a few peculiar proteins, is intimately connected with protein function[58]. Diversity is in fact intrinsic in the energy landscape defining protein structure, that is far from being a smooth funnel, and it is instead a rugged landscape hosting an ensemble of conformers at similar free energy. Also, conformational diversity is a significant propeller of protein evolution. Structural diversity is likely related with functional diversity. Evolution can therefore act on different functions in a single protein chain, and selection may shape the sequence as to “freeze” one function upon the other. The genesis of the first stable, folded proteins from random polypeptide sequences could have been the first example of this kind of evolutionary pathway.

It must be noticed that the practically equivalent expressions found in literature like “intrinsically unstructured protein” , “intrinsically disordered protein” or “natively unfolded protein” (often abbreviated as IUP, IDP and NUP) can be somewhat misleading, since in most cases they define the behaviour of the protein in dilute solution, in absence of molecular partners. Moreover these proteins are not simple random coils: they exist as dynamic, highly flexible structural ensembles that undergo conformational conversions on a wide range of timescales.

4.1.1 Discovery of intrinsically unstructured proteins

IUPs are not a novel discovery. Polypeptide hormones, known from 40 years ago, are IUPs often[56]. Early proton NMR experiments detected IUPs in living cells[57]. However only in the last decade transcriptional regulation, translation and cellular signal transduction[59][60][61] have faced the necessity to recognize the pervasive presence of IUPs.

It must be emphasized that classic biochemical methods show a strong bias that encourages the production and characterization of active *and* folded proteins. Standard protein preparation method require cells to be isolated and homogenized. Unfolded proteins suffer degradation by endogenous proteases under these conditions, and thus the methods select automatically only folded proteins able to survive the homogenization process. Also, unfolded domains are often part of regulatory proteins produced in only a few copies per cell, thus being hard to detect and assay.

The recent shift of molecular biology to a genome-based approach, where we

now have access to a vast library of gene sequences that correlate with known functions and pathologies, helped isolation and characterization of IUPs by isolating function, using mutants and knockouts. The desired protein, in this case, is already chemically known and expressed and purified bottom-up, and it can safely be characterized.

IUPs are also hard to characterize with common structural biology techniques. X-ray crystallography cannot provide information on disordered regions but only evidence their existence as zones lacking electron density. Recently, advances in NMR sensitivity and resolution made them able to characterize, if only by averaging on the bulk sample, structural propensities and dynamics of sizeable disordered proteins. Finally, single molecule techniques, being able to probe molecular structure and dynamics without averaging between different molecules, are now beginning to be employed to map the structural complexity of disordered sequences (see Section 4.5).

4.1.2 Intrinsic unfolding and function

There are different hypothesis on the functional rationale for protein disorder. The so-called “moonlighting” -*i.e.* the ability of a protein to bind vastly different ligands and perform different functions- could be helped by the structural flexibility of IUPs. IUPs often display larger-than-average intermolecular interfaces. In fact, a folded monomeric protein would need to be two or threefold larger to achieve the same intermolecular interface of a IUP. IUPs thus could have evolved to allow large intermolecular interfaces to without increasing protein size (and, consequently, cell size or crowding).[62].

Unstructured sequences are in fact often molecular recognition domains, regulatory partners of other proteins, but also of nucleic acids or membranes¹. For example, IUPs interact with DNA in processes like transcription, transposition, DNA packaging, DNA repair and genome replication, and are extensively involved in transcription regulation[63][64]. As expected in such a context, IUP-encoding genes in the human genome have been found to be linked with cancer[65]. Cell signaling, cell cycle control, endocytosis, scaffolding and development are other

¹Enzymatic function, that requires finely tuned recognition sites, seems instead harder to achieve in an unfolded context

processes where IUPs play a fundamental role. In addition, many disordered regions are present in proteins regulated by phosphorylation[64].

Usually, proteins that fail to unfold face unavoidable destruction in the cell by the ubiquitin-proteasome pathway, that recognize the presence of solvated hydrophobic patches. How can IUPs escape proteasomal degradation in the cells? The simplest explanation is that IUPs are not unstructured in their functional state: they are at least partially folded in being tightly bound with their partner. On the other hand, IUPs disorder could be in fact selected for being subject to fast turnover, creating a dynamic, fast layer of regulation[66]. Transcriptional activation could be linked with the degradation of disordered domains by the ubiquitin-proteasome pathway[67][68][69].

Sequences linking modules in multidomain protein are a peculiar class of intrinsically unstructured sequences. These sequences must combine flexibility and extension with stability in face of degradation and proteolysis. Linker sequences lack often hydrophobic residues, probably enhancing their stability. In fact, disordered polyglutamine repeats resist degradation by eukaryotic proteasomes[70]. This has pathological significance, since in polyglutamine-repeat disorders these sequences tend to accumulate and form pathological aggregates (see below). In general, IUPs display below-average hydrophobicity and a charged surface, that can hinder their recognition by the protein degradation systems.

4.1.3 Prediction of intrinsic disorder

Intrinsically unstructured proteins or segments sequences are characteristic in displaying low average hydrophobicity, high net charge, and poor sequence complexity, with frequent repeats[71][72]. IUPs identified *in vitro* have a distinct amino acid composition, enriched in A, R, G, Q, S, P, E and K while depleted in W, C, F, I, Y, V, L and N. This difference is stronger enough to make them successfully predictable by *a priori*, untrained algorithms like IUPred[73] that estimate the total pairwise interresidue interaction energy of sequences, which is significantly smaller for IUPs than for globular sequences. Other predictions, usually neural-network based, have been implemented[74][75]. There is computational evidence of more than one class of sequences that can lead to intrinsic disorder. Protein predictors trained on one type of protein perform poorly when applied to proteins of a

different class. Three “flavors” of protein disorder are today acknowledged and computationally recognized, distinguished by amino acid composition, sequence location and functional significance[75].

4.1.4 The structural states of IUPs

It is possible to describe protein structures as a continuum that starts from tightly folded globules, to mostly folded proteins that include flexible or disordered regions, to flexible but still compact molten globules to, finally, extended and heterogeneous random coils (see Figure 4.1). The concept of *protein trinity* has been introduced to rationalize this continuum into three main classes (ordered, molten globule and random coil[55]). However,. it is most likely that proteins rarely, if ever, can be described as true random coils, at least in non heavily denaturing media. IUPs, even when highly extended, still display residual propensity to form secondary structure elements and/or clusters of hydrophobic residues[76]. This residual structure could be of fundamental importance in modulating intermolecular interactions. The existence of residual structural elements could pivot first-contact points, which then drive the interaction of the remaining unstructured sequences with the ligand[77].

Many intrinsically disordered proteins tend to acquire order or even fold into stable secondary structures on binding to their targets.[78][79] (Figure 4.1. This disorder-to-order transition has an entropic cost. The key thermodynamic driving force for the binding reaction is consequently a favourable enthalpic contribution. This binding-coupling combination is highly specific but of low affinity. An example of a system where this combination is favourable are signal-transduction proteins, where the signalling process is initiated by specific binding and ended by dissociation.

The pervasive binding-folding association lead to the proposal that IUPs are, in physiological state, no less folded than other proteins, because they always come bound to a partner[80]. They evolve therefore just like other folded proteins, their “intrinsic” disorder being a hidden property that arises only when proteins are purified *in vitro*. There is however also evidence that their function is intimately linked with their lack of a single, compact fold *in vivo*. Unstructured proteins can be predicted and are distinguished, in sequence properties, from ordered proteins

CHAPTER 4: HIDDEN IN A CONFORMATIONAL EQUILIBRIUM: INTRINSICALLY UNSTRUCTURED PROTEINS AND AMYLOIDS

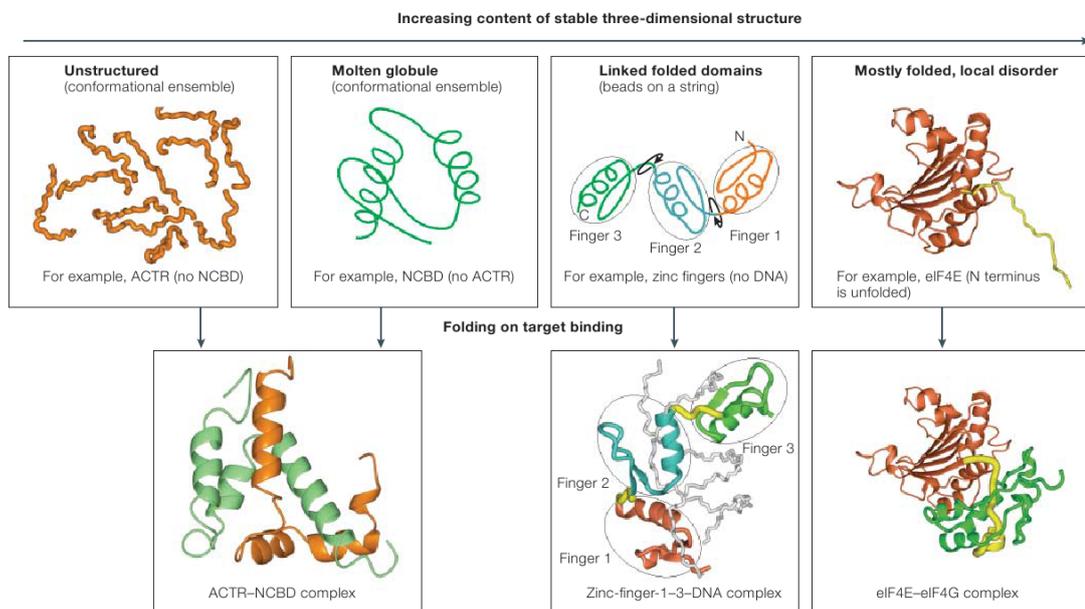


Figure 4.1: Top panels: examples of disordered protein structural states in increasing order, from a loose conformational ensemble to a mostly folded structure with only local disorder. These structures have been characterized to fold upon binding of their molecular partners, as shown in the bottom panels. From [61]

(see Section 4.1.3), *i.e.* they are not a pure *in vitro* artefact. However logically the fact alone that the *in vitro* properties of IUPs can be reliably predicted and rely on factual sequence differences, does not imply that this applies to the *in vivo* protein.

Can simple molecular crowding induce IUP folding? Experimental evidence shows only a limited tendency of simple crowding to induction of IUP fold[81][82][83]. Even while the intracellular environment could be crowded enough to induce IUPs to fold, extracellular IUPs (like milk caseins, salivary glycoproteins and some bacterial fibronectin-binding proteins) should be subject to less crowding, therefore at least the conformational state of those proteins is physiologically unfolded[84].

In general, the energy landscape of IUPs is even more labile than that of other proteins. Subtle microenvironmental conditions, like the presence of a ligand with given surface features and charge, or differences in pH and ionic strength, can allow funnels to move, deepen or flatten in the energy landscape, driving the IUPs towards different structural states and/or differently populated sets of structures.

IUP function, therefore, depends on the full complexity of their conformational equilibria. Such an equilibrium cannot be fully characterized by averaged, bulk techniques. In the near future, single molecule techniques will most probably be essential to complete the picture of IUP structure(s)-function(s) relationships (see Section 4.5).

4.2 Amyloid fibrils

Proteopathies are a class of pathologies whose aetiology is related to the *misfolding* of a protein. Misfolding can lead to pathology both by loss of function (diminishing the quantity of functional protein) or by gain of function (the misfolded protein is toxic).

A large subset of proteopathies share the common feature of dysfunctional, highly regular, fibrillar aggregates called amyloids² that originate from proteins that are natively soluble and functional. Examples of these conditions are neurodegenerative syndromes like Creutzfeld-Jakob, Alzheimer's or Parkinson's diseases, but also type II diabetes or cataract[85]. This correlation to such a wide range of socially relevant and often incurable pathologies led to an enormous biomedical and biophysical interest in this aggregation process. Evidence is now emerging that amyloid aggregation is in fact one of the fundamental structural states of polypeptide chains that arises from very basic physical principles[86]. Amyloid proteins, in fact, do not share any obvious sequence or structural similarity, and are very different in secondary structure composition. Synthetic polymers may form supramolecular assemblies similar to amyloids, thus suggesting that the aggregation of proteins into amyloid fibrils and that of classical polymers share a common ground[85]. A large variety of proteins, in fact, is able of amyloid conversion *in vitro* particular conditions, regardless of their ability to do that *in vivo*[87].

4.2.1 Amyloid fibril structure

Amyloid fibrils are supramolecular structures whose main components are one-dimensional, crystalline, twisted rods made of protein monomers stacked one on another, called *protofilaments*. Protofilaments are about 2-5 nm in diameter[88]

²The term comes from a superficial resemblance of these fibrils with starch crystals.

and are usually rich in β -sheet structure. The β strands are usually perpendicular to the fibril axis[89]. In turn, the protofilaments assemble as fibrils by twisting in a helix fashion into rope-like structures about 7-13 nm wide[88][89] or by joining laterally in ribbon-like structures that can be up to 30 nm wide[90][91][92].

Classical techniques to characterize amyloid fibrils were microscopies such as TEM or AFM, and X-ray fiber diffraction[88][89][93]. Amyloid fibrils were however poorly characterized from the molecular point of view in being too large and insoluble to be amenable for solution NMR. Recently these limitations have been partially overcome thanks to the application of solid-state NMR (SSNMR) spectroscopy[94][95][96] and the characterization of peptidic fragments that model essential structural features of amyloid fibrils and can be studied by mean of X-ray crystallography[97][98]. Solid state NMR permitted to describe the structure of A β 1-40 amyloid fibrils [94][99][100]. The monomer is shaped as a pair of β -strands, comprising approximately residues 12-24 and 30-40 respectively (Figure 4.2). The A β monomers stack in a parallel fashion[99][100].

Electron microscopy was in turn able to characterize the diameter and the mass per unit length of protofilaments[94]. Combining these observations with the SSNMR data and EPR[101], a model for the protofilament as built from a unit of four β -sheets, 10 Å apart, has been built (Figure 4.2 A). The same pattern was observed in crystals of peptides derived from the yeast prion Sup35p, the crystal structure displays pairs of parallel β -sheets where each peptide constitutes each β -strand (see Figure 4.2C).

In general, amyloid fibrils share a cross- β structure featuring polar or hydrophobic repeats that interact along the fibril axis. The core of the fibril is often built from two to four closely interacting sheets. Notably, amyloid sheets are much less twisted (as observed by cryo-EM and FTIR spectroscopy)[102][103] than what should be statistically expected from the collection of analogous β -sheets in globular protein structures. These remarkable structural constants, regardless of the vast diversity in sequence of amyloidogenic proteins, suggests that amyloid formation is an intrinsic propensity of protein chains.

CHAPTER 4: HIDDEN IN A CONFORMATIONAL EQUILIBRIUM: INTRINSICALLY UNSTRUCTURED PROTEINS AND AMYLOIDS

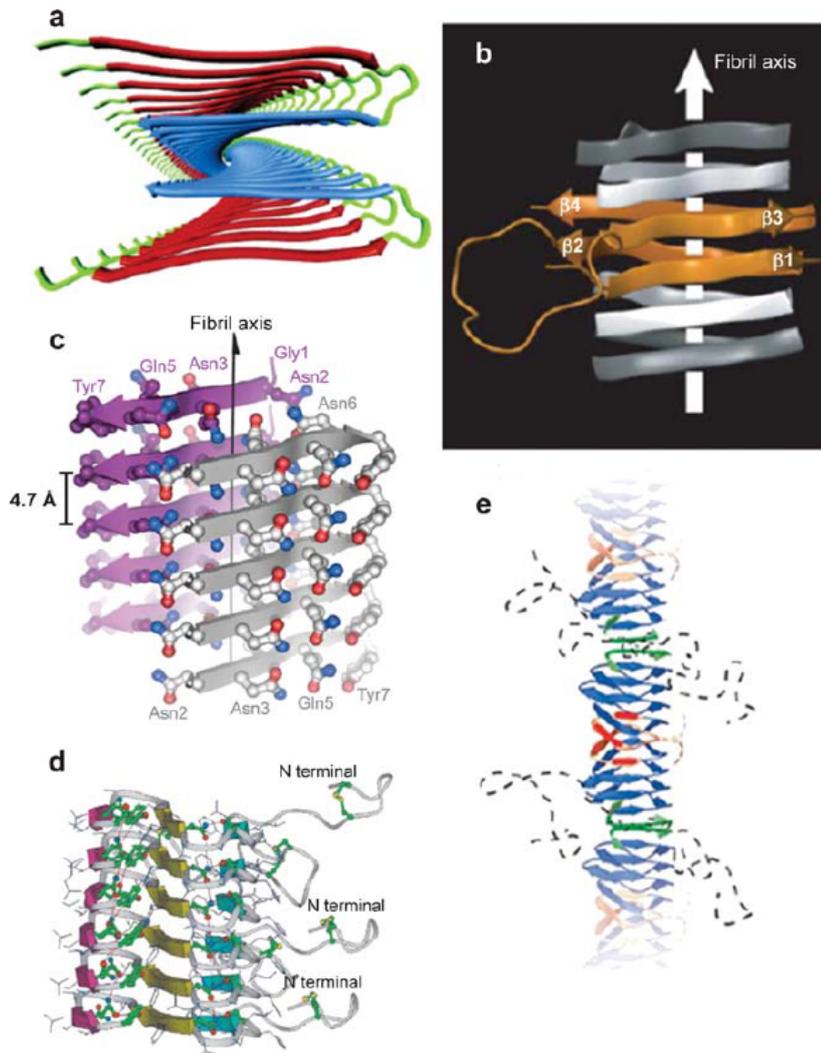


Figure 4.2: 3D models of fibrillar amyloid aggregates. A) The protofilament of the $A\beta$ peptide as viewed along the axis of the fibril. B) The fibril from the C-terminal domain of the fungal protein HET-s. Notice in the central molecule the long loop connecting the four β -strand s. C) Atomic structure of the microcrystals assembled from the GNNQQNY peptide, where each peptide forms one β -strand . D)The protofilament from amylin, comprehending an unstructured N-terminal tail. E)The fibril from the NM region of Sup35p. Caption and figure adapted from [61]

4.2.2 Conformational similarities and diversities of amyloids

Despite the structural constants of the amyloid fibril framework, the sequence details determine a number of basic variables like the length and topology (parallel or antiparallel) of the β -strands, the number of β -sheets themselves in the protofibril and the structure of loop regions not included in the core. These features differ from protein to protein, and even in fibrils coming from the same protein. Interactions between side chains can yield also more profound modifications of the core structure. In fibrils made of poly-Q, β -helices are the prevailing motifs [104].

Even before detailed structural studies, there was striking evidence that the same protein could produce conformationally different fibrils [102][90][105][106]. It is known that templating effects take place in these systems, so that the initial seed dictates (kinetically traps) the structure of the resulting fibril, regardless of the preferred equilibrium structure of the fibril in the given sequence or physico-chemical conditions[94][92][107][108].

This pattern contrast with the usual behaviour of folded, globular proteins, that have been selected by evolution to fold into one specific three-dimensional structure. In the usual cases, complex free-energy landscapes associated with their sequences are usually assumed a single and well-defined minimum, under physiological conditions, corresponding to the native state. Instead, this kind of behaviour matches closely that of IUPs, whose sequence defines a wide variety of local energy minima. These minima will be populated by the protein ensemble in a dynamical equilibrium that can be easily shaped by the local environment. At a first glance amyloids, that are extremely stable, semicrystalline phases, could look extremely remote from the fluidity of IUPs; however these considerations make it is not surprising that a significant proportion of amyloid-forming proteins are IUPs[85].

4.3 Mechanism of amyloid fibril formation

Building a description of the aggregation pathway(s) of a protein is a huge task that is just beginning to be faced. The large diversity of structural phases and early oligomeric and polymeric aggregates has to be deciphered, along with the

thermodynamical and kinetical details of the conformational changes into play at molecular detail.

In this context, significant attention is placed on protofibrils (*i.e.* the aggregate structures preceding the formation of fibrils) because of the increasingly apparent involvement of those structures in the pathogenesis of amyloid diseases.

4.3.1 General features

The aggregation kinetics usually consists of an initial, “lag” phase, during which little or no growth happens, that is due to the time required for the first aggregation nuclei to constitute. After this phase, which can last days or even weeks, rapid exponential growth ensues[109][110][111]. The slow phase is thus nucleus formation: after that, fibrils grow fast by association of monomers or other oligomers. Aggregation of protein into amyloid fibrils is therefore thought to ensue from a nucleated growth mechanism.

In analogy with other nucleated growth systems (for example crystal growth), “seeding” of the solution with ready made fibrils avoids the need for nucleation and consequently can shorten, to the point of eliminate, the lag phase[110][111]. During the lag phase a variety of oligomeric aggregates form, among them the nuclei for the formation of mature fibrils.

Seeding a protein solution with preformed aggregates becomes more and more less efficient with decreasing sequence homology. [107][112], until becoming nil below 30% to 40% sequence identity[112]. Notably, the sequence similarity between neighbouring domain in multimodular proteins is almost always less than 40%. This pattern could have evolved to avoid self-aggregation due to templating effects of sequence identity[112].

4.3.2 Oligomers and protofibrils

The oligomeric species that predate the formation of full-formed amyloid fibrils have been the subject of much attention recently. Their importance is twofold: first, they are the crucial link between monomers and amyloid fibrils, and their comprehension is needed for aggregation aggregation; second, they are ever more implicated as being the real active forms responsible of the proteopathies, the amyloid fibrils looking more of an inert dead-end (see for example [114]). Protofibrils

CHAPTER 4: HIDDEN IN A CONFORMATIONAL EQUILIBRIUM: INTRINSICALLY UNSTRUCTURED PROTEINS AND AMYLOIDS

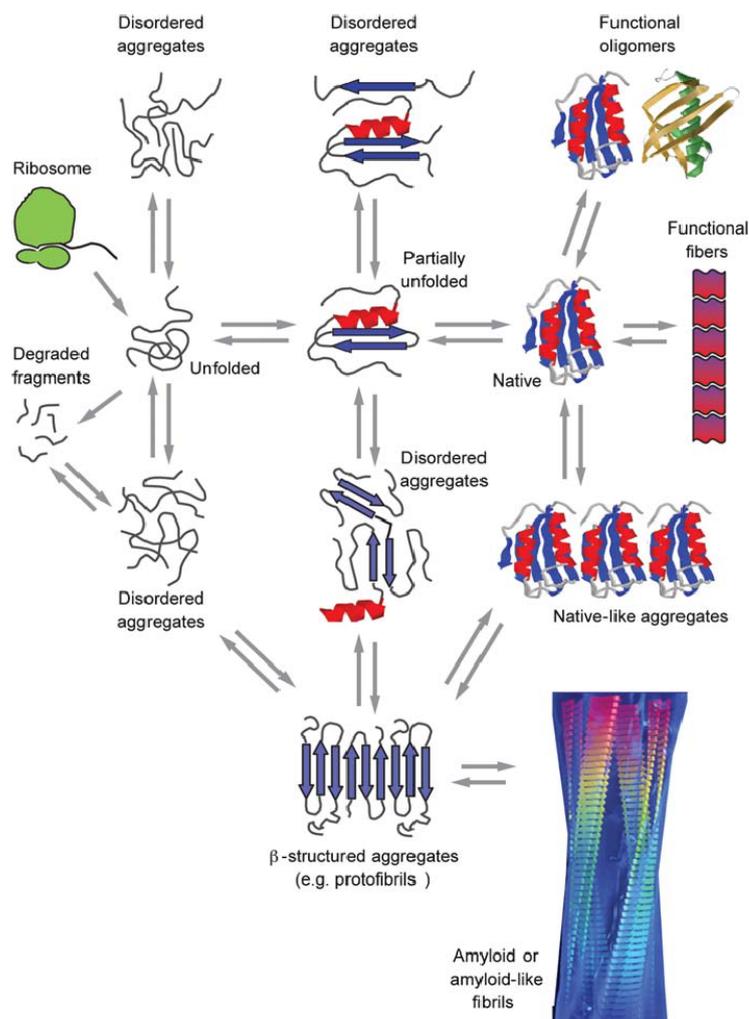


Figure 4.3: A schematic representation of some of the many conformational states that can be adopted by polypeptide chains and of the means by which they can be interconverted. The transition from β -structured aggregates to amyloid fibrils can occur by addition of either monomers or protofibrils (depending on protein) to preformed β -aggregates. All of these different conformational states and their interconversions are carefully regulated in the biological environment, much as enzymes regulate all the chemistry in cells, by using machinery such as molecular chaperones, degradatory systems, and quality control processes. Many of the various states of proteins are utilized functionally by biology, including unfolded proteins and amyloid fibrils, but conformational diseases will occur when such regulatory systems fail, just as metabolic diseases occur when the regulation of chemical processes becomes impaired. Caption and image from [85]

CHAPTER 4: HIDDEN IN A CONFORMATIONAL EQUILIBRIUM: INTRINSICALLY UNSTRUCTURED PROTEINS AND AMYLOIDS

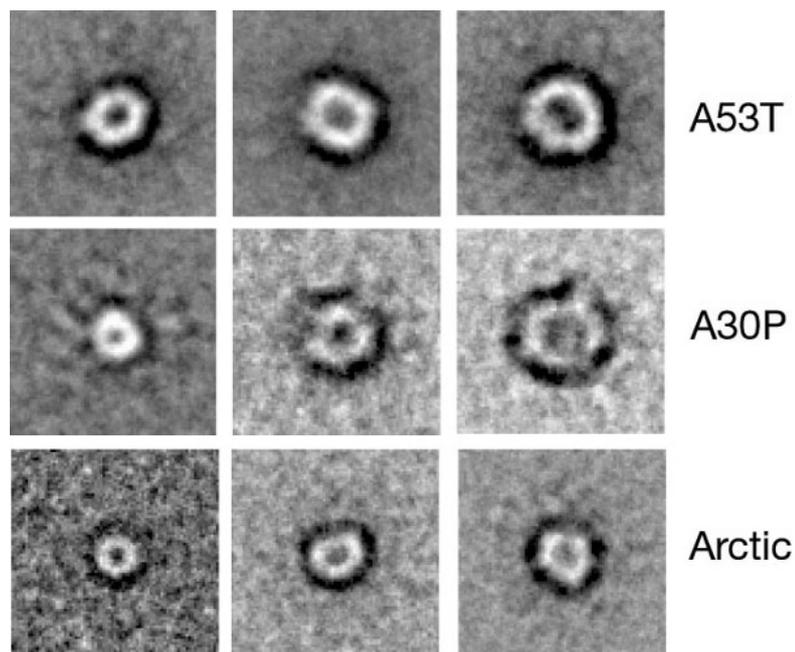


Figure 4.4: Examples of annular protofibrillar aggregates. The images represent the projection averages of pore-like protofibrils formed by α -synuclein mutants A53T (top) and A30P (middle) and by the $A\beta$ amyloid protein (bottom). Each square is 30.5x30.5 nm. These pores resemble bacterial pore toxins, and it is possible that permeabilization of the cell membrane by mean of pore-like protofibrils has a role in amyloid proteopathies pathogenesis[113]. From [113]

can be both on-pathway to fibrils[93][111] or off-pathway[115][116]

In many cases, protofibrils do not need to share the β -sheet structure of the mature fibrils. Oligomers containing flexible structure can form from the NM sequence of Sup35p or in denatured yeast phosphoglycerate kinase. These oligomers in turn are precursors of β -sheet rich forms, which act as the true seeds for amyloid formation[111][117]. There is a direct correlation between the onset of β -sheet structure and oligomer size. The association of oligomers into short, curved protofibrils (similar to those obtained by A β and α -synuclein) happens in turn only when the oligomers reach a certain size.

On the other hand, oligomeric and protofibrillar species containing extensive β -sheet structure are known. Often these species show already significant structural regularity, marking them as true precursors of amyloid fibril formation. The A β peptide aggregation involves the early appearance of a variety of nonfibrillar species, visible microscopically[93][118]. The general “unit” seems to be a spherical aggregate, of about 20 monomers, with large β -sheet content[118][119], that in turn assembles into chains that can be straight, curved, or even circular. Many other amyloid-forming proteins can form similar structures, including α -synuclein[120], poly-glutamine repeats[119] and β 2-microglobulin[115].

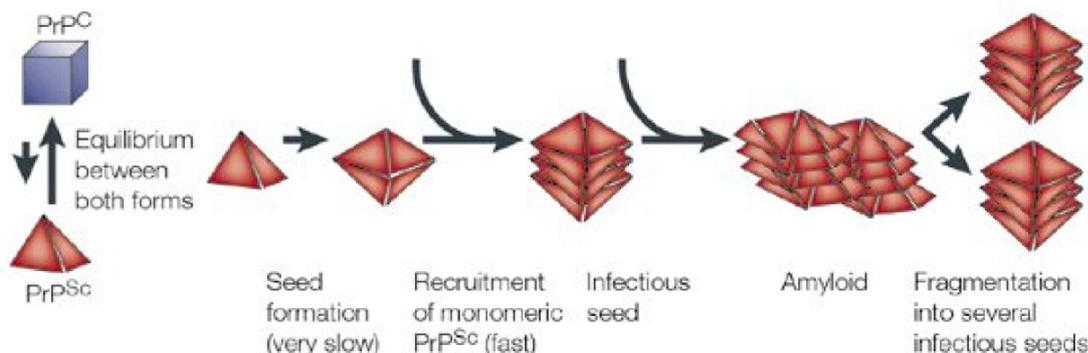
There is evidence that also the oligomeric protofibrils share a number of structural features, regardless of the protein originating them. It had been possible to create specific antibodies that bind to oligomeric forms of *different* proteins, but do not bind to monomers or fibrils of the same proteins[121].

4.3.3 The seeded and conformational models

The structure and growth of already formed amyloid fibrils is relatively well understood, and the problem of protofibrils is currently being thoroughly investigated. A fundamental and still poorly resolved question arises when confronting the problem of how amyloid aggregation first ensues -that is, what are the steps that lead from the protein monomer to the first aggregated states? Resolving this step is of not only theoretical but also practical importance, since blocking the very first steps of aggregation could be the most important aim for prevention and therapy of amyloid-related proteopathies.

Two models can be hypothesized and have found experimental evidence:

A



B

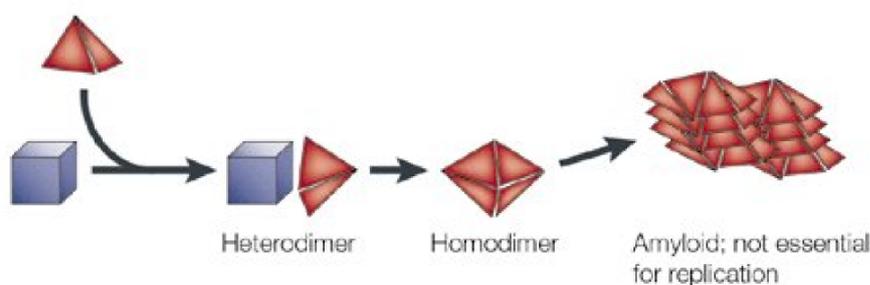


Figure 4.5: Comparison between the conformational and the seeded templating models for amyloid aggregation, with the prion protein PrP as an example. A) In the conformational model, the amyloidogenic PrP^{Sc} conformation (blue cube) and the physiological PrP^{C} conformation (red pyramid) are in equilibrium in solution. In appropriate conditions (*i.e.* high concentration) amyloidogenic conformer recruits other similar conformers and starts aggregation, thus shifting the equilibrium towards more PrP^{Sc} . B) In the seeded templating model, a “seed” of templates a non-amyloidogenic conformer into the amyloid structure, thus propagating the fibril. Adapted from [122]

Seeded templating model. In this case, aggregation starts from a “seed” of a few molecules, able to catalyze the appropriate conformational change on the monomer molecules joining the seed, thus leading to the birth of the fibril. Computer models show some evidence of this process: small artificial peptides can be “templated” by an already existing seed of preformed beta sheet [123] and

the Alzheimer's A β peptide can acquire a β -sheet structure by docking a pre-formed fibril[124][125]. Experimentally it has been found that this model holds in growth of fibrils of the protein barstar[126] and it explains the transmission of prion strains[107][127]. The tau protein fibrils structure suggests a template-mediated filament growth by parallel stacking[128]. A fundamental question however arises in at least some cases (for example, transimission of prion strains): do seeds "select" already existing, even poorly populated, monomeric conformations or do they directly template them?[129]

Conformational model. In this case, aggregation ensues from an aggregation-prone monomeric form that is usually poorly populated in non-pathological conditions. Imbalance in the conformational equilibria of the protein monomer could lead to an increase of the aggregation-prone conformer; when two or more of these conformers meet, the nucleus for protofibril/fibril formation is formed. Also in this case there is evidence. *In silico*, equilibria between discrete conformational states of Alzheimer's A β peptide, populating α -helix and β -sheet conformations, have been found[130]. Experimentally, there have been a few cases in which conformational equilibria shifted towards a distinct, amyloidogenic intermediate has been identified. Most strikingly, in the case of β -2-microglobulin the amyloidogenic intermediate is long-lived enough to be separated, and characterized by mean of mutation models[131][132]; the same for polyQ[133]. Also, transmission of prion strains can be explained in terms of different, overlapping conformational spaces accessible to the different sequences[107] and there is indirect evidence of a partially folded intermediate in α -synuclein [134].

The two models must not be thought as mutually exclusive. Most probably, each model can explain a subset of amyloidogenic proteins. Moreover, it is possible that even in the same protein the two models can be applied. For example, a conformational model could explain the initial formation of the seed, while templating could be most important in the fibril elongation.

Experimental investigation of the models, able to distinguish reliably between the two and/or study their intertwining is however still extremely challenging. In the case of the conformational model, evidence has been obtained only in the cases the aggregation-prone conformer is long-lived enough to be analyzed with standard biochemical techniques. However such techniques cannot easily probe the distinct

structures of conformational equilibria with short life times. It can be argued that single molecule techniques will be the next technological step towards investigation of the origin of amyloid aggregation (see Section 4.5).

4.4 Conformational disorder and amyloid aggregation

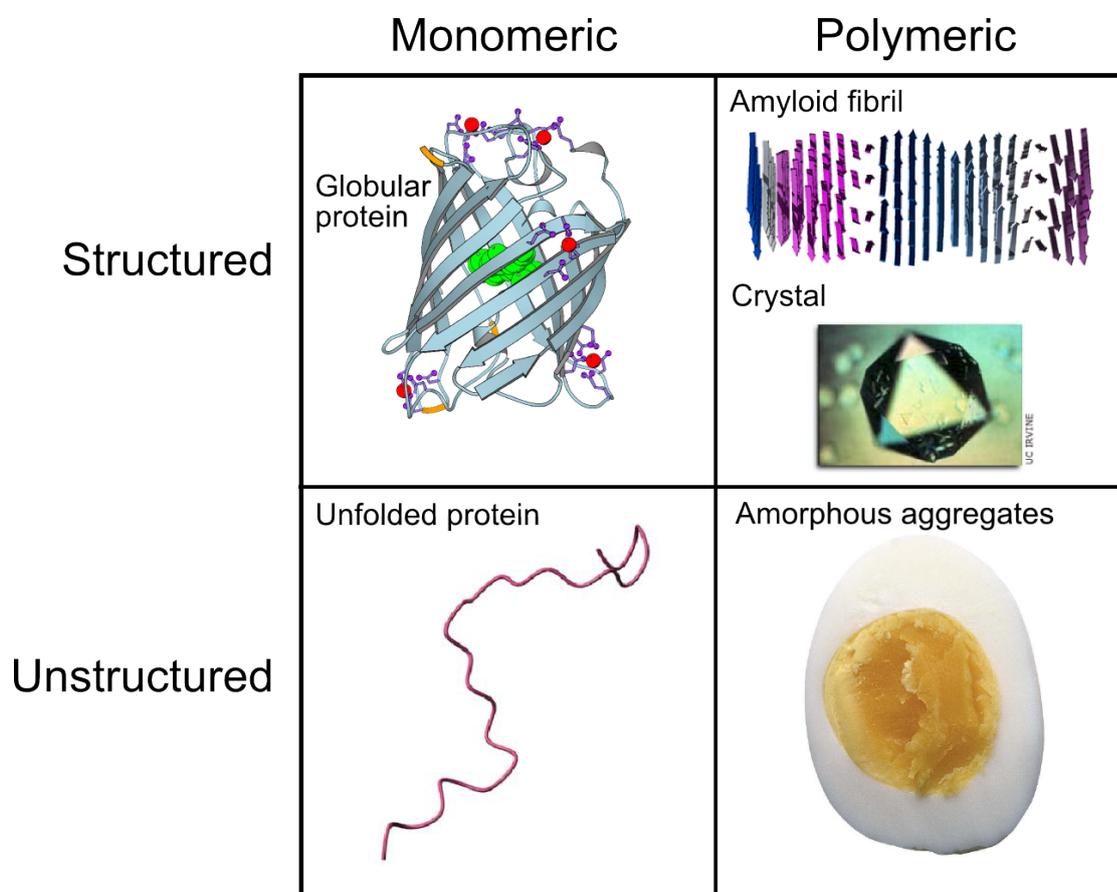


Figure 4.6: Diagram of the possible structural phases of a protein, depending on the acquisition or less of a rigid, folded structure and on them being monomeric or polymeric.

The existence of IUPs and the features of the aggregation processes, described in the previous sections, reveal that polypeptide chains can adopt a multitude of

conformational states and interconvert between them on a wide range of timescales. The conformations available to the protein are of course interconnected by multiple equilibria as shown in Figure 4.3.

There is considerable evidence that the common globular, folded state of proteins is just one of several *phases* in which proteins can be found[86]. Conformational disorder and amyloid semicrystalline phases are just as intrinsically belonging to polypeptidic chains as globular states.

Fully or partially unfolded ensembles have energy landscapes devoid of significant ($> kT$) potential wells. Any modification of this landscape can lower the ΔG of a structure belonging to ordinate phases, and amyloid phases can be selected just as non-amyloid ones. This makes IUPs and unfolded states particularly vulnerable to aggregation[85][80]: in fact, proteins like α -synuclein, $A\beta$ peptide and the prion protein are partially or fully unstructured in their common solution states. The structured polypeptide aggregates can then assemble into full-grown amyloid fibrils, both by self-association or simply the binding of new monomers. Proteins that adopt a folded structure under physiological conditions can also aggregate under some circumstances. To achieve aggregation, folded proteins often must unfold at least in part. Partial unfolding then leads to aggregation directly or the monomers can form on-pathway oligomers before conformational changes ensue, blocking the protein to a fibrillar destiny. In this case, the conformational changes can be driven by the binding surfaces of the different monomers in the oligomer. The routes leading to amyloid formation are thus different depending on which is the original state of the protein, natively folded or intrinsically unstructured[80]. This essential difference in the starting points, reflecting significant differences in the aggregation pathways, is only now beginning to be recognized.

It turns out that the common structured phase of globular proteins, where a sequence folds in a rigid state with very limited flexibility, is not representative at all of the intrinsic polymorphism of the proteion chain. All other structural states (unfolded or amyloid), in fact, show broad conformational diversity even within the same sequence. This conformational flexibility, which is used with parsimony and driven by evolution with amazing precision to fulfill biomolecular functions, is in turn at the heart of the pathological processes of misfolding and amyloid aggregation.

4.5 Single molecule techniques for intrinsically unstructured proteins and aggregation

We have seen that IUPs and amyloid aggregation, apart from being causally related, share a common theme. In general, proteins (IUPs or not) must switch their conformation, at least partially, to trigger amyloid aggregation. The complexity of conformational equilibria of proteins in the monomeric state is therefore not only the essence of IUP behaviour, but it is also at the core of the problem of amyloid aggregation. We stated above in many sections that such equilibria, involving the fast, continuous transition between a number of structures of nearly equal ΔG , cannot be fully understood at the level of the bulk. Doing that is like understanding how birds are by looking at blurred, long exposure pictures of a flock: we know they fly, but we don't see *how*. In particular, it is possible that the structures leading the biological actions are only a minority of the sample. In this case, ensemble-averaged techniques can tell us very little on the properties we are interested in. For example, in the Section 4.3.3, we have seen that the seeds of amyloid aggregation are, possibly, poorly populated conformers lost in a bulk of other, inactive conformations.

For these reasons, single molecule techniques, able to gain information from molecules one by one, are entering the arena of the investigation of these systems. The results are still sparse, but nonetheless interesting. Among the first studies there are single molecule force spectroscopy (SMFS) (see Part II) experiments on the unstructured PEVK and N2B segments of human titin[135][136][137]. These experiments characterized conformers with different elasticity, attributed to the existence of several isoforms of the segments featuring a continuum of *cis-trans* states of proline residues.

AFM-based force spectroscopy has also been used for characterization of the conformational heterogeneity of α -synuclein, the amyloidogenic protein responsible of Parkinson's disease. Ray and Akremitchev[138] described the interaction between single pairs of amyloidogenic peptides derived from α -synuclein. They found that the α -synuclein fragments form conformationally heterogeneous dimer states not significantly affected by temperature. A similar approach was adopted by McAllister et.al.[139] on α -synuclein, A β peptide and lysozyme, as a function of pH. Interestingly, they found that pH values able to induce β -sheet structure

in α -synuclein and A β peptide was accompanied by an increase of the interaction force. The pH value corresponding to the strongest interaction force was also the one leading to highest propensity to fibrillate.

Studies are beginning to directly probe the conformational diversity of IUPs. Single molecule fluorescence resonance energy transfer (SM-FRET) was applied on the study of conformational diversity of the NM domain of the yeast prion Sup35p[140]. The analysis revealed that the NM domain exists as a continuous ensemble of collapsed but rapidly interconverting species, with fast fluctuation on the 20-300 ns time scale. A recent, detailed single molecule force spectroscopy study on α -synuclein by the author has suggested that α -synuclein instead exists as a multiform ensemble separated in three different conformational classes, of which one is correlated with amyloid aggregation (see Part IV and [141]).

The problem of amyloid aggregation is also beginning to be tackled at the single molecule level. The β -sheets of A β peptide amyloid fibrils have been unzipped by mean of SMFS[142]. Surprisingly, the unzipping of β -sheets is elastic and reversible, giving interesting hints on the mechanics of amyloid aggregation. Single molecule fluorescence was employed to monitor the growth of Sup35p amyloid fibrils, confirming that growth occurs prevalently by monomer addition and, possibly, templating of the monomer on the fibril end[143].

Part II

Single molecule force spectroscopy of proteins

Chapter 5

Single molecule force spectroscopy

I know a man who grabbed a cat by the tail and learned 40 per cent more about cats than the man who didn't.

Mark Twain

5.1 Single molecule techniques

Molecules have long been inferred entities, even if they build everything concrete around us (including us): the acceptance of the atomic model was still challenged by Ernst Mach in the first years of the XX century[144]. Molecules were and are purified, characterized, modified, broken and rebuilt, but the experimenter actually never saw them directly. At best, he could have a detailed but average image of their shape from X-ray crystallography or NMR spectroscopy. The actual object of study was not the molecule itself, but the average properties of a bulk of billions of individual molecules. This is far from being an optimal situation: imagine what astronomy would have been if it was not possible to image individual stars, but only an “average” star being the weighted sum of all stars shining in a portion of the sky. Working with an Avogadro’s number of molecules means that rare intermediates or conformers will be irremediably lost in the background of the most populated species, and are overlooked. It also means that individual steps of a multistep process are usually not synchronized, thus making their study and

individuation difficult.

In the last twenty years, experimental techniques have been developed that allow to observe and manipulate molecules one by one[145]. Single molecule techniques made possible a huge conceptual shift in (bio)chemistry. The behaviour of molecules in bulk is an emergent property of the collective behaviour of individual molecules. Working at the single molecule level makes available to investigation the real phenomena underlying these emergent properties. The diversity of individual structures that form an apparently homogeneous bulk sample can now be directly probed. The molecule itself can now be the real, concrete object of study. Most importantly, most of those techniques probe the “dark zone” between the Å and the μm scale. This is one of the most difficult length scale to probe: too large for conventional crystallography, too small for conventional microscopy. Yet it is the scale at which most essential, dynamical processes of molecular biology take place, because it is the length scale of most biological macromolecules supramolecular complexes.

Single molecule techniques include, among others, atomic force microscopy (AFM) imaging, single molecule fluorescence resonance energy transfer (SM-FRET), and single molecule force spectroscopy (SMFS). In the first, the sample is deposited on a surface and then imaged by mean of a probe. Usually little or no modification to the sample is needed. In the second, the energy transfer between fluorophores is monitored on individual molecules. Molecules require large fluorescent moieties to be attached to them, but after that they are left untouched.

In the last, the molecule is directly manipulated in real time by the experimenter by applying a mechanical force on it. SMFS stands out among single molecule techniques in being a both a nanomanipulative technique and an analytical technique. It works by picking up the molecule and directly apply to it a stretching force (or, more rarely, a torque). The mechanical response to the applied force is directly monitored. This opens the way to investigate directly, in real time-fashion, force dependent molecular process and reactions. The behaviour of the molecule is studied along a well-defined mechanical reaction coordinate, allowing to probe alternative regions of the energy landscape.

Force spectroscopy is a general term that includes vastly different techniques. The two most used force spectroscopy techniques are optical tweezers-based SMFS (OT-SMFS) and AFM-based SMFS (AFM-SMFS). In the first case, a laser-based

optical trap is used as the manipulating device. In the second case, the manipulator is a standard AFM tip. Optical tweezers have been widely used to study the mechanics of motor proteins, while AFM has been applied mostly to characterize the mechanical resistance of both individual polypeptides (intramolecular interactions) and protein-biomolecule bonds (intermolecular interactions). AFM has the remarkable advantage, among single molecule techniques, to be commercially available: optical tweezers and other less common force spectroscopy techniques (bendable microneedles, magnetic tweezers etc.) are almost always custom-built.

Compared to other nanomanipulation techniques, AFM has high distance resolution and precision of positioning (of the order of a few nanometers), while it cannot resolve forces of less than 10-20 pN. AFM is particularly suitable for single molecule studies due to the relatively small (about 50 nm) radius of the AFM tip, much smaller than the radius of optical or magnetic beads and thus able to hold a smaller number of molecules or molecular pairs.

In this thesis, only AFM-SMFS will be discussed; for info on OT-SMFS you can read [146] [147].

5.2 Atomic force microscopy

5.2.1 AFM as an imaging device

The atomic force microscope (AFM) is a representative example of instruments known as *scanning probe microscopes*. Scanning probe microscopies work by monitoring the value of a physical variable that depends on the distance between the surface to image and a specific probe. The topography of the surface is then reconstructed in three-dimensional detail. The prototype scanning probe microscope was the scanning tunneling microscope (STM), in which the probe is a microscopic electrode and the physical variable is the current between the electrode and a conductive surface on which the sample is placed. The tunneling current, varying exponentially and very steeply with distance, is a sensible probe of surface topography.

AFM instead works using probe displacement as the variable, being similar conceptually to a miniature phonograph. The probe itself is a solid, sharp microscopic tip located at the end of a flexible cantilever. The tip and the surface

CHAPTER 5: SINGLE MOLECULE FORCE SPECTROSCOPY

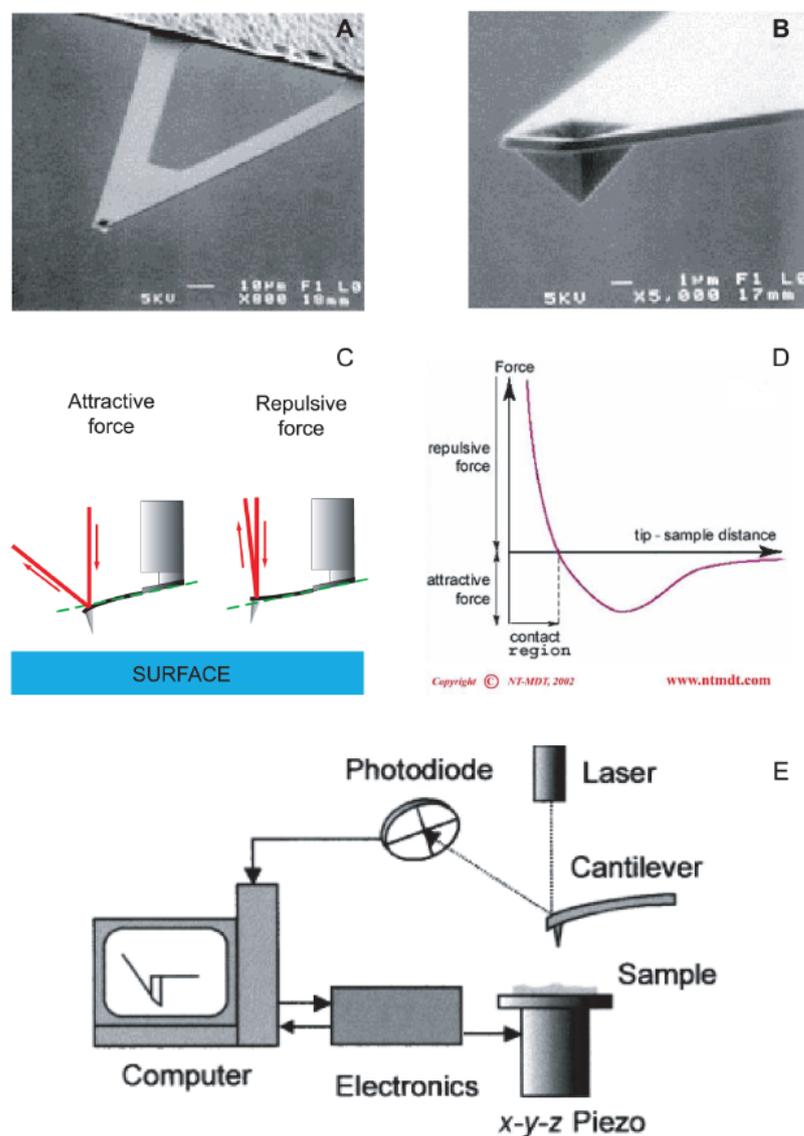


Figure 5.1: **A** SEM image of a V-shaped cantilever (the bar corresponds to $10 \mu\text{m}$). **B** SEM image of the tip of the V-shaped cantilever (the bar corresponds to $1 \mu\text{m}$). **C** Effect of attractive and repulsive interaction forces on the cantilever deflection and on the laser reflection. **D** Force acting on the cantilever as a function of the tip-surface distance. The attractive force preceding the contact is mainly due to Van der Waals and capillarity forces. **E** Scheme of the atomic force microscope, from [148]

are put in contact or close proximity, and the cantilever deflection due to the tip-sample interactions is the variable being monitored. The cantilever, normally, can

be modeled as a simple Hookean spring and therefore its response is linear with applied force. In most cases, cantilever deflection is in turn probed and amplified by a laser beam that reflects on the cantilever very end, forming a so-called optical lever system. The laser beam reflected by the tip hits a photodiode, which records the displacement of the laser beam from a reference point. The system in this way achieves a resolution of the order of the nanometer. A piezoelectric scanner is usually used to move the sample (or the tip) in three dimensions, allowing the tip to scan various portions of the sample. Piezoelectric scanners allow AFMs to reach 1 nm of lateral resolution. Atomic resolution has been reached vertically for hard materials in vacuum conditions. On the other hand, AFM can easily be applied to biological samples because it requires little or no sample preparation and works readily in physiological conditions (liquid buffer, room temperature etc.). Live eukaryotic cells can be extensively analyzed with the AFM[149].

5.2.2 AFM as a dynamometer

As described above, an AFM tip is basically a dynamometer that works on the nanoscale, converting cantilever deflection into force values. Behaving approximately as a Hookean spring, this conversion can be described mostly by the cantilever spring constant. For commercial cantilevers, the spring constant ranges from a few pN/nm to a few hundreds pN/nm (For comparison, an aluminum sheet 3 mm long and 1 mm wide has a spring constant of 1 N/m [150]). Being capable to measure the deflection with the nanometer precision, the scanning force microscope is a force sensor in the piconewton force scale. In practice, thermal noise limits the force resolution to > 20 pN.

5.2.3 AFM calibration

The cantilever elastic constant needs to be measured for each experiment: this is due to unavoidable inhomogeneities in the batches of commercial cantilevers. This is usually done today by the “thermal tune” method, that is by measuring the thermal oscillation spectrum of the free cantilever (not in touch with any surface) and using the equipartition theorem to obtain the actual elastic constant. The cantilever is here treated as a simple oscillator and the Brownian motion of the

main fundamental oscillation mode of the cantilever gives the elastic constant by the simple formula:

$$k = \frac{k_b T}{\langle z_c^2 \rangle} \quad (5.1)$$

where k_b is the Boltzmann constant, T is temperature and $\langle z_c^2 \rangle$ is the square mean cantilever deflection. This latter variable is found by fitting the power spectrum of the cantilever thermal noise, with the advantage of avoiding interference from both non-thermal oscillation at other discrete frequencies and white noise. In practice, it must be noticed that today more complex, empirical equations are used, to correct significant deviations from the ideal oscillator behaviour of the cantilever. The typical error is $\pm 20\%$ [151] [152].

The other essential parameter to be known is the optical lever sensitivity, *i.e.* the ratio between cantilever deflection and the photodiode output voltage difference. This depends from the optical properties of the cantilever, the medium (air or water), the eventually present fluid cell etc. and also needs to be measured for each cantilever. This is usually done by pressing the cantilever on a hard surface (*i.e.* glass or mica) at high forces: assuming that no substrate deformations happen, we know the cantilever deflection must be theoretically equal to the piezo movement and the ratio can be calculated. When this is done, the sample can be put in place under the tip. There is a number of free parameters that the experimenter must decide before starting the experiment (most of them can also be changed during operation): the speed, the length range, the maximum force acting on contact on the surface, the scan rate (number of points recorded per second) are usually the most important (see Table 5.1).

Variable	Typical range
Speed	0.1 - 5 $\mu\text{m} / \text{s}$
Length	100 nm - 5 μm
Maximum pushing force	0 - 10 nN
Scan rate	1000 - 10000 points per second

Table 5.1: Typical ranges of experimental variables used in SMFS.

5.3 Constant speed force spectroscopy

The name of “force spectroscopy” includes a variety of related but different techniques. What defines force spectroscopy techniques is the involvement of a mechanical force that acts on the sample and the recording of the sample mechanical response to this force. Differences lie in how the forces are made to act, in space and time, on the sample. Different force/time patterns give rise to vast differences in the actual signal, its interpretation and the behaviour of the sample, and in the end define related but markedly different analytical techniques.

Constant speed force spectroscopy (CS-FS) is the oldest and most popular of force spectroscopy techniques. This is also the technique with which the results described in this thesis has been obtained. Usually when the term “force spectroscopy” is used without other connotations, this is the technique it refers about. CS-FS is so popular because it is the technically easier force spectroscopy technique. Virtually any commercial AFM equipment is able to perform constant speed force spectroscopy, and basic operation requires no particularly sophisticated electronic feedback.

Other techniques like force clamp FS will be treated in the next chapter, since they need a deeper understanding of protein unfolding for their potential and limits to be understood.

5.3.1 Features of a force curve

The constant speed force spectroscopy raw data output is a piezo displacement Δz_p vs cantilever deflection Δz_c plot, where the distance is usually a fixed range between two fixed, arbitrary Z levels (usually decided by the experimenter). This plot rarely used by itself¹: the actually useful plot (usually called *force curve*) is the tip distance D (distance of the tip from the surface along the Z axis) vs force F plot.

To obtain the force curve from the raw distance vs deflection plot, two corrections need to be made. First, the deflection can be easily converted into the the

¹However the raw data output is often of paramount usefulness for SMFS data processing algorithms; see Part V

force by multiplying it for the measured elastic constant:

$$F = k\Delta z_c \quad (5.2)$$

Second, we are interested in the real distance between the tip and the sample. This distance will differ from the piezo Z displacement Δz_p by the amount of Δz_c - that is, the tip will be closer or farther from the surface depending on the cantilever deflection. So we have to correct each point of the X axis by this amount:

$$D = \Delta z_p - \Delta z_c \quad (5.3)$$

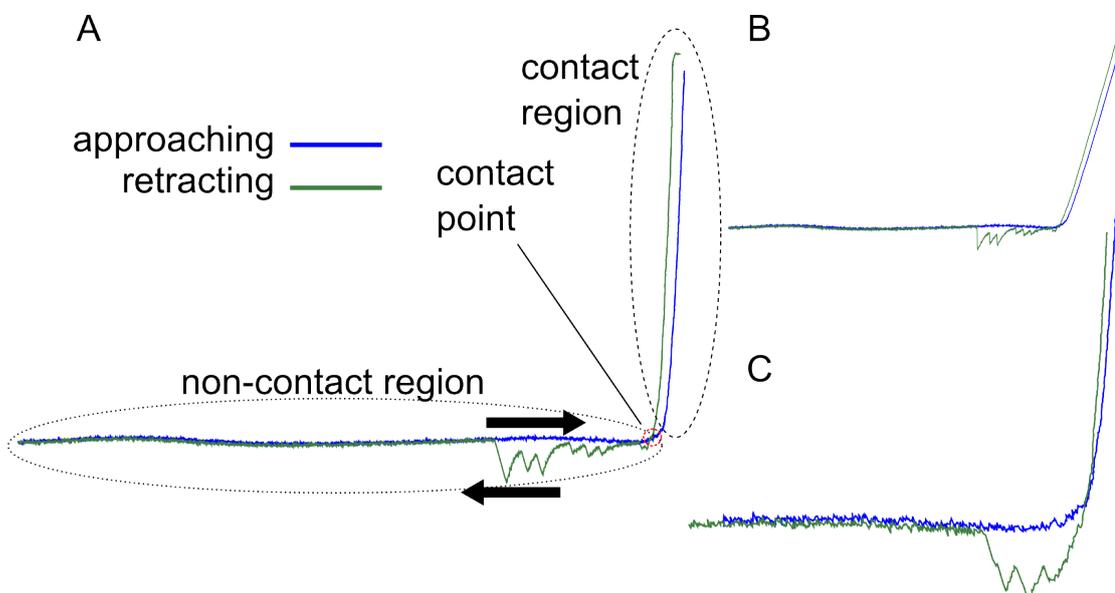


Figure 5.2: A) Example of a real force curve (tip distance *vs* force), with the contact and non-contact region, plus the contact point, evidenced. The retracting curve shows some peaks. The arrows show the time direction for the approaching curve (right-pointing arrow) and retracting curve (left-pointing arrow). The gentle oscillation of the non-contact region is due to optical interference (see text). The whole curve spans about 500 nm of tip distance; peaks are about 100-200 pN high. B) The raw piezo displacement *vs* cantilever deflection plot of the same data. C) Zoom of another force curve, showing specific interaction. Notice also the thermal noise.

The plot is actually composed of two curves: the first is called the *approaching* curve, recorded when the tip is approaching the surface; the second is the *retracting*

curve, recorded when the tip is retracting from the surface. Ideally, if nothing has been captured by the tip, the two curves superimpose perfectly, apart from noise.

² Each of the two curves features two different regions: the *non-contact region* is the horizontal section where the tip is not touching the surface, and ideally stays flat at zero force; and the *contact region* that is the almost vertical portion drawn when the tip is pushing on the surface and the force acting on the cantilever increases while the tip stays at the same level.

The point located at the boundary between these two regions is the *contact point*. The contact point is of paramount importance, since its location determines on the force curve where the surface is located. Finding the contact point may look trivial, but as we will see below, on a real force curve it is often challenging and its determination is subject to substantial errors.

On the retracting force curve, a force acting on the cantilever gives rise to a deviation from the flat profile of the non-contact region. A deviation in the opposite direction of the contact region is an attractive force (that is, something tends to keep the tip close to the surface), while a deviation in the opposite direction is a repulsive force. In practically all cases what is measured is an attractive force, due for example to a molecule tethered between the tip and the surface that is resisting traction. When the tip manages to free itself from the action of the force (that is, for example, the molecule detaches from the tip) the force acting on the cantilever abruptly falls back to the zero level. Thus, the breaking of interactions between the tip and the surface usually appears as a characteristic *peak* on the force curve.

An example of such a peak is the so-called *aspecific peak* that often masks the contact point in the retracting curve. This peak is due to the interactions between the tip and the surface material, and their presence and magnitude are related to the characteristics of the surface and the medium (air or water); charged surfaces like mica in air tend to give very strong (larger than 1 nN) aspecific peaks, while metal surfaces like gold tend to give smaller peaks. The existence of the “blind window” due to the aspecific peak makes the determination of the contact point less reliable and the interpretation of signals in the first 5-15 nm of extension dubious at best. It is thus strongly recommended to design the experiment in such a way the interesting signals are located more than 10-15 nm away from the the

²The cantilever drift can in practice slightly offset the two curves.

contact point.

The shape of peaks due to the sudden rupture of an interaction is usually asymmetric, with a characteristic curved slope before the sudden release of force. This slope contains information on the mechanical properties of the molecule(s) that bridge the tip and the surface, as explained in detail in section 6.3.2.

Another characteristic signal of force curves is the *plateau*. While peaks are due to the sudden rupture of a single set of interactions, far from equilibrium, *plateaux* are due to the continuous peeling of an high number of small interactions in a “velcro-like” fashion, in a near-equilibrium condition. Examples of force curves involving plateaux are the peeling of a polymer chain from the surface[153][154], or the unraveling of long coiled-coil protein domains[155]

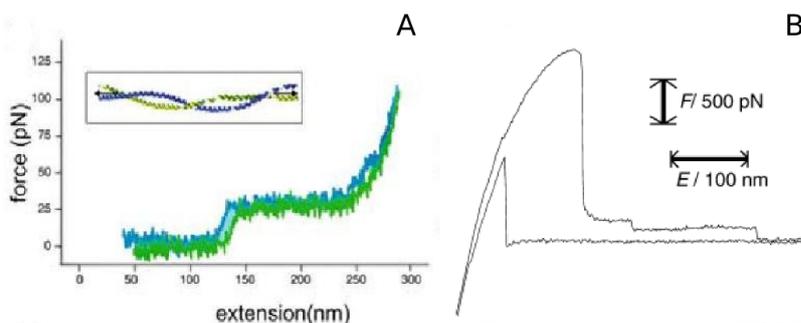


Figure 5.3: Example of two different plateaux coming from different processes. A) Plateau due to the unraveling of coiled-coil domains of myosin[154] and B) Plateaux coming from the forced desorption of polymers from a surface[153].

Instrumental problems can complicate the shape and interpretation of the force curve. *Cantilever drift*, due to the slow equilibration of the cantilever, can in severe cases add a linear slope on the non-contact region. *Optical interference* is an artefact due to laser light reflected by the cantilever and by the gold surface interfering on the photodiode. This is usually revealed by an oscillating pattern superimposed on the non-contact region, whose period is approximately equal to the laser wavelength.

Chapter 6

SMFS investigation of multimodular proteins

Bait the hook well: this fish will bite.

William Shakespeare, “Much Ado About Nothing” , Act II, sc. iii

6.1 Protein force spectroscopy operation

To analyze a molecule by mean of SMFS, the molecule must form a mechanical bridge connecting the cantilever tip to the surface: this bridge must be such that the end points of the bridge must be mechanically stronger than the intramolecular bonds we want to investigate (otherwise the molecule would detach from the tip or the surface before we break the intramolecular bonds). In theory it can seem excruciatingly difficult; in practice it is remarkably easy. The concept behind usual CS-SMFS analysis of proteins is “random fishing” of molecules from the surface, automatically done by the tip.

The protocol itself is simple. A 10-20 μL drop of a 0.5-2 μM protein solution is deposited on the surface (mica, glass or gold) mounted on top of the piezoelectric positioner. After an interval ranging between 15 and 45 minutes, protein molecules will have been adsorbed on the surface by physisorption¹. The protein can sometimes deposit as a layer 20-50 nm thick layer. It must be noticed that proteins can have the tendency to denature on the surface to some extent [156].

¹“Physical” , nonspecific adsorption of protein

Physisorption is a non specific process, and the mechanical resistance of such a bond can vary. To improve attachment of the protein to the surface, two cysteine residues are often added on one end of the protein (usually the C terminus) by genetic engineering; the -SH moieties will then covalently attach to a gold surface. Still, the effective gain in efficiency of this method is disputed[157].

The tip also attaches to the molecule by physisorption. In this case, physisorption is helped by the tremendous pressures the tip exerts on the surface when in contact. To have an idea of the numbers into play, a 30x30 nm square tip pressing 1 nN on the surface exerts a pressure of about 1.1 MPa, and this is a low estimate: pressures are estimated to reach even the GPa range. In these conditions, chemical bonds can actually form between the tip surface and the protein.

In a typical SMFS experiment, therefore, the protein-covered surface is put in a sealed fluid cell, containing the cantilever and appropriate buffer. The piezoelectric positioner cycles up and down: the tip approaches the surface, touches it and pushes on the surface until the pushing force reaches on average 0.5-5 nN. In this phase the tip hopefully adsorbs one protein molecule that establishes a mechanical bridge between the tip and the surface. The piezo then moves away from the tip, stretching whatever is between the tip and the surface. For each cycle the approaching and retracting curves are recorded. Several hundred cycles are made for each zone of the surface; the experimenter moves around on the surface from time to time, trying to “fish” the molecules. An average experiment usually requires to record from 3000 to 10000 force curves: this number of curves can be easily recorded in an afternoon of operation. However, very rarely a single day of data harvesting is enough to build a significant statistic on the system being probed: see 6.8 for discussion.

6.2 Why multimodular?

Very rarely is a single globular protein molecule stretched by mean of SMFS. In most cases, multimodular proteins (*i.e.* proteins made of individual globular modules linked in tandem, like pearls on a necklace) are stretched. There are two essential reasons for that. First, a globular protein is often a short object, with lateral dimensions ranging between 2 and 5 nm. This means that when the AFM begins to unravel it, we are still well into the “blind window” masked by aspecific

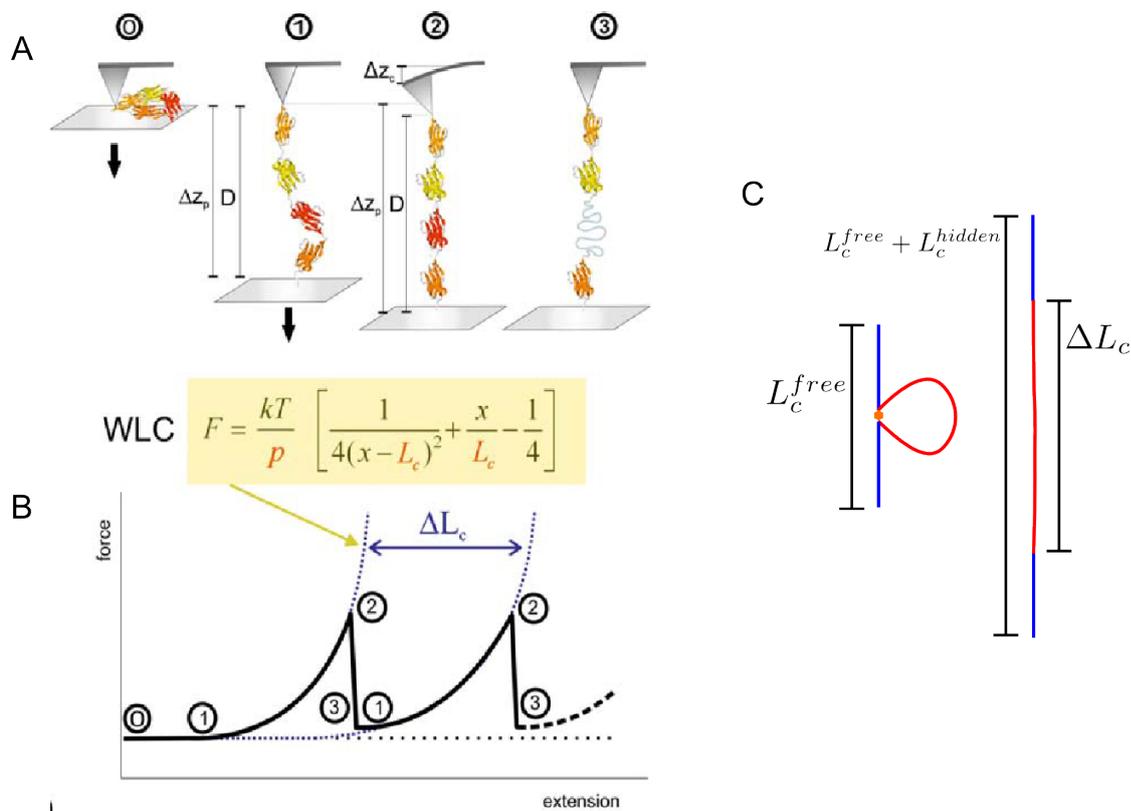


Figure 6.1: A) Example of AFM operation in force spectroscopy of multimodular proteins. The tip goes down on the surface “fishing” a protein molecule. The extension of the piezoelectric positioner Δz_p stretches the protein that bridges it to the tip. Increasing the distance between tip and cantilever, the conformational space available to the molecule decreases, and the reduction in entropy generates a restoring force that deflects the cantilever of Δz_c (states 1 and 2). This entropic force can be described by the WLC model of polymer elasticity. When one module unfolds (state 3), the contour length of the protein increases, relaxing the force on the cantilever to near zero. The cycle repeats until the last module unfolds, thus creating the sawtooth pattern in B). The last peak is created when the protein detaches from the tip or the surface. C), a model of the topology of a protein module and two flanking linker sequences. Left: before unfolding, the module behaves a hidden loop (red) kept together by a point-like interaction (orange); the sequence in blue yields the measured contour length. Right: after the interaction breaks, the hidden loop becomes exposed, and the total contour length increases; the difference between the two contour lengths equals the length of the previously hidden loop. A and B from [158].

interactions: distinguishing reliably the protein signal from aspecific signals is next to impossible². Second, we do not have any control on where the protein is picked up: the tip curvature radius is often about ten times larger than the protein itself; when the tip squashes down the protein, it can be physisorbed to the tip and to the surface from any point. So we not only do not know what portion of the protein we are stretching: we do not even know the direction of the force vector with respect to the protein, which is of fundamental importance for data interpretation.

In a multimodular protein, instead, we can ignore both problems. First, such a protein can be long enough so that the unfolding of first modules will happen well away from the aspecific interactions (for example, a fully folded protein made of 6 Ig-like modules in tandem is long more than 25 nm) Second, given a protein segment bridging tip and surface, we can “sacrifice” the modules forming the ends of the bridge, while studying the mechanical behaviour of the modules in between. In this case we are sure that, if our multimodular protein is made of modules connected by their N and C termini, the force vector pulling the protein is parallel to the N-C direction, allowing us to make considerations on the actual unfolding pathway of the protein.

There is also an historical reason for multimodular proteins being studied by mean of SMFS: many of them have an actual mechanical function. Titin, a huge muscle protein made of tens of folded independent modules that works as a “shock absorber” in muscular fibers, is the prototype multimodular protein studied by mean of SMFS, and most of our knowledge on protein mechanical behaviour comes from experiments on titin or titin modules. SMFS is however today applied also on proteins that are globular or even, as we will see in this thesis, natively unfolded. To study such proteins, we need to insert them inside a multimodular construct (see 6.4 for discussion).

²If the protein is linked to the surface via a self-assembled monolayer of organic molecules (SAM), thus passivating the surface, it is possible to effectively eliminate most aspecific interaction. This kind of setup is beginning to be investigated in SMFS use[159]

6.3 The mechanical unfolding of a multimodular protein

6.3.1 Understanding the sawtooth pattern

When a single multimodular protein has been fished by the tip, a characteristic regular sawtooth pattern appears on the force curve (see Figure 6.2). Why is it so? In understanding force curves, it is always useful to build a model of your system as a simple flexible chain(s) kept together by point-like interactions (see Figure 6.1 C). Regions belonging to two categories can be defined in the chain. The *free chain* is the portions of the chain that are not kept enclosed by an interaction. The *hidden chain* is the portion of the chain that is enclosed by an interaction.

Let us consider a chain where interactions define a number of equal hidden loops, under tension. In each moment, the only portion of the chain contributing to the elastic response is the free chain. The hidden chain, being locked by the interactions, does not participate to a first approximation to the mechanical response of the system. The length of the free chain is also called the *contour length* of the molecule. When one of the interactions breaks, the loop hidden by the interaction becomes instantaneously free. The overall contour length abruptly increases, with a sudden change in the elastic response of the molecule. More precisely, a longer chain is much more compliant than a shorter chain: the system therefore suddenly relaxes, with the force acting on the cantilever coming back to near zero. If the AFM keeps pulling on the molecule, it will stretch the free chain again, until another interaction breaks, and so on. The cycle takes place until the chain detaches from the tip or from the surface. Each cycle draws a peak on the force curve; the last peak corresponds to the detachment of the molecule from the tip or the surface. The distance between peaks *approximately* corresponds to the length of the hidden loop that has been freed.

We have now built a reasonable model of the mechanical unfolding of a multimodular protein. Each module behaves as a hidden loop containing most of the folded protein. Linker segments between the modules form the free chain. The mechanical unfolding of most protein modules studied so far is surprisingly catastrophic: that is, when a bundle of critical bonds break, the whole fold collapses and the chain unravels with little or none residual resistance.

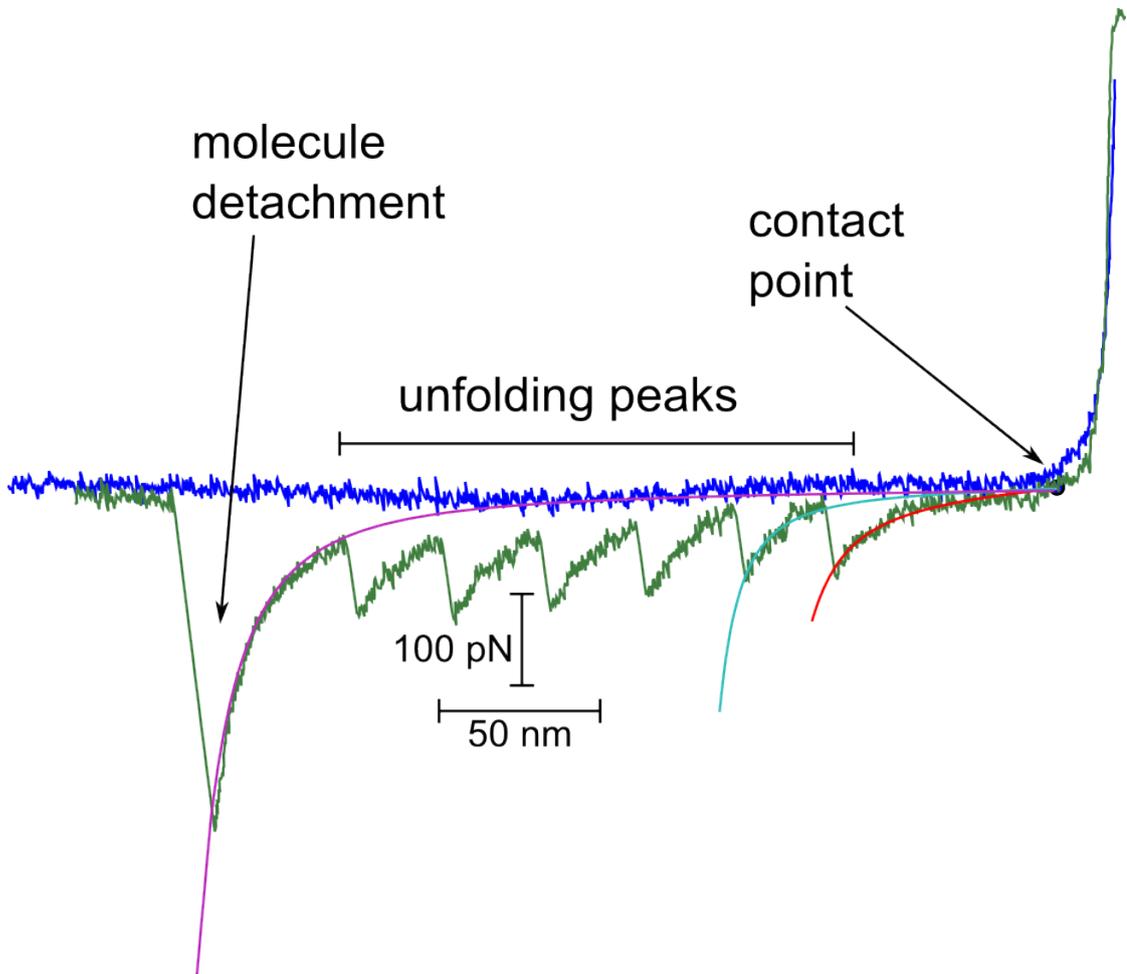


Figure 6.2: Example of an actual force curve from the unfolding of a multimodular protein. Six unfolding peaks coming from individual modules, plus the final detachment peak are clearly visible. A WLC fit is superimposed to the two first unfolding peaks and the detachment peak.

6.3.2 Polymer elasticity and the WLC model

What remains to be explained is the characteristic profile of the rise of each peak, that is, the elastic response of the polypeptidic chain.

The deformability of an ideal random chain molecule (approximations of which are a polyethyleneglycol molecule, a DNA molecule or a non-structured polypeptide chain) is controlled by its *entropic elasticity*. When a stretching force is applied to the polymer, its conformational space is shrunk: less and less con-

formations become available, until at full elongation there would be ideally only one conformation available (a straight, rigid chain). Given that a large majority of conformations are virtually at the same energy, entropy dominates the process, generating an opposing force.

The force (F) versus extension (x) profile of entropic elasticity is can be modeled by the worm-like chain model (WLC) of entropic elasticity, a statistical model for polymer chain conformation that models the chain as a flexible, continuous string [160]:

$$F = \frac{k_b T}{p} \left(\frac{x}{L} + \frac{1}{4} \left(1 - \frac{x}{L} \right)^{-2} - \frac{1}{4} \right) \quad (6.1)$$

The parameters describing the chain in the model are L , the contour length and p , the persistence length of the chain molecule. This latter is defined as the distance at which the curvature of the chain in a single point can influence the curvature of neighbouring point, and it is a measure of chain elasticity: the more the persistence length is high, the more the molecule is rigid. The WLC model describes very well the entropic elasticity of molecules, as long as $L \gg p$.

The WLC model predicts a non-Hookean, highly non-linear elastic response of the molecule (Figure 6.1 B). While extension approaches the complete elongation, the resisting force rises steeply and approaches infinity.

Fitting the force curves with the WLC model is standard practice in SMFS because it yields easily the two essential parameters L and p . It is extremely important to remind that **only by obtaining L from the WLC fit the true contour length of the object can be retrieved**: simply measuring the distance between peaks can give rise to significant errors in the estimate of the contour length. p can give more subtle informations on the nature of the molecule into stretching that are otherwise unavailable. While the expected L are often well known (they usually correspond to the length of the protein modules into study), p is usually unknown and depends both on the size of the polymer unit and other factors. For polypeptides, p is often in the range of 0.3-0.5 nm, next to the size of a single aminoacid unit. Variations in p have been investigated to discover equilibria between *cis-trans* forms of proline residues[135].

6.3.3 Deviations from the WLC model

When the elasticity deviates from the WLC profile, this implies the existence of an *enthalpic* component in the response of the molecule extension. Such deviations can be the result of intramolecular rearrangements and/or of rupture of chemical bonds.

Examples of this are the force-extension profiles of double stranded DNA and polysaccharides (see Figure 6.3). Force curves of individual dsDNA molecules show a WLC profile followed by a characteristic plateau around 60 pN. This profile has been attributed to a force-induced phase transition in the DNA structure: the base pairs separate and change their stacking. Only when the transition propagated to the whole DNA chain, the polymer behaves entropically again.

The forced stretching of polysaccharides induces different “humps” and plateaux, depending on the nature of the polysaccharide. Such deviations from the WLC profile has been attributed to force-induced chair-boat transitions in the sugar units, also thanks to steered molecular dynamics simulations.

In proteins, a classic example of deviation from the WLC profile is the *hump* displayed by the initial stretching of I27 titin modules. This hump corresponds to an elongation of 0.7 nm of the individual modules, happening at a lower force than the unfolding (about 100 pN)[162]. Other deviations show as smaller peaks superimposed on the main force peak, as evidenced for example in angiostatin [163] (see also III).

6.3.4 Basic information extracted from protein SMFS data

Force spectroscopy can basically yield three raw variables related to the behaviour of a protein molecule under tension: the increase in contour length ΔL_c , the unfolding force F_u and the persistence length p (see 6.3.2 for definitions of the first and last parameter).

ΔL_c We have seen the force-extension curve of a homomeric polyprotein features a series of equally spaced force peaks as the protein is stretched between the tip and the surface, with each peak marking usually marking the unfolding of a single domain except for the last, which is the detachment of the protein from the tip or the surface breaking the mechanical bridge.

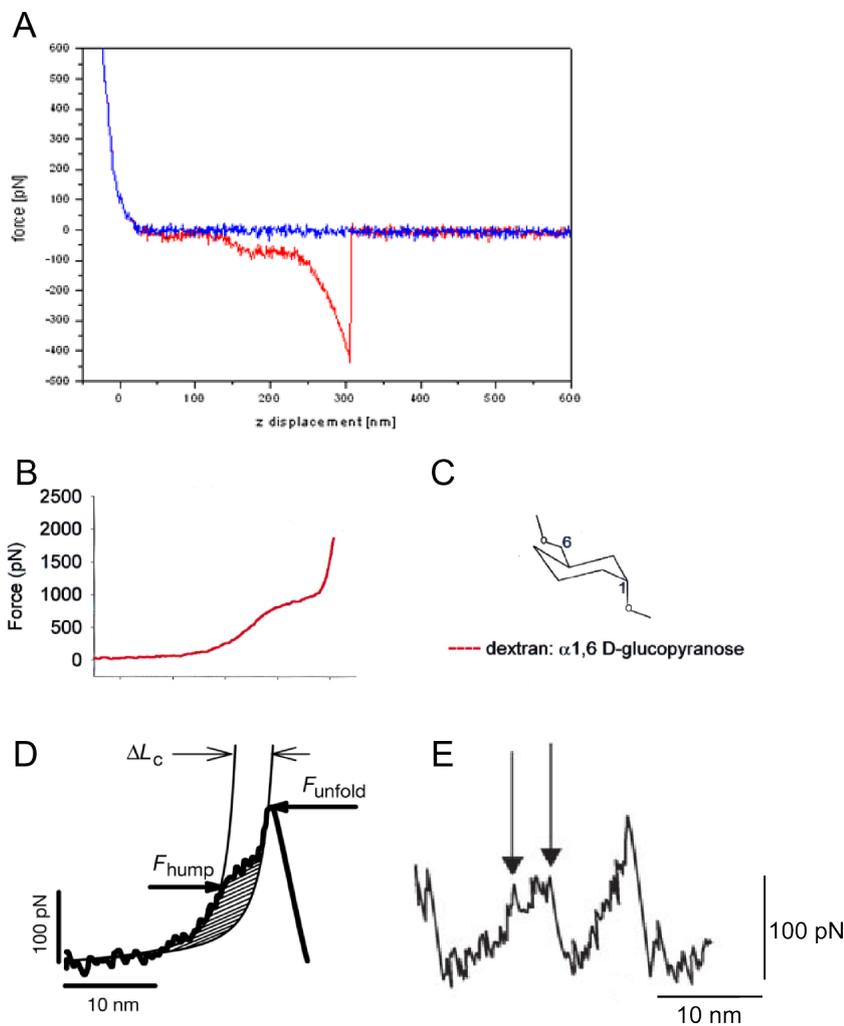


Figure 6.3: Examples of deviation from the WLC model. **A**, stretching of λ DNA molecule, showing a plateau at around 70 pN indicating the overstretching structural transition of DNA (from JPK website). **B** stretching of dextran molecule, showing complex transitions at different heights, due to the distortion of the sugar rings constituting the polymer (depicted in **C**) (from [161]). **D**, characteristic hump indicating the I27 domains structural transition prior to mechanical unfolding (from [162]) **E**, double peak in angiostatin module unfolding, showing an unfolding intermediate [163].

By fitting with the WLC model the rising part of each peak, we can know the length of the hidden loops being exposed by the unfolding of each domain (see 6.3.1). However we have also seen that, formally, each peak indicates the release of an *interaction* that needs to be broken to proceed in unraveling the polymer. That a complex protein fold must unfold by mean in an all-or-none fashion is often only a useful approximation: in practice, unfolding intermediates are quite often discovered. Intermediates will appear therefore as “humps” or smaller peaks superimposed on the force curve (see 6.3.3). Measuring the lengths of the hidden loops uncovered by the intermediates, when combined with molecular dynamics simulations of the unfolding process, can uncover information on the actual mechanical unfolding pathway.

p The WLC fit also yields the persistent length p . A change in p usually indicates a change in the properties of the monomers constituting the polymer. As expected, this value is usually relatively poorly informative for proteins: the p of polypeptide chains is usually in the range between 0.3 and 0.4 nm[164], close to the length of a single aminoacid. A notable exception was the measure of p monitoring the proline *cis-trans* equilibrium in the PEVK unstructured segment of titin [135].

F_u Protein mechanical unfolding is a thermally driven probabilistic process: actually, the force never directly breaks a bond. Instead, force tilts the protein energy landscape until the barrier to unfolding becomes of the order of kT or smaller (see Section 6.6.2) . Therefore, strictly speaking, we cannot talk about an “unfolding force” , but only of a most probable unfolding force. Moreover the actual force distribution depends on the pulling speed (more precisely on the *loading rate*, that is $\frac{dF}{dT}$)³ and therefore any given unfolding force is a relative value.

The distribution of the unfolding forces measured for a given module should contain information on the distribution of thermal energy available to each module prior to unfolding. This distribution is not Gaussian, but it has a characteristic shape skewed towards the lower forces[165][166]. This shape contains also informations on the kinetic parameters that describe the mechanical unfolding energy landscape: more precisely, in a single-barrier, single-well approximation, K_{off}

³Ideally, any protein domain will eventually unfold even at zero force, statistically. In general, the slower the pulling, the lower the most probable unfolding force. See Section 6.7.1.

(interaction lifetime) and X_b (barrier position along the reaction coordinate X). However, in practice experimental errors make it difficult to extract reliably those parameters from force distributions alone. The method of choice to obtain the kinetic parameters is based on the dependence of the unfolding force on the pulling speed (see 6.7.1).

ΔW The difference area between the approaching and retracting curve gives the energy irreversibly dissipated by the system during the unfolding process. This area, by itself, should be expected to contain little information on the effective ΔG of the system. However, exploiting the thermodynamic Jarzynski's equality, approaches have been attempted to obtain free unfolding energy information from the work dissipated in unfolding curves[167].

6.4 The polyprotein strategy for SMFS protein analysis

As explained above (6.2) a non-multimodular protein has to be studied by inserting it into an artificial multimodular construct. Even for natively multimodular protein, building an artificial construct is advisable. In fact, in presence of different modules of approximately equal length, it is practically impossible to assign a specific force peak to a specific module. To analyze the mechanical behaviour of a module, the best strategy is to build a so-called *polyprotein*, a perfectly repetitive, artificial protein made of several identical modules in tandem.

Building a polyprotein not only avoids concerns about the aspecific interactions and the pulling direction, but provides a number of fundamental advantages. Multiplying the number of modules in a molecule means that for each molecule we have many signals to analyze, thus amplifying the rate of data acquisition. The periodicity of the polyprotein signal is a natural fingerprint of the desired molecule that allows better discrimination of the signal of the protein into study from aspecific signals or contaminants. Using polyproteins allows also for “spatial amplification” of otherwise undetectable features: small aminoacid insertions[168], conformational intermediates close to the native state[162] and misfolding events[169].

6.4.1 Building polyproteins

Two strategies have been developed to this end:

1. Genetic engineering of head-to-tail DNA concatamers[170]. First implementations used a single restriction site, concatamerization and selection of the desired number of repeats after gel separation; a cassette strategy relying on different restriction sites has recently been established. This is the most used current strategy.

Advantages:

It allows for specific modules to be inserted at desired position (chimeric polyprotein)

Requires only standard molecular biology techniques

Head to tail directionality automatically obtained

Disadvantages:

Only N-C terminal orientation is possible

Highly repetitive DNA sequences are prone to recombination

Slow, sometimes difficult expression and purification

2. A solid-state strategy where repeats are assembled post-translationally from chemically joining monomers via disulfide bonds [171]. It requires the protein to be easily crystallized. Residues on the protein surface that are in close proximity with other monomers in the unit cell are modified with cysteines. Solid-state oxidation in the crystal leads to assembly of monomers into disulfide-linked strings. The same strategy can be employed without crystallization (by oxidizing the proteins in fluid and engineering exposed cysteines in desired positions), but losing directionality.

Advantages:

Fast; once the protein has been engineered multimers can be obtained in a few days

Permits every kind of orientation, not only N-C

Disadvantages:

Requires a crystal to be readily available or to lose directionality

Highly repetitive DNA sequences are prone to recombination

Impossible to insert a specific module in a specific position

6.4.2 Polyprotein design and heteromeric polyproteins

The simplest polyprotein design is the homomeric polyprotein, *i.e.* a protein construct made only of identical repeats of the same module. Unfortunately, not all proteins are suitable for such a design. The signal of the protein may be particularly complex, and if weak signals add all together at the beginning of the force curve, interpretation can become difficult. Aggregation-prone proteins (like, for example, amyloid-forming proteins) are obviously best not assembled together in a polyprotein. Not all modules, furthermore, guarantee to fold independently in a polyprotein construct. Finally, if the signal of the protein is not guaranteed to be a simple, strong unfolding peak, we lose the “fingerprinting” advantage.

For these reasons, heteromeric polyproteins are today the strategy of choice for the SMFS study of protein modules. In this strategy, two different protein modules are inserted in the construct: here we call them the *markers* and the *analyte*. The marker is a well characterized protein module, proved to independently fold robustly in a multimodular construct, that yields a simple, strong unfolding peak immediately characterizable. Usually the Ig-like titin modules like I27 and I28 satisfy correctly all these requisites. The analyte is the protein module that is the object to the study.

The marker and the analyte are usually alternated in the sequence in a “sandwich” pattern (MMMAMMM) (Figure 6.4.2 A). This pattern has the advantage that we know a single unit of the analyte has been stretched if and only if we observe four or more peaks coming from the unfolding of the markers. By measuring the initial contour length and the separation between marker peaks, one can be sure of the nature of the object pulled, thus making data analysis considerably easier and safer (see 6.8 for discussion).

Another possibility is the alternate (MAMAMAMAM) polyprotein (Figure 6.4.2 B). An alternate polyprotein allows for amplified data acquisition, at the

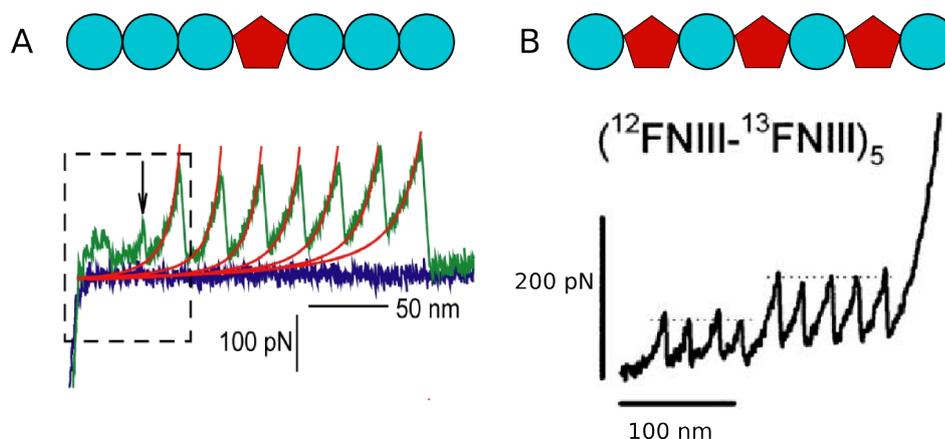


Figure 6.4: Polyprotein strategies for SMFS and corresponding experimental curves. A) Top: Sandwich pattern. The analyte (red pentagon) is flanked by marker modules (blue circles) at left and right. If at least four signals from the markers appear, the analyte will be surely have been stretched. Bottom: example of force curve from a sandwich polyprotein $(\text{I27})_3\text{-synuclein-}(\text{I27})_3$, from the author (see Part IV). B) Top: diagram of alternate pattern polyprotein. Bottom: example of force curve from alternate pattern[172]. Notice that, as expected, the order of peaks reflects the hierarchy of mechanical resistance of different modules and not their actual order in the construct.

expense of interpretation of signals if weak signals sum up in a complex fashion at the beginning of the stretching. A sandwich polyprotein sees the stretching of only one analyte molecule per successful pulling, but any signal coming from the analyte is unmistakably coming from a single analyte molecule. Choice between the approaches thus depends on the nature of the analyte; the sandwich polyprotein is usually the safest alternative.

6.4.3 Open problems in polyprotein design

As of 2008, this is the state of the art in polyprotein design. However, it is clear that many more factors than currently recognized have to be considered for polyprotein design. An ideal marker in fact has to:

1. Be easily distinguished from the analyte
2. Interfere as little as possible with the analyte

Due to the stochastic nature of the bond rupture, the first requirement cannot be fulfilled by a difference in force. The marker requirement about force usually is simply: as strong as possible, to increase the odds of having the analyte signal in a known place in the force curve (*i.e.* before the first marker unfolding peak). A difference in unfolding length, instead, can be measured with high precision, and can be easily used to distinguish a known marker from the analyte. Unfolding length depends only on the nature of the fold and on the module sequence length, and therefore does not change randomly or depending on buffer conditions. Unfolding length can therefore often be predicted, at least roughly, for the analyte and the marker as well. Unfolding length can also be shortened for a given protein module by engineering disulfide bonds in the module structure, mostly retaining mechanical properties, as successfully demonstrated by the group of Julio Fernandez[157]

The second requirement is more subtle. A polyprotein construct puts the analyte and the marker in extremely close proximity, and it can be expected this alters the analyte surrounding environment and therefore the analyte properties. There is evidence that even modules belonging to naturally multimodular proteins, and that therefore have evolved to be independently folding units, are slightly but measurably influenced by the nature of neighbouring modules[173][174]. In any case, the presence of two large moieties flanking the analyte unit has entropic consequences that must be taken into account[175][38].

It can be possible, however, to design the marker to minimize and control these influencing effects (see also Figure 6.4.3). The most important factor is the electrostatic nature of the marker and analyte surfaces. Markers having an electrostatic charge complementing that of the analyte should be avoided, since it is virtually certain that the analyte and the marker will strongly interact, potentially the analyte structure and leading to spurious analyte-marker detachment signals on the force curve. On the same basis, markers and analytes with large exposed hydrophobic patches should be avoided too. It is conceivable that the best combination is that of markers and analyte with roughly the same electrostatic charge, leading to repulsion and minimization of marker-analyte interactions. Tailoring the surface of an already known protein module, although requiring substantial work, can be made, as shown by[176].

Apart from marker tailoring, other improvements could be made. Simple examples of that include additional rigid sequences, for example polyproline stretches,

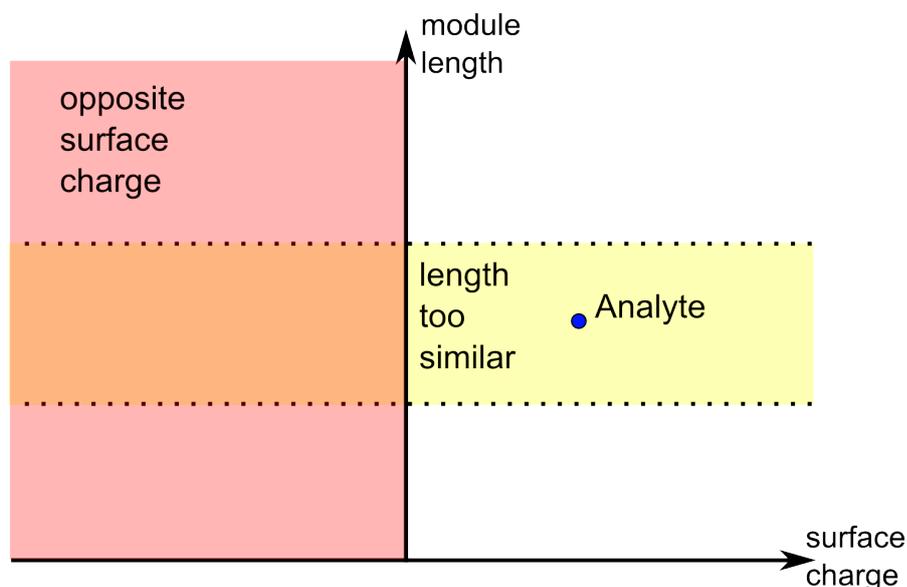


Figure 6.5: Simple length-charge diagram showing two parameters to be taken care of in the choice and design of markers for polyprotein construct for SMFS. Protein modules with lengths too similar to those of the analyte should be excluded because their length (and, consequently, their signal) is at risk of being confused with that of the analyte itself (yellow area). Protein modules of opposite electrostatic surface charge should be avoided too because they can interact too much with the analyte, creating spurious interactions and altering the analyte conformation (pink area). The optimal modules should lie in the areas in white.

could be added between the markers and the analyte to increase separation. The same rigid sequence could be also exploited at the ends of the molecule, to increase the odds of the polyprotein construct to be picked up at the very end.

Advanced hypotheses for marker design include the possibility of building “modular” marker libraries, in which markers are joined to the analyte by mean of click chemistry, so that different suitable markers are exploited for different purposes and proteins, without having to redesign a whole polyprotein each time. Non-protein markers can also be considered, as for example polysaccharide markers: however stretching polysaccharides do not give rise to clear-cut peak signals and can easily adhere to proteins, thus making their usefulness as markers still uncertain.

6.5 Molecular determinants of the mechanical stability of proteins

Despite early claims, SMFS experiments have shown that there is little overlap between chemical and mechanical stability of proteins. Thermodynamical stability ΔG or melting temperature $T_m = \Delta G/\Delta S$ do not correlate with mechanical stability [177]. This is due to chemical and mechanical unfolding basically following different pathways [178], thus having different unfolding barriers. Moreover, mechanical stability is not a property of a protein fold *itself*, but, being force a directional variable, it is a function of both protein structure and of pulling geometry.

6.5.1 Secondary and tertiary structure

In general, beta-rich structures can be more mechanically resistant than alpha helix-rich structures⁴. This has been confirmed both experimentally and by a wide-range survey of the Protein Data Bank by mean of coarse-grained Gō models [179].

The underlying reason of this hierarchy lies in the geometry of the bonds, particularly of the hydrogen bond network, with respect to the direction of the force vector. In this respect, two different general classes of geometries are distinct: *zipping* geometries and *shearing* geometries. In zipping geometries, two different protein strands are connected by hydrogen bonds parallel to the force direction. In this case, force is applied on a single bond that rapidly “unzips” , leaving exposed to force the next bond and so on, until the whole couple of strands has unzipped. In every moment, only a single bond is under direct tension, and therefore the net force required to unfold the protein is never vastly superior to that required to break a single hydrogen bond. Therefore, zipping geometries have very low mechanical resistance (sometimes even under the AFM detection threshold).

In shearing geometries, instead, bonds are perpendicular to force direction. Thus, to unfold the protein many bonds have to break at the same time (and following a different geometry). This event is exponentially less probable, at a

⁴This does not mean that *all* beta structures are stronger than *all* alpha structures: it means that alpha-helix structures can never be as strong as some classes of beta structures

given force, than the breaking of a single bond, and thus much higher forces are needed to make it happen. Shearing geometries have unfolding forces ranging from 100 to more than 200 pN at standard AFM loading rates⁵.

Strong shearing geometries are almost practically always found only among beta-sheet rich proteins. Alpha-helix bundles are not as tightly bound as beta strands, and tend to peel from each other in a zipping-like fashion (as dramatically seen in the case of the double helix coiled-coil of myosin[155]). Alpha-helix proteins have been much less studied by mean of AFM: the ones studied unfold at forces well below 100 pN. The mechanical unfolding of alpha-helix bundles tends to be more complex and less catastrophic than that of beta sheet proteins, as it has been seen for spectrin[181] although it has been disputed when using different setups[182].

6.5.2 Pulling geometry

Force is a vector quantity, therefore the mechanical stability of a protein will depend not only on the magnitude but also on the direction and application point of the force. Zipping and shearing geometries, by definition, are relative to the vector orientation. Therefore it can be expected that by varying the direction of the force vector, we can switch between zipping and shearing geometries in the same protein.

This concept has been first demonstrated not in proteins, but in DNA oligomers, by comparing the rupture forces of two oligomers of the same sequence, in which the orientation of the tails was different. The two configurations were measured to be thermodynamically and kinetically identical when probed with bulk methods, but the complex with shear geometry had an approximately 3-fold higher mechanical stability [183].

In proteins, it has been recently possible to experimentally pull some proteins from directions different from the canonical N-C direction. Such experiments took advantage of the existence of naturally occurring covalent linkages for internal aminoacid residues in the proteins ubiquitin and E2lip3. These experiments showed that, depending on the pulling direction, the mechanical stability of the

⁵The different mechanical resistance between shear and zipping geometries was already predicted on the basis of molecular dynamics simulations of proteins [180].

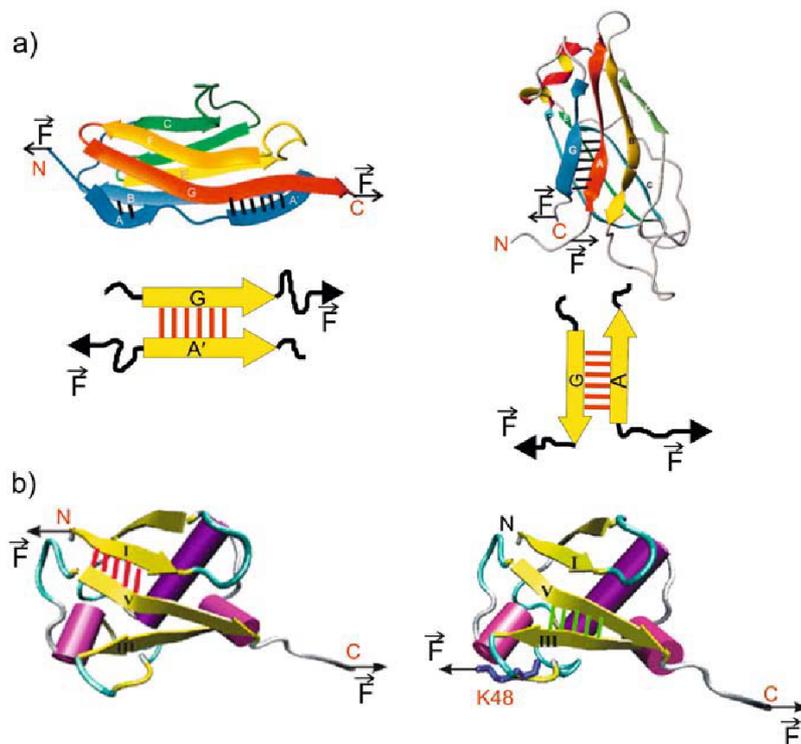


Figure 6.6: Pulling geometry effects on protein mechanical unfolding. **A** Shearing *vs* zipping geometries in proteins stretched from the N and C termini. The mechanical stability determinants of two Ig-like modules (I27 from titin and C2A from synaptotagmin I) are patches of backbone hydrogen bonds that can either be orthogonal (shearing geometry, left: the A'G patch of the I27 module) or parallel to the direction of applied force (zipping geometry, right: AG patch of the C2A module.) The zipping geometry leads to a lower mechanical stability, presumably because the bonds are subject to force one by one. The shear configuration requires instead an all-or-none mechanism of breakage. **B** The mechanical stability of ubiquitin depends on pulling direction. The most probable force required to unfold the ubiquitin domain is higher when the force is applied to the N and C termini (left) than if the force is applied between the K48 residue and the C terminus (right). The application points of the force are indicated in red. Caption and image adapted from [158]

same protein can vary dramatically (from <20 to 170 pN in the case of E2lip3). Coarse-grained simulations have also permitted to observe the same pattern for a wide range of proteins, allowing to map the mechanical resistance of a fold in function of the residues on the sequence where force is applied. This analysis showed

that nearly every protein module shows ample variance in its mechanical stability depending on geometry, with low-resistance geometries (“Achilles’ heels” being potentially selected for use by unfoldases in the cell[184]).

6.5.3 Disulfide bonds

The presence of disulfide bonds can dramatically alter the mechanical properties of a protein. Being covalent bonds, disulfide bonds become mechanically labile only at forces of about 1-1.5 nN, five to more than ten times higher than those necessary for the unraveling of non covalent interactions in a protein fold. Usually the protein is already detached from the tip or the surface when forces become strong enough to overcome the disulfide bond. In practice this means that disulfide bonds act as “locks” that can exclude whole protein segments from the action of force. This has a direct influence on the protein apparent contour length in the SMFS force curves: the unfolded protein domains will appear shorter than they truly are, loops being hidden by the disulfide bond (see Figure 6.7).

This has been dramatically shown experimentally thanks to the work of Discher et.al. [185] and of Bustanji and Samorí [186] on V-CAM and angiostatin, respectively. Both groups showed that the action of reducing agents modulating the redox state of disulfide bonds is readily apparent in the ΔL_c statistics. Bustanji and Samorí also showed that it is possible to discriminate and quantify the different population of partially reduced states, displaying different topologies, in protein modules containing more than a single disulfide bonds, by looking at the contour lengths.

6.6 The physics of protein mechanical unfolding

6.6.1 Equilibrium and non-equilibrium unfolding

In single-molecule mechanics, the superimposition of approaching and retracting force curves usually indicates, at least approximately, that the mechanical process is thermodynamically reversible. More often, the approaching curve is markedly different from the retracting one, if not completely flat, indicating that the mechanical process occurred far from equilibrium. The hysteresis, *i.e.* the area

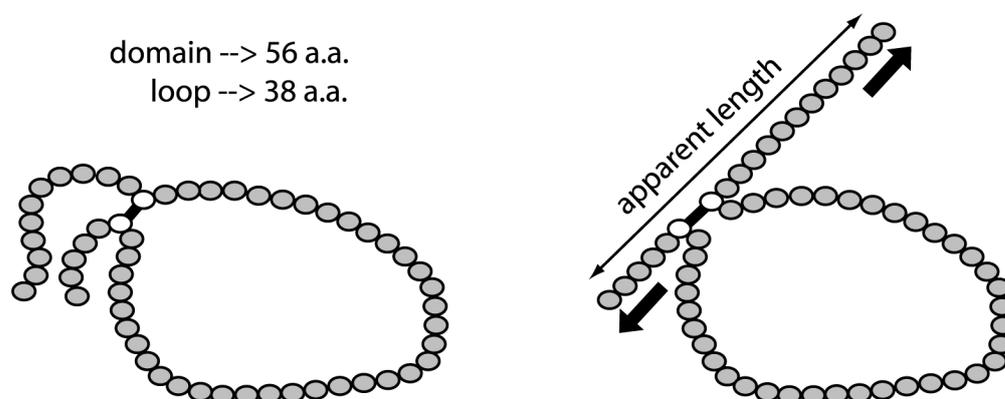


Figure 6.7: Disulfide bonds hide portions of a protein domain from an external force. Left: Two-dimensional sketch of a protein domain of length L containing an internal disulfide bond. The resulting topology of the domain allows to distinguish an inner loop of length L_{loop} enclosed by the disulfide bond. Right: The disulfide bond acts as a barrier to mechanical unfolding. The apparent contour length L_{app} that will be directly accessible to force in the same protein sketched left, is therefore given by $L_{app} = L - L_{loop} + SS_{bond}$, where SS_{bond} is the length of the disulfide bond itself (this can be safely neglected). If L_{loop} encloses a large fraction of total L , the presence or absence of a disulfide bond can effectively switch the protein between two radically different elastic states: a locked, inextensible one and an unlocked, compliant one.

between the approaching and retracting curve, is a measure of the energy dissipated in the process.

In the loading rate range available to SMFS, very few proteins act as true elastic elements, storing and yielding elastic energy reversibly with little heat dissipation. Notable examples are elastomeric structures like that of elastin [187], the PEVK and N2B domains of cardiac titin [135] [137], the EH domain of myomesin [188].

The mechanical unraveling of the myosin II tail is an example of a near-equilibrium peeling of secondary structure elements. The two long coiled alpha helices unzip from each other at next to constant force, producing a plateau (see Figure 6.6.1, center). Relaxing the molecule brings the helices back in coiled-coil, following the same pathway [155].

Most proteins, however, display equilibration rates much slower than the timescales accessible to SMFS experiments. In proteins it is usually impossible to distinguish

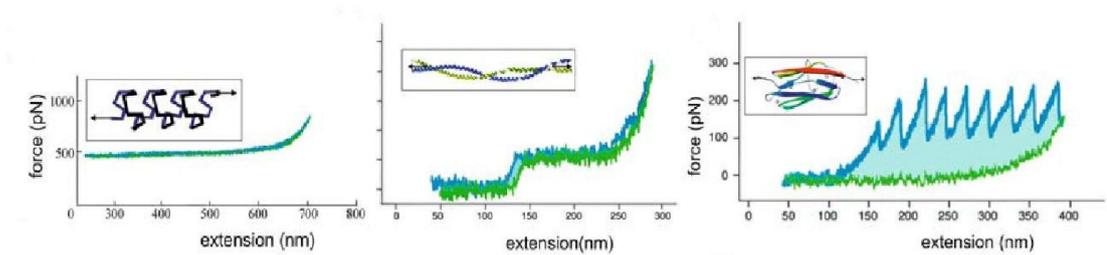


Figure 6.8: Equilibrium and non-equilibrium elastic structures in protein unfolding. Left: The protein elastin behaves a reversible entropic spring, with nearly complete superposition between approaching and retracting curve[187]. Middle: Unraveling of the myosin II coiled-coil displays a reversible plateau region, again demonstrated by the superimposition of approaching and retracting curves[155]. Right: The unfolding of β -sheet rich polyprotein is instead a highly dissipative process, as shown by the large hysteresis between the approaching and retracting curves (light blue). Adapted from [158].

the collapse of tertiary structures and the subsequent unraveling of secondary structures, as for RNA: the unfolding of most protein domains appears as a cooperative process where the module mostly resists to the force until a critical bundle of bonds breaks up: in this case the whole fold collapses and subsequent unraveling takes place with little or no resistance.

6.6.2 Description of single barrier mechanical unfolding

Since most folds are mechanically defined by what looks like a single, strong interaction, a discussion of the behaviour of a non-covalent bond under a stretching force is suited to describe it. As sketched in Figure 6.6.2, a bond can be, in first approximation, described by an energy profile where a barrier confines the bound state away from the dissociated one. The energy difference between the bound and the dissociated state controls the relative populations of the two states. This energy gap can be measured by traditional bulk ensemble techniques, but as pointed out above, by measuring only the energy difference we learn very little about the relations between force, lifetime and affinity that describe the dynamics of a binding event. In general, any bond or interaction has a limited lifetime, so if we wait

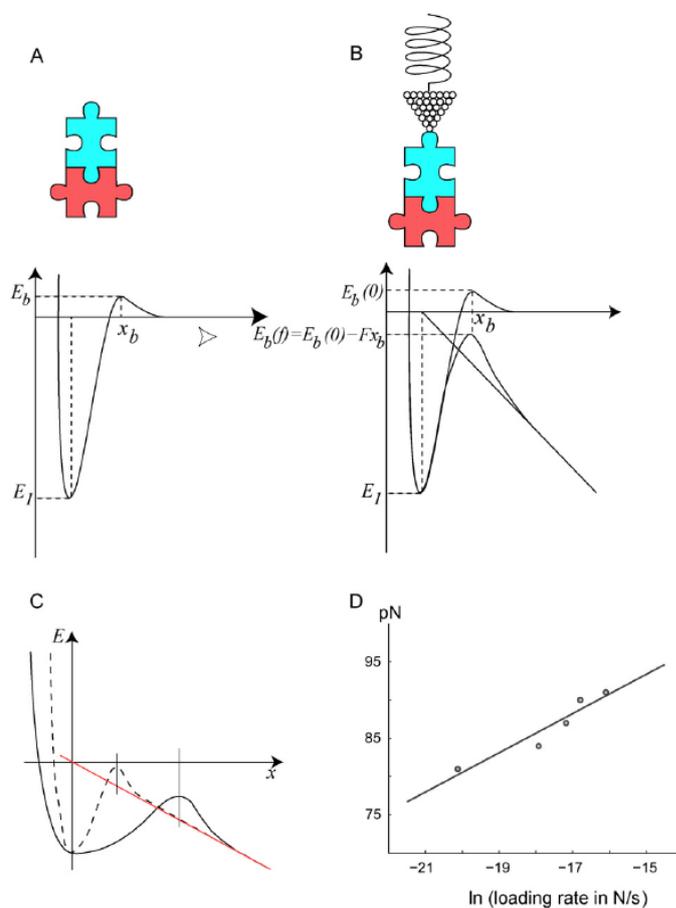


Figure 6.9: The energy landscape of a binding interaction is tilted by the application of an external stretching force. (A) Energy landscape at zero force of a pair of interacting molecules. A barrier whose height is E_b is located at X_b along the reaction coordinate—that is, the distance between the molecules. (B) When a force is applied the free energy profile is tilted downwards by an amount equal to the mechanical work Fdx that is done on the system (represented by the straight tilted line), where F is the stretching force and dx is the elongation along the force vector. (C) The extent of the reduction of the height of a barrier is greater the further is the location x_b from the equilibrium configuration. (D) A force vs. loading rate logarithmic plot of the most probable unfolding force of angiostatin Kringle modules upon a mechanical stress (from [163]). The plot shows a linear dependence between the force the loading rate kv (where k is the elastic constant of the cantilever and v the loading rate).

for a sufficient amount of time we will observe it to break. Whenever an external force f that breaks apart the bound components is applied, the bond lifetime under that force is described by :

$$t_{off}(f) = \frac{1}{k_{off}(f)} = t_D \exp \left[\frac{E_b(f)}{k_b T} \right] \quad (6.2)$$

where k_{off} is the diffusive relaxation time and $E_b(f)$ is the height of the energy barrier that confines the bound from the dissociated state.

The pulling force tilts the energy landscape (Figure 6.6.2 B). Evans and Ritchie [165] were the first to model the change induced by an externally applied force on the energy profile of a bond, starting from a pioneering study by Bell [189], and on this basis they analysed in detail the physical quantities that, as we will see below, can be measured in force spectroscopy experiments. The energy barrier at the transition state is reduced by the application of the external force, the dissociated state is made more favoured, and therefore the lifetime of the bond decreases. The application of an external force properly directed thus supplies some encouragement to the binding to fail. The energy distortion shown in Figure 6.6.2 quantitatively changes the bond lifetime in an exponential way, as described by Equation 6.2, where the height of the barrier is reduced of an amount corresponding to the work performed along the reaction coordinate by the force f along a distance x_b (the distance of the barrier from the minimum).

$$t_{off}(f) = \frac{1}{k_{off}(f)} = t_D \exp \left[\frac{E_b(f)}{k_b T} \right] = t_D \exp \left[\frac{E_b(0) - f x_b}{k_b T} \right] = t_{off} \exp \left[-\frac{f x_b}{k_b T} \right] \quad (6.3)$$

The overall effect is a downward tilting of the energy landscape. The extent of the reduction of the height of the barrier is greater the further the location x_b from the equilibrium configuration (Figure 6.6.2 C).

To relate the bond lifetime to the measured values of the force at which a bond fails, one has to consider the dissociation equation of an isolated pair of interacting molecules. The master equation is

$$\frac{dp_1}{dt} = -k_{off}(t) p_1(t) + k_{on}(t) p_0(t) \quad (6.4)$$

where $p_1(t)$ is the likelihood of being in the bound state and $p_0(t) = 1 - p_1(t)$ is the likelihood of being dissociated. k_{off} and k_{on} are the dissociation and binding

constants of the complex. This equation, when a force is applied as to separate the bound components, reduces to

$$\frac{dp_1}{dt} = -k_{off}(t) p_1(t) \quad (6.5)$$

because, keeping on pulling, they move apart and thus the rebinding probability vanishes.

When an elastic probe (like for example the AFM cantilever) acts as a pulling spring at constant speed, the applied force usually increases with time. The relationship between dynamics and mechanics is better described by the loading rate $r_f(f) = \frac{df}{dt} = k_s(f) v_s$, where k_s is the spring constant of the full system composed of the force sensor and the tethered molecules and v_s is its pulling speed. By using this parameter the force becomes the integration variable, instead of the time. By introducing the force dependent expression of $k_{off}(f)$ in the master equation and solving the differential equation thus obtained, the probability density of the bond survival as a function of the applied force, *i.e.* the distribution of the unbinding forces, is obtained.

$$\omega(f) = \frac{dp_1(f)}{df} = \frac{k_{off}(0)}{r_f} \exp \left[\frac{f x_b}{k_b T} + \frac{k_{off}(0) k_b T}{r_f x_b} \left(1 - \exp \left(\frac{f x_b}{k_b T} \right) \right) \right] \quad (6.6)$$

It is then possible to calculate the value of the force at which this distribution has a maximum. This is the most probable unfolding force measured in the SMFS experiment:

$$f(r) = \omega(f) = \frac{k_b T}{x_b} \ln \left(\frac{r_f}{k_{off}(0)} \frac{x_b}{k_b T} \right) \quad (6.7)$$

6.7 Mapping the mechanical unfolding pathway

6.7.1 Dynamic SMFS

Dynamic SMFS is the most widespread and straightforward technique (albeit time consuming) to obtain energy landscape parameters from force spectroscopy data. The technique simply requires to obtain force curves at different loading rates and finding the relationship between the most probable unfolding force and loading rate.

The most probable rupture force (f^*) of the interaction depends upon the loading rate (r_f) because the likelihood of the bond failing under an external force depends on the time spent at the same force. Under the application of a ramp of force the time spent by the bond at a certain force is controlled by the pulling velocity. If the velocity is so high that this time is shorter than the bond lifetime at the same force, the bond is likely to fail only at a greater force. The value of the most probable rupture force therefore increases with the pulling velocity, *i.e.* with the loading rate. This dependence is expected to be a weak one, like the logarithmic one, (see Equation 6.7) because it comes out from the balance of two effects. On pulling faster, the most probable rupture force becomes greater, but at the same time, the dominant barrier is lowered, thus leading to a decrease of the bond lifetime and also of the most probable force at which the rupture takes place.

The force-induced reduction of the height of the barrier being greater the further is its location along the reaction coordinate (see Figure 6.6.2 C), as predicted by Equation 6.7, a linear plot of the most-probable-rupture-force versus the logarithm of the loading rate is therefore obtained (see Figure 2D). This dependence makes it possible to determine not only the barrier location (x_b), but also the unbinding constant at zero force $k_{off}(0)$. The spatial location of the barrier (x_b) is given by the slope of that linear relationship, and therefore if k_s is assumed to be constant, only knowledge of v is needed. On the other hand $k_{off}(0)$ depends both upon the slope and the constant term. This is the reason why an accurate evaluation of the effective loading rate $ksvs$ is needed. The velocity v is set within the dynamic range accessible to the instrument and is known with very high accuracy. The value of k_s was assumed by many authors to be determined by the spring constant of the cantilever only. Most recently it has become understood that k_s should include also the elasticity of the stretched molecule. On the other hand, the elasticity of the molecule depends on the conformational space it can explore. This space is gradually narrowed on increasing the extent of the stress applied to the molecule. This is the reason why Dettmann et al. [190] proposed to evaluate k_s values from the slope of the final part of the force peaks, *i.e.* when the molecule is in a stretched state that is very close to the bond rupture point.

Unfortunately, in most cases, noise and irregularities of the peak profile often make somewhat arbitrary the definition of a meaningful “slope” . This problem

can be circumvented by obtaining k_{off} and x_b from Monte Carlo simulations of the mechanical unfolding. The simulations normally assume a two-state model for the unfolding potential barrier and the WLC elasticity model for protein elongation. The energy parameters of the barrier are adjusted to fit the experimental force distribution and the dependence of the rupture force data from the pulling speed. In this model the unfolding probability of any module at each step is

$$P = Nk_{off}\Delta t \quad (6.8)$$

where N is the number of unfolded modules and Δt is the time interval. With this method the loading rate is implicit in the calculation, being a function of cantilever elastic constant and of molecule p : these parameters must be given to the simulation as explicit data [164]. If there is some knowledge about the interactions involved in mechanical unfolding, more refined models can be used [191].

Evaluations of not only the intrinsic rate coefficient and location of transition state, but also of the free energy of activation, can be extracted by taking into account that the tilting of the energy landscape not only decreases the height of the barriers, but also subtly shifts them towards the equilibrium minima (an example of Hammond effect)[192]. Because of this shift, the mean rupture force becomes a non linear function of the logarithm of the loading rate. Nevertheless at low to intermediate speeds a linear fit of this model is quite good. Dudko et al. [192] therefore warn that, in this case, one might incorrectly attribute the curvature appearing at higher force loading rates to the switch from one dominant free-energy barrier to another.

6.7.2 Protein engineering analysis

Another way to understand protein mechanical unfolding pathways is perturbing the system by mutations, and observing the result. In this way, unique insights can be obtained. For example, several mutants of the I27 titin domain have been studied. The insertion of glycine loops into I27 modules was used to show how SMFS could capture length differences down to the single aminoacid[168]. Mutants containing proline were employed to demonstrate mechanical unfolding intermediates[162] and to study the effect of mutations on protein mechanical

stability[193]. The existence of the I27 unfolding intermediate was further characterized by mean of substitutions and deletions[178]. Best et. al. developed the mechanical equivalent of the Φ value analysis[194], used to probe the conformational effect of a mutation by measuring the ratio:

$$\Phi = \frac{\Delta\Delta G_{N-\ddagger}}{\Delta\Delta G_{N-D}} \quad (6.9)$$

that is, the ratio between the native-transition energy gap and the native-denatured energy gap. This is a measure of the amount of native structure that is present in the neighbourhood the mutated residue in the transition state. This kind of analysis however assumes Δx_u to stay constant between the wild-type and mutant protein.

6.7.3 Molecular dynamics simulations of mechanical unfolding

SMFS itself can only know of interactions, their forces and the lengths of loops enclosed by them. As such, they cannot provide information at the atomic level on the nature of the probed interactions[35]. As of today, most structural insights on the atomic behaviour of protein chains under an external stress come from steered molecular dynamics (SMD) simulations[195]. SMD can reliably predict the relative mechanical stability of homologous modules and the location of force-bearing residues [196].

Despite useful, SMD simulations suffer of onerous computational requirements. This leads to several limitations in what can be obtained by them. The time scale available to SMD is often of the order of a few nanoseconds. Conversely, the required pulling speed of the simulation must be orders of magnitude faster than the fastest available dynamic range of AFM (which works in the millisecond to second range). The physical correspondence between the simulation and the experimental results is thus unclear. Moreover, simulations conducted in explicit solvent often cannot simulate correctly the water molecules while the molecule extends: the bubble of water surrounding the molecule tends to “evaporate” , and simulations of the latest unfolding phases correspondingly suffer of artefacts.

In spite of these difficulties, experiments and simulations usually agree strikingly with each other, their combination being the approach of choice for dissecting

the mechanical unfolding pathways at atomic level.

Recently, it has been shown that coarse-grained simulations, which mimic the system at the aminoacid level rather than at the atomic level, can provide surprisingly good results[184] in predicting the general features of unfolding pathways and relative force magnitude of protein folds. Being computationally much lighter, they could explore time scales much more similar to the experimental ones and help the experimenter to know what to look for. This permitted to survey known protein folds in the Protein Data Bank, obtaining a proteome-wide scan of mechanical behaviour of proteins[179].

6.8 SMFS data analysis

While much has been written on the experimental, “hands on” side of force spectroscopy, much less attention has been dedicated to the aspect of data analysis (see [197] for an exception). This lack of information is even more visible considering the fact that SMFS data analysis is all but standardized, and that while basic practical skills inherent to SMFS can be learned in less than a month, interpretation of force curves is often the most bewildering and difficult stage for the SMFS experimenter, more so when analyzing a new protein. Force curves, in fact, are rarely if even close to ideal. Force curves are intrinsically different from many spectroscopies in not relying only on the absolute positions of their signals, but mostly on the relationship between its features.

In fact, only about 1% of force curves usually contains data good enough to be used for interpretation. This means that 99% of curves has to be discarded. Most of these curves (about 70-90%) will simply contain no signal at all: this is actually a *desired* feature, since a higher rate of success implies a higher probability of picking up two or more molecules instead of only one. The remaining ones will contain force curves from the stretching of two or more molecules, contaminants, fragments or simply interactions and objects that resist interpretation.

6.8.1 Recognizing aspecific interactions

The first part of SMFS force curves is often obscured by aspecific interactions (see 5.3.1). Moreover, small fragments or contaminants being casually stretched

between the tip and the surface will appear as short, irregular peaks in the initial part of the curve. Such signals are at risk of being confused with genuine signals coming from the molecule into study. In general, interactions should not be measured unless they fall clearly outside this blind window. A rule of thumb is: if a peak is separated by aspecific signals by a significant number of points near the baseline, meaning the system has relaxed, it can be took into consideration.

A second control can be made on lengths. If the interaction comes from the molecule and not from aspecific interaction between the molecule and the surface, the contour length measured on the peak will be next to that of the full molecule. If the measured contour length is unreasonably shorter, the interaction is suspicious.

6.8.2 Recognizing single molecule events

The force spectroscopy “blind fishing” strategy of course does not guarantees that only single molecules will bridge the tip and the surface. Two, three or more protein molecules can easily fit on the large tip surface. The significance of AFM as a single molecule technique relies not on the absolute certainty of picking up only a single molecule, but on the ability to recognize the signal of single molecules from that of multiple molecules.

In principle, signals from multiple molecules contain the same information of multiple signals from a single molecule. In practice however the signals add up and make impossible to assign features to one of the two molecules or both. Therefore only signals from single molecules must be selected. In fact, signals from multiple molecules are a common plague of SMFS experiments. To be avoided, usually the concentration of the molecule on the surface has to be balanced so that actually picking molecules becomes a *rare* event (about 10-1% of all force curves). This way, Poisson statistics guarantees that a large proportion of these events will come from single molecules.

Recognizing single molecule events depends on the features of the molecule itself. In general, however, a very strong suspicious marker of multiple events is a *reverse hierarchy* of forces. When stretching a single object weak bonds fail before strong ones, so usually force peaks are distributed approximately in a crescent hierarchy. When multiple molecules are picked up, instead, shorter molecule segments will resist force before longer ones are fully stretched. Therefore,

larger peaks from the detachment of shorter molecules can easily precede smaller peaks from the unfolding of longer ones.

In the case of polyprotein, another tell-tale pattern is the presence of a “staggered” sawtooth pattern, where the step between two peaks is minor than the length of the domains, but the step between *couples* of peaks is equal to this length. This is usually the fingerprint of two polyprotein molecules being caught and unfolding together.

6.8.3 Recognizing objects

SMFS force curves are exceptionally poorly informative, by themselves, on the nature of the molecule being stretched. SMFS by itself gives no data on the molecular structure, but only on the force of intramolecular interactions and of the lengths of loops enclosed by them; moreover it is often hard to have a priori data on the intramolecular interactions that can be probed mechanically.

For this reason, having a thoroughly characterized marker is of immense help. The tell-tale pattern of marker modules in a polyprotein construct (see 6.4.2) , with its characteristic spacing between peaks of reasonable size, allows to recognize curves containing the signal of the analyte from those due to possible contaminants. Of course, the pattern alone is not enough. Curves with similar patterns can come from dimers, from partially unfolded molecules (due to protein-surface interactions), from fragments and contaminants. Not being able to characterize exactly what molecule has been stretched means, simply, that we do not know what we have stretched and therefore we are unable to do any inference.

For this reason, appropriate controls must be made. Since length is the only structurally constant variable that can be safely probed by mean of SMFS, usually such controls are made on length. In particular, the following lengths have to be controlled, since they are best known *a priori*:

1. *The initial contour length.* This parameter gives the contour length of the object *before* it is being stretched.
2. *The final contour length,* measured on the detachment peak. This parameter gives the length of the whole unraveled object.

It is important that the two lengths are consistent. Picking up a shorter section of a polyprotein, say, 4 modules out of 6, means that (1)four peaks will be observed (2)the final contour length will be equal to 4 modules unraveled (3)also the initial contour length must correspond to that of 4 folded modules.

Applying such simple controls on the choice of the signal to include in the statistics often means to surprisingly increase the value of the data measured.

6.8.4 Recognizing patterns

Of course if everything on SMFS curves was expected *a priori*, SMFS experiments would be of little value. Recognizing unexpected patterns is therefore a crucial (and difficult) skill: but it is the most important, since it allows to discover new conformations otherwise difficult if not impossible to observe with other techniques.

In general, every pattern that (1)repeats regularly between different force curves and (2)belongs to interpretable, recognizable objects should be investigated. Such patterns can range from whole new signals appearing on the force curve to subtle differences in the shape of the peaks.

Usually analyzing force curves requires hundreds, if not thousands, of force curves to be reviewed. In such a sea of data, of which an overwhelming majority is discarded, a rarely occurring but repeating pattern can easily be lost. For this reason a conservative and iterative approach is best. Data are therefore analyzed two times: first, the dataset is cleaned from obviously useless signals, but keeping generously “suspicious” or “dubious” signals together with the clearly “good” ones. These data (usually already less than 10% of the initial dataset) are then kept apart. After a significant number of these curves has been chosen, a second review is made on these, now trying to apply as rigorous as possible criteria for their classification. Both in the first and in the second review, notes are taken on curves that are selected, and in the second *also* on curves that have not been selected. Curves not selected in the second review are not thrown away: they can still contain material and information that will be of interest later, or that contains significant information that will become apparent only after successive rounds of data analysis. It is also good practice to mix review rounds between operators, and/or to do that collectively, so to reduce subjective bias.

6.8.5 Data analysis automation

It is obvious that in SMFS analysis as described there is space for considerable operator bias. A further, important step for SMFS would be an at least partial automation of data analysis. Unfortunately the impressive instrumental achievements of SMFS have not been joined by a parallel development of data analysis software. Most disappointingly, no prominent software platforms emerged as standard mainstream in the SMFS community (most laboratories use home-made data analysis scripts), thus making efforts many efforts in this direction failing to be widely adopted.

As of today, most efforts on data analysis automation have been focused on automatic elimination of the large percentage of curves recognition of unfolding peaks and automatic measurement of peak force and distance[198][199]. Although very useful in reducing the time needed for data analysis, such approaches still do not help in lending rigour to the criteria used for selecting curves to measure and to classify different patterns.

A possible step in this direction could be applying multivariate analysis on a large set of variables describing the force curves (for example, average distance between peaks, average peak force, folded contour length etc.). Even a simple cluster analysis made on the correlation of different variables could help in quantify, and thus rationalize, the criteria used for selecting and classifying force curves.

6.9 Force clamp SMFS of multimodular proteins

Force clamp SMFS (FC-SMFS) is the other main AFM-based SMFS technique. As the name says, whereas CS-SMFS keeps speed constant, FC-SMFS keeps *force* acting on the molecule constant. The evolution of the system under constant force is then followed in time. The output of FC-SMFS is a time *vs* length graph, where the length is the surface-tip distance. In the case of multimodular proteins, the classical plot shows a characteristic “staircase” where each step marks the unfolding of a domain: the height of a step is, approximately, the length of the module (just like the distance between peaks in CS-SMFS)⁶ and the length of the

⁶In practice, the height of a step depends on the clamp force, following the WLC function. At low forces the unfolded domain will not be fully stretched, therefore a domain will appear

step is the time before the next unfolding. Since this time is probabilistic, the only clear parameter identifying the protein domain here is the length[200].

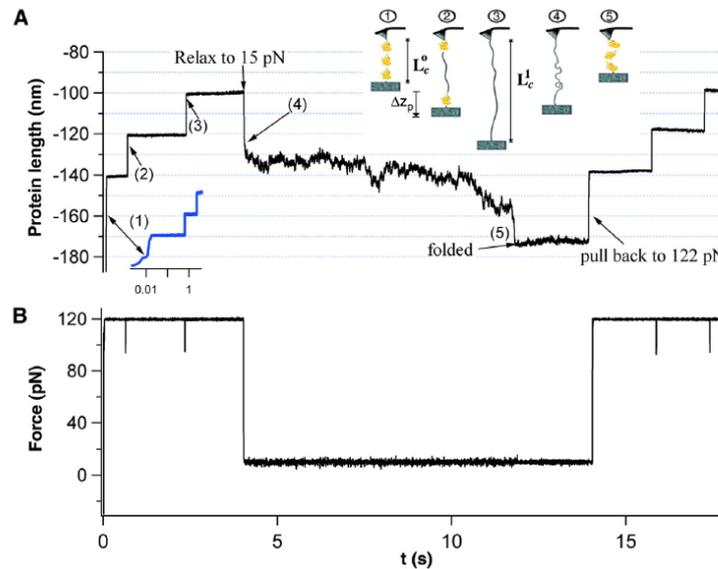


Figure 6.10: Forced unfolding, refolding and unfolding again of a ubiquitin polyprotein by FC-SMFS. (A) Plot of the chain extension as a function of time, (B) corresponding plot of the applied force. The external applied force is first set at a high force (122 pN) to encourage the folded protein to unfold (steps 1-3 correspond to the unfolding of three ubiquitin domains) then released to a lower force (15 pN) to increase the likelihood of the refolding, and finally set to 122 pN again. If in this last phase unfolding signals appear again, this means that refolding was successful. It is possible to see different stages in the refolding trace: the fast elastic recoil of the unfolded polymer (4), and then a continuous but fluctuating collapse as the protein folds (from 4 to 5). At 14 s the polyprotein was again stretched back to 122 pN. A stepwise pattern analogous to that at the beginning is present, confirming that the polyprotein modules had indeed folded again. The initial steplike extension is the elastic stretching of the folded polyubiquitin. The inset line above the time scale shows the different refolding stages. Figure edited from [201].

FC-SMFS has enormous potential, since it allows to obtain kinetic parameters of proteins under force directly, without having to resort to dynamic force spectroscopy or other workarounds. Moreover it shows promise of identifying unfolding intermediates with finer resolution than conventional CS-SMFS: for example, while shorter than at high forces.

the I27 unfolding intermediate on CS-SMFS resolves only as a smooth hump when multiplied for a significant number of domains (see Figure 6.3), it can, at least in theory, show as a well-resolved step if the clamp is set at enough low force. However FC-SMFS is still suffering technological limits, because of the unavoidable cantilever drift and the force feedback noise (about 50 pN); the feedback speed is also limited, in best cases, to approximately 3 ms, which is barely enough to obtain clear-cut unfolding steps. Further improvement in the experimental set up through the employment of smaller force probe[202] [203] in order to strongly decreasing the noise level and/or through faster feedback controls are being developed.

Most strikingly, the possibility of quenching the force at any chosen value allowed observation of the refolding of these multimodular proteins under different mechanical loads[201] (see Figure 6.9). The time taken to fold was shown to be dependent on the contour length of the unfolded protein and the stretching force applied during folding. The folding collapse was marked by large fluctuations in the end-to-end length of the protein, but these fluctuations vanished upon the final folding contraction. Afterwards, the force was raised again in order to control the extent of the refolding. The interpretation of these results was done in the light of a polymer coil-to-globule transition, where the chain starts from an almost fully extended state and then collapse in a compact form as it would do a polymer placed in a poor solvent. In any case the recorded refolding trajectories are presently not clear enough to understand each detail of the folding of a complex object such as a multimodular protein construct. Recent theoretical considerations have shown that protein refolding in this kind of force-quenching experimental set up may not be directly comparable to temperature-quenching protein refolding: the initial state[204] and also the reduced degrees of freedom[205] affect the preferred refolding pathways. More recently, the group of J.Fernandez showed possible evidence of force-clamp followed refolding on a real single protein, instead than on a multimodular construct[159].

6.10 Limits and perspectives

6.10.1 SMFS as a structural technique

SMFS experiments are still far from being a structural technique: they can detect structural differences in a sample of molecules, and large (>100 pN) force peaks are highly suspect of belonging to beta-structured modules, but can do little more. The most detailed analysis achieved so far by mean of SMFS is the sequential pulling of α -helices from monolayers membrane proteins[206]. Therefore, molecular dynamics simulations or time-consuming mutational approaches are needed to gather structural information on mechanical unfolding (see 6.7).

Two routes are currently in development to overcome these limitations. The first is acquiring a better force resolution, as to be able to solve the unraveling of single hydrogen bonds and, ultimately, the secondary structure of the protein into study. Currently the only limit to force resolution is cantilever thermal fluctuation due to thermal noise. A reduction of the size of the force sensor, keeping the same stiffness, should improve the signal-to-noise ratio[207]. Small cantilevers have been tested and allow a resolution of about 7-10 pN, two to three times better than that achieved with standard cantilevers [203][155]. Usage of these cantilevers is however still extremely limited in the SMFS community due to them still being scarcely available commercially, their high cost and technical difficulties (optical interference, fragility).

The second route is that of coupling optical and force measurements on a single molecule. By using appropriate FRET⁷ probes and optical equipment, it could be possible to see what residues are being pulled apart mechanically during a SMFS experiment. The titanic technological challenges still prevent this to be a commonplace approach, but the first attempts have been made[208]. It is not improbable that in the next few years such instrumentation will be put in commerce, thus boosting the detailed molecular comprehension of mechanobiological processes.

6.10.2 Increasing data throughput

SMFS experiments can be extremely time consuming, especially if a single module in a “sandwich” construct is being probed. To build a decent statistical sample,

⁷Fluorescence Resonance Energy Transfer

often weeks of experiments are needed. Speeding up the process would make it easier for SMFS to become a “routine” analytical technique.

The first approach is increasing the raw data throughput by using multiple tips, like in the “tip arrays” [209] and the “Millipede” [210] devices. This method has the advantage of allowing for “reference cantilever” to be used, effectively removing problems due to mechanical noise and thermal drifts.

The second approach is transforming SMFS from a blind fishing expedition to scope-aided hunting. Today there is little control on the general state of the protein on the surface, and none at all on the actual molecules picked up. The heterogeneous mixture on the surface contains fragments, denatured, misfolded, aggregated proteins that at best are a distraction and at worst degrade the quality of the statistical sample. Imaging of the sample, allowing to pick up only “good” molecules leaving behind “bad” ones, would be of exceptional help in increasing both the data throughput and the confidence in data. However, proteins are soft materials that are difficult to reliably image with conventional AFM imaging modes, and therefore, with the exception of systems like 2D crystals[206] it is unclear how this could be achieved with conventional protein deposition. Using self-assembled monolayers as a platform, however, makes it possible to obtain better control on the sparseness and attachment of proteins on the surface, and therefore this method shows some promise about improving singling out single proteins on the surface.

CHAPTER 6: SMFS INVESTIGATION OF MULTIMODULAR PROTEINS

Part III

Hierarchical mechanochemical switches in angiostatin

*The complexity of things - the things within things - just seems to be endless.
I mean nothing is easy, nothing is simple.*
Alice Ann Munro

Chapter 7

Background and theoretical elaboration

7.1 Disulfide bonds as molecular switches

Disulfide bonds are commonly thought to have been selected by evolution to serve one main purpose: to sustain and protect the native conformation of a protein[211][212]. In fact a disulfide bond is rare in intracellular proteins but is a common feature of proteins that work in the extracellular space, i.e. a space that offers a particularly challenging environment for protein folding. The disulfide bonds reduce the conformational space accessible to the native folded structure of the proteins, by linking adjacent strands with a covalent bond (Figure 7.1 A). The disulfide bonds can even form a sort of ladder in a protein structure that must be maintained in a particularly harsh environment. That is for instance the case of laminin γ -chain, a structural protein of the basement membrane[213] (Figure 7.1 B).

This point of view looks at the disulfide bond just as an inert structural feature. On the other hand the disulfide bond can be reversibly cleaved and reformed by various physiological agents. It is therefore natural to imagine that the disulfide bond can also act as a reversible switch able to be turned off and on. In fact, several protein functionalities have been already proposed to depend on switching mechanisms based on disulfide bonds.

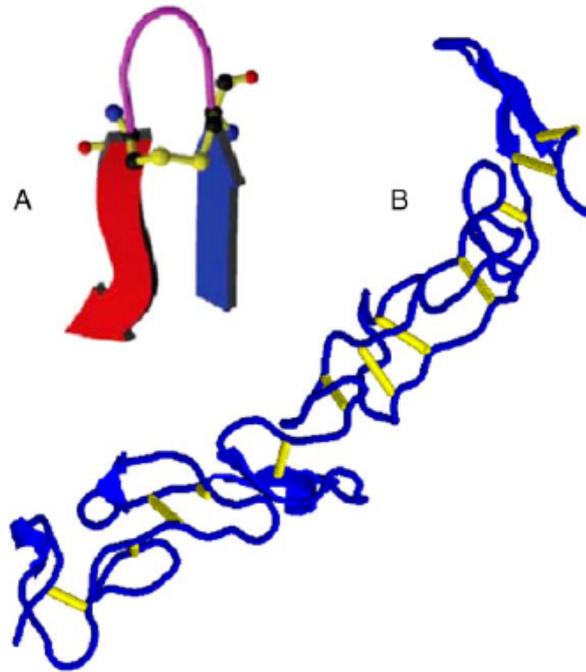


Figure 7.1: Disulfide bonds protect the protein folding. (A) Cross-strand disulfides connecting a beta hairpin in the influenza virus B neuraminidase (from[214]). (B) The structure of the laminin gamma chain (1 KLO in the Protein Data Bank) is an example of a structure with a high concentration of disulfide bonds. The fully oxidized structure is a rigid rod kept together by a ladder of 12 disulfide bonds.

7.1.1 Disulfide bond cleavage as a mechanism for protein function control

Extracellular regulation by disulfide cleavage has been specifically demonstrated in secreted proteins like thrombospondin-1[215], von Willebrand factor (vWF)[216] and plasmin[217], in cell-surface receptors including CD-4 T cell receptor[218], integrins[219] and the HIV gp120[220]. The functional effects of this regulation are different and not always clear.

Disulfide regulation directly affects the multimer size of vWF[216] and triggers the auto proteolysis process that leads to the production of angiostatin from plasmin[217]. In the case of the CD-4 receptor both oxidised and reduced forms of the molecule co-exist in equilibrium on the T-cell surface, and T-cell activation leads to a shift to the reduced form. This suggests a definite but still unknown

functional role of the disulfide reduction in the CD-4 receptor[218] and indicate that the reversible cleavage of disulfide bonds in extracellular proteins may be an important tool for the regulation of their function. CD-4 disulfide reduction has been shown to block HIV infection[221].

In all these cases, except in that of plasmin, disulfide cleavage is triggered by the action of unspecific oxidoreductases of the protein disulfide isomerase (PDI) superfamily like thioredoxin or PDI itself. There is a general evidence that free thiols are exposed on the extracellular face of many plasma membrane proteins and that PDI controls the exofacial thiol/disulfide equilibrium[222]. Thioredoxin and PDI are known to be secreted outside the cytoplasm and to be present and active on the cell surface, despite the lack of obvious secretory signal sequences for both proteins[223][224][225]. It is still unclear if these proteins act as single turnover reductants or if they act in a catalytic-like manner: in the latter case a PDI- or thioredoxin-reductase system fuelled by NAD(P)H, for example, must be present. Evidence for a working extracellular thioredoxin/ thioredoxin-reductase system has been found[226].

The main plasmin reductase has been demonstrated instead, quite surprisingly, to be phosphoglycerate kinase (PGK), the sixth enzyme of the glycolytic pathway[227]. PGK appears to act with a still unknown disulfide independent mechanism unrelated to that of oxidoreductases of the PDI superfamily. There is a definite functional difference between dithiol-disulfide redox exchange and thiol independent mechanism for disulfide reduction. The first class of proteins have broad specificity and require at least two additional factors to act catalitically: this means they are unlikely to act in a specific and efficient fashion. The second mechanism, although still unclear, can act both catalytically and with high specificity.

Disulfide bonds can also be enzymatically reformed in the extracellular matrix. Sulphydryl oxidases are disulfide bonding catalysts that seem to act in the extracellular matrix along with PDI[228].

7.2 Coupling the redox equilibrium of a disulfide bond with an external mechanical force

The cleavage of a disulfide bond therefore seems a fast, specific and reversible switch for protein functions. However, in all the cases known in which a biochemical signal is triggered by disulfide bond cleavage, it is still unclear which is the structural mechanism that underlies this kind of regulation. The common explanation postulates that the cleavage of the disulfide bonds itself can trigger a conformational switch and therefore directly influence the protein function. This purely redox switching mechanism, although it can directly affect protein flexibility, is not expected to be very effective in altering the protein conformation, and it is unclear how widespread it may be[229].

In some cases disulfide bonds somehow trap an otherwise frustrated fold, that relaxes in the minimal energy state only once the disulfide bond is broken. This can be the case of disulfide bonds that have high potential energy stored: examples are cross-strand disulfides[214] and vicinal disulfide turns[230]. However, the cleavage of most extracellular disulfide bonds is known to be reversible[229] and the existence of highly strained disulfide bonds is not expected to be general. After the cleavage of the disulfide bond, it should be difficult for a protein to “come back” and reform spontaneously a disulfide bond that stores high potential energy. Even if there are rare cases of reversible conformational switches triggered by disulfide bonds[231], the general rule is that disulfide bonds do not drive the folding, but merely stabilize protein structures[232].

Therefore, after the cleavage of a disulfide bond it might be desirable to “help” the system to drive the protein to the specific structural change that is required to activate a specific functionality. This help might have a mechanical origin. In the course of my Ph.D. work, a theoretical framework has been elaborated on the possible relationship between disulfide redox regulation and mechanochemical regulation.

7.2.1 Mechanical effects of disulfide bonds

An effective functional switch should be fast, specific, and reliable. A mechanical switch or a redox switch based on a disulfide bond, taken alone, fulfil only partially

these requirements.

We have seen that just the opening of a disulfide bond can hardly lead to the relevant conformational changes that are expected to be required to activate specific functionalities. We have seen also that mechanical processes are able, instead, to deeply affect the structure and, therefore, the functionality of an extracellular protein (see Chapter 3). What these latter processes lack is a reliable control of the induced molecular extension.

Disulfide bonds create topological loops that drastically affect the mechanical behaviour of a protein molecule under external force, by covalently connecting cysteines that are distant in the protein primary sequence. Being covalent, the disulfide bonds can withstand forces up to a few nanonewton[233]. The stretching forces normally generated *in vivo* have been estimated to range from one to a few tens of piconewton[31][33]. These forces would not be enough to break a disulfide bond. The action of the force will therefore bypass the loop enclosed by the disulfide bond (Figure 6.7): the contour length measured in a single molecule force spectrum thus results apparently shorter than that expected simply on the basis of the total number of its aminoacid residues (see also Section 6.5.3).

An experimental evidence of this effect has been given by the human angiostatin K1-5 molecule[186]. Angiostatin is a protein composed of up to five very similar modules, called kringle domains. Each kringle domain has a globular structure defined by three internal disulfide bonds, forming a characteristic triple-loop topology[234] (see Section 10 below). The most external bond of each domain encloses practically all the module.

In the fully oxidized configuration angiostatin behaves as a short, inextensible molecule with an apparent contour length of a few nanometers, contrary to a contour length of the order of 150 nm, in the fully reduced form. By chemically reducing the disulfide bonds the internal topological loops of the kringle domains become accessible to the action of the force and can mechanically unfold. The more the reduction proceeds, the larger is the portion of each module that can unfold under force. The presence or the absence of disulfide bonds therefore modulates the mechanical properties of the polymer among states with different extensibility. In the case of angiostatin, the triple loop topology of kringle domains makes three different mechanical states to be accessible in each domain.

An analogous but simpler system has been described for the Ig-like domains of

melanoma cell-adhesion molecule (Mel-CAM)[185].

7.2.2 The redox modulation of the protein mechanics can switch new protein functionalities

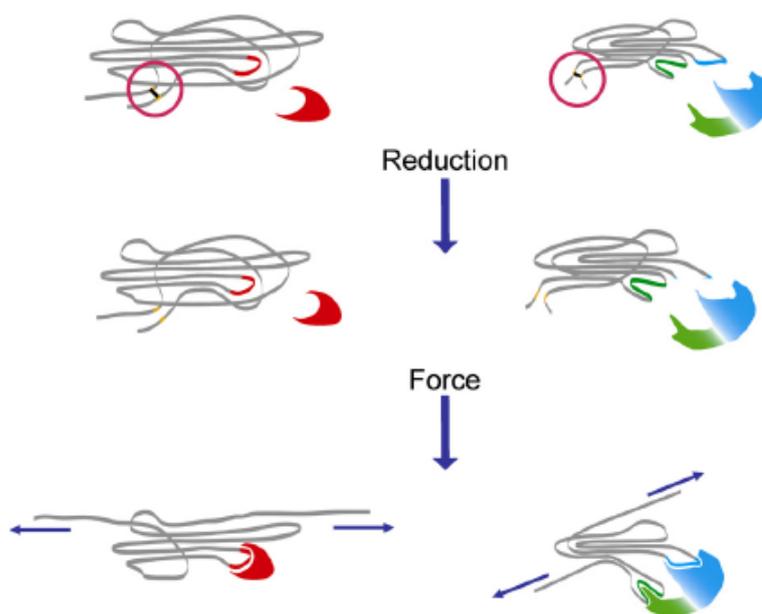


Figure 7.2: The coupling of cysteine redox and mechanical regulation. Left: If a cryptic site is enclosed in a loop defined by a disulfide bond, mechanical force alone is not able to break the constrain and free the site. Only after the reduction of the disulfide bond, the protein fold can be mechanically unraveled and the cryptic site revealed, thus allowing binding. Right: The same logic applies for two synergic binding sites that must be placed at the correct distance to simultaneously bind their target. Also in this case reduction of the disulfide bond makes it possible to deformate the protein fold as to allow the correct displacement of the synergic binding sites.

The mechanisms proposed in 3 by which a mechanical stress can be transduced into a biochemical signal, can meet a finer tuning whenever they are coupled to a redox equilibrium based on disulfide bonds. In fact, in the absence of disulfide bonds, potential cryptic sites are exposed each time there is a tensile stress on the protein. On the other hand, if the cryptic site is enclosed in a loop defined by a

disulfide bond, the site will be hidden by default, and it would be exposed to a mechanical stress only after having unlocked the disulfide bond.

A signalling pathway on the cell surface can be therefore based on the over-expression and/or activation of extracellular disulfide reductases, that can finely control the exposure of cryptic sites in mechanically stressed proteins (Figure 7.2).

7.2.3 Candidate proteins

The structure of laminin strongly suggests the possibility that such a regulation mechanism might act in the basal membrane of vascular endothelium. The basal membrane is the specialized extracellular matrix that sustains the growth and the survival of vascular endothelium [213] and it is normally subjected to the mechanical forces generated by migrating endothelial cells. Laminin is a trimeric protein and is one of the main structural components of the basal membrane. The structure of laminin shows a striking “ladder” of disulfide bonds in the γ chain [235], and also three exposed disulfide bonds in the α -2 chain [236] (Figure 7.1 A).

Thioredoxin has been shown to be able to reduce laminin disulfide bonds and this reduction seems to alter the growth of endothelial cells [237]. Moreover, the binding sites for nidogen (another essential component of the basement membrane) located on the III4 domain of the laminin gamma chain geometrically match with disulfide-connected loops in laminin. These binding sites are at least partially less active when reduced [238], suggesting a mechanism in which binding affinity could be locally regulated by a redox equilibrium.

Cell adhesion molecules and transmembrane molecules are normally subjected to the mechanical stresses generated along the ECM-cytoskeleton pathway. Many of these molecules also present disulfide bonds in their structure and constitute therefore other candidates for a disulfide-controlled regulation mechanism. It has been proved that the CD4 receptor activity is regulated by disulfide reduction [221], and that the mechanical properties of a cell adhesion molecule (CAM) can be modulated by mean of its disulfide bonds [185]. The vascular cell adhesion molecules (VCAMs) and CD2 [239] are two other examples of the many cell adhesion molecules that contain disulfide bonds in their structures.

Integrins are the transmembrane molecules that mediate cell adhesion. Also

these molecules contain disulfide-bonded modules and their reduction has been shown to lead to integrin activation[219].

7.2.4 Mechanoredox controls in the extracellular matrix

The majority of structural proteins present in the various kinds of ECM contain disulfide bonds. In ECM, as any other polymer gel, the degree of crosslinking strongly influences the gel mechanical behaviour. There is evidence that at least some ECM components are crosslinked by means of disulfide bonds[240]. Given also the presence of reductases and disulfide isomerases in the extracellular space we must expect that a disulfide bond mediated crosslinking can be a quite diffused phenomenon.

7.3 Angiostatin

Angiostatin is an internal fragment of plasminogen, and it has been discovered through its antitumoral and antiangiogenic action in mice[241]. Three isoforms of angiostatin are known, depending on the plasminogen kringle domains they contain (K1-3, K1-4 and K1-5). Angiostatin has been among the first endogenous angiogenesis inhibitors known, acting by both inhibiting endothelial cell migration[241][242], and eliciting endothelial cell apoptosis[243], thus suppressing tumour growth in animal models[241].

7.3.1 Kringle structure

Kringle domains are compact, rigid folds that are unusual in having virtually no recognizable secondary structure. The only secondary structure is typically an antiparallel β -sheet composed of 2-3 amino acids per strand (Figure 7.3 A).

Kringle domains share a triple loop topology defined by three disulfide bonds¹ (Figure 7.3 A). Each domain is 78-80 aminoacid long. Angiostatin kringle domains show considerable sequence identity, of about 50% [244], that reflects in nearly equal three-dimensional structure for all domains. The linker sequences between

¹The name of the domains in fact comes from the kringle biscuits, which loosely remind the topological diagram

the kringle domains are non-structured and of variable length (from 3 to 30 amino acids).

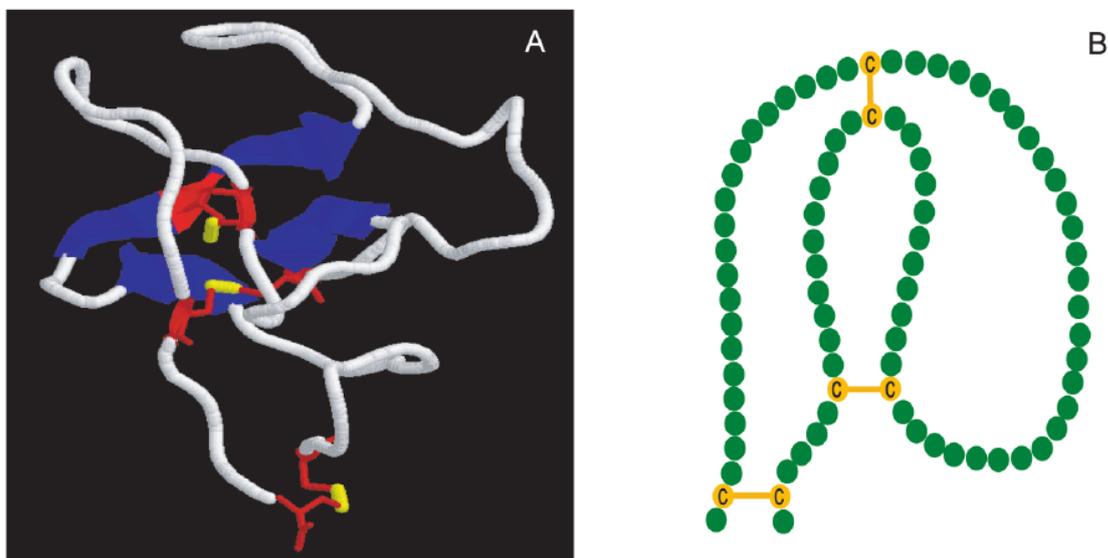


Figure 7.3: Kringle domains structure. (A) Three-dimensional structure of the kringle domain K1 of angiostatin. The β -sheet (blue colored), the disulfide bonds (yellow colored) and the cystein residues (red colored) are evidenced. (B) Sketch of a two-dimensional structure of a kringle domain. The three disulfide bonds (yellow colored) determine the typical triple-loop topology.

The function of kringle domains is usually of protein recognition. Kringle domains contain characteristic lysine binding sites[245], with the notable exception of angiostatin K3 domain.

The only multikringle structure known in atomic detail is that of angiostatin K1-3[234]. In the crystal, the three kringle domains orient themselves to form a triangular bowl-like structure, kept together also by an interkringle disulfide bond between K2 and K3.

7.3.2 Angiostatin production

Angiostatin is produced by step-wise proteolysis of plasminogen[217][246]. First, the N-terminal fragment of plasminogen is cleaved by the plasminogen activators uPA and tPA, or by some metalloelastases. The resulting fragment is known as

plasmin. The reduction of two disulfide bonds of K5 by PGK and the sequential proteolysis of the plasmin protease domain leads to angiostatin K1-4 (plus a fragment of K5). The agent responsible of K5 disulfide reduction has been shown to be phosphoglycerate kinase (PGK)[227]. Other disulfide reductases are however able to cleave K5 disulfide bonds, like thioredoxin and protein disulfide Isomerase (PDI)[217]. The K1-5 fragment can in turn be processed by matrix metalloproteinases resulting in generation of K1-3 and K1-4 isoforms.

7.3.3 Angiostatin proposed mechanisms of action

There is considerable uncertainty on the mechanism of action of angiostatin at the molecular level. Although cell surface receptors for angiostatin have been identified, it is not clear which receptors actually transduce the inhibitory or proapoptotic signals. Also, the intracellular signaling pathway triggered by angiostatin is yet to be characterized.

- Tetranectin and ECM of endothelial cells have been shown to be ligands for angiostatin K1-4[247]. The binding of tetranectin was shown to inhibit the interaction between K1-4 and the ECM, and blocks angiostatin anti-proliferative activity, suggesting that angiostatin action is related to its interaction with the ECM[247].
- Angiostatin can bind the α and β subunits of mitochondrial F1-F0 ATP synthase, inhibiting ATP synthesis. Most surprisingly, the same ATPase is uniquely expressed at the surface of endothelial cells[248]. The inhibition of cell surface ATPase has been proposed to alter pH inside the cell and act as an apoptotic trigger. The binding of K1-5 to the endothelial ATP synthase leads to caspase-mediated endothelial apoptosis. Inhibiting caspase-mediated apoptosis also abolishes K1-5 activity[249]. It is then possible that the specific action of angiostatin on endothelial cells is due to angiostatin recognizing cell surface ATPase.
- Integrin $\alpha v \beta 3$ on the surface of endothelial cells is a ligand of both angiostatin[250] and plasmin[251]. Plasmin is an angiogenic agent, so it has been proposed that angiostatin acts as a negative competitor to plasmin in its binding to integrin.

- Focal adhesion kinase (FAK) is activated by angiostatin, inhibiting endothelial cell migration and promoting endothelial cell apoptosis[243]. FAK activation seems related to the angiomin receptor[252]. Angiomin is able to bind angiostatin on the surface of endothelial cells, despite its sequence not sharing obvious characteristics of a cell surface receptor. Angiomin makes cells able to internalize angiostatin. Angiostatin internalization then leads to FAK activation

7.3.4 Kringle structures activities

Various studies have investigated the activity of single kringle domains and their combinations, to dissect the action of angiostatin modules at molecular level. The studies showed that the activity of multikringle structures is far from being simply the sum of its parts, and it is instead arising from complex synergic and antagonistic effects.

K5 and K1 have been identified to be the most potent domains regarding inhibition of endothelial cell growth, while K4 is virtually inactive[244][253]. The ranking order of endothelial cell growth inhibition is $K5 > K1 > K3 > K2 > K4$.

The antiproliferative activity of K1-4 is boosted by synergic effects, its effect being higher than that of the single domains composing it[244]. K1-3 shows however even more powerful activity. It seems that K4 is not required, or even disadvantageous, for anti-endothelial growth activity.

K5 is especially powerful. It is the most anti-proliferative kringle domain, and by itself it is several times more powerful than K1-4[253]. K1-4 and K5 together acts synergically, and correspondingly the full K1-5 is the most powerful inhibitor of cell growth.

Synergistic and antagonistic effects are evident in kringle domains, their effects being combined as covalently bound dimers or multimers are markedly different even from their effect as combined domains in solution. The K2-3 fragment for example displays a lower activity than the one of K2 and K3 alone. On the contrary, the K1-3 fragment displays higher anti-proliferative activity than the one of K1 and K2-3 alone. When combining the two fragments, however, antagonistic effects appear and the activity of K1+K2-3 is comparable with the one of K2-3.

Synergistic and antagonistic patterns have been found also for antimigratory

activity. Antimigratory and antiproliferative activities seem to be poorly correlated. K4, while being almost inactive against proliferation, is the most active antimigratory domain of angiostatin, both alone or in the K1-4 fragment. K1-3, on the contrary, is a poor antimigratory agent[242]. The K2-3 fragment has antimigratory activity comparable with the one of K4. K1 shows a very low activity and has been found to be responsible of the modest activity of K1-3, which is markedly lower than K2-3 one. The combination of the two individual fragments (K1 and K2-3) leads to an high activity, comparable to the one of K2-3 alone. The latter also shows a higher activity than the combination of the two individual domains K2 and K3. K5 shows an activity comparable to that of K4[254].

7.3.5 Kringle activities and disulfide bonds

Angiostatin K1-4 activity is abolished if disulfide bonds are reduced and kept cleaved by alkylation of cysteines. This has been usually interpreted as proof of the native fold structure in angiostatin activity[244][242]. However this model is based on the assumption that disulfide bonds reduction strongly alters the conformation of kringle domains. Force spectroscopy experiments have shown that kringle domains maintain a significantly folded structure even after all three disulfide bonds have been completely reduced[186]. K5 also displays higher anti-migratory activity when reduced[254]. K4 contains cryptic fragments that are more active than the whole folded domain, but that are buried in the fold in the oxidized state[255].

7.4 A mechanoredox hypothesis for angiostatin

7.4.1 Mechanical forces acting on angiostatin

The non-addictive behaviour of kringle fragments angiostatic activity, with the complex combination of synergistic and antagonistic effects (see 7.3.4) is coherent with a mechanochemical model of angiostatin action. Cryptic angiostatic sites buried in equilibrium multi-kringle conformations[255], or the existence of synergistic sites, would be likely affected even by small mechanical forces acting on the protein.

Angiostatin binding partners suggest it is subject to mechanical forces in phys-

iological conditions. Its ligands include components of the ECM (tetranectin and fibrin) and the integrin $\alpha v \beta 3$ on the surface of the endothelial cells (see 7.3.3). Integrins are a crucial point in the ECM-cytoskeleton mechanical communication[34][256]. Angiostatin therefore binds components that relay forces between cells and the ECM. Another cause of applied force on angiostatin might be the continuous remodelling of the ECM structure and conformation. These relative movements of the ECM components would apply mechanical tension to an angiostatin molecule connecting two different points of the membrane.

Angiostatin, in fact, acts in the dynamic process of angiogenesis, which is a place in which significant mechanical forces take a physiological role. Donald Ingber proposed a mechanism in which the control of blood vessel growth is piloted by local modifications of cellular force balances and the tensegrity² of ECM[34]. At the beginning of angiogenesis, in fact, various proteases are overexpressed. These enzymes digest structural matrix proteins, freeing protein cryptic sites and softening the ECM. The default tension that endothelial cells exert on the ECM stretches and thins the digested regions of the ECM: this in turn enhances the tension acting on the focal adhesions of the cells surrounding the proteolyzed ECM. Endothelial cells react to this modification of the tensile forces network by migrating towards the proteolyzed zone, duplicating and rebuilding the ECM behind them. This proteolysis-migration cycle continues at the apex of the growing blood vessel, that “tunnels” literally in the ECM.

There are a number of experimental results that support this model. Tension forces in the extracellular matrix directly controls the growth of new vessels[257]. Endothelial cells morphogenesis is regulated by the action of mechanical forces on integrin inside the ECM[258]. The nascent blood vessels morphogenesis is regulated by the ratio between ECM stiffness and forces generated by endothelial cells[259].

7.4.2 Forced unfolding accessibility of kringle domains

The stretching forces normally generated in vivo by cells and ECM, ranging from one to a few tens of pN[31][32], would not be high enough to break the disulfide

²Tensegrity is a portmanteau of “tensional integrity” . It refers to the integrity of structures as being based in a synergy between balanced tension and compression components.

bonds enclosed in the kringle structure (see 7.3.1) (see Section 7.2.1 and Section 6.5.3). Importantly, one of these bonds directly connects the N and C termini of the domain. Therefore, disulfide bonds act as mechanical barriers that completely prevent angiostatin mechanical unfolding in oxidizing conditions[186].

A reducing environment, however, could cleave one or more of the disulfide bonds. This in turn would enable mechanical unfolding of the angiostatin domains, thus creating a mechanism similar to that describe in Section 7.2.

The hypothesis is worth investigation since it is known the angiostatin disulfide bonds *are* physiologically reduced. Proteolytic generation of angiostatin from plasmin requires reduction of K5 disulfides (see 7.3.2) by mean of various reducing agents. Kringle disulfide equilibria are therefore fundamental for the existence of angiostatin itself. Moreover, as seen in Section 7.3.5, the activity of kringle domains is known to be modulated by disulfide reduction. While these effects have often been attributed to a loss of kringle structural integrity, SMFS established that disulfide bond integrity was not necessary to maintain the angiostatin folding integrity (otherwise, no unfolding peaks could be observed)[186].

Chapter 8

Rationale

As explained above, there are strong reasons to think of disulfide bond redox state as being intimately connected with mechanical processes leading to biological signalling. There are also strong hints of angiostatin being a significant example of this kind of mechanochemical coupling. Having already characterized angiostatin behaviour in different reducing conditions by mean of SMFS in the same laboratory[186], we continued investigation of the same system to elucidate the following points:

- Characterizing the redox equilibrium and the consequent mechanical properties of angiostatin in the physiologically relevant redox conditions, with the use of physiological reducing agents.
- Characterizing the details of angiostatin mechanical unfolding by mean of SMFS experiments and molecular dynamics simulations (in collaboration with the group of Prof.Rita Casadio, University of Bologna)
- Reconnecting these data with the physiological understanding of angiostatin

The work was aimed at characterizing angiostatin as an example of coupling between redox and mechanical processes. To our knowledge, this is the principal and most detailed work in which this link has ever been established. This work, and the theoretical elaboration surrounding it, have been the first, to our knowledge, to suggest the coupling of redox and mechanical processes as a signalling pathway of potential interest for understanding extracellular processes.

CHAPTER 8: RATIONALE

Chapter 9

Materials and methods

9.1 Force spectroscopy experiments with dithiotreitol

A drop of purified angiotensin K1-5 (Calbiochem) solution (20 mL, 0.5 mg) in HEPES buffer (pH 7.4), previously incubated with dithiotreitol (DTT) (100 mM; 2 h, 37 °C; which is more than enough to reduce all intramolecular disulfide bonds[186][260]) was deposited on a clean glass surface that had been rinsed with water, acetone and ethanol, drawn through a Bunsen flame and rinsed again. After a deposition time of 30 min, the sample was then inserted into a fluid cell. The force spectroscopy experiments were performed in micro filtered PBS buffer with DTT (100 mM) with a Nanoscope IIIa multimode atomic force microscope (AFM; Digital Instruments) using sharpened silicon nitride (Si_3N_4) tips (from Park Scientific), at room temperature.

Cantilever spring constants were measured by their thermal spectrum (range between 40 and 60 pN/nm). The loading rate values reported in Figure 10.2 were computed by taking into account that they are related to the pulling speed through the elasticity of the bridge, which is composed of by both the molecule under stress, and the AFM cantilever. After analyzing the experimental data, we estimated that an equivalent spring constant of 27 ± 3 pN/nm can be assigned to the bridge for the speed range that we used.

9.2 Force spectroscopy experiments with thioredoxin

Angiostatin (0.13 mM) was incubated for 30 min at 37 °C with an activated human thioredoxin (hTRX; from Calbiochem) solution (20 mM, angiostatin/hTRX molar ratio about 1:150). After the reduction, angiostatin was alkylated for 30 min in the dark by adding iodoacetamide (from Fluka) at a 5-10 mM final concentration in order to avoid the oxidation of hTRX-reduced Cys residues, and to block hTRX activity. Angiostatin was finally purified from hTRX and iodoacetamide by using YM-30 ultra free Microcon (from Millipore; 4 °C). We found that this simple purification step was enough to avoid interference in the force spectroscopy experiments, due to engagements of the residual hTRX rather than angiostatin molecules. One drop of the final solution (0.5 mM) was used for the force spectroscopy experiments, following the protocol described above, but without DTT in the fluid cell buffer.

The activation of the enzymatic site of hTRX was obtained by incubation with DTT (from Sigma-Aldrich; 1 h, 37 °C, 1 mM) in HEPES buffer (pH 7.4). The activated hTRX was purified by using YM-3 ultra free Microcon (from Millipore) at 4 °C (final DTT concentration below 50 mM). Control force spectroscopy experiments performed with DTT at 50 mM ruled out any detectable reduction of angiostatin, which would have been contributed by the DTT.

9.3 Force spectroscopy raw data analysis

The unfolding force curves were selected by keeping only those clearly presenting the sawtooth profile that is characteristic of angiostatin single-molecule mechanical unfolding for analysis. More specifically, the measured curves showed a) from 2 to 4 peaks (1-3 unfolding peaks plus final detachment); b) peaks whose height was similar besides the case of the last peak that must be the highest (or at least as high as the others); c) no indication of plateaus or other features superimposed on the sawtooth profile. We observed no significant evidence of the two-stage unfolding described for disulfide-bonded domains[261] probably because in our case the disulfide bonds are already naturally accessible to the solvent; however

we cannot completely rule out that some sporadic occurrence of these effects might have entered our statistics.

The selected curves have been analyzed by a custom MatLab program which performs a double-parameter fit on each unfolding peak. The fit function is a worm-like chain (see equation 6.1) and the two fitted parameters are the persistence length (range of variation from 3.5 to 7 Å) and the contour length. The fit is not normally performed on the entire peak profile: in this case, it is advisable in fact to exclude the final part of the peak, because it can be affected by small enthalpic deviations from the entropic elasticity profile. The analysis software allows one to directly select the fit interval on the curve plot.

For dynamic single-molecule force spectroscopy analysis, the single unfolding peaks have then been separated into three classes according to their contour length increase: I (12.5 ± 4 nm), II (21 ± 2.5 nm) and III (30.5 ± 3 nm), according to their triple Gaussian distribution (see Figure 10.2). Only the peaks that belong to one of these three classes have been used for dynamic SMFS analysis.

9.4 Monte Carlo simulations

Monte Carlo simulations of the stretching experiments were performed according to a previously published procedure[262]. with a time step of 10^{-5} s and a protein composed of two modules (the case of three unfolding modules, although observed, was rare enough in our experiments to be neglected). Three different contour length were used, corresponding to the three observable levels of reduction. The dependence of the unfolding force on the pulling speed was simulated at the same five different pulling speeds used in the experiment. The single stretching cycle was repeated 500 times for each set of conditions. Fit of the experimental speed dependence with the simulation was obtained by a two-step procedure. First, the barrier position leading to the same slope in the linear dependence was identified. This is possible since the slope only depends from the barrier position. Second, once the correct slope was identified, the intercept ensuring the best fit of the data was identified. The Monte Carlo analysis has been used for fitting of the linear force dependence, but fitting the single experimental force histograms with that obtained by mean of Monte Carlo was unsuccessful, probably because they were too noisy.

9.5 Steered molecular dynamics simulations.

SMD simulations were performed according to[263]. The kringle domain was solvated with the TIP3 water model, and the energy of the system was minimized. After this, the kringle water box was heated over 10 ps and then equilibrated with a thermal bath at 300 K for another 10 ps. The SMD simulation was performed with a time step of 1 fs, a uniform dielectric constant of 1, and a cut-off of Coulomb forces with a switching function from 10-13 Å . The molecular dynamics were carried out using the CHARMM19 force field, and by fixing the C- α atom of one terminus and applying an external force to the other C- α terminus. For each kringle, 5 different simulations were performed at 1 Å/ps pulling speed. The pulling direction was chosen along the vector from the fixed atoms to pulled one. The value of the spring constant k was set to $10 K_b T \text{ \AA}^{-2}$, corresponding to the thermal fluctuation of constrained C- α at 300 K.

Simulations with these values of k and v correspond to pulling with a stiff spring in the drift regime. The K2 and K3 structures have been extracted from the human K1-3 protein structure[234] (PDB ID: 1KI0); the K4 structure from the structure labeled 1KRN in the PDB.[264] The SMD simulations were carried out with NAMD 2.5[265] and the solvated kringle domains were obtained by using the TIP3 water model with the GROMACS 3.0 program[266]. Both programs were executed on a double processor Intel Xeon 2.4 GHz machine with the Debian GNU/Linux operating system.

9.6 Protein-protein docking simulation.

According to the procedure described in the previous section, the crystal structure of K1-3, with only one disulfide bond reduced for each kringle, was solvated in water and its energy minimized. The whole system was heated and equilibrated at 300 K and then a SMD simulation of 500 ps was performed. For the docking procedure, only a fragment of the K2-3 of 1KI0 structure was considered. The stretched K2-K3 fragment was obtained after 140 ps of steered molecular dynamic simulation. The two subunits of F1 ATPase were extracted selecting the chains A and E of the crystal structure labeled 1BMF in the PDB[267] Docking of the angiostatin fragment to the ATPase was performed by using Z-DOCK.[268] Patches of 10 Å

CHAPTER 9: MATERIALS AND METHODS

radius centered on the relative binding sites were considered.

CHAPTER 9: MATERIALS AND METHODS

Chapter 10

Results

We have studied the mechanical unfolding of the angiostatin domains in their different redox states at the simulation level by the steered molecular-dynamics (SMD) approach[263] and at the experimental level by new SMFS experiments. Because of the setup of the SMFS experiments (see below), the simulations addressed only the K2, K3 and K4 domains, since no reliable unfolding data for the two terminal K1 and K5 domains could be obtained.

10.1 Steered molecular-dynamics simulations of the angiostatin mechanical unfolding under reducing conditions

The simulations showed that, when the three disulfide bonds of the angiostatin domains are fully reduced, the mechanical unfolding of K4 and K2, is controlled by the sequential rupture of two critical sets of interactions: first the hydrogen bonds sealing the C- to the N-terminal break, then a short antiparallel β -sheet ruptures (see Figure 10.1). The failing of the three hydrogen bonds that sustain this β -sheet corresponds to the dominant peak in the simulated force-elongation profiles (see figure 10.1 A) and to an energy barrier located at an elongation of 23 ± 4 Å from the fully folded, native structure. This peak is preceded by 13 ± 2 Å along the elongation coordinate by another peak or hump, which is due to the rupture of the set of hydrogen bonds connecting the N and C termini. The structure of this

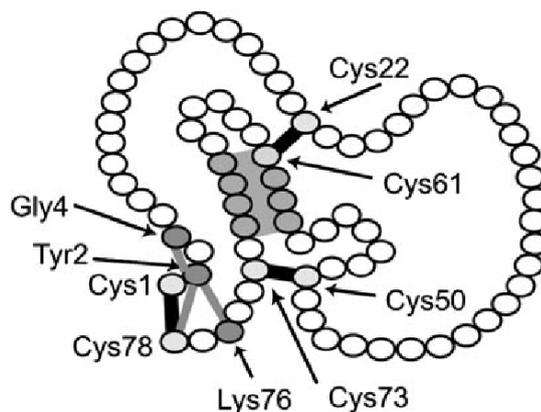


Figure 10.1: The triple-loop topology of an angiotensin kringle domain (K4) defined by the three internal disulfide bonds (black). The two critical noncovalent interactions that control the mechanochemical unfolding of the angiotensin domains are the hydrogen bonds connecting the N and C termini (light gray) and a short β -sheet .

intermediate is shown in Figure 10.1.

The SMD simulations were also performed on the partially reduced states of K2 and K4, in which their two internal Cys22-Cys61, and Cys50-Cys73 bonds are not reduced. As shown in Figure 10.1, their simulated force-elongation profiles do not substantially change with respect to the totally reduced state; the double humped profiles of figure 10.1 are maintained, despite the topological blockades due to the unreduced disulfide bonds (Figure 10.1). In conclusion the simulations showed that, when at least the most external Cys1-Cys78 bond is reduced, the mechanical unfolding of K2 and K4 is controlled by a double-well, free energy landscape like that shown in figure 10.1 B, which has a metastable intermediate, whose structure is depicted in figure 10.1.

In the case of K3, the rupture of the five hydrogen bonds connecting the N and C termini and the failing of the β -sheet occurred without any relaxation into an intermediate in between (see Figure 10.1 A). The intertwining of these two ruptures in K3 is due to the greater number of hydrogen bonds connecting its N and C termini: there are five instead of two or three, as in the case of K2 and K4, respectively. In the case of K3, the rupture of the fifth hydrogen bond takes place when the β -sheet is already under tension.

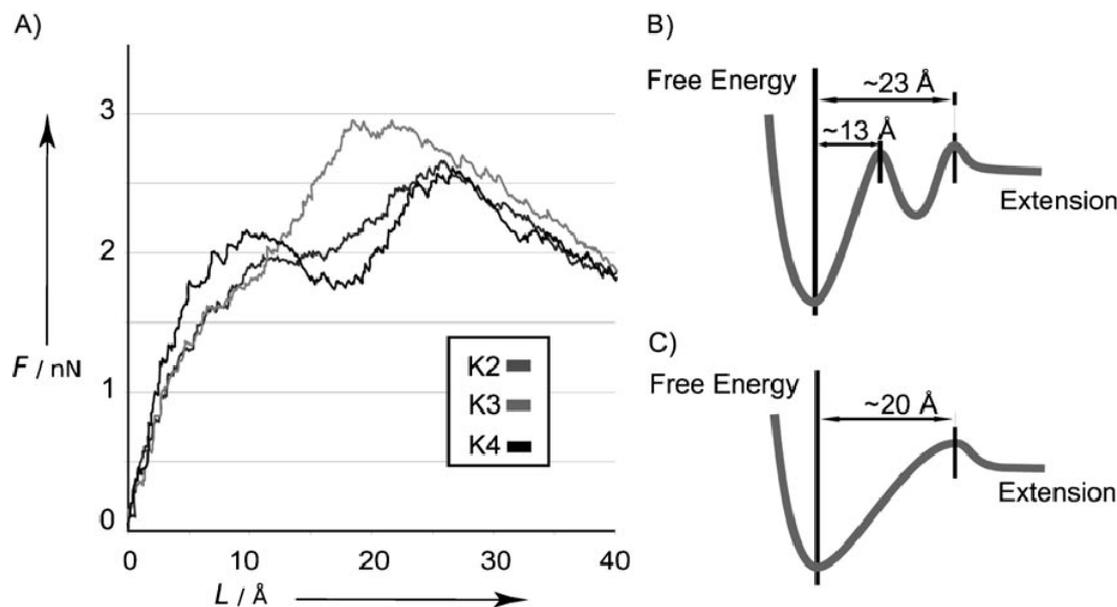


Figure 10.2: Force-elongation profiles of the A) K2, K3 and K4 angiostatin domains, and the corresponding B) double- or C) single-well free-energy landscapes that can be inferred from the SMD simulations on the K2-K4 or K3 domains.

10.2 Stretching single angiostatin molecules under reducing conditions by SMFS.

When single molecules of angiostatin are experimentally stretched by the SMFS methodology, force curves with sawtooth shaped peaks like that in Figure 10.2 are recorded. Therein, each peak corresponds to the unfolding of one single domain.

As demonstrated by a previous study[186] the state of reduction of the relevant domain can be assigned to one of the three reduction stages based on the contour length increments of each peak with respect to the previous one: I (Cys1-Cys80 opened), II (Cys1-Cys80 + Cys22-Cys63 opened) and III (Cys1-Cys80 + Cys22-Cys63 + Cys51-Cys75 opened). This is made possible by the fact that as the reduction of the disulfide bonds proceeds, a larger portion of each domain module can unfold under the external force, and the distance between the peaks increases, as also shown previously[185][186].

About 50% of the recorded peaks exhibit either a shoulder or a double tip at

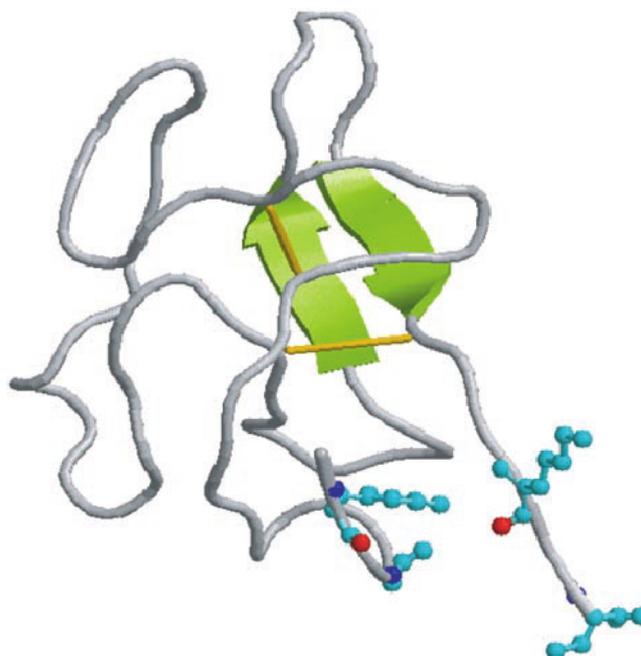


Figure 10.3: Structure of the partially unfolded and partially reduced intermediate that is mechanochemically populated in the K2 and K4 domains. These side chains evidenced are those involved in the hydrogen bonds between the N and C termini that break apart in the first unfolding step.

their crests, like that in Figure 10.2 C-F. This statistic is maintained for all three reduction states. The length increment between the two tips, or between the tip and the shoulder, ranges from 10 to 20 Å, and their force is approximately equal (ratio 1st/2nd tip: 0.95 ± 0.15). Therefore, the profile of these double-tipped peaks matches the energy profile that results from the simulations, and is shown in figure 10.1 for the K2 and K4 domains.

The assignment of these double-tipped peaks to K2 and K4 is confirmed by the frequency of their occurrence, which is in accord with the expected mechanochemical population of the intermediate that was identified by the SMD simulations of those two domains. The other peaks with a single-tipped shape fit the case of K3, which according to the same simulations, should not have any on-path intermediate.

Normally the SMD simulations are used to provide an atomistic interpretation

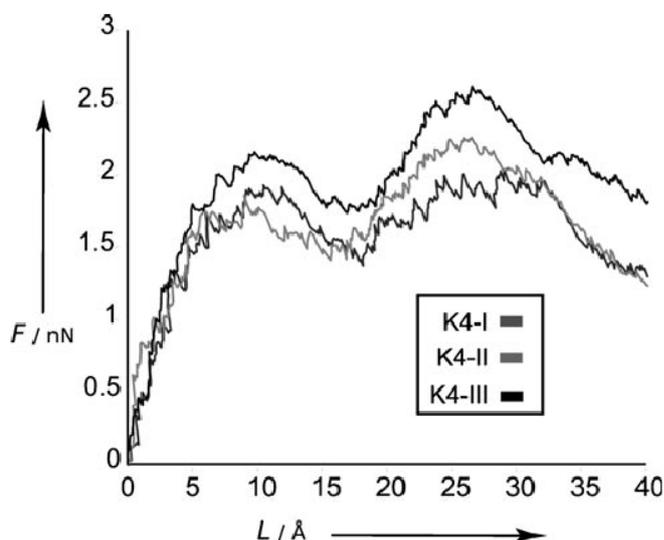


Figure 10.4: The double-humped force-elongation profiles obtained by SMD simulations of the K4 domains in the (I-III) reduction stages. The double-well energy landscape inferred in Figure 2 B for the completely reduced case therefore applies also to the other reduction stages.

for the experimental SMFS force extension data. In this case, due to intrinsic limitations of these data (see below), dynamic SMFS (see 6.7.1) data have been used to support the energy landscape depicted by the SMD simulations. We recorded and analyzed about a thousand force curves under different reducing conditions and at different pulling speeds (from 100 nm s^{-1} to 5600 nm s^{-1}). All the peaks were then assigned to the I, II, or III state of reduction (see [186], Chapter 9 and Figure 10.2) and each data set, corresponding to each state, was analyzed separately.

For each pulling velocity, a histogram like that in Figure 10.2 was obtained, and the most probable unfolding forces, corresponding to each of the three peaks, that is, to each different reduced stage, were reported vs. the natural logarithm of the loading rate (Figure 10.2).

For the reduction stage III, an energy barrier at an extension of $16 \pm 8 \text{ \AA}$ (as shown in Figure 10.2 B) was mapped from the slope of the linear dependences of the force upon the loading rate (Figure 10.2 A). The fit of the experimental unfolding data by Monte Carlo simulations [164] confirmed the value of the energy barrier

CHAPTER 10: RESULTS

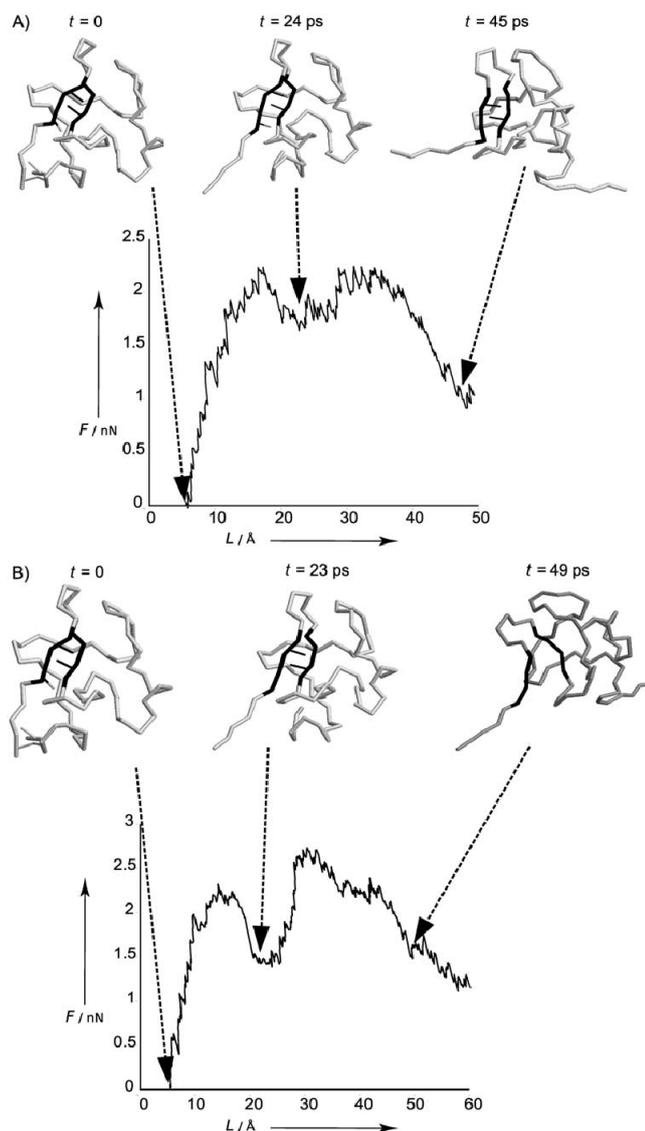


Figure 10.5: Snapshots of the SMD simulation of the mechanical unfolding of the K4 domain with A) only the most external disulfide bond unlocked (state I) B) all three disulfide bonds reduced (state III). In both cases the first barrier to mechanical unfolding is connected to the rupture of the hydrogen bonds between the N- and C-termini. The second barrier is due to the general disruption of the protein core. This takes place in B with the full unzipping of the β strands that is made possible by the reduction of all the three disulfide bonds. The SS bridge in A allows only a distortion and twisting of the β -sheet. On the other hand, this distortion is sufficient to disrupt the core structure of the protein through its invasion by solvent.

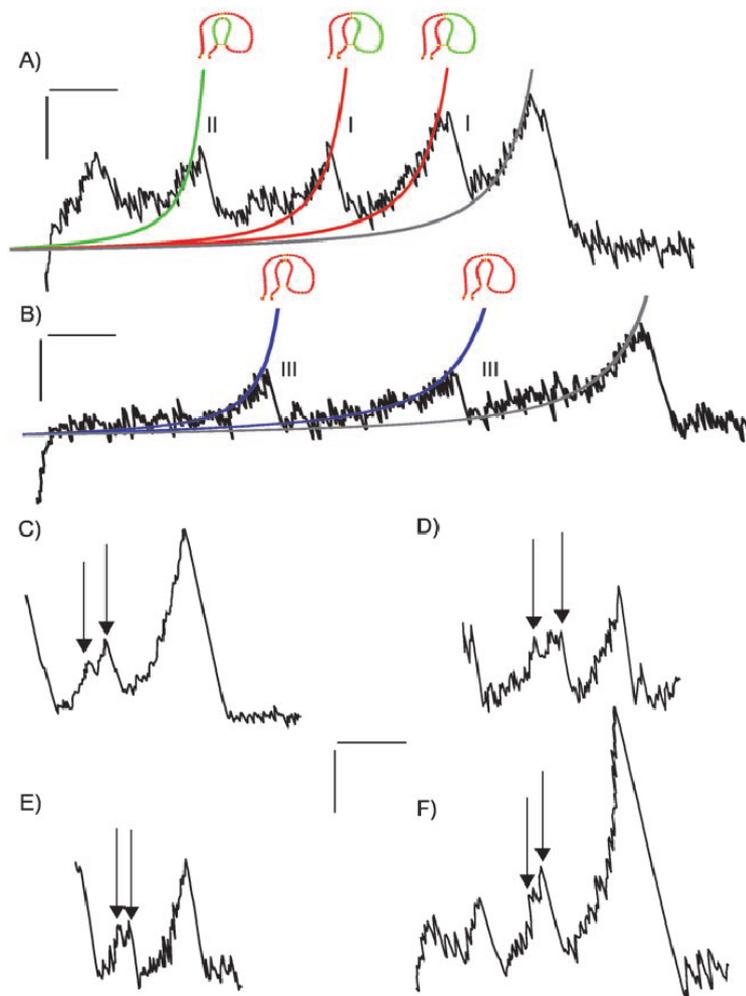


Figure 10.6: A) and B) Typical angiostatin force-elongation curves (black) recorded by AFM-based single-molecule force spectroscopy. The worm-like chain lines are colored according to the state of reduction (I, II and III) that can be assigned to each unfolding peak. The last peak of all curves is the detachment peak (fit line in gray). The corresponding unfolded portion of the kringle domain is colored in red in the 2D sketches over the fitting lines. C) to F) Unfolding curves showing typical double-tipped unfolding peaks (arrows) revealing the population of the on-path intermediate predicted by SMD simulations with the structure reported in Figure 3. Scale bars (in the center) are 100 pN (vertical) and 10 nm (horizontal). The last peak of each curve is the detachment peak. The large vertical portion visible at the left of C, D and E is the final part of the unspecific adhesion signal, and is not related to protein unfolding. All the scale bars are 100 pN (vertical) and 10 nm (horizontal).

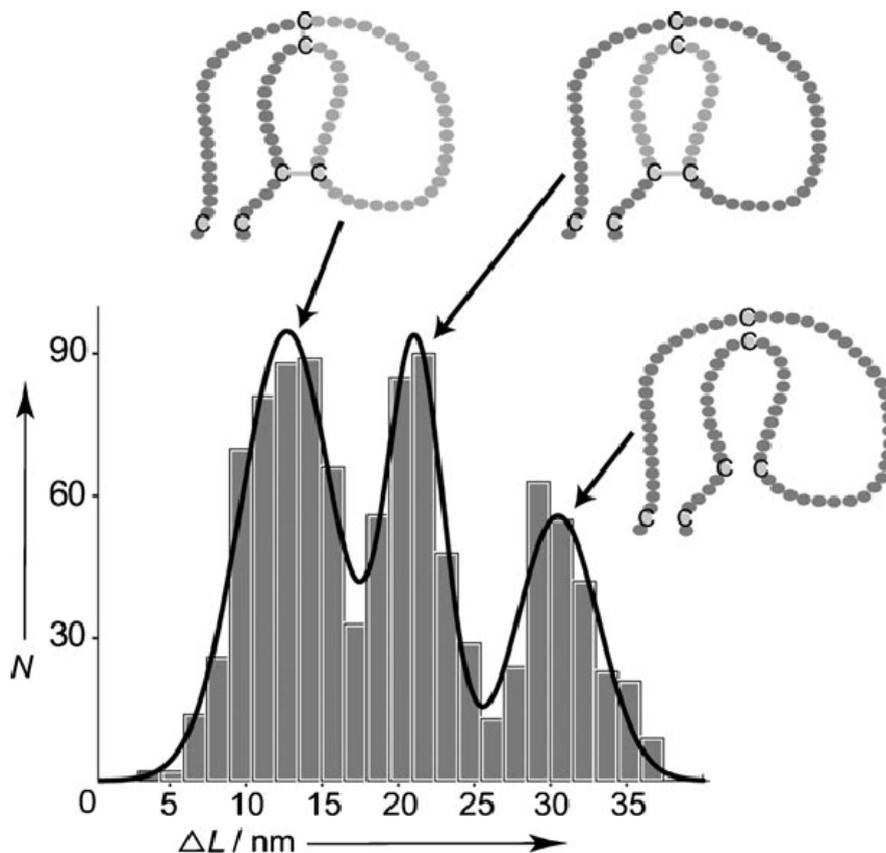


Figure 10.7: A triple Gaussian function with three peaks centered at 12.5, 21 and 30 nm fits the distribution of the unfolding length increments of the unfolded domains, obtained from the contour length increment of each single peak with respect to the previous one. According to Bustanji et al.[186] these values are in good agreement with the unfolding lengths expected for the three (I-II-III) thiol-disulfide reduction stages (about 13.5, 20 and 28 nm). These stages are sketched on the top of each relative band with the portion of the domain accessible to mechanical unfolding (and, therefore, corresponding to the observed average length increment) shaded in dark gray.

position, and estimated a lifetime, extrapolated at zero force, of the order of 10^{10} s, which is consistent with those reported for domains with β -sheet structures[269]. This 16 ± 8 Å position most likely results from the limited capability of the dynamic SMFS to resolve the two barriers that were found by the SMD simulations, and to confine the intermediate identified in K2 and K4 at about 10 and 23 Å. This lack of resolution of the SMFS data is likely due to the intrinsic limitations

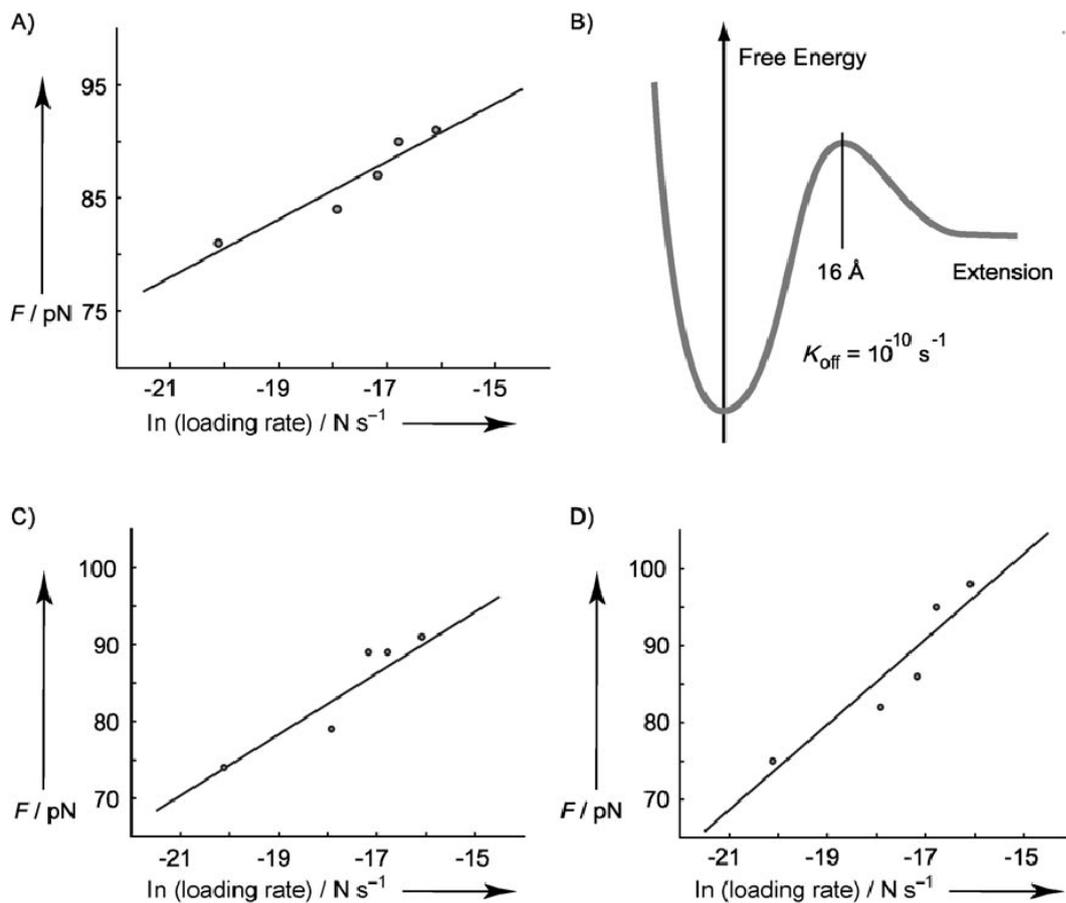


Figure 10.8: Dynamic force spectroscopy of angiostatin domains in the reduction stages (see 6.7.1 for description). The most probable unfolding forces that correspond to the kringle domains in their three reduction stages (I, II and III) were reported vs. the natural logarithm of the loading rate in A (stage III), C (stage I) and D (stage II). The energy landscape extrapolated from the plot in A is sketched in B.

we have met in the curve analysis, which come from the dissimilar structures of the domains.

The dynamic-force spectroscopy analysis for the reduction states I and II (see Figure 10.2C,D) locates the barrier at extensions of 10 ± 8 and 8 ± 4 Å, respectively. The barrier position is therefore almost cut in half when the reduction of the disulfide bonds is not complete. The simulations suggest that, in the reduction states I and II, the β structure is partially blocked by the two internal disulfide bonds (Cys22-Cys61, and Cys50-Cys73), and the nature of the second and final unfolding barrier is markedly different from that of state III: the strands of the β -sheet are not fully broken apart, as in the case of that latter state, but rather distorted, with the residues 14-22 fully exposed to the solvent (see Figure 10.1).

In conclusion, we should expect that when a molecule of this protein experiences a force, once at least the Cys1-Cys78 disulfide bridge has been reduced, the intermediate depicted in Figure 10.1 becomes populated in the K4 and K2 domains. This structure does not correspond to any intermediate in K3.

10.3 Stretching angiostatin in a condition that mimics the redox environment on the surface of a tumor cell.

The physiological significance of the states and the structures identified by the above study of the unfolding energy landscapes of the different reduction states of angiostatin domains, with regard to their antiangiogenic activity, is linked to the possibility that the same protein experiences *in vivo* the conditions that are required to reach them. We therefore translated the SMFS experiments to conditions mimicking the *in vivo* environments that this protein experiences on the surface of a tumor cell. Due to the absence of *in vivo* studies, and because of the extreme difficulty of assigning a reduction state in that environment, there is considerable experimental uncertainty on the physiological reduction state of angiostatin kringle domains. We have seen how the SMFS can make this kind of assignment possible[186] *in vitro*, but not under *in vivo* conditions. We therefore carried out SMFS *in vitro* experiments aimed at reproducing the redox environment that is possibly met by angiostatin on the surface of a tumor cell.

Human thioredoxin (hTRX) is a widespread disulfide reductase present in the extracellular space[225] and it is active in altering the thiol-disulfide equilibrium of cell surface proteins[218][270]. Its reducing activity is ensured by the concurrent extracellular secretion of thioredoxin reductase (TRXR)[226]. Furthermore TRX is greatly overexpressed on the surface of highly metastatic tumors[271][272].

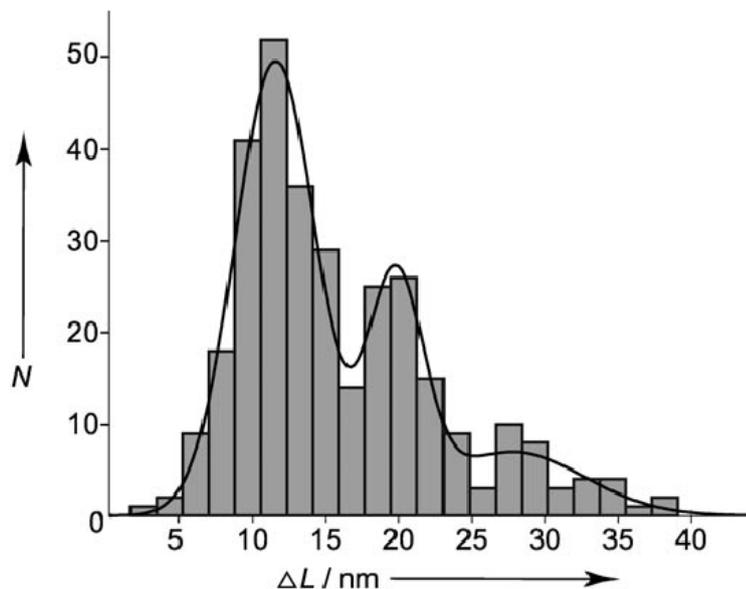


Figure 10.9: The distribution of the unfolding contour lengths obtained by mechanically unfolding angiostatin previously incubated with human thioredoxin. The peak at 12 nm fits the estimated 13.4 nm elongation that corresponds to the reduction of the Cys1-Cys78 disulfide bond only.

We performed the pulling experiments after having treated angiostatin with hTRX at a concentration of 20 μM , which is of the same order of magnitude as that found on the surface of mammalian tissues[272]. After this treatment, we blocked the reduced state by treating the sample with iodoacetamide. Force curves with the characteristic sequence of saw-tooth peaks, like those previously obtained with DTT were obtained. From the length increments in the force curves recorded after the treatment with hTRX, the statistical distribution of the reduction stage of each unfolded kringle domain was obtained (see Figure 10.9). This histogram demonstrates that under those conditions, this enzyme almost selectively reduces the Cys1-Cys78 disulfide bond. The coupling of this reducing enzyme to mechan-

ical stress can therefore lead to the intermediate structure of Figure 10.1.

This result strongly indicates that the angiostatin kringle domains can be present *in vivo*, not only with their native, fully oxidized and folded structure, but also with a series of partially reduced and partially unfolded structures. If we empirically assume that these structures can act as the antiangiogenic active forms of this protein rather than the native one, as commonly believed so far, a set of puzzling activity data can be accounted for, which has been so far unexplained. As explained in the following two paragraphs, one can both explain by using this assumption the triggering mechanism of the cell antimigratory activity of several fragments of angiostatin, and propose a cooperative binding mechanism of angiostatin to the endothelial cell surface ATPase.

10.4 Mechanochemical tuning of the binding of angiostatin with cell surface ATPase.

The binding of angiostatin K1-K5 with the endothelial cell surface ATPase has been recently proposed to mediate its antiproliferative activity by triggering caspase-mediated endothelial cell apoptosis activation[249]. A simple structural rationale for the binding of angiostatin to a single ATPase α -helix has been proposed by Geiger and Cnudde[245]. The docking was based on the kringle ability to bind α -helices.[234][245][273] According to these authors the binding could be sustained by the interactions between exposed helices in the α and β subunits of ATPase, and the surface containing the LBS binding sites of the kringle domains.

By performing docking simulations between ATPase, and a stretched and partially reduced (see below) angiostatin K2-K3 fragment, we found that a cooperative binding between ATPase and adjacent kringle domains can be settled. An increased binding affinity should result from this mechanically-induced, two-site interaction (see legend of Figure 10.12). The simulations showed that when the K2-K3 fragment is stretched, once its external disulfide bond has been reduced, its elongation takes place first at the expense of K3 only (see Figure 10.10). K2 starts unfolding only after the opening of K3, and then reaches the intermediate conformation described above. In this elongated structure of K2-K3, the Glu221 and Arg220 residues of K2 are at an average distance of 51 Å from the Asp309

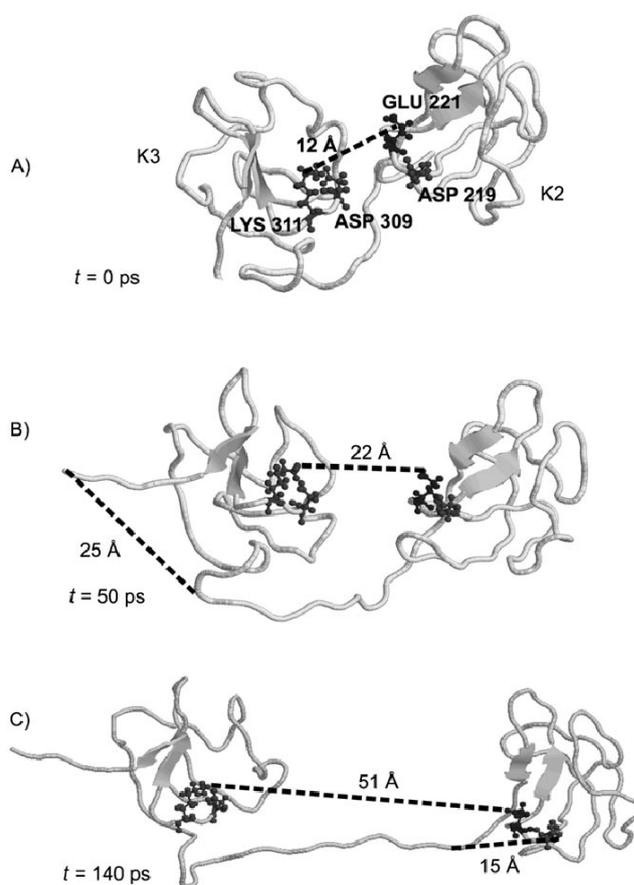


Figure 10.10: Mechanically stretching the K2-K3 fragment. SMD simulations show that when the K2-K3 fragment with the Cys1-Cys78 disulfide bonds reduced in both domains is mechanically stretched, the unfolding starts in K3 first (A). The five hydrogen bonds connecting the C and N termini are broken first. Afterwards, because of the rotation of one of the two strands of the β -sheet with respect to the other, the three hydrogen bonds that sustain this sheet fail (B). The unfolding of K2 starts at this point and reaches the conformation of its intermediate (C). In the meantime the conformation of K3 remains almost unchanged with respect to that in (B). The folding of the two more internal topological loops of the two domains is prevented by the Cys22-Cys61 and the Cys50-Cys73 disulfide bonds, which are not reduced. The preserved integrity of the folding of most sections of the two domains is confirmed by the RMSD values of the structures of the K2 and K3 domains as in (C), with respect to their native structures: 0.8 Å and 2.2 Å, respectively. The resolution of the structure (1KI0) used in the simulations is in fact of the same order of magnitude as these two values.

and Lys311 of K3 (see Figure 10.10). The 51 Å distance makes it possible to those two angiostatin sites to bind in concert to the residues Lys496-Glu499 in chain A and Lys472-Asp471 in chain E of F1-ATPase (Figure 10.11 A and 10.12. Those angiostatin binding sites are instead only 12 Å apart in the native structure of the K2-K3 fragment[234] thus preventing its concerted binding to ATPase to take place.

10.5 Mechanochemical tuning of the exposure of the K4 fragments with the highest antimigratory activity.

The binding of angiostatin to an ectopic, plasma membrane form of ATPase, is one of the mechanisms postulated for its antiangiogenic activity, even if certainly not the only one[275][245][249]. By the same mechanically induced elongation process that can make it possible for the K2-K3 fragment to settle a two-site binding to F1-ATPase, the metastable intermediate of the K4 domain becomes thermally populated. This intermediate can account for a set of puzzling data on the cell antimigratory activity of several fragments of angiostatin.

Among its five kringle domains, K4 is the one with the highest activity[242], and its linear K4-A and cyclic K4-C fragments showed even higher activity than the entire K4 domain[255] (Figure 10.11 B). These two fragments are partially buried in the native structure of the kringle. In the intermediate form, however, the opening of the C-N termini ensures accessibility to the most active linear fragment K4-A, and to the K4-C loop. It should be mentioned, that the looped conformation of the K4-C fragment, which is apparently required for its antimigratory activity, is in this intermediate that is preserved by the integrity of both the β -sheet and the other two Cys22-Cys61 and the Cys50-Cys73 disulfide bonds. We should therefore expect that whenever the reduction has gone beyond the rupture of these latter disulfide bonds, a decrease in activity might result. This prediction is confirmed by the activity data reported by Ji et al.[242].

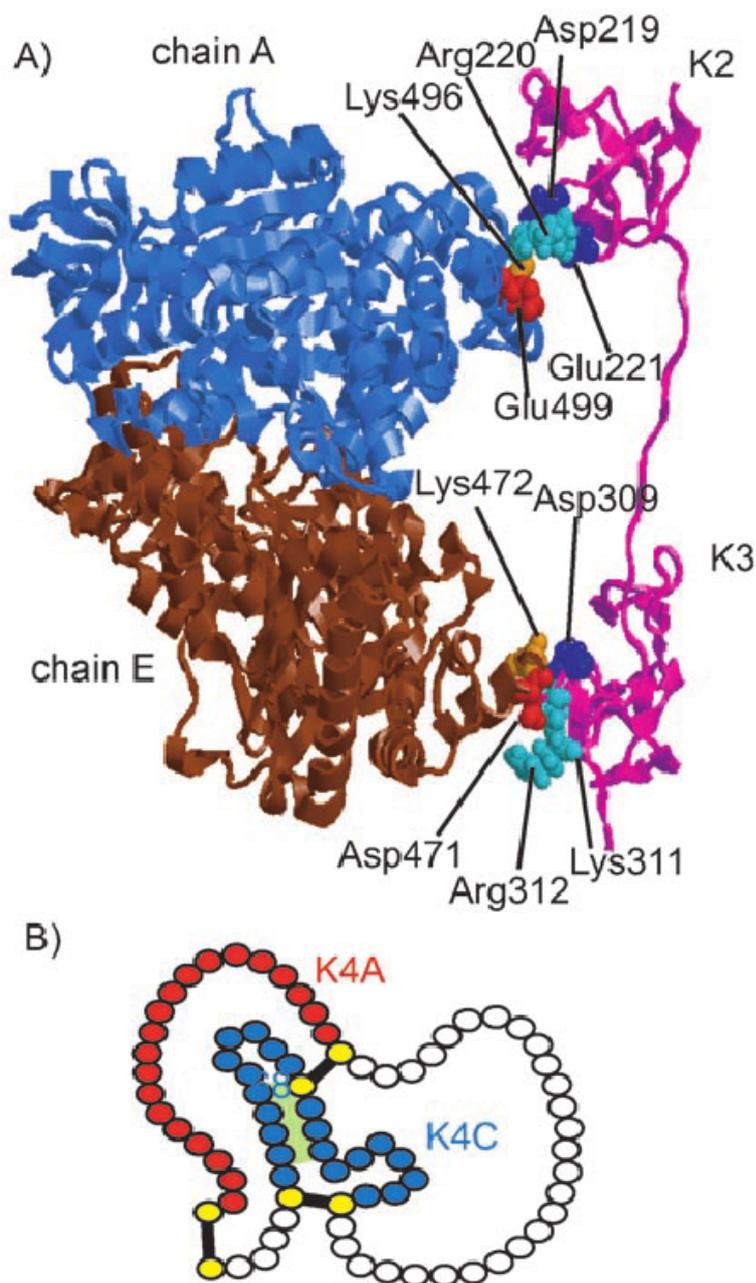


Figure 10.11: A) Mechanochemically induced binding of the β subunit (in blue) and the α subunit (in brown) of the F1-ATPase with the K2-K3 fragment (in magenta), which was elongated according to the energy landscapes in Figure 10.1, and the structure in Figure 10.10. Amino acid numbering refers to the crystal structures used in the simulation. B) Topology of the K4 domain that indicates the fragments (K4-A: red and K4-C: blue) whose antimigratory activity is higher than that of the full domain[255].

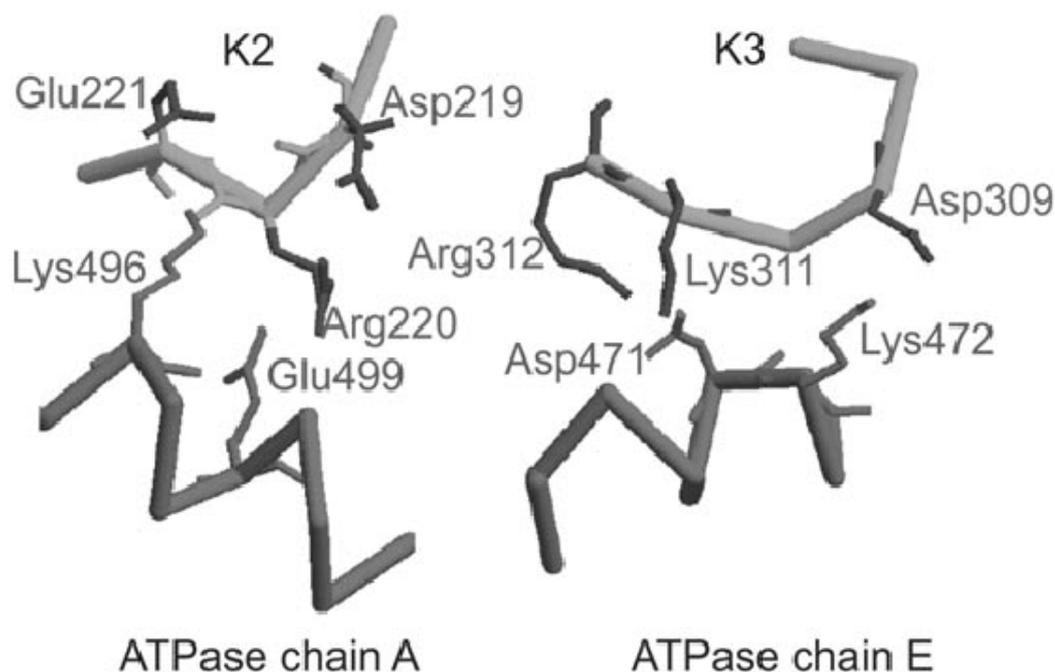


Figure 10.12: Binding energies and specific interactions that sustain the K2 and K3 binding to ATPase. The two most critical amino acid binding determinants of the docking showed in Figure 10.11 A are for K2: Glu221-Lys496, and Arg220-Glu499, whereas for K3 they are: Asp309-Lys472, Asp471-Lys311, and Asp471-Arg312. The K2 and K3 amino acids are numbered following the 1KI0 structure in the Protein Data Bank; the ATPase amino acids are numbered following the 1BMF structure in the PDB. The docking energies were calculated according to S. Liu et al.[274] Binding energy values of -4.99 and -5.20 Kcal/mol were obtained for the interaction of ATPase with K2 or K3 in their native structure, respectively. When a K2-K3 fragment is in its native form, only one of those two interactions can be settled: either with K2 or with K3. When K2-K3 is instead in the elongated form shown in Figures 10.10 and 10.11 A, its binding to ATPase can be sustained by the interactions with both domains. The single binding-energy values were found to increase to -5.14 and -5.61 Kcal/mol for K2 and K3 respectively, leading to an overall estimated binding energy of -10.75 Kcal/mol. In conclusion, the elongation of the K2-K3 fragment not only increases the angiotensin-ATPase binding affinity, by making possible a concerted binding of both domains, but it also increases slightly the binding affinity of each of them.

Chapter 11

Discussion and perspectives

In conclusion, the SMFS experiment carried out with hTRX indicates that the angiostatin kringle domains can be present in vivo not only in their native, fully oxidized and folded structures, but with a series of partially reduced and partially unfolded structures, including that of the intermediate depicted in Figure 10.1. We propose that these structures act as the active antiangiogenic forms of this protein, rather than the native one, as commonly believed so far. In Figure 11.1 a model is proposed for the mechanochemical control of the triggering of antiproliferative and antimigrative activities of angiostatin through the production of those partially reduced and partially unfolded structures. This model follows closely a general model for the combined action of redox and mechanical switches that we have recently hypothesized (see 7.2).

The mechanochemical paradigm proposed here has broad implications for redox biology. It applies in particular to other kringle fragments of multimodular proteins, like prothrombin, apolipoprotein(a) and hepatocyte growth factor, whose role in angiogenesis has been proved[275][276]. Moreover it also applies to the main components of the basement membrane-like thrombospondin-1, laminin or perlecan, which, like angiostatin, are composed of independent modules containing internal disulfide bonds[235][277][278] and experience mechanical stresses in vivo.

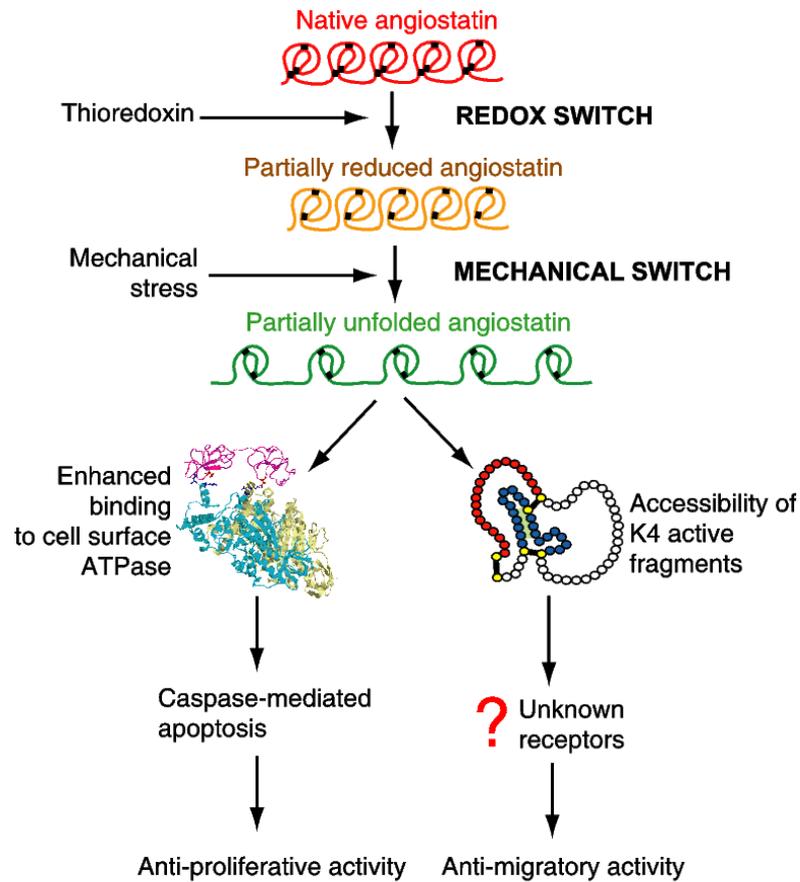


Figure 11.1: Two-switch model of how force and oxidation state can work in tandem for an enhanced antiproliferative and antimigratory activity of angiostatin. A reducing environment[271][272] is able to reduce the most external disulfide bonds of the kringle domains. The mechanical stress that is constantly being developed in the ECM environment, where angiostatin is located[28], now can mechanically unfold a topological loop of 40 amino acids. The resulting partially elongated, and partially unfolded structure of angiostatin (see Figure 10.1) can be biologically active, as shown for the cases of a K2-K3 fragment and the K4 kringle domain. Hierarchical activation of those two switches can lead i) to an enhanced binding of a K2-K3 fragment to ATPase and therefore to an increased antiproliferative activity of angiostatin (see 10.11 A); ii) to exposure of two segment chains that bring antimigratory activities higher than that of the native K4 full domain (see Figure 10.11 B).

11.1 The hierarchy of redox and mechanical regulation

In the scheme in Figure 7.2, Figure 11.1 and the above discussion, a definite hierarchy between redox chemistry and mechanical regulation was assumed. The redox switch has been depicted as the upstream regulator that controls the activation of a downstream mechanical switch. Evidence is emerging however that the also reverse is possible -that is, mechanics can control chemistry. Specifically, redox regulation in the extracellular space can be influenced by mechanical forces. Bhasin et al. [261] have observed that the kinetics of reduction of a disulfide bond buried in modules of the adhesion molecule VCAM-1 can be greatly enhanced by mechanical unfolding of the surrounding protein fold. Steered molecular dynamics simulations indeed showed that the buried disulfide bond becomes exposed in the very first steps of the mechanical unraveling of the modules. This makes it possible for the reducing agent in solution to attack the buried disulfide. In turn, the opening of the disulfide bond allows for further unraveling of the module. In the case of VCAM-1, therefore, multiple redox and mechanical switches may be sequentially activated.

A recent work by Wiita et al. [157] proved that an external mechanical force can directly alter the reduction kinetics of the covalent disulfide bond, independently from its accessibility. In this work, a disulfide bond has been artificially engineered on I27 titin modules and then kept at a constant force by means of the single-molecule force clamp technique, in the presence of DTT. The reduction rate of the bond has been measured to be exponentially dependent on the applied force.

CHAPTER 11: DISCUSSION AND PERSPECTIVES

Part IV

**Probing the conformational
diversity of the intrinsically
unstructured protein α -synuclein
at the single molecule level**

The most exciting phrase to hear in science, the one that heralds new discoveries, is not “Eureka!” , but “That’s funny...”

Isaac Asimov

Chapter 12

Background

12.1 The α -synuclein protein

α -synuclein (α -synuclein) is a 140-amino acid (aa) intrinsically unstructured protein expressed in the human brain, especially in the presynaptic terminals. Its physiological function is uncertain, but there is evidence it is involved in vesicle fusion and trafficking[279]. α -synuclein became subject of much interest as being at the center of the aetiology of various central nervous system pathologies, called synucleopathies. In fact, Lewy bodies and Lewy neurites are intracellular inclusions, mostly containing α -synuclein, found in the brain of Parkinson disease[280] and multiple system atrophy[281] patients.

12.1.1 α -synuclein sequence

α -synucleins is a well conserved protein among many organisms[282]. In humans, the α -synuclein gene encodes at least three α -synuclein isoforms by alternative splicing[283]. α -synuclein -140, the full transcript, is also the majority form of the protein and the most studied. α -synuclein -126 is an isoform deprived of the exon 3 (residues 41-544) while α -synuclein -112 results from the deletion of exon 5 (residues 103-130).

The full sequence of α -synuclein -140, can be divided into three regions:

- The N-terminal region, residues 1-60, is an amphipatic region (consisting of four 11-amino acid, imperfect repeats with the consensus sequence KTKEGV)

with α -helix propensity with similarity to apolipoproteins binding domains[282].

- The central, hydrophobic region (residues 61-95) includes the highly aggregation-prone NAC sequence[284].
- The C-terminal region, residues 96-140, mostly includes acidic residues and contain numerous prolines, and has no structural propensity.

12.1.2 α -synuclein physiological function

α -synuclein is an abundant brain protein (up to 1% of protein content in neuronal cytosol), yet there is still considerable uncertainty on its physiological function.[285]. α -synuclein is located at the pre-synaptic termini, both free and bound to the plasma membrane or vesicles[286]. About 15% of α -synuclein is usually found to be membrane-bound[287]. These data suggest that α -synuclein may play a role in vesicular trafficking and synaptical function in the nervous system[284][282].

There is considerable biophysical evidence suggest that α -synuclein modulates the organization of membrane lipid components. In fact, a yeast genome screening showed that genes correlated with α -synuclein toxicity are involved in lipid metabolism and vesicle turnover[288]. Changes in the expression of α -synuclein are related with modifications in the plasma membrane fluidity and balance of fatty acids[289]. α -synuclein binds vesicle made of phospholipids that expose negatively charged surfaces[289]. The geometry of the vesicles is also important: α -synuclein is more affine to little vesicle than large ones. The reason for this preference can lie in the higher surface-to-volume ratio of small vesicles or, intriguingly, in the difference in local curvature between small and large micelles[290]. Coupling of α -synuclein with negatively charged lipid bilayers or vesicles has profound effects, altering their molecular structure, disrupting the bilayer and leading to the formation of small vesicles [291]. There is also an antioxidant effect of α -synuclein on the lipid membrane and it has been suggested that α -synuclein could be a physiological antioxidant in the nervous system[292].

12.1.3 α -synuclein in pathology

α -synuclein is known to be involved in a group of interrelated central nervous system syndromes called synucleinopathies. These diseases include multiple system

atrophy, Parkinson's disease (PD), and dementia with Lewy bodies (DLB)[293]. In addition to these three diseases, the current list of the synucleinopathies includes (but is not limited to) neurodegeneration with brain iron accumulation, type I (also known as adult neuroaxonal dystrophy or Hallervorden-Spatz diseases), pure autonomic failure and several Lewy body (LB) disorders, including diffuse Lewy body disease (LBD) and the LB variant of Alzheimer's disease[280][293][289].

Synucleopathies share the histological hallmark of neuronal inclusions enriched in amyloid aggregates of α -synuclein molecules. The deposits are found in specific neuronal types, like dopaminergic ones, and in the glia. In Parkinson and multiple system atrophy, these are known as Lewy bodies.

Genetic analysis of familial Parkinsonis unveiled three α -synuclein mutants involved in the disease: A30P, A53T, and E46K [294][295][296]. Apart from causing early onset Parkinsonism in patients, the mutant proteins are also more prone to form nonfibrillar aggregates[297] and Lewy bodies-like fibrils *in vitro*[298]. Moreover, several environmental factors linked to PD have been found to increase *in vitro* propensity of α -synuclein to aggregate into fibrils[289]. Among these there is exposure to organic solvents like carbon disulfide, several pesticides like rotenone and heavy metal ions (in particular divalent cations like copper or iron).

12.2 The problem of α -synuclein conformational equilibrium

α -synuclein has been defined a "molecular chameleon" [299]. The structure of α -synuclein is, in fact, extremely sensitive to its environment and can be easily modulated by even subtle changes in environmental and sequence conditions. The list of known α -synuclein conformations includes both monomeric and oligomeric/multimeric conformations, in both α -helix and β -sheet secondary structures. α -synuclein, in fact, has the uncanny ability to span practically the whole ensemble of conformations available to protein sequences. α -synuclein could therefore prove, in the future, to be a model system to investigate the general folding problem.

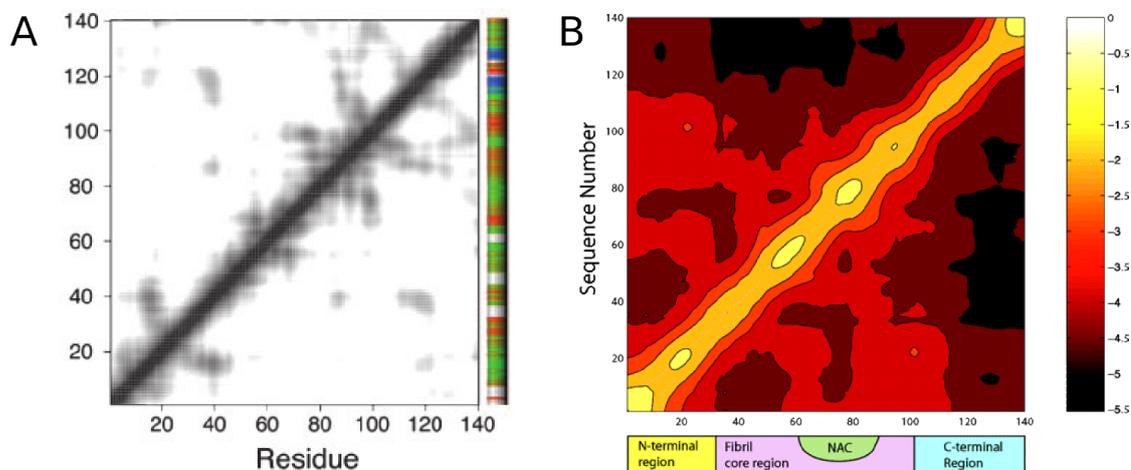


Figure 12.1: Intramolecular contact maps of α -synuclein as found by [301] (A) and [302] (B). Darker colors indicate in both cases residues more probable to be in contact than in a pure random coil.

12.2.1 Monomeric conformers

Disordered conformers Monomeric α -synuclein in dilute buffer solutions characteristically acquires an intrinsically unfolded, extended conformation[300]. This conformational state has been thoroughly characterized by NMR techniques and can be described as a set of non-random, flexible structures kept together by long-range interactions. Bertoni and coworkers[301] found evidence for hydrophobic clusters, long range interactions and calculated several putative native-state clusters from spin-labeled NMR. These clusters consistently showed that the highly acidic C-terminal region contacts the N-terminal region, shielding the central, hydrophobic NAC region. Polyamine binding was found to release long-range interactions and, interestingly, to increase fibrillation rate. Dedmon and coworkers[302] almost at the same time used the same technique, finding similar results but quite a different contact map, in which the C-terminal tail more closely contacts the middle, NAC region instead of the N-terminal region.

Also, a pre-molten globule state has been identified as prevalent in several conditions like high temperature, low pH, metal ions and others[134][300].

Folded conformers α -synuclein, strikingly, can assume both α -helix and β -sheet folded conformations, depending on its environment. α -helix secondary structure is acquired on associating with membranes[303]. Circular dichroism analysis of α -synuclein associated with phosphatidylserine/phosphatidylcholine single-layer vesicles clearly show the loss of random coil signal with a 195 nm minimum and the acquisition of α -helical spectrum[304]. Using NMR, these dramatic changes were attributed to the formation of an α -helix-rich conformation in the N-terminal fragment (100 residues)[305][303]. The α -helix pattern disappears around residues 43 and 44[304]. The structure was then detailed as formed by two α -helices (3-37 and 45-92) joined by a well structured loop[306][303]. In contrast, the acidic C-terminal tail beyond residue 98 never structures, and remains exposed to the solvent[307][306].

Alcohols in solution tend to increase the content of secondary structure in α -synuclein, with a complex pattern in which α -synuclein shifts between different structural states depending on alcohol type and concentration[308]. In general, normal alcohols tended to induce β -sheet structures, while fluorinated alcohols are α -helix forming agents, as expected. The α -helix and β -sheet conformers found in those conditions are initially monomeric, but tended to polymerize with time; in particular β -sheet structures have the distinct tendency to precipitate as amorphous aggregates. α -helix rich globular oligomers, with well-defined tertiary structure, can be formed in high concentrations of trimethylamine N-oxide[309].

It must be noticed that, apart from association with lipid layers, all of these conditions are highly non-physiological and therefore their study, while yielding essential information on the conformational complexity of α -synuclein, cannot be reliably related with physiological structuring of α -synuclein. The conformational diversity of α -synuclein is simply too high and too sensitive to its environment to deduce consistent structural information from such studies.

12.2.2 Polymeric conformers

α -synuclein can associate in a variety of oligomeric and polymeric forms. Simply incubating α -synuclein at high temperature leads to dimerization and then aggregation[134]. A number of conformationally distinct oligomers of α -synuclein are known. Notably, in presence of metal cations α -synuclein can form different

types of oligomers: incubation with Cu^{2+} , Fe^{3+} , and Ni^{2+} yielded 0.8-4 nm spherical particles, similar to α -synuclein incubated without metal ions; incubation with Mg^{2+} , Cd^{2+} , and Zn^{2+} instead resulted in formation of 5-8 nm spherical oligomers; and Co^{2+} and Cd^{2+} generated even larger oligomers, with annular structure of 70-90 nm diameter in the case of Cd^{2+} and 22-30 nm in the case of Co^{2+} [310].

AFM observations consistently detected a variety of oligomeric protofibrils in α -synuclein [311][298][113][312]. Evolution of α -synuclein protofibrils starts from spherical oligomers with sizes in the range of 2-5 nm[311][313] that slowly polymerize into curved, linear assemblies and even full annular structures. Interestingly, α -synuclein oligomers can form pore-like objects when incubated with membranes of brain origin[311].

α -synuclein is capable to form both insoluble amorphous aggregates and insoluble fibrils[299]. The predominance of one form upon the other depends on environment conditions: incubation in alcohols favours the amorphous forms, while incubation with metals generally results in amyloid-like fibrils. There is however an equilibrium between the two forms, as α -synuclein can present both at the same time. These insoluble polymeric forms are of interest, in being comparable with the aggregates that form Lewy bodies into the neurons of patients with Parkinson's disease or other synucleopathies.

EPR experiments showed the extension of α -synuclein included in the main fibril framework: the last residue identified as inside the fibril is 101, while the N-terminal boundary was about at residue 33-34[314]. The detailed conformation of α -synuclein inside amyloid fibrils was partially elucidated by mean of hydrogen-deuterium exchange experiments[315]. The fibrils share the common properties of amyloid structures (see Section 4.2.1), with a cross-beta structure of around 60 aa including mostly residues from 39 to 101. The remaining N and C terminal tails extend outside.

12.3 The problem of α -synuclein aggregation

There is still considerable uncertainty on the aggregation pathway of α -synuclein . The most uncertain factors are the structural nature of the species triggering pathological aggregation and the nature of the equilibrium between the aggregation-prone species and the other structures.

12.3.1 The partially folded intermediate hypothesis

The group of Vladimir N. Uversky and coworkers found evidence of a monomeric partially folded intermediate on pathway to fibrillation[134][300][308]. Various factors known to trigger α -synuclein fibrillation have been found to induce α -synuclein partial folding, including high α -synuclein concentrations[316], molecular crowding[317], metal ions[316]. Pathogenic mutations were shown not to increase the population of the partially folded species, therefore the higher propensity of the mutants to aggregate was attributed to their higher ability to oligomerize[297]. The intermediate was shown to have partial β -sheet structure but to be essentially a molten globule configuration, not a tightly folded species[300].

Interestingly, an elegant NMR experiment in supercooled -15 °C water reported that α -synuclein retains backbone features typical of β -sheet configuration even in extremely cold-denatured conditions, with an hydrodynamic radius corresponding to that displayed in 8 M urea. The location of β strands correlates well with known features of the α -synuclein fibril[318]. It is therefore likely that at least portion of this β -sheet structure may be somewhat more stable in the compact conformations at physiological temperatures.

12.3.2 The extended conformation hypothesis

The group of Tom Jovin, Marcus Zweckstetter and coworkers extensively analyzed α -synuclein conformational properties with NMR and other techniques. Spin-labeled NMR experiments coupled to simulations suggested that long-range interactions create structures where the C-terminal tail shields the NAC region by interacting with the N-terminal region[301]. They have also found that polyamines both release long-range interactions and increase fibrillation rates[301]. The same effect was found and invoked for α -synuclein mutants[319] and other conditions like increase in ionic strength[320][321] known to induce aggregation. In this model, destabilization of the native long-range interactions of α -synuclein lead to exposure of the NAC region that becomes prone to bind other α -synuclein monomers, leading to oligomerization.

CHAPTER 12: BACKGROUND

Chapter 13

Rationale

Our knowledge of IUPs, of amyloid aggregation and, in particular, of α -synuclein, is today mostly due to results obtained by bulk techniques. These techniques work well in the case of a purely homogeneous sample, but suffer the following limits when investigating both IUPs and the first steps of amyloid aggregation:

- Cancel out most of the fluctuations by measuring time- and ensemble-averaged properties. This means that while the fluctuations, their timescales and the involved residues can be at least partially probed, the individual, discrete structures populated by the molecules are hardly resolvable.
- Non-physiological conditions are often required to monitor β -structured conformers possibly prone to aggregation. For example, in the case of α -synuclein, stable monomeric conformers were observed only in presence of high percentages of alcohols[308].
- Monomers are hardly singled out from early oligomers. Spectroscopic techniques like FTIR or CD in particular cannot easily distinguish between monomers, dimers or small oligomers, even if light scattering can help understand the issue. An increase in structured content, therefore, cannot often be attributed with certainty to the monomer or early oligomers.

Single molecule techniques, even if still young, should provide instead substantial advantages (see 5.1):

CHAPTER 13: RATIONALE

- Single out monomeric conformers from early oligomers. If the experiment is correctly designed, oligomerization is simply kept out of the way and only monomers can be probed.
- Classifying differences between single conformers. Since single molecule are probed, signals from each molecule can be analyzed without interference from other molecules. Differences between discrete conformers classes can be classified.
- Probe their conformational equilibria. By shifting the physico-chemical environment, variation in the properties and population of conformational classes can be monitored.

Therefore, to help untangling the problem of the very first steps of α -synuclein aggregation (see 12.3), we decided to analyze the conformational properties of α -synuclein at the single molecule level by mean of CS-SMFS, in collaboration with the group of Luigi Bubacco (Department of Biology, University of Padua). To our knowledge, this is the first time α -synuclein conformational properties were investigated at the single molecule level, and the first time that a IUP is investigated by mean of SMFS techniques. The experiment has been designed as to investigate the monomer without any possible interference from early oligomers.

Moreover, while SMFS has already shown potential to understand force-induced conformational intermediates in systems of known structure (see 6.3.3), this is to our knowledge the first time that SMFS is used as a structural technique *per se* (albeit very rough).

Chapter 14

Materials and methods

14.1 Polyprotein design and expression

To stretch an individual α -synuclein molecule by AFM, we need handles to connect one end of the protein to the tip and the other to the substrate. To this aim, we followed the design proposed by J. Fernandez for the study of the random coiled titin N2B segment[322]. A chimeric polyprotein composed of a single α -synuclein module flanked on either side by three tandem I27 domains (Figure 14.1) was expressed[170][197][323]. These domains act as molecular handles to mechanically stretch a single α -synuclein molecule. They also introduce well-characterized fingerprint signals into the recorded force curves that make it possible to identify the different α -synuclein conformations. To better understand the rationale of the polyprotein strategy, see 6.4.

Chimeric polyproteins were obtained starting from pAFM1-4, pAFM5-8, and pAFM(I27)3mer vectors, kindly provided by Professor Jane Clarke (Cambridge University, United Kingdom) and constructed according to Steward et.al. [323]. Native human α -synuclein or its A30P mutated sequences were amplified by PCR using two different pairs of primers, each containing unique restriction sites. A first pair contained KpnI and XbaI sites, and a second one contained SacI and BssHII sites. The original eight I27 module plasmid was reconstituted from pAFM1-4 and pAFM5-8, obtaining the pAFM8m vector. pAFM8m was then digested with KpnI and XbaI and ligated to the amplified α -synuclein sequence, then cleaved by the same enzymes in substitution of the two central titin modules to give the pAFM3s3

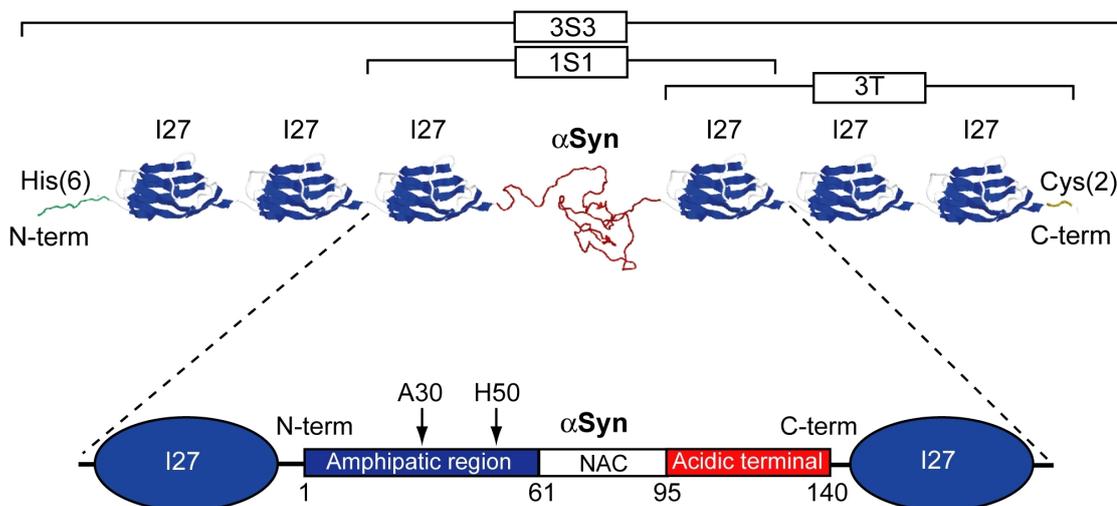


Figure 14.1: Schematic representation of the polyprotein constructs used in this work: 3S3 contains the α -synuclein sequence (red) flanked on either side by three titin I27 modules (blue), the N-terminal His-tag needed for purification purposes (green), and the C-terminal Cys-Cys tail needed for covalent attachment to the gold surface (yellow). In 1S1, the α -synuclein moiety is flanked only by one I27 on both sides; the 3T is made up by three I27s. In the α -synuclein moiety (enlarged), three regions are shown: (i) the amphipathic region, prone to fold in α -helical structures when in contact with phospholipid membranes; (ii) the fibrillogenic NAC region, characteristic of the fibril core of α -synuclein amyloid; and (iii) the acidic C-terminal tail, strongly charged and not prone to fold. The positions of alanine 30, site of the A30P mutation and histidine 50, which is crucial for the binding of Cu^{2+} , are marked.

vector. By a similar strategy, the pAFM(I27)3mer vector was digested with *SacI* and *BssHII*, and the central titin module replaced by α -synuclein sequence, obtaining the pAFM1s1 vector. The obtained expression plasmids, pAFM3s3 and pAFM1s1, code for two chimeric polyproteins composed of a single α -synuclein module flanked on either side by three tandem I27 domains or by just one, named 3S3 and 1S1, respectively. The two pAFM8m and pAFM(I27)3mer vectors (coding for two recombinant poly(I27) proteins named 8T and 3T) were transformed into *Escherichia coli* C41(DE3) cells[324] (obtained from Professor John E. Walker (Medical Research Council Dunn Human Nutrition Unit, Cambridge, United King-

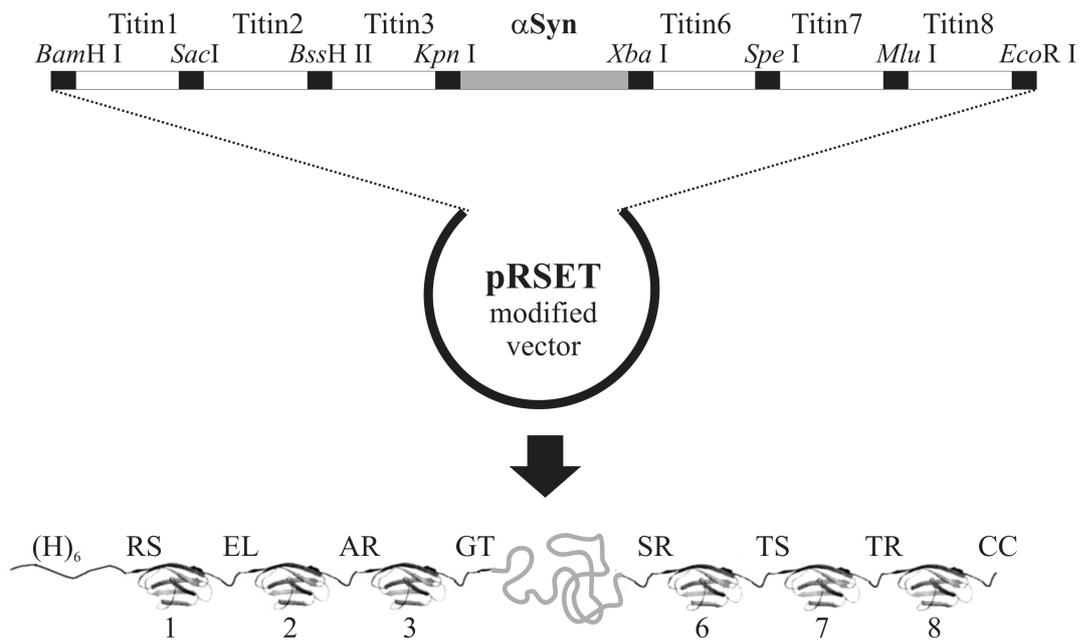


Figure 14.2: Top panel: Schematic diagram of pAFM3s3 vector obtained from cloning α -synuclein sequence in J. Clarke's pRSET-A modified vector and (bottom panel) representation of the chimeric protein 3S3 coded by the cloned DNA sequence. Titin module numbers refers to the original vector described in [323]. pAFM (I27)3mer and pAFM1s1 vectors and corresponding chimeric protein (respectively 3T, 1S1) can be described with similar diagrams

dom) with the agreement of the Medical Research Council center of Cambridge). The cells were grown and the expression of proteins was induced as described in [323]. Recombinant proteins were purified by Ni²⁺-affinity chromatography in 20 mM sodium phosphate buffer pH 8, 500 mM NaCl; the elution from the resin was obtained with 20 mM imidazole. After dialysis, proteins were kept at -80 °C in phosphate buffered saline (PBS) with 15% glycerol.

14.2 Circular dichroism experiments

CD measurements were carried out on a JASCO J-715 spectropolarimeter interfaced with a personal computer. The CD spectra were acquired and processed using the J-700 program for Windows. All experiments were carried out at room temperature using HELLMA quartz cells with Suprasil windows and an optical path length of 0.1 cm. Spectra were recorded in the 190-260 nm wavelength range using a bandwidth of 2 nm and a time constant of 2 s at a scan speed of 50 nm/min. The signal-to-noise ratio was improved by accumulating at least four scans. All spectra are reported in terms of mean residue molar ellipticity $[\Theta]_R$ (deg cm² dmol⁻¹).

14.3 Fluorescence experiments

Fluorescence emission spectra were recorded on a Perkin-Elmer LS 50 spectrofluorimeter equipped with a thermostated cell compartment and interfaced with a personal computer using the FL-WinLab program for Windows. Sample measurements were carried out using a HELLMA ultra-micro cell with Suprasil windows and an optical path length of 10 × 2 mm. Fluorescence spectra were obtained at 25 °C using an excitation wavelength of 288 nm, with an excitation bandwidth of 4 nm and emission bandwidth of 4 nm. Emission spectra were recorded between 290-380 nm at a scan rate of 60 nm/min.

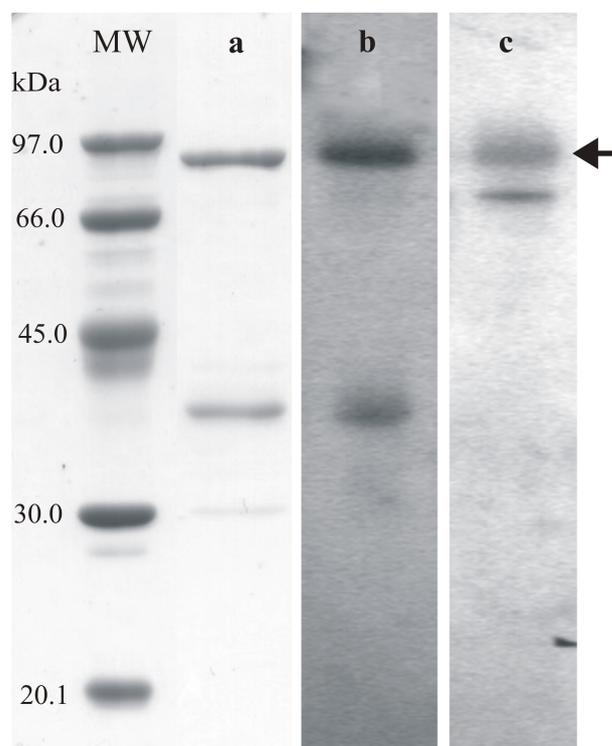


Figure 14.3: Sample of purified 3S3, SDS-10% PAGE and Western blot analysis. MW, molecular mass markers (Amersham Biosciences); *lane a*, final product of immobilized metal affinity chromatography (IMAC) purification of 3S3 chimeric protein; *lane b*, Western blotting with anti-His tag Ab; *lane c*, Western blotting with anti- α -synuclein antibody. The arrow indicate a band corresponding to 3S3 protein with an expected molecular weight of 78119 Da which is in good agreement with observed electrophoretical mobility. The lower band that is copurified with 3S3 is recognised only by the anti-His tag Ab, indicating that it is probably an abortive product of translation that does not contain α -synuclein and, consequently, the last three titin modules with the two final cysteines, necessary for linking the protein to the gold surface. The lower band recognised in lane c is present also in negative controls (data not shown) so it can be considered an aspecific band.

14.4 Buffer elemental analysis

Due to the well-known structuring effects of divalent metal ions on α -synuclein [316], an accurate elemental analysis of the buffer was performed to exclude artifacts in our results due to metal contamination. The high concentration Tris-buffer solution (500 mM) was analyzed for metal contents by atomic absorption spectroscopies. The measured concentrations were $\text{Cu} = 0.2 \pm 0.1$ nM, $\text{Zn} = 3.5 \pm 0.1$ nM, $\text{Fe} = 0.9 \pm 0.1$ nM, and $\text{Ca} = 22.5 \pm 0.1$ nM. These values are two orders of magnitude lower than the concentration required to induce structural effects on α -synuclein [325].

14.5 Surface preparation

Gold (Alfa Aesar, 99.99%) was deposited onto freshly cleaved mica substrates (Mica New York Corp., clear ruby muscovite) in a high-vacuum evaporator (Denton Vacuum, model DV502-A) at 10^{-5} Torr. Before deposition, the mica was preheated to 350 °C by a heating stage mounted behind the mica to enhance the formation of terraced Au(111) domains. The typical evaporation rate was 3 Å/s, and the thickness of the gold films ranged around 300 nm. The mica temperature was maintained at 350 °C for 2 h after deposition for annealing. This method produced samples with flat Au(111) terraces. These films were fixed to a glass substrate with an epoxy (EPO-TEK 377, Epoxy Tech.). They were then separated at the gold-mica interface by peeling immediately before functionalization with the desired molecules. This procedure produced gold substrates with a flat surface morphology due to the templating effect of the atomically flat mica surface [326][327].

14.6 Force spectroscopy experiments

For each experiment, a 20 μl drop of 3S3 construct solution (160 $\mu\text{g}/\text{ml}$) was deposited on the freshly peeled gold surface for about 20 min. SMFS experiments were performed using a commercially available AFM system: Picoforce AFM with Nanoscope IIIa controller (Digital Instruments) using V-shaped silicon nitride cantilevers (NP; Digital Instruments) with a spring constant calibrated by the ther-

mal noise method[152]. The pulling speed was $2.18 \mu\text{M} / \text{s}$ for all experiments. The buffer used was Tris/HCl (10 mM or 500 mM, pH 7.5; the 10 mM buffer was obtained by diluting the 500 mM buffer with milliQ ultrapure water). For CuCl_2 experiments, the protein was deposited in a drop with the addition of a final concentration of $1 \mu\text{M}$ CuCl_2 and left on the surface for about 20 min, and the experiments were carried out in 10 mM Tris/HCl with $1 \mu\text{M}$ CuCl_2 . Control experiments in dithiothreitol (DTT) were made in 50 mM DTT Tris/HCl buffer.

14.7 Data analysis

The force curves were analyzed using the commercially available software from Digital Instrument (Nanoscope v6.12r2), custom Origin scripts and *Hooke*, a Python-based force spectroscopy data analysis program discussed in Part V. Force curves were analyzed fitting each peak with a simple WLC force versus extension model[160] with two free parameters: the contour length L and the persistence length p (see 6.1). The I27 modules were characterized in terms of the length of the polypeptide chain extended after each unfolding event.

The design of the multimodular construct is such that if the number of unfolding signals coming from I27 modules is larger than four, we are sure to have also mechanically stretched the α -synuclein module in the middle (see Fig. 14.1). Among the curves showing mechanical unfolding events, however, only those featuring at least six unfolding peaks were selected and analyzed. This choice reduced the statistical sample even more, but it allowed us to recognize, in a very stringent way, the signatures of the different conformations of the α -synuclein moiety on each construct molecule that had been stretched. Of curves, only those displaying an overall contour length consistent with it resulting from a single molecule and not dimers or unknown fragments have been chosen.

It was also found that length characterization of weak interaction (see 15.8) was much noisier on curves with less than six peaks, probably due to I27 partially unfolded segments next to the tip contributing both to spurious interactions and broadening of the overall lengths.

To assess the statistical validity of the comparison between data obtained in 10 mM Tris/HCl buffer and those obtained in other conditions, standard chi square tests were performed (see 15).

CHAPTER 14: MATERIALS AND METHODS

Chapter 15

Results

15.1 α -synuclein single molecules display conformational diversity on the AFM timescale

To probe the native-like conformer population of α -synuclein, we performed experiments in a 10 mM Tris buffer solution. We found that the profiles of the selected force curves can be classified into three main classes (Figure 15.1). Molecules belonging all three classes were observed in the same experiment, in the same conditions: the diversity is thus an intrinsic characteristic of the sample. Two were unambiguously assigned to well-defined classes of conformers: one with the typical mechanical behavior of random-coil chains and the other of β -like structures. We propose that the profiles of the third class correspond to fairly compact architectures, likely to be sustained also by interactions among different modules of the construct.

15.2 Conformers with the mechanical properties of a random coil

In the class of traces depicted in Figure 15.1 A, the force curve exhibits (from left to right) a long initial region, without any significant deviation from the worm-like chain (WLC) behavior (see 6.3.3), followed by a saw-tooth pattern with six consecutive unfolding events, in addition to the last one that corresponds to the

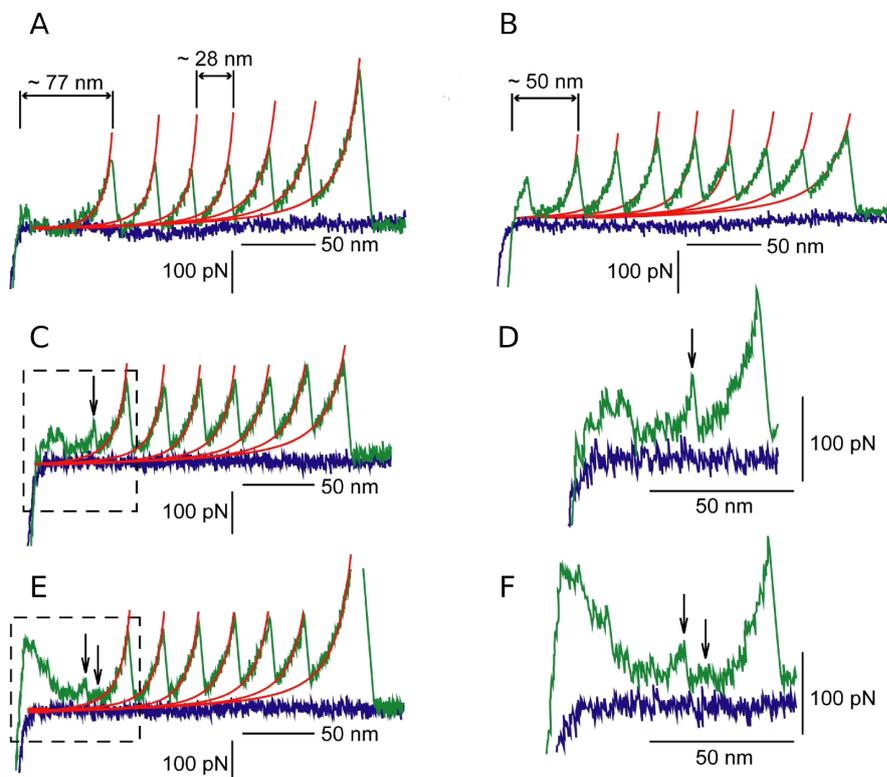


Figure 15.1: Conformational diversity of α -synuclein as seen by CS-SMFS. A) Example of curve characterized by a featureless region assigned to the stretching of α -synuclein moiety having, in this case, the mechanical properties of a random coil. This region is followed (from left to right) by six unfolding peaks of about 200 pN, with about 28-nm gaps between each, assigned to the unfolding of I27 domains. B) Example of the curves featuring the β -like signature of α -synuclein, showing seven practically indistinguishable unfolding events of similar magnitude and spacing. C, E) Curves featuring the signature of mechanically weak interactions, showing single (C) or multiple (E) small peaks (arrows) superimposed on the purely entropic WLC behavior of the trace preceding the six sawtooth-like peaks. D, F) show a zoom of the region enclosed by the dashed squares.

final detachment of the molecule from the tip. The initial region thus corresponds to the extension of a chain that occurs at low force and without significant energy barriers limiting its extensibility.

The six unfolding peaks are spaced by 28 nm. This spacing between the peaks corresponds to an 89-aa chain (0.36 nm per amino acid [164]), and they correspond, as expected, to the increase in contour length of the protein after the unfolding of one I27 domain. These six unfolding peaks are the characteristic fingerprint of the mechanical unfolding of the six I27 modules[322]. We can therefore infer that, in this case, the AFM tip picked up the 3S3 construct molecules at one end (probably the His-tag terminus), while the other end was tethered to the gold surface by the C-terminal cysteines.

The location of the first unfolding peak of I27, corresponding to the contour length of the construct molecules prior to any unfolding event, proves that the preceding featureless part of the trace can be unambiguously assigned to the α -synuclein chain. In fact, the measured contour length that fits this peak is 77 ± 4 nm (Figure 15.3). Subtracting the length of the six, still folded, I27 domains from this value (4.5 nm each [169]), a value of 48 ± 4 nm is obtained. This length corresponds to the chain of 140 aa of the α -synuclein . Therefore, this featureless initial part is the signature of α -synuclein conformers with the mechanical properties of a random coil. Their average persistence length was estimated by fitting the WLC model at 0.36 ± 0.05 nm, consistent with that of an unfolded polypeptdic chain. About 38% of the molecules showed this mechanical behavior in Tris/HCl buffer 10 mM (Figure 15.3)).

15.3 Conformers with the mechanical properties of a chain containing a β -like structured segment

A significant proportion of force curves with seven regularly spaced unfolding peaks in the 200-pN range (in addition to the last one corresponding to the final detachment) (Figure 15.1 C) was also recorded.

15.3.1 Assignment of seven-peaked force curves to the pulling of a single 3S3 molecule

The occurrence of *seven* apparently identical unfolding peaks was an unexpected occurrence. The similarity between peaks is highly suspect, and thus great care must be taken to ensure these curves are not an artifact. However, the following data permit us to attribute this class of curves to the unfolding of a single α -synuclein domain inside the 3S3 construct:

- The presence of a number of unfolding events greater than that of the I27 modules in the construct cannot be ascribed to a possible simultaneous pulling of more than one 3S3 molecule, because pulling two multidomain constructs at the same time would not likely lead to a uniform, 28-nm spaced separation between the I27 unfolding events (see 6.8).
- The appearance of seven unfolding events cannot come from a construct accidentally expressed with seven, instead of six, I27 domains because of the cloning strategy (see 14).
- Curves containing eight or more unfolding peaks were seldomly recorded. All of these curves showed a total contour length of the molecule much larger than that possible for a single 3S3 construct, and were often irregular in appearance. These curves could be due to occasional dimers of 3S3 forming in solution. In contrast, most curves showing seven unfolding peaks showed characteristics consistent with the pulling of a single 3S3 molecule. The occasional seven-peaked curves with characteristic incompatible with the pulling of a single 3S3 molecule were discarded. To further dismiss the possibility that curves with seven peaks can be due to dimers in solution, the sample was tested using dithiothreitol (DTT) to avoid any disulfide-bonded dimer formation (see 14). Under these conditions, the statistics of different populations was comparable to those in the standard buffer, and we still recorded a significant proportion (about 10%) of seven-peaked curves.
- Finally, curves with seven unfolding events were well reproducible, and their statistics were unambiguously modulated by conditions able to trigger aggregation: e.g., ionic strength, the presence of Cu^{2+} ions and, most importantly,

pathogenic mutations (see below).

15.3.2 Features of the β -like α -synuclein segment

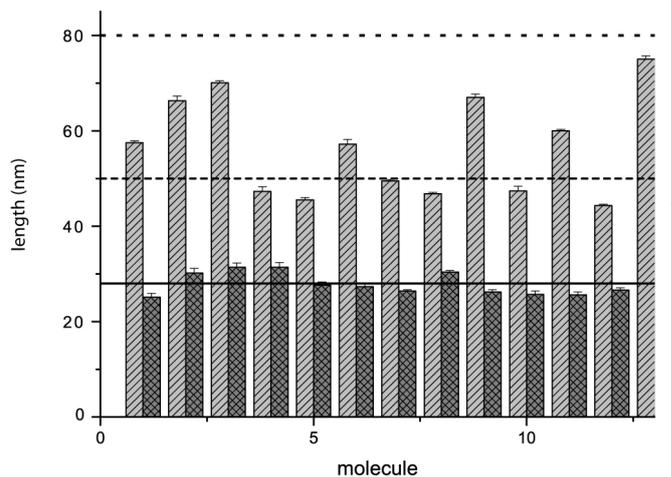


Figure 15.2: Values of the first peak position in force curves showing seven unfolding peaks. The height of each bar corresponds to the initial contour length of a single curve, obtained by fitting the first unfolding peak by means of the WLC model. The dashed line is the length corresponding to a protein construct with six I27 folded modules plus 95 aa of α -synuclein folded into a β -like structure, and the remaining 50 aa of α -synuclein unstructured (see Discussion section). The dashed-and-dotted line is the length corresponding to a protein construct with six I27 modules plus the 140 aa of α -synuclein completely unstructured. The lengths of twelve randomly chosen I27 modules have also been reported (dark gray columns) for comparison. The solid line is the nominal I27 contour length. The larger spreading of the α -synuclein data confirms the higher conformational heterogeneity. Side quotas show the difference between the maximum and minum observed length value for I27 (bottom) and β -like structures (top).

Because of the previous considerations, we therefore assign one of the seven peaks to the unfolding of the α -synuclein moiety. The length (95 aa) and unfolding force of this α -synuclein β -like folded section accidentally coincides with that of the I27 domain. This coincidence hinders the possibility to discriminate the peak of the α -synuclein from the six of the I27 domains. Nevertheless, the assignment of these curves to the unfolding of the α -synuclein moiety is confirmed by the position of the first unfolding peak, *i.e.* by the contour length of the construct molecules prior

to any unfolding event. As shown in Figure 15.2, the position values correspond to a chain composed of the six I27 folded modules, plus the α -synuclein moiety with its C-terminal segment of 50 aa fully unfolded, and the remaining 95 amino acids folded into a structure with the same contour length as a folded I27 domain (solid line). The low propensity to fold of the 50 aa of the very acidic α -synuclein C-terminal tail has been extensively documented[305][303][328]. Segments of the remaining 95 amino acids are instead known to fold under different conditions into an α -helix [305][303] or, in the amyloid, into a β -sheet structure[329]. It must be noted that in about 40% of the molecules, the contour length of the same folded section is larger than that corresponding to 95 aa. The α -synuclein structural diversity therefore includes also β -like chain portions with different lengths. This interpretation is confirmed by comparing the variance of the folded section of seven-peaked curves with that of I27 modules (Figure 15.2).

These curve profiles provide evidence that in 10 mM Tris/HCl buffer, about 7% of the molecules (Figure 15.3) contain a segment of the α -synuclein chain of about 95 aa folded into a structure with the mechanical property of the I27 β -sandwich structure. This percentage of the β -like structures, as we will see below, can be related with conditions leading to pathogenic aggregation.

15.4 Conformers displaying mechanically weak interactions

The remaining force spectroscopy curves (Figure 15.1 C) show single or multiple small peaks (sometimes with a plateau- or dome-like appearance) superimposed on the purely entropic WLC behavior of the trace preceding the six sawtooth-like peaks.

15.4.1 Assignment of the mechanically weak interactions to the pulling of the α -synuclein module

The geometry of our construct made it possible to exclude that those small peaks might correspond to the rupture of aspecific α -synuclein -gold interactions. In fact, if the unstructured α -synuclein was adsorbed on the surface, upon pulling

the construct, we would have recorded the first event at a distance from the tip contact point corresponding to the length of the three I27 modules (about 13.5 nm). The mechanically weak events we observed instead took place at an average distance from the contact point of 60 ± 26 nm with no events below 20 nm. They are therefore not compatible with α -synuclein-gold interactions. We assign these signals to the rupture of mechanically weak interactions placed at short and long distances along the chain.

15.4.2 Features of the mechanically weak interactions

The average forces of those single or multiple small peaks of the profiles are in the 64 ± 30 pN range (well above the noise level), without a defined hierarchy; often stronger peaks precede weaker ones, evidence of topologically nested interactions. From the difference between the contour length estimated at those small peaks and that at the first I27 unfolding peak, one can measure the size of the topological loop enclosed by the interactions whose rupture is monitored by the different peaks. For a detailed discussion of the topology of these interactions, see below (15.8)

More than 50% of the molecules showed short- and long-distance mechanically weak interactions in 10 mM Tris/HCl buffer (see Figure 15.3).

15.5 Shift of conformational equilibrium due to sequence and buffer alterations

To test the physiological significance of the conformational equilibrium as probed by SMFS in the 3S3 construct, we did experiments in three markedly different conditions whose *trait d'union* is the ability to enhance α -synuclein aggregation.

15.5.1 1 μ M Cu^{2+} buffer

It is well established that multivalent metal cations like Cu^{2+} can accelerate α -synuclein aggregation[330][316]. To validate our approach and to investigate how metal cations influence the conformer equilibrium of α -synuclein, we performed SMFS experiments on the 3S3 construct in 10 mM Tris/HCl buffer in the presence of 1 μ M CuCl_2 . The low concentration of copper was chosen to target the His

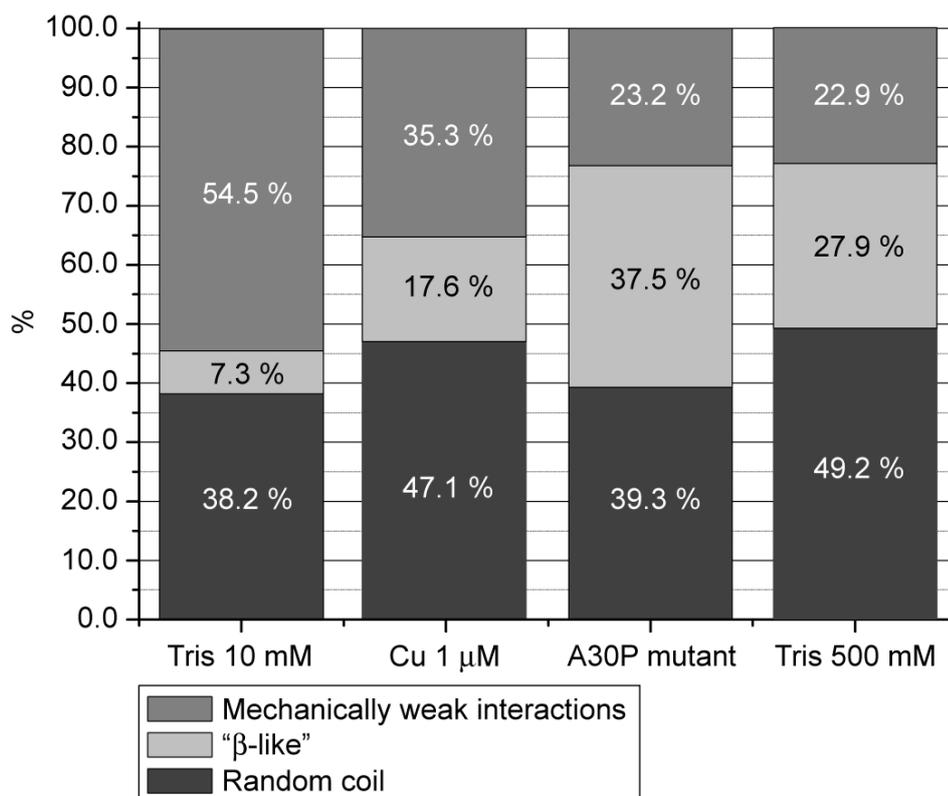


Figure 15.3: Population of α -synuclein conformers in the four different conditions tested in the present work. Percentages observed for each curve type (see 15.1) at 10 mM Tris/HCl ($n = 55$), 10 mM Tris/HCl with 1 μ M Cu^{2+} ($n = 34$), the A30P mutant in 10 mM Tris/HCl ($n = 56$), and 500 mM Tris/HCl ($n = 61$).

50 specific copper binding site of α -synuclein (dissociation constant $K_d = 0.1 \mu\text{M}$) [330].

The presence of 1 μM Cu^{2+} moderately, but significantly (χ^2 statistical significance $p < 0.01$), alters the relative distribution of the α -synuclein conformers with respect to plain 10 mM Tris/HCl (see Figure 2). In particular, the relative population of the β -like conformers more than doubles (from 7.2% to almost 18%), with a parallel decrease of the signals coming from mechanically weak interactions. An increase (from 38% to 47%) of random coil-like curves is also observed.

15.5.2 A30P mutant

The A30P mutation is a pathogenic, naturally occurring human α -synuclein variant, that correlates with familial Parkinsonism[295]. The mutant protein displays an increased rate of oligomerization[298] and impaired degradation by chaperone-mediated autophagy [331]. We tested the 3S3 α -synuclein -A30P construct to evaluate the capability of our methodology to probe different conformational propensities in mutants of the same protein. We found that the A30P mutation induces a striking shift in the conformational equilibrium of α -synuclein with β -like curves being around 37% of the sample and again, a corresponding decrease of signals coming from mechanically weak interactions (see Fig. 15.3). In contrast with wild-type α -synuclein incubated with Cu^{2+} , the A30P mutant does not induce an increase of random coil curves that are exactly in the same proportion observed in wild-type α -synuclein .

15.5.3 500 mM Tris

Another condition known to speed up α -synuclein aggregation is high ionic strength[320][321]. SMFS experiments on the 3S3 wild type construct were performed also in 500 mM Tris/HCl buffer. As reported in Fig. 15.3, the frequency of the three types of profiles radically changed in different ionic strength conditions. The most remarkable result is, again, the significant increase in the population of the β -like structures with buffer concentration (up to about 28%) and the parallel decrease of the percentage of the mechanically weak structures. An increase of random coil curves is also observed, as occurs in the presence of Cu^{2+} , but unlike the case of the A30P mutant.

15.6 Fluorescence spectroscopy shows evidence of I27- α -synuclein interactions

To probe α -synuclein -I27 interactions, ensemble-averaged fluorescence spectroscopy was performed. The autofluorescence of tryptophan residues of the I27 domains, which are absent in the α -synuclein sequence, was used. The fluorescence spectra reported in Figure 15.6 A prove that interactions between the I27 handles and

tracts of the α -synuclein moiety do take place, as shown by the broadening of the spectrum of 1S1 with respect to that of the 3T construct (see 14) and by the 5-nm shift of the λ_{\max} . The possibility of partial I27 unfolding leading to Trp exposure and broadening of the spectrum is ruled out by the CD data and by our force curves, which show that I27 domains are as tightly folded in the 3S3 construct as in an I27 homopolymer. A broadening due to subtle conformational effects on the I27 domain that expose the I27 Trp residue is theoretically possible. However, the same experiments done with I27 and α -synuclein mixed in the solution (instead than being covalently connected) show that this broadening happens only when the α -synuclein moiety is inserted in the construct: only in that case, direct interaction is taking place.

15.7 Circular dichroism of the 3S3 construct

CD spectra of 1S1 and 3T were recorded in which 1S1 shows some α -helical content in the α -synuclein moiety (Figure 4B). . Subtraction of the contribution of the I27 linkers (2/3 of the CD of 3T recorded in the same 10 mM Tris/HCl buffer) from the CD spectrum of 1S1 reveals a profile that is different from that of α -synuclein in the same buffer condition (Figure 4C) but similar to that of the same protein in the α helix structure induced by the addition of SDS [332]. This α -helical content might be induced by the interactions between the α -synuclein moiety and the I27 domains (see Section 16.4)

15.8 The distribution of mechanically weak interaction lengths is differently affected by variations in the environment or protein sequence.

The mechanically weak interactions were characterized by measuring the distance between the contour length of events attributed to them and the contour length measured on the first I27 peak in the same curves, both obtained by a double-parameter WLC fit. This value should correspond to the chain segment length

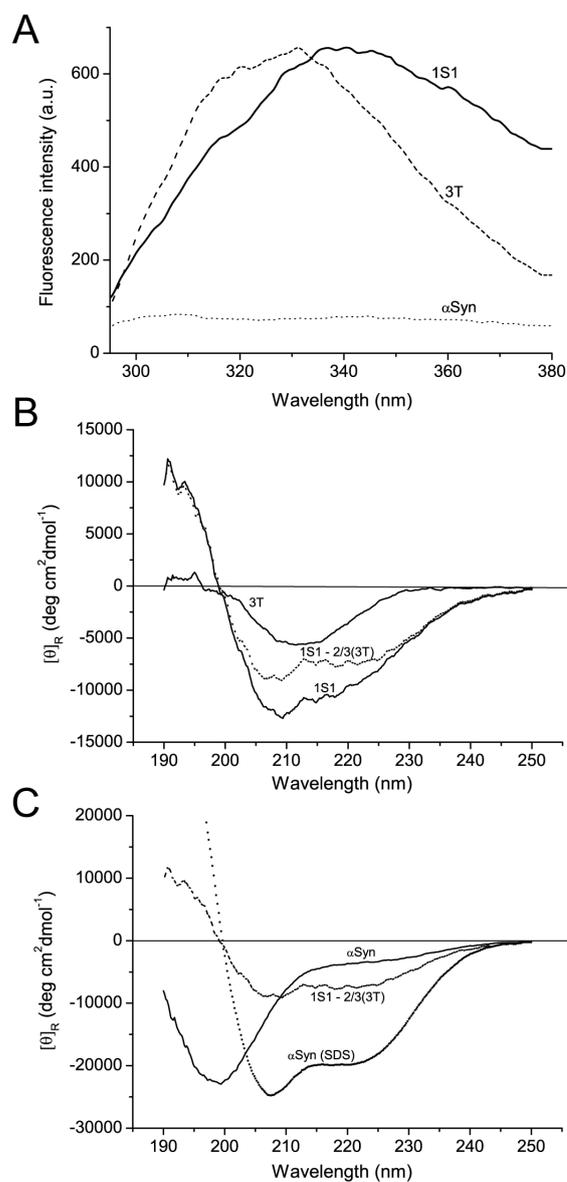


Figure 15.4: A) Fluorescence spectra of α -synuclein , 3T, and 1S1 (dotted, dashed, and solid line, respectively). B) Circular dichroism spectra in PBS buffer of 3T, 1S1 (solid lines). The α -synuclein contribution in 1S1 (dashed line) is calculated by subtracting the relative contribution of the I27 domains from the CD spectrum of 1S1. C) CD spectra of α -synuclein in PBS (solid line) and 250 mM SDS (dotted line). The α -synuclein contribution in 1S1 (dashed line) is reported as in (B).

enclosed by the mechanically weak interactions. In case of multiple interactions, often only the larger one was measured (the smaller was often too weak and irregular to obtain a reliable fit).

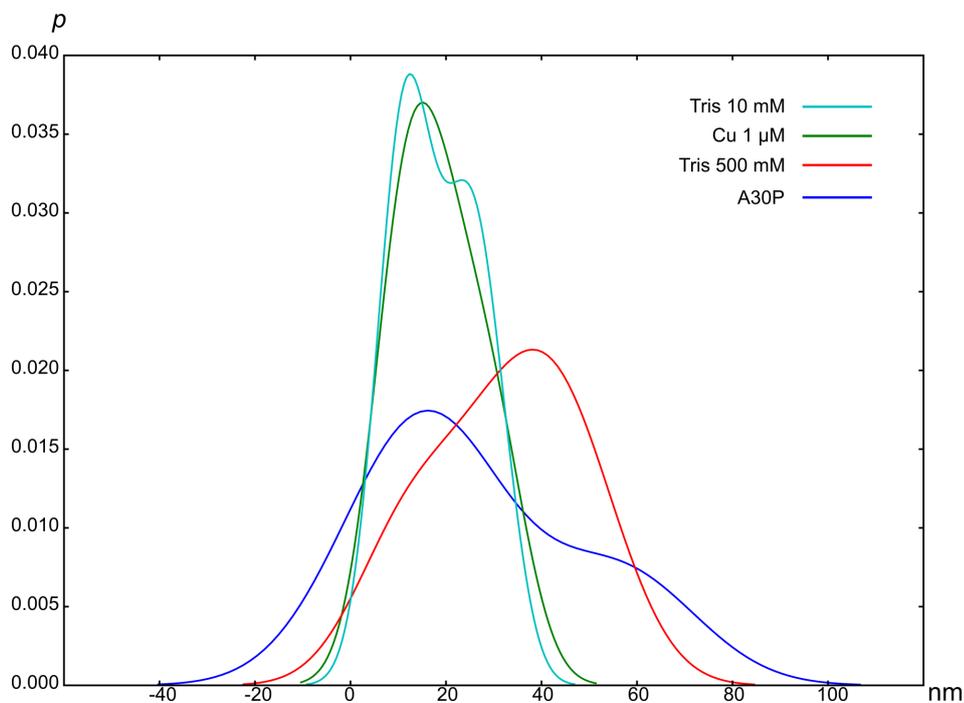


Figure 15.5: Gaussian kernel density estimation of the mechanically weak interactions lengths under different conditions. Briefly, a Gaussian function (kernel) has been centered on each data point for each data set. The sum of the kernels, normalized to have unitary integral, is the KDE plot. Kernel bandwidth h (*i.e.* the Gaussian kernel standard deviation) was automatically calculated for each data set such that it minimizes the asymptotic mean integrated square error: $h = \sigma \left(\frac{4}{3n}\right)^{\frac{1}{5}}$, where σ is the standard deviation of data and n is the size of the data set. The plot has been calculated using Statistics for Python (<http://bonsai.ims.u-tokyo.ac.jp/mdehoon/software/python/Statistics/>).

The larger set of mechanically weak interactions events is that obtained on Tris 10 mM ($n = 19$). Data recorded in other conditions show a significant decrease in the occurrence of mechanically weak events. The resulting data sets are too small for a significant statistical comparison to be made for each condition, but general trends may be inferred.

We tried to understand the underlying trends by using kernel density estimation (KDE)[333] to obtain a sketch of the probability density function of interaction

lengths. This refined method does not suffer of biases in the choice of binning boundaries and allows for a more objective visualization of small data sets than a classical histogram. The KDE plot (Figure 15.8) of the interaction lengths found in Tris 10 mM and the Tris 10 mM + 1 μ M CuCl₂ for mechanically weak interactions events are almost perfectly superimposable, despite the difference in the size of the data set ($n = 5$). The larger number of data in plain Tris 10 mM allows nonetheless to distinguish a bimodal distribution distinguishing mechanically weak interactions enclosing loops of 11 and 25 nm. Experiments conducted with 500 mM buffer concentration ($n = 8$) and the A30P mutant ($n = 6$) show instead a broadening of the KDE that may contain a shift from short- to long-distance interactions.

CHAPTER 15: RESULTS

Chapter 16

Discussion

16.1 Conformational equilibria of α -synuclein as probed by AFM-based CS-SMFS

The first clear, striking evidence coming from the CS-SMFS experiments is that stretching of α -synuclein molecules in the same sample, conditions and experiments generates three (possibly four) discrete classes of signals. These classes correspond to vastly different mechanical behaviours (from random coil behaviour to β -sheet-like behaviour) and, correspondingly, to different underlying α -synuclein structures¹. It is most probable (and safe) to describe these signals as coming from different conformational clusters, each displaying different conformations, rather than coming from a few distinct conformers. In fact, as described below each class of conformers displays measurable diversity.

The properties of the α -synuclein bulk of molecules therefore emerge from an underlying conformational equilibrium between discrete conformational clusters. The existence of a complex conformational equilibrium in a classical IUP like α -synuclein was expected, but their measurability by mean of the AFM-CS-SMFS

¹Technically, it can be argued that such diversity comes from a vast diversity of unfolding pathways underlying a much more narrow set of structures. However, despite the recent findings of the existence of multiple unfolding pathways[334][335], such a clear partitioning in discrete classes was never clearly observed before in the unfolding of folded proteins. Given the unique conformational properties of α -synuclein, it is therefore safe to attribute the differences to really different conformations.

technique means that these conformers have significant structure and a considerable life time (of the same range of that measured by the instrument: 10^{-3} - 10^0 s). However, see also Section 16.5 for a discussion on the effect of I27 handlers.

Most importantly, these conformers are not dispersed as a broad multitude of flexible conformations, but clearly divide themselves between a few discrete structural classes. This is a novel concept in the study of IUPs, although a similar scenario was recently described by molecular dynamics simulations on the $A\beta$ Alzheimer peptide[130]. From the point of view of folding biophysics, this result is of theoretical interest. If our structural interpretations of conformational classes hold correct, α -synuclein is not only a protein sequence able to assume practically all possible structural states of protein (unfolded, α -helical monomer, β -sheet monomer, amyloid polymer), depending on particular conditions: given appropriate conditions, the α -synuclein ensemble can populate *all these states at the same time* in solution. α -synuclein therefore explores a peculiar conformational energy landscape, which contains the standard energy minima associated with protein folding, with barriers between those states below kT . These energy minima and/or energy barriers, in turn, shift rapidly in position and magnitude depending on environmental conditions or subtle sequence variants. α -synuclein can therefore prove to be a useful model system for the investigation of folding.

16.2 The population β -like conformers is correlated with the α -synuclein propensity to pathological aggregation

We observed a marked increase of the population of β -like conformers under three very different conditions known to accelerate α -synuclein aggregation. This result links the population of those α -synuclein monomeric conformers to the process of α -synuclein aggregation. The first condition is the presence of a μM concentration of Cu^{2+} . Our results in this condition agree with the observation of a metal-induced partially folded intermediate[134][300]. Also Rasia et al. suggested a compact set of metal-induced conformations, noticing that the specific binding of Cu^{2+} to the α -synuclein N terminus requires the formation of a metal-binding interface (pivoted on His 50), which possibly involves residues that are widely separated in

the primary amino acid sequence[330].

The second condition is the A30P mutation. Nuclear magnetic resonance (NMR) experiments have observed a much more flexible average conformation of the α -synuclein mutants A30P and A53T. The increased average flexibility of α -synuclein allows the protein to sample a larger conformational space[319]. Interestingly, the mean hydrodynamic radius of α -synuclein is not affected by the A30P and A53T mutations[297][81], thus showing that the increased flexibility is compatible with the population of compact folded structures like those singled out by our experiments.

The third condition is a radical increase of the ionic strength. Our results in 500 mM Tris/HCl can be reconciled with the model proposed by Hoyer et al.[320] and by Bernado et al.[321] to explain the well-documented phenomenon of the increased α -synuclein fibril formation with increasing ionic strength. According to that model, the increased fibril formation is explained just on the basis on an increased freedom of the fibrillogenic NAC region caused by the release of its interaction with the negatively charged C-terminal tail. The increased ionic strength of the buffer leads to a more efficient charge shielding of the strongly acidic C-terminal tail, thus relieving its electrostatic self-repulsion. This in turn leads to the lowering of the protein-excluded volume and increases its flexibility. According to our data in Figure 2, we should add to this model a shift of the conformational equilibrium toward the β -like structures that takes place on increasing the charge shielding.

16.3 Are β -like structures really β ?

16.3.1 Experimental evidence

The 200-pN unfolding force of all the seven peaks indicates that the folded section of α -synuclein has the same mechanical properties of the I27 β -sandwich structure. At the moment, without any independent structural characterization, we consider and label this folded structure of the α -synuclein moiety just as β -like, in accordance with its mechanical behavior. Nevertheless, its mechanical behavior is in agreement with a β -sheet content in the β -like class of conformers. It is unlikely that the α -helical content we observed by means of circular dichroism (CD) (see

below) correlates with the β -like conformers. In fact, whereas β -structures, like those of titin modules, such as I27, or tenascin, unfold at forces in the range of 100-300 pN (at loading rates of the same order), the α -helix domains, in the same conditions, are always observed to unfold at forces almost one order of magnitude smaller (see 6.5). We thus have labeled as β -like those conformers with a mechanical behavior closely matching those of structures rich in β -sheet content.

16.3.2 Evidence from the literature

The correlation of the population of these structures with aggregation conditions, which enrich β -sheet content in α -synuclein, supports this labeling. Evidence of some β -sheet content in the monomeric state of α -synuclein was previously reported in the literature. Most recently by means of NMR spectroscopy in supercooled water at minus 15 °C, it was found that the α -synuclein chain, cold-denatured to an hydrodynamic radius equivalent to that displayed by the same protein in 8 M urea, retains a surprising amount of unpacked β strand content that correlates with the amyloid fibril β structure[318]. The packing of these β strands into compact structures like those observed by us is thus likely to occur in non-denaturing conditions and at physiological temperatures. This NMR result supports the observation of β -like conformers in the monomeric state of α -synuclein and links them to the amyloid β structure. The presence of β sheet structures was indicated also by Raman spectra of this protein in aqueous solution[332]. In the same investigation, CD spectroscopy proved unable to detect any β content. Correspondingly, the CD spectra of α -synuclein recorded by us in 10 mM and 500 mM Tris were practically superimposable. We conclude that CD is not a technique sensitive enough to detect partial β -sheet content in the α -synuclein sample. A fraction of β -sheet /extended structure of about 19% was also detected, again not by CD, but by Fourier transform infrared (FTIR) spectroscopy in dried films of α -synuclein [292]. This fraction is much larger than that estimated by our experiments in 10 mM Tris/HCl buffer (see Figure 2). However, the conditions of the SMFS and FTIR experiments were markedly different, and in the latter case, some template-mediated formation of β structures due to the packing of the α -synuclein molecules in the dried films required by the FTIR measurements cannot be ruled out.

In conclusion, despite the fact that force spectroscopy data cannot directly assign a specific secondary structure to the conformers we have labeled as β -like, it is most likely that they have significant β sheet content. Further investigation will be needed to consolidate this point.

16.4 The nature of mechanically weak interactions

By now, any structural characterization of the mechanically weak interactions events monitored by the small peaks in force curves as in Figure 15.1 C is at best tentative and falls outside the focus of the present work. A more detailed characterization of these events is, however, within the range of capabilities of the techniques proposed here and is being currently addressed in our laboratory.

16.4.1 An electrostatic model of I27- α -synuclein interaction

A plausible explanation of the short- and long-distance mechanically weak interactions we observed cannot exclude the interaction between positively charged residues on the α -synuclein N-terminal and the negatively charged surface of I27 modules[263]. It has been documented that α -synuclein in contact with negatively charged surfaces assumes an α helix structure[306][305][303]. We might expect a similar structural transition in the α -synuclein moiety also from the contact with the I27 modules within the 3S3 or 1S1 constructs. This transition is indicated by the CD spectra of the 1S1 construct in 10 mM Tris/HCl (see Figure 15.6 B). We propose that the small peaks like those shown in Figure 1D and assigned to the mechanically weak interactions can be the signature of the interaction between α -synuclein, possibly in α helical form, and the flanking I27 modules. It is not surprising that more than one of those signals are present in the same force curves, because multiple interactions of this type can occur at the same time in the same molecule. It should be noted that the same transition does not take place when free α -synuclein is mixed in solution with I27 modules of the 3T construct.

Considering electrostatic and hydrophobic features of the I27 and α -synuclein

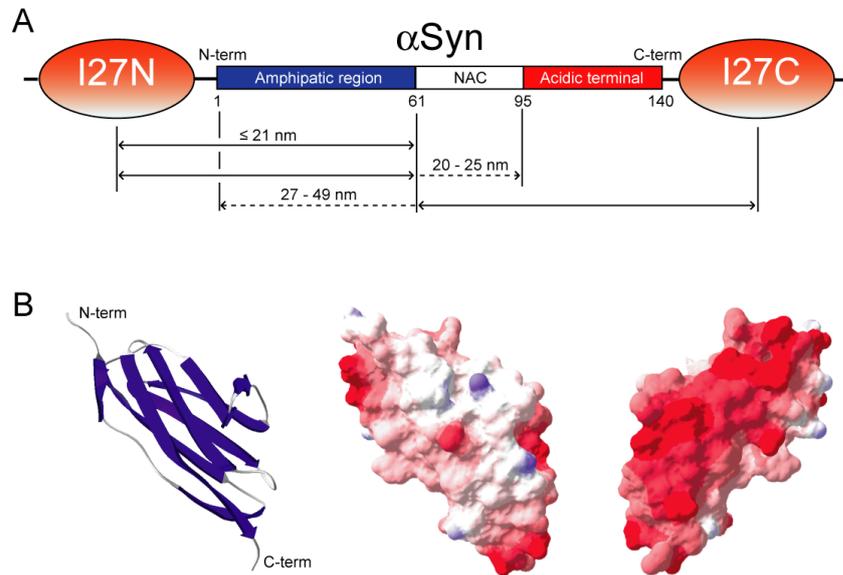


Figure 16.1: An electrostatic model that may explain mechanically weak interactions. (A) Schematic representation of the central portion of 3S3 construct, evidencing the I27 modules flanking the α -synuclein element (named I27N and I27C, see text). In the α -synuclein moiety three regions are evidenced: i) the amphipatic region, prone to fold in α -helical structures when in contact with phospholipid membranes; ii) the fibrillogenic NAC region, characteristic of the fibrils core of α -synuclein amyloid; and iii) the acidic C-terminal tail, strongly charged and not prone to assume any fold. The reported quotes correspond to interactions that may lead to the small peaks observed in curves featuring mechanically weak interactions (see text). The colors of the different regions depict the electrostatic potential: red for the negatively charged I27 (see text) and for the acidic C-terminal region of α -synuclein ; white for the hydrophobic NAC region of α -synuclein ; and blue for the positively charged amphipatic region of α -synuclein . (B) Cartoon (left panels) and solid surface representations of the electrostatic potential (central and right panels) of titin I27 domain obtained using the program DeepView. I27 coordinates were taken from the Protein Data Bank (PDB code: 1WAA). In the cartoon, the secondary structure elements are colored in blue for β -strands. Surfaces were colored according to the calculated electrostatic potential contoured from $-5.0 kT/e$ (intense red) to $+5.0$ (where k = Boltzmann constant, T = absolute temperature, and e = electron charge) (intense blue). The molecular orientation in the central panel is the same as that in the cartoon (left panel), whereas that in the right panel is rotated by 180 about the vertical axis.

domains, we can try, very tentatively, to associate the mechanically weak interactions with those of different domains within the 3S3 construct. We can sketch and divide the overall structure of α -synuclein into three main regions: the amphipatic N-terminal tail (aa 1-61), positively charged because of its Lys residues; a hydrophobic central (aa 61-95) region containing the fibrillogenic NAC segment; and the highly negatively charged acidic C-terminal tail. In the 3S3 construct the α -synuclein N- and C- termini are flanked by partially negatively charged I27 modules (Figure 16.1), that we label I27N and I27C, respectively. Contacts between the negatively charged region of I27N and the positive α -synuclein N-terminal would appear as peaks that correspond to short-range distances along the primary structure: lower than or equal to 21 nm. The peaks in the middle-range distance of 21-33 nm, and those in the long-range distance of 27-49 nm might correspond to contacts of I27N with the central region of α -synuclein, and to contacts of the α -synuclein N-terminal with either the I27C or the α -synuclein C-terminal, respectively.

16.4.2 The influence of sequence and chemical conditions on I27- α -synuclein interactions

If the interactions are electrostatic in nature, we expect a marked difference in their population and characteristic by changing the ionic strength, thus increasing electrostatic shielding. In fact, the increase of the electrostatic shielding with the buffer concentration has two main effects. First, the population of the mechanically weak interaction featuring structures decreases from about 55% to 23% (see text, Fig. 2), most likely because the interactions with the I27 domains that drive α -synuclein towards α -helical structures are less favored. Second, the rigidity of the C-terminal tail is decreased as also evidenced by the shift from the short- to the long-distance intermolecular interactions with the increase of the buffer concentration (see Supplementary Material Figure 3). This reduction of rigidity is associated to an increased conformational freedom of the whole α -synuclein moiety, and most importantly of its NAC region. The intertwining of both effects can shift the conformational equilibrium towards the β -like structures. In fact, their population increases from about 7% to almost 30% on passing from 10 to 500mM Tris/HCl buffer (Figure 15.3).

Notably, these short- and long-distance mechanically weak interactions are observed to be in equilibrium with the β -like structures. The population of the former always decreases while that of the latter increases. This result is in accord with the observation by Zhu et al. that a driving force to α helical structures inhibits α -synuclein fibril formation[290] and also rule out any template-mediated β -sheet imprinting by the I27 linkers. This conclusion is confirmed by the data on 500 mM Tris/HCl buffered solutions showing that when electrostatic interactions between the α -synuclein moiety and the flanking I27 linkers are decreased, the population of β -like conformers increase. We can also expect entropic effects due to the presence of the flanking I27 domains to drive the protein toward more extended conformations rather than compact conformations[175][38].

16.4.3 Other possible structural assignments

Again, the structural assignment of these signals to helical structures is only a working model. It is true that circular dichroism shows evidence of helical structures (15.7) and fluorescence of I27- α -synuclein interactions. However, there is no data correlating these findings with the mechanically weak interactions, apart from the electrostatic model. While strong peaks from the β -like curves have only been related to β -sheet -rich structures, this does not hold for smaller (< 100 pN) peaks. On the other hand, it is possible that the detaching of small helical portions yields only very weak signals, and are therefore invisible to the CS-SMFS. If this is the case (and currently we have no way to exclude it), others possibilities must be taken into account.

A different, possible explanation is that these signals come from the unfolding of other β -sheet structures that are less folded and/or less mechanically resistant than the β -like one, maybe being precursors of them. The decrease of the mechanically weak interactions correlated with the increase of β -like structures could indicate both that these structures giving rise mechanically weak interactions are on the folding pathway to β -like structures, or that they instead compete. It is however difficult to correlate such a hypothesis with the differences found in the length distribution of these structures in different ionic strengths, which is instead well explained by the electrostatic model.

It is also possible that mechanically weak interaction events contain the sig-

nature of the unlocking of α -synuclein intramolecular interactions. These latter interactions can sustain the fairly compact structures reported by many authors already[302][134][301]. It must be noticed this is less likely due to the weak and “fluid” nature of these interactions, that are even more unlikely to give clear peaks on the force curves than α helices. The pattern of those interactions should also be disrupted by the inclusion of α -synuclein in the construct: the electrostatic repulsion between the negatively charged I27 modules should hinder the long-range contacts between the C- and N-terminal domains proposed by several authors.

16.4.4 Multimodular constructs for the measure of inter-protein interactions?

If the assignment of these signals to I27- α -synuclein interactions is confirmed, this would represent an unexpected methodological side effect of this study. This would represent, to the best of our knowledge, the first time a SMFS experiments exploits the polyprotein strategy to measure interaction *between protein modules* instead of interactions intrinsic to the protein modules themselves.

Such a methodology could have interesting advantages. The two proteins of which the interaction is analyzed would be in a known orientation in the construct, and their geometry would be constrained by their embedding in polyprotein. This could allow to gain finer structural information from protein-protein interaction assays. Using alternative linking strategies (for example, by using disulfide linking), alternative orientations, and thus interaction between alternative interfaces, could be explored.

An example application of such a technique could be the analysis of the interaction between two amyloid-forming proteins. Using a polyprotein construct containing *two* analytes in tandem, one could look for additional signals beyond those coming from the unfolding of the analytes, indicating protein-protein interactions, and for structuring of the analyte modules subsequent to their being in close contact.

16.5 Effect of the I27 handlers

When discussing the results obtained by mean of the polyprotein approach for SMFS (6.4), one must never forget that this is a highly non-native condition for the protein, especially if the analyte module was not evolved to fold correctly into a multimodular protein.

For α -synuclein this is even more problematic, given the extreme sensibility of α -synuclein structure to environmental conditions. This does not mean that it is impossible to extract meaningful information on the properties of the protein in solution. While the conformational equilibria of α -synuclein as measured in the 3S3 in a given physico-chemical environment are hardly representative of the behaviour of α -synuclein in solution, the *comparison* between different conditions can be informative. In our case, conditions leading to α -synuclein aggregation consistently lead to a dramatic increase of the β -like structures. It is unlikely that the equilibrium of these structures is quantitatively similar in the 3S3 construct and in the free protein in solution. Nevertheless the identification of a conformer correlated, also structurally, with α -synuclein aggregation is a strong indicator that α -synuclein, in aggregation-prone conditions, is significantly shifted towards a β -sheet - rich phase.

The presence of bulky objects at the N and C termini of the α -synuclein module has profound consequences on the conformational behaviour of α -synuclein. Volume exclusion effects, combined with the electrostatic repulsion of the I27 modules, could lead to constant pulling on the α -synuclein module with respect to conditions in solution[38][175]. The steric and electrostatic repulsion of I27 modules hinders also the conformational space of α -synuclein N and C termini. In general, the presence of the handlers should therefore favour more extended conformations.

On the other hand, the presence of I27 modules mimics a strong molecular crowding. Crowding alters the conformational properties of α -synuclein, leading to more fibrillation-prone states[317][336]. This means that the equilibrium probed in this work could be much more shifted towards fibrillation-prone structures by default.

A recent study also shows that crowding slows down proteins' intramolecular motions[8]. This finding is notable, because it could mean that the I27 mod-

CHAPTER 16: DISCUSSION

ules not only provide linkers and fingerprints, but, increasing the lifetime of the α -synuclein conformers, allow us probe their conformational equilibria “in slow motion” , while the interchange between structures would be too fast if equivalent to that in solution. While many techniques can probe fast ($< 10^{-3}$ s) timescales, probing slower timescales is more difficult[8]. AFM-SMFS on the contrary can only probe timescales well longer than the sampling interval (usually $> 10^{-3}$ s).

These considerations indicate that the design and use of alternative linkers or experimental strategies may prove useful in the future to further discriminate the effective conformational distribution of α -synuclein from alterations due to the interaction with the linkers.

CHAPTER 16: DISCUSSION

Chapter 17

Conclusion and perspectives

17.1 SMFS for the intrinsically unfolded proteins

For the first time, to our knowledge, we applied the AFM-based single-molecule mechanical unfolding methodology to a multimodular protein containing the α -synuclein moiety. This approach brings into play three main methodological capabilities inaccessible to the bulk ensemble-averaged spectroscopies previously applied to study the structure of α -synuclein and other natively unstructured proteins:

- The possibility to work strictly at the single-molecule level, thus ensuring that the conformer distribution of the monomeric α -synuclein is detected and quantified without interference from oligomeric soluble forms of the protein and therefore of any possible intermolecular imprinting toward the amyloidogenic β structures.
- The ability of spanning the conformational space of the protein under investigation and of directly catching and quantifying all of its conformers with a lifetime longer than $< 10^{-3}$ s. These conformers, because of their longer life time, might be the most biologically relevant. Three classes of the monomeric α -synuclein conformations, including random coil, mechanically weakly folded and β -like, were characterized by our experiments. They could be detected even in low concentration without the necessity of selectively enhancing one of them by adding specific agents to unbalance the con-

formational equilibrium, as most commonly done so far with bulk ensemble-averaged experiments.

- The possibility of following shifts in the population of these classes of conformers in response to changing the solution conditions or the protein sequence and to detect them even if scarcely populated. In the case of α -synuclein, conditions known to promote oligomerization and aggregation -like the presence of Cu^{2+} , the A30P mutation, or a radical increase of ionic strength- markedly shift its conformational equilibrium toward the β -like form at the expense of other structures. These results indicate that the β -like curves contain the signature of structural precursors to α -synuclein oligomerization. We suggest that the different aggregation propensities and, ultimately, the pathogenicity displayed by α -synuclein under different environmental conditions or point mutations can be triggered by unbalancing the delicate equilibria among α -synuclein conformers.

These capabilities can be applied also to other proteins with similar structural properties. The human prion protein, the Alzheimer's $A\beta$ peptide and *tau* protein are just the more prominent examples. In the near future, single-molecule methodologies could play a crucial role in studies of the folding equilibria of the IUP monomers and, in particular, in the detection and quantification of the conformers that can lead to aggregation of those proteins.

17.2 AFM-based SMFS as a structural technique

In this work, moreover, SMFS has for the first time been used as a structural technique. SMFS at the current state can hardly discriminate secondary or tertiary structures (with the possible exception of discriminating between β -rich and non- β structures), but has proven able to discriminate between conformational classes of a complex conformational ensemble.

In the future, it is possible that SMFS will permit also basic structural characterization. The long-sought (and yet to be fully achieved) combination of SMFS and SM-FRET techniques on the same molecule can allow to visualize in real time the separation between known residues and the mechanical response of a protein,

allowing to model the unfolding pathway and to infer structural properties of the unfolded conformer.

17.3 Testing therapies against proteopathies by mean of SMFS

Our results suggest the feasibility of single-molecule approaches to the testing of novel pharmacological or biophysical therapies for pathologies involving the conformational equilibria of IUPs.

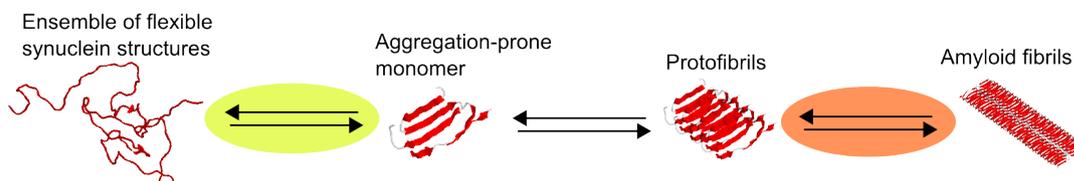


Figure 17.1: Simplified diagram of α -synuclein amyloid aggregation, with equilibria being possible targets for therapies. The equilibrium between protofibrils and fibrils is currently the one most available to investigation, but it should be avoided as a target, since it could bring to the formation of toxic species. The equilibrium between monomeric forms instead, probed by mean of single-molecule techniques, would keep the protein in its functional state, and avoid the formation of both protofibrils and fibrils.

The work herein reported is, in fact, an addition to the very few techniques that can hope to spot the appearance of aggregation-prone conformers before aggregation actually ensues. This is of medical interest, because there is still considerable debate on what is the actual toxic phase in amyloid-related proteopathies, the oligomeric protofibrils or the fully formed amyloid fibril (4.3.2). Disrupting the fibril is likely to form a plethora of small, oligomeric species which can be more toxic than the fibril itself. In the words of Vladimir Uversky, “Halting the fibrillation process at early stages associated with the formation of such small oligomers can create more harm than cure, as these species are neurotoxic rather than neuroprotective” [289].

CHAPTER 17: CONCLUSION AND PERSPECTIVES

By using the approach herein reported, one can in principle test the effect on α -synuclein conformational equilibria of a number of pharmacological agents, selecting the ones that inhibit the formation of conformers correlated with aggregation. However, a feasible application of this technique to mass-screening of potential drugs requires a substantial improvement in the speed of data harvesting, which currently requires from a few weeks to a few months, depending on the construct sample purity and the kind of statistics to be built.

Part V

**Hooke: an open source software
platform for SMFS data analysis**

Control over the use of one's ideas really constitutes control over other people's lives; and it is usually used to make their lives more difficult.

Richard Stallmann

Chapter 18

Background

A little more than 10 year old, force spectroscopy is still in its infancy. This is evident in the theoretical understanding of the technique (which is far from being complete: for example, only recently is arising that the long-held Evans model[165] is only an approximation of deeper, more complex models[192]), with serious consequences for energy landscape reconstruction) in the experimental data interpretation (see 6.8 for example) and in the design of appropriate constructs (see 6.4), as examples.

This is also evident in the more mundane tasks of the SMFS experimenter. SMFS often requires the interpretation and analysis of thousands of raw force curves. The analysis, in turn, relies on the recognition of features and the fitting of these features to physical non-linear models. It is self-evident that SMFS is essentially a computer-dependent technique. Despite this obvious truth, there is considerable immaturity of informatic tools as applied to SMFS.

The current situation suffers of the following general problems:

- **Data format balkanization.** There are currently a number of manufacturers building AFM equipment for SMFS. These manufacturers all use different and incompatible data formats, which are often poorly documented and relatively difficult to reverse engineer. No single data format has ever emerged as a standard (much less an open, freely documented standard) for SMFS data.
- **Poor quality of commercial software.** Each AFM manufacturer usually releases its own software for microscope control and data analysis. However

the capabilities of such software, even if helpful, are often limited and not easily extensible by the user. In general, the user cannot expect to use the vendor-provided software for serious, custom data analysis.

- **No standard software platform.** While AFM imaging can rely on standard, good quality software like WSxM[337] and Gwyddion (<http://www.gwyddion.net>), a single, standard software platform for SMFS data analysis has not been developed. There have been a few attempts ([338] , SPIP (<http://www.imagemet.com>) PUNIAS (<http://site.voila.fr/punias/klmenu/punias0.htm>) but none of them has been widespread.
- **No open and/or extensible software platform.** Each SMFS experiment requires its own kind of data analysis. This in turn requires that the user must be able to script or extend the data analysis software to his own needs. Today, no SMFS software allows this kind of flexibility
- **No standard library.** However, data analysis often (even if not always) relies on a set of standard operations (*e.g.* identifying peaks, WLC fitting, finding the contact point) which are then combined together for the needs of each experiment. However, no standard library or algorithm for these needs emerged in the SMFS community, even if there have been a few studies on that[198].

The problems listed above usually result in a continuous “reinventing the wheel” in the SMFS community. Each laboratory implements its own quick-and-dirty scripts that on one hand all do the same things (reading the data format, fitting data, etc.) and on the other hand are often strictly bound to their immediate needs and are often hard to adapt to new experiments with different requirements or new data formats. This situation is frustrating because a lot of work is unnecessary duplicated and each laboratory has to lose time writing its own tools.

Chapter 19

Design and implementation

When I had to rewrite my data analysis scripts again, due to new needs arising, I decided to shift from the usual quick-and-dirty scripting towards the developing of a more complex, extensible application.

A data analysis software was (and is currently) developed by me, called *Hooke*¹. It has currently been developed on GNU/Linux, and is known to run on Linux and Windows. It should also run without modification on Apple OS X, but this has not been tested.

Hooke has become the default data analysis application in our laboratory for force spectroscopy, and has been already used for research published in high-impact journals[141][339]. The software is still actively developed, “rough around the edges” and will be published online when the documentation will be complete.

19.1 An open approach to SMFS data analysis

19.1.1 A platform for SMFS data analysis

Given the need in SMFS analysis, of both shared foundations (operations and utilities that are of general interest for the SMFS community) and of strong flexibility (each experiment requires its own analysis tools), an interesting approach could be not that of implementing a *tool*, but that of implementing a *platform* on the top

¹The name was chosen in honour of Robert Hooke (1635-1703), the scientist who coined the word “cell”, microscopist and one of the first investigators of elasticity. The name *Hooke* can also refer to the capability to “hook” functionality to the software backbone.

of which tools are integrated and shared. This way, both problems are solved at the same time: a shared collection of tools that can avoid “reinventing the wheel” over and over, and a flexible way to build new tools, using the existing ones or adding completely new ones.

Therefore, the main tool should be made of a simple, slim *kernel* providing only the necessary infrastructure and essential, general features, and an *interface* on top of which tools can be built and easily plugged-in by the user.

19.1.2 The Open Source philosophy

How should these tools accessed and exchanged? Science is (at least ideally) an open process, where information is openly and freely exchanged by participants and where each result is built on the foundation and improvement of the precedent ones. There is no reason to not apply a similar approach not only to science itself, but also to the tools used for science. In fact this approach has been used for non-scientific software development, under the name of Free Software and/or Open Source[340]. In the open source approach, the details of the software, to the level of source code, are available to the public and can be freely exchanged and improved. This way, each user can improve the software or extend it to his/her own needs. The improvements can be released again to the public, thus creating a virtuous circle in which the software is continuously extended and improved by mean of collaboration.

Open source, albeit utopistic in its philosophy, proved surprisingly and exceptionally practical and successful and today plays a significant part in modern computing infrastructure, including the Internet².

Therefore, the best way to ensure the tools and the platform remain valuable for the community in the long term and can be extended and improved by everyone to every need is to open them and allowing free modification and redistribution of these tools by the whole SMFS community.

²As an additional example, all software used to write this Ph.D. thesis, including the operating system on top of which these softwares run, is free and open.

19.1.3 Platform independence

A poorly confronted problem is that of platform independence, that is: the software should not tie the user to a defined operating system hardware architecture, especially to commercial ones. The choice of the platform to use should therefore be independent from the choice of the software.

In a Windows®-centric world, that of platform independence could seem a minor problem. However the Windows platform (like everyone else) is not optimal for all situations. Requirements due to other softwares, security or simply user habits and preference could make the availability of a tool running on other operating systems, such as Apple®OS X or GNU/Linux, desirable.

19.1.4 Independence from commercial software

Most SMFS data analysis tools and scripts so far often rely on commercial scientific software platforms, like Matlab®, Mathematica® or IGOR®. This used to offer several advantages in terms of reliability and features, but has also specific disadvantages. Commercial licenses for these platforms are usually expensive, and there is no guarantee that programs developed with a previous version of the platform will run on future versions of the platforms, nor that the current version will run on future operating systems and machines. This forces the user to a “Red Queen’s race”³ situation: the user has to constantly buy new versions of the commercial platform just to ensure his own software will continue to run.

Moreover, there is often little compatibility between the platforms, and this is a strong obstacle to sharing: one used to one software platform (say, Matlab®) will not buy another expensive commercial software just to run software developed on another platform (say, IGOR®).

Finally, these platforms are usually very easy to program for quick, one-purpose data analysis requirements, but being not designed as general-purpose program-

³The Red Queen’s race is an incident appears in Lewis Carroll’s *Through the Looking-Glass* and involves the Red Queen and Alice constantly running but remaining in the same spot:

“Well, in our country,” said Alice, still panting a little, “you’d generally get to somewhere else if you run very fast for a long time, as we’ve been doing.”

“A slow sort of country!” said the Queen. “Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!”

ming languages, the construction of complex tools with these programs is usually difficult and hard to maintain.

Today however a number of high-quality, multiplatform and most importantly free and open tools exist that have capabilities of the same level of commercial packages. These tools can take the shape of free analogues to the commercial tools described (like Octave and Scilab, that aim to be free replacements for Matlab®) or of libraries made to be used by general-purpose languages like C, Fortran, Perl or Python, like the BLAS[341] and LAPACK[342] scientific libraries. Ironically enough, many commercial packages actually rely on these free libraries for their working.

Apart from highly specialized applications, it is safe to say that there is no more the need to build data analysis tools on expensive, incompatible commercial software platforms. Free programming languages and libraries can well be used for this need, without significant losses on features and power.

19.2 The *Hooke* implementation

Hooke has a very simple interface, made of a command line and a window showing the plot. This design was chosen because it is very common in several bioinformatics programs, like RasMol and PyMol[343], and allows high flexibility and extensibility. Writing a command-line function is also much less onerous on the programmer than writing complex graphical user interfaces, and this has to be taken into account for easy,fast plugin writing.

By itself, *Hooke* is a program made to manage and visualize data sets made of multiple independent data files encoding one or more 2D plots. These files are organized by the operator in a *playlist*, analogue to the functioning of a music player. Data files can be visualized, exported as text files or images, and annotated by the operator. Most other *Hooke* functionality, included functions able to read data files, is encoded by separate plugins.

19.2.1 Choice of the programming environment

Hooke is written in Python (<http://www.python.org>). The choice was made depending on a variety of advantages:

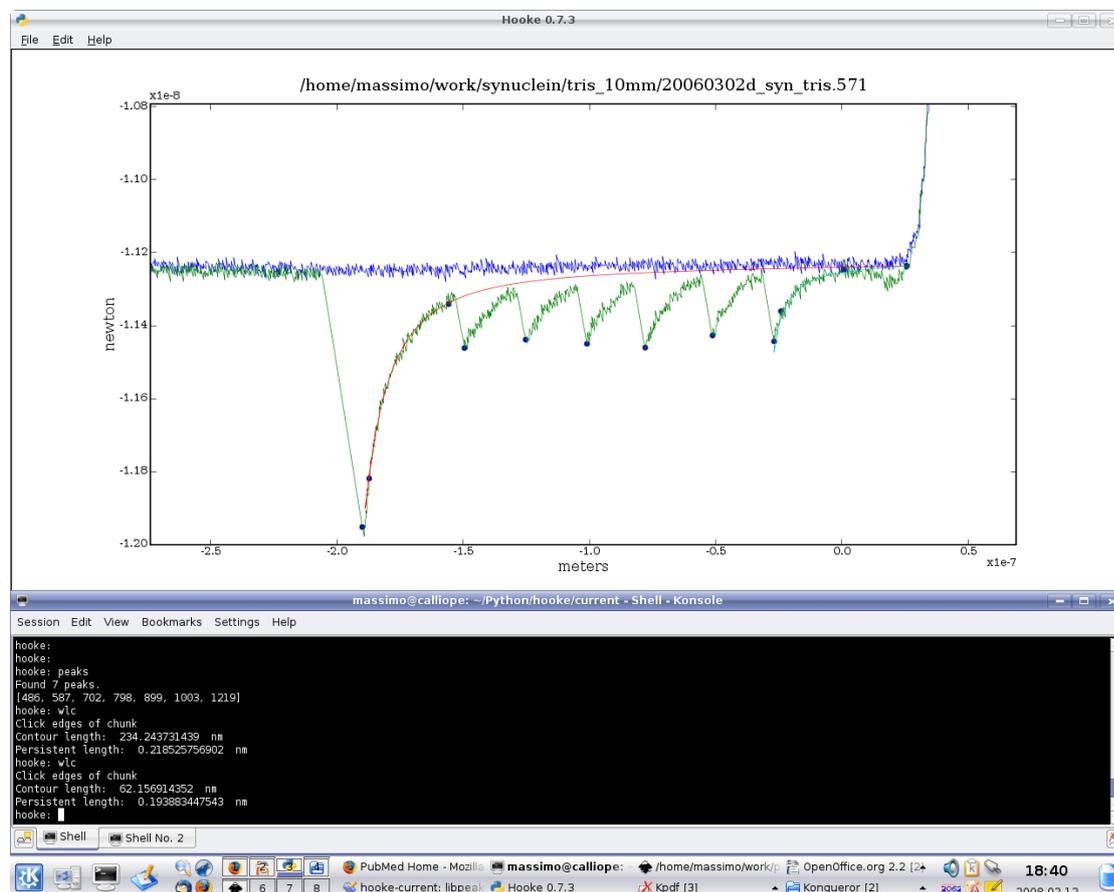


Figure 19.1: A typical *Hooke* session, from the author’s GNU/Linux desktop. The top window shows the plot, the bottom window is the *Hooke* command line session. In the image, the **peaks** command validates the convolution algorithm (see 19.3.1): the blue dots on the tip of each peak have been recognized as peak maxima by the command. Other blue dots mark the contact point and the intervals chosen by the user for two WLC fits (red and cyan lines)

- It is open source and available for free.
- It is object-oriented
- It is extremely portable, supporting a large number of operating systems and hardware architectures: on many of these systems it is already preinstalled⁴.
- It has a simple, fast syntax that can be grasped easily by non-programmers, but it is powerful enough to have been deployed for full-blown scientific applications like PyMol[343] and MMTK[344].
- It does not need to be compiled, saving development time and allowing the same files to run on every operating system.
- It features free numerical and scientific libraries of professional quality. These libraries in turn hook to fast, compiled Fortran and C based libraries, therefore using Python allows to combine speed of execution with speed of development.

For scientific and numerical programming I choosed the NumPy and SciPy libraries (<http://scipy.org>), the two standard numerical and scientific Python libraries. For the graphical user interface (GUI) part I choosed wxPython/wxWidgets (<http://www.wxpython.org>), a platform-independent library that is native-looking on all the main operating systems. This way (unlike in the case of for example Java) the application would look integrated in the operating system without having to rewrite the graphical part. For plotting, I relied on Matplotlib (<http://matplotlib.sourceforge.net>), an advanced 2D-plotting Python library. wxPython and matplotlib are glued together by mean of the little utility wxMPL (<http://agni.phys.iit.edu/~kmcivor/wxmpl/>).

All these additional components are as free and multiplatform as Python itself, and can be easily found on the Internet.

19.2.2 The *Hooke* architecture

The *Hooke* backbone consists of two concurrent threads, the command-line interface (CLI) and the graphical user interface (GUI). The CLI and the GUI are

⁴A notable exception is Windows®

independent programs and communicate by passing messages: this means that, in principle, one could substitute one or the other if the message passing protocol is kept the same. However, one cannot run without the other. The CLI thread runs in a command line terminal of the operating system. The backbone manages the basic command interpretation, general plot drawing and interaction facilities, a few very basic commands (like measuring the distance between two points) and other very basic stuff like loading playlists and configuration file (see above).

All other functionality is done by mean of *drivers* and *plugins*. Drivers and plugins use a publicly accessible application programming interface (API) and are completely separate files. *Hooke* reads from the configuration file what plugins and drivers have to be loaded at startup and loads them automatically. Plugins are coded as Python classes defining methods encoding CLI commands or GUI extensions. Each plugin has access to the whole runtime *Hooke* data structures, including these of other plugins.

Drivers are a bit different, since they do not have to code additional functionality, but to provide a common interface to different data formats. Each driver must contain two essential methods. The first, `is_me()`, is a method that allows the data format to be self-identified. When *Hooke* meets a new data file in the playlist, it feeds it, sequentially, to the `is_me()` function of each driver, which sees if it can be correctly assigned to that driver. The first driver recognizing the file format takes control, reads the file and transforms it in data vectors containing the plot informations: these are accessed by *Hooke* using the method `default_plots()`. Since of course each data format can encode additional information, every driver can implement additional methods that the appropriate plugins can call and use, if necessary. This way a common interface to all data formats is provided, while the *Hooke* user has no need of previous knowledge on the data type he has to analyze (or to convert it to a custom data type); however the whole content of the data file can be accessed if needed.

As mentioned above, *Hooke* manages data sets as “playlists” . A playlist can be saved as an XML file (which is by itself a plain textfile), and reloaded later. The user navigates the playlist, sequentially or jumping around, and can annotate each data file. Annotations can then be saved in a separate text file, and annotated data files can be copied automatically in a folder choosed by the user. The playlist file format has the capability to save unlimited information about each data file

under examination

19.3 Example plugins and features

Hooke plugins have been developed to provide the general essential capabilities required for CS-SMFS analysis. These include automatic correction from piezo movement-deflection to tip distance-force, automatic recognition of the contact point and worm-like chain fit of peaks

A simple median window filter has been implemented to smooth out noise if needed. To correct optical interference, a simple but effective solution was that of implementing a command that simply subtracts the (median-filtered) approach curve from the retracting curve. Since the two curves shape, in most cases, are practically identical expect for the signals found in the retracting curve, this method effectively flattens the retracting curve, at the expense of some added noise. A more refined algorithm, using polynomial fits to smoothly subtract optical interference, is currently being developed.

A couple of more advanced capabilities have been developed or are in development to help data analysis.

19.3.1 Peak recognition and filtering

As previously discussed (6.8) only a small fraction of force curves contains interesting signals. Recognizing these curves by hand in the large data sets produced by SMFS techniques is a tedious and time-consuming process. To automatize this process, two filters have been developed.

The first, **flatfilt**, simply recognizes curves that possess some kind of significant feature above the thermal noise. The filter cycles all the curves in the playlist. For any given curve:

- A median window filter (window size = 7) is applied on the retracting curve
- A difference vector, containing the Y difference between contiguous points, is calculated.
- The number of points in the difference vector exceeding a user-defined threshold is counted

- If there are enough points exceeding the user-defined threshold, then the curve is kept in the playlist, otherwise it is discarded.

This filter is of course quite raw, and requires relatively conservative settings to safely avoid false negatives (that is, to avoid discarding interesting curves), but on experiments on the 3S3 construct discussed in this thesis (see 6.4) it has been found to reduce the data set to analyze by hand of $\sim 80\text{-}90\%$.

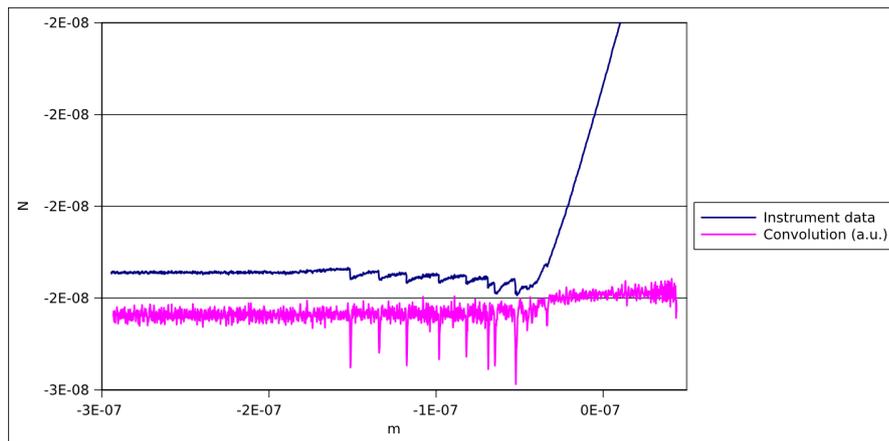


Figure 19.2: Example of the convolution used by the **convfilt** filter on a force curve. The blue graph is a retracting curve. The magenta graph represents the convolution of the blue graph. The spikes in the magenta graph correspond precisely to the peaks of the blue graph.

The second filter, **convfilt**, relies on automatic recognition of force peaks. Recognition of peaks is based on a simple convolution algorithm that is an extreme simplification of the approach of Kasas *et.al.* [198]. Again, the filter cycles all curves in the playlist. For any given retracting curve:

- The contact point is found and only data before the contact point is used
- The data before the contact point are multiplied by a L-shaped vector $(-6,-1,-1,-1,-1,-1,-1)$ that encodes the approximate L-shape that forced unfolding peaks show after the maximum. Each unfolding peak should be a clearly visible spike on the convolution (see Figure 19.3.1)

- The noise level of the convolution is calculated, eliminating the highest values (which are presumably the spikes) until noise seems converging. Average value and absolute deviation of noise is calculated.
- Data points exceeding the user-defined threshold of absolute deviation are counted. Usually these data will show up as “clusters” around a single force peak: each cluster is then reduced to a single data point.
- If there are enough points exceeding the user-defined threshold, then the curve is kept in the playlist, otherwise it is discarded.

The convolution algorithm, with appropriate thresholds, recognizes true unfolding peaks almost 99% of the time (of course, it also recognizes other similarly shaped peaks). The filter based on it can reduce the dataset to analyze by hand of 90-95%.

19.3.2 Mass analysis and selection of force curve data

As proposed in 6.8.5, to analyze the behaviour of the force experiment and the force curve patterns objectively, it could be useful to collect automatically a large number of variables and to analyze them by mean of multivariate statistics (like analysis of principal components). Moreover, better and more objective data filters could be based on selecting data clusters which are located within the range of desired variables (for example, below the total expected molecule contour length, or within a desired range of average peak-peak distance).

To this aim, a plugin is currently in development that scans the playlist and collects a number of potentially interesting variables. These variables can then be exported to be reviewed with other softwares or can be plotted in scatterplots by *Hooke* itself. In the future the user may choose a region in the scatterplot and select the curves by choosing appropriate clusters in the variable space.

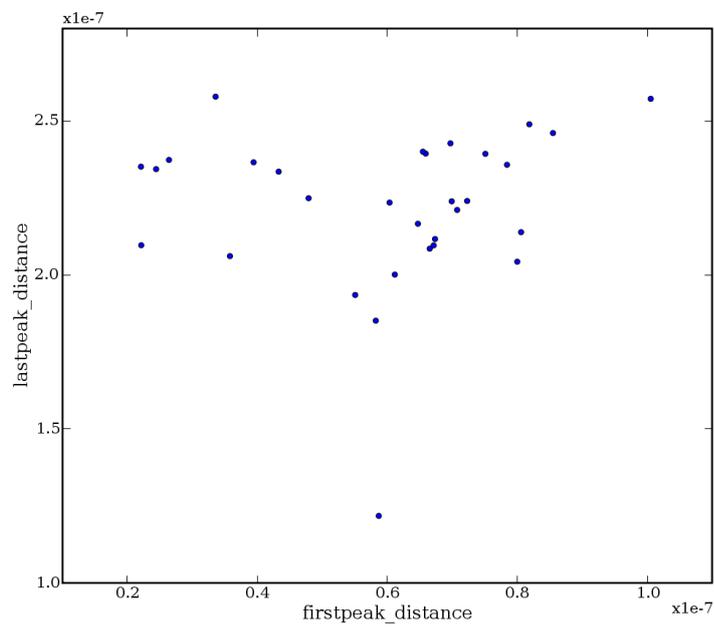


Figure 19.3: Example of scatterplot (distance of the first peak from the contact point *vs* distance of the last) obtained by running the mass analysis plugin on a small set of curves.

CHAPTER 19: DESIGN AND IMPLEMENTATION

Part VI

Final remarks

Chapter 20

Sources

Most of the material presented in the thesis finds its sources in original material and the published literature from the author.

The literature review portions of this thesis are in debt with the following works not from the author (in random order): [61], [158], [60], [345], [17], [85], [80], [289] and the Fabio Grandi Ph.D. thesis.

CHAPTER 20: SOURCES

Chapter 21

Thank you

Gratitude is merely the secret hope of further favors.

Pierre La Rouchefoucauld

Before beginning. This is of course the most read chapter of every thesis on the planet: it could be argued that thesis are an object whose only purpose is the interpolation of acknowledgements (just like Jorge Luis Borges conjectured that the Divine Comedy's only purpose is to provide a framework where to interpolate the meeting between Dante and Beatrix). Unfortunately, it's also the last one I'm writing, and also in a relative hurry. So, please forgive me if I forgot you, or if what I wrote of you isn't something you expect. You can always look for revenge in real life.

21.1 Family

To my family, I dedicate this work. They would deserve much, much, much more, but that's what I can do right now. What you did for me is a miracle that always amazes me. I love you all. Special mention of course for my grandmother Wilma, a terrific, steel woman, the undisputed leader of my family. Don't worry, grandma: *I will cut that hair and that beard* (maybe).

21.2 Laboratory

This thesis is, among other things, the compendium of years of work spent in a little, apparently insignificant, laboratory hidden in a fatigued building, that revealed instead to be a treasure of outstanding, brilliant scientists and amazing persons. I couldn't have hoped to spend my undergraduate thesis first and my Ph.D. later in a place with better people.

First of all, of course, thanks to my supervisor, Prof. Bruno Samorì, who chose to bet on me when I still was a young, confused undergraduate student, offering me an early chance to dip my toes in the misty sea of research. It was an enormous, positive shock for my self-esteem, and for my career (or at least, I hope so). His good nature, irresistible enthusiasm (perhaps even *too much* enthusiasm!), and persistent, bold willingness to break the boundaries of established research were the constant leitmotiv of these years in the laboratory. Also, a special mention is due to his fine taste for wines and foods, of course.

Then, I have to thank all the guys who embraced me as their little kid when I started working in the laboratory. Fabio Grandi, who as a graduate student followed my undergraduate thesis and my first times of graduate studies. Clever, exquisitely nice, beer expert and extremely patient tutor. He taught me the basis of experimental work, data analysis and day-to-day random fun in the lab. Thanks to you a lot. Really a lot. Anna “Verzia” Bergia, the laboratory's mother-in-law, lifesaver, and infinitely sweet friend. She made us feel all warm and fuzzy. Andrea “Turno” Giro, who behaved as a (damn funny) cartoon character and that now creates cartoon characters. Giampaolo Zuccheri, essential guide in the Dante's Inferno of experimental work, and the deeper Cocitus of day-to-day laboratory bureaucracy and correct administration. Giovanni Guarguaglini, native of the funniest corner of Italy (Livorno) and holding high its flag. We won't ever forget the alcoholic (to say the least) lunches we had. Carlotta Guiducci: perennial candidate to the Stakanov prize, yet everything you would *not* expect from an engineer, and that's what makes her special.

Other guys then joined, not a bit less special than the first. A billion special thanks go to the Supreme Archyandrite, the Blind Explorer of the Proteiform Amphisbaena, a.k.a. dr. Marco Brucalè, someone I wish I had known before and I wish to keep in touch for a long time. There is an uncanny, weird intellectual,

CHAPTER 21: THANK YOU

witty feeling between us, mostly concentrated on surreal humour and obscure subjects, that made the last years a constant source of (mostly incomprehensible for others) damn fun. Oh, also he is an exquisite musician, tough FreeCiv opponent and a tremendously brilliant scientist. Shoud I also mention the unfathomable mystery of the...No, that I shouldn't. An exquisite companion of me and Marco was Mauro Lombardi, whose mix of experimental boldness, encyclopaedic knowledge of chemistry and epileptic mental processes was always ready to explode. In laughs, but also in fire. Flaviana Mosca: we shared the lab for a few weeks only, but there are a couple of things we will never forget of you. Trust us. Francesco Valle, one of the finest, most clever guys I've ever met in my life. I will forever, ever live in the envy of his charme and social skills. And of his equally good taste for food and physics. I felt honoured to share most of my Ph.D. with you. We cannot obviously forget Sandro "Plasmone" Carrara, thanks for all the nice talk and for always reminding that we have to love the Good Things of Life. One in particular, of course *grin*. But while Sandro aimed to holiness, I had the rare honour of sharing my office with a first-class daemon. Behold, ye mortal, the infernal greatness of Francesco Musiani, Master of Sandworms, who gnaws aeternally in front of his desk at the monotonous whine of accursed heavy metal MP3s! I also shared for the office with a female rugby player (one of these creatures you usually think live only on the Internet, until you meet them. Like me.), Alessandra Vinelli, a damn brilliant girl that holds the secret of an incredible source of energy. What I do in a week, she has done in a day. Twice. Kudos. And thanks to Anna Ferretti, seductive biotechnologist with fantastic sense of humour, good taste for comics and cinema, and an infinite joy of life. You always make me smile when you enter the lab. Here's a kiss. Fabrizio Benedetti, undergraduate student everyone would be proud of, capable of killing armies with his naked hands but nonetheless incredibly peaceful. His enthusiasm for science and adamantine sense of rigour and ethics always teach me something. Thanks for keeping me alive despite my subpar abilities in undergraduate mentoring. Thanks a lot to Selena di Maio, the Pencil Angel, tremendously good friend, who was my years long bench neighbour in the high school, and now, years and years after, again shares a work room with me. I would never have finished high school, let alone university, without her supporting me by lending me her pens and pencils. Thanks to Aldo Rampioni, a nice, joyful fellow who has just now joined the laboratory and whose undestructible

optimism and peaceful attitude are the necessary contrapunct to my pessimistic, hysterical attitudes. And also, thanks to the “guys from the other laboratory” , namely Emanuele, who ectoplasmatically comes and goes for his master thesis, Andrea and Daniele, in all a team of nice guys I’d like to share more time with. Hope to meet you at the chessboard, Daniele!

Oh, we cannot forget the three guys from abroad that, inexplicably, decided to spend some time here. Dr.Salama, exquisite Egyptian guy whose brief trip to Italy could have been luckier. All the best. Bo Chen, a mysterious, eerie presence from the Far East that nicely haunted the lab. And dr.Vijayender, who introduced us gently and warmly to Indian culture. Thanks for the time we spent together.

21.3 Significant others

Special thanks to Ambra Galassi, of course, who incredibly tolerated two months of my anxiety and grief for writing this stuff, making me always feel loved and quiet. *Balú, Maurina*. And thanks to Duke the cat, too, of course.

Then, in more or less random order. Thanks a lot to my acquired sister Susanna Raule, one of the smartest and nicest girls I’ve ever met. Oh, but you already know that, don’t you? Thanks also to Giovanni Ferretti, sweet and clever guy, someone with which I discovered the true meaning of the abused word “friendship” . The same of course is for Gianluca Paredi, who I’d like to see more. But he’s too much of a womanizer to worry about bearded men like me... I also should have mentioned my roommate and comrade Cristiano Belloni aka DJ Jane’s Conference, but he, as a good La Spezia native, refuses any social honour and explicitly requires to not be mentioned in this thesis. Thanks a lot to Martinka, for all the good time we spent together, and for still being a close friend. Thanks also to Carla, for the same reasons, and for Topper. Thanks a lot to Mauro Bonacini, perfect friend and perfect geek, the guy who forever ruined my life introducing me to the alchemic mysteries of Linux. Thanks to Cristina, Aulin and Napo, terrific crew from the Padanian badlands: INDIE BAD, GABBER GOOD. Thanks to Illy, not a coffee but a living monument to black humour. And thanks to Martina and Il Poeta, who spent a lot of funny weekends in our home. Come back! And the whole Electronic Drums Contest, the best music festival. Ever. And the only one where I exhibited. And of course, I must tip my hat to the whole Asphalto crew, for the

constant fun on the net spent in endless, pointless, remorseless flamewars. Lot of thanks to anyone also following my rants on the Net. And of course, to the whole ton of people I forgot but that surely deserved a mention here.

21.4 Software

This thesis could not have been written (or at least, it would have been a lot different) without the incredible phenomenon that is the Free Software/Open Source community (<http://www.gnu.org>). An entire army of people devoted to the concept of breaking information free. Nothing better could happen. This thesis had been written on two x86 Gentoo GNU/Linux systems (<http://www.gentoo.org>), running on my desktop and on my Macbook Pro laptop, respectively. The Gentoo mailing list, forums and wiki are enormous sources of documentation (and nice talks) that cannot be overestimated. The thesis has been written using L^AT_EX (<http://www.latex-project.org/>), the well known professional free typesetting language. Kudos to Leslie Lamport and coworkers for this incredible tool, although I must say sometimes its old age can be felt in the dirty quirks of the syntax. Time for a L^AT_EX3 maybe? The L^AT_EXcode was written using Kile (<http://kile.sourceforge.net>), a splendid, featureful but elegant and easy L^AT_EX development environment. A lot of time has been saved thanks to this program. In turn, Kile is part of the KDE 3 software suite (<http://www.kde.org>), probably the best desktop environment available for Unix systems. Images were processed and created using Inkscape (<http://www.inkscape.org>) and The GIMP (<http://www.gimp.org>), two state-of-the-art tools for vectorial and raster image development, respectively. Really good stuff. PDF files containing the papers read and used for this thesis were read and printed thanks to KPDF (<http://kpdf.kde.org>), the KDE PDF reader, showing that a good PDF reader can be made without suffering of Acrobat Reader disgusting elephantiasis. The papers were in turn downloaded using Mozilla Firefox (<http://www.mozilla.org>), which needs no further words.

There's also online services I have to thank. The Pubmed-to-BibTex translator TeXMed (<http://www.sbg.bio.ic.ac.uk/~mueller/TeXMed/>) and Google Scholar (<http://scholar.google.com>) allowed me to build the bibliography in a reasonable, decent time.

A mention is also owed to /b/. Who knows, knows; otherwise, *lurk moar*.

CHAPTER 21: THANK YOU

Bibliography

- [1] JC KENDREW, G. BODO, HM DINTZIS, RG PARRISH, H. WYCKOFF, and DC PHILLIPS. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610):662–6, 1958.
- [2] MF PERUTZ. The hemoglobin molecule. *Sci Am*, 211:64–76, 1964.
- [3] C.I. Brändén and J. Tooze. *Introduction to Protein Structure*. Garland, 1991.
- [4] CN Pace. The stability of globular proteins. *CRC Crit Rev Biochem*, 3(1):1–43, 1975.
- [5] PL Privalov. Stability of proteins: small globular proteins. *Adv Protein Chem*, 33:167–241, 1979.
- [6] R.J. Ellis. Macromolecular crowding: obvious but underappreciated. *Trends Biochem. Sci.*, 26:597–604, Oct 2001.
- [7] S.M. Simon, F.J. Sousa, R. Mohana-Borges, and G.C. Walker. Regulation of *Escherichia coli* SOS mutagenesis by dimeric intrinsically disordered umuD gene products. *Proc. Natl. Acad. Sci. U.S.A.*, 105:1152–1157, Jan 2008.
- [8] L. Makowski, D.J. Rodi, S. Mandava, D.D. Minh, D.B. Gore, and R.F. Fischetti. Molecular crowding inhibits intramolecular breathing motions in proteins. *J. Mol. Biol.*, 375:529–546, Jan 2008.
- [9] L. Stagg, S.Q. Zhang, M.S. Cheung, and P. Wittung-Stafshede. Molecular crowding enhances native structure and stability of alpha/beta protein flavodoxin. *Proc. Natl. Acad. Sci. U.S.A.*, 104:18976–18981, Nov 2007.

BIBLIOGRAPHY

- [10] M. Bokvist and G. Grbner. Misfolding of amyloidogenic proteins at membrane surfaces: the impact of macromolecular crowding. *J. Am. Chem. Soc.*, 129:14848–14849, Dec 2007.
- [11] M. Perham, L. Stagg, and P. Wittung-Stafshede. Macromolecular crowding increases structural content of folded proteins. *FEBS Lett.*, 581:5065–5069, Oct 2007.
- [12] M.S. Cheung and D. Thirumalai. Effects of crowding and confinement on the structures of the transition state ensemble in proteins. *J Phys Chem B*, 111:8250–8257, Jul 2007.
- [13] A. Roque, I. Ponte, and P. Suau. Macromolecular crowding induces a molten globule state in the C-terminal domain of histone H1. *Biophys. J.*, 93:2170–2177, Sep 2007.
- [14] K. Richter, M. Nessling, and P. Lichter. Experimental evidence for the influence of molecular crowding on nuclear architecture. *J. Cell. Sci.*, 120:1673–1680, May 2007.
- [15] Massimo Sandal, Fabio Grandi, and Bruno Samori. Single molecule force spectroscopy discovers mechanochemical switches in biology: The case of the disulfide bond. *Polymer*, 47(7):2571–2579, 2006. TY - JOUR.
- [16] C.P. Johnson, H.Y. Tang, C. Carag, D.W. Speicher, and D.E. Discher. Forced unfolding of proteins within cells. *Science*, 317:663–666, Aug 2007.
- [17] V. Vogel. Mechanotransduction involving multimodular proteins: Converting force into biochemical signals. *Annu Rev Biophys Biomol Struct*, 35:459–488, 2006.
- [18] B. Chaqour, P.S. Howard, C.F. Richards, and E.J. Macarak. Mechanical stretch induces platelet-activating factor receptor gene expression through the NF-kappaB transcription factor. *J. Mol. Cell. Cardiol.*, 31:1345–1355, Jul 1999.
- [19] M. Papadaki and S.G. Eskin. Effects of fluid shear stress on gene regulation of vascular cells. *Biotechnol. Prog.*, 13:209–221, 1997.

BIBLIOGRAPHY

- [20] J.A. Carson and L. Wei. Integrin signaling's potential for mediating gene expression in hypertrophying skeletal muscle. *J. Appl. Physiol.*, 88:337–343, Jan 2000.
- [21] M. Chiquet. Regulation of extracellular matrix gene expression by mechanical stress. *Matrix Biol.*, 18:417–426, Oct 1999.
- [22] M. Chiquet, V. Tun-Civelek, and A. Sarasa-Renedo. Gene regulation by mechanotransduction in fibroblasts. *Appl Physiol Nutr Metab*, 32:967–973, Oct 2007.
- [23] E.J. Blain. Mechanical regulation of matrix metalloproteinases. *Front. Biosci.*, 12:507–527, 2007.
- [24] A.J. Maniotis, C.S. Chen, D.E. Ingber, and D.E. Ingber. Demonstration of mechanical connections between integrins, cytoskeletal filaments, and nucleoplasm that stabilize nuclear structure. *Proc. Natl. Acad. Sci. U.S.A.*, 94:849–854, Feb 1997.
- [25] H.R. Muñoz and C.M. Sacco. Cardiac mechanical energy and effects on the arterial tree. *J. Cardiothorac. Vasc. Anesth.*, 11:289–298, May 1997.
- [26] N. Resnick, S. Einav, L. Chen-Konak, M. Zilberman, H. Yahav, and A. Shay-Salit. Hemodynamic forces as a stimulus for arteriogenesis. *Endothelium*, 10:197–206, 2003.
- [27] B. Geiger, A. Bershadsky, R. Pankov, and K.M. Yamada. Transmembrane crosstalk between the extracellular matrix–cytoskeleton crosstalk. *Nat. Rev. Mol. Cell Biol.*, 2:793–805, Nov 2001.
- [28] T. Ohashi, D. P. Kiehart, and H. P. Erickson. Dynamics and elasticity of the fibronectin matrix in living cell culture visualized by fibronectin-green fluorescent protein. *Proc Natl Acad Sci U S A*, 96(5):2153–8, 1999. 0027-8424 Journal Article.
- [29] G. Baneyx, L. Baugh, and V. Vogel. Fibronectin extension and unfolding within cell matrix fibrils controlled by cytoskeletal tension. *Proc. Natl. Acad. Sci. U.S.A.*, 99:5139–5143, Apr 2002.

BIBLIOGRAPHY

- [30] M.E. Chicurel, C.S. Chen, D.E. Ingber, and D.E. Ingber. Cellular control lies in the balance of forces. *Curr. Opin. Cell Biol.*, 10:232–239, Apr 1998.
- [31] N.Q. Balaban, U.S. Schwarz, D. Riveline, P. Goichberg, G. Tzur, I. Sabanay, D. Mahalu, S. Safran, A. Bershadsky, L. Addadi, and B. Geiger. Force and focal adhesion assembly: a close relationship studied using elastic micropatterned substrates. *Nat. Cell Biol.*, 3:466–472, May 2001.
- [32] U.S. Schwarz, N.Q. Balaban, D. Riveline, A. Bershadsky, B. Geiger, and S.A. Safran. Calculation of forces at focal adhesions from elastic substrate data: the effect of localized force and the need for regularization. *Biophys. J.*, 83:1380–1394, Sep 2002.
- [33] O. du Roure, A. Saez, A. Buguin, R.H. Austin, P. Chavrier, P. Silberzan, P. Siberzan, and B. Ladoux. Force mapping in epithelial cell migration. *Proc. Natl. Acad. Sci. U.S.A.*, 102:2390–2395, Feb 2005.
- [34] D. E. Ingber. Mechanical signaling and the cellular response to extracellular matrix in angiogenesis and cardiovascular physiology. *Circ Res*, 91(10):877–87, 2002. 1524-4571 Journal Article Review Review, Tutorial.
- [35] V. Vogel, W. E. Thomas, D. W. Craig, A. Krammer, and G. Baneyx. Structural insights into the mechanical regulation of molecular recognition sites. *Trends Biotechnol*, 19(10):416–23, 2001. 0167-7799 Journal Article Review.
- [36] J.Y. Shyy and S. Chien. Role of integrins in endothelial mechanosensing of shear stress. *Circ. Res.*, 91:769–775, Nov 2002.
- [37] B. Martinac. Mechanosensitive ion channels: molecules of mechanotransduction. *J. Cell. Sci.*, 117:2449–2460, May 2004.
- [38] F Valle, M Sandal, and B Samori. The interplay between chemistry and mechanics in the transduction of a mechanical signal into a biochemical function. *Phys. Life Rev.*, doi:10.1016/j.plprev.2007.06.001, 2007.
- [39] A. Krammer, D. Craig, W.E. Thomas, K. Schulten, and V. Vogel. A structural model for force regulated integrin binding to fibronectin’s RGD-synergy site. *Matrix Biology*, 21(2):139–147, 2002.

BIBLIOGRAPHY

- [40] M. Gao, D. Craig, O. Lequin, I.D. Campbell, V. Vogel, and K. Schulten. Structure and functional significance of mechanically unfolded fibronectin type III1 intermediates. *Proc. Natl. Acad. Sci. U.S.A.*, 100:14784–14789, Dec 2003.
- [41] E. Puklin-Faucher, M. Gao, K. Schulten, and V. Vogel. How the headpiece hinge angle is opened: New insights into the dynamics of integrin activation. *J. Cell Biol.*, 175:349–360, Oct 2006.
- [42] A. Krammer, D. Craig, W.E. Thomas, K. Schulten, and V. Vogel. A structural model for force regulated integrin binding to fibronectin’s RGD-synergy site. *Matrix Biol.*, 21:139–147, Mar 2002.
- [43] K.J. Langenbach and J. Sottile. Identification of protein-disulfide isomerase activity in fibronectin. *J. Biol. Chem.*, 274:7032–7038, Mar 1999.
- [44] J. Schnepel, J. Unger, and H. Tschesche. Recombinant cryptic human fibronectinase cleaves actin and myosin: substrate specificity and possible role in muscular dystrophy. *Biol. Chem.*, 382:1707–1714, Dec 2001.
- [45] J. Schnepel and H. Tschesche. The proteolytic activity of the recombinant cryptic human fibronectin type IV collagenase from E. coli expression. *J. Protein Chem.*, 19:685–692, Nov 2000.
- [46] B.L. Smith, T.E. Schaeffer, M. Viani, J.B. Thompson, N.A. Frederick, J. Kindt, A. Belcher, G.D. Stucky, D.E. Morse, and P.K. Hansma. Molecular mechanistic origin of the toughness of natural adhesives, fibres and composites. *Nature*, 399(6738):761–763, 1999.
- [47] X.X. Qin, K.J. Coyne, and J.H. Waite. Tough tendons. Mussel byssus has collagen with silk-like domains. *J. Biol. Chem.*, 272:32623–32627, Dec 1997.
- [48] J.H. Waite, X.X. Qin, and K.J. Coyne. The peculiar collagens of mussel byssus. *Matrix Biol.*, 17:93–106, Jun 1998.
- [49] AP Jackson, JFV Vincent, and RM Turner. Comparison of nacre with other ceramic composites. *Journal of Materials Science*, 25(7):3173–3178, 1990.

BIBLIOGRAPHY

- [50] G.E. Fantner, T. Hassenkam, J.H. Kindt, J.C. Weaver, H. Birkedal, L. Pechenik, J.A. Cutroni, G.A. Cidade, G.D. Stucky, D.E. Morse, and P.K. Hansma. Sacrificial bonds and hidden length dissipate energy as mineralized fibrils separate during bone fracture. *Nat Mater*, 4:612–616, Aug 2005.
- [51] H. Heslot. Artificial fibrous proteins: a review. *Biochimie*, 80:19–31, Jan 1998.
- [52] J.B. Thompson, J.H. Kindt, B. Drake, H.G. Hansma, D.E. Morse, and P.K. Hansma. Bone indentation recovery time correlates with bond reforming time. *Nature*, 414(6865):773–776, 2001.
- [53] P. F. Lenne, A. J. Raae, S. M. Altmann, M. Saraste, and J. K. H. Horber. States and transitions during forced unfolding of a single spectrin repeat. *FEBS Letters*, 476(3):124–128, 2000. TY - JOUR.
- [54] J. Liphardt, B. Onoa, S.B. Smith, I. Tinoco, and C. Bustamante. Reversible Unfolding of Single RNA Molecules by Mechanical Force, 2001.
- [55] A.K. Dunker, J.D. Lawson, C.J. Brown, R.M. Williams, P. Romero, J.S. Oh, C.J. Oldfield, A.M. Campen, C.M. Ratliff, K.W. Hipps, J. Ausio, M.S. Nissen, R. Reeves, C. Kang, C.R. Kissinger, R.W. Bailey, M.D. Griswold, W. Chiu, E.C. Garner, and Z. Obradovic. Intrinsically disordered protein. *J. Mol. Graph. Model.*, 19:26–59, 2001.
- [56] C. Boesch, A. Bundi, M. Oppliger, and K. Wuthrich. ¹H nuclear-magnetic-resonance studies of the molecular conformation of monomeric glucagon in aqueous solution. *Eur. J. Biochem.*, 91:209–214, Nov 1978.
- [57] A.J. Daniels, R.J. Williams, and P.E. Wright. The character of the stored molecules in chromaffin granules of the adrenal medulla: a nuclear magnetic resonance study. *Neuroscience*, 3:573–585, 1978.
- [58] L. C. James and D. S. Tawfik. Conformational diversity and protein evolution—a 60-year-old hypothesis revisited. *Trends Biochem Sci*, 28(7):361–8, 2003. 0968-0004 (Print) Journal Article.

BIBLIOGRAPHY

- [59] P. Tompa. Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays*, 25:847–855, Sep 2003.
- [60] P. Tompa. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.*, 579:3346–3354, Jun 2005.
- [61] H.J. Dyson and P.E. Wright. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.*, 6:197–208, Mar 2005.
- [62] K. Gunasekaran, C.J. Tsai, S. Kumar, D. Zanuy, and R. Nussinov. Extended disordered proteins: targeting function with less scaffold. *Trends in Biochemical Sciences*, 28(2):81–85, 2003.
- [63] KI Nakayama, S. Hatakeyama, and K. Nakayama. Regulation of the Cell Cycle at the G1-S Transition by Proteolysis of Cyclin E and p27Kip1. *Biochemical and Biophysical Research Communications*, 282(4):853–860, 2001.
- [64] JJ Ward, JS Sodhi, LJ McGuffin, BF Buxton, and DT Jones. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *Journal of Molecular Biology*, 337(3):635–645, 2004.
- [65] L.M. Iakoucheva, C.J. Brown, J.D. Lawson, Z. Obradović, and A.K. Dunker. Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins. *Journal of Molecular Biology*, 323(3):573–584, 2002.
- [66] P.E. Wright and H.J. Dyson. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*, 293(2):321–331, 1999.
- [67] S.E. Salghetti, A.A. Caudy, J.G. Chenoweth, and W.P. Tansey. Regulation of transcriptional activation domain function by ubiquitin. *Science*, 293:1651–1653, Aug 2001.
- [68] A.H. Huber and W.I. Weis. The structure of the beta-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by beta-catenin. *Cell*, 105:391–402, May 2001.

BIBLIOGRAPHY

- [69] A.H. Huber, D.B. Stewart, D.V. Laurents, W.J. Nelson, and W.I. Weis. The cadherin cytoplasmic domain is unstructured in the absence of beta-catenin. A possible mechanism for regulating cadherin turnover. *J. Biol. Chem.*, 276:12301–12309, Apr 2001.
- [70] P. Venkatraman, R. Wetzell, M. Tanaka, N. Nukina, and A.L. Goldberg. Eukaryotic proteasomes cannot digest polyglutamine sequences and release them during degradation of polyglutamine-containing proteins. *Mol. Cell*, 14:95–104, Apr 2004.
- [71] V.N. Uversky, J.R. Gillespie, and A.L. Fink. Why are?natively unfolded? proteins unstructured under physiologic conditions? *Proteins Structure Function and Genetics*, 41(3):415–427, 2000.
- [72] P. Romero, Z. Obradovic, X. Li, E.C. Garner, C.J. Brown, and A.K. Dunker. Sequence complexity of disordered protein. *Proteins*, 42(1):38–48, 2001.
- [73] Z. Dosztányi, V. Csizmók, P. Tompa, and I. Simon. The Pairwise Energy Content Estimated from Amino Acid Composition Discriminates between Folded and Intrinsically Unstructured Proteins. *Journal of Molecular Biology*, 347(4):827–839, 2005.
- [74] C. Bracken, L.M. Iakoucheva, P.R. Romero, and A.K. Dunker. Combining prediction, computation and experiment for the characterization of protein disorder. *Current Opinion in Structural Biology*, 14(5):570–576, 2004.
- [75] S. Vucetic, C.J. Brown, A.K. Dunker, and Z. Obradovic. Flavors of protein disorder. *Proteins Structure Function and Genetics*, 52(4):573–584, 2003.
- [76] H.J. Dyson and P.E. Wright. Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. *Adv. Protein Chem.*, 62:311–340, 2002.
- [77] M. Fuxreiter, I. Simon, P. Friedrich, and P. Tompa. Preformed Structural Elements Feature in Partner Recognition by Intrinsically Unstructured Proteins. *Journal of Molecular Biology*, 338(5):1015–1026, 2004.

BIBLIOGRAPHY

- [78] H.J. Dyson and P.E. Wright. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.*, 12:54–60, Feb 2002.
- [79] A.P. Demchenko. Recognition between flexible protein molecules: induced and assisted folding. *J. Mol. Recognit.*, 14:42–61, 2001.
- [80] V.N. Uversky. Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: which way to go? *Cell. Mol. Life Sci.*, 60:1852–1871, Sep 2003.
- [81] A. S. Morar, A. Olteanu, G. B. Young, and G. J. Pielak. Solvent-induced collapse of alpha-synuclein and acid-denatured cytochrome *c*. *Protein Sci*, 10(11):2195–9, 2001. 0961-8368 (Print) Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.
- [82] Y. Qu and DW Bolen. Efficacy of macromolecular crowding in forcing proteins to fold. *Biophysical Chemistry*, 101:155–165, 2002.
- [83] S.L. Flaugh and K.J. Lumb. Effects of macromolecular crowding on the intrinsically disordered proteins c-Fos and p27 (Kip1). *Biomacromolecules*, 2(2):538–540, 2001.
- [84] P. Tompa. Intrinsically unstructured proteins. *Trends in Biochemical Sciences*, 27(10):527–533, 2002.
- [85] F. Chiti and C. M. Dobson. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem*, 75:333–66, 2006. 0066-4154 (Print) Journal Article Review.
- [86] T.X. Hoang, L. Marsella, A. Trovato, F. Seno, J.R. Banavar, and A. Maritan. Common attributes of native-state structures of proteins, disordered proteins, and amyloid. *Proceedings of the National Academy of Sciences*, 103(18):6883–6888, 2006.
- [87] V. N. Uversky and A. L. Fink. Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochim Biophys Acta*, 1698(2):131–53, 2004. 0006-3002 (Print) Journal Article Review.

BIBLIOGRAPHY

- [88] L.C. Serpell, M. Sunde, M.D. Benson, G.A. Tennent, M.B. Pepys, and P.E. Fraser. The protofilament substructure of amyloid fibrils. *Journal of Molecular Biology*, 300(5):1033–1039, 2000.
- [89] M. Sunde and C. Blake. The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv Protein Chem*, 50:123–59, 1997.
- [90] H.H. Bauer, U. Aebi, M. Häner, R. Hermann, M. Müller, T. Arvinte, and H.P. Merkle. Architecture and Polymorphism of Fibrillar Supramolecular Assemblies Produced by in Vitro Aggregation of Human Calcitonin. *Journal of Structural Biology*, 115(1):1–15, 1995.
- [91] M. Saiki, S. Honda, K. Kawasaki, D. Zhou, A. Kaito, T. Konakahara, and H. Morii. Higher-order Molecular Packing in Amyloid-like Fibrils Constructed with Linear Arrangements of Hydrophobic and Hydrogen-bonding Side-chains. *Journal of Molecular Biology*, 348(4):983–998, 2005.
- [92] J.S. Pedersen, D. Dikov, J.L. Flink, H.A. Hjuler, G. Christiansen, and D.E. Otzen. The Changing Face of Glucagon Fibrillation: Structural Polymorphism and Conformational Imprinting. *Journal of Molecular Biology*, 355(3):501–523, 2006.
- [93] J.D. Harper, S.S. Wong, C.M. Lieber, and P.T. Lansbury. Observation of metastable A β amyloid protofibrils by atomic force microscopy. *Chemistry & Biology*, 4(2):119–125, 1997.
- [94] A.T. Petkova, Y. Ishii, J.J. Balbach, O.N. Antzutkin, R.D. Leapman, F. Delaglio, and R. Tycko. A structural model for Alzheimer’s beta-amyloid fibrils based on experimental constraints from solid state NMR. *Proceedings of the National Academy of Sciences*, 99(26):16742, 2002.
- [95] C.P. Jaronec, C.E. MacPhee, N.S. Astrof, C.M. Dobson, and R.G. Griffin. Molecular conformation of a peptide fragment of transthyretin in an amyloid fibril. *Proceedings of the National Academy of Sciences*, 99(26):16748, 2002.
- [96] C. Ritter, M.L. Maddelein, A.B. Siemer, T. Lührs, M. Ernst, B.H. Meier, S. Saupe, and R. Riek. Correlation of structure and infectivity of the HET-s prion. *Nature*, 435(7043):844–848, 2005.

BIBLIOGRAPHY

- [97] O.S. Makin, E. Atkins, P. Sikorski, J. Johansson, and L.C. Serpell. Molecular basis for amyloid fibril formation and stability. *Proceedings of the National Academy of Sciences*, 102(2):315, 2005.
- [98] R. Nelson, M.R. Sawaya, M. Balbirnie, AO Madsen, C. Riek, R. Grothe, and D. Eisenberg. Structure of the cross-beta spine of amyloid-like fibrils. *Nature*, 435(7043):773–778, 2005.
- [99] O.N. Antzutkin, J.J. Balbach, R.D. Leapman, N.W. Rizzo, J. Reed, and R. Tycko. Multiple quantum solid-state NMR indicates a parallel, not antiparallel, organization of beta-sheets in Alzheimer’s beta-amyloid fibrils. *Proceedings of the National Academy of Sciences*, page 230315097, 2000.
- [100] J.J. Balbach, A.T. Petkova, N.A. Oyler, O.N. Antzutkin, D.J. Gordon, S.C. Meredith, and R. Tycko. Supramolecular Structure in Full-Length Alzheimer’s β -Amyloid Fibrils: Evidence for a Parallel β -Sheet Organization from Solid-State Nuclear Magnetic Resonance. *Biophysical Journal*, 83(2):1205–1216, 2002.
- [101] M. Torok, S. Milton, R. Kaye, P. Wu, T. McIntire, C.G. Glabe, and R. Langen. Structural and Dynamic Features of Alzheimer’s $A\beta$ Peptide in Amyloid Fibrils Studied by Site-directed Spin Labeling. *Journal of Biological Chemistry*, 277(43):40810–40815, 2002.
- [102] J.L. Jiménez, JI ñaki Guijarro, E. Orlova, J. Zurdo, C.M. Dobson, M. Sunde, and H.R. Saibil. Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. *The EMBO Journal*, 18:815–821, 1999.
- [103] G. Zandomenighi, M.R.H. Krebs, M.G. McCammon, and M. Fandrich. FTIR reveals structural differences between native {beta}-sheet proteins and amyloid fibrils. *Protein Science*, 13(12):3314, 2004.
- [104] MF Perutz, JT Finch, J. Berriman, and A. Lesk. Amyloid fibers are water-filled nanotubes. *Proceedings of the National Academy of Sciences*, 99(8):5591, 2002.

BIBLIOGRAPHY

- [105] J.L. Jimenez, E.J. Nettleton, M. Bouchard, C.V. Robinson, C.M. Dobson, and H.R. Saibil. The protofilament structure of insulin amyloid fibrils. *Proceedings of the National Academy of Sciences*, 99(14):9196, 2002.
- [106] C.S. Goldsbury, G.J.S. Cooper, K.N. Goldie, S.A. Müller, E.L. Saafi, WTM Gruijters, M.P. Misur, A. Engel, U. Aebi, and J. Kistler. Polymorphic Fibrillar Assembly of Human Amylin. *Journal of Structural Biology*, 119(1):17–27, 1997.
- [107] M. Tanaka, P. Chien, K. Yonekura, and J.S. Weissman. Mechanism of Cross-Species Prion Transmission An Infectious Conformation Compatible with Two Highly Divergent Yeast Prion Proteins. *Cell*, 121(1):49–62, 2005.
- [108] E.M. Jones and W.K. Surewicz. Fibril Conformation as the Basis of Species- and Strain-Dependent Seeding Specificity of Mammalian Prion Amyloids. *Cell*, 121(1):63–72, 2005.
- [109] V. N. Uversky, J. Li, P. Souillac, I. S. Millett, S. Doniach, R. Jakes, M. Goedert, and A. L. Fink. Biophysical properties of the synucleins and their propensities to fibrillate: inhibition of alpha-synuclein assembly by beta- and gamma-synucleins. *J Biol Chem*, 277(14):11970–8, 2002. 0021-9258 (Print) Journal Article.
- [110] H. Naiki, N. Hashimoto, S. Suzuki, H. Kimura, K. Nakakuki, and F. Gejyo. Establishment of a kinetic model of dialysis-related amyloid fibril extension in vitro. *Amyloid*, 4:223–232, 1997.
- [111] T.R. Serio, A.G. Cashikar, A.S. Kowal, G.J. Sawicki, J.J. Moslehi, L. Serpell, M.F. Arnsdorf, and S.L. Lindquist. Nucleated Conformational Conversion and the Replication of Conformational Information by a Prion Determinant. *Science*, 289(5483):1317, 2000.
- [112] C. F. Wright, S. A. Teichmann, J. Clarke, and C. M. Dobson. The importance of sequence diversity in the aggregation and evolution of proteins. *Nature*, 438(7069):878–81, 2005. 1476-4687 (Electronic) Journal Article.

BIBLIOGRAPHY

- [113] H. A. Lashuel, D. Hartley, B. M. Petre, T. Walz, and Jr. Lansbury, P. T. Neurodegenerative disease: amyloid pores from pathogenic mutations. *Nature*, 418(6895):291, 2002. 0028-0836 (Print) Journal Article.
- [114] M. S. Goldberg and Jr. Lansbury, P. T. Is there a cause-and-effect relationship between alpha-synuclein fibrillization and parkinson's disease? *Nat Cell Biol*, 2(7):E115–9, 2000. 1465-7392 (Print) Journal Article.
- [115] W.S. Gosal, I.J. Morten, E.W. Hewitt, D.A. Smith, N.H. Thomson, and S.E. Radford. Competing Pathways Determine Fibril Morphology in the Self-assembly of β 2-Microglobulin into Amyloid. *Journal of Molecular Biology*, 351(4):850–864, 2005.
- [116] L.A. Morozova-Roche, V. Zamotin, M. Malisauskas, A. Ohman, R. Chertkova, M.A. Lavrikova, I.A. Kostanyan, D.A. Dolgikh, and M.P. Kirpichnikov. Fibrillation of carrier protein albebetin and its biologically active constructs. Multiple oligomeric intermediates and pathways. *Biochemistry*, 43(30):9610–9619, 2004.
- [117] AJ Modler, K. Gast, G. Lutsch, and G. Damaschun. Assembly of Amyloid Protofibrils via Critical Oligomers? A Novel Pathway of Amyloid Formation. *Journal of Molecular Biology*, 325(1):135–148, 2003.
- [118] D.M. Walsh, D.M. Hartley, Y. Kusumoto, Y. Fezoui, M.M. Condron, A. Lomakin, G.B. Benedek, D.J. Selkoe, and D.B. Teplow. Amyloid β -Protein Fibrillogenesis STRUCTURE AND BIOLOGICAL ACTIVITY OF PROTOFIBRILLAR INTERMEDIATES. *Journal of Biological Chemistry*, 274(36):25945–25952, 1999.
- [119] R. Kaye, Y. Sokolov, B. Edmonds, T.M. McIntire, S.C. Milton, J.E. Hall, and C.G. Glabe. Permeabilization of Lipid Bilayers Is a Common Conformation-dependent Activity of Soluble Amyloid Oligomers in Protein Misfolding Diseases. *Journal of Biological Chemistry*, 279(45):46363, 2004.
- [120] K. A. Conway, S. J. Lee, J. C. Rochet, T. T. Ding, J. D. Harper, R. E. Williamson, and Jr. Lansbury, P. T. Accelerated oligomerization by parkinson's disease linked alpha-synuclein mutants. *Ann N Y Acad Sci*, 920:42–5, 2000. 0077-8923 (Print) Journal Article Review.

BIBLIOGRAPHY

- [121] M. Malisauskas, V. Zamotin, J. Jass, W. Noppe, C.M. Dobson, and L.A. Morozova-Roche. Amyloid Protofilaments from the Calcium-binding Protein Equine Lysozyme: Formation of Ring and Linear Structures Depends on pH and Metal Ion Concentration. *Journal of Molecular Biology*, 330(4):879–890, 2003.
- [122] A. Aguzzi and CJ Sigurdson. Antiprion immunotherapy: to suppress or to stimulate? *Nat. Rev. Immunol*, 4(9):725–36, 2004.
- [123] F. Ding, J. J. LaRocque, and N. V. Dokholyan. Direct observation of protein folding, aggregation, and a prion-like conformational conversion. *J Biol Chem*, 280(48):40235–40, 2005. 0021-9258 (Print) Journal Article Research Support, Non-U.S. Gov't.
- [124] P.H. Nguyen, M.S. Li, G. Stock, J.E. Straub, and D. Thirumalai. Monomer adds to preformed structured oligomers of Abeta-peptides by a two-stage dock-lock mechanism. *Proceedings of the National Academy of Sciences*, 104(1):111, 2007.
- [125] N.L. Fawzi, Y. Okabe, E.H. Yap, and T. Head-Gordon. Determining the Critical Nucleus and Mechanism of Fibril Elongation of the Alzheimer's A β 1–40 Peptide. *Journal of Molecular Biology*, 365(2):535–550, 2007.
- [126] S. Kumar, S.K. Mohanty, and J.B. Udgaonkar. Mechanism of Formation of Amyloid Protofibrils of Barstar from Soluble Oligomers: Evidence for Multiple Steps and Lateral Association Coupled to Conformational Conversion. *Journal of Molecular Biology*, 367(4):1186–1204, 2007.
- [127] J. Safar, H. Wille, V. Itri, D. Groth, H. Serban, M. Torchia, F.E. Cohen, and S.B. Prusiner. Eight prion strains have PrP Sc molecules with different conformations. *Nature Medicine*, 4:1157–1165, 1998.
- [128] M. Margittai and R. Langen. Template-assisted filament growth by parallel stacking of tau. *Proc Natl Acad Sci US A*, 101(28):10278–10283, 2004.
- [129] A. Aguzzi. Understanding the diversity of prions. *Nature Cell Biology*, 6(4):290–292, 2004.

BIBLIOGRAPHY

- [130] J. Ikebe, N. Kamiya, J. Ito, H. Shindo, and J. Higo. Simulation study on the disordered state of an alzheimer's beta amyloid peptide abeta(12 36) in water consisting of random-structural, beta-structural, and helical clusters. *Protein Sci*, 16(8):1596–608, 2007. 0961-8368 (Print) Journal Article.
- [131] VJ McParland, AP Kalverda, SW Homans, and SE Radford. Structural properties of an amyloid precursor of beta (2)-microglobulin. *Nat Struct Biol*, 9(5):326–31, 2002.
- [132] C.M. Eakin, A.J. Berman, and A.D. Miranker. A native to amyloidogenic transition regulated by a backbone trigger. *Nature Structural & Molecular Biology*, 13:202–208, 2006.
- [133] Y. Nagai, T. Inui, H.A. Popiel, N. Fujikake, K. Hasegawa, Y. Urade, Y. Goto, H. Naiki, and T. Toda. A toxic monomeric conformer of the polyglutamine protein. *Nature Structural & Molecular Biology*, 14:332–340, 2007.
- [134] V. N. Uversky, H. J. Lee, J. Li, A. L. Fink, and S. J. Lee. Stabilization of partially folded conformation during alpha-synuclein oligomerization in both purified and cytosolic preparations. *J Biol Chem*, 276(47):43495–8, 2001. 0021-9258 (Print) Journal Article.
- [135] H. Li, A. F. Oberhauser, S. D. Redick, M. Carrion-Vazquez, H. P. Erickson, and J. M. Fernandez. Multiple conformations of pevk proteins detected by single-molecule techniques. *Proc Natl Acad Sci U S A*, 98(19):10682–6, 2001. 0027-8424 Journal Article.
- [136] A. Sarkar, S. Caamano, and J. M. Fernandez. The elasticity of individual titin pevk exons measured by single molecule atomic force microscopy. *J Biol Chem*, 280(8):6261–6264, 2005. 0021-9258 (Print) Journal Article.
- [137] M.C. Leake, A. Grtzner, M. Krger, and W.A. Linke. Mechanical properties of cardiac titin's N2B-region by single-molecule atomic force spectroscopy. *J. Struct. Biol.*, 155:263–272, Aug 2006.
- [138] C. Ray and B. B. Akhremitchev. Conformational heterogeneity of surface-grafted amyloidogenic fragments of alpha-synuclein dimers detected by

BIBLIOGRAPHY

- atomic force microscopy. *J Am Chem Soc*, 127(42):14739–44, 2005. 0002-7863 (Print) Journal Article.
- [139] C. McAllister, M. A. Karymov, Y. Kawano, A. Y. Lushnikov, A. Mikheikin, V. N. Uversky, and Y. L. Lyubchenko. Protein interactions and misfolding analyzed by afm force spectroscopy. *J Mol Biol*, 354(5):1028–42, 2005. 0022-2836 (Print) Journal Article.
- [140] S. Mukhopadhyay, R. Krishnan, E. A. Lemke, S. Lindquist, and A. A. Deniz. A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. *Proc Natl Acad Sci U S A*, 104(8):2649–54, 2007. 0027-8424 (Print) Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't.
- [141] M. Sandal, F. Valle, I. Tessari, S. Mammi, E. Bergantino, F. Musiani, M. Brucale, L. Bubacco, and B. Samor. Conformational equilibria in monomeric alpha-synuclein at the single molecule level. *PLoS Biology*, 6(1):e6.
- [142] M.S.Z. Kellermayer, L. Grama, A. Karsai, A. Nagy, A. Kahn, Z.L. Datki, and B. Penke. Reversible Mechanical Unzipping of Amyloid {beta}-Fibrils. *Journal of Biological Chemistry*, 280(9):8464, 2005.
- [143] S.R. Collins, A. Douglass, R.D. Vale, and J.S. Weissman. Mechanism of prion propagation: amyloid growth occurs by monomer addition. *PLoS Biol*, 2(10):e321, 2004.
- [144] D. Lindley and J. O'Connell. Boltzmann's Atom: The Great Debate That Launched a Revolution in Physics. *American Journal of Physics*, 69:1020, 2001.
- [145] M. Sotomayor and K. Schulten. Single-molecule experiments in vitro and in silico. *Science*, 316(5828):1144–8, 2007. 1095-9203 (Electronic) Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.
- [146] P.V. Cornish and T. Ha. A survey of single-molecule techniques in chemical biology. *ACS Chem. Biol.*, 2:53–61, Jan 2007.

BIBLIOGRAPHY

- [147] A.D. Mehta, M. Rief, J.A. Spudich, D.A. Smith, and R.M. Simmons. Single-molecule biomechanics with optical methods. *Science*, 283:1689–1695, Mar 1999.
- [148] A. Janshoff, M. Neitzert, Y. Oberdrfer, and H. Fuchs. Force Spectroscopy of Molecular Systems-Single Molecule Spectroscopy of Polymers and Biomolecules. *Angew Chem Int Ed Engl*, 39:3212–3237, Sep 2000.
- [149] A. B. Mathur, A. M. Collinsworth, W. M. Reichert, W. E. Kraus, and G. A. Truskey. Endothelial, cardiac muscle and skeletal muscle exhibit different viscous and elastic properties as determined by atomic force microscopy. *J Biomech*, 34(12):1545–53, 2001. 0021-9290 Journal Article.
- [150] P.K. Hansma, V.B. Elings, O. Marti, and C.E. Bracker. Scanning tunneling microscopy and atomic force microscopy: application to biology and technology. *Science*, 242:209–216, Oct 1988.
- [151] J.L. Hutter and J. Bechhoefer. Calibration of atomic-force microscope tips. *Review of Scientific Instruments*, 64:1868, 1993.
- [152] E-L Florin, M Rief, H Lehmann, M Ludwig, C Dornmair, VT Moy, and HE Gaub. Sensing specific molecular interactions with the atomic force microscope. *Biosensors and Bioelectronics*, 10:895–901, 1995.
- [153] M. Conti, Y. Bustanji, G. Falini, P. Ferruti, S. Stefoni, and B. Samorì. The Desorption Process of Macromolecules Adsorbed on Interfaces: The Force Spectroscopy Approach. *J. Chem. Phys*, 109:42, 1998.
- [154] F. Khner, M. Erdmann, L. Sonnenberg, A. Serr, J. Morfill, and H.E. Gaub. Friction of single polymers at surfaces. *Langmuir*, 22:11180–11186, Dec 2006.
- [155] I. Schwaiger, C. Sattler, D. R. Hostetter, and M. Rief. The myosin coiled-coil is a truly elastic protein structure. *Nat Mater*, 1(4):232–5, 2002. 1476-1122 Journal Article.
- [156] V. Hlady and J. Buijs. Protein adsorption on solid surfaces. *Current Opinion in Biotechnology*, 7(1):72–77, 1996.

BIBLIOGRAPHY

- [157] A. P. Wiita, S. R. Ainarapu, H. H. Huang, and J. M. Fernandez. Force-dependent chemical kinetics of disulfide bond reduction observed with single-molecule techniques. *Proc Natl Acad Sci U S A*, 103(19):7222–7, 2006. 0027-8424 (Print) Journal Article.
- [158] M. Carrion-Vazquez, AF Oberhauser, H. Diez, R. Hervas, J. Oroz, J. Fernandez, and D. Martinez-Martin. Protein nanomechanics as studied by AFM single-molecule force spectroscopy. *Advanced Techniques in Biophysics*, 10:163–245, 2006.
- [159] S. Garcia-Manyes, J. Brujia, C.L. Badilla, and J.M. Fernandez. Force-clamp spectroscopy of single-protein monomers reveals the individual unfolding and folding pathways of I27 and ubiquitin. *Biophys. J.*, 93:2436–2446, Oct 2007.
- [160] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith. Entropic elasticity of lambda-phage dna. *Science*, 265(5178):1599–600, 1994. 0036-8075 Letter.
- [161] P. E. Marszalek, H. Li, and J. M. Fernandez. Fingerprinting polysaccharides with single-molecule atomic force microscopy. *Nat Biotechnol*, 19(3):258–62, 2001. 1087-0156 (Print) Journal Article.
- [162] P. E. Marszalek, H. Lu, H. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten, and J. M. Fernandez. Mechanical unfolding intermediates in titin modules. *Nature*, 402(6757):100–3, 1999. 0028-0836 (Print) Journal Article.
- [163] F. Grandi, M. Sandal, G. Guarguaglini, E. Capriotti, R. Casadio, and B. Samori. Hierarchical mechanochemical switches in angiostatin. *Chem-biochem*, 7(11):1774–82, 2006. 1439-4227 (Print) Journal Article Research Support, Non-U.S. Gov't.
- [164] M. Rief, J. M. Fernandez, and H. E. Gaub. Elastically coupled two-level systems as a model for biopolymer extensibility. *Physical Review Letters*, 81(21):4764–4767, 1997. 1.
- [165] E. Evans and K. Ritchie. Dynamic strength of molecular adhesion bonds. *Biophys J*, 72(4):1541–55, 1997. 0006-3495 Journal Article.

BIBLIOGRAPHY

- [166] M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernandez. Mechanical and chemical unfolding of a single protein: a comparison. *Proc Natl Acad Sci U S A*, 96(7):3694–9, 1999. 0027-8424 Journal Article.
- [167] N.C. Harris, Y. Song, and C.H. Kiang. Experimental Free Energy Surface Reconstruction from Single-Molecule Force Spectroscopy using Jarzynski's Equality. *Physical Review Letters*, 99(6):68101, 2007.
- [168] M. Carrion-Vazquez, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez. Atomic force microscopy captures length phenotypes in single proteins. *Proc Natl Acad Sci U S A*, 96(20):11288–92, 1999. 0027-8424 Journal Article.
- [169] A. F. Oberhauser, P. E. Marszalek, M. Carrion-Vazquez, and J. M. Fernandez. Single protein misfolding events captured by atomic force microscopy. *Nat Struct Biol*, 6(11):1025–8, 1999. 1072-8368 Journal Article.
- [170] M. Carrion-Vazquez, A. F. Oberhauser, T. E. Fisher, P. E. Marszalek, H. Li, and J. M. Fernandez. Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Prog Biophys Mol Biol*, 74(1-2):63–91, 2000. 0079-6107 Journal Article Review.
- [171] G. Yang, C. Cecconi, W.A. Baase, I.R. Vetter, W.A. Breyer, J.A. Haack, B.W. Matthews, F.W. Dahlquist, and C. Bustamante. Solid-state synthesis and mechanical unfolding of polymers of T4 lysozyme. *Proceedings of the National Academy of Sciences*, 97(1):139, 2000.
- [172] A. F. Oberhauser, C. Badilla-Fernandez, M. Carrion-Vazquez, and J. M. Fernandez. The mechanical hierarchies of fibronectin observed with single-molecule afm. *J Mol Biol*, 319(2):433–47, 2002. 0022-2836 Journal Article.
- [173] S. J. Hamill, A. E. Meekhof, and J. Clarke. The effect of boundary selection on the stability and folding of the third fibronectin type iii domain from human tenascin. *Biochemistry*, 37(22):8071–9, 1998. 0006-2960 Journal Article.

BIBLIOGRAPHY

- [174] H. Li, A.F. Oberhauser, S.B. Fowler, J. Clarke, and J.M. Fernandez. Atomic force microscopy reveals the mechanical design of a modular protein. *Proceedings of the National Academy of Sciences of the United States of America*, 97(12):6527, 2000.
- [175] DE Segall, PC Nelson, and R Phillips. Volume-exclusion effects in tethered-particle experiments: Bead size matters. *Physical Review Letters*, 96:088306, 2006.
- [176] M.S. Lawrence, K.J. Phillips, and D.R. Liu. Supercharging Proteins Can Impart Unusual Resilience. *J. Am. Chem. Soc.*, 129(33):10110–10112, 2007.
- [177] R. B. Best, B. Li, A. Steward, V. Daggett, and J. Clarke. Can non-mechanical proteins withstand force? stretching barnase by atomic force microscopy and molecular dynamics simulation. *Biophys J*, 81(4):2344–56, 2001. 0006-3495 Journal Article.
- [178] S. B. Fowler, R. B. Best, J. L. Toca Herrera, T. J. Rutherford, A. Steward, E. Paci, M. Karplus, and J. Clarke. Mechanical unfolding of a titin ig domain: structure of unfolding intermediate revealed by combining afm, molecular dynamics simulations, nmr and protein engineering. *J Mol Biol*, 322(4):841–9, 2002. 0022-2836 Journal Article.
- [179] J.I. Sulkowska and M. Cieplak. Stretching to understand proteins - a survey of the protein data bank. *Biophys. J.*, 94:6–13, Jan 2008.
- [180] R. Rohs, C. Etchebest, and R. Lavery. Unraveling proteins: a molecular mechanics study. *Biophys J*, 76(5):2760–8, 1999. 0006-3495 (Print) Journal Article.
- [181] R. Law, S. Harper, D.W. Speicher, and D.E. Discher. Influence of lateral association on forced unfolding of antiparallel spectrin heterodimers. *J. Biol. Chem.*, 279:16410–16416, Apr 2004.
- [182] L.G. Randles, R.W. Rounsevell, and J. Clarke. Spectrin domains lose cooperativity in forced unfolding. *Biophys. J.*, 92:571–577, Jan 2007.

BIBLIOGRAPHY

- [183] C. Albrecht, K. Blank, M. Lalic-Multhaler, S. Hirler, T. Mai, I. Gilbert, S. Schiffmann, T. Bayer, H. Clausen-Schaumann, and H. E. Gaub. Dna: a programmable force sensor. *Science*, 301(5631):367–70, 2003. 1095-9203 Journal Article.
- [184] D.K. West, D.J. Brockwell, P.D. Olmsted, S.E. Radford, and E. Paci. Mechanical Resistance of Proteins Explained Using Simple Molecular Models. *Biophysical Journal*, 90(1):287–297, 2006.
- [185] P. Carl, C. H. Kwok, G. Manderson, D. W. Speicher, and D. E. Discher. Forced unfolding modulated by disulfide bonds in the ig domains of a cell adhesion molecule. *Proc Natl Acad Sci U S A*, 98(4):1565–70, 2001. 0027-8424 Journal Article.
- [186] Y. Bustanji and B. Samori. The mechanical properties of human angiostatin can be modulated by means of its disulfide bonds: A single-molecule force-spectroscopy study. *Angewandte Chemie International Edition*, 41(9):1546–1548, 2002. 2002 Article English.
- [187] D. W. Urry, T. Hugel, M. Seitz, H. E. Gaub, L. Sheiba, J. Dea, J. Xu, and T. Parker. Elastin: a representative ideal protein elastomer. *Philos Trans R Soc Lond B Biol Sci*, 357(1418):169–84, 2002. 0962-8436 Journal Article Review Review, Tutorial.
- [188] R. Schoenauer, P. Bertoncini, G. Machaidze, U. Aebi, J.C. Perriard, M. Hegner, and I. Agarkova. Myomesin is a molecular spring with adaptable elasticity. *J. Mol. Biol.*, 349:367–379, Jun 2005.
- [189] GI Bell. Models for the specific adhesion of cells to cells. *Science*, 200(4342):618–627, 1978.
- [190] W. Dettmann, M. Grandbois, S. Andre, M. Benoit, A. K. Wehle, H. Kaltner, H. J. Gabius, and H. E. Gaub. Differences in zero-force and force-driven kinetics of ligand dissociation from beta-galactoside-specific proteins (plant and animal lectins, immunoglobulin g) monitored by plasmon resonance and dynamic single molecule force microscopy. *Arch Biochem Biophys*, 383(2):157–70, 2000. 0003-9861 Journal Article.

BIBLIOGRAPHY

- [191] D. E. Makarov, P. K. Hansma, and H. Metiu. Kinetic monte carlo simulation of titin unfolding. *Journal of Chemical Physics*, 114(21):9663–9673, 2002.
- [192] O.K. Dudko, G. Hummer, and A. Szabo. Intrinsic Rates and Activation Free Energies from Single-Molecule Pulling Experiments. *Physical Review Letters*, 96(10):108101, 2006.
- [193] H. Li, M. Carrion-Vazquez, A. F. Oberhauser, P. E. Marszalek, and J. M. Fernandez. Point mutations alter the mechanical stability of immunoglobulin modules. *Nat Struct Biol*, 7(12):1117–20, 2000. 1072-8368 Journal Article.
- [194] R. B. Best, S. B. Fowler, J. L. Toca-Herrera, and J. Clarke. A simple method for probing the mechanical unfolding pathway of proteins in detail. *Proc Natl Acad Sci U S A*, 99(19):12143–8, 2002. 0027-8424 Journal Article.
- [195] H. Lu and K. Schulten. Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins*, 35(4):453–63, 1999. 0887-3585 Journal Article.
- [196] B. Isralewitz, M. Gao, and K. Schulten. Steered molecular dynamics and mechanical functions of proteins. *Curr Opin Struct Biol*, 11(2):224–30, 2001. 0959-440x Journal Article Review Review, Tutorial.
- [197] R. Rounsevell, J. R. Forman, and J. Clarke. Atomic force microscopy: mechanical unfolding of proteins. *Methods*, 34(1):100–11, 2004. 1046-2023 Journal Article Review Review, Tutorial.
- [198] S. Kasas, B.M. Riederer, S. Catsicas, B. Cappella, and G. Dietler. Fuzzy logic algorithm to extract specific interaction forces from atomic force microscopy data. *Review of Scientific Instruments*, 71(5):2082–2086, 2000.
- [199] M. Kuhn, H. Janovjak, M. Hubain, and D. J. Muller. Automated alignment and pattern recognition of single-molecule force spectroscopy data. *J Microsc*, 218(Pt 2):125–32, 2005. 0022-2720 (Print) Journal Article Research Support, Non-U.S. Gov't.
- [200] A. F. Oberhauser, P. K. Hansma, M. Carrion-Vazquez, and J. M. Fernandez. Stepwise unfolding of titin under force-clamp atomic force microscopy. *Proc Natl Acad Sci U S A*, 98(2):468–72, 2001. 0027-8424 Journal Article.

BIBLIOGRAPHY

- [201] J. M. Fernandez and H. Li. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science*, 303(5664):1674–8, 2004. 1095-9203 Journal Article.
- [202] B. Samori, G. Zuccheri, and R. Baschieri. Protein unfolding and refolding under force: methodologies for nanomechanics. *Chemphyschem*, 6(1):29–34, 2005. 1439-4235 (Print) Journal Article Review.
- [203] M.B. Viani et al. Small cantilevers for force spectroscopy of single molecules. *Journal of Applied Physics*, 86(4):2258, 1999.
- [204] M.S. Li, C.K. Hu, D.K. Klimov, and D. Thirumalai. Multiple stepwise refolding of immunoglobulin domain I27 upon force quench depends on initial conditions. *Proceedings of the National Academy of Sciences of the United States of America*, 103(1):93, 2006.
- [205] M. Cieplak and P. Szymczak. Protein folding in a force clamp. *The Journal of Chemical Physics*, 124:194901, 2006.
- [206] D. J. Muller, M. Kessler, F. Oesterhelt, C. Moller, D. Oesterhelt, and H. Gaub. Stability of bacteriorhodopsin alpha-helices and loops analyzed by single-molecule force spectroscopy. *Biophys J*, 83(6):3578–88, 2002. 0006-3495 Journal Article.
- [207] F. Gittes and CF Schmidt. Thermal noise limitations on micromechanical experiments. *European Biophysics Journal*, 27(1):75–81, 1998.
- [208] A. Sarkar, R. B. Robertson, and J. M. Fernandez. Simultaneous atomic force microscope and fluorescence measurements of protein unfolding using a calibrated evanescent wave. *Proc Natl Acad Sci U S A*, 101(35):12882–6, 2004. 0027-8424 (Print) Journal Article.
- [209] SC Minne et al. Automated parallel high-speed atomic force microscopy. *Applied Physics Letters*, 72(18):2340, 1998.
- [210] P. Vettiger, G. Cross, M. Despont, U. Drechsler, U. Durig, B. Gotsmann, W. Haberle, MA Lantz, HE Rothuizen, R. Stutz, et al. The” millipede”-nanotechnology entering data storage. *Nanotechnology, IEEE Transactions on*, 1(1):39–55, 2002.

BIBLIOGRAPHY

- [211] J. M. Thornton. Disulphide bridges in globular proteins. *J Mol Biol*, 151(2):261–87, 1981. 0022-2836 Journal Article.
- [212] Ronald Wetzel. Harnessing disulfide bonds using protein engineering. *Trends in Biochemical Sciences*, 12:478–482, 1987. TY - JOUR.
- [213] R. Kalluri. Basement membranes: structure, assembly and role in tumour angiogenesis. *Nat Rev Cancer*, 3(6):422–33, 2003. 1474-175x Journal Article Review.
- [214] M. A. Wouters, K. K. Lau, and P. J. Hogg. Cross-strand disulphides in cell entry proteins: poised to act. *Bioessays*, 26(1):73–9, 2004. 0265-9247 Journal Article.
- [215] K. A. Hotchkiss, C. N. Chesterman, and P. J. Hogg. Catalysis of disulfide isomerization in thrombospondin 1 by protein disulfide isomerase. *Biochemistry*, 35(30):9761–7, 1996. 0006-2960 Journal Article.
- [216] L. Xie, C. N. Chesterman, and P. J. Hogg. Control of von willebrand factor multimer size by thrombospondin-1. *J Exp Med*, 193(12):1341–9, 2001. 0022-1007 Journal Article.
- [217] P. Stathakis, A. J. Lay, M. Fitzgerald, C. Schlieker, L. J. Matthias, and P. J. Hogg. Angiostatin formation involves disulfide bond reduction and proteolysis in kringle 5 of plasmin. *J Biol Chem*, 274(13):8910–6, 1999. 0021-9258 Journal Article.
- [218] L. J. Matthias, P. T. Yam, X. M. Jiang, N. Vandegraaff, P. Li, P. Pombouros, N. Donoghue, and P. J. Hogg. Disulfide exchange in domain 2 of cd4 is required for entry of hiv-1. *Nat Immunol*, 3(8):727–32, 2002. 1529-2908 Journal Article.
- [219] B. Yan and J. W. Smith. Mechanism of integrin activation by disulfide bond reduction. *Biochemistry*, 40(30):8861–7, 2001. 0006-2960 Journal Article.
- [220] E. Fenouillet, R. Barbouche, J. Courageot, and R. Miquelis. The catalytic activity of protein disulfide isomerase is involved in human immunodeficiency virus envelope-mediated membrane fusion after cd4 cell binding. *J Infect Dis*, 183(5):744–52, 2001. 0022-1899 Journal Article.

BIBLIOGRAPHY

- [221] L. J. Matthias, P. T. Yam, X. M. Jiang, and P. J. Hogg. Disulfide exchange in cd4. *Biofactors*, 17(1-4):241–8, 2003. 0951-6433 Journal Article Review Review, Tutorial.
- [222] X. M. Jiang, M. Fitzgerald, C. M. Grant, and P. J. Hogg. Redox control of exofacial protein thiols/disulfides by protein disulfide isomerase. *J Biol Chem*, 274(4):2416–23, 1999. 0021-9258 Journal Article.
- [223] T Yoshimori, T Semba, H Takemoto, S Akagi, A Yamamoto, and Y Tashiro. Protein disulfide-isomerase in rat exocrine pancreatic cells is exported from the endoplasmic reticulum despite possessing the retention signal. *J. Biol. Chem.*, 265(26):15984–15990, 1990.
- [224] K. Terada, P. Manchikalapudi, R. Noiva, H. O. Jauregui, R. J. Stockert, and M. L. Schilsky. Secretion, surface localization, turnover, and steady state expression of protein disulfide isomerase in rat hepatocytes. *J Biol Chem*, 270(35):20410–6, 1995. 0021-9258 Journal Article.
- [225] A. Rubartelli, A. Bajetto, G. Allavena, E. Wollman, and R. Sitia. Secretion of thioredoxin by normal and neoplastic cells through a leaderless secretory pathway. *J Biol Chem*, 267(34):24161–4, 1992. 0021-9258 Journal Article.
- [226] A. Soderberg, B. Sahaf, and A. Rosen. Thioredoxin reductase, a redox-active selenoprotein, is secreted by normal and neoplastic cells: presence in human plasma. *Cancer Res*, 60(8):2281–9, 2000. 0008-5472 Journal Article.
- [227] A. J. Lay, X. M. Jiang, O. Kisker, E. Flynn, A. Underwood, R. Condrón, and P. J. Hogg. Phosphoglycerate kinase acts in tumour angiogenesis as a disulphide reductase. *Nature*, 408(6814):869–73, 2000. 0028-0836 Journal Article.
- [228] C. Thorpe, K. L. Hooper, S. Raju, N. M. Glynn, J. Burnside, G. K. Turi, and D. L. Coppock. Sulfhydryl oxidases: emerging catalysts of protein disulfide bond formation in eukaryotes. *Arch Biochem Biophys*, 405(1):1–12, 2002. 0003-9861 Journal Article Review Review, Tutorial.
- [229] P. J. Hogg. Disulfide bonds as switches for protein function. *Trends Biochem Sci*, 28(4):210–4, 2003. 0968-0004 Journal Article Review Review, Tutorial.

BIBLIOGRAPHY

- [230] O. Carugo, M. Cemazar, S. Zahariev, I. Hudaky, Z. Gaspari, A. Perczel, and S. Pongor. Vicinal disulfide turns. *Protein Eng*, 16(9):637–9, 2003. 0269-2139 Journal Article Review Review, Tutorial.
- [231] S. Lobov, M. Wilczynska, F. Bergstrom, L. B. Johansson, and T. Ny. Structural bases of the redox-dependent conformational switch in the serpin pai-2. *J Mol Biol*, 344(5):1359–68, 2004. 0022-2836 Journal Article.
- [232] T. E. Creighton, A. Zapun, and N. J. Darby. Mechanisms and catalysts of disulfide bond formation in proteins. *Trends Biotechnol*, 13(1):18–23, 1995. 0167-7799 Journal Article Review Review, Tutorial.
- [233] M. Grandbois, M. Beyer, M. Rief, H. Clausen-Schaumann, and H. E. Gaub. How strong is a covalent bond? *Science*, 283(5408):1727–30, 1999. 0036-8075 Journal article.
- [234] M. C. Abad, R. K. Arni, D. K. Grella, F. J. Castellino, A. Tulinsky, and J. H. Geiger. The x-ray crystallographic structure of the angiogenesis inhibitor angiostatin. *J Mol Biol*, 318(4):1009–17, 2002. 0022-2836 Journal Article.
- [235] Jorg Stetefeld, Ulrike Mayer, Rupert Timpl, and Robert Huber. Crystal structure of three consecutive laminin-type epidermal growth factor-like (le) modules of laminin [gamma]1 chain harboring the nidogen binding site. *Journal of Molecular Biology*, 257(3):644–657, 1996. TY - JOUR.
- [236] D. Tisi, J. F. Talts, R. Timpl, and E. Hohenester. Structure of the c-terminal laminin g-like domain pair of the laminin alpha2 chain harbouring binding sites for alpha-dystroglycan and heparin. *Embo J*, 19(7):1432–40, 2000. 0261-4189 Journal Article.
- [237] A. R. Farina, A. Tacconelli, L. Cappabianca, G. DeSantis, A. Gulino, and A. R. Mackay. Thioredoxin inhibits microvascular endothelial capillary tubule formation. *Exp Cell Res*, 291(2):474–83, 2003. 0014-4827 Journal Article.
- [238] E. Poschl, J. W. Fox, D. Block, U. Mayer, and R. Timpl. Two non-contiguous regions contribute to nidogen binding to a single egf-like motif of the laminin gamma 1 chain. *Embo J*, 13(16):3741–7, 1994. 0261-4189 Journal Article.

BIBLIOGRAPHY

- [239] C. Chothia and E. Y. Jones. The molecular structure of cell adhesion molecules. *Annu Rev Biochem*, 66:823–62, 1997. 0066-4154 Journal Article Review.
- [240] P. D. Yurchenco, P. S. Amenta, and B. L. Patton. Basement membrane assembly, stability and activities observed through a developmental lens. *Matrix Biol*, 22(7):521–38, 2004. 0945-053x Journal Article Review Tutorial.
- [241] M. S. O'Reilly, L. Holmgren, Y. Shing, C. Chen, R. A. Rosenthal, M. Moses, W. S. Lane, Y. Cao, E. H. Sage, and J. Folkman. Angiostatin: a novel angiogenesis inhibitor that mediates the suppression of metastases by a lewis lung carcinoma. *Cell*, 79(2):315–28, 1994. 0092-8674 Journal Article.
- [242] W. R. Ji, F. J. Castellino, Y. Chang, M. E. Deford, H. Gray, X. Villarreal, M. E. Kondri, D. N. Marti, M. Llinas, J. Schaller, R. A. Kramer, and P. A. Trail. Characterization of kringle domains of angiostatin as antagonists of endothelial cell migration, an important process in angiogenesis. *Faseb J*, 12(15):1731–8, 1998. 0892-6638 Journal Article.
- [243] L. Claesson-Welsh, M. Welsh, N. Ito, B. Anand-Apte, S. Soker, B. Zetter, M. O'Reilly, and J. Folkman. Angiostatin induces endothelial cell apoptosis and activation of focal adhesion kinase independently of the integrin-binding motif rgd. *Proc Natl Acad Sci U S A*, 95(10):5579–83, 1998. 0027-8424 Journal Article.
- [244] Y. Cao, R. W. Ji, D. Davidson, J. Schaller, D. Marti, S. Sohndel, S. G. McCance, M. S. O'Reilly, M. Llinas, and J. Folkman. Kringle domains of human angiostatin. characterization of the anti-proliferative activity on endothelial cells. *J Biol Chem*, 271(46):29461–7, 1996. 0021-9258 Journal Article.
- [245] J. H. Geiger and S. E. Cnudde. What the structure of angiostatin may tell us about its mechanism of action. *J Thromb Haemost*, 2(1):23–34, 2004. 1538-7933 Journal Article Review.
- [246] S. Gately, P. Twardowski, M.S. Stack, D.L. Cundiff, D. Grella, F.J. Castellino, J. Enghild, H.C. Kwaan, F. Lee, R.A. Kramer, et al. The

BIBLIOGRAPHY

- mechanism of cancer-mediated conversion of plasminogen to the angiogenesis inhibitor angiostatin. *Proceedings of the National Academy of Sciences*, 94(20):10868, 1997.
- [247] T. Mognes, M. Etzerodt, C. Hall, G. Engelich, J. H. Graversen, and K. L. Hartshorn. Tetranectin binds to the kringle 1-4 form of angiostatin and modifies its functional activity. *J Biomed Biotechnol*, 2004(2):73–78, 2004. 1110-7243 Journal article.
- [248] T. L. Moser, M. S. Stack, I. Asplin, J. J. Enghild, P. Hojrup, L. Everitt, S. Hubchak, H. W. Schnaper, and S. V. Pizzo. Angiostatin binds atp synthase on the surface of human endothelial cells. *Proc Natl Acad Sci U S A*, 96(6):2811–6, 1999. 0027-8424 Journal Article.
- [249] N. Veitonmaki, R. Cao, L. H. Wu, T. L. Moser, B. Li, S. V. Pizzo, B. Zhivotovsky, and Y. Cao. Endothelial cell surface atp synthase-triggered caspase-apoptotic pathway is essential for k1-5-induced antiangiogenesis. *Cancer Res*, 64(10):3679–86, 2004. 0008-5472 Journal Article.
- [250] T. Tarui, L. A. Miles, and Y. Takada. Specific interaction of angiostatin with integrin alpha(v)beta(3) in endothelial cells. *J Biol Chem*, 276(43):39562–8, 2001. 0021-9258 Journal Article.
- [251] T. Tarui, M. Majumdar, L.A. Miles, W. Ruf, and Y. Takada. Plasmin-induced Migration of Endothelial Cells A POTENTIAL TARGET FOR THE ANTI-ANGIOGENIC ACTION OF ANGIOSTATIN. *Journal of Biological Chemistry*, 277(37):33564–33570, 2002.
- [252] B. Troyanovsky, T. Levchenko, G. Mansson, O. Matvijenko, and L. Holmgren. Angiomotin: an angiostatin binding protein that regulates endothelial cell migration and tube formation. *J Cell Biol*, 152(6):1247–54, 2001. 0021-9525 Journal Article.
- [253] Y. Cao, A. Chen, S.S.A. An, R.W. Ji, D. Davidson, Y. Cao, and M. Llinas. Kringle 5 of Plasminogen is a Novel Inhibitor of Endothelial Cell Growth. *Journal of Biological Chemistry*, 272(36):22924–22928, 1997.

BIBLIOGRAPHY

- [254] W. R. Ji, L. G. Barrientos, M. Llinas, H. Gray, X. Villarreal, M. E. DeFord, F. J. Castellino, R. A. Kramer, and P. A. Trail. Selective inhibition by kringle 5 of human plasminogen on endothelial cell migration, an important process in angiogenesis. *Biochem Biophys Res Commun*, 247(2):414–9, 1998. 0006-291x Journal Article.
- [255] M. Dettin, S. Bicciato, C. Scarinci, E. Cline, M. W. Lingen, and C. Di Bello. Synthetic peptides derived from the angiostatin k4 domain inhibit endothelial cell migration. *Chembiochem*, 4(11):1238–42, 2003. 1439-4227 Journal Article.
- [256] C. Li and Q. Xu. Mechanical stress-initiated signal transductions in vascular smooth muscle cells. *Cell Signal*, 12(7):435–45, 2000. 0898-6568 Journal Article Review.
- [257] T. Korff. Tensional forces in fibrillar extracellular matrices control directional capillary sprouting, 1999.
- [258] G.E. Davis and C.W. Camarillo. Regulation of Endothelial Cell Morphogenesis by Integrins, Mechanical Forces, and Matrix Guidance Pathways. *Experimental Cell Research*, 216(1):113–123, 1995.
- [259] AL Sieminski, RP Hebbel, and KJ Gooch. The relative magnitudes of endothelial force generation and matrix stiffness modulate capillary morphogenesis in vitro. *Experimental Cell Research*, 297(2):574–584, 2004.
- [260] M. Trexler and L. Patthy. Residues cys-1 and cys-79 are not essential for refolding of reduced-denatured kringle 4 fragment of human plasminogen. *Biochim Biophys Acta*, 787(3):275–80, 1984. 0006-3002 Journal Article.
- [261] N. Bhasin, P. Carl, S. Harper, G. Feng, H. Lu, D. W. Speicher, and D. E. Discher. Chemistry on a single protein, vascular cell adhesion molecule-1, during forced unfolding. *J Biol Chem*, 279(44):45865–74, 2004. 0021-9258 (Print) Journal Article.
- [262] R. B. Best and J. Clarke. What can atomic force microscopy tell us about protein folding? *Chem Commun (Camb)*, (3):183–92, 2002. 1359-7345 Journal Article.

BIBLIOGRAPHY

- [263] H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys J*, 75(2):662–71, 1998. 0006-3495 Journal Article.
- [264] B. Stec, A. Yamano, M. Whitlow, and M. M. Teeter. Structure of human plasminogen kringle 4 at 1.68 a and 277 k. a possible structural role of disordered residues. *Acta Crystallogr D Biol Crystallogr*, 53(Pt 2):169–78, 1997. 0907-4449 Journal Article.
- [265] L Kalé, R Skeel, M Bhandarkar, R Brunner, A Gursoy, N Krawetz, J Phillips, A Shinozaki, K Varadarajan, and K. Schulten. NAMD2: Greater scalability for parallel molecular dynamics. *Journal of Computational Physics*, (151):283–312, 1999.
- [266] E Lindahl, B Hess, and D van der Spoel. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *Journal of Molecular Modelling*, (7):306–317, 2001.
- [267] J. P. Abrahams, A. G. Leslie, R. Lutter, and J. E. Walker. Structure at 2.8 a resolution of F₁-ATPase from bovine heart mitochondria. *Nature*, 370(6491):621–8, 1994. 0028-0836 Journal Article.
- [268] R. Chen, L. Li, and Z. Weng. ZDOCK: an initial-stage protein-docking algorithm. *Proteins*, 52(1):80–7, 2003. 1097-0134 Evaluation Studies Journal Article.
- [269] David J. Brockwell, Godfrey S. Beddard, John Clarkson, Rebecca C. Zinnober, Anthony W. Blake, John Trinick, Peter D. Olmsted, D. Alastair Smith, and Sheena E. Radford. The effect of core destabilization on the mechanical resistance of I27. *Biophys. J.*, 83(1):458–472, 2002.
- [270] M. Kwon, C. S. Yoon, W. Jeong, S. G. Rhee, and D. M. Waisman. Annexin A2-s100A10 heterotetramer, a novel substrate of thioredoxin. *J Biol Chem*, 280(25):23584–92, 2005. 0021-9258 (Print) Journal Article.
- [271] D. T. Lincoln, E. M. Ali Emadi, K. F. Tonissen, and F. M. Clarke. The thioredoxin-thioredoxin reductase system: over-expression in human cancer. *Anticancer Res*, 23(3B):2425–33, 2003. 0250-7005 Journal Article.

BIBLIOGRAPHY

- [272] G. Powis and W. R. Montfort. Properties and biological activities of thioredoxins. *Annu Rev Biophys Biomol Struct*, 30:421–55, 2001. 1056-8700 Journal Article Review.
- [273] J. L. Rios-Steiner, M. Schenone, I. Mochalkin, A. Tulinsky, and F. J. Castellino. Structure and binding determinants of the recombinant kringle-2 domain of human plasminogen to an internal peptide from a group a streptococcal surface protein. *J Mol Biol*, 308(4):705–19, 2001. 0022-2836 Journal Article.
- [274] S. Liu, C. Zhang, H. Zhou, and Y. Zhou. A physical reference state unifies the structure-derived potential of mean force for protein folding and binding. *Proteins*, 56(1):93–101, 2004. 1097-0134 (Electronic) Journal Article.
- [275] Y. Cao, R. Cao, and N. Veitonmaki. Kringle structures and antiangiogenesis. *Curr Med Chem Anti-Canc Agents*, 2(6):667–81, 2002. 1568-0118 Journal Article Review Review, Tutorial.
- [276] S. Sengupta, E. Gherardi, L. A. Sellers, J. M. Wood, R. Sasisekharan, and T. P. Fan. Hepatocyte growth factor/scatter factor can induce angiogenesis independently of vascular endothelial growth factor. *Arterioscler Thromb Vasc Biol*, 23(1):69–75, 2003. 1524-4636 Journal Article.
- [277] J Lawler and RO Hynes. The structure of human thrombospondin, an adhesive glycoprotein with multiple calcium-binding sites and homologies with several different proteins. *J. Cell Biol.*, 103(5):1635–1648, 1986.
- [278] B. Schulze, T. Sasaki, M. Costell, K. Mann, and R. Timpl. Structural and cell-adhesive properties of three recombinant fragments derived from perlecan domain iii. *Matrix Biol*, 15(5):349–57, 1996. 0945-053x Journal Article.
- [279] A. A. Cooper, A. D. Gitler, A. Cashikar, C. M. Haynes, K. J. Hill, B. Bhullar, K. Liu, K. Xu, K. E. Strathearn, F. Liu, S. Cao, K. A. Caldwell, G. A. Caldwell, G. Marsischky, R. D. Kolodner, J. Labaer, J. C. Rochet, N. M. Bonini, and S. Lindquist. Alpha-synuclein blocks er-golgi traffic and rab1 rescues neuron loss in parkinson’s models. *Science*, 313(5785):324–8, 2006. 1095-9203 (Electronic) Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov’t.

BIBLIOGRAPHY

- [280] M. G. Spillantini, M. L. Schmidt, V. M. Lee, J. Q. Trojanowski, R. Jakes, and M. Goedert. Alpha-synuclein in lewy bodies. *Nature*, 388(6645):839–40, 1997. 0028-0836 (Print) Letter.
- [281] G. K. Wenning and K. A. Jellinger. The role of alpha-synuclein in the pathogenesis of multiple system atrophy. *Acta Neuropathol (Berl)*, 109(2):129–40, 2005. 0001-6322 (Print) Journal Article Review.
- [282] D.F. Clayton and J.M. George. The synucleins: a family of proteins involved in synaptic function, plasticity, neurodegeneration and disease. *Trends in Neurosciences*, 21(6):249–254, 1998.
- [283] K. Beyer. α -Synuclein structure, posttranslational modification and alternative splicing as aggregation enhancers. *Acta Neuropathologica*, 112(3):237–251, 2006.
- [284] K. Ueda, H. Fukushima, E. Masliah, Y. Xia, A. Iwai, M. Yoshimoto, DA Otero, J. Kondo, Y. Ihara, and T. Saitoh. Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. *Proc Natl Acad Sci US A*, 90(23):11282–11286, 1993.
- [285] A. Iwai, E. Masliah, M. Yoshimoto, N. Ge, L. Flanagan, HA de Silva, A. Kitzel, and T. Saitoh. The precursor protein of non-A beta component of Alzheimer's disease amyloid is a presynaptic protein of the central nervous system. *Neuron*, 14(2):467–475, 1995.
- [286] P.J. McLean, H. Kawamata, S. Ribich, and B.T. Hyman. Membrane Association and Protein Conformation of α -Synuclein in Intact Neurons. *Journal of Biological Chemistry*, 275(12):8812–8816, 2000.
- [287] H.J. Lee, C. Choi, and S.J. Lee. Membrane-bound α -Synuclein Has a High Aggregation Propensity and the Ability to Seed the Aggregation of the Cytosolic Form. *Journal of Biological Chemistry*, 277(1):671–678, 2002.
- [288] S. Willingham, T.F. Outeiro, M.J. DeVit, S.L. Lindquist, and P.J. Muchowski. Yeast Genes That Enhance the Toxicity of a Mutant Huntingtin Fragment or α -Synuclein. *Science*, 302(5651):1769, 2003.

BIBLIOGRAPHY

- [289] V.N. Uversky. Neuropathology, biochemistry, and biophysics of alpha-synuclein aggregation. *Journal of Neurochemistry*, 103(1):17–37, 2007.
- [290] M. Zhu, J. Li, and A. L. Fink. The association of alpha-synuclein with membranes affects bilayer structure, stability, and fibril formation. *J Biol Chem*, 278(41):40186–97, 2003. 0021-9258 (Print) In Vitro Journal Article Research Support, U.S. Gov't, P.H.S.
- [291] J. Madine, A.J. Doig, and D.A. Middleton. A Study of the Regional Effects of-Synuclein on the Organization and Stability of Phospholipid Bilayers. *Biochemistry*, 45(18):5783–5792, 2006.
- [292] M. Zhu, Z. J. Qin, D. Hu, L. A. Munishkina, and A. L. Fink. Alpha-synuclein can function as an antioxidant preventing oxidation of unsaturated lipid in vesicles. *Biochemistry*, 45(26):8135–42, 2006. 0006-2960 (Print) Journal Article Research Support, N.I.H., Extramural.
- [293] MG. Spillantini, R. Anthony Crowther, R. Jakes, N.J. Cairns, P.L. Lantos, and M. Goedert. Filamentous α -synuclein inclusions link multiple system atrophy with Parkinson's disease and dementia with Lewy bodies. *Neuroscience Letters*, 251(3):205–208, 1998.
- [294] M. H. Polymeropoulos, C. Lavedan, E. Leroy, S. E. Ide, A. Dehejia, A. Dutra, B. Pike, H. Root, J. Rubenstein, R. Boyer, E. S. Stenroos, S. Chandrasekharappa, A. Athanassiadou, T. Papapetropoulos, W. G. Johnson, A. M. Lazzarini, R. C. Duvoisin, G. Di Iorio, L. I. Golbe, and R. L. Nussbaum. Mutation in the alpha-synuclein gene identified in families with parkinson's disease. *Science*, 276(5321):2045–7, 1997. 0036-8075 (Print) Journal Article Research Support, Non-U.S. Gov't.
- [295] R. Kruger, W. Kuhn, T. Muller, D. Voitalla, M. Graeber, S. Kosel, H. Przuntek, J. T. Epplen, L. Schols, and O. Riess. Ala30pro mutation in the gene encoding alpha-synuclein in parkinson's disease. *Nat Genet*, 18(2):106–8, 1998. 1061-4036 (Print) Letter Research Support, Non-U.S. Gov't.
- [296] J. J. Zarranz, J. Alegre, J. C. Gomez-Esteban, E. Lezcano, R. Ros, I. Ampuero, L. Vidal, J. Hoenicka, O. Rodriguez, B. Atares, V. Llorens,

BIBLIOGRAPHY

- E. Gomez Tortosa, T. del Ser, D. G. Munoz, and J. G. de Yébenes. The new mutation, e46k, of alpha-synuclein causes parkinson and lewy body dementia. *Ann Neurol*, 55(2):164–73, 2004. 0364-5134 (Print) Case Reports Journal Article Research Support, Non-U.S. Gov't.
- [297] J. Li, V. N. Uversky, and A. L. Fink. Conformational behavior of human alpha-synuclein is modulated by familial parkinson's disease point mutations a30p and a53t. *Neurotoxicology*, 23(4-5):553–67, 2002. 0161-813X (Print) Journal Article Research Support, U.S. Gov't, P.H.S.
- [298] K. A. Conway, J. D. Harper, and P. T. Lansbury. Accelerated in vitro fibril formation by a mutant alpha-synuclein linked to early-onset parkinson disease. *Nat Med*, 4(11):1318–20, 1998. 1078-8956 (Print) Journal Article.
- [299] VN Uversky. A protein-chameleon: conformational plasticity of alpha-synuclein, a disordered protein involved in neurodegenerative disorders. *J Biomol Struct Dyn*, 21(2):211–34, 2003.
- [300] V. N. Uversky, J. Li, and A. L. Fink. Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J Biol Chem*, 276(14):10737–44, 2001. 0021-9258 (Print) Journal Article.
- [301] C. W. Bertocini, Y. S. Jung, C. O. Fernandez, W. Hoyer, C. Griesinger, T. M. Jovin, and M. Zweckstetter. Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. *Proc Natl Acad Sci U S A*, 102(5):1430–5, 2005. 0027-8424 (Print) Journal Article.
- [302] M. M. Dedmon, K. Lindorff-Larsen, J. Christodoulou, M. Vendruscolo, and C. M. Dobson. Mapping long-range interactions in alpha-synuclein using spin-label nmr and ensemble molecular dynamics simulations. *J Am Chem Soc*, 127(2):476–7, 2005. 0002-7863 (Print) Journal Article.
- [303] M. Bisaglia, A. Trolino, M. Bellanda, E. Bergantino, L. Bubacco, and S. Mammi. Structure and topology of the non-amyloid-beta component fragment of human alpha-synuclein bound to micelles: implications for the aggregation process. *Protein Sci*, 15(6):1408–16, 2006. 0961-8368 (Print) Journal Article.

BIBLIOGRAPHY

- [304] S. Chandra, X. Chen, J. Rizo, R. Jahn, and T.C. Sudhof. A Broken α -Helix in Folded α -Synuclein. *Journal of Biological Chemistry*, 278(17):15313–15318, 2003.
- [305] M. Bisaglia, I. Tessari, L. Pinato, M. Bellanda, S. Giraud, M. Fasano, E. Bergantino, L. Bubacco, and S. Mammi. A topological model of the interaction between alpha-synuclein and sodium dodecyl sulfate micelles. *Biochemistry*, 44(1):329–39, 2005. 0006-2960 (Print) Journal Article.
- [306] T. S. Ulmer and A. Bax. Comparison of structure and dynamics of micelle-bound human alpha-synuclein and parkinson disease variants. *J Biol Chem*, 280(52):43179–87, 2005. 0021-9258 (Print) Journal Article Research Support, N.I.H., Intramural Research Support, Non-U.S. Gov't.
- [307] C.C. Jao, A. Der-Sarkissian, J. Chen, R. Langen, and H.B. Gray. Structure of Membrane-Bound α -Synuclein Studied by Site-Directed Spin Labeling. *Proceedings of the National Academy of Sciences of the United States of America*, 101(22):8331–8336, 2004.
- [308] L. A. Munishkina, C. Phelan, V. N. Uversky, and A. L. Fink. Conformational behavior and aggregation of alpha-synuclein in organic solvents: modeling the effects of membranes. *Biochemistry*, 42(9):2720–30, 2003. 0006-2960 (Print) Journal Article.
- [309] V.N. Uversky, J. Li, and A.L. Fink. Trimethylamine-N-oxide-induced folding of α -synuclein. *FEBS Letters*, 509(1):31–35, 2001.
- [310] R. Lowe, D.L. Pountney, P.H. Jensen, W.P. Gai, and N.H. Voelcker. Calcium (II) selectively induces {alpha}-synuclein annular oligomers via interaction with the C-terminal domain. *Protein Science*, 13(12):3245, 2004.
- [311] T. T. Ding, S. J. Lee, J. C. Rochet, and Jr. Lansbury, P. T. Annular alpha-synuclein protofibrils are produced when spherical protofibrils are incubated in solution or bound to brain-derived membranes. *Biochemistry*, 41(32):10209–17, 2002. 0006-2960 (Print) Journal Article.

BIBLIOGRAPHY

- [312] H. A. Lashuel, B. M. Petre, J. Wall, M. Simon, R. J. Nowak, T. Walz, and Jr. Lansbury, P. T. Alpha-synuclein, especially the parkinson's disease-associated mutants, forms pore-like annular and tubular protofibrils. *J Mol Biol*, 322(5):1089–102, 2002. 0022-2836 (Print) Journal Article.
- [313] K. A. Conway, S. J. Lee, J. C. Rochet, T. T. Ding, R. E. Williamson, and Jr. Lansbury, P. T. Acceleration of oligomerization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset parkinson's disease: implications for pathogenesis and therapy. *Proc Natl Acad Sci U S A*, 97(2):571–6, 2000. 0027-8424 (Print) Journal Article.
- [314] A. Der-Sarkissian, C.C. Jao, J. Chen, and R. Langen. Structural Organization of α -Synuclein Fibrils Studied by Site-directed Spin Labeling*. *Journal of Biological Chemistry*, 278(39):37530–37535, 2003.
- [315] C. Del Mar, E. A. Greenbaum, L. Mayne, S. W. Englander, and Jr. Woods, V. L. Structure and properties of alpha-synuclein and other amyloids determined at the amino acid level. *Proc Natl Acad Sci U S A*, 102(43):15477–82, 2005. 0027-8424 (Print) Journal Article.
- [316] V. N. Uversky, J. Li, and A. L. Fink. Metal-triggered structural transformations, aggregation, and fibrillation of human alpha-synuclein. a possible molecular link between parkinson's disease and heavy metal exposure. *J Biol Chem*, 276(47):44284–96, 2001. 0021-9258 (Print) Journal Article.
- [317] V. N. Uversky, M. Cooper E, K. S. Bower, J. Li, and A. L. Fink. Accelerated alpha-synuclein fibrillation in crowded milieu. *FEBS Lett*, 515(1-3):99–103, 2002. 0014-5793 (Print) Journal Article.
- [318] H. Y. Kim, H. Heise, C. O. Fernandez, M. Baldus, and M. Zweckstetter. Correlation of amyloid fibril beta-structure with the unfolded state of alpha-synuclein. *Chembiochem*, 8(14):1671–1674, 2007. 1439-4227 (Print) Journal article.
- [319] C. W. Bertocini, C. O. Fernandez, C. Griesinger, T. M. Jovin, and M. Zweckstetter. Familial mutants of alpha-synuclein with increased neurotoxicity have a destabilized conformation. *J Biol Chem*, 280(35):30649–52, 2005. 0021-9258 (Print) Journal Article Research Support, Non-U.S. Gov't.

BIBLIOGRAPHY

- [320] W. Hoyer, D. Cherny, V. Subramaniam, and T. M. Jovin. Impact of the acidic c-terminal region comprising amino acids 109-140 on alpha-synuclein aggregation in vitro. *Biochemistry*, 43(51):16233–42, 2004. 0006-2960 (Print) Journal Article Research Support, Non-U.S. Gov't.
- [321] P. Bernado, C. W. Bertoncini, C. Griesinger, M. Zweckstetter, and M. Blackledge. Defining long-range order and local disorder in native alpha-synuclein using residual dipolar couplings. *J Am Chem Soc*, 127(51):17968–9, 2005. 0002-7863 (Print) Journal Article.
- [322] H. Li, W. A. Linke, A. F. Oberhauser, M. Carrion-Vazquez, J. G. Kerkvliet, H. Lu, P. E. Marszalek, and J. M. Fernandez. Reverse engineering of the giant muscle protein titin. *Nature*, 418(6901):998–1002, 2002. 0028-0836 Journal Article.
- [323] A. Steward, J. L. Toca-Herrera, and J. Clarke. Versatile cloning system for construction of multimeric proteins for use in atomic force microscopy. *Protein Sci*, 11(9):2179–83, 2002. 0961-8368 (Print) Journal Article.
- [324] B. Miroux and J. E. Walker. Over-production of proteins in escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol*, 260(3):289–98, 1996. 0022-2836 (Print) Journal Article Review.
- [325] A. Binolfi, R. M. Rasia, C. W. Bertoncini, M. Ceolin, M. Zweckstetter, C. Griesinger, T. M. Jovin, and C. O. Fernandez. Interaction of alpha-synuclein with divalent metal ions reveals key differences: a link between structure, binding specificity and fibrillation enhancement. *J Am Chem Soc*, 128(30):9893–901, 2006. 0002-7863 (Print) Journal Article.
- [326] M. Hegner, P. Wagner, and G. Semenza. Ultralarge atomically flat template-stripped au surfaces for scanning probe microscopy. *Surface Science*, 291(1-2):39–46, 1993. Lk470 Times Cited:161 Cited References Count:28.
- [327] P. Wagner, M. Hegner, H. J. Guntherodt, and G. Semenza. Formation and in-situ modification of monolayers chemisorbed on ultraflat template-stripped gold surfaces. *Langmuir*, 11(10):3867–3875, 1995. Tc040 Times Cited:145 Cited References Count:40.

BIBLIOGRAPHY

- [328] A. P. Pawar, K. F. Dubay, J. Zurdo, F. Chiti, M. Vendruscolo, and C. M. Dobson. Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *J Mol Biol*, 350(2):379–92, 2005. 0022-2836 (Print) Journal Article.
- [329] A. Dusa, J. Kaylor, S. Edridge, N. Bodner, D. P. Hong, and A. L. Fink. Characterization of oligomers during alpha-synuclein aggregation using intrinsic tryptophan fluorescence. *Biochemistry*, 45(8):2752–60, 2006. 0006-2960 (Print) Comparative Study Journal Article Research Support, N.I.H., Extramural.
- [330] R. M. Rasia, C. W. Bertoncini, D. Marsh, W. Hoyer, D. Cherny, M. Zweckstetter, C. Griesinger, T. M. Jovin, and C. O. Fernandez. Structural characterization of copper(ii) binding to alpha-synuclein: Insights into the bioinorganic chemistry of parkinson's disease. *Proc Natl Acad Sci U S A*, 102(12):4294–9, 2005. 0027-8424 (Print) Journal Article.
- [331] A. M. Cuervo, L. Stefanis, R. Fredenburg, P. T. Lansbury, and D. Sulzer. Impaired degradation of mutant alpha-synuclein by chaperone-mediated autophagy. *Science*, 305(5688):1292–5, 2004. 1095-9203 (Electronic) Journal Article.
- [332] N. C. Maiti, M. M. Apetri, M. G. Zagorski, P. R. Carey, and V. E. Anderson. Raman spectroscopic characterization of secondary structure in natively unfolded proteins: alpha-synuclein. *J Am Chem Soc*, 126(8):2399–408, 2004. 0002-7863 (Print) Journal Article.
- [333] E. Parzen. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.
- [334] M. Mickler, R.I. Dima, H. Dietz, C. Hyeon, D. Thirumalai, and M. Rief. Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations. *Proc. Natl. Acad. Sci. U.S.A.*, 104:20268–20273, Dec 2007.
- [335] Q. Peng and H. Li. Atomic force microscopy reveals parallel mechanical unfolding pathways of T4 lysozyme: evidence for a kinetic partitioning mechanism. *Proc. Natl. Acad. Sci. U.S.A.*, 105:1885–1890, Feb 2008.

BIBLIOGRAPHY

- [336] M. D. Shtilerman, T. T. Ding, and Jr. Lansbury, P. T. Molecular crowding accelerates fibrillization of alpha-synuclein: could an increase in the cytoplasmic protein concentration induce parkinson's disease? *Biochemistry*, 41(12):3855–60, 2002. 0006-2960 (Print) Journal Article.
- [337] I. Horcas, R. Fernández, JM Gómez-Rodríguez, J. Colchero, J. Gómez-Herrero, and AM Baro. WSXM: A software for scanning probe microscopy and a tool for nanotechnology. *Review of Scientific Instruments*, 78:013705, 2007.
- [338] F. Mueller, D.J. Muller, and D. Labudde. Analysis assistant for single-molecule force spectroscopy data on membrane proteins-MPTV. *Bioinformatics*, 22(14):1796–1799, 2006.
- [339] F. Valle, G. Zuccheri, A. Bergia, L. Ayres, A.E. Rowan, R.J.M. Nolte, and B Samori. A polymeric molecular "handle" for multiple and simultaneous afm-based single molecule force measurements. *Angewandte Chemie International Edition*, (in press), 2007.
- [340] E.S. Raymond. *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. O'Reilly, 2001.
- [341] K.A. Remington. The NIST Fortran Sparse BLAS Library Reference Implementation. *National Institute of Standards and Technology, Maryland, Aug*, 1997.
- [342] E. Angerson, Z. Bai, J. Dongarra, A. Greenbaum, A. McKenney, J. Du Croz, S. Hammarling, J. Demmel, C. Bischof, and D. Sorensen. LAPACK: A portable linear algebra library for high-performancecomputers. *Supercomputing'90. Proceedings of*, pages 2–11, 1990.
- [343] W.L. DeLano. Pymol: An open-source molecular graphics tool. *CCP4 NEWSLETTER ON PROTEIN CRYSTALLOGRAPHY*, 40, 2002.
- [344] K. Hinsen. The molecular modeling toolkit: A new approach to molecular simulations. *Journal of Computational Chemistry*, 21(2):79–85, 2000.

BIBLIOGRAPHY

- [345] A. L. Fink. Natively unfolded proteins. *Curr Opin Struct Biol*, 15(1):35–41, 2005. 0959-440X (Print) Journal Article Review.

BIBLIOGRAPHY