Molecular interpretation of single-molecule force spectroscopy experiments with computational approaches

Guillaume Stirnemann

Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Molecular interpretation of single-molecule force spectroscopy experiments with computational approaches

Guillaume Stirnemann*[a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Single molecule force-spectroscopy techniques have granted access to unprecedented molecular-scale details about biochemical and biological mechanisms. However, the interpretation of the experimental data is often challenging and it benefits from the perspective brought by computational approaches. In many cases, these simulations (all-atom steered MD simulations in particular) are key to provide molecular details about the associated mechanisms, to help test different hypotheses and to predict experimental results. In this review, particular recent efforts directed towards the molecular interpretation of single-molecule force spectroscopy experiments on proteins and protein-related systems will be discussed, often in close collaboration with experimental groups. These results will be discussed in the broader contexts of the field, highlighting the recent achievements and the ongoing challenges for computational biophysicists and biochemists. In particular, I will focus on the input gained from molecular simulations approaches to rationalize the origins for the unfolded protein elasticity and the protein conformational behavior under force, to understand how force denaturation differs from chemical, thermal or shear unfolding, and to unravel the molecular details of unfolding events for a variety of systems. I will also discuss the use of models based on Langevin dynamics on a 1-D free-energy surface to understand the effect of protein segmentation on the work exerted by a force, or, at the other end of the spectrum of computational techniques, how quantum calculations can help to understand the reactivity of disulfide bridges exposed under force.

1 Introduction

Most physical chemists studying biomolecular objects and processes use experimental techniques based on *bulk* measurements, where the studied sample contains many replicas of the system. There are two main reasons for this: first, owing to the small, molecular-scale of biomolecules, even a tiny physical sample would contain a very large number of molecules; and second, while working in a very dilute environment could virtually enable the study of a small subset of molecules, this often does not yield to enough measurable signal. As a consequence, the recorded measurements are ensemble-averaged, and often, most of the detailed mechanistic aspects of the corresponding stochastic events are lost because they are not synchronized in space and time.

In contrast, single-molecule techniques allow direct access to the dynamical evolution of individual biomolecules[1]. This contribution will exclusively focus on the specific case of single-molecule force spectroscopies, which have opened a whole new and exciting research field in the last 25 years[2,3]. In such experiments, biomolecules can be manipulated with nano- to micrometer size objects that enable the application of directional mechanical force[1]. Force can be used as a probing tool that alters the thermodynamics, the structure and the dynamics of biomolecules, providing valuable kinetic and mechanistic insights into a variety of phenomena, including enzyme catalysis[4], protein-ligand interactions[5], or folding/unfolding events[6], just to name a few important examples. For other systems, mechanical force is actually of direct biological relevance. Hence, many of the cell functions are controlled by mechanotransduction, which is the cascade of biomolecular events leading to the conversion of mechanical stimuli into chemical signals[7]. Muscles generate macroscopic forces through the extension and the contraction of their biomolecular units[8]. Biological mechanisms such as blood coagulation[9] or urinary infections are regulated by shear forces from the surrounding fluid[10].

Despite their great power, these techniques also face some limitations. First, cooperative and multi-body effects that would involve several biomolecules operating together might be lost when isolating and focusing on a single object; however, such an approach precisely allows to disentangle the effects of each of the biomolecular players. Second, because the measured signal is typically of very small amplitude and thus amplified, the experimental noise is usually large compared to that of bulk techniques[1]. Another major issue is that the experimental observables

[a] CNRS Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, PSL University, Université de Paris, 13 rue Pierre et Marie Curie, 75005, Paris, France; E-mail: stirnemann@ibpc.fr

are not as rich and detailed as those accessible with other types of spectroscopies, such as nuclear magnetic resonance (NMR), ultrafast infrared, or UV-visible spectroscopies. These are for example simply an on/off activity in fluorescence-based techniques, ionic currents in patch-clamp spectroscopy, or the biomolecule end-to-end distance in force-spectroscopies. As a consequence, a detailed mechanistic picture, down to the molecular level of these biological objects, is typically obtained indirectly, by systematically varying the experimental conditions (pH, temperature, presence of other biomolecular partners or substrates, mutations, etc.).

In this Feature article, I will highlight the interest in complementing single-molecule force-spectroscopy studies with computational approaches, in particular with quantum calculations, particle-based molecular simulations, or stochastic dynamics on simplified free-energy surfaces. This discussion will be illustrated with some of our significant achievements made over the years, using joint experimental/computational studies or detailed computational investigations of previously-available experimental results. While the experimental single-molecule force-spectroscopy measurements are almost exclusively the starting point of these studies, often complemented with other experimental techniques as well, I will show how the results of computational approaches are instrumental (despite many limitations) in providing a rationalization and a detailed mechanistic picture of the experimental results at the molecular level. Comparing simulation and experimental results also enables to understand some of the effects of the experimental setup on the measured signals, which is of crucial importance when trying to interpret these measurements.

In the following, this short overview is organized into different sections corresponding to several key aspects of our recent works, focusing here almost exclusively on proteins. I will thus start by discussing how proteins unfold upon force, and how this differs from that under more conventional perturbations (temperature or chemical denaturation); I will then detail the molecular origins for protein elasticity under mechanical force, and the implications in terms of e.g. biomolecular recognition. A third section is devoted to the impact of protein segmentation into polyprotein constructs that are often used in the experiments on the measured free-energy landscapes. Finally, I will address the important topic of chemical reactivity upon force, which, in the context of proteins, has often focused on disulfide bond reduction.

2 Biographical sketch

Guillaume Stirnemann was born in France in 1987. After undergraduate studies at the École normale supérieure (ENS) and a Master's thesis with Pablo Debenedetti at Princeton University, he received his Ph.D. (2011) from ENS and Sorbonne University (Paris, France), under the supervision of Damien Laage. He was then a postdoctoral fellow at Columbia University with Bruce Berne and later at the Institut de Biologie Physico-Chimique (IBPC) in Paris with Fabio Sterpone. In 2014, he was recruited as a CNRS researcher at IBPC where the current interests of his group include the stability, the mechanical properties and the reactivity of biomolecules, and transport phenomena in aqueous solutions, with a special emphasis on questions related to the origins of life.

3 Simulation vs experimental approaches

Three techniques[1] are extensively used to apply forces on biomolecular objects in the pN-nN range: the atomic force microscope (AFM), optical tweezers (OT), and magnetic tweezers (MT). A biomolecular object (a protein, a polyprotein, a nucleic acid, or a chimeric structure with protein(s) and nucleic acid handles) is typically attached in a covalent or non-covalent manner to micrometer-size tethers (such as beads, tips, or a piezzo-electric crystal surface), which are stretched apart, thereby exerting mechanical force on the biomolecular construct. Most of the time, these are used either in a constant-velocity, or in a constant-force mode[1].

Molecular dynamics (MD) simulations, on which this article will exclusively focus, consist in the numerical propagation of Newton's equations of motion for an ensemble of particles. In all-atom strategies, each atom of the protein is described by one particle. Interactions between particles are calibrated on quantum calculations or on experimental observables; the water solvent is either described explicitly at the same level of precision (all-atom), or implicitly (for example, as a dielectric continuum). In other approaches that are computationally less expensive, one particle in the simulation can correspond to several atoms of the protein (for example, to one whole residue, or to one side-chain).

In a MD simulation set-up, both experimental strategies to apply directional force can easily be implemented in a *steered* mode[11–14] (SMD): in the constant-velocity mode, a unidirectional harmonic restraint is applied to the protein extremities and its equilibrium length is shifted at a constant linear rate. In the constant-force mode, a unique, constant and unidirectional force is applied to one end of the protein while the other one is kept fixed. Because force can directly be applied on individual atoms during the propagation of the equations of motion, there is no need of microscopic tethers or handles in the simulations.

The outcome of experimental and simulation studies regarding a variety of biomolecular phenomena will be discussed in detail, ranging from protein unfolding to disulfide bond reduction, focusing on the molecular interpretation provided by the simulations. However, two important points regarding the comparison of experimental and computational results should be raised. First, the use of tethers in the experiments, that are typically several orders of magnitude larger and heavier than the biomolecule itself, necessarily affects the dynamics of the end-to-end motion[15–17]. In MD simulations, the protein is "free" and the dynamics of motion along the end-to-end distance is a direct reporter of the protein

diffusion coefficient along the end-to-end distance. While this might seem obvious, this effect has long been ignored. We clearly demonstrated that upon a decrease in force, an unfolded protein chain collapses much faster in the simulations as compared to the experiments[16].

A second issue one has to keep in mind is that usually, the forces employed in MD simulations are significantly larger than their experimental counterparts, because the respective timescales of these approaches are very different. Experiments are typically performed on the millisecond to second timescale, and are limited by the time-resolution of the recording devices. Simulations are typically propagated on the nanosecond to microsecond timescale, limited by the extensive computational resources that are required to numerically solve the equations of motion. This implies that when a force is used to trigger e.g. a conformational change, it has to be larger in the simulations in order to cross the free-energy barrier faster[18]. When force is of physiological relevance (typically in the pN range), simulations also need to employ higher forces to observe significant changes. However, as will be detailed later, care should be taken when comparing experiments and simulation when the employed forces are very different, because the explored pathways on the free-energy landscape might be sensitive to force[19–23].

Interestingly, simulations employing enhanced sampling strategies specifically adapted to the study of biomolecules under force, such as infinite switch simulated tempering in force[24], boxed molecular dynamics[25], or accelerated steered molecular dynamics[26], could offer the possibility to access to conformational changes occurring at experimental forces but usually not in the limited timescale of unperturbed simulations. As recently argued[27], another promising approach could be to combine SMD at experimental forces, with enhanced sampling algorithms designed for the estimation of kinetic rates[28]. A proof-of-concept was recently published on a model system[29].

4 Protein conformational changes upon force

In single-molecule force-spectroscopy studies, the protein resistance and eventually its unfolding are probed along one particular reaction coordinate, the end-to-end distance. Some early work supported the idea that high-force unfolding data could be extrapolated at zero force, implying that mechanical and "bulk" (chemical or thermal) unfolding were proceeding along the same pathways[30]. However, it is now clear, in particular thanks to coarse-grained and all-atom simulations studies, that the folding/unfolding pathways are markedly different[19,22,23,31]. The unfolded state ensembles also exhibit very distinct protein conformations[31]. Not surprisingly, experimental[32,33] and simulation[34] studies have shown that the mechanical and the thermal resistance of protein mutants are not necessarily correlated. The molecular details gained from these simulation studies regarding the force unfolding pathways are now discussed in detail.

4.1 Deformation and unfolding

Unfolding upon force necessarily occurs upon stretching the end-to-end distance, whereas chemical or thermal unfolding involve different collective variables. In the experiments, unfolding intermediates are not routinely observed, especially at high forces, and the experimental signal typically records force unfolding as single events[3], with some exceptions[35]. Yet, the time-resolution of the molecular simulations allows to unravel the microscopic details of these unfolding events, the presence of short-lived intermediates that cannot be observed in the experiments, and to understand the origins of mechanical resistance. These strategies are exemplified for four different systems that we recently investigated and that illustrate the interest in the molecular modeling of proteins under force, as now detailed.

While in the experiments, protein usually unfold under force in a two-state manner, azurin, a copper protein, exhibits a well-identified unfolding intermediate[36]. SMD simulations showed that the protein can actually stochastically unfold along either the N-C or the C-N direction, suggesting that the barriers to unfolding from both termini are largely equivalent at the probed pulling velocity[36]. Using the simulations, we could ascribe each of the observed unfolding intermediates to a given unfolding scenario. Indeed, each individual unfolding pathway is unambiguously distinguished by the position of the subsequent rupture of the copper-ligand bond, as further supported by directed mutagenesis. These observations were expanded to the topologically similar plastocyanin protein, exhibiting analogous behavior.

In another joint experimental-computational effort, we unravelled the molecular details for the force-opening of a highly conserved binding site between two proteins involved in mechanotransduction, β-catenin and E-cadherin. Experimentally, it is known that force triggers the phosphorylation of a binding site residue, leading to irreversible dissociation of the two proteins. All-atom MD simulations under a force mimicking a 6 pN physiological mechanical strain predicted a local 45% stretching between the two bound α-helices and a 15% increase in accessibility of the residue that would be phosphorylated[37].

MD simulations also allowed to rationalize how the binding of TG-rich single stranded nucleic acid to the RRM1 domain of the TDP-43 protein could entirely change its mechanical stability[38]. Experiments had suggested that in the absence of nucleic acid binding, this protein does not resist to mechanical force and unfolds very easily. When a nucleic acid was bound, it triggered mechanical resistance for forces up to 40 pN (Fig. 1). The simulations demonstrated that in the apo form, the protein mechanical unfolding occurred upon disruption of a set of hydrogen-bonds (HBs) between two key β-sheets. For mechanically-resistant proteins, HBs motifs are perpendicular to the direction of the force (therefore linking β-sheets that are parallel to the force direction), which provides a high degree of cooperativity and thus high resistance against force. In the case of the TDP-43 protein, these HBs are parallel to the force and can be broken one by one when the two β-sheets unzip. Interestingly, nucleic acids bind on top of these β-sheets, in the region where unfolding is initiated, and they can therefore act as a molecular "lid" that prevents unfolding. It is only after the nucleic acid strand has been removed that the protein can rapidly unzip and unfold. As a further validation of this mechanism, additional simulations were performed where the nucleic acid was maintained fixed, resulting in no unfolding

of the protein on a timescale where all traces exhibited full unfolding in the presence of a flexible nucleic acid.

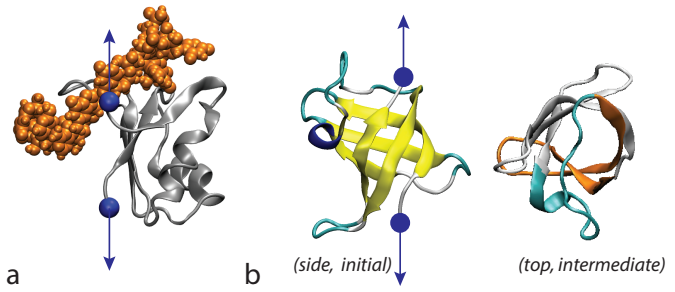

Fig. 1 Molecular mechanisms of protein unfolding upon force. (a) Mechanical unfolding of the RRM1 domain of the TDP-43 protein (gray) is much harder when a nucleic acid (orange) is bound, preventing the unzipping of two key β-sheets. After the nucleic acid strand has been removed, the protein can rapidly unzip and unfold. Here, the protein was stretched apart from its extremities (blue spheres) that were subject to mechanical force. This effect is reminiscent of that observed for a small globular protein, the cold shock CspA (b), for which mechanical and thermal stabilities were not correlated [34]. Indeed, the mesophilic variant was found to be mechanically more resistant, because of the position of a large loop (top view, cyan) that prevents unzipping of the two β strands that provide mechanical stability to the protein (top view, orange).

Finally, we explored and compared the unfolding mechanism of a small globular protein (cold shock CspA) under three external perturbations: mechanical force, shear flow, and thermal denaturation [39]. We had already shown, based on all-atom MD simulations, that the mechanical and thermal resistances of CspA homologues were not correlated at all, with a thermophilic variant being weaker mechanically-speaking than the mesophilic one [34]. Shear flow is of physiological importance in many biological processes, but it was not clear how it differed from a directional mechanical constraint more usually studied using single-molecule force spectroscopies. Modelling shear flow in all-atom MD with explicit solvent is not straightforward [18], and we used a coarse-grained representations of the protein in implicit solvent with inclusion of hydrodynamic interactions in a lattice-Boltzmann scheme [39–41]. While we could show that a shear rate of $\approx 10^9$ s^{-1} would lead to protein unfolding on a timescale similar to that of 300-pN directional force, the unfolding mechanisms were markedly different. The unfolding pathways under shear have strong similarities with those observed for thermal unfolding, which is not the case of a directional mechanical force which imposes unfolding along the end-to-end distance. Therefore, a shear perturbation probes a very different weakness of the protein fold as compared to mechanical force [39].

I have so far discussed the succession of molecular events leading to protein deformation and eventually unfolding upon force. Ongoing work in my group is addressing the differences in mechanical resistance of a protein depending on the direction of pulling, or the impact of redox conditions on crystallin mechanical unfolding [42]. I now discuss how some other studies have helped to characterize the unfolded state ensemble for a protein stretched upon force. How different is it from that of chemical/denaturation? What are the molecular degrees of freedom responsible for the chain elasticity? And how can this be used as a reporter of biomolecular recognition?

4.2 Unfolded state ensemble

Using all-atom MD simulations, we have studied the local and long range structures of unfolded ubiquitin either under force or following chemical/thermal denaturation [31]. The resulting structures were very different: a fully extended polypeptidic chain whose length followed a worm-like chain model above a few tens of piconewton, and a molten-globule, collapsed structure not much different from the native folded state in terms of spatial extension under thermal and chemical denaturation (Fig. 2). Even at relatively low forces of 30 pN, the formation of extensive secondary structure was not observed and no contacts usually formed between non-neighboring residues. In contrast, chemically-denatured structures exhibit many non-native contacts as well as non-native α-helices. Mechanical force alters the backbone dihedral degrees of freedom. As a consequence, the Ramachandran plots of chemically and force denatured proteins substantially differ from each other.

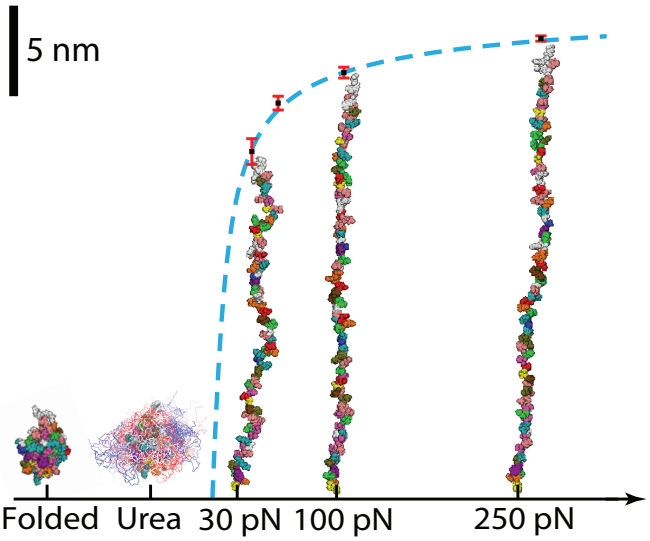

Fig. 2 Protein unfolded state ensemble. Comparison between representative structures for the native state (left), chemically- and thermally-denatured configurations (whose spatial extension is illustrated by blue to red ribbons that are superimposed to the native structure), and force-extended conformations at three different forces. In this regime, the chain extension (black squares, standard deviations in red) follows a worm like-chain model for polymer elasticity (dashed blue line). Figure adapted from *Proc. Natl. Acad. Sci. U.S.A.*, 111, 3413–3418 (2014).

4.3 Protein elasticity

The further characterization of the unfolded state ensemble could shed light on the molecular origins for the protein elasticity under force [43]. Usually, it is described at the mesoscopic level within the framework of the worm-like chain (WLC) model [2]. Although it has been successful for the interpretation of experimental data [3], especially at high forces, the WLC model lacks structural and dy-

namical molecular details associated with protein relaxation under force, that are key for the understanding of the molecular mechanisms behind the force's effects on the protein flexibility and reactivity. The first important check was that MD simulations in explicit solvent, using a state-of-the-art biomolecular forcefield, could reproduce the force extension profiles and the potential of mean force of a WLC model above ≈ 30 pN, with chain parameters (contour length and persistence length) in remarkable agreement with experimental results[30,44,45], providing a further validation of the approach and of the molecular models[43]. However, the WLC model alone is expected to break down at lower forces, when the barrier to collapse to a more compact and stable state is low enough to be observed on the experimental second timescale[46,47] (but not in the simulations). We then used our simulations to understand the molecular details for the WLC behavior. Bond lengths and bend angles appeared to be insensitive to the sub-nN forces usually used in experiments, but the backbone dihedral angles ($\phi = C_{(-1)}NC_\alpha C$ and $\psi = NC_\alpha CN_{(+1)}$) are dramatically affected by force in this regime (Fig. 3a), as clearly seen in their distributions (Fig. 3b and c) and in the Ramachandran plots that are changing with force. To provide a further validation of the connection between the dihedral degrees of freedom and protein elasticity, we then found that artificially altering the dihedral free-energy surfaces in the MD simulations lead to different persistence lengths, with larger free-energy barriers corresponding to a stiffer chain. Logically, residues with side-chains that significantly alter the backbone dihedral angles exploration (such as glycine) were seen to have an influence on the persistence length as well. Interestingly, the connection between the protein chain elasticity and the dihedral degrees of freedom could help design peptide-based tension sensors with tuned characteristics. From a dynamical perspective, we demonstrated that the applied force has no intrinsic effect on the diffusion coefficient along the pulling coordinate, in contrast to prior claims[48]. The applied force certainly changes the free-energy surface the protein is moving along, but it does not affect its intramolecular friction.

4.4 Impact for biomolecular recognition

We had thus clearly demonstrated the connection between the elastic properties of an unfolded protein, and its backbone dihedral degrees of freedom. We realized that these concepts allowed to rationalize very peculiar force-dependences that were observed in separate sets of experiments, as now discussed for two such systems.

In the first study, single-molecule force-spectroscopy experimental results suggested that a protein chaperone, DnaJ[49], was binding an initially extended and unfolded protein substrate with high specificity for a well-defined sequence[50]. Crucially, the interaction between the unfolded substrate and DnaJ was largely modulated by a force-dependent binding constant, that was surprisingly maximal at intermediate forces (150-170 pN), but not favored at lower or higher forces. The force-dependence first appeared surprising, but was to be understood within the framework we had developed before. We could indeed estimate, at a given force, the free-energetic cost to bring each dihedral angle from its stable value in the free substrate-protein fragment under a given force to its new constrained conformation imposed by DnaJ binding (Fig. 4a and b). This was obtained by estimating the probabilities $P(\phi,\psi)[F]$, for each couple of (ϕ,ψ) backbone dihedral angles of residues in the substrate ubiquitin binding sequence, to observe, at each force, the most probable (ϕ_0,ψ_0) combination measured in the bound complex in the absence of force. These were finally translated into a dihedral free-energy contribution:

$$\Delta G_{dih} = -k_B T \sum_{i \in complex} \ln P(\phi_0,\psi_0)[F] \qquad (1)$$

This free-energy cost was indeed higher at low or high forces in the MD simulations as compared to the optimal, intermediate force regime (Fig. 4d), in agreement with the experimental results (Fig. 4c). The MD trajectories were then instrumental in obtaining a structural interpretation for the measured values based on the phase space exploration of the Ramachandran plot for each of the substrate residues involved in the protein-protein binding complex. This work could thus demonstrate the important role that the unfolded and extended protein structure has in modulating the binding and thus the mechanical protein folding by a protein chaperone. Mechanical tension determines the local protein conformations able to bind to the chaperone, which could be understood in terms of backbone dihedral phase space exploration under force.

We have successfully applied similar concepts of conformational selection in another joint experimental–simulation study to explain the peculiar force-dependence of the TeV protease[51] cleavage of a protein substrate, which again suggested an optimal force-range where binding and thus reaction were optimal[52]. At low force, the large spectrum of accessible states slowed down the association, and above the optimal force the probability of

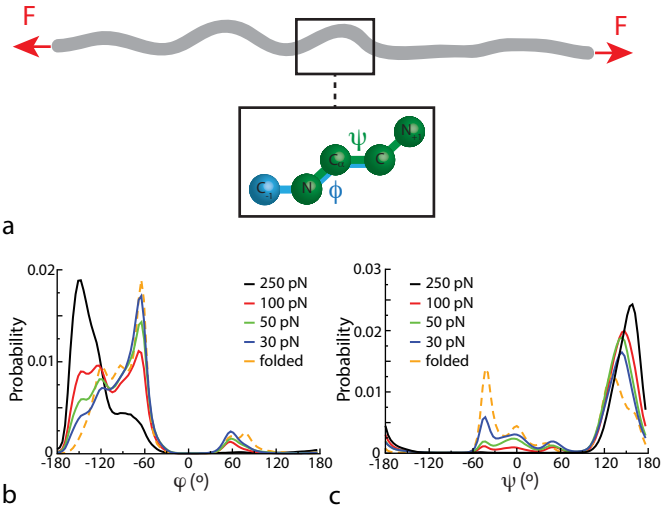

Fig. 3 Protein elasticity upon force extension. (a) When a protein chain (gray) is extended upon force (red arrows), it involves significant distortion of its backbone dihedral angles $\phi = C_{(-1)}NC_\alpha C$ (blue) and $\psi = NC_\alpha CN_{(+1)}$ (green), as clearly visible in the distributions (b and c) that are very sensitive to force. Figure adapted from *Proc. Natl. Acad. Sci. U.S.A.*, 110, 3847–52 (2013).

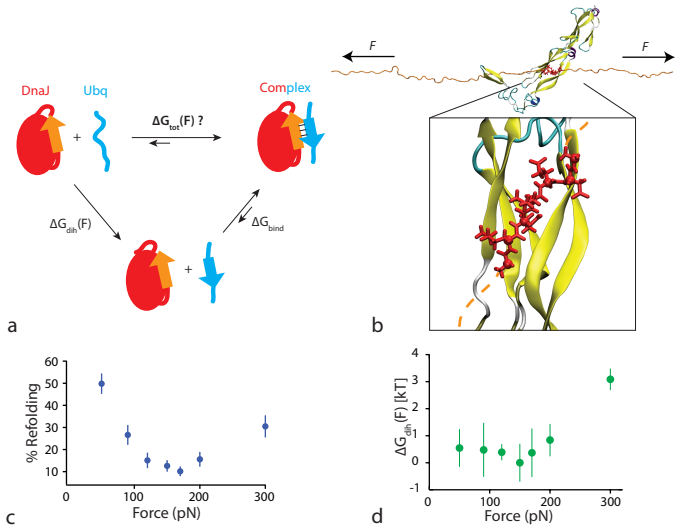

Fig. 4 Model for the force-dependence of DnaJ binding. (a) Schematic picture of the decomposition of binding free-energy ΔG_{tot} into a force-dependent, conformational term $\Delta G_{dih}(F)$ and a force-independent (because of the absence of change in the end-to-end distance) interaction term ΔG_{bind}. (b) Snapshot of the interaction of DnaJ (PDB: 1NLT) with the unfolded and mechanically stretched ubiquitin (orange ribbon). Inset: zoom on the reconstructed DnaJ-ubiquitin fragment system. The bonds of the amino-acids from the consensus sequence are shown as red tubes and the corresponding C_α as red balls. (c) Dependency of the refolding % (which is a proxy of the binding affinity, a lower refolding fraction corresponding to a higher affinity) with the pulling force. (d) Free energy contribution $\Delta G_{dih}(F)$ corresponding to the overall energetic cost underlying the force-induced remodelling of the dihedral angles of the ubiquitin interaction fragment upon DnaJ binding. Error bars correspond to the standard deviations observed among independent trajectrories, and the data is normalized with respect to the lowest free-energy value (observed at 150 pN).

visiting a compatible conformation dropped abruptly. At higher forces, the loss of substrate flexibility slowed the association via a strong free-energy penalty. As compared to our previous study, we have considerably improved our model to make it transferable to any substrate sequence[52]. In particular, we have used enhanced sampling methods to obtain the potential of mean force along the ϕ and ψ backbone dihedral angles in the absence of force, and similar maps were generated at any force by estimating the additional work performed by force along these coordinates (whose projection along the end-to-end distance, on which the force acts, can be estimated). Therefore, the backbone dihedral free-energy model can be used to estimate, for any binding sequence and with the knowledge of the binding structure — and thus the target backbone dihedral angles (ϕ_0,ψ_0), the force-dependence of the protein–protein interaction in the pN–force range of the experiments.

The examples discussed so far have illustrated the power of particle-based approaches (either coarse-grained or all-atom) in order to provide molecular-scale interpretation of single-molecule force-spectroscopy data. Computational approaches allowing to tackle different questions are now addressed. These include either systems that are too large and/or that require timescales that are out or reach for molecular dynamics simulations, or some involving chemical reactivity under force, for which the a classical particle-based approach would fail as bonds cannot be broken or formed during a forcefield-based simulation. I will thus now focus on these two extreme regimes: first,, I will discuss how much more simplistic models based on Langevin dynamics on a 1-D free-energy surface to understand the effect of protein segmentation on the work exerted by force, and second, how quantum calculations can be used in order to understand the reactivity of disulfide bridges exposed under force.

5 Protein segmentation and free-energy landscape of a polyprotein

As mentioned in the previous sections, the elasticity of unfolded proteins or nucleic acids under force is successfully described by the worm-like chain (WLC) model. The chain is seen as a semi-flexible robe of total (or contour) length L_c, and with a persistence length p, which reports on the local flexibility of the chain. Under a force L, the free-energy of the chain by a length L is given by:

$$U_{WLC}(L;p,L_c) = \frac{k_B T}{p}\left(\frac{L_c}{4}\left[\frac{1}{1-\frac{L}{L_c}}-1\right]-\frac{L}{4}+\frac{L^2}{2L_c}\right). \quad (2)$$

In contrast, folded proteins are very stiff and usually cannot extend without losing their native structure, and thus their mechanical stability. For example, in a typical constant-force experiment, no change in the end-to-end distance would be observed until the mechanical clamp maintaining the native fold breaks, releasing the entire chain up to the equilibrium WLC distance at this force (minimum in Equation 2).

The situation gets more complicated when the chain is made of a polyprotein with several protein domains. Some interest was raised to study the effect of this segmentation of the chain on the underlying free-energy landscape along the end-to-end distance, which can be accessed in the experiments. Understanding the peculiar patterns measured in the PMF as a function of the extension requires to go beyond the simple traditional picture of the WLC behavior of a polymer chain. We derived a model in order to express, at a given force, the free-energy of a polyprotein as a function of its extension. This model could be used both to understand the specific profile of a polyprotein PMF, and later to run Langevin dynamics simulations on this 1-D free-energy landscape in order to run numerical experiments[53,54].

5.1 High force limit for extension

Let's consider a polymer (or a polyprotein) with N domains, out of which m domains are unfolded and n domains are folded. After unfolding, each domain will behave as a worm-like chain with a persistence length p (considered as constant) and a contour length L_c. This polymer is pulled under a force F. We call L_F the equilibrium length of an unfolded domain at this force F, which can be found by taking the first derivative of Equation 2. At high force, the length of the folded domain L_{fold} is considered as negligible as compared to L_F or L_c. The unfolding barriers of the native domains are also small compared to the elastic free-energy terms and can be neglected.

Under these assumptions, for $x \in [0; mL_F]$, the energy of the chain is given by (for simplification, the p dependence of U_{WLC} is now omitted in the notations) :

$$U(x) = U_{WLC}[x; mL_c] - F \times x \quad (3)$$

as the m unfolded domains will extend up to their total contour length. After extending this chain, one of the folded domains may unfold and will be subsequently extended. In that case, we have a chain of contour length $(m+1)L_c$ which will be extended between mL_F (the final point of the previous extension) and $(m+1)L_F$. The process is repeated until all domains unfold and extend upon force. As a consequence, U can be written in a recursive way (Fig. 5). For any $x \in [(m+i-1)L_F; (m+i)L_F]_{i=1,\ldots,n}$, we have:

$$\begin{aligned} U(x) = U_{WLC}[x; (m+i)L_c] - U_{WLC}[(m+i-1)L_F; (m+i)L_c] \\ + U((m+i-1)L_F) - F \times x \end{aligned} \quad (4)$$

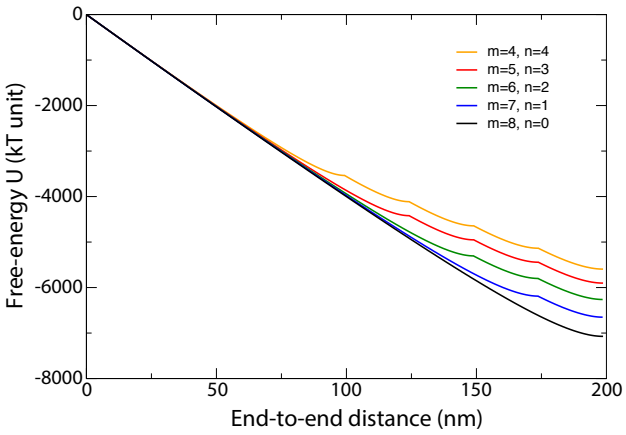

Fig. 5 Model free-energy of for a typical polyprotein under experimental force, as a function of its extension, using the expression of Equation 4 and varying the number of initially-unfolded domains.

5.2 Low force and refolding

In other experiments involving magnetic tweezers, polyproteins could be manipulated under a very low force (typically, 2–20 pN) for a long period of time, which allowed dynamically following single domain unfolding/refolding under force. Equation 4 becomes no longer valid for several reasons: first, L_F is not significantly larger than the folding length L_{fold}, and second, one has to somehow include in the model a term reporting on the stabilization of the native state, and on the free-energy barrier to unfold the protein before it extends.

To simplify the expressions, we assumed all domains are initially folded ($m = 0$). In order to model the folding of the protein into a stable structure, whose relative free-energy difference with the unfolded state becomes relevant at low forces, as well as the folding/unfolding free-energy barrier, we used a simple activated Morse (aM) potential, consisting of a regular Morse potential together with a Gaussian barrier :

$$U_{aM}(x) = E_0 \left(1 - e^{-2bx/x_0}\right)^2 + Ae^{-(x-x_b)^2/2\sigma^2} \quad (5)$$

where x_0 is the position of the minimum, E_0 its depth, b a dimensionless parameter; A is the amplitude of the Gaussian barrier, x_b the position of the barrier, σ its width. We finally obtained that for any $x \in [nL_{fold} + (i-1)(L_F - L_{fold}); nL_{fold} + i(L_F - L_{fold})]_{i=1,\ldots,n}$:

$$\begin{aligned} U(x) \quad = \quad & U_{WLC}[x; iL_c] - U_{WLC}[(i-1)L_F + L_{fold}; iL_c] \\ & + U_{aM}(x - (i-1)(L_F - L_{fold}) - (n-1)L_{fold})) \\ & + U((i-1)L_F + L_{fold}) \\ & - F \times (x - nL_{fold}) \end{aligned} \quad (6)$$

This model could be used to rationalize the force-extension profiles of polyproteins over a wide range of forces in the experimental measurements, as well as to run 1-D Brownian or Langevin dynamics simulations on the underlying free-energy surface in order to generate and analyze force-extension traces very similar to those of the experiments[54].

6 Chemical reactivity of exposed disulfide groups

An entire subfield of single-molecule force-spectroscopy has been devoted to the study of mechanochemistry, i.e., the impact of mechanical force on covalent bond formation/rupture[55,56]. As mentioned above, in the force-range usually employed in the experiments that involve proteins, chemical bonds do not break, and forces mainly alter the dihedral degrees of freedom of the extended polypeptidic chain. However, there is at least one specific case where force has an effect on chemical reactivity in proteins, which is the disulfide bond reduction. It is thus not surprising if this reaction has been extensively studied using experimental approaches[4,57] as well as calculations and simulations[58–62]. Beyond the fundamental interest in studying this phenomenon, it also has real biological implications. Indeed, a widespread natural strategy to regulate protein extensibility lies in the presence of disulfide bonds, which enhance protein stiffness, and whose presence or not can be regulated in an organism.

Force can actually affect disulfide bond rupture in several, and sometime complementary, ways: first, it usually allows to expose cryptic disulfide bonds that are buried into the native protein state and thus not solvent-exposed in the absence of force; second, once the disulfide bond is exposed, it can trigger local conformational changes by acting on dihedral degrees of freedom that in turn tune the accessibility of the disulfide bond for a nuclephilic attack[60]; finally, force also directly lowers the free-energy barrier for the reaction, which in general, implies that higher forces result in faster measured kinetics[58,60].

Tackling these systems with simulations is challenging, especially because chemical reactivity is involved. This typically requires a quantum treatment of, at least, the chemical groups involved in the reaction. A number of pioneering simulation studies employed state-of-the-art ab-initio molecular dynamics[60,61] of a

model reactive system upon force (typically neglecting the protein environment and focusing on a model compound with a disulfide bond). In these simulations, the molecular interactions are calculated at the quantum level for the electronic degrees of freedom, with classical evolution of the nuclei on the corresponding energy landscape. However, they are presently too costly to be deployed at a large scale on a real protein system.

Our own approach to the problem relied on much simpler strategies that used static DFT calculations to rationalize the kinetics and thermodynamics of some intriguing experimental results[63,64]. The experiments worked as follows. A constant force was applied to a polyprotein composed of eight identical domains, each containing two cysteines, which initially form a structurally buried disulfide bond. A specifically designed five-pulse force protocol measured the time elongation of the protein with organic nucleophiles in solution, which allowed to capture each individual disulfide bond rupture and reformation event. We initially compared these events in the presence of two different nucleophiles in solution, L-cysteine (Cys) and Homo-cysteine (HCys)[63]. The protein was first partially unfolded up to a point where the disulfide bond became exposed. This bond could then be attacked by a solution nucleophile (Cys or HCys), resulting in the full chain release. Upon force-quenching, reduced folded protein was observed to be dramatically higher for HCys as compared to Cys, suggesting that disulfide bond was not reforming in the presence of HCys, while it was in the presence of Cys.

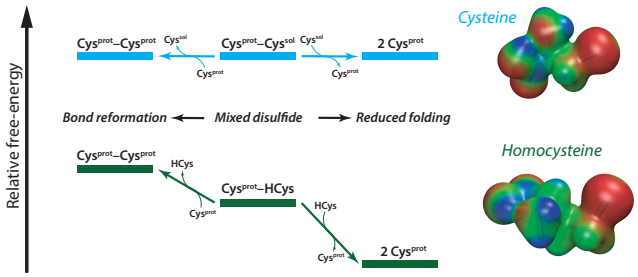

Fig. 6 Thermodynamics and kinetics of disulfide bond formation and reduction by thiol nucleophiles. The relative free-energy costs were estimated using DFT calculations. Once a mixed disulfide is formed between a protein cysteine and a free nucleophile (Cys or HCys), further reduction of the mixed disulfide is favored in the presence of free HCys, leading to reduced folding, while bond reformation and reduced folding are equiprobable in the presence of free Cys. On the right, the charge distribution of the free-nucleophile is indicated, from red (negative) to green (0) to blue (positive). We demonstrated that the charge was correlated with the kinetics of disulfide bond reduction.

To obtain a more precise understanding of the thermodynamically allowed reactivity between the different thiol species (i.e. (i) the initial native disulfide bond; (ii) the mixed disulfide bond species with a bound solution nucleophile; (iii) the free cysteine thiol after disulfide bond cleavage and (iv) the free small thiol nucleophile species in solution), we performed DFT calculations on model species using the expanded 6-311+G(d,p) basis set and the B3LYP functional, as used before in the context of thiol-disulfide exchange[63]. The structures were first optimized in an implicit water solvent consisting of a dielectric medium, and frequency calculations were performed on these optimized structures. We finally obtained the free-energies by summing the gas-phase energy at 0 K, the solvation free-energy component, the zero-point energy, and a thermal correction at ambient temperature. Of course, such a simplistic approach cannot account for free-energy contributions that would stem from a protein conformational change upon formation of a native or mixed disulfide, but it could still provide very valuable information and help rationalize the experimental observations. For example, we found that the rupture of the mixed non-equivalent disulfide between Cys_{prot}–HCys by free HCys in solution is favorable (−4 kcal/mol). However, the attack of the same mixed disulfide by the freed protein cysteine is not favorable (+2.6 kcal/mol). This thus provided an explanation for the reduced protein refolding in the presence of HCys (Fig. 6).

In a subsequent work, we improved our original approach[64]. For example, it is known that explicitly taking into account explicit water molecules improves the agreement between calculated and experimental pKa values on these species. We therefore repeated our calculations with one explicit water molecule donating an hydrogen-bond to the sulfur atom in the thiolate and thiol species. We also used another functional that was shown to perform well in these conditions and a more extended and diffuse bases set. We obtained overall better agreement with the experiments, and the correct trends were successfully predicted for almost 10 different organic thiols. Such a naive strategy could thus reasonable well account for the experimental results. Further free-energy calculations at an increased level of description (such as a mixed quantum/classical approach coupled with a force perturbation) could bring a decisive molecular point of view in order to rationalize the effects of e.g. solvation or bond orientation and accessibility.

7 Conclusions and perspectives

In this Feature article, some of our recent work devoted to the molecular interpretation of single-molecule force-spectroscopy experiments has been reviewed. A critical aspect in such an approach is to choose the right level of description and usually, simplification, that allow to tackle the biophysical and chemical phenomena probed in the experiments. Using a variety of computational techniques spanning several orders of magnitude of length- and timescales, we have addressed questions ranging from the chemical reactivity of exposed disulfide groups to the mechanisms of protein unfolding upon force. Our effort is of course not isolated, and the work of many other groups should also be acknowledged.

While they can offer a powerful framework to guide the interpretation of experimental data, one should keep in mind that each computational approach also faces many limitations. Choosing the right strategy to tackle a given biophysical or biochemical question involves a critical assessment of what can be done, and what cannot. However, the increase in computational power and the development of better molecular models and algorithms have enabled to address challenges of growing complexity since the first hallmark steered MD simulations some 25 years ago[11–13].

The future certainly offers some bright perspectives for the field of mechanobiology in general, which includes a continued

interest in using computational approaches as a complement to experimental investigations. Some of our current efforts in my group are devoted to understanding the molecular mechanisms of catch-bond phenomena, whereby a biomolecular interaction is increased upon application of tensile force. The complex molecular mechanisms associated with mechanosensing and mechanotransduction also offer a challenging (by the involved length- and timescales) but fascinating playground in which the computational biophysicists and biochemists will continue to play an important role.

Acknowledgements

This work was supported by the "Initiative d'Excellence" program from the French State (Grant "DYNAMO", ANR-11-LABX-0011-01). I am also grateful for many experimental collaborations developed over the years which have made the work presented here possible. I am particularly thankful to Sergi Garcia-Manyes, Julio Fernandez, Emmanuel Farge, David Giganti and Ronen Berkovich, and to their students. I also thank collaborators and coworkers from the computational field, Bruce Berne and Fabio Sterpone, as well as students involved in some of the computational investigations discussed here.

Notes and references

1 K. C. Neuman and A. Nagy, *Nat. Methods*, 2008, **5**, 491–505.
2 C. Bustamante, J. F. Marko, E. D. Siggia and S. Smith, *Science*, 1994, **265**, 1599–1600.
3 M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez and H. E. Gaub, *Science*, 1997, **276**, 1109–1112.
4 J. Alegre-Cebollada, R. Perez-Jimenez, P. Kosuri and J. M. Fernandez, *J. Biol. Chem.*, 2010, **285**, 18961–18966.
5 R. Merkel, P. Nassoy, A. Leung, K. Ritchie and E. Evans, *Nature*, 1999, **397**, 50–53.
6 M. Schlierf, H. Li and J. M. Fernandez, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 7299–7304.
7 T. Iskratsch, H. Wolfenson and M. P. Sheetz, *Nat. Rev. Mol. Cell Biol.*, 2014, **15**, 825–833.
8 J. K. Freundt and W. A. Linke, *J. Applied Physiol.*, 2019, **126**, 1474–1482.
9 T. A. Springer, *Blood*, 2014, **124**, 1412–1425.
10 W. E. Thomas, E. Trintchina, M. Forero, V. Vogel and E. V. Sokurenko, *Cell*, 2002, **109**, 913–923.
11 H. Grubmüller, B. Heymann and P. Tavan, *Science*, 1996, **271**, 997–999.
12 H. Lu and K. Schulten, *Proteins: Struct., Funct., Bioinf.*, 1999, **35**, 453–463.
13 B. Isralewitz, M. Gao and K. Schulten, *Curr. Opin. Struct. Biol.*, 2001, **11**, 224–230.
14 F. Franz, C. Daday and F. Gräter, *Curr. Opin. Struct. Biol.*, 2020, **61**, 132–138.
15 P. Cossio, G. Hummer and A. Attila, *Proc. Natl. Acad. Sci. U.S.A.*, 2015, **112**, 14248–14253.
16 R. Berkovich, R. I. Hermans, I. Popa, G. Stirnemann, S. Garcia-Manyes, B. J. Berne and J. M. Fernandez, *Proc. Natl. Acad. Sci. U.S.A.*, 2012, **109**, 14416–14421.
17 G. M. Nam and D. E. Makarov, *Protein Sci.*, 2016, **25**, 123–134.
18 G. Stirnemann, *J. Phys. Chem. B*, 2022, **126**, 1365–1374.
19 T. G. Graham and R. B. Best, *J. Phys. Chem. B*, 2011, **115**, 1546–1561.
20 B. Mondal, D. Thirumalai and G. Reddy, *J. Phys. Chem. B*, 2021, **125**, 8682–8689.
21 P. I. Zhuravlev, M. Hinczewski and D. Thirumalai, *J. Phys. Chem. B*, 2021, **125**, 1799–1805.
22 R. Tapia-Rojo, J. J. Mazo and F. Falo, *J. Chem. Phys.*, 2019, **151**, 185105.
23 L. Sun, J. K. Noel, J. I. Sulkowska, H. Levine and J. N. Onuchic, *Biophys. J.*, 2014, **107**, 2950–2961.
24 M. J. Hartmann, Y. Singh, E. Vanden-Eijnden and G. M. Hocky, *J. Chem. Phys.*, 2020, **152**, 244120.
25 J. J. Booth and D. V. Shalashilin, *J. Phys. Chem. A*, 2016, **120**, 700–708.
26 C. Mücksch and H. M. Urbassek, *J. Chem. Theo. Comput.*, 2016, **12**, 1380–1384.
27 D. Gomez, W. J. Peña Ccoa, Y. Singh, E. Rojas and G. M. Hocky, *J. Phys. Chem. B*, 2021, **125**, 12115.
28 P. Tiwary and M. Parrinello, *Phys. Rev. Lett.*, 2013, **111**, 230602.
29 W. J. P. Ccoa and G. M. Hocky, *J. Chem. Phys.*, 2022, **156**, 125102.
30 M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke and J. M. Fernandez, *Proc. Natl. Acad. Sci. U.S.A.*, 1999, **96**, 3694–3699.
31 G. Stirnemann, S. G. Kang, R. Zhou and B. J. Berne, *Proc. Natl. Acad. Sci. U.S.A.*, 2014, **111**, 3413–3418.
32 K. M. Tych, M. Batchelor, T. Hoffmann, M. C. Wilson, M. L. Hughes, E. Paci, D. J. Brockwell and L. Dougan, *Langmuir*, 2016, **32**, 7392–7402.
33 K. M. Tych, M. Batchelor, T. Hoffmann, M. C. Wilson, E. Paci, D. J. Brockwell and L. Dougan, *Soft Matt.*, 2016, **12**, 2688–99.
34 G. Stirnemann and F. Sterpone, *J. Phys. Chem. Lett.*, 2017, **8**, 5884–5890.
35 J. Schönfelder, R. Perez-Jimenez and V. Muñoz, *Nat. Commun.*, 2016, **7**, 1–8.
36 A. E. Beedle, A. Lezamiz, G. Stirnemann and S. Garcia-Manyes, *Nat. Commun.*, 2015, **6**, 1–9.
37 J. C. Röper, D. Mitrossilis, G. Stirnemann, F. Waharte, I. Brito, M. E. Fernandez-Sanchez, M. Baaden, J. Salamero and E. Farge, *eLife*, 2018, **7**, e33381.
38 Y. J. Wang, P. Rico-Lastres, A. Lezamiz, M. Mora, C. Solsona, G. Stirnemann and S. Garcia-Manyes, *J. Phys. Chem. Lett.*, 2018, **9**, 3800–3807.
39 O. Languin-Cattoën, S. Melchionna, P. Derreumaux, G. Stirnemann and F. Sterpone, *J. Phys. Chem. B*, 2018, **122**, 11922–11930.
40 F. Sterpone, P. Derreumaux and S. Melchionna, *J. Chem. Theo. Comput.*, 2015, **11**, 1843–1853.
41 O. Languin-cattoën, E. Laborie, D. O. Yurkova, S. Melchionna,

P. Derreumaux, A. V. Belyaev and F. Sterpone, *Polymers*, 2021, **13**, 3912.

42 M. Mora, S. Board, O. Languin-Cattoën, L. Masino, G. Stirnemann and S. Garcia-Manyes, *Nano Lett.*, 2022, 10.1021/acs.nanolett.2c00043.

43 G. Stirnemann, D. Giganti, J. M. Fernandez and B. J. Berne, *Proc. Natl. Acad. Sci. U.S.A*, 2013, **110**, 3847–52.

44 S. R. K. Ainavarapu, J. Brujic, H. H. Huang, A. P. Wiita, H. Lu, L. Li, K. A. Walther, M. Carrion-Vazquez, H. Li and J. M. Fernandez, *Biophys. J.*, 2007, **92**, 225–233.

45 M. Carrion-Vazquez, H. B. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser and J. M. Fernandez, *Nat. Struct. Mol. Biol.*, 2003, **10**, 738–743.

46 R. Berkovich, S. Garcia-Manyes, M. Urbakh, J. Klafter and J. M. Fernandez, *Biophys. J.*, 2010, **98**, 2692–2701.

47 I. Popa, J. A. Rivas-Pardo, E. C. Eckels, D. J. Echelman, C. L. Badilla, J. Valle-Orero and J. M. Fernández, *J. Am. Chem. Soc.*, 2016, **138**, 10546–10553.

48 B. S. Khatri, K. Byrne, M. Kawakami, D. J. Brockwell, D. A. Smith, S. E. Radford and T. C. B. McLeish, *Faraday Discuss.*, 2008, **139**, 35–51.

49 H. H. Kampinga and E. A. Craig, *Nat. Rev. Mol. Cell Biol.*, 2010, **11**, 579–592.

50 J. Perales-Calvo, D. Giganti, G. Stirnemann and S. Garcia-Manyes, *Sci. Adv.*, 2018, **4**, eaaq0243.

51 C. López-Otín and C. M. Overall, *Nat. Rev. Mol. Cell Biol.*, 2002, **3**, 509–519.

52 M. E. Guerin, G. Stirnemann and D. Giganti, *Proc. Natl. Acad. Sci. U.S.A*, 2018, **115**, 11525–11530.

53 R. Berkovich, V. I. Fernandez, G. Stirnemann, J. Valle-Orero and J. M. Fernández, *J. Phys. Chem. B*, 2018, **9**, 4707–4713.

54 J. Valle-Orero, E. Eckels, G. Stirnemann, I. Popa, R. Berkovich and J. M. Fernandez, *Biochem. Biophys. Res. Commun.*, 2015, **460**, 434.

55 J. Ribas-Arino and D. Marx, *Chem. Rev.*, 2012, **112**, 5412–5487.

56 D. E. Makarov, *J. Chem. Phys.*, 2016, **144**, 030901.

57 S. Garcia-Manyes, J. Liang, R. Szoszkiewicz, T.-L. Kuo and J. M. Fernández, *Nat. Chem.*, 2009, **1**, 236–242.

58 I. B. Baldus and F. Gräter, *Biophys. J.*, 2012, **102**, 622–629.

59 W. Li, I. B. Baldus and F. Gräter, *J. Phys. Chem. B*, 2015, **119**, 5386–5391.

60 P. Dopieralski, J. Ribas-Arino, P. Anjukandi, M. Krupicka, J. Kiss and D. Marx, *Nat. Chem.*, 2013, **5**, 685–691.

61 P. Dopieralski, J. Ribas-Arino, P. Anjukandi, M. Krupicka and D. Marx, *Angew. Chem. Int. Ed. Engl.*, 2016, **55**, 1304–1308.

62 F. Hofbauer and I. Frank, *Chem. Eur. J.*, 2012, **18**, 16332–16338.

63 A. E. Beedle, M. Mora, S. Lynham, G. Stirnemann and S. Garcia-Manyes, *Nat. Commun.*, 2017, **8**, 1–11.

64 A. E. Beedle, M. Mora, C. T. Davis, A. P. Snijders, G. Stirnemann and S. Garcia-Manyes, *Nat. Commun.*, 2018, **9**, 3155.