Single molecule studies of folding and misfolding mechanisms of proteins using optical tweezers.

By

Mohsin Mubarak Naqvi

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR OBTAINING THE DEGREE OF

DOCTOR OF PHILOSOPHY

In

PHYSICAL SCIENCES

THE SCHOOL OF GRADUATE STUDIES

CYCLE XXVII

(2012-2015)

UNIVERSITY OF MODENA AND REGGIO EMILIA
Department of Physics, Informatics and Mathematics

Doctoral Supervisor

Prof. Ciro Cecconi

Università degli Studi di Modena e Reggio Emilia
Dipartimento di Scienze Fisiche Informatiche e Matematiche

Via Campi 213/A – 41122 Modena (Italy)
Tel. +39 0592055243 - Fax +39 0592055235 – http: www.unimore.it
P.IVA 00427620364

Dottorato di ricerca in : Physics and Nanosciences

(nell'ambito della Scuola di dottorato in Physics and Nanosciences)

Ciclo : **XXVII**

Titolo tesi

"Single molecule studies of folding and misfolding mechanisms of proteins using optical tweezers".

Candidato: **Mohsin Mubarak Naqvi**

Relatore (Tutor): **Prof. Andrea Alessandrini**

Correlatore (Tutor): **Prof. Ciro Cecconi**

Coordinatore del Dottorato: **Prof. Marco Affronte**

Direttore della Scuola di dottorato: **Prof. Marco Affronte**

Marco Affronte
Prof. Marco Affronte

Prof. Ciro Cecconi

Dedicated to my
Father: Mr. Syed Mubarak,
Mother: Mrs Huma Tausif Naqvi,
Brother: Hasan Mubarak Naqvi,
Sister : Batool Naqvi.

Preface

In this PhD thesis, three projects are explained that were conducted under the supervision of Prof. Ciro Cecconi at Department of Physics, Informatics and Mathematics, University of Modena and Reggio Emilia Italy. Two projects, based on single molecule manipulation of the protein - human neuronal calcium sensor -1 (NCS-1) using optical tweezers, are the result of collaboration with Prof. Birthe B. Kragelund at Structural Biology and NMR laboratory, University of Copenhagen, Denmark. The preparation of protein samples for single molecule manipulation studies and experiments involving bulk measurements in the projects were conducted and analyzed at Structural Biology and NMR laboratory University of Copenhagen, Denmark. The single molecule manipulation experiments were conducted at optical tweezers lab, Department of Physics, Informatics and Mathematics, University of Modena and Reggio Emilia Italy and were analyzed in collaboration with Dr. Alessandro Mossa, Department of Physics, University of Bari Italy. Another project involving the amyloidogenesis studies of β2-Microglobulin was carried out in collaboration with Prof. Vittorio Bellotti, Wolfson Drug Discovery Unit, University College London. In this project theoretical calculations were done at Department of Physics, Informatics and Mathematics, University of Modena and Reggio Emilia Italy.

Acknowledgements

First of all I would like to acknowledge my supervisor Prof. Ciro Cecconi for helping me immensely during the course of my PhD work. He worked long hours with me and taught me the fundamentals of single molecule force spectroscopy and devoted lot of his time in helping me understand the recent studies done in this field. He guided me during the tough times of my PhD and always shown great interest in making me a better researcher. In the field of single molecule protein folding studies my supervisor trained me in understanding three key problems – protein folding, protein misfolding and amyloidogenesis which has opened numerous avenues for me to work ahead. I feel very privileged that I did my PhD under his supervision and learned so many things from him.

I want to also acknowledge Dr. Alessandro Mossa who helped me a lot in learning the various statistical tools such as hidden Markov model and Crooks Fluctuation theorem. He helped me in developing theoretical skills for analyzing and interpreting experimental data. He always took interest in clearing any doubt or questions I had and explained them in detail to me. Without his support it would have been very difficult for me to publish during my PhD.

I would now like to acknowledge Prof. Birthe B. Kragelund in University of Copenhagen with whom our group collaborated for many years and I got the opportunity to be involved in two interesting projects (NCS-1 folding and misfolding), in collaboration with her. I want to particularly thank Dr. Pétur O. Heidarsson in the research group of Prof. Birthe B. Kragelund with whom I worked in two projects of my PhD. He helped me in understanding the biochemistry work involved in my thesis projects and in the successful completion of the project relating to NCS-1 misfolding, which is published in PNAS.

I also would like to acknowledge Prof. Vittorio Bellotti at University College London, who collaborated with our group in the project amyloidogenesis of β2-Microglobulin which was published in journal of biological sciences and in other projects which we are working on at present.

Lastly, I want to acknowledge my colleagues Punam Sonar, Dr. Mariela R. Otazo and Dr. Immanuel Valpapuram, who helped me in carrying out my single molecule experiments and other lab activities. I also would like to thank Punam Sonar for the memorable time we spent in our various trips in Italy and other parts of Europe and in the conferences and lectures we attended together.

Mohsin Mubarak Naqvi,
Department of Physics, Mathematics and Informatics,
CNR NANO,
University of Modena and Reggio Emilia,
Via Campi, 213/A - 41125 Modena, Italy.

Table of Contents

Abstract

The PhD work presented in this thesis is devoted to the understanding of the basic rules governing three major biophysical processes – protein folding, protein misfolding and amyloidogenesis. We have used a custom-built dual-beam optical tweezers set up to investigate the off- and on-pathway trajectories of the protein – human neuronal calcium sensor -1 (NCS-1). A novel experimental approach has been used to investigate the effect of Ca^{2+} concentration on the misfolding probability of the protein. The analysis of the single molecule data was carried out using advanced statistical tools, such as the hidden Markov model, to gain insight into the salient features of the energy landscapes underlying on- and off-pathway trajectories of NCS-1. In a separate project, theoretical models have been developed to describe quantitatively the shear and hydrophobic forces acting on a protein in a laminar flow, with the aim of elucidating the molecular mechanisms triggering amyloidogenesis of the protein β2-microglobulin. The work presented in this thesis is divided in 5 chapters, includes 16 figures and a total of 163 pages.

1. INTRODUCTION

In the introduction of the thesis an overview has been provided regarding the protein folding problem which includes brief discussions on protein misfolding and amyloidogenesis of proteins. In the later sections, single molecule manipulation of proteins using optical tweezers has been explained which includes the theory of optical tweezers set up and the recent progress in its instrumentation. The subsequent section of the introduction includes the theoretical models employed in single molecule manipulation studies. In last sections, the structure and functionality of the model proteins studied in the dissertation work along with the objectives behind each project are explained.

1.1 *Protein folding problem*

1.1.1 Protein Folding: its relation to structure and functionality of a protein.

How biomolecules self assemble and attain a particular structure to carry out their respective functions with great precision, is a major mystery for modern science to unravel. The mechanism that governs the folding of proteins into unique three dimensional structures is a complex phenomenon of biomolecular self assembly. These unique structures which protein molecules attain during their folding process, involves the assembly of functional groups in close proximity that enable them to carry out numerous biological functions. Moreover, for proteins to be functionally active they need to correctly fold into their lowest energy native state configuration and incorrect or misfolding of proteins result in many types of diseases [1]. The stochastic mechanism by which a polypeptide chain, acquire its lowest energy configuration (in a time scale of microseconds), among an ensemble of numerous possible structures is still elusive. Moreover, if a molecule reaches its lowest energy native state, through all its possible conformations then it should take an astronomical amount of time, as stated in the Levinthal's paradox [2]. Understanding this complex phenomenon will not only help in the cure of various diseases related to misfolding of proteins but will give us great insight into how such tiny molecular machines carry out their functions in our body. The protein folding problem can be summarized in terms of two key questions: (1) for a given sequence of amino acids and protein environment what are the factors that determines it's native or functional state configuration? (2) What is the thermodynamics and kinetics of the protein folding reaction and how the energy landscape of the protein look like? The second problem can be understood from the energy landscape theory of protein folding.

1.1.2 Energy landscape theory; the funnel view of protein folding:

The Levinthal's paradox of protein folding [2], encouraged researchers to see this problem as two mutually exclusive events – reaching the global minimum state and doing that at a fast speed [3]. These two simultaneous events were described as thermodynamic and kinetic parameters of protein folding reaction. The thermodynamic viewpoint describes the folding process as path independent where the molecule attains its lowest free energy state without depending on the initial denaturing conditions while the kinetic viewpoint states that the speed of folding is path dependent and is fast because it depends on the initial denatured condition of the protein. The combination of these two processes results in the energy landscape theory, that explains the fundamental mechanism of structure formation of proteins and nucleic acids [4, 5]. The energy landscape describes protein folding as a statistical mechanics process involving an ensemble of allowed conformational states of the molecule. In this theory, protein folding can be viewed as a stochastic diffusive process, in which the molecule tumbles towards its native state defined as the global minimum energy state configuration [6]. This stochastic search of the native state takes place over a multidimensional surface representing a configuration space with free energy of each possible conformation. Fig 1 (a) describes a multidimensional view of the energy landscape with a funnel shaped energy minima, involving various parallel pathways towards it [3]. Since the folding of a molecule results in a decrease in its entropy, the search for the low energy minima on the energy landscape is a competitive process with it. The multidimensional energy landscape can have many local minimas depicting a rugged energy surface, which can slow down the folding of the molecules due to the phenomenon of kinetic partitioning [7, 8]. The ruggedness or local minimas in the energy landscape occur due to the competitive process of low entropy and free energy of the molecules. The energy traps or minimas act as hurdle in the folding process resulting in the intermediate states during the native state journey Fig. 1 (A). The intermediate states which the protein traverses, can be on pathway towards its native state or off pathway misfolded states causing aggregation and resulting in numerous diseases subsequently [1]. Fig. 1 (B) depicts the one dimensional representation of the energy landscape in which the full multidimensional characteristics are projected onto a single reaction coordinate that can possibly define the folding reaction. The choice of the reaction coordinate depends on the experimental observable that can best explain the folding behavior. This simple approach has helped in extracting various features like barrier heights, intermediate states, multiple pathways and ruggedness of energy landscape of numerous proteins. However the full multidimensional energy landscape of a protein is yet to be determined and it is believed that to reconstruct such energy landscape, single molecule reaction at atomic scale and with temporal resolution of bond rotations will be required [9]. In this thesis two projects have been explained in which the energy landscape of both folding and misfolding trajectories of a protein human neuronal calcium sensor-1 were reconstructed using single molecule mechanical manipulation.

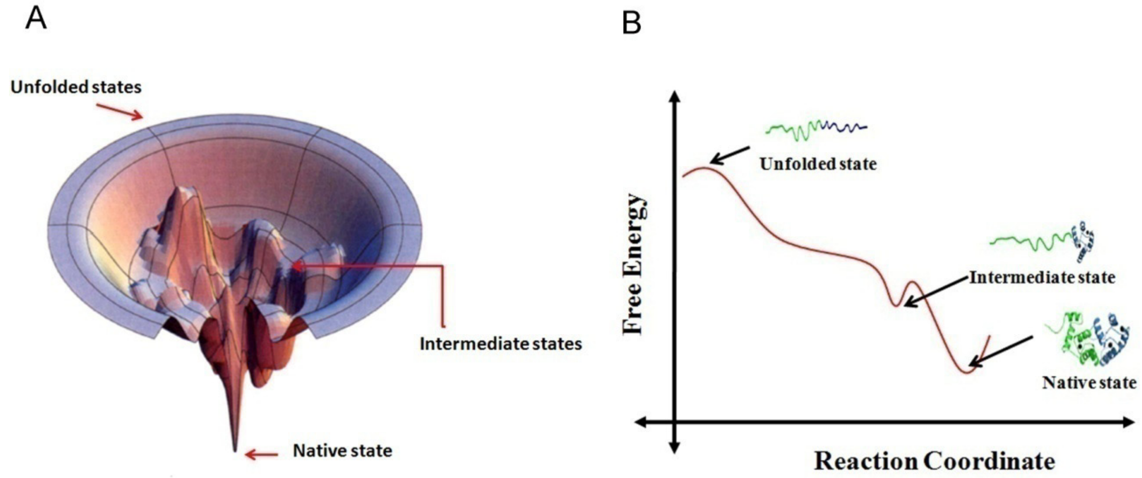


Figure 1: Energy landscapes governing protein folding reaction. (A) Multidimensional energy landscape involving various pathways towards the native state of the protein from its fully unfolded state. The local minimas in the trajectories are the intermediate states which slow down the overall folding process of the protein [3]. (B) Energy landscape represented on a one dimensional reaction coordinate. The choice of the reaction coordinate is made according to the experimental conditions. The strategy of one dimensional energy landscape representation is generally used in extracting parameters like activation energy barriers, kinetic distances and free energy change of unfolding of the protein.

1.1.3 Different models of protein folding:

Apart from the energy landscape theory, various models have been proposed in understanding protein folding problem such as: a frame work mechanism, hydrophobic collapse and nucleation condensation model [10, 11]. In the framework model the protein molecule is considered as made of numerous microdomains that search for their respective native state configuration. The microdomains in their native states then diffuse and collide with each other under the influence of external forces that leads to the formation of micro domain pairs which then collapse into the final native state of the protein. In the hydrophobic model, the initial collapse of side chains of the molecule takes place due to hydrophobic interactions resulting in a molten globule state which then transform into the tertiary structure before the final native state configuration. The nucleation condensation model explains the formation of a partially stable nucleus which acts as a seed for the formation of other structures around it.

1.1.4 Forces that govern the folding process of a protein:

The folding process of a protein is initially governed by the hydrophobic collapse of the molecule in which the hydrophobic residues get buried resulting in a molten globule like structure. This initial collapse is more entropically driven where the water molecules surrounding the protein play crucial role in its structuring [12]. The exact mechanism of the initial hydrophobic collapse of the molecule is still elusive however studies have shown that hydrophobic effect is the main driving force in the protein folding reaction

[13]. Once the molecule attains this initial molten globule like structure, hydrogen bonding within its secondary structures and between other regions comes into play. Along with the hydrogen bond formation, van der Waals interactions between atoms, electrostatic interactions between charged residues, backbone structuring and orientation of bond angles to overcome the loss in entropy due to native state configuration also takes place. The forces mentioned above are inversely proportional to the distances between the atoms and thus their role becomes significant after the initial hydrophobic collapse of the molecule. The details of how these forces act during the folding process are complex in nature and considering that the environment of the protein plays a crucial role during its folding [12], many forces are thermodynamically driven and are still elusive in the explanation of protein folding [14].

1.1.5 Protein misfolding:

The funnel view of the multidimensional energy landscape of proteins Fig 1 (A), shows numerous pathways towards the native state configuration. As explained before in [section (1.1.2)], these pathways include local minimas also known as kinetic traps which can be both on pathway to the native state or off pathway misfolded state. The off pathway states occur due to wrongly folded molecules having non native contacts. These non native contacts happen when the off pathway trajectories become kinetically more favorable than their on pathway counterparts. The misfolding of proteins causes several diseases such as Alzhiemers's, Parkinson's and Cretzfeldt-Jakob's and thus has gained a lot of attention over the years [15]. It has been shown that numerous proteins enter misfolded states among which many are not related to any disease [15-18]. Recently lot of studies has been done in understanding the characteristics of the misfolded states; the structural studies of misfolded PDZ domain revealed that it consisted of a misfolded packed hairpin on a native like scaffold, suggesting that misfolded states can have compact shapes [19] . Moreover it is proposed that multi domain proteins are more prone to misfolding due to large number of interacting peptide sequences during their folding process [20]. However, the initial stages of the misfolding process of proteins is still elusive and factors responsible for the aggregation and amyloidogenesis of several proteins, that results into amyloid fibrils formation is still a matter of debate [21, 22].

1.1.6 Amyloidogenesis and Amyloid Fibrils:

Amyloidogenesis is a disease in which proteins misfold and aggregate into long insoluble structures known as amyloid fibrils that get deposited in organs and cause diseases like Parkinson's, Alzheimer's and type 2 diabetes [21, 23]. Amyloidogenesis is caused by several proteins and peptides such as β2-microglobulin (β2-m), lysozyme, fibrinogen, transthyretin (TTR), islet amyloid polypeptide (IAPP), serum amyloid A (SAA) etc [21]. A lot of structural studies have been done on the amyloid fibrils using transmission electron microscopy, X -ray diffraction and atomic force microscopy. These studies have revealed that the fibrils usually contain numerous protofilaments [24] that jointly forms structures like long ribbons or rope like fibrils [24-26]. Moreover it is found that in each protofilament the protein molecules arrange to form β strands structures, perpendicular to the axis of fibrils [26]. Apart from structural studies significant progress has been made in understanding the detailed mechanism of the formation of amyloid fibrils. To elucidate the aggregation process, numerous models have been proposed such as nucleation growth mechanism, partial unfolding resulting in aggregation of globular proteins, native like oligomers formation, interaction of proteins with hydrophobic or hydrophilic surfaces etc [21, 22, 27-33]. However the exact mechanism of initial growth of fibrils is still elusive. In this thesis the amyloidogenesis of a protein β- 2 Microglobulin has been described in the paper II [chapter 2].

1.2 Single molecule mechanical manipulation using optical tweezers.

1.2.1 Why single molecule study: Its benefits over bulk studies:

To understand the folding dynamics of a protein, numerous techniques like circular dichroism, nuclear magnetic resonance (NMR), fluorescence spectroscopy, Small angle X-ray and neutron scattering etc. have been employed to give information about the structure and functionality of a protein both in its free and bound forms [34-36]. These bulk studies have been very informative but often limited to the description of the overall properties of a large population of proteins. Due to the ensemble averaging over large number of molecules, the measured properties gives no temporal information of the processes and they lack details regarding the forces involved in the reactions. Moreover thermal fluctuations present in the systems are cancelled out in the bulk measurements. A big assumption in the bulk studies is that the studied processes are homogeneous in nature, which is very ideal for many systems. However, with the advent of single molecule techniques the above mentioned limitations have been overcome and these advanced techniques have revealed unprecedented details of bimolecular processes.

Using single molecule techniques, the real-time folding (u) trajectories of individual molecules can now be observed in which the details regarding the intermediate states involved in the reactions are available, that are generally masked in the ensemble averaging of traditional bulk studies [17, 37]. Research in single-molecule protein folding is mainly due to the recent technological advancements in the instrumentation, particularly in single-molecule fluorescence (FRET) and mechanical manipulation methods like atomic force microscopy (AFM), and optical and magnetic tweezers. The applications of single-molecule techniques range from protein and nucleic acid folding dynamics to structural studies of protein molecules and biomolecular machines [38]. The advancement of single molecule techniques have reached a level where temporal information of protein folding processes can now be measured using ultra fast lasers. Moreover interaction of proteins with water molecules and the role of water molecules in the folding process can now be investigated [39, 40]. Due to the constraint of space, the full details about the recent progress in single molecule techniques cannot be provided here, however the reader is referred to some latest reviews to learn about the advancement in instrumentation and analysis methods in this field [41-43]. Here our focus will be on single molecule optical tweezers studies, where I will initially discuss the optical tweezers methodology, the experimental procedure of single molecule mechanical manipulation, recent advancement of instrumentation in this field and the theoretical methods generally applied in the analysis of single molecule manipulation data.

1.2.2 Biological significance of protein folding studies using mechanical manipulation techniques:

Protein folding studies using mechanical manipulation techniques like AFM and optical tweezers not only give insight into the folding trajectories and energy landscapes of molecules but they also provide information about the mechanical stability of proteins. The mechanical manipulation of proteins is allowing us to answer many key questions relevant to structure and functionality of individual proteins. Some of these questions are : Is it true that proteins that carry out mechanical functions like tenascin and titin should be more resistant to external forces than non mechanical proteins like enzymes ? How mechanical stability dictates the structure of a protein? How the functionality of a protein is related to its mechanical stability? The mechanical stability of proteins play key role in many biological processes such as in the translocation of mitochondrial proteins, that involves the electrophoresis of the targeting sequence through the membrane channels and in subsequent pulling of the target sequence by the proteins, the mechanical stability of the proteins dictates their translocation [44]. The translocation of misfolded proteins in the ATP dependent proteases 26S proteosome and ClpXP show similar unfolding process and dependence of mechanical stability of the degraded protein on the ATP consumption rate has been shown [45, 46]. The role of mechanical manipulation is significant in the protein folding mediated by GroEl chaperonin, that undergo conformational changes during encapsulation process resulting in the unfolding of tethered misfolded proteins which helps in their correct folding [47]. For proteins like fibronectin and tenascin, their mechanical stability is very significant for their functionality as these proteins extend and contract in response to mechanical forces during cell migration in vertebrate embryos

[48]. The mechanical properties of giant proteins like titin are of great significance for the tension developed by muscle cells during stretching [49]. The mechanical stability of many proteins also governs their structures. Proteins with high mechanical stability such as titin [50], FN-III domain of tenascin [48] and ubiquitin [51] are usually β-proteins, in which hydrogen bonding network protects the tertiary structure of the protein from unfolding. Whereas proteins that are prone to unfolding under tension are generally α-helical in which the only resistance to force is hydrophobic interactions between helices such as α-spectrin [52], barnase [53] and T4 lysozyme [54] .

1.2.3 Protein folding studies using optical tweezers:

Optical Tweezers:

From early 70s and mid 80s [55-57] , Arthur Ashkin and colleagues published some remarkable papers that lead to the invention of optical tweezers as a device which can be used to manipulate tiny objects with high dielectric constant. They found that these particles with high refractive index are attracted to high intensity regions of laser beams and thus can be held at a focal point. The basic principle of optical tweezer can be explained on the basis that the light interacting with a particle can be seen as both refracted and reflected by the particle. Due to the momentum P = hk (where h is Plank's constant and k is the wave vector) of light photons impinging on the particle, a reaction force is experienced by the particle which is equal and opposite to the change in light momentum. When the dimensions of the particle are much greater than the wavelength of light (narrow waist or focal point of beam), then the particle will act as positive lens which refracts photons in a direction towards its focal point and thus it will be trapped in the light beam. The above phenomenon can be explained by a simple ray diagram Fig. 2 (A):

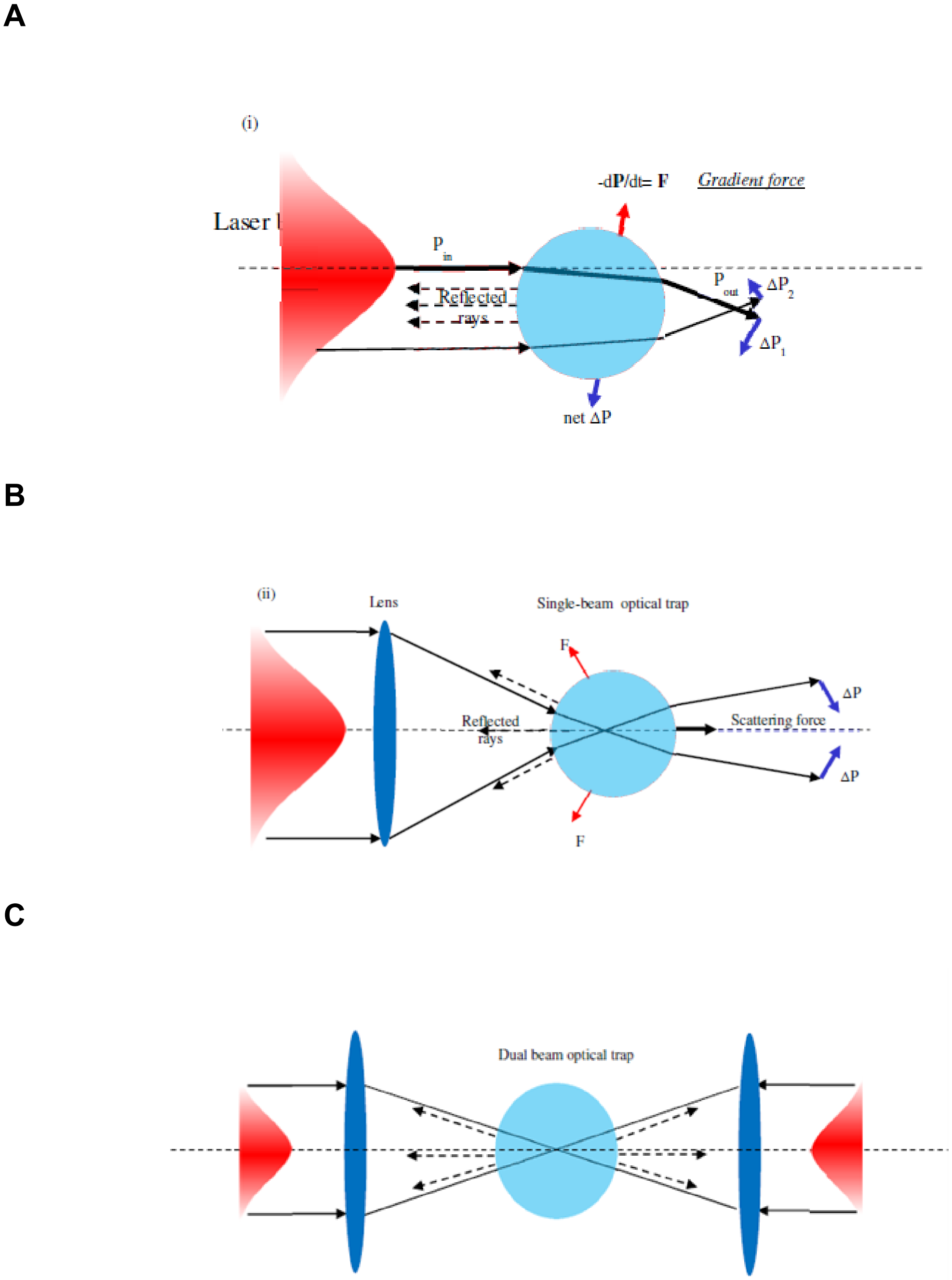


Figure 2: Light momentum effects on a refractive sphere (A). The particle is moved into the intense part of the light beam, but it is also pushed away from the light source. (B) The scattering force caused by reflected rays is balanced by the refraction of the high angle rays. (C) Dual-beam trap. The scattering forces caused by reflected rays are canceled, while the gradient forces are doubled [58].

In the above diagram for the particle with refractive index ‘n’ greater than that of surrounding medium ‘n_B’ the rays of light either get absorbed or reflected by the object. As the net momentum is conserved, the rate of change of momentum of the light caused by its interaction with the object must be equal and opposite to the rate of change in momentum of the object. For collimated beam of light, with a Gaussian intensity profile hitting the particle, for each ray in the beam the components of momentum flux parallel and perpendicular to the optical axis are given as [59].

$$\frac{dp}{dt}_{parallel} = n_B\left(\frac{W}{c}\right)\cos(\theta) \tag{1.2.1}$$

.

$$\frac{dp}{dt}_{transverse} = n_B\left(\frac{W}{c}\right)\sin(\theta) \tag{1.2.2}$$

Where W is the power of the light ray, θ is the angle which the rays make with the optical axis and c is the speed of light.

The reaction force (*F*) experienced by the particle is equal and opposite to the change in the light momentum flux summed over all rays passing through the interacting particle is given as:

$$F_{bead} = -\sum\left(\frac{dp}{dt}\right)_{in} - \left(\frac{dp}{dt}\right)_{out} \tag{1.2.3}$$

Where $\left(\frac{dp}{dt}\right)_{in}$ and $\left(\frac{dp}{dt}\right)_{out}$ are the rate of change of momentum flux for entering and leaving the trap respectively. The net change of these fluxes can be calculated using equation (1.2.1) and (1.2.2). Fig. 2 (A) shows that when the particle in the beam acts as converging lens then the rays of light in the central part of the light beam gets refracted downwards from the optical axis while rays interacting at the edge of the particle gets refracted towards the axis. Due to the above effect, a force due to change of light momentum acts downward on the particle which causes the particle to move towards the axis as a reactionary effect. Apart from this, the particle experiences a downward force due to scattered rays after refraction. Moreover, the impulsive force due to the reflected rays moves the particle forward in the light beam. The net effect of above phenomenon is that the particle moves towards the intense part of the light beam and it is moved forward due to reflected rays. A better geometry for the trapping of the object is illustrated in Fig 2 (B) in which a more focused light beam is used in which the impulsive force due to the reflected rays is balanced by the scattering force which acts along the optical axis. The dielectric object thus gets trapped slightly downwards of the optical axis where the large focal angle of the beam causes strong backward impulsive force due to scattered rays. For a single beam trap, thus an objective with high numerical aperture is required such that it can collect all the rays in the light beam resulting in efficient trapping of the object. However, the above geometry is not very robust when external forces are applied on the trapped object and photo detectors are used for measuring the applied force as a function of change in momentum of output rays from the trap. In this case, the applied force will deflect the output rays from the trap farther off axis then the input rays and objectives with even very high numerical aperture will not be able to collect all leaving rays resulting in poor measurement of applied forces. A more efficient way of countering the above limitations is the use of double beam optical tweezers set up as shown in Fig 2 (C). In this setup low beams with high numerical objectives are used such that the trapping rays form a narrow cone and the most marginal of output rays can be easily collected by the opposite objective. Even when the object is moved from its trapped position due to an external force, the deflected rays can be easily collected by the objectives situated near the photodetector. Due to the dual beam geometry, the scattering forces generated by reflected rays are canceled, while the transverse gradient forces are doubled resulting in a more focused position of the trapped object.

1.2.4 Force sensor theory:

In this section, how the external forces applied on a trapped particle can be measured from the change in light momentum of the impinging beam is described. In Fig (3), the geometry of the trapped object is shown. Light with a low beam enters from the left through an objective with high numerical aperture and is focused at a spot and the light that leaves the trap is collected by a similar objective. When an external force is applied on the particle, light exiting the trap has a modified angular distribution due to asymmetric refraction of interacting rays. A position sensitive detector (PSD) placed at right collects all the light momentum from the trap which is then used to measure the force applied on the particle as $F = \left(\frac{W}{c}\right)(\frac{x}{RL})$, where W is the light intensity of the collected rays at the PSD, c is speed of light, x is deflection of particle in the trap and RL is the focal length of the lens. This method of measuring the external force on the molecule from the change in light momentum is developed by Steve Smith [58].

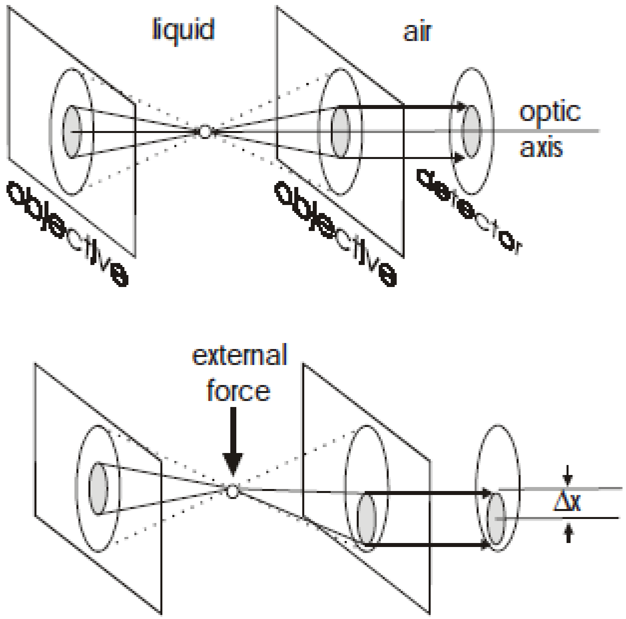


Figure 3: Light-momentum force sensor. The laser beam entering from the left passes through the bead, is imaged by the right objective, and exit. Application of an external force to the bead causes it to equilibrate slightly off center in the trap so that the change in momentum of the deflected beam exactly balances the external force. The angular deflection, θ, of a ray leaving the trap is transformed by the right objective into an offset distance, x, such that x/RL= $n_B$sin(θ), where n_B is the refractive index of the liquid and RL is the local length of the lens. The transverse light force F felt by a bead as it deflects a light ray of intensity W through an angle θ is given by F = (n_BW/c)sin(θ), where c is the speed of light. Therefore F= (W/c)(x/RL) [58].

The details of the force sensor theory can be understood from the following derivation:

When light interacts with an object immersed in a transparent liquid of refractive index n_1 the momentum flux of the light waves can be written as [59]:

$$d\left(\frac{dP}{dt}\right) = \left(\frac{n_1}{c}\right) S.dA \tag{1.2.4}$$

Where S is Poynting Vector, dA is element of area normal to S and c is speed of light. The force experienced by the object due to the light waves is the difference in momentum flux of light waves entering P_{in} and leaving the trap P_{out}. This force can be calculated by integrating the light intensity S_{in} entering and exiting S_{out} a spherical surface with radius R>>λ centered at the focal point. As the incoming waves are normal and the outgoing waves emanates from a point source the surface element dA is normal to S_{in} and S_{out}. Thus the light force acting on the object can be written as:

$$F = \frac{dP_{in}}{dt} - \frac{dP_{out}}{dt} = \left(\frac{n_1}{c}\right) \iint (S_{in} - S_{out})\, dA \qquad (1.2.5)$$

To calculate the angular intensity distribution for light entering or leaving the focus we can write:

$$I(\theta, \varphi) r\, d\gamma = SdA \qquad (1.2.6)$$

where θ *and* φ are angles shown in Fig. 4, r is a unit vector from the focus, and $d\gamma$ = dA/R ^{2} = $d\theta$ sin θ $d\varphi$ is an element of the solid angle.

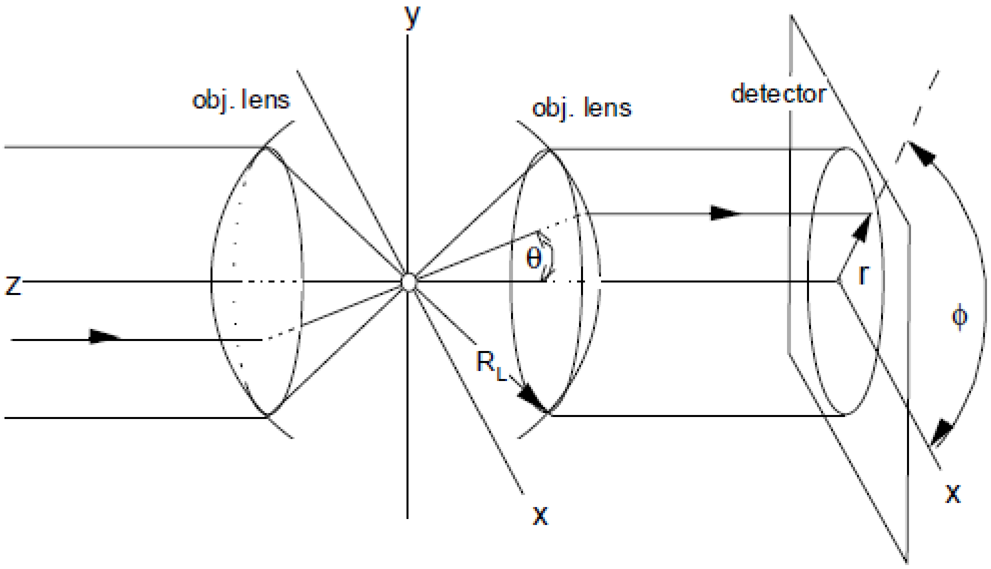


Figure 4: Light ray diagram showing the geometry of the incident beams through the objective and the exiting beams after interacting with the trapped object. Using the Abbe sine condition the intensity profile entering the detectors can be calculated from the above geometry [58].

Using the angular distribution of light intensity, the force experienced by the object due to light can be written as:

$$F = \frac{n_1}{4\pi c} \iint I(\theta, \varphi)(i\, sin\theta cos\varphi + j\, sin\theta sin\varphi + k\, cos\theta) d\gamma \qquad (1.2.7)$$

Here $I(\theta,\varphi)$ is measured in watts/steradian, which is considered negative for rays entering the trap and positive for rays leaving it. The value for $I(\theta,\varphi)$ can be calculated using Abbe sine condition according to which any ray emanating from the principal focus of objective lens inclined at angle θ to optic axis but still hitting the lens with radius R_L will exit the image side of the lens at a radial distance r Fig. (4) given by:

$$r = R_L n_1 sin\theta \tag{1.2.8}$$

The rays leaving the trap in a small element of solid angle $\frac{d\gamma}{4\pi}$ can be projected onto an area element dA' (rd φdr) on a plane placed at the image side then by conservation of energy the net irradiance at dA' is given by:

$$\mathrm{E}(r,\varphi)dA' = I(\theta,\varphi)\,\frac{d\gamma}{4\pi} \tag{1.2.9}$$

Here irradiance is used instead of intensity because it is a measurable quantity. Assuming that the lenses intercept all the light exiting the trap then equation (1.2.7) can be written as

$$F = \frac{1}{c}\iint E(r,\varphi)(i\,\frac{r}{R_L}cos\varphi + j\,\frac{r}{R_L}sin\varphi + k\,\sqrt{n_1^2 - \left(\frac{r}{R_L}\right)^2}\,\Big)\,\mathrm{rd}\varphi\mathrm{dr} \tag{1.2.10}$$

Here the integration is done on the image side principal plane of the axis. The transverse components (i and j) in the above equation can be measured using position sensitive photo detectors at those principal planes. A dual axis detector gives two different signals Dx and Dy which is proportional to the sensitivity Ψ of the silicon detectors and to the sum of local irradiances E(x,y) weighted by their relative distances $\frac{x}{R_D}$ and $\frac{y}{R_D}$ from the detector center where R_D is the detectors half width.

$$\mathrm{Dx} = \Psi\iint E(x,y)\left(\frac{x}{R_D}\right)dA' = \Psi\iint E(r,\varphi)\left(\frac{rcos\varphi}{R_D}\right)dA' \tag{1.2.11}$$

$$\mathrm{Dy} = \Psi\iint E(x,y)\left(\frac{y}{R_D}\right)dA' = \Psi\iint E(r,\varphi)\left(\frac{r sin\varphi}{R_D}\right)dA' \qquad (1.2.12)$$

Where integrals are over the surface of the detector. Thus the transverse components of light force can be measured from above relations as:

$$\mathrm{F_x} = \frac{\mathrm{D_x}R_D}{c\Psi R_L} \text{ and } \mathrm{F_y} = \frac{\mathrm{D_y}R_D}{c\Psi R_L} \qquad (1.2.13)$$

The above equations show that the transverse force component does not depend on the size of the trapped object but it depends only on the instrumental parameters. The force experienced by the object along the Z axis can be calculated from the following methods:

Z - Force measurements:

The Z force component from equation (10) can be written as:

$$F_z = \frac{1}{c}\iint E(r,\varphi)\sqrt{n_1^2 - \left(\frac{r}{R_L})^2\right)}\,\mathrm{rd}\varphi\mathrm{dr} \qquad (1.2.14)$$

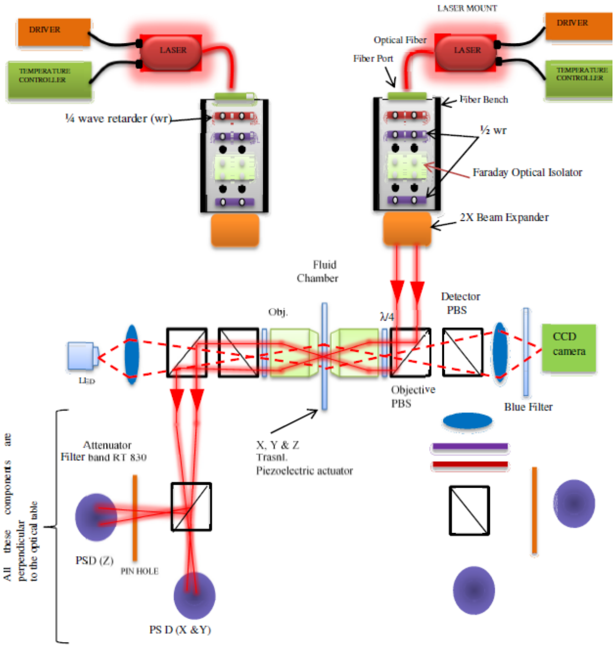
For the above component of force the light intensity should be modulated such that the signal from the photo detectors can be converted into force. To carry out the above task a transmission filter is used whose transmission pattern T can be expressed as:

$$T = \sqrt{1 - \left(\frac{r}{n_1 R_L}\right)^2\}} \qquad (1.2.15)$$

Using the above equation the transmission filter can be used to extract the z force experienced by the object.

1.2.5 Dual beam optical tweezers set up at University of Modena and Reggio Emilia:

The schematic drawing of optical tweezer set up, which we used in our single molecule manipulation experiments at Department of Physics, Informatics and Mathematics, University of Modena and Reggio Emilia, is shown in Fig. 5 below. We used a custom built dual beam set up in which the optical trap is formed using two single beam counter propagating diode lasers (845 nm) with 150 mW light power. The intensity of the laser diodes are controlled by two drivers and two temperature controllers are used to monitor the heating of the lasers. Inorder to collimate the laser beams to a size of 1.6 mm optical fibers are connected to single lens fiber ports. The linear polarization of the beams is maintained by feeding the output of fiber ports to a quarter wave plate. The beam then enters, a half wave plate that changes the linear polarization of the light such that it can enter a Faraday optical Isolator that prevents the entry of any reflected rays which can damage the laser diodes. The polarization of the beam is further rotated by a half wave plate such that it is perpendicular to the optical table. Further a beam expander is used to double the beam waist to a diameter of 3.2 nm. This beam then enters an objective with aperture size 8mm, which is large enough to collect all the rays leaving the optical trap despite of significant deflection of the beam after interacting with the trapped object. The expanded laser beam then enters a polarizing beam splitter objective (PBS) that directs it towards a quarter-wave plate (qwp) and a microscope objective lens (OBJ, Nikon Plan Apo VC, 60x, 1.2 NA),that focuses the beam to a spot where the bead can be trapped. The quarter wave plate restricts the entry of rays that are reflected from the bead in the trap towards the laser. Once the beam leave the trap after interacting with the trapped object it is collected by another objective and then it is again transformed into horizontally polarized light from where it is directed to position sensitive detectors (PSD) via polarizing beam splitter (PBS). A relay lens is then used to reimage the principal plane onto the photo-detectors because the PSDs cannot be placed at the output principal plane of the objective, which lies inside the objective. Since the detectors measure only the light exiting the optical trap, so for each detector only half of the integration required in Eq. (1.2.11, 1.2.12) is performed. This problem is solved by aligning the detectors on the optical axis so that when no bead is present in the trap, the output beams are centered on the detectors and the difference signals vanish, corresponding to zero volts at outputs Dx and Dy. Since the introduction of a particle has no effect on the beam entering the trap thus only the rays leaving the object in trap should be considered. For measuring all the force components, two PSDs are placed at each side of the dual beam optical tweezers. The light from the trap enters these detectors after being equally divided by a 50:50 non polarizing beam splitter. A transmission filter is placed before the z-direction photo-detector in order to make the signal voltage output proportional to the z-force component. Using this geometry, single molecules are manipulated by placing them between two polystyrene beads, one held in optical trap (3 μm) and another placed at the end of a micropipette by suction (2.1 μm) (Fig 5 (b). The changes in the extension of the molecule were measured by the movement of the two beads. The position of the bead on the micropipette is monitored by a “light lever” system that measures the position of the fluid chamber with which the micropipette is attached (Fig 5 (b)). Light emitting diode (LED) (465 nm) and a charge-coupled-device (CCD) camera, is used to visualize the bead in the trap and the fluid chamber with micropipette. The force sensor of this set up is calibrated for Stokes drag force and thermal fluctuations using the method explained in [58] and the force resolution of our set up is 0.1 pN.



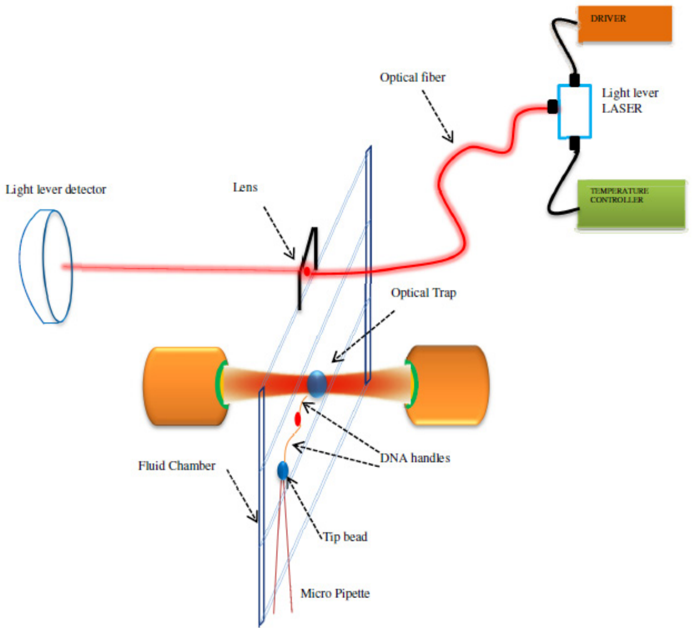


Figure 5 Block diagram of dual beam optical tweezers at University of Modena and Reggio Emilia (A). Two diode laser beams with vertical linear polarization pass through polarizing beam splitters (pbs), quarter-wave plates (qwp) and microscope objective lenses (OBJ). A bead is trapped at their common foci in circular polarized light. The exiting beams are collected by the opposite objective, converted to horizontal-polarized light and directed to position-sensitive photodetectors. (B) The flow chamber comprises two coverslips spaced 200 um apart by Parafilm layers and sealed by heat. The pipette is drawn from 100-um glass tubing down to a point with an opening of ~0.5 um. The light lever detection system monitors the position of the fluid chamber.

1.2.6 General procedure of protein folding studies using optical tweezers:

Sample Preparation:

As the size of biomolecules such as proteins is too small (nm) compared to the micrometer sized polystyrene beads, a big issue in the mechanical manipulation of these molecules is to prevent the interaction between the tethering surfaces once they come close. To solve this problem, a method has been developed in which two DNA molecular handles are used for attaching the proteins with the two beads [60]. One end of each DNA handle (500 bp) is covalently attached to the side chain of a cysteine residue and the other end is attached to the polystyrene beads as shown in Fig (6). The bead in the trap is attached to the end of one of the handle by means of digoxigenin/antibody interactions while the end of another handle is attached to the micropipette bead with biotin-streptavidin interactions. The DNA molecular constructs in this geometry act as spacers between the molecule and the beads thereby preventing unspecific interactions during the mechanical manipulation experiments. Full details regarding the attachment procedure of DNA handles can be attained here [61]. After the first successful application of this method in the mechanical manipulation of single protein molecules [37], this methodology has been extensively used in various single molecule optical tweezers experiments worldwide [17, 62-64].

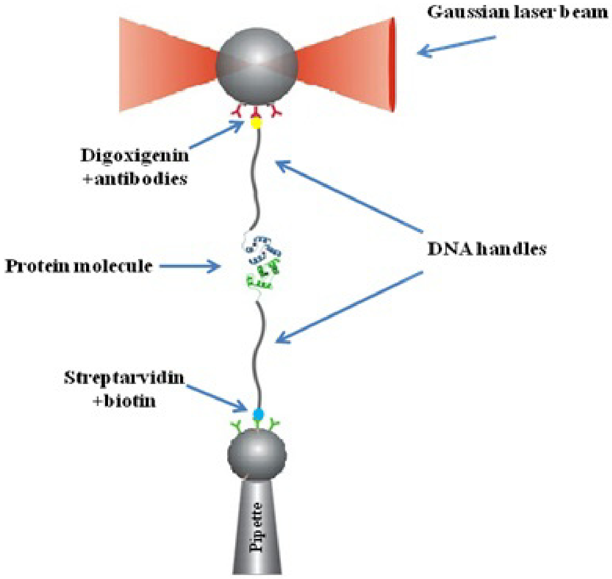


Figure (6): Experimental strategy of manipulating single protein molecules. Molecule is attached to functionalized polystyrene beads by means of 500bp DNA handles. One handle is attached to a steptarvidin bead placed at the end of a pipette and another handle is attached to digoxygen bead optical trapped. The molecule is pulled and relaxed by moving the pipette downwards and upwards respectively. Using this strategy applied force and change in extension of the molecule is recorded simultaneously [37].

1.2.7 Various methods of single molecule manipulation using optical tweezers:

The mechanical manipulation of proteins can be done using three main experimental strategies namely: Force ramp or constant velocity, force jump and constant force or hopping experiments. In this section I have explained these methodologies along with the references of the recent studies in which they have been employed.

Force Ramp or constant velocity:

In this method, single molecules are pulled and relaxed at constant velocity (nm/sec), which results in force extension curves as shown in Fig 7 (A). Here when the molecule is pulled, we initially observe a monotonic increase in force with extension, due to stretching of the DNA handles and then a sharp rip is observed corresponding to the complete unfolding of two state folding (u) molecule. On further pulling, the stretching of DNA handles along with the molecule can be observed. The sharp rip during unfolding of the protein happens such that the extension of the molecule increases with a drop in force due to upward movement of bead in the trap. Similar process is observed when the molecule is relaxed back to its native state configuration in which the force increase during the refolding transition. The values of the unfolding and refolding forces depend on the pulling speeds and distribution of these forces can be attained at different loading rates (pN/sec) from several pulling and relaxation cycles. These distributions of forces can then be fitted by a probability density function to extract the rates of folding and unfolding at zero force and distance to transition states [see section (1.3)].The above methodology has been applied in several recent studies [65, 66]. From the force extension curve, we can also measure the change in contour length between the native and unfolded state by fitting the branches before and after the transition with worm like chain model [see section (1.3)]. The change in contour length also gives the information about the number of amino acids involved in the unfolding and refolding of the protein. From these set of data we can also understand whether the protein folds / unfolds at equilibrium or away from it. If the folding and unfolding trajectories are reversible then the free energy change of unfolding can be directly measured from the area under the transition. This value can then be compared with the stability value measured in bulk studies after correction for the stretching free energy of the molecule. However if the folding/unfolding trajectories are not reversible, we observe hysteresis in the force extension curves as shown in Fig 7 (A). For measuring the free energy values from such data statistical tools like Crooks fluctuation theorem are used to extract equilibrium information [67, 68] [see section (1.3)].

Constant Force or Hopping experiments:

In this method, constant force is applied on the molecule by means of force feedback mechanism. When a molecule folds/unfolds, the force applied on the molecule increases/ decreases with the change in extension of the molecule. The feedback mechanism moves the position of the pipette in response to the above changes and thus maintains the force at a constant value. The change in extension of the molecule between its different states is measured in this process as a function of time. In Fig 7 (B), the hopping of a two state molecule at equilibrium is shown at a constant force along with the probability distribution of extensions corresponding to the native and unfolded states. From such traces, the life times of each state can be extracted (τ_U and τ_N) and the cumulative distribution of the lifetime values display single exponential behavior, suggesting first order kinetics of the reaction; see Fig. 7 (C) [69]. The fitting of these distributions provide the rates of folding and unfolding of the molecule at each force value. These rate constant values can then be linearly fit using the Bell's model to extract the rates at zero force and the distances to the transition states [see section (1.3)]. However for fast fluctuations of the molecule and for multi state folding process the above method of rate constant extraction is not very robust. For such processes, hidden Markov model (HMM) analysis is used that can mimic the ultra fast transitions of the molecule and give the full details of kinetics of the equilibrium fluctuations [70]. The full details of the HMM method of analysis is explained in [section (1.3)]. The method of feedback mechanism has several limitations mainly due to the relaxation time of the piezo stage that moves the pipette for keeping the force constant. Moreover, when the pipette is moved the feedback requires relaxation time for its components to become stable before it can respond to another transition of the molecule. Thus the response time of the feedback limits the timescales at which the force is kept constant [71]. Moreover, the spatial and temporal resolution of the instrument is very significant for extracting the kinetics of short lived states of the molecule.

Constant distance experiments:

In this method, the position of the pipette is kept constant and the force values due to movement of bead in the trap in response to folding/unfolding of the molecule is acquired. Since the distance between the trap and the pipette is kept constant, the force values indirectly describe the different states of the molecule instead of extension values in constant force experiments. In this method the temporal resolution is higher than in constant force traces which depend on the response time of the bead in the trap rather than the force feedback mechanism as used in constant force experiments.

Force Jump experiments:

When the rate of folding and unfolding of a molecule is slow, it becomes difficult to acquire lot of transitions in a short amount of time. For such cases, collecting a statistically good number of folding and unfolding events requires large amount of time which is experimentally very demanding considering the stability of the set up. This problem is circumvented by using force jump experiments. In this method, for observing the unfolding and folding events, the force is quickly increased (jumped) and decreased (dropped) respectively, to a preset force value and is kept constant with the feedback mechanism. The force range in these experiments is chosen such that the probability of folding and unfolding is high in that range. Fig. 7 (D), depicts a typical force jump trace with extension time and force time cycles. Initially, the force is at 6pN and the molecule is in its folded state from which it is pulled suddenly by increasing the force to 8.7 pN. The force is kept at this value and the corresponding stretching of the DNA handles is observed until a sudden transition occurs due to the unfolding of the molecule. Once the molecule is unfolded, the force is suddenly dropped to its earlier preset value of 6 pN. The relaxation of the DNA handles then takes place at this new force which then terminates with a sharp decrease in extension due to refolding of the molecule. This process is repeated for number of cycles until a good number of transitions are acquired. The dwell times of the unfolded (τ_U) and native states (τ_N) are then acquired from these cycles and the single exponential fitting of the cumulative distribution of the lifetime values then provide the rates of folding and unfolding transitions [69].

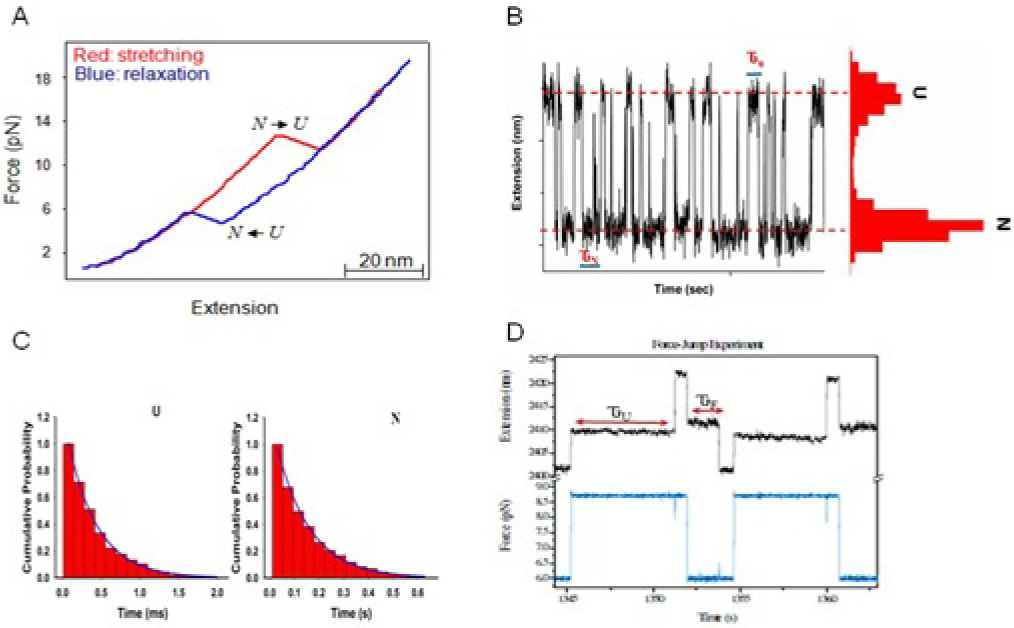


Figure 7: Single molecule mechanical manipulation data (A) Force extension curve of a two state folding (un) single molecule obtained from force ramp experiments. The unfolding of the protein is at high force and refolding occurs at lower forces showing non equilibrium trajectory that depends on the pulling speed[65]. (B) Extension time traces recorded using constant force methodology where the molecule can be seen as hopping between its two states. (C) Dwell time distributions of each state of the molecule depicting single exponential behavior from which the rates of folding an unfolding can be calculated. (D) Force jump trace depicting the extension time and force time data. The molecule is forced to jump between its different states in a range of forces such that the force is increased or decreased quickly to a preset values as the molecule folds or unfolds [69].

1.2.8 Recent advancement in optical tweezers instrumentation:

The recent progress in single molecule optical tweezers instrumentation has been focused in acquiring more details about the biological processes by improving the temporal and spatial resolution of the set up. Particularly, for single molecule protein folding studies, stress has been laid in developing ultrafast optical tweezers set ups that can provide more details about the complex energy landscape of the molecules [17]. A major change has been provided by the development of dual trap optical tweezers set ups, in which the thermal drift and fluctuations of the trapped bead is cancelled out due to the dual traps

[72]. For single molecule manipulation experiments, novel optical tweezers set ups have been developed which can be used to measure new parameters such as angular deflections due to applied torque or rotation of molecules [73]. Torques can now be applied to the molecules using magnetic tweezers set up which can now reveal details about many biological processes like transcription, replication and protein folding [74-76]. The magnetic tweezers have been integrated with dual beam optical tweezers that can be used in the study of wide range of complex systems such as DNA-DNA interactions, DNA bound proteins, intermolecular friction etc [77]. Moreover, the recent combination of optical and magnetic tweezers with fluorescence microscopy have recently provided unprecedented details regarding DNA supercoiling, dynamics of diffusion, hopping of plectonemics in DNA [78, 79]. The development of Quad – trap optical tweezers instrument has opened new avenues in which four polystyrene beads could be simultaneously trapped independently, with which mechanical manipulation of two molecules can be done at the same time [80].

1.3 Studying thermodynamics and kinetics of protein folding using single molecule force spectroscopy.

In biochemical reactions the interacting molecules undergo large conformational changes which involve both linear and rotational motions. These conformational changes help in understanding the progress of the reactions where parameters like end to end distance of the molecule acts as a convenient reaction coordinate. In this section the effect of force on thermodynamics and kinetics of single molecule protein folding reaction is explained considering the end to end distance of the molecule as the well defined reaction coordinate.

1.3.1 Effect of force on the energy landscape of a molecule:

The change in energy dE of a molecule when it is stretched using optical tweezers or AFM can be explained using the first law of thermodynamics as:

$$dE = \partial q_{rev} + \partial W_{rev} \tag{1.3.1}$$

$\partial q_{rev} = TdS$ and $\partial W_{rev} = -PdV + \int F\,dx\,)$ are the reversible heat exchange with surrounding and reversible work done on or by the molecule in quasi static equilibrium conditions. Where T is temperature, S is entropy, P is pressure, V is volume, F is force, and dx is change in extension of the molecule. Thus equation (1.3.1) can be expressed as:

$$dE = (T(dS)) + (-PdV + \int F\,dx\,) \tag{1.3.2}$$

As T and P are the independent variables in mechanical manipulation experiments the change in Gibbs free energy gives more relevant information:

$$dG = -SdT + VdP + \int F\,dx \tag{1.3.3}$$

At constant T and P , if the molecule is pulled a distance Δx in quasi static equilibrium conditions then the work done on the molecule will be reversible and will be equal to free energy change of the system

$$dG_{stretc\,h} = \int_{x_0}^{x_0+\Delta x} F\,dx \tag{1.3.4}$$

The energy landscape of a two state protein folding reaction where the molecule goes to its native state N from its fully unfolded state U at zero force is shown in Fig 8 (A). The two states of the protein occupy local free energy minima, separated by a distance Δx along the reaction coordinate that is end to end distance of the molecule. X_f and X_U are the distance to the transition state from N and U state respectively while $\Delta G_u^{\neq}$ and $\Delta G_f^{\neq}$ are the activation energy barriers for the two states of the molecule. The thermodynamic free energy change of the molecule is shown as ΔG^0. The free energy change of this two state reaction is given by:

$$\varDelta G(F = 0) = \varDelta G^0 + k_B T ln\left(\frac{P\{U\}}{P\{N\}}\right) \tag{1.3.5}$$

In the above equation $P\{N\}$ and $P\{U\}$ are the probability of occupying the native and unfolded states respectively in single molecule experiments. When force is applied to the molecule, to a first approximation, it can be assumed that each point on the energy landscape is tilted by an amount proportional to its distance from the native state [81] Fig 8 (B).

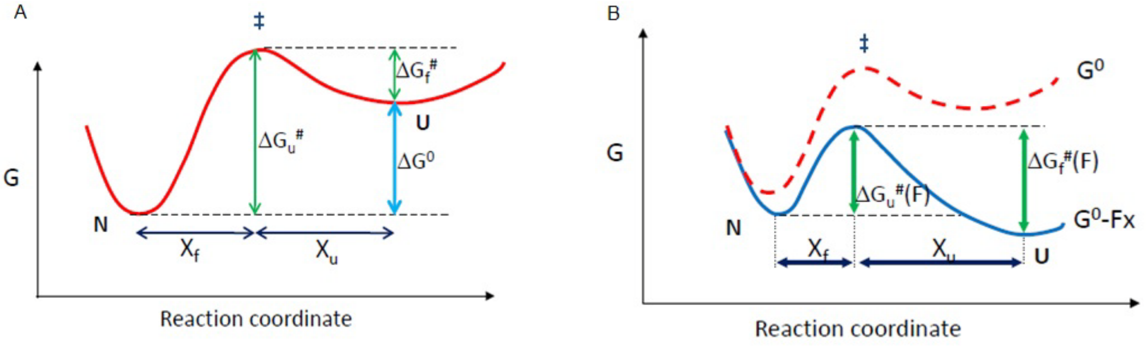


Figure 8: (A) Energy landscape of a two state folding unfolding molecule at Force = 0. The reaction coordinate is the end to end distance of the molecule. (B) The energy landscape of the molecule is tilted when force is applied, such that each point on the landscape moves downwards proportional to its distance from the native state which acts as a pivot [82]. Here it is assumed that the force does not affect the positions of the native and unfolded states.

Thus equation (1.3.5) can be written as:

$$\varDelta G(F) = \varDelta G^0 - F(\Delta x) + k_B T ln\left(\frac{P\{U\}}{P\{N\}}\right) \tag{1.3.6}$$

Since at equilibrium $\varDelta G(F) = 0$ thus

$$\varDelta G^0 = F(\Delta x) - k_B T ln\big(K_{eq}(F)\big) \tag{1.3.7}$$

Where

$$K_{eq}(F) = \exp - \frac{(\varDelta G^0 - F(\Delta x))}{k_B T} \tag{1.3.8}$$

is the equilibrium constant of the folding reaction which depends exponentially on the applied force F. Thus by varying force the population of the two states of the molecule can be changed by shifting the equilibrium to native state by decreasing force and towards the unfolded state by increasing the tension as shown in Fig 9 (A). As we pull the molecule we can find a force = F ½ at which the molecule can be observed to have equal probability of occupying the two states. At this force the equilibrium constant K_{eq}(F ½) $= 1$.

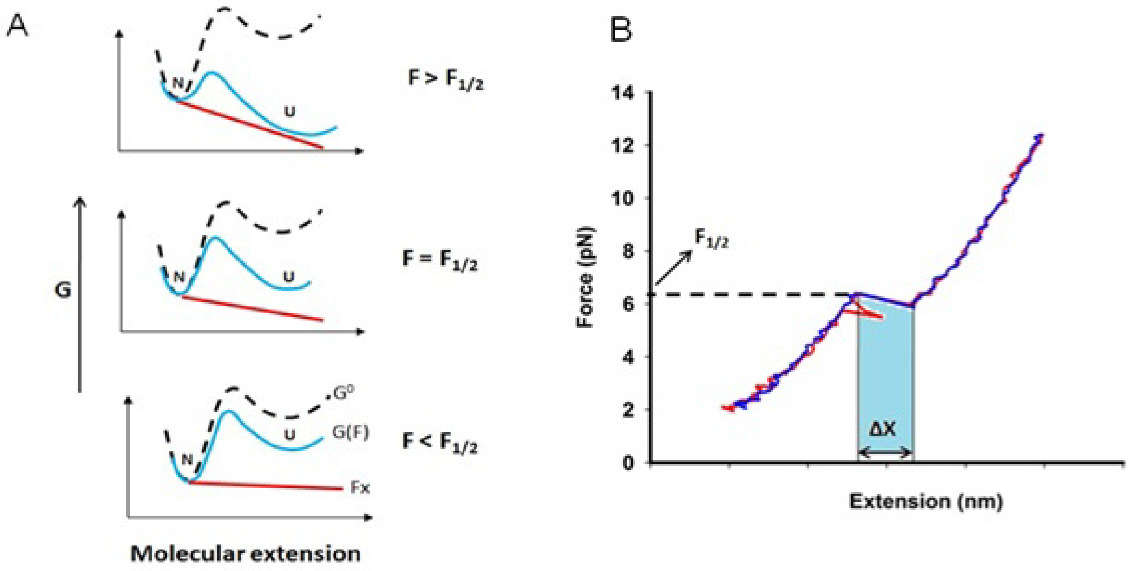


Figure 9: (A) Effect of force on the energy landscape of a two state folding unfolding of a molecule. At a particular force F= F ½ the molecule has equal probability of occupying the native and unfolded states. The probability of occupation shifts towards the unfolded state if F > F ½ and towards the native state for F < F ½ . (B) The force extension curve of a two state folding (un) of a molecule at equilibrium. Here the unfolding and refolding trajectories overlap and the molecule folds - unfolds reversibly about an equilibrium force. From the force extension curve the area under the transition gives the free energy change of unfolding the molecule [83].

In the above description, an assumption is made that the positions of the N and U states are unaffected by the applied force and the energy landscape is tilted with position of native state acting as a pivot. However this assumption is not purely valid as we know that the position of the U state does change, because at force F the molecule becomes a stretched polypeptide chain from a random coil configuration at force = 0. The shift in the position of the states depends on the local curvature of the states where a steeper potential well will have lesser change in position. Thus a correction in free energy change of the molecule should be done in equation (1.3.7) with respect to the new position of U state where the shift in the N state can be considered as negligible due to more curvature of its potential well.

$$\Delta G^0 = -k_B T \ln K_{eq}(F) + F\Delta x - \Delta G_{stretc\,h}(F) \tag{1.3.9}$$

Where

$$\Delta G_{stretc\,h}(F) = \int_{X_U(F=0)}^{X_U(F=F)} F.dX_U \tag{1.3.10}$$

At force = F ½ since K_{eq} (F ½) $= 1$ thus equation (9) can be rewritten as:

$$\Delta G^0 = F_{1/2}\Delta x - \Delta G_{stretc\,h}(F_{1/2}) \tag{1.3.11}$$

In the equilibrium conditions, $F_{1/2}\Delta x$ can be calculated as the area under the reversible folding unfolding transitions of the molecule in the force extension curve as shown in (Fig. 9 b) shaded portion. Whereas the $\Delta G_{stretc\,h}(F_{1/2})$ can be calculated from the WLC interpolation formula as explained next.

Entropic elasticity of the molecule:

Inorder to calculate $\Delta G_{stretc\,h}(F_{1/2})$ using equation (1.3.10) various polymer elasticity models can be used, however in single molecule manipulation experiments the most widely used model is Worm like chain model (WLC) [84, 85]. The WLC model is based on the entropic elasticity of polymers. Due to thermal fluctuations, a polymer in solution acquires a random coiled configuration with maximum entropy through steps of low entropy bends and curves. If the polymer is pulled from this maximum entropy configuration then a restoring force is generated known as entropic elasticity force. The WLC model assumes the polymer as a flexible rod of length L that can bend as a result of thermal fluctuations. The directional correlation between two segments in the polymer decreases exponentially with their separation *s* according to *e(-s/P),* where the decay length *P* is the persistence length of the polymer which is proportional to the stiffness of the polymer. Thus the force required to stretch a polymer of contour length L can be calculated from WLC interpolation formula as:

$$\frac{FP}{k_B T} = \frac{1}{4\left(1 - \frac{x}{L}\right)^2} + \frac{x}{L} - \frac{1}{4} \tag{1.3.12}$$

Where x is the end to end distance of the polymer. The $\Delta G_{stretc\,h}\left(F_{1/2}\right)$ can be calculated as the area under the WLC curve generated by integrating equation (1.3.12) from zero to extension of the unfolded molecule at $F_{1/2}$ Fig 2 (B). The ΔG^0 calculated after correcting for stretching free energy of the molecule can then be compared with values obtained in bulk experiments from the equilibrium unfolding of a protein [37, 83].

1.3.2 Effect of Force on the kinetics of single molecule reactions:

According to transition state theory, for a two state protein folding reaction the rate of folding (k_f^0) and unfolding (k_u^0) at zero force, depends on the activation energy barriers $\Delta G_u^{\neq}$ and $\Delta G_f^{\neq}$ for unfolding and folding respectively as :

$$k_u^0 = A\, exp\left(\frac{-\Delta G_u^{\neq}}{k_B T}\right) \tag{1.3.13}$$

$$k_f^0 = A\, exp\left(\frac{-\Delta G_f^{\neq}}{k_B T}\right) \tag{1.3.14}$$

The pre exponential factor A, depends on the rate at which the molecule diffuses to the transition state from state N and U [86] as:

$$A = \frac{\omega_s \omega^{\neq}}{2\pi\gamma/m} \tag{1.3.15}$$

Where ω_s is the characteristic potential of the harmonic wells at states N and U, which determines the rate at which the molecule approaches the transition state. $\omega^{\neq}$ defines the rate at which the barrier is crossed once it is reached from N and U states. While the ratio $\frac{\gamma}{m}$, of friction coefficient over the mass of the molecule determines the ruggedness of the energy landscape. As the applied force tilts the energy landscape of the molecule, the Bell's model [82] states that the activation energy barrier for unfolding $\Delta G_u^{\neq}$ decreases by a factor $F.X_u$ and the activation energy barrier for refolding is increased by a factor $F.X_f$ Fig 1b. Thus the unfolding and refolding rates at a given can be written as:

$$k_u(F) = A\, exp - \left(\frac{\Delta G_u - F\, \Delta x_u}{k_B T}\right) \tag{1.3.16}$$

$$k_f(F) = A\, exp - \left(\frac{\Delta G_f + F\, \Delta x_f}{k_B T}\right) \tag{1.3.17}$$

Using equations (1.3.14), (1.3.15) and (1.3.16) (1.3.17) we can write:

$$k_u(F) = k_u^0\, exp\left(\frac{F\, \Delta x_u}{k_B T}\right) \tag{1.3.18}$$

$$k_f(F) = k_f^0\, exp\left(\frac{-F\, \Delta x_f}{k_B T}\right) \tag{1.3.19}$$

The ratio of $k_u(F)$ and $k_f(F)$, the equilibrium constant of the reaction at a given force can be calculated as:

$$K_{eq}(F) = K_{eq}^0\, exp\left(\frac{F\, \Delta x}{k_B T}\right) \tag{1.3.20}$$

Where $\Delta x = \Delta x_u + \Delta x_f$ is the extension of the molecule and K_{eq}^0 is the equilibrium constant at zero force.

On equating equations (1.3.18) and (1.3.19) we can find a relation for the force $F_{1/2}$ at which the molecule will have equal probability of occupying the two states:

$$k_u^0\, exp\left(\frac{F_{1/2}\, \Delta x_u}{k_B T}\right) = k_f^0\, exp\left(\frac{-F_{1/2}\, \Delta x_f}{k_B T}\right)$$

$$F_{1/2} = -\frac{k_B T}{\Delta x} \ln \frac{k_u^0}{k_f^0} \tag{1.3.21}$$

Equations (1.3.18) and (1.3.19) can be linearized by taking the natural log as follows:

$$\ln\big(k_u(F)\big) = \ln(k_u^0) + \frac{F\,\Delta x_u}{k_B T} \tag{1.3.22}$$

$$\ln\left(k_f(F)\right) = \ln\left(k_f^0\right) - \frac{F\,\Delta x_f}{k_B T} \tag{1.3.23}$$

As we can see from above equations, $\ln\big(k_u(F)\big)$ increases linearly with the applied force while $\ln\left(k_f(F)\right)$ decreases linearly with increase in tension. The plots of $\ln\left(k_u(F) or k_f(F)\right)$ vs force is called chevron plot (Fig. 10) from which rate constants at zero force (k_u^0 *and* k_f^0) can be extrapolated and distance to transition states (Δx_u *and* Δx_f) are evaluated from the slope values.

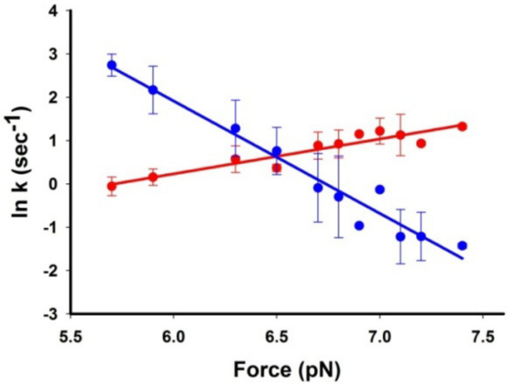


Figure 10: Plot of log of rate constants of folding and unfolding of the molecule calculated at each force. The experimental data is fitted with linearized form of Bell's model from which the distance to transition states and rates at zero force can be calculated.

Using, equations (1.3.13) and (1.3.14) the activation energy barriers $\Delta G_u^{\neq}$ and $\Delta G_f^{\neq}$ can then be calculated by approximating the preexponential factor A for the given system. In the above kinetic model the position of the transition state is considered as unchanged with the application of force which is a good approximation only for small range of forces in which the molecule is manipulated. However experiments in which large range of forces are used the transition barrier position will shift with the applied forces and Bell's model with a force independent transition barrier position cannot be used in such cases. To account for the shift in the transition state position the dependence of folding and unfolding rates on an additional energy barrier associated with changes of bead deflection in optical trap or spring constant of AFM cantilever and extension of DNA handles is considered as explained in these papers [87, 88]. Apart from these other authors have also proposed theoretical methods that account for shift in transition states with force [89-91].

1.3.3 Energy landscape reconstructions from equilibrium and non - equilibrium force manipulation data:

The experimental data from single molecule mechanical manipulation experiments can be divided into two categories – equilibrium and non equilibrium folding (un) of proteins. In this section I will talk about these two types of data and the theoretical models generally applied to analyze the thermodynamic and kinetic parameters of protein folding reactions from them.

1.3.4 Equilibrium fluctuations in single molecule manipulation experiments:

As discussed in [section (1.2.7)], using constant force technique single molecules can be observed hopping between there different molecular configurations at equilibrium. These fast fluctuations observed in the extension time traces, [see section (1.2.7)], contain wealth of information about the kinetic and thermodynamics of the folding unfolding reaction. On plotting the probability distributions of these traces, by Gaussian fitting, one can estimate the change in extensions of the molecule corresponding to transitions between different states as shown in Fig. 11 below.

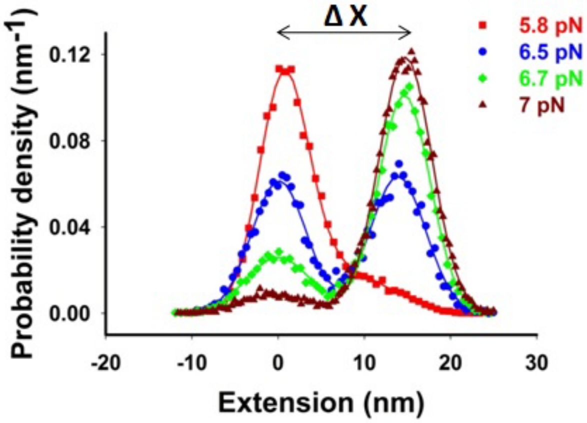


Figure 11: The probability density distributions of the extensions corresponding to the different states of the molecule. The Gaussian fit of the data provide the extensions corresponding to the transition between the states. The density distribution can also be used to acquire free energy change of unfolding using the deconvulation method [87, 92].

These probability distributions P(x) of extensions can be used to reconstruct the energy landscape of the molecule where the free energy function G(x) for the desired energy landscape is related to P(x)as [92]:

$$G(x) = -k_B T . ln[P(x)] \qquad (1.3.24)$$

This approach of energy landscape reconstruction has several requirements like high resolution optical trapping system and constant load over full temporal bandwidth of the folding process. For systems in which the time required to measure the position and adjust the force in response is slower as compared to the timescale of transitions the above methodology is not applicable. Moreover it is also essential that instrumental effects and noise in the experiments be considered, because the dynamics of the molecule are convolved with the dynamics of the force probe. The effect of the force probe can be removed by deconvolution methods to recover the intrinsic energy landscape of the proteins [87, 93]. The above method helps in extracting not only the on pathway trajectories of the protein but also the non native states which the molecule occupies due to kinetic traps in the energy landscape, which are also known as misfolded states of the protein [18]. However for systems in which multi state transitions take place at equilibrium, the above technique may not be feasible and may be applicable for only two state or single domain proteins. Proteins having two domains like CaM and NCS-1, are observed to show complex fluctuations between multiple states [17, 94] and merely characterizing a state on the basis of its extension signal is not feasible in such cases, as two states can give same extension values but different kinetics in the traces [17].

Hidden Markov Modeling (HMM) is another method for analyzing such ultra fast fluctuations which is based on probabilistic theory. In this section I will initially discuss the basics of HMM and then its application in the analysis of hopping traces of single molecule manipulation experiments.

1.3.5 Hidden Markov Model

Signals and Signal Models:

Real world processes produce signals of different types like discrete signals, continuous signals, stationary source signals, dynamic source signals, noisy signals etc. In order to find a theoretical description of the signal processing system and for understanding the properties of the signal source, signal models are used for simulating the source. HMM lies in the class of statistical models in which the basic assumption is that the signal is governed by a random or stochastic process whose parameters can be characterized in a well defined manner. HMM was first introduced by Baum and his colleagues in late 1960s and early 1970s and was applied for speech processing applications.

Discrete Markov Process:

HMM is based on the Markovian theory of signal processing which can be described as follows:

Let us consider a system with N possible states S1, S2,…SN. The transitions between the N states, which the system can undergo at discrete interval of time say, can be defined by an N x N transition matrix A. Where the probability that the system can switch from its present state at time $t : q_t = S_i$ into a new state $q_{t+1} = S_j$ at time $t + 1$is defined as:

$$A_{ij} = P\left(q_{t+1} = S_j \mid q_t = S_i\right) \qquad (1.3.25)$$

A_{ij} is the element of the transition matrix A where for each i,j $A_{ij} \geq 0$ and $\sum_{j=1}^{N} A_{ij} = 1$. Moreover eq (1.3.1) implies the assumption:

$$P\left(q_{t+1} = S_j \mid q_t = S_i, q_{t-1} = S_k, \ldots\ldots\right) = P\left(q_{t+1} = S_j \mid q_t = S_i\right) \qquad (1.3.26)$$

That the probability of transition to the next state depends only on the present state of the system thus it is history independent process. This property of the system is known as Markovian. Another assumption in this model is that matrix A is time independent. The above properties describe a first order discrete Markov process. Although a theory of continuous Markov process is possible but we will consider only the discrete time process.

Hidden States and Model:

The above Markov model is observable, in which the probability of transition between the states can be directly measured. However in processes where the states of the system are 'hidden' and we can only observe a sequence of events, we use HMM to understand the dynamics of these states or the underlying process which generates the observable events. For such systems we can assume that there is a correlation between the hidden states and the observable events where the former causes the latter.

Model:

In HMM, we assume that the observation sequence is a probabilistic function of the hidden states of the system where HMM is a doubly embedded stochastic process. For a discrete time process we can assume that at time t the system is in an unknown or hidden state q_t which generates an observation event O_t from a sequence of M possible events $O_1, O_2, \dots O_M$. The probability of occurrence of each observation event $O_t = O_1, O_2, \dots O_M$ corresponding to hidden states $q_t = S_1, S_2 \dots S_N$ are defined by a NxM observation matrix B whose elements are:

$$B_i(j) = P(O_t = V_j \mid q_t = S_i) \qquad (1.3.27)$$

The above equation shows that at time , the probability that the system in state S_i among its N states, generates an observable event V_j among the M possible events is given as $B_i(j)$ where $B_i(j) \geq 0$ and the sum of $B_i(j)$ over j must be equal to 1.

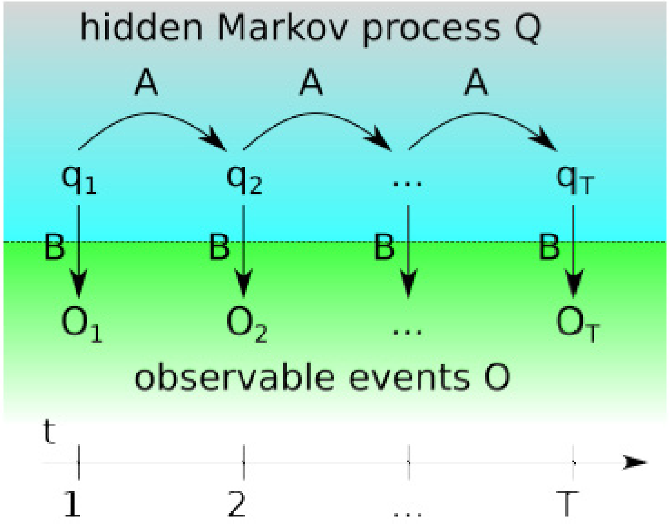


Figure 12: A general HMM 'Q', where the hidden state sequence q_t is related to the observation sequence O_t by observation probability matrix B and transition probability matrix A.

The roles of transition matrix A eq (1.3.25) and observation matrix B eq (1.3.27) are pictorially described in Fig. 12 Another information which is required for constructing the model is the initial conditions of the hidden system at time $t = 1$. These conditions are defined by the vector π where for all $1 \leq i \leq N$:

$$\pi_i = P\,(q_1 = S_i) \tag{1.3.28}$$

that gives the probability π_i that the hidden system is in state S_i at the beginning of the considered time interval.

Thus using equations (1.3.25), (1.3.27) and (1.3.28) we can construct the model as $\lambda = (A, B, \pi)$ which can precisely define the statistics related to both the visible and the hidden part of the system.

Three Basic Problems of HMM:

A big issue with the model discussed above is the method of determining the parameters in $\lambda = (A, B, \pi)$ that best describes the observation sequence $O_t = O_1, O_2, .. O_M$. For tackling this issue HMM deals with solving the three problems :

(1) For a given observation sequence $O_t = O_1, O_2, .. O_M$ if we assume a model $\lambda = (A, B, \pi)$ that can generate the observation sequence then what is the probability of the observation sequence given the model λ defined as $(O|\lambda)$?

(2) Given an observation sequence $O_t = O_1, O_2, .. O_M$ and a model $\lambda = (A, B, \pi)$ how we can choose a sequence of states $Q = Q_1, Q_2, .. Q_N$ that best describes the observations or is optimal in some sense?

(3) For the observation sequence $O_t = O_1, O_2, .. O_M$ how we can adjust the model parameters to maximize $P(O|\lambda)$?

Forward- Algorithm:

For solving problem (1), HMM applies this algorithm in which given the model parameter λ , the probability of occurrence of an observation event at each time step is computed which is much less computationally demanding than considering all possible hidden processes. The procedure makes use of an auxiliary forward variable $\alpha_t\ (i)$ which is defined as:

$$\alpha_t\ (i) = P\ (O_1, O_2\ \ldots, O_t, \quad q_t = S_i\ |\lambda) \tag{1.3.29}$$

Which gives the probability that at time, the observation sequence is observed And the hidden state is S_i given the model is λ. The next step is of **initialization** where the value of $\alpha_t\ (i)$ at $t = 1$ is given as

$$\alpha_1\ (i) = \pi_i\ B_i\ (O_1);\ \ 1 \le i \le N \tag{1.3.30}$$

Now the values for $\alpha_t\ (i)$ at each time step are calculated inductively as:

Induction step:

$$\alpha_{\ t+1}\ (j) = \left[\sum_{i=1}^{N} \alpha_t\ (i)\, A_{ij}\right] B_j\ (O_{t+1});\ \ 1 \le t \le T-1; 1 \le j \le N \tag{1.3.31}$$

In the above step all possible ways at time $t + 1$ are considered, in which the hidden states S_i from $= 1\ to\ N$, can make the transition to state S_j according to the matrix elements A_{ij} AND state S_j generates the observable O_{t+1} given by matrix element $B_j\ (O_{t+1})$.

Termination: After calculating α for all t and j values the sum of $\alpha_T\ (j)$ over all j will give the solution of problem one mentioned before:

$$P\ (O|\lambda)\ = \sum_{j=1}^{N} \alpha_T\ (j) \tag{1.3.32}$$

As $\alpha_T\ (j)$ gives the probability of observing the sequence $O_1 , O_2\ \dots , O_T$ and the hidden state at T as S_j the sum over it will include all possible hidden states which are equally likely possible to generate the sequence.

For solving problem (2), it is important to first understand what we mean by optimal state sequence. Here optimal means a state sequence $Q =\ q_1 , q_2 \ldots q_T$ which has the maximum probability of generating the observation sequence $O\ =\ O_1 , O_2 , \ldots O_T$. There are various methods for finding such optimal state sequence like one way is to look for all possible Q and choose the one with maximum probability of generating the observations $O = O_1 , O_2 , \ldots O_T$. However here we will only discuss the case where we can choose the individually most like states. This method maximizes the number of correct states corresponding to the target observation sequence O. HMM applies the method of backward algorithm to carry out the procedure of finding individually most like states.

Backward Algorithm:

here an auxiliary variable is initially defined as :

$$\beta_t\ (i) = P\ (\ O_{t+1} , O_{t+2}\ \dots , O_T \mid q_t\ =\ S_i\ , \lambda) \tag{1.3.33}$$

The above equation gives the probability of observing the events $O_{t+1}, O_{t+2} \ldots, O_T$ subsequent to time t given that the hidden state is S_i and the model is λ.

The initialization of this algorithm is done as $\beta_T\,(i) = 1;\;\; 1 \le i \le N$

Next the **induction** of the variable $\beta_t\,(i)$ is defined as:

$$\beta_t\,(i) = \sum_{j=1}^{N} A_{ij}\; B_j\,(O_{t+1})\,\beta_{t+1}\,(j)\,;\;\; t = T-1, T-2, \ldots 1;\; 1 \le i \le N \quad (1.3.34).$$

The above equation shows that for the system to have been in state i at time t and the observation sequence subsequent to time t $O_{t+1}, O_{t+2} \ldots, O_T$ is taken into account, then we have to consider all the possible states at time $t+1$ S_j given by the transition element A_{ij} and the observation element O_{t+1} expressed by $B_j\,(O_{t+1})$ and the probability of those subsequent to $t+1$ $\beta_{t+1}\,(j)$. To find the most optimal state sequence we define the term:

$$\gamma_t(i) = P\,(\,q_t = S_i \mid O, \lambda) \qquad (1.3.35)$$

Which gives the probability of the hidden state as S_i given the whole observation sequence O and the model λ. This is the actual quantity we are interested in, and which we want to maximize. In other words, in our current choice of the "optimality criterion", for each time t we are looking for the state q_t the system is most likely to be found in, given our knowledge of the observed sequence O. This can be mathematically expressed as:

$$q_t = argmax[\,\gamma_t(i)]; \qquad 1 \le i \le N;\; 1 \le t \le T.$$

Equation (35) can be written in terms of α and β terms as:

$$\gamma_t(i) = \frac{\alpha_t\,(i)\beta_t\,(i)}{P\,(O|\lambda)} = \frac{\alpha_t\,(i)\beta_t\,(i)}{\sum_{j=1}^{N}\alpha_t\,(j)\beta_t\,(j)} \tag{1.3.36}$$

As $\alpha_t\,(i)$ accounts for the observations till time t while $\beta_t\,(i)$ accounts for the events subsequent to time $t+1$ the normalization factor $\sum_{j=1}^{N}\alpha_t\,(j)\beta_t\,(j)$makes $\gamma_t(i)$ a probability measure. Thus using (1.3.36) HMM chooses the individually most likely state q_t at time t.

Search for the best fitting Model:

the third problem of HMM is the most difficult one which requires the solution of the first two problems. In this problem, given an observation sequence $O = O_1, O_2, .. O_T$, we need to find the model $\lambda = (A, B, \pi)$ having the maximum probability for the sequence O to occur. Although there is no efficient way to solve this problem but we can find a model $\lambda = (A, B, \pi)$ such that $P\,(O|\lambda)$ is locally maximized. This procedure is carried out using iterative techniques like Baum-Welch method also known as expectation modification EM. In this method, in order to begin the iterative procedures we need to guess initial model parameter values A_{ij}, $B_i\,(j)$ and π_i . Also, we have to define the number of states of the system N and the number of observation events M. These initial choices for modeling must be made after having a prior knowledge of the system under study which is subsequently corrected by the expectation maximization procedures. To do that we first define a parameter

$$\varepsilon_t\,(i,j) = P\left(q_t = S_i,\, q_{t+1} = S_j \;\middle|\, O, \lambda\right) \tag{1.3.37}$$

Which gives the probability of the system to be in state S_i at time t and state S_j at time $t+1$, given the model and the observation sequence. Eq (1.3.13) is somewhat the two-times analogous of the $\gamma_t(i)$ variable defined in eq (1.3.35, 1.3.36). Thus by using the forward $\alpha_t\ (i)$ and backward variables $\beta_t\ (i)$ we can rewrite eq (1.3.37) as:

$$\varepsilon_t\ (i,j) = \frac{\alpha_t\ (i) A_{ij} B_j\ (O_{t+1}) \beta_{t+1}\ (j)}{P\ (O|\lambda)} \tag{1.3.38}$$

$$\varepsilon_t\ (i,j) = \frac{\alpha_t\ (i)\ A_{ij} B_j\ (O_{t+1}) \beta_{t+1}\ (j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t\ (i) A_{ij} B_j\ (O_{t+1}) \beta_{t+1}\ (j)} \tag{1.3.39}$$

The studied process involves all the correct observations up to time t together with a particular hidden state at t, then the transition to another particular hidden state at $t+1$ and finally all the correct observations from $t+1$on. From above equations and eq (1.3.36) we can see that:

$$\gamma_t(i) = \sum_{j=1}^{N} \varepsilon_t\ (i,j) \tag{1.3.40}$$

The sums of $\gamma_t(i)$ and $\varepsilon_t\ (i,j)$ over time, represent the expected number of transitions from a given hidden state i, and the expected number of transitions from a hidden state i to a hidden state j, respectively. Which can be expressed as:

$\sum_{t=1}^{T-1} \gamma_t(i)$ = expected number of transitions from S_i.

$\sum_{t=1}^{T-1} \varepsilon_t\ (i,j)$ = expected number of transitions from S_i to S_j

where the summation goes only upto $t = T-1$, since T is the maximum time index and transition from T to $T+1$ are not considered. From above equations we can write a set of reestimation formulas for A, B and π. Where

$\pi_i{}'$ = expected frequency (number of times) in state S_i at time $t = 1$= $\gamma_1(i)$ (1.3.41)

$$A_{ij}{}' = \frac{\text{expected frequency of transitions from state } \mathrm{S_i} \text{ to state } \mathrm{S_j}}{\text{expected frequency of transitions from state } \mathrm{S_i}} =$$

$$\frac{\sum_{t=1}^{T-1} \varepsilon_t\,(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{1.3.42}$$

$$B_j\,(k)' = \frac{\text{expected frequency in state } \mathrm{S_j} \text{ and observing the event } O_k}{\text{expected frequency in state } \mathrm{S_j}}$$

$$= \frac{\sum_{t=1}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(i)} \tag{1.3.43}$$

Using an initial guess model $\lambda = (A, B, \pi)$ with elements A_{ij}, $B_i\,(j)$ and π_i chosen according to some prior information about the system, we can carry out the reestimation procedure discussed above to find the optimal state sequence which best describes the observation events. However there are two possible solutions: (a) the initial λ is a critical point of the system defined by the re-estimation rules above, and thus does not change, implying $\lambda = \lambda'$ or (b) the new model λ' is more likely than λ, i.e. $P\,(O|\lambda') > P\,(O|\lambda)$. If b) occurs, we can assign $\lambda \leftarrow \lambda'$, i.e. the new model becomes the "old" one and we can reiterate this procedure from using the reestimation eq (1.3.41, 1.3.42 and 1.3.43). This goes on until we are satisfied by the level of likelihood $P\,(O|\lambda)$ achieved. The final model is called a maximum likelihood estimate of the HMM.

HMM Scaling:

The three HMM solutions discussed earlier all involve induction steps requiring products of probabilities. From eq (1.3.31) it is evident that, $\alpha_t\,(i)$ tends to 0 exponentially as T increases. The final result of such chains of operations is likely to be a very small number which can be smaller than the machine precision a problem known as underflow. The solution to this underflow problem is to scale the numbers. However, care must be taken to insure that the re-estimation formulae remain valid.

Using eq (1.3.31):

$$\alpha_t\,(j) = \left[\sum_{i=0}^{N-1} \alpha_{t-1}\,(i)\,A_{ij}\right] B_j\,(O_t); \qquad 0 \le t \le T-1; 0 \le j \le N-1$$

From above equation it looks possible to normalize $\alpha_t\,(j)$ by dividing it by its sum over j. To check whether the reestimation formulae remain valid lets assume that at $t = 0$ $\alpha_0\,(j)' = \alpha_0\,(j)$ for $j = 0,1,\dots N-1$. If we define a variable

$$c_0 = \frac{1}{\sum_{j=0}^{N-1} \alpha_0\,(j)'} \tag{1.3.44}$$

And $\alpha_0\,(j)'' = c_0\alpha_0\,(j)'$ for $j = 0,1,\dots N-1$. Then for each $t = 1, 2, \dots T-1$ we can do the following:

(1) For $j = 0,1,\dots N-1$ we can compute $\alpha_t\,(j)'' = \left[\sum_{i=0}^{N-1} \alpha_{t-1}(i)' A_{ij}\right] B_j\,(O_t)$

(2) Then if we define : $c_t = 1/\sum_{j=0}^{N-1} \alpha_t\,(j)'$

(3) For $= 0,1,\dots N-1$ $\alpha_t\,(j)'' = c_t\alpha_t\,(j)'$.

As $\alpha_0\,(j)'' = c_0\alpha_0\,(j)'$ we can write:

$$\alpha_t\,(j)'' = c_0c_1\dots c_t\alpha_t\,(j) \tag{1.3.45}$$

Then $\alpha_{t+1}\,(j)'' = c_{t+1}\alpha_{t+1}\,(j)'$

$$= c_{t+1}\left[\sum_{i=0}^{N-1} \alpha_t(i)'' A_{ij}\right] B_j\,(O_t)$$

$$= c_0c_1..c_tc_{t+1}\left[\sum_{i=0}^{N-1} \alpha_t(i)\,A_{ij}\right] B_j\,(O_t) = c_0c_1 \dots c_{t+1}\alpha_{t+1}(i)$$

Thus eq (45) holds due to induction for all t and by using the definitions of $\alpha_t\,(j)''$ and $\alpha_t\,(j)'$ we can write:

$$\alpha_t\ (j)'' = \frac{\alpha_t(j)}{\sum_{i=0}^{N-1}\ \alpha_t\ (i)'} \tag{1.3.46}$$

Which gives $\alpha_t\ (j)''$ as the required scaled values of $\alpha_t(j)$. From eq (1.3.45 and 1.3.46) it can be concluded that

$\sum_{i=0}^{N-1} \alpha_{T-1}(i)'' = 1$ and $\sum_{i=0}^{N-1} \alpha_{T-1}(i)'' = c_0 c_1 \ldots c_{T-1} \sum_{i=0}^{N-1} \alpha_{T-1}(i)$

$$= c_0 c_1 \ldots c_{T-1} P\ (O|\lambda)$$

Using these results we can write :

$$P\ (O|\lambda) = \frac{1}{\sum_{j=0}^{T-1}\ c_j}$$

We can avoid underflow in the computation by using the log of above equation as:

$$log[P\ (O|\lambda)] = -\sum_{j=0}^{T-1} log\ c_j \tag{1.3.47}$$

Similarly we can scale $\beta_t\ (j)$ variables by using c_t such that $\beta_t\ (j)'' = c_t\ \beta_t\ (j)$. We can thus replace $\alpha_t(j)$, $\beta_t\ (j)$ by $\alpha_t\ (j)''$, $\beta_t\ (j)''$ for calculating $\gamma_t(i)$ and $\varepsilon_t\ (i,j)$ in eq (1.3.36) and (1.3.39) respectively. Moreover, since in the reestimation formulae $P\ (O|\lambda)$ get cancelled thus eq (1.3.47) is used to check whether $P\ (O|\lambda)$ is increasing in every iteration.

1.3.6 Application of HMM in single molecule force spectroscopy:

The hopping of single molecule between its different molecular states is a typical phenomenon observed in single molecule force spectroscopy experiments [17, 37, 63, 64, 87, 94]. In these experiments step like transitions of the molecule are observed with respect to time. The observables in these hopping traces can be extension of the molecule in constant force measurements or the force values in constant distance experiments. These observable values are characterized by some probability distributions which are generally assumed as Gaussian [92]. To characterize the kinetics and thermodynamics of these fast fluctuations of molecules, HMM is considered as the most efficient statistical tool [70, 95]. The basic assumption in the HMM analysis of single molecule data is that each state of the molecule is a discrete variable evolving under Markovian dynamics [96] with a constant transition probability matrix A eq (1.3.25). The matrix of observation sequence B eq (1.3.27) corresponding to each state of the molecule is deduced from the Gaussian signals from extension or force time traces. The initialization of the HMM is done by feeding a rough estimate of the transition matrix A and the observation matrix B which is based on finding the initial guess of the number of states in the extension time traces. The HMM algorithm then deduces the maximum likelihood estimate of the transition matrix A and the observation matrix B (characterizing the mean and variance of each state) from the hopping traces. The transition matrix is then used for calculating the rates of transitions between different states of the molecule. The rates are calculated using the equation

$$k_{ij} = \frac{\ln(T_{ij})}{\Delta t} \tag{1.3.48}$$

Where Δt is the data acquisition time of the system. The transition matrix A and the observation matrix B is optimized using the equations (1.3.42) and (1.3.43) respectively. The observation matrix evaluated by HMM can then be used to calculate the change in extension of the molecule at each force in constant force measurements. The HMM algorithm provides an idealized trace of extension values, which can be fitted to the experimental data in order to check the robustness of the model used, as shown below (red curve).

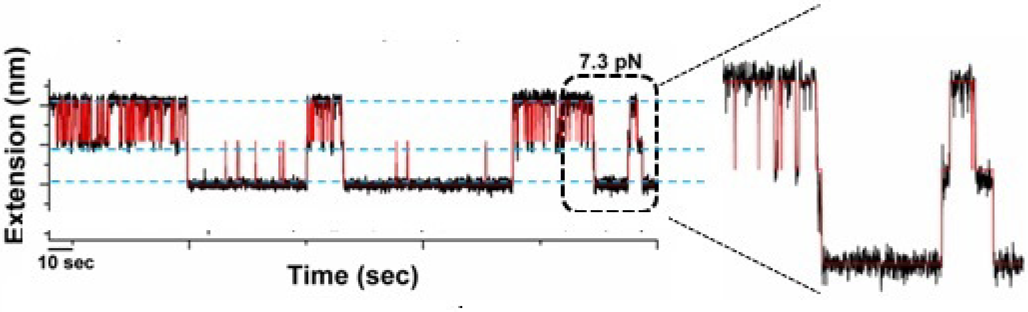


Figure 13: HMM fitting using a three state model for characterizing the kinetics of a multi state extension time trace. The red line shows the idealized trace generated by HMM.

From HMM analysis the lifetimes of each state can be calculated from which the probability of unfolding curves for each state can be generated (Fig. 14). For a two state reaction the probability of unfolding can be fitted with sigmoidal function as [83] :

$$P_U(F) = \frac{1}{(1 + \exp(\frac{\Delta G - F\Delta x}{k_B T}))} \tag{1.3.49}$$

Where ΔG and Δx are the fitting parameters as shown below:

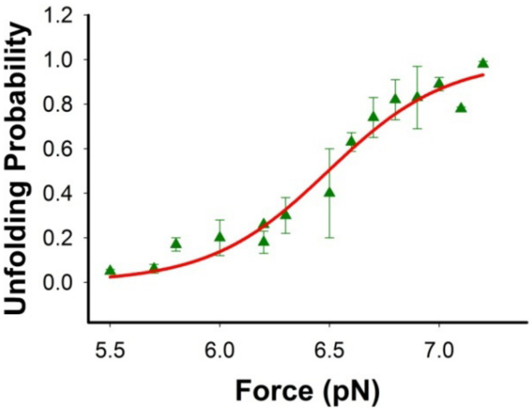


Figure 14: Probability of unfolding of a two state system. The experimental data can be fitted with a sigmoidal function from which the extension of the molecule and free energy of unfolding can be calculated [83].

1.3.7 Energy landscape reconstruction from Equilibrium measurements:

An advantage of equilibrium measurements of protein folding is that, the sum of transition state distances (Δx_f and Δx_u) as measured from the linearized forms of Bells model equations (1.3.22 and 1.3.23) give the extension of the molecule at $F_{1/2}$. Since the folding and unfolding trajectories in equilibrium measurements are identical, the energy landscape of the molecule at zero force can be reconstructed by keeping the change in extensions as the reaction coordinate [94] where the positions of the activation energy barriers are described by the kinetic distances Δx_f and Δx_u . For calculating the height of the activation energy barriers, equations (1.3.13) and (1.3.14) can be solved as:

$$\Delta G_u^{\neq} = k_B T\,(\ln A - \ln k_u^0) \qquad (1.3.50)$$

$$\Delta G_f^{\neq} = k_B T\,(\ln A - \ln k_f^0) \qquad (1.3.51)$$

Where the values for k_u^0 and k_f^0 are extracted from HMM analysis and the chevron plot (Fig. 3) and the pre exponential factor A, should be considered depending upon the experimental conditions.

1.3.8 Extracting kinetic parameters from force distributions:

When the rate at which force is applied on a molecule is faster than its slowest relaxation rate, the unfolding and refolding processes occur out of equilibrium. For such type of systems, generally force ramp and force jump methodologies [see section (1.2.7)] are used to extract the kinetics and thermodynamics of the folding/unfolding behavior of the molecule. In force ramp optical tweezers experiments [see section (1.2.7)], the rate at which the force is applied on the molecule, loading rate *r* (dF/dt in units of pN/s), typically becomes constant above 3-4 pN [37, 83]. For constant velocity data the kinetics of the folding reaction can be extracted using the model proposed by Evans and Ritchie [97]. For a first order reaction with negligible refolding rate, the probability that the molecule has not unfolded is [98].

$$\frac{dP_f(t)}{dt} = -k_u(t)P_f(t) \tag{1.3.52}$$

If force varies linearly with time *t* as *F* = *r t*, where *r* is the loading rate, then variable *t* can be changed to *F* in the above equation as

$$\frac{dP_f(F)}{dF} = -\frac{k_u(F)}{r}P_f\,(\mathrm{F}) \tag{1.3.53}$$

Integrating the above equation from 0 to *F* and using equation (1.3.18) we get

$$ln\{P_f(F)\} = \frac{k_u^0 k_B T}{r x_u}\left(1 - exp(F x_u/k_B T)\right) \tag{1.3.54}$$

From equation (1.3.54), the probability of unfolding as a function of force ($P_u(F)$) can be derived as

$$P_u(F) = 1 - P_f(F) = 1 - exp\left(-\frac{k_u^0 k_B T}{r x_u}(exp(F x_u/k_B T) - 1)\right) \tag{1.3.55}$$

By differentiating *Pu(F)*, which is a sigmoidal function of force, we can calculate the probability density as

$$\frac{dP_u}{dF} = \frac{k_u^0}{r}exp(F x_u/k_B T) * exp\left(-\frac{k_u^0 k_B T}{r x_u}(exp(F x_u/k_B T) - 1)\right) \tag{1.3.56}$$

The distance to the transition states measured from the above analysis is inversely proportional to the width of force distribution whereas k_u^0 determines the unfolding force required for loading rate r. From equation (1.3.56) we can calculate the relation for the most probable force for the unfolding of the molecule at which $\frac{dP_u}{dF} = 0$ as:

$$F_{mp} = \frac{k_B T}{x_u} ln(\frac{r.x_u}{k_B T k_u^0}) \tag{1.3.57}$$

The above relation shows that F_{mp} increases logarithmically with the loading rate.

Following similar steps the probability density function for the refolding force can be calculated as :

$$\frac{dP_f}{dF} = \frac{k_f^0}{r} exp\big(-Fx_f/k_B T\big) * \frac{exp\left(-\frac{k_f^0 k_B T}{r x_f}\left(\exp(-\frac{F x_f}{k_B T}) - 1\right)\right)}{exp\left(\left(-\frac{k_f^0 k_B T}{r x_f}\right) - 1)\right)} \tag{1.3.58}$$

Normalized distributions of the unfolding and refolding forces of a molecule manipulated at constant loading rate can be fit to equations (1.3.57) and (1.3.58) in order to estimate rate constants at zero force (k_u^0, k_f^0) and distances to the transition state (x_u, x_f) [65, 66]. Often however, experimental force distributions are analyzed with a slightly different method. When $exp(Fx_u/k_B T) > 10$, which is usually the case with biomolecules, equation (1.3.54) can be written as

$$ln\{P_f(F)\} = -\frac{k_u^0 k_B T}{r x_u}\big(exp(F x_u/k_B T)\big) \tag{1.3.59}$$

which can then be linearized as

$$ln\{r\, ln\big(1/P_f(F)\big)\} = ln\frac{k_u^0 k_B T}{x_u} + (x_u/k_B T)\, F \tag{1.3.60}$$

Through similar considerations, for the refolding process, we have

$$ln\{-r\, ln(1/P_u(F))\} = ln\frac{k_f^0 k_B T}{x_f} + (x_f/k_B T)\, F \tag{1.3.61}$$

Equations (1.3.60) and (1.3.61) are often used to fit *ln[rln[1/N]]* and *ln[-rln[1/U]]* vs. force graphs, where *N* and *U* are the folded and unfolded fractions, respectively, which are calculated by integrating the histograms of the force distributions over the corresponding range of forces [65, 66]. The above method assumes that the shift in the position of transition state barrier due to an applied force is negligible, however new theoretical models have improved the above formalism to include the shift in transition state positions. The probability density distribution of the unfolding forces are described by the Dudko model [99] which includes transition barrier which move with the applied force as:

$$\frac{dP_u}{dF} = \frac{k(F)}{r} * exp\left(\frac{k_u^0 k_B T}{r x_u} - \frac{k(F).k_B T}{r x_u} \left(1 - \frac{F.x_u}{\Delta G_u^{\neq}} v \right)^{1-\frac{1}{v}} \right) \qquad (1.3.62)$$

$$k(F) = k_u^0 \left(1 - \frac{F.x_u}{\Delta G_u^{\neq}} v \right)^{\frac{1}{v}-1} \exp\left\{ \frac{\Delta G_u^{\neq}}{k_B T} \left[1 - \left(1 - \frac{F.x_u}{\Delta G_u^{\neq}} v \right)^{\frac{1}{v}} \right] \right\} \qquad (1.3.63)$$

Where v defines the shape of the landscape like $v = 2/3$ gives linear cubic potential, $v = 1/2$ gives cusp shape and $v = 1$ gives equation (1.3.56). Using the above equations, new parameters like diffusion constant D and the transition path time Ʈ_{tp} over the barrier can also be extracted [100] which are defined as:

$$D = \frac{\pi}{3} \left\{ \frac{[k_u^0 x_u]^2}{\beta \Delta G_u^{\neq}} \right\} \exp(\beta \Delta G_u^{\neq}) \qquad (1.3.64)$$

$$\text{Ʈ}_{tp} = \frac{\ln(2\beta \exp(\gamma) \Delta G_u^{\neq}}{2\pi k_u^0 \sqrt{\frac{\mathrm{k_b}}{\mathrm{k_w}}}} \qquad (1.3.65)$$

Where k_w is the stiffness (curvature) of the potential well, k_b the stiffness of the barrier and γ is the eulers constant.

1.3.9 Extracting equilibrium information from non equilibrium single molecule manipulation data:

When a molecule is pulled and relaxed at a speed faster than its slowest relaxation rate, we observe hysteresis in its force extension curves [see section (1.2.7)], the size of which is proportional to the loading rate of pulling and relaxing the molecule. From second law of thermodynamics, we know that the work done on a system that is moved from one equilibrium state to another is related to free energy change of the system as $\Delta W \geq \Delta G$ and for a cyclic process $\Delta W \geq 0$. For a two state folding/ unfolding molecule out of equilibrium we can write

$$W_U \geq \Delta G = G_U - G_F$$

$$W_F \geq -\Delta G = G_F - G_U$$

Which shows that: $-W_F \leq \Delta G \leq W_U$. Here the equalities are attained if and only if the process is reversible. From the above relation we can write that $W_{cyclic} = W_F + W_U \geq 0$. If the cyclic process of folding and unfolding is repeated many times we get different values of work due to thermal fluctuations and role of water molecules in the process. As at microscopic level these fluctuations are not negligible, we can also observe rare events in which $W_{cyclic} < 0$ thus the second law should be used in terms of expectation values of work done. Thus the above relations can be written in terms of average values of work done in folding $\langle W_F \rangle$ and unfolding $\langle W_U \rangle$ as:

$$\langle -W_F \rangle \leq \Delta G \leq \langle W_U \rangle. \text{ and}$$

$$<W_{cyclic}> \; = \; <W_F + W_U> \; \geq 0$$

Here average values of work are realized by many repetitions of folding unfolding cycles. Based on the thermal fluctuations of the work values, various fluctuation theorems have been developed to extract the free energy change of unfolding a molecule from the average work done in the microscopic processes [101, 102] . The first of these theorems which was successfully applied in single molecule force spectroscopy was Jarzynski's equality[103]. For a system with states positioned at 0 and z along a reaction coordinate, the free energy change between the states $\Delta G\ (z)$, is related to the Boltzmann weighted work values obtained by repeated irreversible switching of the system between its states by the Jarzynski's equality as:

$$\exp[-\beta\Delta G(z)] = LimN_{N_{\rightarrow}\infty} < \exp[-\beta\ wi(z,r)] >_N \qquad (1.3.66)$$

where < > denotes averaging over N work trajectories, $w_i(z,r)$ represents the work of the *i-th* of N trajectories, and r is the switching rate. The above equality was first successfully verified, in the single molecule manipulation of P5ab RNA hairpin out of equilibrium using optical tweezers [104]. However, since the above equality includes Boltzmann averaging of work thus only low work values will contribute which demands extensive number of experiments and many pulling cycles. This method is convenient to apply for molecules like P5ab RNA, that (un) folds close to equilibrium however in far from equilibrium systems, Jarzynski's equality is hampered by large statistical uncertainties due to the exponential averaging of low work values [105]. A more robust method for extracting thermodynamic free energies of molecules that unfold far from equilibrium is the Crooks Fluctuation theorem (CFT) [106]. CFT state that if $P_U(W)$ and $P_R(W)$ denote the probability distributions of the work performed on a molecule that is pulled (U) and relaxed (R) an infinite number of times, then the work done on the molecule is related to free energy change as:

$$\frac{P_U(W)}{P_R(W)} = \exp\left(\frac{W - \Delta G}{k_B T}\right) \qquad (1.3.67)$$

To apply CFT in single molecule manipulation experiments it is important that the molecule should be pulled and relaxed with same speeds (time reversal symmetry) and the molecule is always pulled from an equilibrium position (native state) and relaxed from an equilibrium unfolded state. CFT states that although $P_U(W)$ and $P_R(W)$ depends on the pulling protocol their ratio depends only on ΔG. The point of intersection of the two distributions then give the ΔG of the molecule. CFT has already been applied in several single molecule manipulation studies, as in [68, 87, 107]. It can be shown that the Jarzynski's equality is a consequence of the CFT by multiplying equation (1.3.67) by a generic function Φ (W) and integrating over W to get:

$$\exp[-\beta\Delta G] \;=\; \frac{<\Phi\,(\mathrm{W})\exp[-\beta\,W]>_U}{<\Phi\,(-\mathrm{W})>_R} \qquad (1.3.68)$$

If we put Φ (W) = 1, then we get the Jarzynski's equality eq (1.3.64) from above. Since CFT utilizes only the local behavior of work distributions around ΔG thus it is not an accurate estimator of free energy change and for non overlapping distributions due to systems far from equilibrium the uncertainties in the calculated values are high. For such cases Bennet's function [108] should be used in equation (1.3.68) to minimize the statistical variance in the estimator.

$$\Phi\,(\mathrm{W}) \;=\; \{1 + \frac{n_u}{n_f}\exp\left(\frac{W - \Delta G}{k_B T}\right)\}^{-1} \qquad (1.3.69)$$

Where n_u and n_f are the number of forward and reverse events respectively [109]. For systems in which the molecule occupies partially equilibrium states, CFT has been modified into extended fluctuation theorems [110] and these methods have been successfully verified in single molecule optical tweezers experiments recently [67].

1.4 Model Proteins.

To understand how mechanical force and changes in the environment of proteins effect their folding (unfolding) behavior and cause misfolding, we chose human neuronal calcium sensor 1 (NCS-1) and β-2 Microglubulin as our model systems. Considering that a lot of studies have been done on these proteins, information regarding their structure and functionality was available for us to correlate with our results. Particularly, studies done in bulk on NCS-1 allowed us to directly match our single molecule data and confirm the validity of our conclusions. Using optical tweezer set up we performed single molecule manipulation experiments on NCS-1 molecule in presence of Mg^{2+} and in apo form (no divalent ions). Whereas forces involved in the amyloidogenesis of β-2 Microglubulin were theoretically studied by us. In this section the structure and functions of the model proteins studied are discussed.

1.4.1 Neuronal calcium sensor 1 (NCS-1): Structure and Functions

Structure:

NCS-1 is an EF hand protein with 190 residues, which belongs to the family of Ca^{2+} binding proteins such as GCAP, troponin C, CaM, and recoverin [111]. Fig. 15 depicts the structure of NCS-1 in presence of Ca^{2+} in which the molecule has two domains – the

N and C domain [112]. The N domain consist of two EF hand motifs EF 1 and EF 2 while the C domain consist of EF 3 and EF 4 motif hands. The EF hand motifs are basically two α helices that are linked by a short loop consisting of 12 amino acids that can bind with divalent ions like Ca^{2+} and Mg^{2+}. EF 1 motif hand does not bind with any ions due to a conserved cys /pro mutation. EF 2 and EF 3 motifs bind with both Ca^{2+} and Mg^{2+} ions and they are known as structural sites whereas EF4 motif bind with only Ca^{2+} and is known as sensory site [94, 113]. NCS-1 in presence of Ca^{2+} acquires an open structure revealing large amount of hydrophobic patches which allows it to interact with target proteins for carrying out its functions [112]. Recently our single molecule studies have confirmed that the folding of NCS-1 in presence of Ca^{2+} takes place sequentially, where the C domain folds first followed by the complete folding of the N domain [94]. Moreover the folding process corresponds to the binding of the Ca^{2+} ions to the EF hand motifs, such that Ca^{2+} binds first with EF3 and then EF 4 resulting in the complete folding of C domain and finally to EF2 that causes the folding of the N domain. The structure of NCS-1 in presence of Ca^{2+} has been studied in great details using X ray crystallography and NMR [112, 114-116] however bulk studies have shown that NCS-1 can also exist as two other conformers - Mg^{2+} bound and apo form (no divalent ions) at physiological conditions of cells [113]. The structures of these conformers are still elusive and only few details are available from NMR studies where it is proposed that the apo form of NCS1 has a molten globule like structure [117] while in presence of Mg^{2+} NCS-1 adopts a very closed form revealing almost no hydrophobic patches [118].

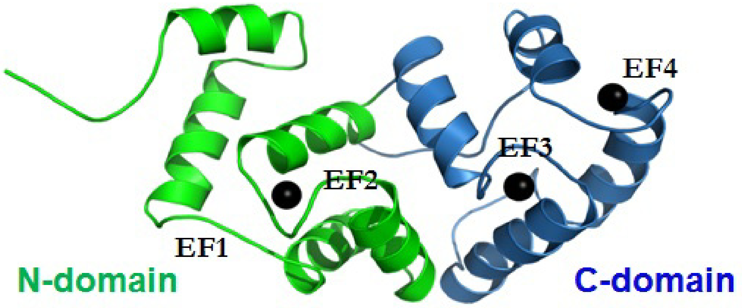


Figure 15: Structure of NCS-1 with two domains N (in green) and C (in blue) in presence of Ca $^{2+}$. The N domain has EF hands motifs EF 1 and EF2 and C domain has EF 3 and EF 4 motif hands [112].

Functions:

NCS-1 belongs to the family of neuronal calcium sensor (NCS) proteins that are expressed inside neurons and photoreceptor cells [119, 120]. NCS-1 carries out diverse neuronal functions such as membrane trafficking, neuro transmitter release, learning and neuronal growth [119]. NCS-1 on binding with Ca^{2+} undergoes large conformational changes which allow it to bind with different target proteins for carrying out its functions [121]. At present 22 binding partners of NCS-1 has been reported [119, 122]. Moreover NCS-1 also binds with many target proteins in its Mg^{2+} bound and apo form conformers and it is proposed that these conformers are also responsible for many neuronal functions [119]. NCS-1 interacts with dopamine receptor D2 [123] and in the activation of P14Kβ [124] in its Mg^{2+} bound form. Moreover NCS-1 is also responsible for many neuronal disorders such as Schizophrenia, bipolar disorder and autism [119, 125]. Since Ca^{2+} levels frequently fluctuates inside neurons [126] it is believed that these disorders are related to the varying Ca^{2+} levels. In this thesis one project has been explained in which the misfolding mechanisms of NCS1 are studied at single molecule level at different Ca^{2+} concentrations.

1.4.2 β -2 Microglobulin (β-2 M)

Structure:

β-2 M is the light chain constituent of the class I human lecocyte antigen (HLA class I). It has 99 residues and its structure is similar to the Ig family having a seven β stranded structure [127]. Till now about 80 crystal structures of β-2 M bound to HLA complex have been reported and they all display the seven stranded β sandwitched native state (see Fig. 16) [128]. Among the seven β strands four of them comprise one β sheet while the remaining three forms the second β sheet in the structure. The structure is stabilized by a single disulphide bond linking the two β sheets [129]. The structure of monomeric β-2 M studied using X ray diffraction combined with NMR [130] revealed distinct changes compared to the HLA β-2 M [127] and the structure of monomeric β-2 M studied in solution [131]. These changes correspond to a β bulge that serves as separation between two β strands resulting in a structure comprising of six β strands at one side of the protein [130]. Moreover these structural changes are also proposed to be responsible for making monomeric β-2 M more prone to aggreagtion [130, 132].

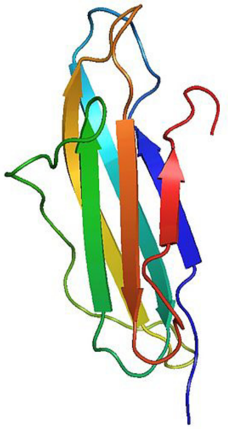


Figure 16: Structure of β-2 M having a seven β strands that forms two β sheets.

Amyloidogenesis of β-2 M :

In adults, β-2 M is produced at a rate of 200mg/day and it continuously dissociates from cell surface into the serum and from where it is then cleared by the kidneys. In patients with long term dialysis or renal failure, the concentration of β-2 M in the serum increases by 60 fold [133] resulting in the aggregation of free β-2 M molecules into amyloid fibrils. The amyloid fibrils get deposited in bones and joints causing serious problems like arthropathy and bone cysts which is known as disorder dialysis related amyloidogenesis [21]. Several studies have been done in understanding the strcuture and function of β-2 M however the mechansim of its amyloid fibril formation and the forces responsible for it are still elusive [134].

1.5 Objectives

To gain more information about how proteins fold and misfold, we have used single molecule mechanical manipulation using optical tweezers [see section (1.2)], as our experimental strategy for studying the model protein NCS-1 [see section (1.4.1)]. Two projects *(papers 1,3)* are explained in this thesis in which the folding and misfolding trajectories of NCS-1 were studied in varying ionic conditions. Using single molecule force spectroscopy and advanced statistical tools like hidden Markov model the full energy landscapes of folding and misfolding trajectories of NCS-1 were studied in these projects. In another project the amyloidogenesis of β-2 Microglubulin, [see section (1.4.2)] has been explained in which we have theoretically calculated the forces responsible for the amyloid fibril formation of β-2 Microglubulin in laminar flow and in presence of hydrophobic surfaces *paper2***.** The objectives behind each of these projects are listed below:

1.5.1 Direct observation of calcium-dependent misfolding in single Neuronal calcium sensor-1 molecules:

As discussed in [see section (1.4.1)], NCS-1 is associated with various neuronal disorders such as Schizophrenia, bipolar disorder and autism [119, 125]. Moreover inside neurons it has been shown that the concentration of Ca^{2+} varies frequently [126] and it is believed that these varying Ca^{2+} levels could be responsible for the misfolding of NCS-1 resulting in the diseases mentioned above. The mechanism by which protein molecules enter off pathway trajectories and by non native interactions forms misfolded structures is an important subject to understand [see section 1.1.5]. Recently single molecule mechanical manipulation has helped in understanding the misfolding behavior of several proteins [18, 64, 135]. In this project our objective is to study the off pathway trajectories of NCS-1 at varying calcium concentrations similar to in vivo conditions. Moreover in this project we planned our experiments to characterize the thermodynamic and kinetics of the misfolding transitions two Ca^{2+} concentrations (10mM and 10 μM). From the kinetics and thermodynamic information we can reconstruct the possible energy landscapes of the off pathway transitions which will help in understanding the factors responsible for the misfolding of NCS-1 due to variation in calcium concentrations in cells.

1.5.2 Structure, folding dynamics, and amyloidogenesis of D76N β2-microglobulin. Roles of shear flow, hydrophobic surfaces and α crystalline:

This project was devoted in understanding the factors responsible for the amyloidogenesis of the model protein β2-Microglobulin *see sections 1.4.2*. Numerous studies have been done in elucidating the structure and function of β2-Microglobulin [129-131] and the structural characteristics of its amyloid fibrils [136], however little is known about the early events of its fibrillization process. In this study, the effect of factors like shear flow and interaction with hydrophobic surfaces on the rate of fibrillization of β2-Microglobulin and its structural variant D76N β2-Microglobulin [137] were investigated. Considering that in cells the proteins do interact with various hydrophobic surfaces and there are conditions involving laminar flows [138], it is interesting to experimentally study what roles these factors play in the early events of fibrillization of β2-Microglobulin. Moreover, since the WT β2-Microglobulin has a more stable native state than its variant D76N, this study will help in elucidating the role of the thermodynamic stability of the molecule in its fibrillization process [139]. In this study theoretical calculations were done to estimate the range of the shear and hydrophobic forces involved in the destabilization of the native state of β2-Microglobulin. The estimation of the forces will elucidate the most dominant factor in the early events of the fibrillzation process.

1.5.3 Single-molecule folding mechanisms of the apo- and Mg^{2+}-bound forms of human Neuronal Calcium Sensor-1:

Neuronal calcium sensor 1 (NCS-1) is the primordial member of the NCS family and is well studied for various functional roles such as regulating neuronal ion channels, membrane traffic, learning and neuronal growth [119]. NCS-1 interacts with multiple target proteins in executing these functions [111]. In vitro, NCS-1 has been experimentally observed to have three conformers – Ca^{2+} bound, Mg^{2+} bound and in apo form. It is believed that these conformers exist in nature in different cell conditions [113, 117, 140]. In neurons, calcium binds with NCS proteins and plays a major role in actuating neurotransmitter release [141, 142]. However due to chemical similarity and high intracellular concentrations, Mg^{2+} has been found as a major competitor for Ca^{2+} binding sites in Ca^{2+} sensor proteins [140, 143]. Although, extensive studies have been done on the Ca^{2+} bound state [144, 145] little is known about the structure and functionality of NCS-1 - Mg^{2+} bound and apo form conformers. Recently we have examined the folding (un) behavior of Ca^{2+} bound form of NCS1 and reconstructed its full energy landscape [94]. In this project we have used optical tweezers to study the folding mechanism of NCS-1 in Mg^{2+} bound and apo-form states at single molecule level. Our objective in this study was to understand the folding mechanisms and reconstruct the energy landscapes of Mg^{2+} bound and apo-form states of NCS-1. The comparison of energy landscapes of each NCS-1 conformer will not only help in elucidating their respective physiological roles but will also give us insight into the effect of varying environmental conditions of a protein on its folding behavior.

2. PAPERS AND MANUSCRIPTS

The manuscripts of the research projects carried out in this PhD work are arranged in the following order:

Paper 1 : *Direct Observation of Calcium-dependent Misfolding in Single Neuronal Calcium Sensor-1 Molecules.*

Paper 2 : *Structure, Folding Dynamics, and Amyloidogenesis of D76N* β*2-Microglobulin. Roles of shear flow, hydrophobic surfaces and* α *crystalline.*

Paper 3 : *Single-molecule folding mechanisms of the apo- and* Mg^{2+}*-bound forms of human Neuronal Calcium Sensor-1.*

PAPER I

Direct single-molecule observation of calcium-dependent misfolding in human neuronal calcium sensor-1

Pétur O. Heidarsson[a], Mohsin M. Naqvi[b,c], Mariela R. Otazo[b,d], Alessandro Mossa[e], Birthe B. Kragelund[a,1], and Ciro Cecconi[b,c,1]

[a]Structural Biology and NMR Laboratory, Department of Biology, University of Copenhagen, 2200 Copenhagen, Denmark; [b]Institute of Nanoscience S3, Consiglio Nazionale delle Ricerche, 41125 Modena, Italy; [c]Department of Physics, Informatics and Mathematics, University of Modena and Reggio Emilia, 41125 Modena, Italy; [d]Department of Physics, Center of Applied Technologies and Nuclear Development, Havana, Cuba; and [e]Department of Physics, University of Bari and Istituto Nazionale di Fisica Nucleare, Sezione di Bari, 70126 Bari, Italy

Edited by Martin Gruebele, University of Illinois at Urbana-Champaign, Urbana, IL, and approved August 8, 2014 (received for review February 5, 2014)

Neurodegenerative disorders are strongly linked to protein misfolding, and crucial to their explication is a detailed understanding of the underlying structural rearrangements and pathways that govern the formation of misfolded states. Here we use single-molecule optical tweezers to monitor misfolding reactions of the human neuronal calcium sensor-1, a multispecific EF-hand protein involved in neurotransmitter release and linked to severe neurological diseases. We directly observed two misfolding trajectories leading to distinct kinetically trapped misfolded conformations. Both trajectories originate from an on-pathway intermediate state and compete with native folding in a calcium-dependent manner. The relative probability of the different trajectories could be affected by modulating the relaxation rate of applied force, demonstrating an unprecedented real-time control over the free-energy landscape of a protein. Constant-force experiments in combination with hidden Markov analysis revealed the free-energy landscape of the misfolding transitions under both physiological and pathological calcium concentrations. Remarkably for a calcium sensor, we found that higher calcium concentrations increased the lifetimes of the misfolded conformations, slowing productive folding to the native state. We propose a rugged, multidimensional energy landscape for neuronal calcium sensor-1 and speculate on a direct link between protein misfolding and calcium dysregulation that could play a role in neurodegeneration.

protein folding | NCS-1 | off-pathway intermediate | conformational dynamics | optical trapping

Most proteins have evolved to fold rapidly into a specific and functional 3D structure immediately after translation from the ribosome. The folding process is, however, not adequately efficient to prevent the occurrence of misfolded states in vivo (1), especially in the case of larger multidomain proteins which comprise roughly 75% of the human proteome (2, 3). Normally, to tackle and destroy these unproductive structures, cells are equipped with competent clean-up machinery, such as chaperones, proteasomes, and unfoldases (4). If misfolding cannot be ameliorated, these nonnative states accumulate in the cell to form aggregates with potential pathophysiological consequences (5).

The emerging view that protein misfolding is a common phenomenon in living cells is still largely unsubstantiated, as detecting and characterizing misfolded states has been experimentally challenging (2, 6). The mechanistic details that have accumulated over the last decades on misfolding have mostly come from studies on the resulting oligomeric structures and amyloid formation (1), whereas our understanding of the structural rearrangements and pathways leading to precursory misfolded states is still highly incomplete. Importantly, the formation of prefibrillar monomeric and oligomeric misfolded states is, contrary to amyloids, reversible and thus these states provide a potential target for drug design.

Sparse populations and their associated weak signals limit the use of traditional bulk methods for monitoring the early events of misfolding, and relatively few systems have been studied in detail (7–13). Now, single-molecule force spectroscopy techniques, such as optical tweezers, enable detection of rare alternative folding pathways and short-lived misfolded states by direct mechanical manipulation (14–19). Although aggregation requires more than one molecule, nonnative structural rearrangements within a single molecule only report on monomeric misfolded states. Recent works have exploited these properties to study misfolding of well-known disease-related proteins, such as the prion protein, as well as proteins not generally associated with misfolding, such as the EF-hand calcium sensor calmodulin (CaM) (20–22).

The EF-hand superfamily of calcium sensors is responsible for translating changing levels of intracellular Ca^{2+} concentration into a biochemical signal through conformational changes that allow them to interact with an array of binding targets (23). The subfamily of neuronal calcium sensors (NCS) is mostly expressed in neurons and currently includes 15 members (24, 25). Neuronal calcium sensor-1 (NCS-1) is the most ancient member of this family (Fig. 1*A*), and it has been functionally associated with

Significance

Protein misfolding can lead to neurodegeneration. Yet, the mechanistic details of this deleterious phenomenon are largely unknown, as experimental portrayals of the early and reversible molecular events leading to misfolded conformations have so far remained highly limited. Here we use single-molecule optical tweezers to monitor the structural rearrangements leading to misfolded conformations of human neuronal calcium sensor-1, a protein linked to serious neurological disorders. We identified two distinct and calcium-dependent misfolding trajectories originating from an on-pathway folding intermediate. Remarkably for a calcium sensor, pathologically high calcium concentrations impede correct folding by increasing the occupation probabilities of the misfolded states. The results open ostensible links between protein misfolding and calcium dysregulation that could be important in neurodegeneration and its potential inhibition.

Author contributions: P.O.H., M.R.O., B.B.K., and C.C. designed research; P.O.H., M.M.N., and M.R.O. performed research; A.M. contributed new reagents/analytic tools; P.O.H., M.M.N., M.R.O., A.M., B.B.K., and C.C. analyzed data; and P.O.H., M.M.N., A.M., B.B.K., and C.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: ciro.cecconi@gmail.com or bbk@bio.ku.dk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1401065111/-/DCSupplemental.

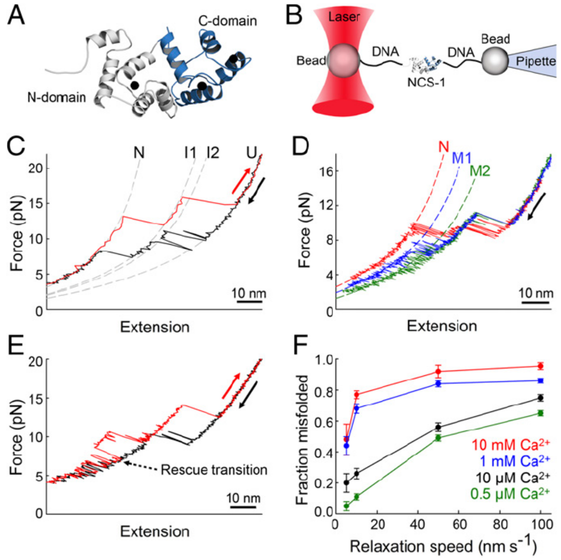


Fig. 1. Misfolding pathways of NCS-1. (*A*) The NMR structure of NCS-1 (PDB 2LCP), with the N domain (EF1/EF2) depicted in gray and the C domain (EF3/EF4) in blue. Black spheres represent Ca^{2+} ions. EF1 does not bind Ca^{2+} because of a conserved cysteine-proline mutation (35). (*B*) Sketch of the experimental setup. NCS-1 was tethered between functionalized beads via DNA handles and stretched and relaxed by moving the pipette relative to the optical trap (16, 62). (*C*) Native folding pathway of NCS-1. After being mechanically stretched and unfolded (red trace), NCS-1 refolds upon relaxation of the applied force into its native state via two intermediate states, I1 and I2. Dashed lines are worm-like-chain fits to the data. Color-coded arrows indicate the pulling/relaxing directions. (*D*) Misfolding pathways of NCS-1. During refolding, NCS-1 sometimes follows alternative pathways leading to misfolded states M1 (blue) or M2 (green), which are less compact than the native state (red). Dashed lines are worm-like-chain fits to the data. (*E*) Rescue of the native state of NCS-1. During relaxation (black), the molecule misfolded. During stretching (red), nonnative contacts of the misfolded conformation are progressively broken, until the molecule can find its native folding pathway ("rescue transition"). (*F*) Fraction of folding pathways leading to either M1 or M2 as a function of Ca^{2+} concentration and relaxation speed. Higher Ca^{2+} concentrations and relaxation speeds facilitate NCS-1 misfolding. Error bars indicate SEs of mean. At least five different molecules were used for each calcium concentration.

cognitive processes, such as learning and memory (26, 27), and with a number of cellular processes such as neurotransmitter release (28, 29), and regulation of ion channels, and G protein coupled receptors (GPCRs) (24, 30), including the dopamine receptor D2 (31). NCS-1 has also been linked to serious neurodegenerative disorders including schizophrenia, bipolar disorder (BD) (32), and autism (33, 34). However, the dysfunctions of NCS-1 are poorly characterized on the molecular level, and whether they involve altered functional profiles or loss of function due to formation of misfolded states is currently unknown.

Because only a few systems have been studied experimentally, little is known about folding and/or misfolding mechanisms of members of the EF-hand superfamily (21, 35–37). The extensively studied CaM has been shown on the single-molecule level to frequently visit misfolded states that slow down the overall folding rate of the protein (21). The physiological consequences of CaM misfolding have not yet been explored. NCS-1 shares modest sequence homology with CaM, mostly within and around the calcium binding sites (24). Similar to CaM, NCS-1 contains four EF hands organized in two EF domains (Fig. 1*A*) yet it exhibits a larger number of interdomain contacts (38), a feature that has been suggested to increase the probability of misfolding in proteins (2). The formation of misfolded states along the folding pathway of NCS-1 may have important consequences with regards to its function as a calcium sensor and might also play a role in disease pathologies.

Using optical tweezers, we have recently characterized the native folding pathway of NCS-1 (39). Here we use a similar experimental approach to monitor individual NCS-1 molecules as they populate nonnative misfolded states in real time. We identified two misfolding trajectories leading to two distinct misfolded conformations, characterized by different extensions and different pathways on the energy landscape. Both misfolding pathways originated from a partially folded on-pathway intermediate state, and they competed with native folding. The occupancy probability of both misfolded states could be controlled by modulating either the relaxation rate of the applied force and/or the calcium concentration. Remarkably for a calcium sensor, higher calcium concentrations, even within physiologically relevant conditions, lead to an increased probability of NCS-1 misfolding.

Results

Individual NCS-1 molecules were manipulated with polystyrene beads by means of DNA molecular handles covalently attached to cysteine residues engineered at positions 4 and 188, as previously reported (39) (Fig. 1*B* and *Methods*). To identify and characterize misfolding pathways of NCS-1, we first performed constant-velocity experiments, where the molecule is stretched and relaxed by moving the pipette at constant speed relative to the optical trap (16, 39, 40). Under these experimental conditions, we have previously shown that NCS-1 unfolds in an apparent three-state manner and folds back into its native state through a process that starts with the folding of the C domain, undergoing a major conformational change (U→I2) followed by a minor rearrangement (I2→I1), and ends with the folding of the N domain at lower forces (I1→N transition) (Fig. 1*C*). The I2→I1 transition is the rate-limiting step of the overall folding process and is mandatory for the subsequent folding of the N domain. Here we show that, in addition to the pathway leading to the native state, NCS-1 folding can follow alternative pathways leading to nonnative (misfolded) structures (Fig. 1*D*). In these cases, the protein folds into I2 but fails to transit into I1, thus never reaching the natively structured C domain. At lower forces, the molecule starts fluctuating between I2 and misfolded conformations, before being trapped in either of two main misfolded states, M1 and M2, differentiated by their extension (Fig. 1*D*, green and blue traces). A fit of the force vs. extension traces to the worm-like-chain (WLC) model of polymer elasticity (41) yielded a contour length change (ΔL_c) of 29 ± 2 nm ($n = 35$; errors are given as SD) for the I2–M1 transition, and $\Delta L_c = 18 \pm 3$ nm ($n = 31$) for the I2–M2 transition. These states are clearly different from the native state as the native I2–I1–N transition has an overall associated contour length change of 34 ± 2 nm (39). Furthermore, M1 and M2 displayed drastically reduced mechanical stability compared with the native state and unfolded at lower forces (Fig. S1). During stretching, kinetically trapped misfolded molecules were occasionally (<1%) observed to snap into their native state through a "rescue transition" that suddenly shortened the extension of the molecule (Fig. 1*E*, red trace). The rescue transition occurred exclusively from I2, not from M1 or M2, and usually nonnative interactions had to be mechanically broken in an unfolding step before the molecule could fold into N, supporting the off-pathway nature of the kinetically trapped conformations. Upon further increase of force, the molecule displayed the signature of high-force unfolding transitions of both the N and C domains, verifying a native fold.

To examine any possible effects of Ca^{2+} concentration and relaxation speed on the misfolding behavior of NCS-1, we performed experiments in 0.5 μM, 10 μM, 1 mM, and 10 mM $CaCl_2$,

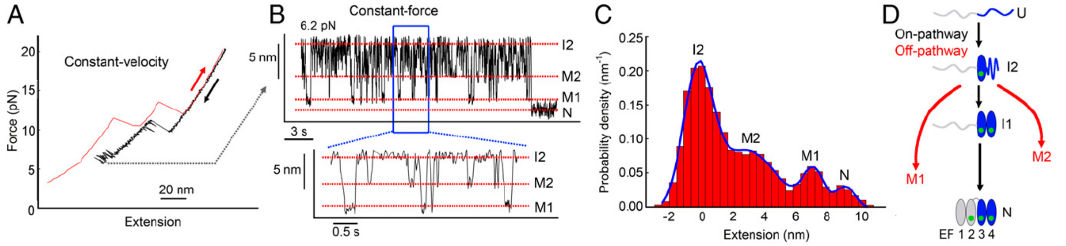


Fig. 2. Misfolding of NCS-1 at equilibrium. (*A*) Force vs. extension cycle showing a mechanically denatured NCS-1 molecule (red trace) that, during relaxation of the applied force, follows a misfolding pathway (black trace). Relaxation was stopped at 6.2 pN, and (*B*) the force was clamped via a feedback mechanism to observe NCS-1 fluctuate at equilibrium between I2 and misfolded states M1 and M2, until it eventually folded into its native state. (*Inset*) A magnified view of the data in the blue box. (*C*) Probability density distribution of the extension values of *B*, resolving misfolded states M2 and M1. The distribution was fit by the sum of four Gaussian functions (blue line) to show that M2 and M1 are located at 3.2 ± 2.4 nm and 7.6 ± 1.1 nm, respectively, from I2. (*D*) Diagram of the different molecular states populated during the folding of NCS-1. Fully folded EF hands are shown as gray or blue ellipses, for the N or C domain, respectively, and Ca^{2+} ions are shown as green circles.

at various relaxation speeds. At all calcium concentrations, the misfolding probability of NCS-1 increased with increasing relaxation speed (Fig. 1*F*). Strikingly, the misfolding probability also increased with increased Ca^{2+} concentration. At a physiologically relevant calcium concentration (0.5 μM) and at slow relaxation speed (5 $nm \cdot s^{-1}$), only 5% of the molecules misfolded. When the calcium concentration was raised to 10 mM, almost 50% of the molecules misfolded. These data are highly interesting, as various physiological or pathophysiological phenomena can cause a transient or sustained increase in neuronal calcium concentrations (42–44), as further addressed in *Discussion*. The two misfolded states, M1 and M2, were also diversely populated at different Ca^{2+} concentrations (Fig. S2). The M1 state was highly populated at 10 mM Ca^{2+} and less populated at 1 mM Ca^{2+}, whereas it was absent at lower Ca^{2+} concentrations. The M2 state instead was equally populated under all four experimental conditions. Rescuing transitions were observed at all calcium concentrations, and their frequency increased with lower calcium concentrations (~1% at 10 mM to ~5% at 10 μM).

To characterize the kinetics and thermodynamics of the misfolded states, we performed constant-force measurements, both at high (10 mM) and low (10 μM) calcium concentrations. Under both conditions, the calcium concentration is above bulk dissociation constants (4–400 nM) (35), suggesting the calcium binding sites of NCS-1 to be saturated. In these experiments, a molecule was stretched and relaxed multiple times until it was observed to populate a pathway leading to a misfolded state (Fig. 2*A*). At this point, the relaxation of the molecule was stopped and the applied force was clamped via a force-feedback mechanism to a specific tension to observe NCS-1 fluctuate at equilibrium between I2 and misfolded conformations, before eventually folding into N (typically after 20–100 s) (Fig. 2*B*). In all cases (>300 events), the transition to N took place from I2, providing additional evidence for the off-pathway nature of the misfolded states. After the molecule had folded into the native state, it did not unfold again within the measuring time (>5 min). The probability density distribution of the extension signal revealed that in 10 mM Ca^{2+}, NCS-1 primarily populates two misfolded states positioned at 3.2 ± 2.4 nm and 7.6 ± 1.1 nm from I2 (Fig. 2*C*). These extensions compare well to those of the misfolded states M2 and M1 observed in constant-velocity measurements (Fig. 1*D*), suggesting that the most probable nonnative conformations populated by NCS-1 at equilibrium are also the conformations in which it remains kinetically trapped at lower forces. The kinetic network of the on- and off-pathway states of NCS-1 is shown in Fig. 2*D*.

To characterize the energies and transition kinetics of M1 and M2, extension-time traces were recorded at different forces, in the range between 5.4 pN and 7.0 pN. A population shift from I2 to the misfolded states was observed when the applied force was reduced (Fig. 3*A*). The force-dependent rates of the I2–M2 and I2–M1 transitions were subsequently determined by analyzing each extension trace with a hidden Markov model (HMM) algorithm (45–47). Experimental data are well modeled by a four-state Markov system where M1 and M2 are only connected to I2, while the transition from I2 to N (in our force range, the transient occupation of state I1 is too short-lived to be detected) is only one way. The position of the transition states along the reaction coordinate and the zero force rates were then estimated by fitting the data to a linearized form of the Bell model (39, 48)

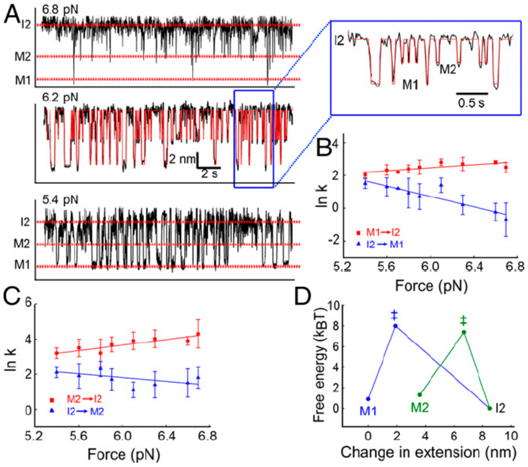


Fig. 3. Energetics and kinetics of NCS-1 misfolding in 10 mM Ca^{2+}. (*A*) Extension vs. time traces of NCS-1 fluctuations at indicated forces. As the force is lowered, the M1 and M2 states are increasingly populated. (*Inset*) A magnified view of the misfolding transitions in the blue box, and the corresponding fit based on the HMM (red line). *B* and *C* show unfolding (red) and refolding (blue) rate constants at different forces for the I2–M1 and I2–M2 transitions, respectively. Solid lines are fits with a linearized form of the Bell model (40, 48). Error bars represent SDs. (*D*) Sketch of the free-energy landscapes for the I2–M2 (green) and I2–M1 (blue) transitions at 5.5 pN and 10 mM Ca^{2+}, reconstructed using the HMM analysis. The I2 state was considered the reference state at zero energy.

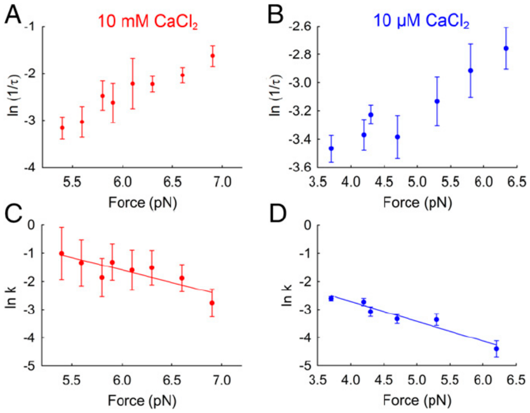


Fig. 5. Occupation of misfolded states slows the overall folding rate of NCS-1. In *A* and *B*, τ is the total time elapsed from the beginning of the force clamp until the protein eventually folds into N (this takes into account also the time spent visiting misfolded states). In *C* and *D*, we plot the force dependence of the rate constants for the I2–I1 transition in 10 mM and 10 μM CaCl$_2$, respectively. The solid line represents fits to the data according to the Bell model. The extrapolated zero-force folding rate at 10 mM CaCl$_2$ is 40 ± 4 s^{-1}, in remarkable agreement with the value of 38 ± 5 s^{-1} that we found performing experiments at higher force (around 10 pN) (39). The folding rate at 10 μM is lowered by a factor of ∼40 (1.1 ± 1.6 s^{-1}), whereas the distance to the transition state is similar at both calcium concentrations (x_f = 3.3 ± 1.0 nm at 10 mM CaCl$_2$; x_f = 2.9 ± 0.4 nm at 10 μM CaCl$_2$).

modulating the relaxation rate of the applied force, a phenomenon previously observed with RNA hairpins (52) but not, to our knowledge, with proteins (16, 21, 40, 53). Precision control over protein folding energy landscapes may find practical applications in the study of molecules that have established links to misfolding diseases. The fate of a protein during folding strictly depends on the topography of its energy landscape. Our results suggest a complicated and rugged energy landscape for NCS-1, with several local maxima and minima that can act as kinetic traps. During folding, nonnative contacts can become accessible before native contacts form (54). If the force is decreased faster than the rate at which nonnative contacts can break, the molecule becomes kinetically trapped in misfolded structures. Mechanical force can then be used to pull out misfolded species from their energy minima and rescue native folding.

Our results add a new layer of complexity to the folding process of NCS-1, revealing a rugged multidimensional folding energy landscape governed by a kinetic partitioning mechanism. The process invariably starts with the folding of the polypeptide chain into the intermediate state I2. From here, it can proceed toward the native state by transiting into intermediate state I1 or, in a calcium-dependent manner, take alternative pathways leading to misfolded states M1 and M2. Two distinct misfolded states have previously been observed in the structurally related two-domain EF-hand protein CaM (21), yet those states appear to be different from the ones observed for NCS-1. First, the modulation of misfolding by calcium has, to the best of our knowledge, not yet been reported for CaM, or for any other protein. Second, for both misfolded states of NCS-1, the C domain is only partially folded, whereas both CaM misfolded states consist of EF hands that are either unfolded or fully folded, although incorrectly paired. The difference in the misfolding mechanisms may be a reflection of the different structural architectures of the two proteins. CaM has a rather symmetrical structure, with a long α-helix separating the two almost identical N and C domains, each binding two Ca^{2+} ions (55). In contrast, NCS-1 is characterized by a certain asymmetry, as the N domain binds only one Ca^{2+} ion, and by significantly more interdomain contacts (38). Despite these differences, the occupation of misfolded states has been shown to slow down the overall folding rates of both NCS-1 and CaM, with a potential impact on their functions (21). Our constant-velocity and constant-force data suggest partial or near-complete folding of the N domain in misfolded states M2 and M1, respectively, and indicate that both the EF2 and EF4 Ca^{2+}-binding sites play important roles in misfolding. The ability of EF2 and EF4 to bind calcium appears to be a prerequisite for wild-type NCS-1 to populate M1 and M2. If either of the two EF hands is disabled, NCS-1 behaves differently. Subsequent to calcium binding to EF3, the EF2 and EF4 sites may compete both for calcium and for successive interaction with a folded EF3. This slows productive folding because native folding requires EF3 to interact with a folded EF4 and hence any nonnative EF2–EF3 contacts need consequently to be broken. We have previously shown how the crosstalk between different EF hands is important for the folding process of NCS-1 (39). Here we report data that highlight the importance of EF-hand crosstalk for the misfolding of NCS-1.

We observed NCS-1 misfolding even at low calcium levels, suggesting a physiological relevance. Intracellular Ca^{2+} levels are tightly regulated in neurons, maintained at 40–100 nM at resting levels and rising to about 1–10 μM under excited conditions (42). However, even under normal conditions, Ca^{2+} concentrations can transiently (seconds to minutes) reach hundreds of micromolars, especially in the vicinity of Ca^{2+} channels (42, 43). When calcium homeostasis fails, the resulting and sustained high concentrations of cytosolic Ca^{2+} can have severe consequences for neurons, including deleterious effects on synaptic function and eventually cell death (44). A hypothesis gaining increasing support in recent years postulates that aging is a result of Ca^{2+} dysregulation (56, 57). Many factors in aging neurons, such as increased release of Ca^{2+} from intracellular stores and increased Ca^{2+} influx through L-type voltage-gated calcium channels, lead to sustained elevation of free Ca^{2+} concentration (56, 58). Moreover, calcium dysregulation has been implicated in the development of major neural disorders such as Alzheimer's disease, BD, and schizophrenia, and drugs that reduce Ca^{2+} signaling activity have in some cases proved successful in alleviating symptoms (59, 60). Interestingly, NCS-1 is highly up-regulated in schizophrenic and BD patients, and substantial evidence suggests that resting and activated levels of Ca^{2+} are elevated in BD (61). Whether NCS-1 is up-regulated to counteract loss of function as a consequence of Ca^{2+}-induced increased misfolding is unknown but might be a highly interesting avenue of research.

Despite both Ca^{2+} homeostasis and protein misfolding being linked to neurodegeneration (5, 59), a direct causation between the two former has been missing. Here, using single-molecule optical tweezers, a putative link has been revealed between Ca^{2+} dysregulation, misfolding, and an NCS protein involved in neurodegenerative disorders. Thus, Ca^{2+} may not only shape the function of the NCS family but also cause its dysfunction in vivo.

Methods

The engineering, expression, and purification of the double-cysteine variants of unmyristoylated human NCS-1, the preparation of protein–DNA chimeric constructs, and the coupling of protein–DNA chimeras to polystyrene beads were performed exactly as described (39, 62). All experiments were performed using a custom-built optical tweezers instrument with a dual-beam laser trap of 840-nm wavelength. Measurements were conducted at ambient temperatures in 10 mM Tris, 250 mM NaCl at pH 7.0 with varying concentrations of CaCl$_2$ in the range 0.5 μM to 10 mM (*SI Methods*).

Changes in contour length associated with (un)folding transitions were estimated by fitting constant-velocity traces to the WLC model, as previously described (16, 39). The thermodynamics and kinetics of the transitions between molecular states I2, M1, and M2 in constant-force measurements were characterized using a HMM algorithm, as previously described (39). The force dependence of the transition rates was analyzed with the phenomenological Bell model (63) to estimate the position of the barriers and the barriers' heights, using a preexponential factor of 1.2×10^{-4} Hz (39, 49). A total of 84 and 62 extension vs. time traces, from roughly 20 and 10 molecules, were used in our energy landscape reconstruction at 10 mM and 10 μM Ca^{2+}, respectively (*SI Methods*).

ACKNOWLEDGMENTS. We thank Dr. P. T. X. Li for critical comments on the manuscript. P.O.H. and B.B.K. acknowledge the Carlsberg Foundation and the Lundbeck Foundation for financial support. C.C. gratefully acknowledges financial support from Fondazione Cassa di Risparmio di Modena, the European Union (EU) through Marie Curie International Re-Integration Grant 44952, the Italian Ministry of Education, University, and Research (MIUR) Grant 17DPXLNBEK, and partial support from Italian MIUR Basic Research Investment Fund (FIRB) RBPR05JH2P "ITALNANONET." M.R.O. gratefully acknowledges financial support from the Abdus Salam International Centre for Theoretical Physics, of United Nations Educational, Scientific and Cultural Organization (UNESCO), and the International Atomic Energy Agency (IAEA), Programme for Training and Research in Italian Laboratories (TRIL).

1. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75:333–366.
2. Zheng W, Schafer NP, Wolynes PG (2013) Frustration in the energy landscapes of multidomain protein misfolding. *Proc Natl Acad Sci USA* 110(5):1680–1685.
3. Borgia MB, et al. (2011) Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature* 474(7353):662–665.
4. Hartl FU, Bracher A, Hayer-Hartl M (2011) Molecular chaperones in protein folding and proteostasis. *Nature* 475(7356):324–332.
5. Soto C (2003) Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat Rev Neurosci* 4(1):49–60.
6. Pastore A, Temussi P (2012) Protein aggregation and misfolding: Good or evil? *J Phys Condens Matter* 24(24):244101.
7. Teilum K, et al. (2009) Transient structural distortion of metal-free Cu/Zn superoxide dismutase triggers aberrant oligomerization. *Proc Natl Acad Sci USA* 106(43):18273–18278.
8. Chiti F, Dobson CM (2009) Amyloid formation by globular proteins under native conditions. *Nat Chem Biol* 5(1):15–22.
9. Danielsson J, et al. (2013) Global structural motions from the strain of a single hydrogen bond. *Proc Natl Acad Sci USA* 110(10):3829–3834.
10. Eichner T, Kalverda AP, Thompson GS, Homans SW, Radford SE (2011) Conformational conversion during amyloid formation at atomic resolution. *Mol Cell* 41(2):161–172.
11. DeMarco ML, Daggett V (2004) From conversion to aggregation: Protofibril formation of the prion protein. *Proc Natl Acad Sci USA* 101(8):2293–2298.
12. Calamai M, Chiti F, Dobson CM (2005) Amyloid fibril formation can proceed from different conformations of a partially unfolded protein. *Biophys J* 89(6):4201–4210.
13. Gianni S, et al. (2010) Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. *Nat Struct Mol Biol* 17(12):1431–1437.
14. Zoldák G, Rief M (2013) Force as a single molecule probe of multidimensional protein energy landscapes. *Curr Opin Struct Biol* 23(1):48–57.
15. Moffitt JR, Chemla YR, Smith SB, Bustamante C (2008) Recent advances in optical tweezers. *Annu Rev Biochem* 77:205–228.
16. Cecconi C, Shank EA, Bustamante C, Marqusee S (2005) Direct observation of the three-state folding of a single protein molecule. *Science* 309(5743):2057–2060.
17. Heidarsson PO, Naqvi MM, Sonar P, Valpapuram I, Cecconi C (2013) Conformational dynamics of single protein molecules studied by direct mechanical manipulation. *Advances in Protein Chemistry and Structural Biology*, ed Tatyana K-C (Academic, New York), Vol 92, pp 93–133.
18. Schlierf M, Yew ZT, Rief M, Paci E (2010) Complex unfolding kinetics of single-domain proteins in the presence of force. *Biophys J* 99(5):1620–1627.
19. Mashaghi A, et al. (2013) Reshaping of the conformational search of a protein by the chaperone trigger factor. *Nature* 500(7460):98–101.
20. Yu H, et al. (2012) Direct observation of multiple misfolding pathways in a single prion protein molecule. *Proc Natl Acad Sci USA* 109(14):5283–5288.
21. Stigler J, Ziegler F, Gieseke A, Gebhardt JC, Rief M (2011) The complex folding network of single calmodulin molecules. *Science* 334(6055):512–516.
22. Xi Z, Gao Y, Sirinakis G, Guo H, Zhang Y (2012) Single-molecule observation of helix staggering, sliding, and coiled coil misfolding. *Proc Natl Acad Sci USA* 109(15):5711–5716.
23. Chazin WJ (2011) Relating form and function of EF-hand calcium binding proteins. *Acc Chem Res* 44(3):171–179.
24. Weiss JL, Hui H, Burgoyne RD (2010) Neuronal calcium sensor-1 regulation of calcium channels, secretion, and neuronal outgrowth. *Cell Mol Neurobiol* 30(8):1283–1292.
25. Reyes-Bermudez A, Miller DJ, Sprungala S (2012) The neuronal calcium sensor protein Acrocalcin: A potential target of calmodulin regulation during development in the coral *Acropora millepora*. *PLoS ONE* 7(12):e51689.
26. Gomez M, et al. (2001) Ca2+ signaling via the neuronal calcium sensor-1 regulates associative learning and memory in *C. elegans*. *Neuron* 30(1):241–248.
27. Saab BJ, et al. (2009) NCS-1 in the dentate gyrus promotes exploration, synaptic plasticity, and rapid acquisition of spatial memory. *Neuron* 63(5):643–656.
28. Pongs O, et al. (1993) Frequenin—A novel calcium-binding protein that modulates synaptic efficacy in the *Drosophila* nervous system. *Neuron* 11(1):15–28.
29. McFerran BW, Graham ME, Burgoyne RD (1998) Neuronal Ca2+ sensor 1, the mammalian homologue of frequenin, is expressed in chromaffin and PC12 cells and regulates neurosecretion from dense-core granules. *J Biol Chem* 273(35):22768–22772.
30. Navarro G, et al. (2012) NCS-1 associates with adenosine A(2A) receptors and modulates receptor function. *Front Mol Neurosci* 5(5):53.
31. Kabbani N, Negyessy L, Lin R, Goldman-Rakic P, Levenson R (2002) Interaction with neuronal calcium sensor NCS-1 mediates desensitization of the D2 dopamine receptor. *J Neurosci* 22(19):8476–8486.
32. Koh PO, et al. (2003) Up-regulation of neuronal calcium sensor-1 (NCS-1) in the prefrontal cortex of schizophrenic and bipolar patients. *Proc Natl Acad Sci USA* 100(1):313–317.
33. Piton A, et al.; S2D team (2008) Mutations in the calcium-related gene IL1RAPL1 are associated with autism. *Hum Mol Genet* 17(24):3965–3974.
34. Handley MT, Lian LY, Haynes LP, Burgoyne RD (2010) Structural and functional deficits in a neuronal calcium sensor-1 mutant identified in a case of autistic spectrum disorder. *PLoS ONE* 5(5):e10534.
35. Aravind P, et al. (2008) Regulatory and structural EF-hand motifs of neuronal calcium sensor-1: Mg 2+ modulates Ca 2+ binding, Ca 2+ -induced conformational changes, and equilibrium unfolding transitions. *J Mol Biol* 376(4):1100–1115.
36. Stigler J, Rief M (2012) Calcium-dependent folding of single calmodulin molecules. *Proc Natl Acad Sci USA* 109(44):17814–17819.
37. Mukherjee S, Mohan PM, Kuchroo K, Chary KV (2007) Energetics of the native energy landscape of a two-domain calcium sensor protein: Distinct folding features of the two domains. *Biochemistry* 46(35):9911–9919.
38. Heidarsson PO, et al. (2012) The C-terminal tail of human neuronal calcium sensor 1 regulates the conformational stability of the Ca^{2+}- activated state. *J Mol Biol* 417(1-2):51–64.
39. Heidarsson PO, et al. (2013) Single-molecule folding mechanism of an EF-hand neuronal calcium sensor. *Structure* 21(10):1812–1821.
40. Heidarsson PO, et al. (2012) A highly compliant protein native state with a spontaneous-like mechanical unfolding pathway. *J Am Chem Soc* 134(41):17068–17075.
41. Bustamante C, Marko JF, Siggia ED, Smith S (1994) Entropic elasticity of lambda-phage DNA. *Science* 265(5178):1599–1600.
42. Augustine GJ, Santamaria F, Tanaka K (2003) Local calcium signaling in neurons. *Neuron* 40(2):331–346.
43. Burgoyne RD (2007) Neuronal calcium sensor proteins: Generating diversity in neuronal Ca2+ signalling. *Nat Rev Neurosci* 8(3):182–193.
44. Seaton G, Hogg EL, Jo J, Whitcomb DJ, Cho K (2011) Sensing change: The emerging role of calcium sensors in neuronal disease. *Semin Cell Dev Biol* 22(5):530–535.
45. Chodera JD, et al. (2011) Bayesian hidden Markov model analysis of single-molecule force spectroscopy: Characterizing kinetics under measurement uncertainty. arXiv:1108:1430.
46. McKinney SA, Joo C, Ha T (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys J* 91(5):1941–1951.
47. Gao Y, Sirinakis G, Zhang Y (2011) Highly anisotropic stability and folding kinetics of a single coiled coil protein under mechanical tension. *J Am Chem Soc* 133(32):12749–12757.
48. Bell GI (1978) Models for the specific adhesion of cells to cells. *Science* 200(4342):618–627.
49. Gebhardt JC, Bornschlögl T, Rief M (2010) Full distance-resolved folding energy landscape of one single protein molecule. *Proc Natl Acad Sci USA* 107(5):2013–2018.
50. Cremades N, et al. (2012) Direct observation of the interconversion of normal and toxic forms of α-synuclein. *Cell* 149(5):1048–1059.
51. Dobson CM (2003) Protein folding and misfolding. *Nature* 426(6968):884–890.
52. Li PT, Bustamante C, Tinoco I, Jr (2007) Real-time control of the energy landscape by force directs the folding of RNA molecules. *Proc Natl Acad Sci USA* 104(17):7039–7044.
53. Shank EA, Cecconi C, Dill JW, Marqusee S, Bustamante C (2010) The folding cooperativity of a protein is controlled by its chain topology. *Nature* 465(7298):637–640.
54. Linse S, Linse B (2007) Protein folding through kinetic discrimination. *J Am Chem Soc* 129(27):8481–8486.
55. Chattopadhyaya R, Meador WE, Means AR, Quiocho FA (1992) Calmodulin structure refined at 1.7 Å resolution. *J Mol Biol* 228(4):1177–1192.
56. Oliveira AM, Bading H (2011) Calcium signaling in cognition and aging-dependent cognitive decline. *Biofactors* 37(3):168–174.
57. Khachaturian ZS (1989) The role of calcium regulation in brain aging: Reexamination of a hypothesis. *Aging (Milano)* 1(1):17–34.
58. Toescu EC, Vreugdenhil M (2010) Calcium and normal brain ageing. *Cell Calcium* 47(2):158–164.
59. Berridge MJ (2012) Calcium signalling remodelling and disease. *Biochem Soc Trans* 40(2):297–309.
60. Berridge MJ (2013) Dysregulation of neural calcium signaling in Alzheimer disease, bipolar disorder and schizophrenia. *Prion* 7(1):2–13.
61. Warsh JJ, Andreopoulos S, Li PP (2004) Role of intracellular calcium signaling in the pathophysiology and pharmacotherapy of bipolar disorder: Current status. *Clin Neurosci Res* 4(3-4):201–213.
62. Cecconi C, Shank EA, Dahlquist FW, Marqusee S, Bustamante C (2008) Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers. *Eur Biophys J* 37(6):729–738.
63. Liphardt J, Onoa B, Smith SB, Tinoco I, Jr, Bustamante C (2001) Reversible unfolding of single RNA molecules by mechanical force. *Science* 292(5517):733–737.

Supporting Information

Heidarsson et al. 10.1073/pnas.1401065111

SI Methods

Experimental Procedures. The double-cysteine construct of NCS-1, placing cysteine residues at positions 4 and 188, was engineered using a pseudowild-type pET-16b expression plasmid (with Cys38 replaced by serine), by standard genetic techniques. This variant has previously been shown to have comparable stability, structure, and calcium-binding properties to the wild-type protein (1). The *Escherichia coli* strain BL21 (DE3) was used to express unmyristoylated human NCS-1 and was grown at 37 °C in Luria–Bertani medium.

Individual proteins were manipulated by means of ~500 bp DNA molecules covalently attached to exposed cysteine residues through a disulfide bond. One DNA handle bears a 5′ digoxigenin moiety that interacts with an optically trapped 3.1-μm polystyrene bead coated with anti-digoxigenin antibodies (Spherotec), while the other handle bears a 5′ biotin moiety that interacts with a 2.18-μm streptavidin-coated bead (Spherotech) held in place at the end of a pipette by suction (2). The force applied to the molecule was varied by moving the micropipette toward or away from the optical trap by means of a piezoelectric flexure stage (MAX311/M, Thorlabs). The applied force was determined by measuring the change in light momentum of the laser beams leaving the trap, whereas changes in the extension of the molecule were determined by measuring the distance between the two beads (3). Force vs. extension traces were collected at constant speeds from 5 to 1,000 nm·s^{-1}. During constant-velocity experiments, data were recorded at a rate of 40 Hz. In constant-force experiments, the force was kept constant through a force-feedback mechanism, where the average force was measured and compared with the set-point force value every 1 ms. Differences between forces were compensated for by movement of the micropipette. In constant-force experiments, data were recorded at a rate of 100 Hz.

Hidden Markov Model. Hidden Markov model (HMM) algorithms have recently replaced threshold-based methods as the tool of choice to study the time series obtained in single-molecule experiments (4). Originally developed during the 1960s by mathematicians and engineers interested in artificial intelligence, HMM is based upon the idea of modeling the system of interest as a Markov chain that is hidden in the sense that the state of the system is not directly accessible: Each state, however, emits a signal according to a discrete or continuous distribution. The (time-independent) transition matrix characterizing the Markov process as well as the parameters of the signal emission distributions may be inferred from a sequence of observations by applying an iterative procedure introduced by Baum and Petrie (5). We refer to the excellent tutorial written by Rabiner (6) for theoretical and practical details about this statistical method. Here we just outline the various passages from the data to the kinetic and thermodynamic parameters reported in Table 1.

Each extension vs. time trace is individually analyzed by means of a four-state (I2, M1, M2, N) HMM routine where we assume that the end-to-end extension associated with each state is normally distributed. The logarithm of the reconstructed transition probability matrix, divided by the time interval between two consecutive measurements, yields the transition rates among the states of our Markovian model. By means of the simplified Kramers–Bell theory derived, for instance, in ref. 7, we can estimate from the force dependence of the rates the position of the barriers and the free energy of the states. An estimation of the barriers' heights requires the knowledge of the preexponential factor, which we cannot measure directly with our experimental apparatus. We adopt, therefore, the value 1.2×10^{-4} s^{-1}, which has been measured in an experiment with a similar setup as ours, albeit using a different protein (8).

1. Heidarsson PO, et al. (2013) Single-molecule folding mechanism of an EF-hand neuronal calcium sensor. *Structure* 21(10):1812–1821.
2. Cecconi C, Shank EA, Dahlquist FW, Marqusee S, Bustamante C (2008) Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers. *Eur Biophys J* 37(6):729–738.
3. Smith SB, Cui Y, Bustamante C (2003) Optical-trap force transducer that operates by direct measurement of light momentum. *Methods Enzymol* 361:134–162.
4. McKinney SA, Joo C, Ha T (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys J* 91(5):1941–1951.
5. Baum LE, Petrie T (1966) Statistical inference for probabilistic functions of finite state Markov chains. *Ann Math Stat* 37(6):1554–1563.
6. Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE* 77(2):257–286.
7. Mossa A, Manosas M, Forns N, Huguet JM, Ritort F (2009) Dynamic force spectroscopy of DNA hairpins: I. Force kinetics and free energy landscapes. *J Stat Mech* 2009(2): P02060.
8. Gebhardt JCM, Bornschlögl T, Rief M (2010) Full distance-resolved folding energy landscape of one single protein molecule. *Proc Natl Acad Sci USA* 107(5):2013–2018.

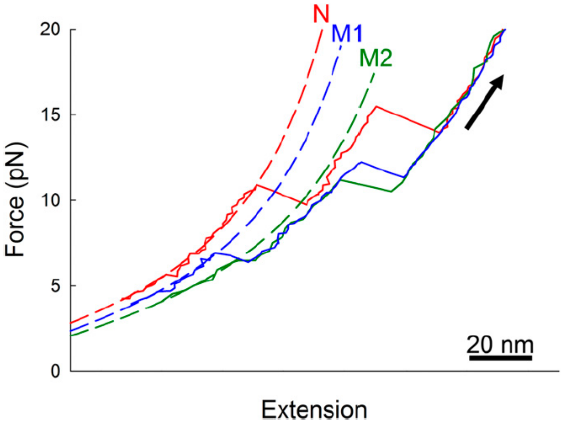


Fig. S1. Unfolding trajectories of different NCS-1 molecular states. The native state (N) of NCS-1 unfolds through two high-force transitions corresponding to the mechanical denaturation of the N and C domains, respectively (red trace). In contrast, misfolded states M1 (blue) and M2 (green) display drastically reduced mechanical stability. Dashed lines are worm-like-chain fits to the data. The arrow indicates the pulling direction.

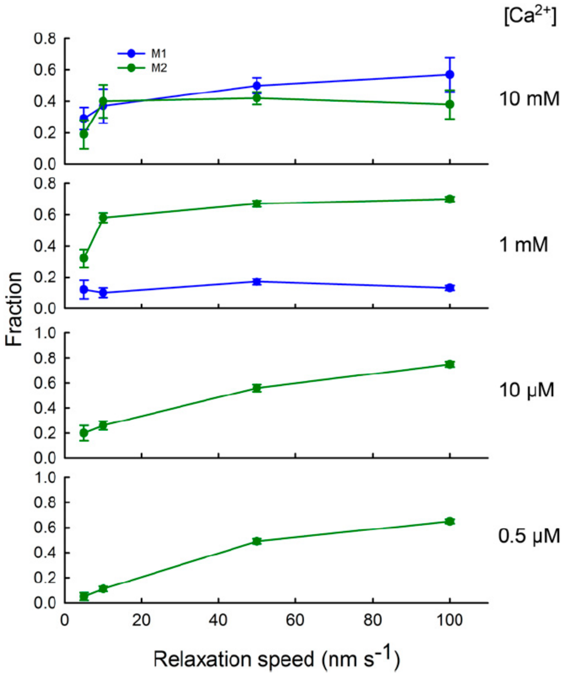


Fig. S2. Occupation probability of misfolded states M1 and M2 as a function of relaxation speed and Ca^{2+} concentration. The M1 state is observed only at high Ca^{2+} concentrations. The M2 state is observed at all experimental conditions.

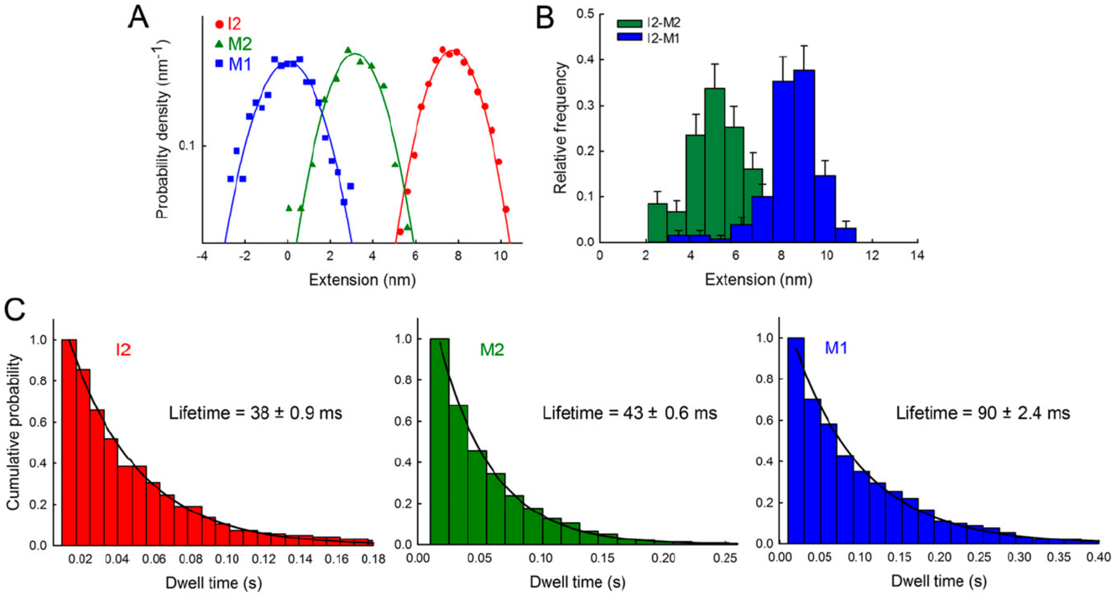


Fig. S3. Extension and lifetimes of NCS-1 misfolded states at 10 mM Ca^{2+}. (*A*) Probability density distributions of the extension of the molecular states I2 (red), M1 (blue), and M2 (green) identified by the HMM algorithm in an extension vs. time trace acquired at 5.9 pN. The distributions were fitted to Gaussian functions, which estimated changes in extension of 4.5 ± 1.8 nm for the I2–M2 transition and of 7.7 ± 1.8 nm for the I2–M1 transition. (*B*) Distribution of the changes in extension for the I2–M2 and I2–M1 transition, calculated using all extension vs. time traces. The medians of these distributions are 5.1 ± 2 nm and 8.6 ± 2 nm, respectively. (*C*) Dwell time distributions of I2 (red), M2 (green), and M1 (blue) at 6.2 pN. The distributions could be well fitted to single-exponential functions.

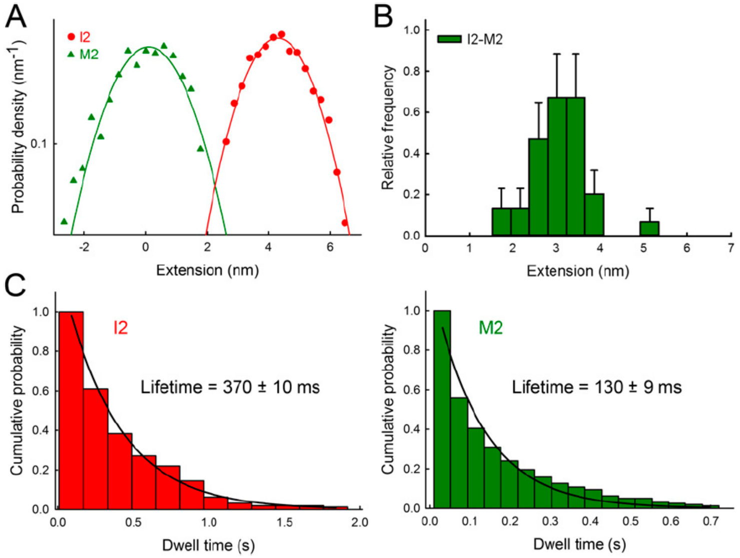


Fig. S4. Extension and lifetimes of NCS-1 misfolded states at 10 μM Ca^{2+}. (*A*) Probability density distributions of the extension of the molecular states I2 (red) and M2 (green) identified by HMM analysis in an extension vs. time trace acquired at 4.7 pN. The distributions were fitted to Gaussian functions, which estimated a change in extension of 4.2 ± 1.6 nm for the I2–M2 transition. (*B*) Distribution of the changes in extension for the I2–M2 transition, calculated using all extension vs. time traces. The median of this distribution is 3.3 ± 1.4 nm. (*C*) Dwell time distributions of I2 (red) and M2 (green) at 3.7 pN. The distributions could be well fitted to single-exponential functions.

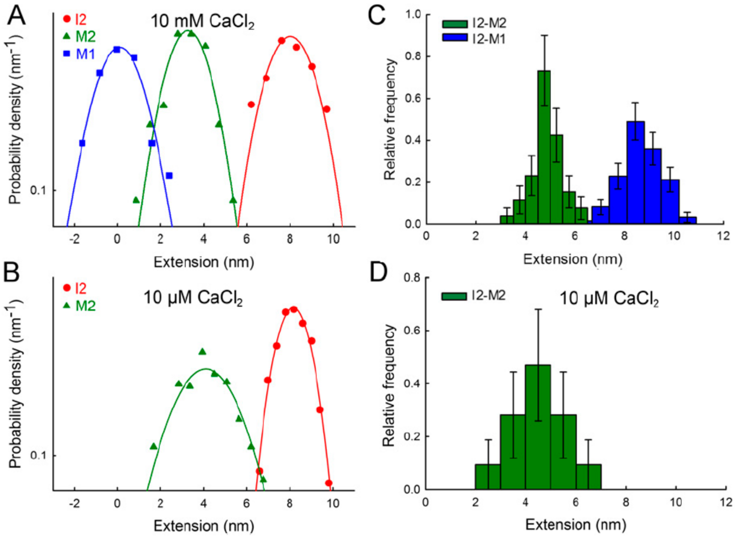


Fig. S5. Comparative extension analysis of NCS-1 misfolded states in the same force range at high and low Ca^{2+}. *A* and *B* show probability density distributions of the extensions of the molecular states I2 (red), M1 (blue), and M2 (green) identified by HMM analysis in extension vs. time traces acquired at 5.8 pN in either 10 mM CaCl$_2$ (*A*) or 10 µM CaCl$_2$ (*B*). The distributions were fitted to Gaussian functions, which estimated a change in extension of 4.8 ± 2.0 nm and 4.2 ± 1.8 nm for the I2–M2 transition at 10 mM and 10 µM CaCl$_2$, respectively, and 8.0 ± 2.0 nm for the I2–M1 transition at 10 mM CaCl$_2$. *C* and *D* show distribution of the changes in extension for the I2–M1 and I2–M2 transitions, calculated using extension vs. time traces in the force range 5.4–6.2 pN at both 10 mM and 10 µM CaCl$_2$. The median of the distribution for the I2–M2 transition is 4.8 ± 1.0 nm and 4.5 ± 1.0 nm at 10 mM and 10 µM CaCl$_2$, respectively, and that of the distribution for the I2–M1 transition is 8.6 ± 2.0 nm at 10 mM CaCl$_2$.

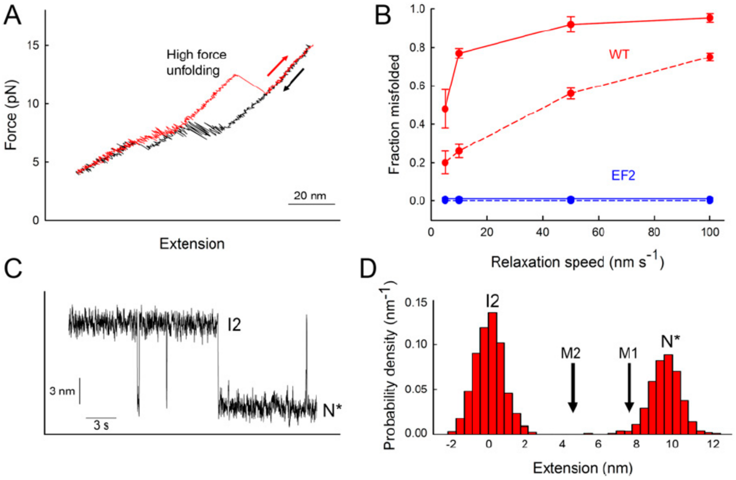


Fig. S6. Folding behavior of the NCS-1^{EF2} and NCS-1^{EF4} variants. (*A*) Force vs. extension trace for NCS-1^{EF2} at 10 mM CaCl$_2$ and 100 nm·s^{-1}. In this variant, the C domain always folds correctly as evident from the high-force unfolding transition (average unfolding force of 13.7 ± 1.0 pN), signifying a natively folded domain (1). Color-coded arrows indicate pulling direction. (*B*) Fraction of relaxation traces in which the C domain misfolds in constant-velocity experiments as a function of relaxation speed for NCS-1 (red) and NCS-1^{EF2} (blue). Data acquired in both 10 mM (solid lines; 230 events, six molecules) and 10 µM (dashed lines; 180 events, three molecules) CaCl$_2$ are shown. The C domain of the NCS-1^{EF2} variant does not misfold at any calcium concentration or relaxation speed. Data for WT NCS-1 is from Fig. 1. (*C*) Extension vs. time traces for NCS-1^{EF4} at 5.4 pN and 10 mM CaCl$_2$. The molecule shows only fluctuations between I2 and a nonnative folded state (N*) with a change in extension that is significantly larger than those associated with the M1 and M2 states (five molecules). Very similar results were obtained at 10 µM CaCl$_2$ (four molecules). (*D*) Probability density distribution of the extension values in the extension-time trace in *C*. Two main populations cluster around 0 nm and ~9.5 nm. The expected average positions of the two misfolded states are indicated.

PAPER II

jbc The Journal of Biological Chemistry
AFFINITY SITES
ASBMB
American Society for Biochemistry and Molecular Biology

Molecular Bases of Disease:

Structure, Folding Dynamics, and Amyloidogenesis of D76N β_2-Microglobulin: ROLES OF SHEAR FLOW, HYDROPHOBIC SURFACES, AND α-CRYSTALLIN

MOLECULAR BASES OF DISEASE
PROTEIN STRUCTURE AND FOLDING

P. Patrizia Mangione, Gennaro Esposito, Annalisa Relini, Sara Raimondi, Riccardo Porcari, Sofia Giorgetti, Alessandra Corazza, Federico Fogolari, Amanda Penco, Yuji Goto, Young-Ho Lee, Hisashi Yagi, Ciro Cecconi, Mohsin M. Naqvi, Julian D. Gillmore, Philip N. Hawkins, Fabrizio Chiti, Ranieri Rolandi, Graham W. Taylor, Mark B. Pepys, Monica Stoppini and Vittorio Bellotti

J. Biol. Chem. 2013, 288:30917-30930.
doi: 10.1074/jbc.M113.498857 originally published online September 6, 2013

⌘ *Author's Choice*

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 288, NO. 43, pp. 30917–30930, October 25, 2013
© 2013 by The American Society for Biochemistry and Molecular Biology, Inc. Published in the U.S.A.

Structure, Folding Dynamics, and Amyloidogenesis of D76N β_2-Microglobulin

ROLES OF SHEAR FLOW, HYDROPHOBIC SURFACES, AND α-CRYSTALLIN[*]

Received for publication, July 5, 2013, and in revised form, August 28, 2013 Published, JBC Papers in Press, September 6, 2013, DOI 10.1074/jbc.M113.498857

P. Patrizia Mangione[‡,§], Gennaro Esposito[¶], Annalisa Relini[‖], Sara Raimondi[§], Riccardo Porcari[‡], Sofia Giorgetti[§], Alessandra Corazza[¶], Federico Fogolari[¶], Amanda Penco[‖], Yuji Goto[**], Young-Ho Lee[**], Hisashi Yagi[**], Ciro Cecconi[‡‡], Mohsin M. Naqvi[§§], Julian D. Gillmore[‡], Philip N. Hawkins[‡], Fabrizio Chiti[¶¶], Ranieri Rolandi[‖], Graham W. Taylor[‡], Mark B. Pepys[‡], Monica Stoppini[§], and Vittorio Bellotti[‡,§,1]

From the [‡]Wolfson Drug Discovery Unit, Centre for Amyloidosis and Acute Phase Proteins, Division of Medicine, University College London, London NW3 2PF, United Kingdom, the [§]Department of Molecular Medicine, Institute of Biochemistry, University of Pavia, 27100 Pavia, Italy, the [¶]Department of Medical and Biological Sciences, University of Udine, 33100 Udine, Italy, the [‖]Department of Physics, University of Genoa, 16146 Genoa, Italy, the [**]Institute for Protein Research, Osaka University, Osaka 565-0871, Japan, the [‡‡]Consiglio Nazionale delle Ricerche-Nanoscience Institute S3, 41125 Modena, Italy, the [§§]Department of Physics, University of Modena and Reggio Emilia, 41125 Modena, Italy, and [¶¶]Department of Biochemical Sciences, University of Florence, 50134 Florence, Italy

Background: We recently discovered the first natural human β_2-microglobulin variant, D76N, as an amyloidogenic protein.
Results: Fluid flow on hydrophobic surfaces triggers its amyloid fibrillogenesis. The α-crystallin chaperone inhibits variant-mediated co-aggregation of wild type β_2-microglobulin.
Conclusion: These mechanisms likely reflect *in vivo* amyloidogenesis by globular proteins in general.
Significance: Our results elucidate the molecular pathophysiology of amyloid deposition.

Systemic amyloidosis is a fatal disease caused by misfolding of native globular proteins, which then aggregate extracellularly as insoluble fibrils, damaging the structure and function of affected organs. The formation of amyloid fibrils *in vivo* is poorly understood. We recently identified the first naturally occurring structural variant, D76N, of human β_2-microglobulin (β_2m), the ubiquitous light chain of class I major histocompatibility antigens, as the amyloid fibril protein in a family with a new phenotype of late onset fatal hereditary systemic amyloidosis. Here we show that, uniquely, D76N β_2m readily forms amyloid fibrils *in vitro* under physiological extracellular conditions. The globular native fold transition to the fibrillar state is primed by exposure to a hydrophobic-hydrophilic interface under physiological intensity shear flow. Wild type β_2m is recruited by the variant into amyloid fibrils *in vitro* but is absent from amyloid deposited *in vivo*. This may be because, as we show here, such recruitment is inhibited by chaperone activity. Our results suggest general mechanistic principles of *in vivo* amyloid fibrillogenesis by globular proteins, a previously obscure process. Elucidation of this crucial causative event in clinical amyloidosis should also help to explain the hitherto mysterious timing and location of amyloid deposition.

β_2m[2] (mass, 11,729 Da), the invariant light chain of the human HLA class I complex, is produced at ~200 mg/day in adults and is cleared only via the kidney. In patients with end stage renal failure on dialysis, the plasma concentration of β_2m therefore rises from the normal 1–2 mg/liter to persistently raised values of ~ 50–70 mg/liter, leading to the serious and intractable condition of dialysis-related amyloidosis with β_2m amyloid fibrils deposited in bones and joints, causing painful arthropathy, bone cysts, pathological fractures, and rarely visceral β_2m amyloid deposits. The normal structure and function of β_2m are well characterized, and although wild type β_2m is poorly amyloidogenic *in vitro*, its fibrillogenesis and its tissue-specific deposition have been intensively investigated (1). Despite much progress, there is neither general agreement about the underlying molecular mechanisms nor an understanding of the forces involved *in vivo* during the destabilization and subsequent amyloid aggregation of either β_2m or any of the other natively folded globular proteins that form amyloid fibrils

---

[*] This work was supported by United Kingdom Medical Research Council Grant MR/K000187/1 (to V. B.), the University College London (UCL) Amyloidosis Research Fund and UCL Wolfson Drug Discovery Unit Funds, Cariplo Foundation Projects 2009-2543 and 2011-2096 (to V. B.), Italian Ministry of University and Research Project FIRB RBFR109EOS (to S. G.), Regione Lombardia (to V. B.), and the International Collaborative Research Program of the Institute for Protein Research, Osaka University (to R. P.). Core support for the Wolfson Drug Discovery Unit is provided by the United Kingdom National Institute for Health Research Biomedical Research Centre and Unit Funding Scheme.

⌘ *Author's Choice*—Final version full access.

[1] To whom correspondence should be addressed: Wolfson Drug Discovery Unit, Centre for Amyloidosis and Acute Phase Proteins, Division of Medicine, University College London, Rowland Hill St., London NW3 2PF, UK. Tel.: 44-20-7433-2773; Fax: 44-20-7433-2803; E-mail: v.bellotti@ucl.ac.uk or vbellot@unipv.it.

[2] The abbreviations used are: β_2m, β_2-microglobulin; ΔG, free energy change; U, unfolded state; I, intermediate; N, native state; subscript C, *cis*-His^{31}-Pro^{32}; subscript T, *trans*-His^{31}-Pro^{32}; TS, transition state; $\Delta G^{‡}$, free energy barrier; t_m, mixing time for TOCSY, NOESY, and ROESY; BLUU-Tramp, Biophysics Laboratory University of Udine temperature ramp; Gdn-HCl, guanidine hydrochloride; ThT, thioflavin T; AFM, atomic force microscopy; ΔN6 β_2m, truncated β_2m isoform lacking the 6 N-terminal residues; TOCSY, total correlation spectroscopy; ROESY, rotating frame nuclear Overhauser enhancement spectroscopy.

in disease. We lately reported (2) the first naturally occurring structural variant of β_2m, D76N, discovered in members of a French family who developed progressive bowel dysfunction with extensive visceral β_2m amyloid deposits despite normal renal function and normal circulating β_2m concentrations and with none of the osteoarticular deposits characteristic of dialysis-related amyloidosis. Here we elucidate in detail the biophysical parameters of amyloid fibrillogenesis by this uniquely tractable protein and develop a paradigm that could be applicable generally to the *in vivo* pathophysiology of amyloidogenesis by the whole range of globular proteins that cause clinically significant systemic amyloidosis.

EXPERIMENTAL PROCEDURES

Production of Recombinant Proteins—Recombinant wild type and variant β_2m were expressed and purified as described previously (2).

Differential Scanning Calorimetry—Differential scanning calorimetry experiments were carried out with a VP-DSC instrument (MicroCal, Northampton, MA) with protein at 0.5 mg/ml in 25 mM sodium phosphate buffer, pH 7.4 and scans from 10 to 90 °C at a scanning rate of 60 °C/h. The reversibility of protein denaturation was assessed by repeating heating and cooling cycles. After normalization and base-line subtraction, the thermal unfolding curves were analyzed using MicroCal Origin 7.0 software with a two-state unfolding model.

Equilibrium Denaturation Experiments and Folding Kinetics—Guanidine hydrochloride (Gdn-HCl) equilibrium denaturation, unfolding, and refolding kinetics were performed as described previously (3). All experiments were carried out at 30 °C in 20 mM sodium phosphate buffer, pH 7.4 at a 0.02 mg/ml final protein concentration. Refolding of acid-denatured protein and double jump experiments were performed at 4 °C as described previously (4).

Energy Diagram—All free energy changes (ΔG) were determined in J mol^{-1} and then converted into kcal mol^{-1}; throughout we use the following abbreviations: U, unfolded state; I, intermediate; N, native state; subscript C, *cis*-His^{31}-Pro^{32}; subscript T, *trans*-His^{31}-Pro^{32}; TS, transition state. The U_T state was arbitrarily given a free energy (G) of 0 J mol^{-1} and was considered as a reference for all reported ΔG values. The ΔG from the U_T to the N_C states was determined from Gdn-HCl unfolding equilibrium curves as reported (3). The ΔG from the N_T to the N_C states was determined using $\Delta G = -RT\ln(k_u/k_{slow})$ where R is the universal gas constant, T is the absolute temperature, and k_u and k_{slow} are the rate constants (in s^{-1} units) for unfolding and for the slow phase of folding, respectively, extrapolated to the absence of Gdn-HCl. The ΔG from the I_T to the U_T states was determined by plotting the fluorescence of the I_T state (corresponding to the fluorescence at time 0 of a kinetic trace of folding) against Gdn-HCl concentration and by plotting the fluorescence of the U_T state against Gdn-HCl concentration (in the latter case, the values at low Gdn-HCl concentration were obtained by linear extrapolation from the values at high Gdn-HCl concentration). The fluorescence of the I_T state decreased with increasing Gdn-HCl concentration until it approached the fluorescence of the U_T state, thus providing an approximate measure of the conformational stability of the I_T state relative to U_T. The ΔG from the I_T to the TS_2 state was determined using $\Delta G^{\ddagger} = -RT\ln(k_{fast}/k^{\ddagger})$ where $\Delta G^{\ddagger}$ is the free energy barrier, k_{fast} is the rate constant for the fast phase of folding, and $k^{\ddagger}$ is the pre-exponential term taken as $4.8\ 10^8$ s^{-1} as reported (5, 6). Similarly, the ΔG values from N_T to TS_3 and from N_C to TS_3 were determined using $\Delta G^{\ddagger} = -RT\ln(k_{slow}/k^{\ddagger})$ and $\Delta G^{\ddagger} = -RT\ln(k_u/k^{\ddagger})$, respectively, where k_{slow} and k_u are the rate constants for the slow phase of folding and for unfolding, respectively. The ΔG values from U_T to TS_1 and from I_T to TS_1 were not determined. All other ΔG values not explicitly mentioned in the study can be determined by arithmetic linear combination of the ΔG parameters described above.

NMR Measurements—NMR spectra were obtained at 500.13 MHz with a Bruker Avance 500 spectrometer on 0.1–0.3 mM protein samples dissolved in H_2O/D_2O 90:10 or 95:5 with 20–70 mM sodium phosphate buffer and pH* (pH meter reading without isotope effect correction) in the range 6.6–7.2. Unlabeled and uniformly ^{15}N- or ^{15}N,^{13}C-labeled protein samples, expressed as described previously (2), were used. The spectra were collected mostly at 25 °C with only a few experiments obtained also at 30 or 37 °C. Homonuclear two-dimensional TOCSY (7), NOESY (8), and ROESY (9) spectra were acquired. The adopted experimental schemes included solvent suppression by WATERGATE (10) and excitation sculpting (11); 1-s steady state recovery time; mixing times (t_m) of 40–50 ms for TOCSY, 100–150 ms for NOESY, and 100 ms for ROESY; *t*1 quadrature detection by time-proportional phase incrementation (12); and gradient-assisted coherence selection (echo-antiecho) (13). The spin-lock mixing of the TOCSY and ROESY experiments was obtained with MLEV17 (14) pulse trains or single long pulse, respectively, at $\gamma B_2/2\pi$ ~10 kHz (TOCSY) and ~ 2.5 or 5 kHz (ROESY). The acquisitions were performed over a spectral width of 8012.82 Hz in both dimensions with matrix size of 1024–2048 points in *t*2 and 256–400 points in *t*1 and 32–64 scans per each *t*1 free induction decay. The BLUU-Tramp experiments were conducted using the procedure described previously over the temperature range 22–42 °C (15, 16). Measurements were performed on samples that had undergone complete deuterium substitution for hydrogen with two cycles of exchange at 4 °C in D_2O containing 10 mM NH_4HCO_3 and subsequent lyophilization. The solvent for the back-exchange was used for the preliminary shimming to enable quick start after dissolving the protein (~5 min dead time before starting the acquisition). The ^{15}N{^{1}H} NOE data were obtained at 25 and 37 °C by standard sequence using a 3-s relaxation interval. The spectra with (NOE) and without (no NOE) proton saturation were acquired in an interleaved manner.

Three-dimensional HNCA (17–19) and HNCOCA (19, 20) were typically acquired with 64 scans and 64 × 40 × 1024 data points in *t*1 (^{13}C), *t*2 (^{15}N), and *t*3 (^{1}H), respectively, over spectral windows of 40, 33.5, and 16 ppm for ^{13}C, ^{15}N, and ^{1}H, respectively. HNCO spectra (17, 19) were acquired using 128 × 40 × 1500 data points and only 32 scans for each *t*1 × *t*2 experiment over spectral windows of 22.1 (^{13}C), 33.5 (^{15}N), and 14 (^{1}H) ppm. Processing of three-dimensional data ended up with real matrices of 512 × 256 × 1024 points in F1, F2, and F3, respectively, except for the HNCO spectra where the carbon dimension (F1) was limited to 256 points. All data, except those from BLUU-Tramp, were processed with Topspin (Bruker Biospin) and analyzed with Sparky (T. D. Goddard and D. G. Kneller, University of California). BLUU-Tramp data were processed using NMRPipe and analyzed by NMRView (21).

Electrostatic Calculations—For the calculation of both surface potential and pK_a shifts, we used the recently developed program BLUUES (22) available also as a server utility (23). For the calculation of isopotential surfaces, we used the program UHBD, and we displayed the isopotential surfaces using the program VMD. To assess effects that could arise from slightly different arrangement in the structural models used for calculation, an alternative structure for the D76N variant was generated using the program SCWRL4.0 by alternative schemes: (i) only the side chain of the mutated residue is allowed to change conformation, and (ii) only the side chain of the mutated residues and contacting residues are allowed to change conformation. Despite numerical differences, the results from the homology-modeled structures are in agreement with the experimental data reported in the study, confirming that the effects are mainly due to the mutation rather than other minor conformational differences.

Molecular Dynamics Simulations—The force field used in the simulations was CHARMM v.27 (24) with the CMAP (two-dimensional dihedral energy grid correction map approach) correction (25). The minimized system was further relaxed, keeping the solute molecules (including ions) fixed, by molecular dynamics simulation. The system was heated to 47 °C in 2 ps, and a further 18-ps simulation was run to let water molecules reorient, consistent with the average lifetime of a hydrogen bond in water of 1–2 ps (26). The system without restraints on solute molecules was energy-minimized by 300 conjugate gradient minimization steps. The system was then heated to 47 °C in 2 ps, and a further 3.0-ns simulation was run to reach equilibrium. The simulations lasted 250 ns, and snapshots were saved every 0.1 ns. All molecular dynamics simulations were performed in the NPT ensemble using the Nosé-Hoover Langevin piston method (27, 28). The Langevin damping coefficient for temperature control was 10 ps^{-1}. For all simulations, the size of the box was fluctuating around its average value within fractions of Å.

H-bond Network Analysis—The mutation D76N within the β_2m sequence is likely to affect the molecular hydrogen bond network. To identify indirect effects of the mutation, 2500 snapshots (at 100-ps intervals) were taken from molecular dynamics simulations, and hydrogen bonds were listed using the program MOLMOL (29). Only hydrogen bonds involving at least one side chain group were selected to remove nonspecific effects. Each hydrogen bond was taken as representative of the proximity of the two involved residues, r_1 and r_2. In particular, a distance, $d_{r1,r2}$, was assigned based on the ratio between the number of snapshots where the hydrogen bond is present (n_s) and the total number of snapshots (n_{tot}).

$$d_{r1,r2} = -\log\left(\frac{n_s}{n_{tot}}\right) \quad \text{(Eq. 1)}$$

Once a set of pairwise distances had been assigned in this way, the shortest path between all nodes was found using the Floyd-Warshall algorithm, which outputs the path and the nodes along the path. In this way, the shortest (in the sense of most frequently observed) network of hydrogen bonds connecting different residues was identified. This definition has the obvious disadvantage of not considering whether hydrogen bonds are observed simultaneously or not. On the other hand, it takes advantage of being based only on pairwise connection, enabling the applicability of the fast Floyd-Warshall algorithm. The output of the program readily identifies stable or fluctuating networks of hydrogen bonds.

Fibrillogenesis—Fibrillogenesis experiments were performed in standard quartz cells stirred at 1500 rpm and 37 °C using 40 μM β_2m isoforms in 25 mM sodium phosphate, pH 7.4 containing 10 μM thioflavin T (ThT). Aggregation was carried out without seeds of preformed fibrils. ThT emission was monitored at 480 nm after excitation at 445 nm using a PerkinElmer Life Sciences LS 55 spectrofluorometer. Fibrillogenesis experiments were also conducted without agitation or in the absence of the air-water interface and with addition of 6 μM elastin isolated from human aorta (Sigma-Aldrich). β_2m, which remained soluble during fibrillogenesis experiments, was monitored by native gel electrophoresis (30). The soluble fractions were separated by centrifugation at 17,000 × *g* for 15 min before loading onto 1% agarose gel, and bands were quantified with Quantity One software (Bio-Rad). The effects of 1 and 40 μM α-crystallin (Sigma-Aldrich) on fibrillogenesis by an equimolar mixture of 40 μM wild type and D76N β_2m, respectively, were investigated, and soluble fractions of the two isoforms were quantified at 24 h by native agarose gel electrophoresis as described above.

Atomic Force Microscopy—Tapping mode AFM images were acquired in air using a Dimension 3000 scanning probe microscope equipped with a "G" scanning head (maximum scan size, 100 μm) and driven by a Nanoscope IIIa controller and a Multimode scanning probe microscope equipped with an "E" scanning head (maximum scan size, 10 μm) and driven by a Nanoscope V controller (Digital Instruments, Bruker). Single beam uncoated silicon cantilevers (type OMCL-AC160TS, Olympus) were used. The drive frequency was between 320 and 340 kHz; the scan rate was 0.5–2.0 Hz. All aggregation experiments were carried out with a 0.4 mg/ml protein concentration in 25 mM sodium phosphate, pH 7.4 at 37 °C with stirring at 750 rpm.

Aggregation Time Course by AFM—Aliquots (2 μl) were withdrawn at time 0, 1, 2, 8, and 24 h, respectively. After 500-fold dilution, 10 μl were finally deposited on freshly cleaved mica and dried under mild vacuum.

Effect of Graphite Sheets by AFM—Aggregation was conducted with and without agitation. Aliquots (2 μl) of samples under agitation were diluted 100-fold, and 10 μl were deposited on freshly cleaved mica and dried under mild vacuum. A graphite sheet was removed from the non-stirred protein solution, gently rinsed with purified (Milli-Q) water, fixed on a metallic disk, and dried under mild vacuum.

Ultrasonication—The effect of carbon nanotubes (0.01 mg/ml) on D76N β_2m fibrillogenesis was evaluated by a combination of a water bath type ultrasonicator and microplate

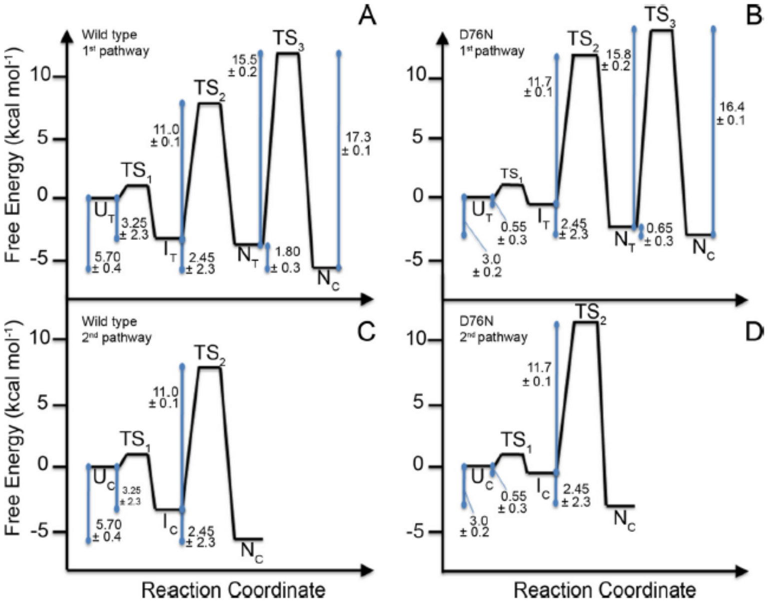


FIGURE 1. **Free energy diagram for wild type and variant β_2m on the two parallel pathways from the unfolded to the native state.** *A* and *B*, free energy values related to the slower pathways limited by the *trans* → *cis* isomerization of the His^{31}-Pro^{32} bond. *C* and *D*, free energy values associated with the rapid formation of native-like molecules via an intermediate containing the native isomer *cis*-His^{31}-Pro^{32} (I_C).

reader (HANABI, Elekon Science Co. Ltd. and Corona Electric Co., Japan). Protein at 40 μM in 25 mM sodium phosphate, pH 7.4, 10 μM ThT with or without nanotubes was aliquoted in a 96-well plate. Cycles of ultrasonication with simultaneous shaking (3 min) followed by 7-min quiescence were sequentially applied to the plate. The frequency and output of the sonication were set to 20 kHz and 700 watts, respectively, and the temperature was kept at 37 °C. Emission of ThT fluorescence was recorded at the end of ultrasonication/shaking treatment.

Protein Adsorption at Hydrophobic Surfaces—The structure of a protein adsorbed at a hydrophobic-hydrophilic interface is perturbed by interfacial, intermolecular, and hydrophobic interactions (31–34). However, studies suggest that among these destabilizing interactions, hydrophobic forces play a dominant role (35–37). Many authors have shown that the interaction of a protein with an apolar surface can be triggered by exposed hydrophobic domains (37–40). If we consider the exposed hydrophobic domains present on β_2m, we can estimate the forces acting on the protein once it is absorbed on an apolar surface using a model recently proposed to calculate hydrophobic interaction energies. According to this model, the hydrophobic interaction energies (E_{hydro}) between two apolar surfaces can be described by the following model (41).

$$E_{\text{hydro}} = -2 * \gamma(a - a_0)\exp(-d/D_{\text{hydro}}) \qquad \text{(Eq. 2)}$$

where γ is the interfacial tension, d is the distance between the two surfaces, a is the exposed area of the molecule at distance d, a_0 is the optimum exposed area of the molecule, which we consider to be equal to the area of one amino acid, and D_{hydro} is the hydrophobic decay length. Thus, the hydrophobic force (F_{hydro}) acting on the molecule can be calculated as follows.

$$\begin{aligned} F_{\text{hydro}} &= -(dE_{\text{hydro}}/dd) \\ &= (-2 * \gamma(a - a_0)\exp(-d/D_{\text{hydro}}))/D_{\text{hydro}} \qquad \text{(Eq. 3)} \end{aligned}$$

For our system, we can consider that $\gamma = 50$ mJ/m^2 (42), $a_0 = 50$ Å^2, $D_{\text{hydro}} = 10$ Å (43), and $a(d) = a_0(1 - \exp(-d/D_{\text{hydro}}))^{-1/2}$. Using these values, it can be calculated that when the distance between the protein and the surface ranges between 1 and 10 Å the interaction energies and forces acting on the molecule vary from 14.7 to 0.7 kcal mol^{-1} and from 4.8 to 102 piconewtons, respectively. These forces/energies should be large enough to perturb the three-dimensional structure of β_2m molecules (2, 44, 45).

RESULTS

Structural Basis of Amyloidogenicity of D76N β_2m Variant—The Asp^{76} mutation to Asn substantially destabilizes β_2m, and differential scanning calorimetry reveals a melting temperature 10.26 °C lower than that of the wild type, corresponding to mean ± S.D. ($n = 3$) ΔH values of 63.9 ± 1.2 and 86.2 ± 1.5 kcal mol^{-1} for D76N and wild type β_2m, respectively. The calculated ΔG values (46) for unfolding of the D76N variant at 37 and at 30 °C were thus 2.73 and 2.86 kcal mol^{-1} lower, respectively, than for wild type β_2m.

The complex folding mechanisms shared by wild type and D76N β_2m involve multiple intermediates and parallel folding routes with two major exponential phases observed during refolding (3): an initial fast phase that was 3 times slower for the D76N variant at 0.2 M Gdn-HCl and a subsequent prolyl *trans-cis* isomerization-dependent slow phase (4) (data not shown). Unresolved fluorescence changes take place in the dead time of the experiments, showing that a burst phase occurs on the sub-millisecond time scale with the same amplitude for both pro-

TABLE 1

Thermodynamic and kinetic values

All values are mean ± S.D. ($n = 3$) · C_m (M), midpoint concentration of Gdn-HCl; $\Delta G^0(H_2O)$ (kcal mol^{-1}), free energy of unfolding in the absence of denaturant; m (kcal mol^{-1} M^{-1}), dependence of $\Delta G^0(H_2O)$ on denaturant concentration; k_0 (s^{-1}), value of rate constant extrapolated to zero denaturant concentration. These values were used to calculate the free energy changes shown in Fig. 1. In addition, ΔG values from the I_T to the U_T states, 3.25 ± 2.3 and 0.55 ± 0.3 kcal mol^{-1} for wild type and D76N β_2m, respectively, were determined with the procedure described under "Experimental Procedures." m_u, m_{fast}, and m_{slow} (kcal mol^{-1} M^{-1}), dependence of ln$k$ on Gdn-HCl concentration for unfolding and fast and slow phases of folding, respectively.

	Gdn-HCl denaturation at equilibrium		
	C_m	$\Delta G°(H2O)$	m
Wild type	2.0 (± 0.2)	5.7 (± 0.4)	2.73 (± 0.7)
D76N	1.15 (± 0.1)	3.0 (± 0.15)	2.64 (± 0.3)

	Chevron plots					
	Unfolding		Fast phase		Slow phase	
	k_0	m_u	k_0	m_{fast}	k_0	m_{slow}
Wild type	1.52 (± 0.2) $\times10^{-4}$	1.34 (± 0.17)	5.17 (± 1.1)	-2.37 (± 0.47)	3.01 (± 1.3) $\times10^{-3}$	0.076 (± 0.03)
D76N	6.52 (± 1.3) $\times10^{-4}$	1.20 (± 0.24)	1.78 (± 0.37)	-2.26 (± 0.50)	1.95 (± 0.8) $\times10^{-3}$	0.070 (± 0.03)

teins. In contrast to folding, unfolding appears to be a monophasic process. D76N β_2m unfolded faster than wild type with a 2-fold increase at 5.4 M Gdn-HCl. The free energy diagrams of (un)folding of the two proteins at pH 7.4, 30 °C in the absence of any denaturant (Fig. 1) derived from combined equilibrium unfolding and kinetic data (Table 1) show that the N_C native state of the variant is (mean ± S.D.; $n = 3$) 2.7 ± 0.25 kcal mol^{-1} less stable than that of native wild type β_2m, thereby promoting the population of partially folded, typically amyloidogenic states of the variant (47). The native-like state of β_2m with the His^{31}-Pro^{32} peptide bond in a non-native *trans* configuration (N_T), previously shown to be highly related to the amyloidogenic pathway of wild type β_2m and populated at (mean ± S.D.; $n = 3$) 4.8 ± 3.0% at equilibrium, was remarkably more abundant in the D76N variant (mean ± S.D.; $n = 3$) at ~25 ± 9%.

Despite the notably reduced stability of D76N β_2m, its solution structure did not differ significantly from wild type. Other than obvious changes at the mutation site and neighboring residues, the NMR signature of the variant at 25 °C was nearly the same as that of the wild type protein (Fig. 2).

Unequivocal assignment at 11.7 teslas (500-MHz ^{1}H frequency) could be obtained for 85% of the backbone ^{1}H, ^{15}N, and ^{13}C nuclei, but no major chemical shift difference, that is no major structure deviation, was observed compared with the wild type protein, consistent with the crystallographic findings (2). For instance, the average difference in deviation from random coil values of the H$^\alpha$ chemical shifts between D76N and wild type β_2m, $\Delta(\Delta\delta)$, is -0.012 ± 0.017 ppm with a value of -0.01 ppm even at the mutation site. This reflects essentially invariant local secondary and tertiary structure. One relevant effect related to the mutation is that the same chemical shifts are observed for the carboxyamide resonances of both Asn^{42} and Asn^{76} (Fig. 3*A*), suggesting that the two side chains share the same chemical environment, consistent with the occurrence of reciprocal H-bonds between the carboxyamides. The presence of an interaction between the side chains of residues 42 and 76 in the variant protein is supported by several lines of evidence including two-dimensional ^{1}H NOESY (Fig. 3*B*) and ROESY cross-peaks (data not shown).

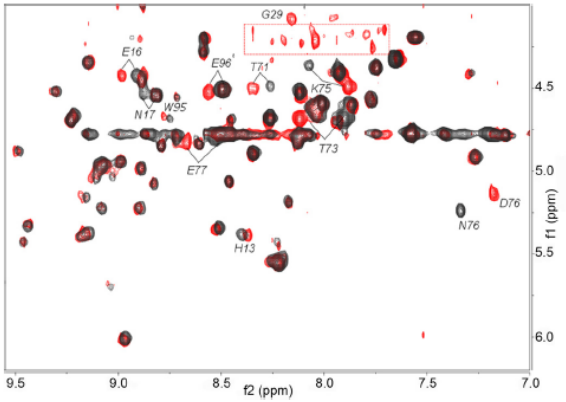


FIGURE 2. **Overlay of the 500-MHz TOCSY fingerprints.** Spectra of D76N (*black* contours) and wild type β_2m (*red* contours) were monitored at 25 °C. The different cross-peaks of the mutation site, *i.e.* Asp^{76} and Asn^{76}, are indicated. Other differences are observed for the H^N-H^α connectivities of residues Lys^{75} and Glu^{77} flanking the mutation site, although changes also occur at the close Thr^{71} and Thr^{73} residues. Additional shifts can be detected for the backbone amides of His^{13}, Glu^{16}, Asn^{17}, Trp^{95}, and Glu^{96} that are located in the same apical region as the loop including the mutation site and are likely to depend on local solvation changes. The *boxed* region in the spectrum of the wild type protein encloses resonances from a limited population of non-natively folded structures arising from the partial unfolding occurring in aged samples. The Gly^{29} H^N-H^α cross-peak is undetectable for the variant protein because of fast exchange under the experimental conditions.

In the wild type protein, the side chain amide of Asn^{42} is H-bonded to the Asp^{76} side chain carboxylate that also forms salt bridges with the side chain ammoniums of Lys^{41} and Lys^{75}. The latter salt bridges contribute electrostatic stability and a computed pK_a shift of -1.2 units for the Asp^{76} carboxylate with respect to the standard value (Table 2). Despite the survival of the residue 42-76 interaction, the asparagine substitution for aspartate has a substantial impact in the variant protein. At pH

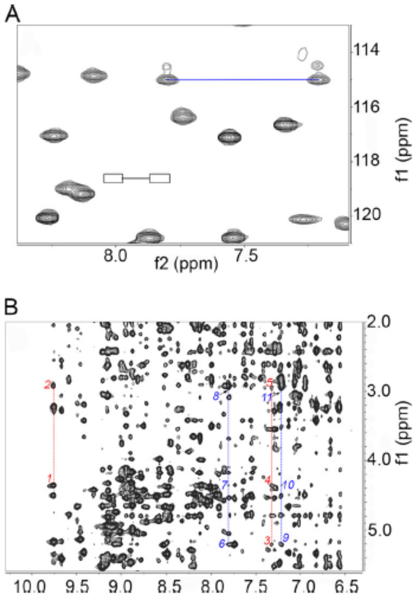


FIGURE 3. **NMR patterns of Asn^{42} and Asn^{76} carboxyamides.** *A*, region of ^{15}N-^{1}H heteronuclear single quantum correlation spectrum of D76N β_2m. The *boxed* extremes of the *black line* correspond to the locations expected for the carboxyamide resonances of Asn^{42} in the wild type protein spectrum; the extremes of the *blue line* indicate the new signals observed for D76N β_2m. The new signal pair can be attributed to Asn^{76} and Asn^{42} side chain amides from NOESY evidence. *B*, details from ^{1}H two-dimensional NOESY spectrum of D76N β_2m. The *dashed red lines* cross through the backbone amide NOEs of Asn^{42} (9.75 ppm) and Asn^{76} (7.31 ppm), whereas the *dashed blue lines* cross through the NOEs of the carboxyamide NHs at 7.80 and 7.21 ppm that are attributed to Asn^{42} and Asn^{76} and tentatively assigned as $H^{\delta21}$ and $H^{\delta22}$, respectively. To avoid notation crowding, the relevant connectivities are highlighted by *numbers* (*1*, Asn^{42} NH-Asn^{42} H^{α}; *2*, Asn^{42} NH-Asn^{42} H^{β}; *3*, Asn^{76} NH-Asn^{76} H^{α}; *4*, Asn^{76} NH-Asn^{42} H^{α}; *5*, Asn^{76} NH-Asn^{76} H^{β}; *6*, Asn^{76}/Asn^{42} $H^{\delta21}$-Asn^{76} H^{α}; *7*, Asn^{76}/Asn^{42} $H^{\delta21}$-Asn^{42} H^{α}; *8*, Asn^{76}/Asn^{42} $H^{\delta21}$-Asn^{76}/Asn^{42} H^{β}; *9*, Asn^{76}/Asn^{42} $H^{\delta22}$-Asn^{76} H^{α}; *10*, Asn^{76}/Asn^{42} $H^{\delta22}$-Asn^{42} H^{α}; *11*, Asn^{76}/Asn^{42} $H^{\delta21}$-Asn^{76}/Asn^{42} H^{β}, with the slash indicating the lack in stereospecific assignment). The NOESY cross-peaks observed for the backbone amides of Asn^{76} and Asn^{42} as well as for the carboxyamide resonances at 7.21 and 7.80 ppm support the attribution of the latter signal pair to the carboxyamides of the same residues. Observing the same chemical shifts for the carboxyamide resonances of both Asn^{76} and Asn^{42} suggests that the two side chains experience the same environment, which in turn is consistent with the occurrence of reciprocal H-bonds between Asn^{42} and Asn^{76} carboxyamides.

7.0, the net charge of D76N β_2m is +0.3 units, whereas wild type carries an average −1.4 elementary charge (Fig. 4). This is consistent with a decreased intermolecular repulsion that facilitates aggregation at pH around neutrality. In addition, the lower stability of D76N β_2m compared with wild type as shown by microcalorimetry was confirmed at single residue resolution by NMR using BLUU-Tramp (15, 16) even under critical conditions for the variant due to its instability (Fig. 5).

The lower thermal resistance of D76N β_2m can be explicitly tracked from $^{15}N\{^{1}H\}$ NOE data measured at 25 and 37 °C, respectively (Fig. 6). These data indicate an extended loss of rigidity consistent with the H-bond analysis (Fig. 7). The average NOE values at 25 °C are consistent with a higher mobility of the variant compared with the wild type. The loss of rigidity observed when temperature increases at 37 °C is slightly more marked for the wild type protein (Fig. 6). However, the fluctuations of the thermally induced mobility increment that derive

TABLE 2

pK_a values for groups titrating below pH 10 in wild type and D76N β_2m

Values were calculated with the software BLUUES (22) and compared with the corresponding random coil model values (limit). The pK_a values for Lys^{41} side chain, which establishes an electrostatic interaction with Asp^{76} in wild type β_2m, are also given. Asp^{76} is the lowest titrating residue with a pK_a of 2.6, corresponding to a pK_a shift of 1.2 compared with the free amino acid pK_a of 3.8. This shift is primarily due to the interaction with Lys^{41} and Lys^{75}. Among other groups with a significant shift of pK_a are Asp^{34} (more exposed to the solvent than Asp^{53}) and Asp^{38} (similar degree of exposure limitation and interactions as Asp^{76}). Among histidine residues, His^{31} and His^{84} both have a pK_a shifted toward acidic pH by ~2 pK units.

Residue	pK_a (limit)	pK_a WT β_2m	pK_a D76N β_2m
Ile^{1}	8.000	7.388	7.401
His^{13}	6.500	6.078	6.006
Glu^{16}	4.500	3.679	3.668
His^{31}	6.500	4.641	4.630
Asp^{34}	3.800	3.117	3.127
Glu^{36}	4.500	4.211	4.207
Asp^{38}	3.800	3.492	3.495
Glu^{44}	4.500	4.678	4.599
Glu^{47}	4.500	4.119	4.120
Glu^{50}	4.500	4.927	4.940
His^{51}	6.500	6.967	6.956
Asp^{53}	3.800	3.564	3.589
Asp^{59}	3.800	3.727	3.711
Glu^{69}	4.500	4.579	4.535
Glu^{74}	4.500	4.191	4.100
Asp^{76}	3.800	2.619	
Glu^{77}	4.500	4.056	4.089
His^{84}	6.500	4.471	4.525
Asp^{96}	3.800	5.350	5.338
Asp^{98}	3.800	4.374	4.408
Met^{99}	3.200	4.863	4.836
Lys^{41}	10.500	11.173	10.470

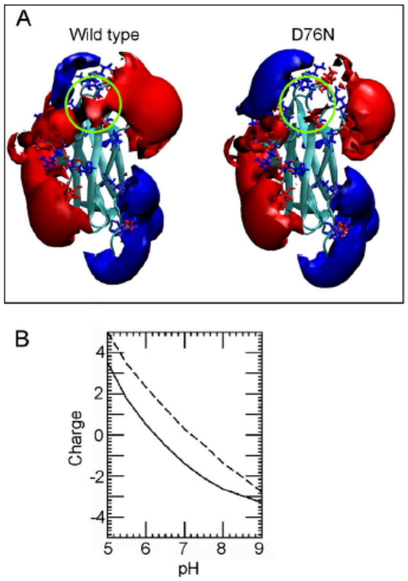


FIGURE 4. **Electrostatic properties of wild type and D76N β_2m.** *A*, isopotential curves are displayed at $+0.5kT/q$ (*blue*) and $-0.5kT/q$ (*red*). The regions around Asp^{76} and Asn^{76} are *circled*. The molecular structure of β_2m (Protein Data Bank code 3HLA, chain B) (74) and the D76N β_2m homology model were prepared for electrostatic calculations using the program PDB2PQR, which adds hydrogens and assigns charges and atomic radii according to different force fields. The CHARMM force field parameters were used except that the minimum radius was set at 1.0 Å. Similar results are obtained from the crystallographic structure of the variant (Protein Data Bank code 4FXL) (2). *B*, relationship between total net charge (elementary charge units) and pH for wild type (*solid line*) and D76N β_2m (*dashed line*). Total charge was calculated by summing the charges of all ionizable groups for which individual pK_a values had been calculated using BLUUES (22).

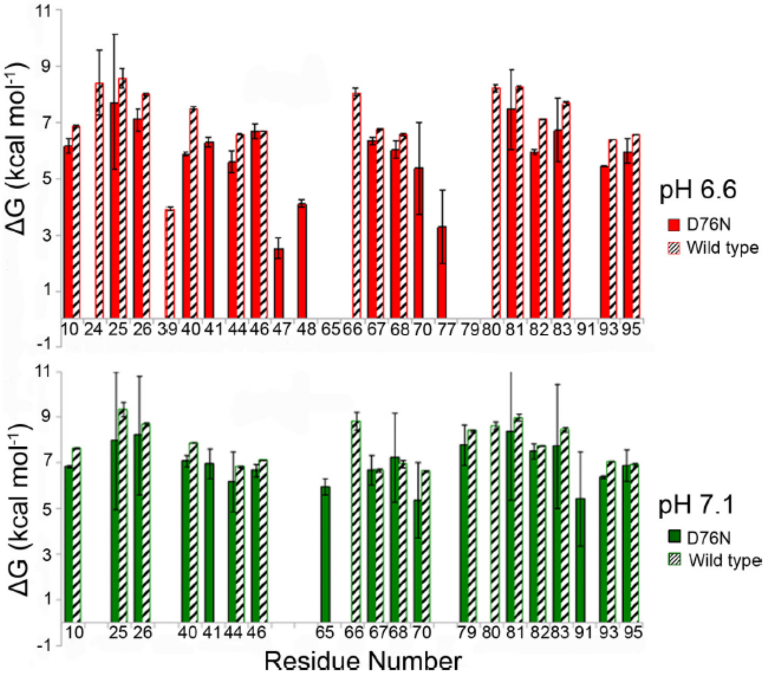


FIGURE 5. **BLUU-Tramp analysis.** ΔG_{unfol} values according to residue number from the amide exchange-based approach BLUU-Tramp (15, 16) are shown as histograms for D76N and wild type β_2m at different pH values. Despite the instability of the variant during the experiment, a ΔG_{unfol} difference of 1.8 ± 1.2 kcal mol^{-1} at pH 7.1, 30 °C was assessed between wild type and variant β_2m. This value is consistent, within the experimental error, with the destabilization (2.86 kcal mol^{-1}) estimated more precisely by calorimetry under similar conditions. *Error bars* represent S.E. Both species show similar patterns of structural determinants, that is residues exhibiting the largest opening ΔG values and that presumably represent the global opening process. Note that these residues were selected among the resolved resonances exhibiting an observable isotope exchange pattern within the designated temperature interval. Larger *error bars* for some residues of the variant compared with wild type are due to experimental error in the quantification of the cross-peak of the unstable variant. Nonetheless, in addition to reduction of ΔG_{unfol}, further destabilization was evident in residues 40–44 of the variant (C strand end, CC′ loop). Destabilization of the weaker connection between loops EF and CC′ should also affect the hydrophobic packing of residues 39–40 and 79–80 with an instability propagation path linked to the F strand bearing the disulfide bridge to the B strand, which is the core of the immunoglobulin domain architecture.

from the $NOE_{25\,°C}/NOE_{37\,°C}$ ratio are uniformly spread over the whole molecule of the wild type protein compared with a distinctive uneven pattern in the variant. When related to the spatial structure, this pattern delineates an instability propagation path (illustrated by the backbone mobility changes with temperature shown in Fig. 8). Decreased conformational stability and reduction of repulsive electrostatic interactions make D76N β_2m extremely sensitive to aqueous boundary conditions where the preferential interface partitioning of the protein and the subsequent surface tension fluctuations overcome the determining role for the effective force field of hydrophobic folding drives, thereby enhancing unfolding and fibrillogenic nucleation events.

Fibrillogenesis of D76N β_2m Occurs under Physiological Conditions—In marked contrast to the wild type protein, D76N β_2m rapidly aggregates as shown by ThT fluorescence (Fig. 9*A*) and atomic force microscopy (Fig. 9*B*) when agitated at pH 7.4 and 37 °C in the presence of an air-water interface. We had already shown that neither protein aggregates in the absence of agitation (2). Furthermore, replacement of the air-water interface with Teflon-water, which reduces the interfacial tension from about 70 to 50 millinewtons/m (43) at fixed interfacial area, also completely suppressed aggregation of the D76N variant (Fig. 9*A*).

The water-air boundary is known to behave as a hydrophobic interface (43, 48, 49). To evaluate interfacial effects on D76N β_2m fibrillogenesis, the air-water interface was removed and replaced with a hydrophilic-hydrophobic interface. We used the prototypic hydrophobic surface provided by graphite or elastin, the very hydrophobic ubiquitous insoluble fibrillar component of the extracellular matrix (50) (Fig. 10). In the absence of the air boundary, graphite triggered fibril formation on the sheet surfaces (Fig. 10*A*) without massive conversion of the bulk β_2m in solution (data not shown). Ultrasonication triggered fibrillogenesis by D76N β_2m even in the absence of an air-water interface (Fig. 10*B*, *blue lines*), and aggregation was accelerated by addition of carbon nanotubes (Fig. 10*B*, *green lines*). In these conditions, ultrasonication had no effect on aggregation of wild type β_2m (*red* and *dashed black lines*). Inclusion of fibrillar elastin in the protein solution, kept at 37 °C under agitation in the absence of an air boundary, strongly promoted fibril formation by D76N β_2m (Fig. 10*C*, *red triangles*). Elastin did not promote aggregation by wild type β_2m under the same conditions (Fig. 10*C*, *black triangles*).

Under conditions that suppress aggregation, the absence of an air-water interface or any agitation of the protein solution, the consistent enhancement of D76N β_2m fibrillogenesis by graphite or elastin clearly demonstrates the impact of an

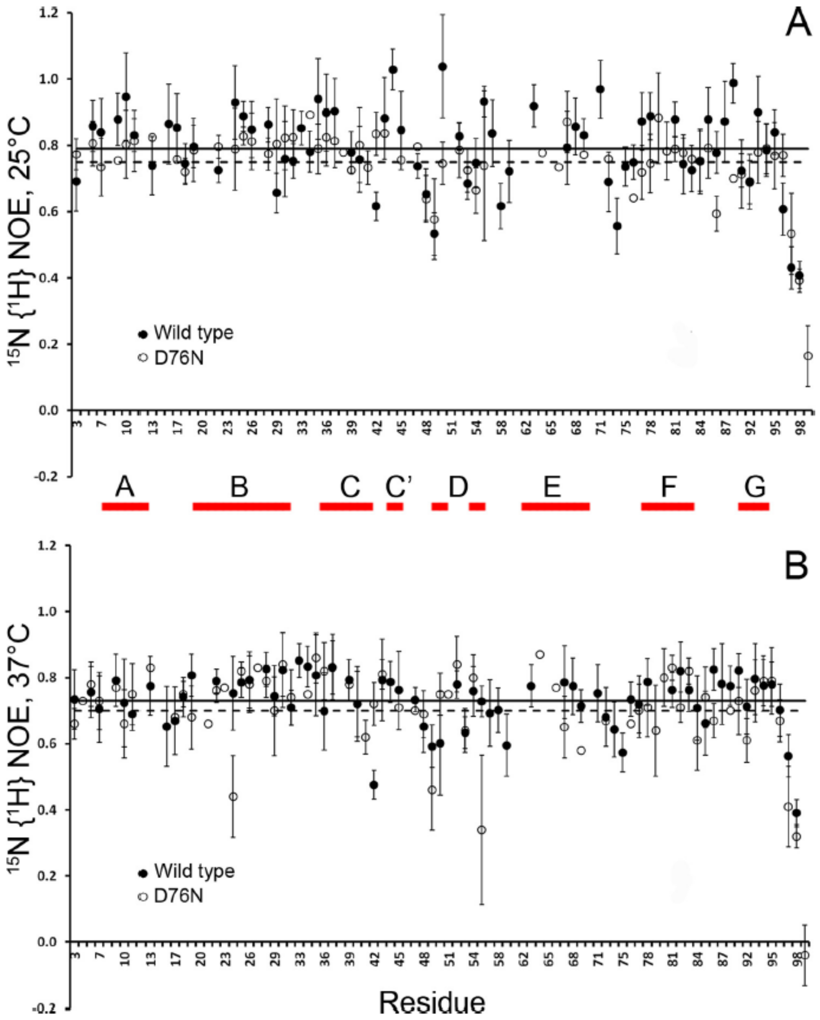


FIGURE 6. **Temperature and heteronuclear NOE.** ^{15}N{^{1}H} NOE values measured at 11.7 teslas (500.13-MHz ^{1}H frequency) at 25 (*A*) and 37 °C (*B*) for wild type and D76N β_2m. The *horizontal lines* represent the average values for wild type (*solid line*) and variant (*dashed line*) species; *error bars* indicate S.E. The *abscissa axes* do not include the positions 5, 14, 32, 72, and 90 corresponding to proline residues, therefore lacking in secondary amides. Only the pairs of well resolved cross-peaks observed at both temperatures were selected for each species. The strand naming scheme is drawn *parallel* to *abscissas* between the two panels.

increased interfacial area. In addition to the crucial role of hydrophobic-hydrophilic interfaces, shaking of the solution is required for amyloid conversion of D76N β_2m in the bulk, and the mechanism by which agitation influences the kinetics of fibril formation is clearly important.

Role of Shear Forces and Hydrophobic Surfaces in β_2m Amyloidogenesis—Agitation of a protein solution applies hydrodynamic shear stress, which in principle could also contribute to protein destabilization leading to denaturation (51). We have therefore calculated the shear forces acting on the β_2m molecule using the equation (52)

$$T = F_s/A = \mu * (dv/dx) \qquad \text{(Eq. 4)}$$

where T is the shear stress, F_s is the shear force, A is the cross-sectional area of the molecule, μ is the dynamic viscosity of the fluid, and dv/dx is the shear rate, that is the fluid velocity gradient. F_s clearly depends greatly on molecular shape (for example, see Refs. 51 and 53). With uniform fluid flow, the major forces acting on a molecule in the bulk are elongational forces along the flow axis that depend on both protein length and shape. We have used the model proposed by Shankaran and Neelamegham (54) that assumes that the molecule has a dumbbell shape, which yields a force coefficient derived from the radius of the two ends of the dumbbell and the distance between them. F_s is then calculated from

$$F_s = \alpha \ . \ \mu \ . \ \gamma \ . \ R^2 \qquad \text{(Eq. 5)}$$

where α is the force coefficient, μ is the dynamic viscosity, γ is the shear rate (dv/dx), and R is the radius of the molecule. The shear rate of the flow can be calculated as

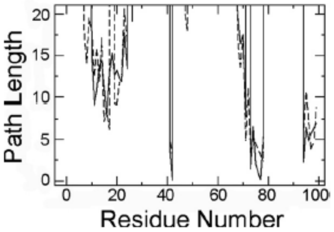


FIGURE 7. **Hydrogen bonding relationship of the residue 76 side chain.** Shortest hydrogen bond path lengths involving at least one side chain atom for residue 76 in wild type (*solid line*) and D76N (*dashed line*) β_2m from the Floyd-Warshall algorithm are shown. The variant shows only minor differences from wild type, specifically with a slightly stronger connection of residue 76 to loop AB and overall slightly weaker connection to other residues of the loop EF itself, to loop CC′D, and to the C terminus. The wild type Asp^{76} residue is strongly linked to Asn^{42} with the hydrogen bond Asn^{42} H$^{\delta 2}$-Asp^{76} O$^{\delta 1}$, which is found in 2222 of 2500 snapshots. In turn, Asn^{42} is hydrogen-bonded with Glu^{77} via two nearly completely conserved hydrogen bonds (Asn^{42} H^{N}-Glu^{77} O and Glu^{77} H^{N}-Asn^{42} O$^{\delta 1}$). The connection to the C-terminal residue Lys^{94} is due to the salt bridge with Glu^{77}. The connection with the loop AB is weaker and involves the hydrogen bonds of Asp^{76} with Thr^{73} (Thr^{73} H$^{\gamma 1}$-Asp^{76} O$^{\delta 1}$), the fluctuating hydrogen bond of Arg^{97} side chain with Thr^{73} backbone, and the salt bridge of Arg^{97} with Glu^{16}. The connection between Thr^{73} and Arg^{97} is very weak. In D76N β_2m, all connections are weaker except for the path to loop AB with a hydrogen bond of the side chain of Arg^{97} to the side chain of Asn^{76} and the hydrogen bond of the side chain of Arg^{97} to the side chain of Asn^{17}.

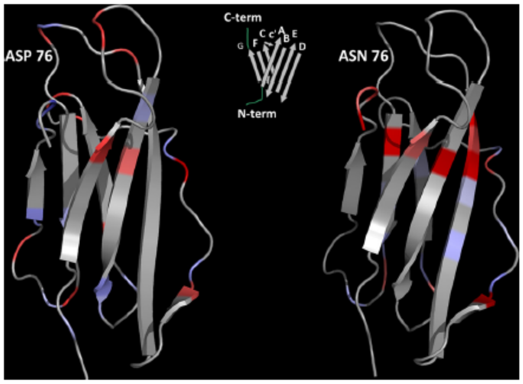


FIGURE 8. **Changes in backbone mobility.** The mobility of wild type and D76N β_2m based on the heteronuclear NOE ratios derived from values shown in Fig. 6 with increments (*red* intensity) and decrements (*blue* intensity) is shown. The weaker connection between EF and CC′ loops of the variant spreads all around the underlying region with flexibility gains involving the intervening edges of strands A, B, E, and F. The solution structure of β_2m (Protein Data Bank code 1JNJ) was also used to model the D76N variant. The *central* diagram shows the strand naming scheme.

$$\gamma = V/L \qquad \text{(Eq. 6)}$$

where V is the translational velocity of the fluid and L is the half-length of the cell. For our system (Fig. 9*A*), $\gamma = 94.2$/s, and using an α value of 10 as reported previously for a similar system (54), the calculated shear force, F_s, of 3.3×10^{-17} newton is much lower than the force of $\sim 10^{-12}$ newton typically required to unfold proteins (44, 45). Shear stress alone is thus unlikely to destabilize native β_2m, but liquid agitation increases the sampling frequency of natively folded monomers at the hydrophobic-hydrophilic interface and facilitates the exchange of misfolded monomers and interface-formed nuclei of aggregation with the bulk solution. All these entities are then available for the recruitment of other protein units, thereby increasing the efficiency of aggregation (51).

At the hydrophobic-hydrophilic interface, the native protein fold is perturbed by the combined action of interfacial, intermolecular, and hydrophobic interactions (31–34) of which the latter is apparently dominant (35–37) with much evidence that exposed hydrophobic domains trigger the interaction of a protein with an apolar surface (37–40). Based on the known exposed hydrophobic domains present on β_2m, we estimate that the forces acting on the protein at the hydrophobic-hydrophilic interface are in the range of 5–100 piconewtons and are sufficient to perturb its three-dimensional structure (2, 44, 45) (see "Experimental Procedures" for details).

The crucial role of interfacial forces in protein destabilization and fibrillogenesis must vary with the different electrostatic charges and thermodynamic stabilities of individual proteins because efficient adsorption at a hydrophobic-hydrophilic interface depends on overcoming the energy barriers of surface pressure and electrostatic repulsion (39, 55–58). The almost neutral D76N β_2m molecule is likely to adsorb more rapidly than the more charged wild type β_2m, and once at the interface, it should undergo larger structural perturbations as it is thermodynamically less stable (59, 60) (Fig. 9*A*).

D76N β_2m Primes the Fibrillar Conversion of Wild Type β_2m in Vitro—The D76N β_2m variant in solution in physiological buffered saline converts into fibrils at the highest rate ever reported for an amyloidogenic globular protein under these conditions. When mixed in equimolar proportions with native wild type β_2m, all the latter was also transformed into insoluble fibrils (Fig. 11*A*) with a much shorter lag phase than reported previously for seeding by the truncated isoform lacking the six N-terminal residues (ΔN6 β_2m) (61) (Fig. 11*B*). Fibrillogenesis was monitored by quantifying the soluble protein by native 1% agarose gel electrophoresis in which the soluble forms of wild type, D76N, and ΔN6 β_2m are readily distinguished by their different respective electrophoretic mobilities (Fig. 12). The duration of the lag phase depended on the aggregation state of D76N variant, which can potently promote aggregation of wild type β_2m only when it is assembled into elongated oligomers and filaments (Fig. 11*C*).

Unexpectedly, our previous proteomic characterization of *ex vivo* natural amyloid fibrils from the tissue deposits of patients carrying the amyloidogenic D76N mutation showed only the presence of full-length variant protein (2). Because wild type β_2m is intrinsically amyloidogenic *in vivo* and forms abundant amyloid fibrils in patients affected by dialysis-related amyloidosis, the absence of any wild type β_2m in the hereditary variant β_2m deposits indicates that *in vivo* fibrillogenesis is more complex than the simple *in vitro* experiment containing just wild type and variant β_2m. A likely physiological factor modulating misfolding, aggregation, and fibrillogenesis could be the presence of extracellular chaperones. Indeed, we show here that α-crystallin (62, 63) prevented amyloid conversion of wild type β_2m induced by D76N β_2m fibrils without interfering with fibrillogenesis of the variant at the lowest chaperone concentration used (1 μM). However, at 40 μM α-crystallin, even the conversion of D76N β_2m is significantly reduced (Fig. 13). The effect of this prototypic chaperone strongly suggests mecha-

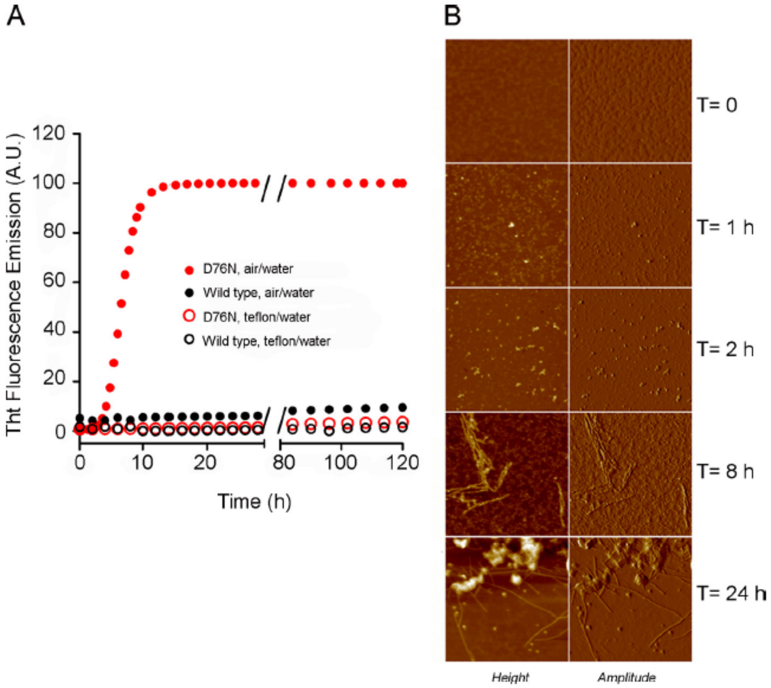


FIGURE 9. **Fibrillogenesis of D76N and wild type β_2m.** *A*, the time course of aggregation of D76N β_2m (*red*) and wild type (*black*) was monitored under stirring conditions by fluorescence emission of ThT (using 445 and 480 nm as excitation and emission wavelengths, respectively). Proteins were dissolved at 40 μM in 25 mM sodium phosphate buffer, pH 7.4 at 37 °C. Aggregation experiments were monitored in the presence of air-water (*filled circles*) and Teflon-water interfaces (*empty circles*). *B*, tapping mode AFM images of different stages of aggregation of D76N β_2m carried out under stirring conditions and in the presence of an air-water interface. Oligomers formed after 1 h, prefibrillar aggregates coexisted with oligomers after 2 h, filaments were observed after 8 h, and fibril clusters were observed after 24 h. Scan size, 1 μm; Z range, 15 (times 0 and 24 h), 8 (times 1 and 2 h), and 3 nm (time 8 h). *a.u.*, arbitrary units.

nisms responsible for the observed composition of naturally occurring fibrils in the affected kindred (2).

DISCUSSION

β_2m is among the most extensively studied globular protein precursors of human amyloid fibrils. The discovery of the first natural variant of human β_2m as the cause of hereditary systemic amyloidosis uniquely enables a very informative comparison of two different types of β_2m amyloidosis with distinctly different clinical and pathological features. The D76N residue substitution allows a fully folded three-dimensional structure almost identical to that of the wild type protein that forms amyloid fibrils in dialysis-related amyloidosis. However, dissection of the mechanism of D76N β_2m fibrillogenesis confirmed our previously established paradigm that the amyloidogenicity of monomeric globular proteins is intimately connected to destabilization of the native fold (64). Importantly, a specific intermediate of the folding pathway of wild type β_2m, which was previously structurally characterized and shown to play a crucial role in priming the amyloid transition (47), is particularly abundantly populated by the D76N variant. It is therefore possible that this specific residue substitution facilitates the molecular mechanism responsible for the inherent amyloidogenicity of wild type β_2m and thereby enables the variant to cause clinical pathology even at a normal plasma concentration rather than the grossly increased abundance of wild type β_2m responsible for dialysis-related amyloidosis.

Our elucidation of the structural properties and folding dynamics of the highly amyloidogenic D76N variant has validated several earlier interpretations of the molecular basis of the amyloid transition of the wild type β_2m. The characterization of conditions for rapid fibrillogenesis of the variant in a physiological milieu is therefore particularly significant. D76N β_2m forms amyloid fibrils within a few hours in physiological buffers *in vitro* that is enhanced by fluid agitation and exposure to a hydrophobic surface. In contrast, fibrillogenesis of wild type β_2m is extremely slow under physiological conditions, being minimal or absent after 100 days of incubation (61). Fluid agitation has been shown previously to be crucial in priming amyloid fibrillogenesis of other polypeptides including amyloid β (31), insulin (65), apolipoprotein C-II (66), and α-synuclein (67), but all these precursors were either natively unfolded (apolipoprotein-CII, amyloid β, and α-synuclein) or induced to unfold by a denaturing buffer (insulin).

Our present demonstration that the interfacial forces, acting in a physiologically relevant fluid flowing over natural hydrophobic surfaces, can prime fibrillar conversion of D76N β_2m monomers identifies this protein as a genuine paradigm for amyloidogenic globular proteins causing systemic amyloidosis. Although critically destabilized by comparison with the wild type protein, it nevertheless folds in the wild type native conformation and evades intracellular quality control so that it is secreted at a physiological rate. Nevertheless, when, like all globular proteins, its stability is challenged by the physiological extracellular environment (51), the variant's propensity to misfold and aggregate as amyloid fibrils becomes evident. Within the extracellular space where amyloid is deposited in the systemic amyloidoses, the interstitial fluid flows over the extensive surfaces of the fibrous network of elastin, collagen, and leucine-rich proteoglycans (68), the high hydrophobicities of which play a key role in promoting local unfolding of globular proteins. We have previously reported the capacity of collagen to

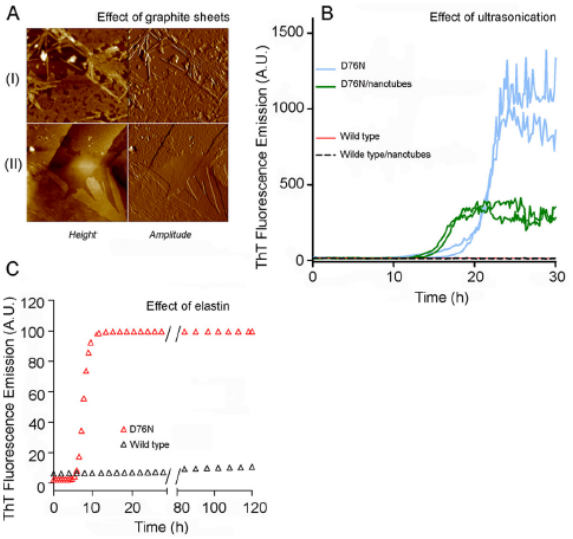


FIGURE 10. **Fibrillogenesis of D76N and wild type β_2m in presence of hydrophilic-hydrophobic interfaces.** *A*, tapping mode AFM images of fibrils formed by D76N β_2m in the presence of graphite sheets under stirring conditions (*I*) or without agitation (*II*). Scan size, 2 μm, Z range, 15 (*I*) and 25 nm (*II*). *B*, time course of fibril formation by D76N β_2m under ultrasonication (*light blue lines* show replicate experiments) in contrast to the absence of fibrillogenesis by wild type β_2m (*red line*) under the same conditions. D76N β_2m fibril formation was accelerated in the presence of carbon nanotubes (*green lines*), whereas wild type β_2m (*dashed black line*) did not aggregate. *C*, fibrillogenesis of D76N (*red triangles*) and wild type β_2m (*black triangles*) carried out under stirring conditions at 37 °C in the presence of a Teflon-water interface with 6 μM human elastin. *a.u.*, arbitrary units.

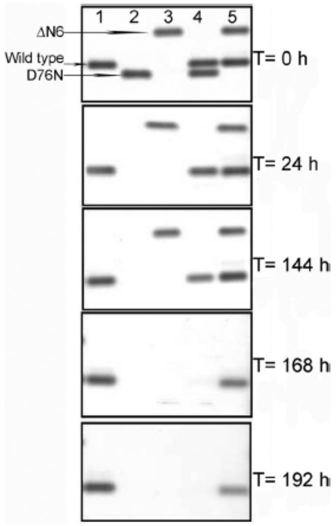


FIGURE 12. **Residual soluble β_2m during aggregation.** Agarose gel electrophoresis analysis of supernatants from fibrillogenesis of wild type β_2m alone (*lane 1*), D76N β_2m alone (*lane 2*), ΔN6 β_2m alone (*lane 3*), an equimolar mixture of wild type and D76N β_2m (*lane 4*), and an equimolar mixture of wild type and ΔN6 β_2m (*lane 5*) is shown. The *arrows* show the electrophoretic mobility of each isoform.

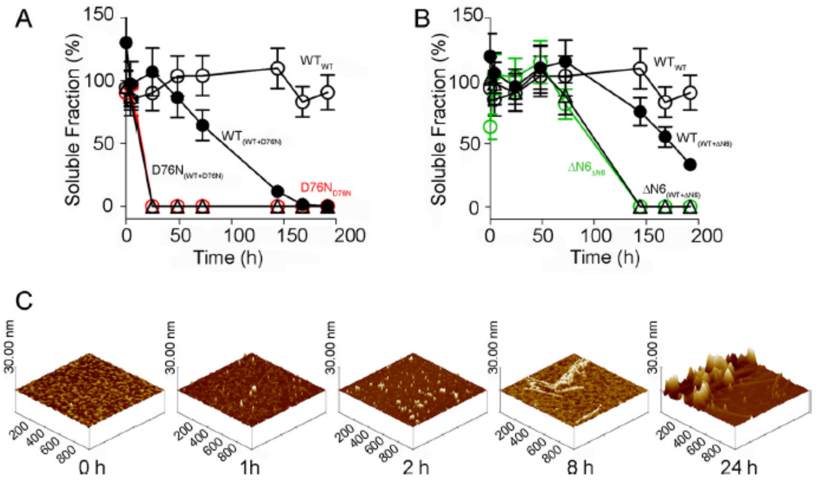


FIGURE 11. **Wild type β_2m elongates D76N fibrils *in vitro*.** *A*, soluble fractions of wild type β_2m either alone (*black empty circles*; WT_{WT}) or in an equimolar mixture with the variant (*black filled circles*; $WT_{(WT+D76N)}$) and of D76N variant either alone (*red empty circles*; $D76N_{D76N}$) or in the mixture (*black empty triangles*; $D76N_{(WT+D76N)}$). *B*, soluble fractions of wild type either alone (*black empty circles*; WT_{WT}) or in an equimolar mixture with the variant (*black filled circles*; $WT_{(WT+\Delta N6)}$) and of ΔN6 β_2m either alone (*green empty circles*; $\Delta N6_{\Delta N6}$) or in the mixture (*black empty triangles*; $\Delta N6_{(WT+\Delta N6)}$). Values are mean and S.D. (*error bars*) from three independent experiments. *C*, surface plots of AFM images showing different steps of the aggregation process of D76N β_2m. At 1 h, oligomers are present; at 2 h, they coexist with short prefibrillar aggregates; and at 8 h, filaments can be observed, whereas at 24 h, fibrils and complex fibril assemblies are seen. The surface plots were obtained from topographic tapping mode AFM images (Fig. 9*B*).

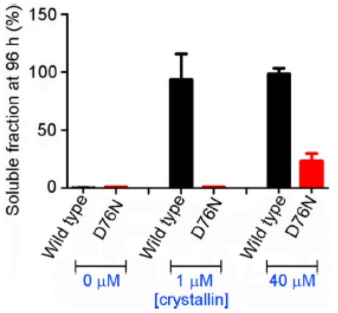


FIGURE 13. **Modulation by α-crystallin of fibril formation by wild type and D76N β_2m.** Soluble fractions of wild type and D76N variant β_2m from an equimolar mixture in the presence and absence of α-crystallin quantified at 96 h by native agarose gel electrophoresis are shown. Values are mean ± S.D. (*error bars*) from three independent experiments.

prime the formation of wild type β_2m amyloid fibrils stacked on the collagen surface, and here we show that elastin is a potent promoter of massive amyloid conversion of the D76N variant in solution. Massive enhancement by graphite nanotubes of variant β_2m amyloid fibrillogenesis further confirms the role of hydrophobic surfaces. Although Linse *et al.* (69) have previously noted an effect of hydrophobic surfaces on fibrillization of wild type β_2m with accelerated nucleation induced by nanoparticles covering a range of sizes and hydrophobicity patterns, their experiments were done at the grossly non-physiological pH of 2.5.

The kinetics of fibril formation by wild type β_2m and its truncated form ΔN6 β_2m depend on a critical nucleation step and can be accelerated by the presence of amyloid fibril seeds. In particular, the truncated form ΔN6 β_2m can catalyze the oligomerization of the wild type (70) and even prime the fibrillogenesis of the wild type protein in physiological buffer (61) although with a slower rate and lower yield than when primed by D76N β_2m. The D76N variant is also much more potent than ΔN6 β_2m in promoting formation of actual amyloid fibrils by wild type β_2m. The apparent capacity of monomeric ΔN6 β_2m to induce conformational rearrangement of the wild type protein structure has previously been ascribed to a prion-like effect (71). In our hands, however, monomeric D76N variant and ΔN6 β_2m do not prime fibrillogenesis by wild type β_2m, which only occurs when it is exposed to filaments and fibrils of the priming species. Such a mechanism is more consistent with a surface nucleation process (72) rather than a genuine prion-like effect.

The contrast between the potent *in vitro* priming and enhancement by D76N β_2m of amyloid fibril formation by wild type β_2m and the proteomic evidence that the wild type protein is not present in the *in vivo* amyloid deposits are intriguing, especially as wild type β_2m clearly does form amyloid *in vivo* in dialysis-related amyloidosis. Furthermore, in other types of hereditary systemic amyloidosis in which the wild type precursor protein is mildly amyloidogenic, for example transthyretin, most patients are heterozygotes for the causative mutation, expressing both amyloidogenic variant and wild type, and both proteins are present in the amyloid fibrils. However, as we have shown here, the capacity of D76N β_2m to catalyze fibrillogenesis by wild type β_2m can be modulated and even blocked by typical chaperones such as crystallin, and this inhibition depends on the stoichiometric chaperone/β_2m ratio. A role for extracellular chaperone-like proteins in the inhibition of wild type β_2m amyloidogenesis has been proposed previously (73), and it is plausible that the persistent, extremely high concentration of wild type β_2m in renal failure patients on dialysis may overcome the natural protective role of physiological chaperones that otherwise protect against deposition of this rather weakly amyloidogenic protein when it circulates at its normal serum concentration. In addition to illuminating the critically important biophysical features of the physiological milieu where amyloid formation takes place, our results thus open up novel avenues for exploration of hitherto unanswered questions about amyloidosis: why only a handful of all proteins ever form amyloid *in vivo*, and when, why, and where amyloid is deposited in disease.

Acknowledgments—We thank Marco Vuano for assisting with NMR data and Beth Jones for processing the manuscript.

REFERENCES

1. Bellotti, V., and Chiti, F. (2008) Amyloidogenesis in its biological environment: challenging a fundamental issue in protein misfolding diseases. *Curr. Opin. Struct. Biol.* **18,** 771–779
2. Valleix, S., Gillmore, J. D., Bridoux, F., Mangione, P. P., Dogan, A., Nedelec, B., Boimard, M., Touchard, G., Goujon, J.-M., Lacombe, C., Lozeron, P., Adams, D., Lacroix, C., Maisonobe, T., Planté-Bordeneuve, V., Vrana, J. A., Theis, J. D., Giorgetti, S., Porcari, R., Ricagno, S., Bolognesi, M., Stoppini, M., Delpech, M., Pepys, M. B., Hawkins, P. N., and Bellotti, V. (2012) Hereditary systemic amyloidosis due to Asp76Asn variant β_2-microglobulin. *N. Engl. J. Med.* **366,** 2276–2283
3. Chiti, F., Mangione, P., Andreola, A., Giorgetti, S., Stefani, M., Dobson, C. M., Bellotti, V., and Taddei, N. (2001) Direct measurement of the thermodynamic parameters of amyloid formation by isothermal titration calorimetry. *J. Mol. Biol.* **307,** 379–391
4. Kameda, A., Hoshino, M., Higurashi, T., Takahashi, S., Naiki, H., and Goto, Y. (2005) Nuclear magnetic resonance characterization of the refolding intermediate of β_2-microglobulin trapped by non-native prolyl peptide bond. *J. Mol. Biol.* **348,** 383–397
5. Lapidus, S., Han, B., and Wang, J. (2008) Intrinsic noise, dissipation cost, and robustness of cellular networks: the underlying energy landscape of MAPK signal transduction. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 6039–6044
6. Jahn, T. R., Parker, M. J., Homans, S. W., and Radford, S. E. (2006) Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat. Struct. Mol. Biol.* **13,** 195–201
7. Braunschweiler, L., and Ernst, R. R. (1983) Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Magn. Reson.* **53,** 521–528
8. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71,** 4546–4553
9. Bothner-By, A. A., Stephens, R. L., Lee, J., Warren, C. D., and Jeanloz, R. W. (1984) Structure determination of a tetrasaccharide: transient nuclear Overhauser effects in the rotating frame. *J. Am. Chem. Soc.* **106,** 811–813
10. Piotto, M., Saudek, V., and Sklenár, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **2,** 661–665
11. Hwang, T.-L., and Shaka, A. J. (1995) Water suppression that works. Excitation sculpting using arbitrary waveforms and pulsed field gradients. *J. Magn. Reson.* **112,** 275–279
12. Marion, D., and Wüthrich, K. (1983) Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H-^{1}H

spin-spin coupling constants in proteins. *Biochem. Biophys. Res. Commun.* **113,** 967–974
13. Keeler, J., Clowes, R. T., Davis, A. L., and Laue, E. D. (1994) Pulsed-field gradients: theory and practice. *Methods Enzymol.* **239,** 145–207
14. Bax, A., and Davis, D. G. (1985) MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **65,** 355–360
15. Rennella, E., Corazza, A., Codutti, L., Bellotti, V., Stoppini, M., Viglino, P., Fogolari, F., and Esposito, G. (2012) Determining the energy landscape of proteins by a fast isotope exchange NMR approach. *J. Am. Chem. Soc.* **134,** 4457–4460
16. Rennella, E., Corazza, A., Codutti, L., Causero, A., Bellotti, V., Stoppini, M., Viglino, P., Fogolari, F., and Esposito, G. (2012) Single-shot NMR measurement of protein unfolding landscapes. *Biochim. Biophys. Acta* **1824,** 842–849
17. Kay, L. E., Ikura, M., Tschudin, R., and Bax, A. (2011) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. 1990. *J. Magn. Reson.* **213,** 423–441
18. Farmer, B. T., 2nd, Venters, R. A., Spicer, L. D., Wittekind, M. G., and Müller, L. (1992) A refocused and optimized HNCA: increased sensitivity and resolution in large macromolecules. *J. Biomol. NMR* **2,** 195–202
19. Grzesiek, S., and Bax, A. (1992) Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J. Magn. Reson.* **96,** 432–440
20. Bax, A., and Ikura, M. (1991) An efficient 3D NMR technique for correlating the proton and ^{15}N backbone amide resonances with the α-carbon of the preceding residue in uniformly ^{15}N/^{13}C enriched proteins. *J. Biomol. NMR* **1,** 99–104
21. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–293
22. Fogolari, F., Corazza, A., Yarra, V., Jalaru, A., Viglino, P., and Esposito, G. (2012) Bluues: a program for the analysis of the electrostatic properties of proteins based on generalized Born radii. *BMC Bioinformatics* **13,** (Suppl. 4) S18
23. Walsh, I., Minervini, G., Corazza, A., Esposito, G., Tosatto, S. C., and Fogolari, F. (2012) Bluues server: electrostatic properties of wild-type and mutated protein structures. *Bioinformatics* **28,** 2189–2190
24. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4,** 187–217
25. Mackerell, A. D., Jr., Feig, M., and Brooks, C. L., 3rd. (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25,** 1400–1415
26. Keutsch, F. N., and Saykally, R. J. (2001) Water clusters: untangling the mysteries of the liquid, one molecule at a time. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 10533–10540
27. Martyna, G. J., Tobias, D. J., and Klein, M. L. (1994) Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101,** 4177–4189
28. Feller, S. E., Zhang, Y., Pastor, R. W., and Brooks, B. R. (1995) Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* **103,** 4613–4621
29. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14,** 51–55, 29–32
30. Jeppsson, J.-O., Laurell, C.-B., and Franzén, B. (1979) Agarose gel electrophoresis. *Clin. Chem.* **25,** 629–638
31. Jean, L., Lee, C. F., Lee, C., Shaw, M., and Vaux, D. J. (2010) Competing discrete interfacial effects are critical for amyloidogenesis. *FASEB J.* **24,** 309–317
32. Kowalewski, T., and Holtzman, D. M. (1999) *In situ* atomic force microscopy study of Alzheimer's β-amyloid peptide on different substrates: new insights into mechanism of β-sheet formation. *Proc. Natl. Acad. Sci. U.S.A.* **96,** 3688–3693
33. Rabe, M., Verdes, D., and Seeger, S. (2011) Understanding protein adsorption phenomena at solid surfaces. *Adv. Colloid Interface Sci.* **162,** 87–106
34. Yu, X., Wang, Q., Lin, Y., Zhao, J., Zhao, C., and Zheng, J. (2012) Structure, orientation, and surface interaction of Alzheimer amyloid-β peptides on the graphite. *Langmuir* **28,** 6595–6605
35. Chandler, D. (2005) Interfaces and the driving force of hydrophobic assembly. *Nature* **437,** 640–647
36. Hammer, M. U., Anderson, T. H., Chaimovich, A., Shell, M. S., and Israelachvili, J. (2010) The search for the hydrophobic force law. *Faraday Discuss.* **146,** 299–308
37. Dyson, H. J., Wright, P. E., and Scheraga, H. A. (2006) The role of hydrophobic interactions in initiation and propagation of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 13057–13061
38. Wierenga, P. A., Meinders, M. B. J., Egmond, M. R., Voragen, A. G. J., and de Jongh, H. H. J. (2003) Protein exposed hydrophobicity reduces the kinetic barrier for adsorption of ovalbumin to the air-water interface. *Langmuir* **19,** 8964–8970
39. Narsimhan, G., and Uraizee, F. (1992) Kinetics of adsorption of globular proteins at an air-water interface. *Biotechnol. Prog.* **8,** 187–196
40. Sethuraman, A., Vedantham, G., Imoto, T., Przybycien, T., and Belfort, G. (2004) Protein unfolding at interfaces: slow dynamics of α-helix to β-sheet transition. *Proteins* **56,** 669–678
41. Donaldson, S. H., Jr., Lee, C. T., Jr., Chmelka, B. F., and Israelachvili, J. N. (2011) General hydrophobic interaction potential for surfactant/lipid bilayers from direct force measurements between light-modulated bilayers. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 15699–15704
42. Johnson, K. L., Kendall, K., and Roberts, A. D. (1971) Surface energy and the contact of elastic solids. *Proc. R. Soc. Lond. A Math. Phys. Sci.* **324,** 301–313
43. Israelachvili, J. N. (2011) *Intermolecular and Surface Forces*, 3rd Ed., pp. 291–314; 361–378; 536–626 Elsevier Academic Press, Waltham, MA
44. Cecconi, C., Shank, E. A., Bustamante, C., and Marqusee, S. (2005) Direct observation of the three-state folding of a single protein molecule. *Science* **309,** 2057–2060
45. Borgia, A., Williams, P. M., and Clarke, J. (2008) Single-molecule studies of protein folding. *Annu. Rev. Biochem.* **77,** 101–125
46. Kardos, J., Yamamoto, K., Hasegawa, K., Naiki, H., and Goto, Y. (2004) Direct measurement of the thermodynamic parameters of amyloid formation by isothermal titration calorimetry. *J. Biol. Chem.* **279,** 55308–55314
47. Chiti, F., De Lorenzi, E., Grossi, S., Mangione, P., Giorgetti, S., Caccialanza, G., Dobson, C. M., Merlini, G., Ramponi, G., and Bellotti, V. (2001) A partially structured species of β_2-microglobulin is significantly populated under physiological conditions and involved in fibrillogenesis. *J. Biol. Chem.* **276,** 46714–46721
48. Krasowska, M., Zawala, J., and Malysa, K. (2009) Air at hydrophobic surfaces and kinetics of three phase contact formation. *Adv. Colloid Interface Sci.* **147–148,** 155–169
49. Hoernke, M., Falenski, J. A., Schwieger, C., Koksch, B., and Brezesinski, G. (2011) Triggers for β-sheet formation at the hydrophobic-hydrophilic interface: high concentration, in-plane orientational order, and metal ion complexation. *Langmuir* **27,** 14218–14231
50. Urry, D. W., and Parker, T. M. (2002) Mechanics of elastin: molecular mechanism of biological elasticity and its relationship to contraction. *J. Muscle Res. Cell Motil.* **23,** 543–559
51. Bekard, I. B., Asimakis, P., Bertolini, J., and Dunstan, D. E. (2011) The effects of shear flow on protein structure and function. *Biopolymers* **95,** 733–745
52. Munson, B. R., Young, D. F., Okiishi, T. H., and Huebsch, W. W. (2009) *Fundamentals of Fluid Mechanics*, 6th Ed., John Wiley & Sons, Hoboken, NJ
53. Szymczak, P., and Cieplak, M. (2011) Hydrodynamic effects in proteins. *J. Phys. Condens. Matter* **23,** 033102
54. Shankaran, H., and Neelamegham, S. (2004) Hydrodynamic forces applied on intercellular bonds, soluble molecules, and cell-surface receptors. *Biophys. J.* **86,** 576–588
55. Wierenga, P. A., Egmond, M. R., Voragen, A. G., and de Jongh, H. H. (2006) The adsorption and unfolding kinetics determines the folding state of proteins at the air-water interface and thereby the equation of state. *J. Colloid Interface Sci.* **299,** 850–857
56. Marshall, K. E., Morris, K. L., Charlton, D., O'Reilly, N., Lewis, L., Walden, H., and Serpell, L. C. (2011) Hydrophobic, aromatic, and electrostatic

interactions play a central role in amyloid fibril formation and stability. *Biochemistry* **50,** 2061–2071

57. Wang, Q., Shah, N., Zhao, J., Wang, C., Zhao, C., Liu, L., Li, L., Zhou, F., and Zheng, J. (2011) Structural, morphological, and kinetic studies of β-amyloid peptide aggregation on self-assembled monolayers. *Phys. Chem. Chem. Phys.* **13,** 15200–15210
58. Yanagi, K., Ashizaki, M., Yagi, H., Sakurai, K., Lee, Y. H., and Goto, Y. (2011) Hexafluoroisopropanol induces amyloid fibrils of islet amyloid polypeptide by enhancing both hydrophobic and electrostatic interactions. *J. Biol. Chem.* **286,** 23959–23966
59. Tripp, B. C., Magda, J. J., and Andrade, J. D. (1995) Adsorption of globular proteins at the air/water interface as measured via dynamic surface tension: concentration dependence, mass-transfer considerations, and adsorption kinetics. *J. Colloid Interface Sci.* **173,** 16–27
60. Lad, M. D., Birembaut, F., Matthew, J. M., Frazier, R. A., and Green, R. J. (2006) The adsorbed conformation of globular proteins at the air/water interface. *Phys. Chem. Chem. Phys.* **8,** 2179–2186
61. Eichner, T., Kalverda, A. P., Thompson, G. S., Homans, S. W., and Radford, S. E. (2011) Conformational conversion during amyloid formation at atomic resolution. *Mol. Cell* **41,** 161–172
62. Basha, E., O'Neill, H., and Vierling, E. (2012) Small heat shock proteins and α-crystallins: dynamic proteins with flexible functions. *Trends Biochem. Sci.* **37,** 106–117
63. Garvey, M., Griesser, S. S., Griesser, H. J., Thierry, B., Nussio, M. R., Shapter, J. G., Ecroyd, H., Giorgetti, S., Bellotti, V., Gerrard, J. A., and Carver, J. A. (2011) Enhanced molecular chaperone activity of the small heat-shock protein αB-crystallin following covalent immobilization onto a solid-phase support. *Biopolymers* **95,** 376–389
64. Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Hawkins, P. N., Dobson, C. M., Radford, S. E., Blake, C. C., and Pepys, M. B. (1997) Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **385,** 787–793
65. Webster, G. T., Dusting, J., Balabani, S., and Blanch, E. W. (2011) Detecting the early onset of shear-induced fibril formation of insulin in situ. *J. Phys. Chem. B* **115,** 2617–2626
66. Teoh, C. L., Bekard, I. B., Asimakis, P., Griffin, M. D., Ryan, T. M., Dunstan, D. E., and Howlett, G. J. (2011) Shear flow induced changes in apolipoprotein C-II conformation and amyloid fibril formation. *Biochemistry* **50,** 4046–4057
67. Pronchik, J., He, X., Giurleo, J. T., Talaga D. S. (2010) *In vitro* formation of amyloid from α-synuclein is dominated by reactions at hydrophobic interfaces. *J. Am. Chem. Soc.* **132,** 9797–9803
68. Swartz, M. A., and Fleury, M. E. (2007) Interstitial flow and its effects in soft tissues. *Annu. Rev. Biomed. Eng.* **9,** 229–256
69. Linse, S., Cabaleiro-Lago, C., Xue, W. F., Lynch, I., Lindman, S., Thulin, E., Radford, S. E., and Dawson, K. A. (2007) Nucleation of protein fibrillation by nanoparticles. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 8691–8696
70. Piazza, R., Pierno, M., Iacopini, S., Mangione, P., Esposito, G., and Bellotti, V. (2006) Micro-heterogeneity and aggregation in β_2-microglobulin solutions: effects of temperature, pH, and conformational variant addition. *Eur. Biophys. J.* **35,** 439–445
71. Eichner, T., and Radford, S. E. (2011) A diversity of assembly mechanisms of a generic amyloid fold. *Mol. Cell* **43,** 8–18
72. Cohen, S. I., Vendruscolo, M., Dobson, C. M., and Knowles, T. P. (2012) From macroscopic measurements to microscopic mechanisms of protein aggregation. *J. Mol. Biol.* **421,** 160–171
73. Ozawa, D., Hasegawa, K., Lee, Y.-H., Sakurai, K., Yanagi, K., Ookoshi, T., Goto, Y., and Naiki, H. (2011) Inhibition of β_2-microglobulin amyloid fibril formation by α_2-macroglobulin. *J. Biol. Chem.* **286,** 9668–9676
74. Saper, M. A., Bjorkman, P. J., and Wiley, D. C. (1991) Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J. Mol. Biol.* **219,** 277–319

PAPER III

Manuscript

Single-molecule folding mechanisms of the apo and Mg^{2+}-bound states of human Neuronal Calcium Sensor-1

Mohsin M. Naqvi[1,3,a], Pétur O. Heidarsson[2,a], Mariela R. Otazo[3,4], Alessandro Mossa[5], Birthe B. Kragelund[2,b], and Ciro Cecconi[1,3,b]

[1]Department of Physics, Informatics and Mathematics, University of Modena and Reggio Emilia, Via Giuseppe Campi, 41125 Modena, Italy.[2] Structural Biology and NMR Laboratory, Department of Biology, University of Copenhagen, Ole MaaløesVej 5, 2200 Copenhagen N, Denmark. [3]CNR Institute of Nanoscience S3, Via Giuseppe Campi, 41125 Modena, Italy. [4]Center of Applied Technologies and Nuclear Development (CEADEN), Department of Physics, calle 30, No. 502, Miramar, La Habana, Cuba. [5]Department of Physics, University of Bari and INFN, Sezione di Bari, Via Amendola 173, 70126 Bari, Italy.[a] These authors contributed equally to the work. [b]To whom correspondence should be addressed, E-mail: ciro.cecconi@gmail.com; bbk@bio.ku.dk

ABSTRACT

Neuronal calcium sensors-1 (NCS-1) is the primordial member of a family of proteins responsible primarily for sensing changes in neuronal Ca^{2+} concentration. NCS-1 is a multi-specific protein interacting with a number of binding partners in both calcium-dependent and independent manners, and acting in a variety of cellular processes in which it has been linked to a number of disorders such as schizophrenia and autism. Despite extensive studies on the Ca^{2+}-activated state of NCS proteins, little is known about the conformational dynamics of the Mg^{2+}-bound and apo states, both of which are populated, at least transiently, at resting Ca^{2+} conditions. Here we used optical tweezers to study the folding behavior of individual NCS-1 molecules in the presence of Mg^{2+} and in the absence of divalent ions. Under tension, the Mg^{2+}-bound state of NCS-1 unfolds and refolds in a three-state manner by populating one intermediate state comprising a folded C-domain and an unfolded N-domain. The interconversion at equilibrium between the different molecular states populated by NCS-1 was monitored in real time through constant-force measurements and the energy landscapes underlying the observed transitions were reconstructed through hidden Markov model analysis. Differently from what has been observed with the Ca^{2+}-bound state, the presence of Mg^{2+} allows both the N- and C-domain to fold through all-or-none transitions with similar refolding rates. In the absence of divalent ions, NCS-1 unfolds and refolds reversibly in a two-state reaction involving only the C-domain, whereas the N-domain has no detectable transitions. Overall, the results allow us to trace the progression of NCS-1 folding along its energy landscapes and provide a solid platform for understanding the conformational dynamics of similar EF-hand proteins.

INTRODUCTION

EF-hand calcium sensor proteins respond to cellular variations in calcium concentrations through conformational changes leading to activated states that interact with a large array of binding partners. Calcium binding *in vivo* occurs in the presence of a relatively constant thousand-fold-excess of Mg^{2+} ions and, perhaps unsurprisingly, many calcium-binding proteins also bind Mg^{2+} under non-activating concentrations of calcium (1, 2). The role of Mg^{2+} binding has been investigated for several EF hand proteins. In the invertebrate sarcoplasmic calcium-binding proteins and troponin C, Mg^{2+} binding to structural sites has been shown to be important for maintaining the structure of the molecules at any calcium concentrations (3). Mg^{2+} binding has also been suggested to modulate Ca^{2+} binding affinity of different proteins (4-6) and to switch off calcium-dependent signaling at resting Ca^{2+} levels (2). Importantly, it is becoming increasingly clear that the Mg^{2+}-bound states of many EF-hand proteins have direct functional roles independent of calcium, underscored by the observation that the single EF-hand of the cytoplasmic C-terminal domain of the cardiac CaV1.2 channel binds only magnesium (7). The Mg^{2+}-bound state of the EF-hand protein DREAM (downstream regulatory element antagonist modulator) binds DNA targets in a sequence-specific manner whereas binding of Ca^{2+} disrupts this interaction (8). The Mg^{2+}-bound state of GCAP1 has been shown to activate retinal guanylyl cyclase

(RetGC) in rod and cone cells (9). Moreover, and of relevance to the current study, the interactome of neuronal calcium sensor-1 (NCS-1) has been proposed to be modulated by ions, suggesting possible roles of the Mg^{2+}-bound state as well as of the apo state in neuronal functions (10).

Neuronal calcium sensor (NCS) proteins are a family of EF-hand calcium binding proteins expressed primarily in neurons, and NCS-1 is its primordial member. NCS-1 comprises four EF-hand motifs organized in two domains, where the N-domain consists of EF1 and EF2, and the C-domain of EF3 and EF4 (11). Nuclear magnetic resonance (NMR) spectroscopy and isothermal titration calorimetry (ITC) have been used to confirm that EF2 and EF3 bind Ca^{2+} and Mg^{2+} and they are therefore identified as structural sites, whereas EF4 binds only Ca^{2+} and is identified as a regulatory site (4). EF1, on the other hand, is unable to bind ions due to a conserved Cys-to-Pro mutation (2). In humans, NCS-1 can interact with multiple and seemingly unrelated target proteins during the execution of its functions, both in a calcium-dependent and independent manner (12). The molecular basis of this multi-specificity, exemplified by an interactome of more than twenty interaction partners reported to date, is currently unknown but suggests that NCS-1 can go through significant molecular rearrangements to accommodate different binding partners. NCS-1 has well-established functional roles such as regulating ion channels and G-protein coupled receptors (13-15), membrane trafficking (16), and learning and neuronal growth (10). Several of these functions involve the Mg^{2+}-bound state, which interacts for example with the dopamine receptor D2 (13) and mediates the activation of P14Kβ at resting cell conditions (17). Despite their obvious physiological importance, however, the structural features and conformational dynamics of the apo- and Mg^{2+}-bound states of NCS-1 are still poorly understood.

The Ca^{2+}-bound state of NCS-1 is well studied in terms of structure and has been shown, by X-ray crystallography and NMR spectroscopy, to have an open structure exposing large hydrophobic patches that mediate the interaction of the molecule with most of its target proteins (18-20). Likewise, the folding mechanism of NCS-1 in the presence of Ca^{2+} has been studied extensively using both traditional ensemble methods (4, 19) and single molecule optical tweezers experiments (11, 21). Through mechanical manipulation, we have recently characterized the energy landscape of the Ca^{2+}-bound state of NCS-1. These experiments revealed a sequential folding process where the C-domain first collapses into a partially folded structure and then undergoes a minor rearrangement to transit into its native conformation prior to the folding of the N-domain (11). We also showed that NCS-1 can become transiently kinetically trapped in two distinct misfolded states where pathologically high calcium concentrations increased the lifetimes of the misfolded structures (21). While significant information is available on the structure, stability, and folding of NCS-1 in its Ca^{2+}-bound state, little is still known about its structural and dynamical properties in the absence of Ca^{2+}. Previous studies have suggested that the structure of the apostate of NCS-1 resembles that of a molten globule (22), and that upon Mg^{2+} binding the protein adopts a closed form in which almost all hydrophobic regions are buried (4, 23). Apart from this marginal information, however, the structural details of NCS-1 in the absence of divalent ions are still elusive and our mechanistic understanding of the conformational dynamics of the apo- and Mg^{2+}-bound states is highly limited.

Single-molecule techniques, such as optical tweezers, have proven powerful for dissecting the folding mechanisms of protein molecules, providing non-averaged information inaccessible to more traditional bulk techniques (24, 25). Through mechanical manipulation it is nowadays possible to monitor in real time the dynamics of single proteins as they follow different routes and populate different intermediate states on their journey to the native structure (26-28). Analysis of the experimental data with advanced statistical methods allows the characterization of the kinetics and thermodynamics of the observed molecular transitions, and ultimately the reconstruction of the energy landscape of the protein (11, 26, 29, 30). These studies have lately proved very useful for the description of the conformational dynamics of different proteins, revealing crucial information for a better understanding of the molecular mechanisms underlining their biological function (11, 26, 29-33).

Here we use optical tweezers to characterize the folding processes of the apo- and Mg^{2+}-bound states of NCS-1 at the single molecule level. In the presence of Mg^{2+}, NCS-1 folds into an intermediate state comprising a properly folded C-domain, and at lower forces it then transits to the native state through the folding of its N-domain. In contrast, in the absence of divalent ions, NCS-1 folds in a reversible two state reaction involving only the C-domain with no observable transitions of the N-domain. Through constant-force measurements and hidden Markov model analysis the energy landscapes underlining the observed folding reactions were reconstructed. The results reported in this paper shed light on the ion-dependent dynamics of NCS-1 and provide insights into the molecular basis of its conformational plasticity.

RESULTS

Folding mechanism of the Mg^{2+}-bound state of NCS-1

Single NCS-1 molecules were manipulated as depicted in Fig. 1A. DNA handles were attached to cysteine residues engineered at position 4 and 188 and the protein was stretched and relaxed by moving the pipette relative to the optical trap. A typical force vs. extension cycle is depicted in Fig.1B.

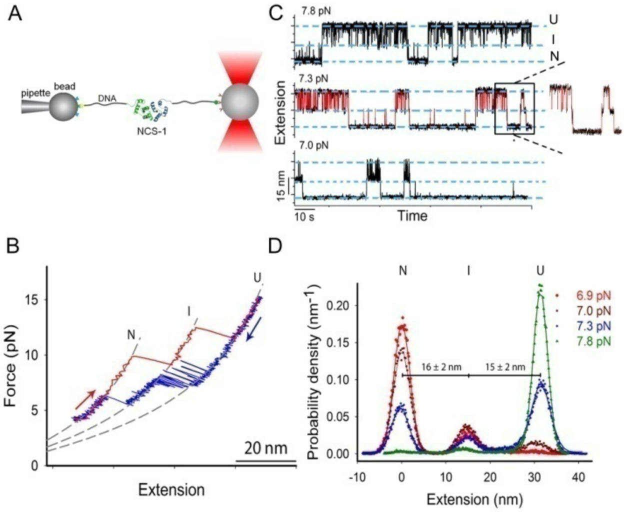


Figure 1.Folding transitions of the Mg^{2+}-bound state of NCS-1. A) Sketch of the experimental setup. Single NCS-1 molecules were manipulated with functionalized polystyrene beads, as previously described (29, 34). B) Force vs extension cycle acquired by stretching (red, 100 nm/sec) and relaxing (blue, 20 nm/sec) a NCS-1 molecule in presence of 10 mM Mg^{2+}. The native state (*N*) transits into the unfolded state (*U*) by populating an intermediate state (*I*), which is visited by the protein also during refolding. During the slow relaxation process fast fluctuations between *U* and *I* are detectable in a narrow range of forces. Color-coded arrows indicate the pulling (red) and relaxing (blue) directions. The stretching trace was fit to the WLC model (dashed lines) to estimate the changes in contour length associated to each unfolding transition. C) Extension vs time traces for the Mg^{2+}-bound state of NCS-1, acquired at three different constant forces. The molecule hops between *N*, *I* and *U* at equilibrium, with the unfolded state being increasingly populated at higher forces. A close-up view of the transitions in the dashed box is shown, along with the corresponding fit based on HMM (red line). D) Probability density distributions (symbols) of the extensions of traces acquired at different constant forces. The distributions were fitted to triple-Gaussian functions (lines) to estimate a change in extension of 16 ± 2 nm for the *N*-*I* transition and

of 15 ± 2 nm for the *I-U* transition.

During both stretching and relaxation, at a force of about 10 pN, the molecule populates an intermediate state *I*, located by extension half way between the native (*N*) and the unfolded (*U*) states. At slow relaxation speed, fast fluctuations between *U* and *I* can be observed in a narrow range of forces. The C- and N-domains of NCS-1 comprise a similar number of residues and in the Ca^{2+}-bound state they unfold and refold sequentially (11). It is thus conceivable that the transitions observed in our experiments are the result of the sequential unfolding and refolding of the two domains. To confirm this, we analyzed 110 stretching traces with the worm-like-chain model of polymer elasticity (29, 35, 36) to estimate the change in contour length (ΔLc) associated with the different transitions. This analysis provided a ΔLc of 30 ± 3 nm for the *N-I* transition, and a ΔLc of 31 ± 2 nm for the *I-U* transition. These values are very close to the nominal ΔLc for the unfolding of the C-domain ($\Delta L_c^0 = 33$ nm) and N-domain ($\Delta L_c^0 = 30$ nm) of the Ca^{2+}-bound state of NCS-1 (11) and are thus consistent with a sequential unfolding of the two domains in the Mg^{2+}-bound state. Nonetheless, the measured $\boldsymbol{\Delta L_c}$ values are undistinguishable within the experimental errors and cannot be used for structural assignment. We thus investigated the behavior of two NCS-1 variants, NCS-1^{10Gly} and NCS-$1^{38\text{-}188}$, where the size of the two domains was selectively altered. In NCS-1^{10Gly} the size of the C-domain was increased by ~10 % by inserting 10 glycine residues into an unstructured loop connecting EF-3 and EF-4. In NCS-$1^{38\text{-}188}$ one handle is attached to position 38 and force is applied to only 56 residues of the N-domain (from Ser38 to Arg94), effectively reducing the size of the N-domain by ~ 40%. Compared to NCS-$1^{4\text{-}188}$, both for unfolding and refolding, the high force transitions were longer in NCS-1^{10Gly}

and the low force transitions shorter in NCS-$1^{38\text{-}188}$ (Fig.S2). These data, along with the WLC-model fitting results, suggest that the C-domain is mechanically more stable than the N-domain, and unfolds and refolds at higher forces. This behavior is consistent with that of the Ca^{2+}-bound state of NCS-1 (11).

Equilibrium unfolding/refolding transitions of the Mg^{2+}-bound state of NCS-1

The equilibrium unfolding/refolding processes of the Mg^{2+}-bound state of NCS-1 were characterized through constant force experiments, where the applied tension is kept constant by a feedback mechanism while the extension of the protein, as it jumps between different conformations, is monitored over time (Fig.1C). We acquired 50 extension-time traces in the force range between 6.7 pN and 8.0 pN using 20 different molecules, and we analyzed each trace with a three-state HMM (37-39) to determine the force-dependent transition rates. The resulting data were fit to a linearized form of the Bell's model (40) to estimate the position of the transition states along the reaction coordinate and the zero force rates (Fig. 2A). This analysis yielded a kinetic distance of 15 ±1 nm between *N* and *I*, and a kinetic distance of 13 ±1 nm between *I* and *U* (Table 1).

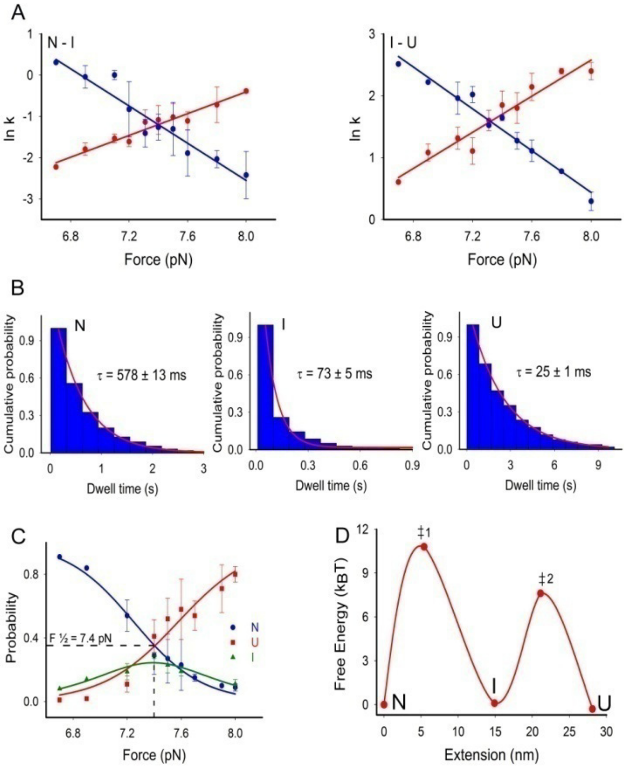


Figure 2. Energetics and kinetics of the Mg^{2+}-bound state of NCS-1. A) Unfolding (red) and refolding (blue) rate constants for the *N-I* (left panel) and *I-U* (right panel) transitions estimated from extension-time traces acquired at different forces. Error bars represent standard deviations. The data were fit to the Bell's model (solid lines) to estimate the positions of the transition states and the zero force rates. B) The dwell time distributions of the *N*, *I* and *U* states follow single exponential distributions. C) Force-dependence of the occupation probability of the *N*, *I* and *U* states (color coded), and their corresponding fits based on a two-state model (Supporting Material). The best fit values for the differences in extension and free energy between *N* and *U*, ΔX_{N-U} = 31.2 ± 2.6 nm and ΔG_{N-U} =16.6 ± 2.8 kcal/mol, are in agreement with the results of the HMM analysis, Table 1 and 2. The equilibrium force ($F_{1/2}$=7.4 pN) corresponding to a probability of 0.33 is indicated. D) Sketch of the free energy landscape of the Mg^{2+}-bound state of NCS-1 at $F_{1/2}$. The positions and free energies of the *U*, *I*, *N* and transition states (dagger1 and dagger2) states are shown.

Transition	x_u	x_f	k^0_u (s^{-1})	k^0_f (s^{-1})	$k_{1/2}$ (s^{-1})	$\Delta G^{\ddagger}_{1/2/}$
N-I	5.4±0.8	9.5±0.8	2(±0.8)x10^{-5}	5.4(±0.8) x10^{6}	0.31(±0.8)	10.8±
I-U	6.2±0.5	7±0.5	1.1(±1) x 10^{-4}	1.1(±1) x 10^{6}	4.98(±0.2)	7.5±0.

Table 1.Kinetic parameters of the folding reactions of the Mg^{2+}-bound state of NCS-1. The positions of the transition state for the unfolding and refolding reactions (x_u *and* x_f, respectively), and the rates at zero force (k^0_u *and* k^0_f, respectively) were estimated from

the force-dependence of the transition rates. The transition rate and the activation free energy at $F_{1/2}$ are indicated as $k_{1/2}$ and $\Delta G^{\ddagger}_{1/2}$, respectively. Errors are estimated from fit parameters' uncertainties. $k_B T = 0.6$ kcal/mol.

NCS-1 variant	$\Delta G_{N\text{-}U}$ *(kcal/mol)*	$\Delta X_{N\text{-}U}$
Mg^{2+}-bound NCS-$1^{4\text{-}188}$	16.6 ± 2.8	31.2 ± 2.6
Apo NCS-$1^{4\text{-}188}$	8.5 ± 1.4	15 ± 1.5
Apo NCS-$1^{95\text{-}188}$	5.96 ± 0.94	14.8 ± 2.3

Table 2. Thermodynamic parameters of the folding reactions of the apo and Mg^{2+}-bound states of NCS-1. The change in free energy and extension between the folded and unfolded states ($\Delta G_{N\text{-}U}$ and $\Delta X_{N\text{-}U}$, respectively) were estimated from the force-dependence of the occupation probabilities of the different molecular states. Errors are estimated from fit parameters' uncertainties

These distances compare well with the jumps in the molecule's extension observed in our constant-force traces (Fig. 1D), showing the overall consistency of our HMM analysis. The dwell time distributions of the *N*, *I* and *U* states as determined from the HMM analysis could be well fitted to single exponentials, in agreement with the kinetics of two consecutive two-state processes (Fig.2B). To further characterize the thermodynamics of the observed transitions we analyzed the force-dependence of the occupation probability of each molecular state, Fig. 2C (see Supplementary information for details). The results of this analysis (the equilibrium force $F_{1/2}$, the change in free energy ΔG, and the distances between the states) are in agreement with those of the HMM analysis (Table 2). From the rates and transition state positions, the salient features of the energy landscape of the Mg^{2+}-bound state of NCS-1 at the equilibrium force $F_{1/2}$ were reconstructed (Fig. 2D). The total change in free energy between *N* and *U* at zero external force ($\Delta G_{N\text{-}U}$), calculated by taking into account the free energy reduction of the unfolded state due to tethering (11, 41), is -16 ± 3 kcal/mol (Supporting

Material). This value compares well with the $\Delta G_{N\text{-}U}$ of -11 ± 2 kcal/mol measured in bulk experiments (Fig. S1) and with the $\Delta G_{N\text{-}U}$ of -16 ± 3 kcal/mol determined from the force-dependence of the occupation probabilities (Fig. 2C, Supporting Material), showing the overall consistency of our analysis.

Folding mechanism of the apo state of NCS-1

In order to investigate the effect of Ca^{2+} and Mg^{2+} on the folding process of NCS-1 we performed constant-velocity experiments in the absence of divalent ions. Remarkably, under these experimental conditions, the apo-protein unfolds and refolds reversibly in a two-state process near 6.5 pN, displaying fast fluctuations between two different molecular extensions, Fig. 3A.

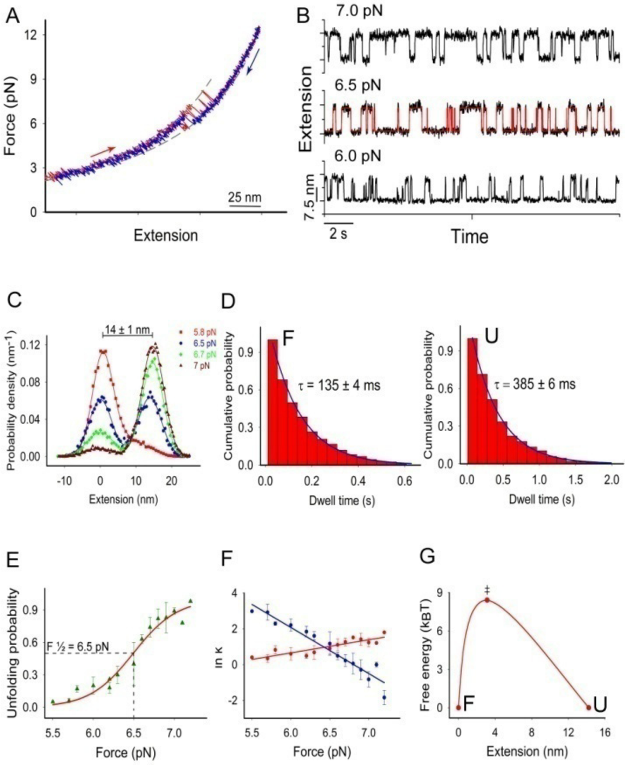


Figure 3. Folding transitions of the apo state of NCS-1. A) Force-extension cycle acquired by stretching (red) and relaxing (blue) a NCS-1 molecule in presence of no

divalent ions. Around 6.5 pN the molecule displays bistability, fluctuating between two different molecular conformations. The associated changes in contour lengths (ΔLc= 31 ± 2 nm) indicate that the discontinuities (rips) observed in the recorded traces are due to the unfolding and refolding of only half of the protein, the rest of the molecule giving rise to no detectable transitions. B) Time-extension traces acquired at the indicated constant forces. The protein hops at equilibrium between a compact (folded) and an extended (unfolded) conformation, with the latter one being favored at higher forces. The data were analyzed with a two-state HMM, which produced idealized state transition traces like the one shown in red at 6.5 pN. C) Extension probability density distributions of traces acquired at four different forces, and their corresponding double-Gaussian fittings (solid lines). The folded and unfolded states of the apo state of NCS-1 differ in extension by 14 ± 1 nm, and are equally populated at 6.5 pN. D) Dwell-time distributions of the unfolded and folded states at 6.2 pN and their corresponding single exponential fittings. E) Force-dependence of the occupation probability of the folded state and the corresponding two-state fitting. The best fit values, $\Delta X_{N\text{-}U}$ = 15 ± 1.5 nm and $\Delta G_{N\text{-}U}$ = 8.5 ± 1.4 kcal/mol, are in agreement with the results of the HMM analysis, Table 2 and 3. The equilibrium force ($F_{1/2}$ = 6.5 pN) corresponding to a probability of 0.5 is indicated. F) Unfolding (red) and refolding (blue) rate constants at different forces as determined from the HMM analysis, and fits with the Bell model (solid lines). Error bars represent standard deviations. G) Sketch of the free energy landscape of NCS-1 at $F_{1/2}$ in presence of no divalent ions, reconstructed using the HMM analysis.

The change in contour length associated with the unfolding/refolding events is 31 ±2 nm, suggesting that only half of the protein, possibly only one domain, gives rise to the observed transitions. The thermodynamics and kinetics of the bi-stability displayed by the molecule around 6.5 pN was analyzed through constant force experiments (Fig. 3B). Fifty two extension-time traces from 12 different molecules were acquired in the force range between 5.7 and 7.2 pN. In each trace, the protein hops between two distinct conformations differing in extension by 14 ±1 nm (Fig. 3C). The more extended one (unfolded) is favored by high forces. The life time distributions of both conformations can be well fit by single exponentials (Fig. 3D), and the unfolding probability displays a sigmoidal dependence on force (Fig. 3E); these data are consistent with a two-state unfolding/refolding mechanism. The transition rates and position of the transition state were estimated by fitting each time-extension trace with a two-state HMM (Fig. 3B, Table 3).

NCS-1 variants	x_u	x_f	$k_u^0(s^{-1})$	$k_f^0(s^{-1})$	$k_{1/2}\,(s^{-1})$	$\Delta G^{\ddagger}_{1/2}/k_BT$
NCS1$^{4-188}$	3.13 ± 0.4	11.1± 0.6	0.02 (±0.6)	8.9 (±1) x 10^{7}	2.72(±0.3)	8.4 ± 0.6
NCS1$^{95-188}$	3.3 ± 0.4	10.7 ± 0.7	0.01 (±0.6)	3.9 (±1) x 10^{7}	2.0(±0.3)	8.7 ± 0.6

Table 3. Kinetic parameters of the folding transitions of the apo state of NCS-1. The kinetic quantities were measured from the force-dependence of the transition rates. Errors are estimated from fit parameters' uncertainties. k_BT = 0.6 kcal/mol.

From the position of the transition state we deduced a difference in extension between the folded and unfolded states of 14 ±1 nm, which compares well with the experimental values (Fig. 3C). Using the kinetic parameters of Table 3 we reconstructed the energy landscape of the molecule at $F_{1/2}$. After correction for the free energy loss of the unfolded state due to tethering (Supporting Material), we estimated a $\Delta G_{U\text{-N}}$ of 7±1 kcal/mol, which compares well with the $\Delta G_{U\text{-}N}$ of 6±1 kcal/mol measured in bulk experiments (Fig. S3), and with the $\Delta G_{U\text{-}N}$ of 8.5±1.4 kcal /mol estimated from the unfolding probability (Fig. 3E), thereby supporting the robustness of the HMM analysis.

The C-domain of NCS-1 has proved to be mechanically more resistant (11) and thermodynamically more stable (19) than the N-domain, and it is thus conceivable that the reversible fluctuations observed in Fig. 3A originate from the unfolding and refolding of the C-domain only. To test this hypothesis we pulled on the mutant NCS-1$^{95-188}$, where the cysteine at position 4 is moved to position 95 to effectively apply force only on the C-domain. The resulting force-extension traces show reversible two-state unfolding/refolding transitions near 6.5 pN that closely resembled those observed when pulling on the entire molecule (Fig. 4A), with very similar changes in contour length ($\Delta L_c = 32 \pm 2$).

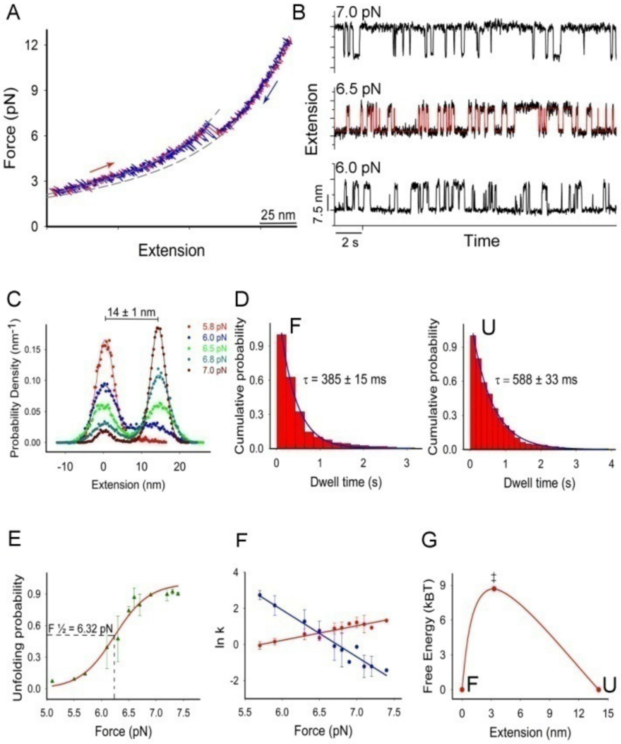


Figure 4. Folding transitions of NCS-1$^{95\text{-}188}$ in absence of divalent ions. A) Under these experimental conditions, when stretched (red trace) and relaxed (blue trace), NCS-1$^{95\text{-}188}$ displays reversible two-state fluctuations A) that closely resemble those observed with NCS-1$^{4\text{-}188}$. In fact, when held at constant force B), NCS-1$^{95\text{-}188}$ hops between two molecular conformations that are equal within experimental error to those described in Figure 3, in terms of extensions C), dwell time distributions D), force-dependent unfolding probability E), and transition rates F), Table 2 and 3. Error bars represent standard deviations. It follows that, in absence of divalent ions, the energy landscape of NCS-1$^{95\text{-}188}$ G) is nearly identical to that of NCS-1$^{4\text{-}188}$.

In fact, the unfolded and folded states of NCS-1$^{95\text{-}188}$ in the constant force measurements (Fig. 4B) showed the same extensions (Fig.4C), energies (Fig. 4E) and transition rates (Fig. 4F) of the corresponding states of NCS-1$^{4\text{-}188}$, revealing nearly identical energy landscapes for the two variants in the absence of divalent ions (Fig. 4G). These results clearly show that the unfolding and refolding events displayed by NCS-1 in the absence of divalent ions originate only from the C-domain, suggesting that under these experimental conditions the N-domain is either unstructured, or only loosely folded, and when stretched it unravels through transitions that are not detectable in our recordings. These data are quite similar to those obtained with the apo state of calmodulin, where only one domain exhibits reversible two-state transitions that can be easily observed in the force-extension curves, while the other gives rise to a signal that can only be vaguely inferred at very low forces (33).

DISCUSSION

Many members of the vast superfamily of EF-hand calcium binding proteins populate Mg^{2+}-bound states. Some of these states have distinct structural roles, where the binding of Mg^{2+} maintains the structure in the absence of Ca^{2+} and presumably protects against aggregation or premature degradation, whereas others have a direct functional role. In the case of NCS-1, the Mg^{2+}-bound state is the primary interacting conformation for many of its targets (10, 11) and its ion-dependent conformational dynamics therefore represents an important part of its functional repertoire. Our characterization of the folding of NCS-1 in the Mg^{2+}-bound and apo-states re-iterates our previous results obtained on the Ca^{2+}-bound state and show that the two EF-domains appear to fold as separate units regardless of ionic conditions. Using our results, we can now start to trace the progression of fold for NCS-1, from the likely transiently populated apo state, to resting Ca^{2+}-levels in a predominantly Mg^{2+}-bound state, to a fully Ca^{2+}-activated molecule during transient or sustained calcium elevation in neurons. The C-domain can already fold under apo conditions and is stabilized further by binding of Mg^{2+}, forming an intermediate state followed by Mg^{2+}-assisted folding of the N-domain into the Mg^{2+}-bound conformation. During transient or sustained elevation of Ca^{2+}, the molecule goes through a conformational change from a Mg^{2+}-bound state into a Ca^{2+}-activated state, where the largest structural changes are likely to be a consequence of the binding of Ca^{2+} into the ion free EF4 which could involve a reorientation of the structural domains for tighter coupling to expose the hydrophobic crevice, as in the related NCS protein recoverin (42). Whether or not this involves unfolding and subsequent Ca^{2+}-assisted refolding or conformational changes within a competitive ion-exchange process is not known from our current data. Nonetheless, as shown in this paper, the energy landscape of the protein significantly changes depending on the ionic environment to which it is exposed.

The Mg^{2+}-bound state of NCS-1

Our results indicate that Mg^{2+} promotes even more structuring than was previously proposed from NMR experiments (4). These latter studies suggested that Mg^{2+} binding led to considerably less structuring than Ca^{2+} binding as significant signal broadening was observed in the heteronuclear single-quantum coherence (^1H-^{15}N-HSQC) spectrum of NCS-1 and many peaks were missing, indicating conformational exchange (4). Remarkably, in this respect, the near-UV CD spectrum of NCS-1, which reports primarily on tertiary structure, does not change much upon Mg^{2+} binding to the apo state, whereas Ca^{2+} binding causes drastic changes (4). It is therefore quite surprising that our results indicate that the entire N-domain obtains structure upon Mg^{2+} binding, at least it is compacted to a state similar in shape as the Ca^{2+}-loaded state with unknown heterogeneity. However, even though the N-domain is unfolded under apo conditions in our accessible experimental force range (>2 pN), it cannot be excluded that it folds to a compact state at forces approaching zero. The mechanical properties of the Mg^{2+}-bound state closely resemble those of the Ca^{2+}-bound state. The N- and C-domains unfolds at forces (9 ± 1.2 pN and 12 ± 1.5 pN, respectively) that are only slightly lower than those observed with NCS-1 in the presence of Ca^{2+} (10.6 ± 0.6 pN and 13.8 ± 1 pN, respectively), and the associated changes in contour length are equal within the experimental error in the two cases. These data suggest that Mg^{2+} binding leads to a compact structure for NCS-1. It might also be that Mg^{2+} binding leads to a closed conformation that resembles that sampled by the apo state of NCS-1, at least with the C-domain. Mg^{2+} likely restricts the conformational space of the N-domain, but the dynamics of the ion-free EF4 could still lead to conformational exchange, reconciling at least some of the conflicting results obtained with NCS-1. In fact, almost 75% of the expected peaks are visible in the ^{1}H-^{15}N-HSQC of myristoylated NCS-1 (120 out of 174, (4)), consistent with one out of four EF-hands experiencing conformational exchange. We envision that Mg^{2+} binding leads to a fully compact globular yet dynamic structural ensemble with a partially open binding site and Ca^{2+} binding leads to reorientation and docking of the two domains adding an additional stability term, and exposing fully the binding site. This is discussed further below.

The folding reactions of the Mg^{2+}- and Ca^{2+}-bound states of NCS-1 share common features, but also display remarkable differences. In both cases, the N-domain folds via a two-state process, crossing transition-state barriers located at comparable positions along the reaction coordinate (11). Conversely, the C-domain follows quite different folding pathways depending on the ionic environment. In the presence of Mg^{2+}, the C-domain folds through a single all-or-none reaction, with a rate constant (k^{0}_{f}) of ~ 10^{6} s^{-1}. In contrast, in the presence of Ca^{2+}, the C-domain folds first into a partially folded conformation and then transits into its native state through a rate-limiting transition ($k^0{}_f \sim 38\ s^{-1}$) that involves calcium binding into the EF4 site. This latter process slows down the overall folding rate by several orders of magnitude and is linked to NCS-1's misfolding (21). These data suggest that in neurons, NCS-1 folds into its native structure more efficiently and safely at resting Ca^{2+}-levels in a predominantly Mg^{2+}-bound state, compared to its folding occurring during Ca^{2+}elevation.

The apo state of NCS-1

Our results suggest that in absence of divalent ions NCS-1 has a folded C-domain and an unstructured N-domain. One obvious difference between the two domains is a conserved Cys–Pro mutation in the N-domain that renders EF1 unable to bind divalent ions (19). It is well-known that EF-hands are commonly found in pairs and binding of ions makes the EF-domain pair stable. For NCS-1, however, the folding or unfolding of the EF-hand pairs seems conserved even in the absence of ions, as the EF hands of the N- and C-domains are either both folded or unfolded, displaying a considerable crosstalk between them.

In the apo-state, NCS-1 has been previously shown to have significant secondary structure yet little tertiary structure, and the ^1H-^{15}N HSQC spectrum has been shown to be severely signal-broadened, suggesting lack of stable structures (4, 22). Our equilibrium unfolding data (Fig. S3) shows an early decrease in fluorescence intensity at very low denaturant concentration. This decrease occurs gradually, lacking a pre-transition baseline, and indicates the unfolding of a structure with only marginal tertiary interactions, followed by a cooperative unfolding transition. These data acquired in bulk appear to be consistent with our single molecule data. In fact, under tension and in absence of divalent ions, NCS-1 unfolds and refolds through reversible two-state transitions that involved only the C-domain. The collapsed state of the N-domain suggested by the equilibrium unfolding experiments may be too unstable to be observable in the force range used in our experiments. We notice that the behavior under tension of the apo state is clearly different from that displayed by the Mg^{2+}- and Ca^{2+}-bound states of NCS-1 and by other native state structures stabilized by tertiary contacts, which unfold and refold out of equilibrium through transitions accompanied by hysteresis (11, 30, 43, 44). On the basis of our data, we speculate that the apo state of NCS-1 consist of a highly dynamic ensemble of similar structures, stabilized by weak interactions, which interconvert on intermediate NMR timescales, giving rise to the observed signal broadening in the HSQC spectrum (4), and to the reversible transitions in single molecule experiments. Thus, similar to intrinsically disordered proteins (45) and to that of the calcium binding S100 proteins (46), neural calcium sensors may exploit ligand-driven folding-upon binding to drive the thermodynamics for binding and the selection of binding partners.

Functional, ion-dependent dynamics of NCS-1

NCS-1 has a broad interactome currently consisting of more than 20 reported binding partners (47). The different protein partners can be bound predominantly in either domain or both at the same time, in either a calcium-dependent or independent manner. This multi-specificity requires a certain degree of conformational plasticity that is poorly understood on the molecular level. The different stabilities and folding rates of the two domains may reflect the conformational heterogeneity needed to allow for accommodation of multiple different ligands in the binding pocket, which spans through both domains. In addition, NCS-1 can gain functional versatility not only through the full or partial folding or unfolding of the two EF-domains but also through their coupling or decoupling, which affects their relative orientation and dynamics. This allostery may fine-tune the exposure of the binding sites for the various interaction partners of NCS-1.

The interaction of NCS-1 with the D2 dopamine receptor (D2R) is believed to include both the Mg^{2+}-bound and Ca^{2+}-bound states (13). NCS-1 interacts with the C-terminal tail residues of D2R independent of calcium and upon calcium activation it binds the G-protein coupled receptor kinase (GRK2) and partially inhibits it from phosphorylating the third intracellular loop of D2R, thus being responsible for its desensitization (13). Presumably, D2R and GRK2 are bound to separate NCS-1 domains. No detailed structural information on the interactions are available but biochemical and structural models have suggested different scenarios regarding the stoichiometry and domain involvement in binding (48, 49). Our results indicate that the extension and folding rate of the N-domain is similar in its Mg^{2+}-bound and Ca^{2+}-bound states It could therefore be speculated that the N-domain is responsible for binding D2R since its properties may not change much when transiting from Mg^{2+} to Ca^{2+} conditions. Upon calcium activation, the N-domain-D2R interaction may become tighter and the conformational changes of the C-domain, mostly through binding of calcium into EF4, enable further coupling with the N-domain, which may fully expose the second binding site for GRK2. This would be consistent with evidence which suggest that occupancy of calcium in EF2 is sufficient for the NCS-1-D2R interaction (49). In this respect, an important next step to understand these processes in detail will be to directly observe and characterize the activation of NCS-1 from a predominantly Mg^{2+}-bound state to the fully calcium-bound state. This would elucidate for example whether it is possible to traverse directly from an Mg^{2+}-bound state to a Ca^{2+}-bound state, or whether partial or full unfolding is a prerequisite for calcium activation. This will be especially important for the D2R-NCS-1 interaction as it has been suggested as a potential target for antipsychotic drug development (50). To this end our results on the conformational dynamics of NCS-1 may form a platform for drug design.

In conclusion, we have determined the salient features of the folding free energy landscapes of the Mg^{2+}-bound and apo states of NCS-1. We found these states to have distinct folding properties from the calcium-bound state, where NCS-1 in presence of magnesium folds each domain separately through all-or-none transitions whereas it in the apo state only folds its C-domain while the N-domain remains unfolded or highly destabilized. We found that the Mg^{2+}-bound state is more structured than previously described, displaying mechanical properties, cooperativity and end-to-end distances that closely resemble those of the Ca^{2+}-bound state. Our results represent an important detailed decomposition of states for the deciphering of the complete conformational as well as functional dynamics of NCS-1.

EXPERIMENTAL PROCEDURES

Protein –DNA chimera preparation

To manipulate NCS-1 with optical tweezers we engineered double-cysteine constructs using either a wild-type pET-16b or a pseudo-wild-type pET-16b expression plasmid (with Cys38 replaced by serine), by standard genetic techniques. The *E. coli* strain BL21 (DE3) was used to express un-myristoylated each NCS-1 construct, and cells were grown at 37 °C in Luria-Bertani medium. The protein was purified as described elsewhere (19, 51), and attachment of DNA to proteins and coupling of protein-DNA chimeras to beads was performed exactly as in previous work (34).

Force Spectroscopy Measurements

All measurements were carried out using a custom-built dual-beam optical tweezers set up that operates by direct measurement of light momentum (43, 52). Measurements were performed either in 10 mM Tris, 250 mM NaCl, 10 mM $MgCl_2$, pH 7.0 or, to study the apo state of NCS-1, in the same buffer without $MgCl_2$ and with 0.5 mM ethylene glycol tetraacetic acid (EGTA). The DNA-protein chimeras were manipulated with a 3.10 μm antidigoxigenin coated bead (Spherotec) held in the optical trap, and a 2.18 μm streptavidin-coated bead (Spherotec) held at the end of a micropipette by suction. The molecules were stretched and relaxed by moving the pipette toward the optical trap or away from it, respectively, by means of a piezoelectric flexure stage (MAX311/M, Thorlabs, Newton, NJ). Only those molecules displaying the characteristic DNA overstretching transition at 67 pN were used for acquiring data (29). In constant-speed experiments the data were acquired at a rate of 40 Hz, while in constant-force experiments at a rate of 100 Hz (43).

Hidden Markov Model Analysis

We use a hidden Markov model (HMM) algorithm (53, 54) to reconstruct from the constant-force hopping traces of our experiments the distribution (which we assume to be Gaussian) of the extension of each state and the transition probabilities between different states. The transition rate matrix K is computed from the transition probability matrix T by means of the relation $K = (\ln T) / \Delta t$, where $\Delta t = 0.01$ s for our data acquisition system. By collecting data obtained at different values of the force, we derive the chevron plots shown in Figs. 2A, 3F, 4F, which we interpret according to the Kramers-Bell theory (40, 55) to deduce the free energy differences among the states and the position of the barriers. The thermodynamics and kinetics information obtained in this way are summarized by the free-energy landscape sketch in Fig. 2D. Further details and references about HMM and the landscape reconstruction procedure are given in the Supporting Material.

Author contributions

M.M.N, P.O.H., B.B.K. and C.C. designed research; M.M.N, P.O.H., and M.R.O. performed experiments; A.M. contributed new analytic tools; B.B. contributed protein samples; M.M.N, P.O.H., A.M., M.R.O. and C.C. analyzed data; M.M.N. performed hidden Markov model analysis and M.M.N, P.O.H., A.M., B.B.K., and C.C. wrote the manuscript;

Acknowledgments

P.O.H. and B.B.K. acknowledge the Carlsberg Foundation and the Lundbeck Foundation for financial support. C.C. gratefully acknowledges financial support from Fondazione Cassa di Risparmio di Modena, the EU through a Marie Curie International Re-Integration Grant (n. 44952), the Italian MIUR (Grant n. 17DPXLNBEK), and partial support from Italian MIUR FIRB RBPR05JH2P "ITALNANONET". M.R.O. gratefully acknowledges financial support from the Abdus Salam International Centre for Theoretical Physics, of UNESCO and IAEA, (ICTP), TRIL program. Dr. Luca Bellucci and Dr. Stefano Corni are thanked for critical comments on the manuscript.

Supporting Citations

References (53-55) appear in the supporting material.

REFERENCES

1. Gifford, J. L., M. P. Walsh, and H. J. Vogel. 2007. Structures and metal-ion-binding properties of the Ca2+-binding helix-loop-helix EF-hand motifs. The Biochemical journal 405:199-221.
2. Grabarek, Z. 2011. Insights into modulation of calcium signaling by magnesium in calmodulin, troponin C and related EF-hand proteins. Biochimica et biophysica acta 1813:913-921.
3. Zot, H. G., and J. D. Potter. 1982. A structural role for the Ca2+-Mg2+ sites on troponin C in the regulation of muscle contraction. Preparation and properties of troponin C depleted myofibrils. The Journal of biological chemistry 257:7678-7683.
4. Aravind, P., K. Chandra, P. P. Reddy, A. Jeromin, K. V. R. Chary, and Y. Sharma. 2008. Regulatory and structural EF-hand motifs of neuronal calcium sensor-1: Mg2+ modulates Ca2+ binding, Ca2+-induced conformational changes, and equilibrium unfolding transitions. J Mol Biol 376:1100-1115.
5. Peshenko, I. V., and A. M. Dizhoor. 2006. Ca2+ and Mg2+ binding properties of GCAP-1. Evidence that Mg2+-bound form is the physiological activator of photoreceptor guanylyl cyclase. The Journal of biological chemistry 281:23830-23841.
6. Wingard, J. N., J. Chan, I. Bosanac, F. Haeseleer, K. Palczewski, M. Ikura, and J. B. Ames. 2005. Structural analysis of Mg2+ and Ca2+ binding to CaBP1, a neuron-specific regulator of calcium channels. The Journal of biological chemistry 280:37461-37470.
7. Brunet, S., T. Scheuer, R. Klevit, and W. A. Catterall. 2005. Modulation of CaV1.2 channels by Mg2+ acting at an EF-hand motif in the COOH-terminal domain. The Journal of general physiology 126:311-323.
8. Osawa, M., A. Dace, K. I. Tong, A. Valiveti, M. Ikura, and J. B. Ames. 2005. Mg2+ and Ca2+ differentially regulate DNA binding and dimerization of DREAM. The Journal of biological chemistry 280:18008-18014.
9. Peshenko, I. V., E. V. Olshevskaya, S. Yao, H. H. Ezzeldin, S. J. Pittler, and A. M. Dizhoor. 2010. Activation of retinal guanylyl cyclase RetGC1 by GCAP1: stoichiometry of binding and effect of new LCA-related mutations. Biochemistry 49:709-717.
10. Burgoyne, R. D., and L. P. Haynes. 2012. Understanding the physiological roles of the neuronal calcium sensor proteins. Mol Brain 5.
11. Heidarsson, P. O., M. R. Otazo, L. Bellucci, A. Mossa, A. Imparato, E. Paci, S. Corni, R. Di Felice, B. B. Kragelund, and C. Cecconi. 2013. Single-Molecule Folding Mechanism of an EF-Hand Neuronal Calcium Sensor. Structure 21:1812-1821.

12. Burgoyne, R. D. 2007. Neuronal calcium sensor proteins: generating diversity in neuronal Ca2+ signalling. Nat Rev Neurosci 8:182-193.
13. Kabbani, N., L. Negyessy, R. Lin, P. Goldman-Rakic, and R. Levenson. 2002. Interaction with neuronal calcium sensor NCS-1 mediates desensitization of the D2 dopamine receptor. The Journal of neuroscience : the official journal of the Society for Neuroscience 22:8476-8486.
14. Navarro, G., J. Hradsky, C. Lluis, V. Casado, P. J. McCormick, M. R. Kreutz, and M. Mikhaylova. 2012. NCS-1 associates with adenosine A(2A) receptors and modulates receptor function. Frontiers in molecular neuroscience 5:53.
15. Weiss, J. L., H. Hui, and R. D. Burgoyne. 2010. Neuronal Calcium Sensor-1 Regulation of Calcium Channels, Secretion, and Neuronal Outgrowth. Cell Mol Neurobiol 30:1283-1292.
16. Haynes, L. P., G. M. Thomas, and R. D. Burgoyne. 2005. Interaction of neuronal calcium sensor-1 and ADP-ribosylation factor 1 allows bidirectional control of phosphatidylinositol 4-kinase beta and trans-Golgi network-plasma membrane traffic. The Journal of biological chemistry 280:6047-6054.
17. Burgoyne, R. D., D. W. O'Callaghan, B. Hasdemir, L. P. Haynes, and A. V. Tepikin. 2004. Neuronal Ca2+-sensor proteins: multitalented regulators of neuronal function. Trends in neurosciences 27:203-209.
18. Bourne, Y., J. Dannenberg, V. Pollmann, P. Marchot, and O. Pongs. 2001. Immunocytochemical localization and crystal structure of human frequenin (neuronal calcium sensor 1). Journal of Biological Chemistry 276:11949-11955.
19. Heidarsson, P. O., I. J. Bjerrum-Bohr, G. A. Jensen, O. Pongs, B. E. Finn, F. M. Poulsen, and B. B. Kragelund. 2012. The C-Terminal Tail of Human Neuronal Calcium Sensor 1 Regulates the Conformational Stability of the Ca2+-Activated State. Journal of Molecular Biology 417:51-64.
20. Bellucci, L., S. Corni, R. Di Felice, and E. Paci. 2013. The Structure of Neuronal Calcium Sensor-1 in Solution Revealed by Molecular Dynamics Simulations. Plos One 8.
21. Heidarsson, P. O., M. M. Naqvi, M. R. Otazo, A. Mossa, B. B. Kragelund, and C. Cecconi. 2014. Direct single-molecule observation of calcium-dependent misfolding in human neuronal calcium sensor-1. Proceedings of the National Academy of Sciences of the United States of America 111:13069-13074.
22. Cox, J. A., I. Durussel, M. Comte, S. Nef, P. Nef, S. E. Lenz, and E. D. Gundelfinger. 1994. Cation binding and conformational changes in VILIP and NCS-1, two neuron-specific calcium-binding proteins. The Journal of biological chemistry 269:32807-32813.
23. Kuboniwa, H., N. Tjandra, S. Grzesiek, H. Ren, C. B. Klee, and A. Bax. 1995. Solution structure of calcium-free calmodulin. Nature structural biology 2:768-776.
24. Bustamante, C., Y. R. Chemla, N. R. Forde, and D. Izhaky. 2004. Mechanical processes in biochemistry. Annual review of biochemistry 73:705-748.
25. Ferreon, A. C., and A. A. Deniz. 2011. Protein folding at single-molecule resolution. Biochimica et biophysica acta 1814:1021-1029.
26. Heidarsson, P. O., M. M. Naqvi, P. Sonar, I. Valpapuram, and C. Cecconi. 2013. Conformational dynamics of single protein molecules studied by direct

mechanical manipulation. Advances in protein chemistry and structural biology 92:93-133.

27. Moffitt, J. R., Y. R. Chemla, S. B. Smith, and C. Bustamante. 2008. Recent advances in optical tweezers. Annual review of biochemistry 77:205-228.
28. Zoldák, G., and M. Rief. 2013. Force as a single molecule probe of multidimensional protein energy landscapes. Curr Opin Struc Biol 23:48-57.
29. Cecconi, C., E. A. Shank, C. Bustamante, and S. Marqusee. 2005. Direct observation of the three-state folding of a single protein molecule. Science 309:2057-2060.
30. Stigler, J., F. Ziegler, A. Gieseke, J. C. Gebhardt, and M. Rief. 2011. The complex folding network of single calmodulin molecules. Science 334:512-516.
31. Schlierf, M., F. Berkemeier, and M. Rief. 2007. Direct observation of active protein folding using lock-in force spectroscopy. Biophysical journal 93:3989-3998.
32. Schlierf, M., H. Li, and J. Fernandez. 2004. The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. Proceedings of the National Academy of Sciences of the United States of America 101:7299-7304.
33. Stigler, J., and M. Rief. 2012. Calcium-dependent folding of single calmodulin molecules. Proceedings of the National Academy of Sciences of the United States of America 109:17814-17819.
34. Cecconi, C., E. Shank, F. Dahlquist, S. Marqusee, and C. Bustamante. 2008. Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers. Eur Biophys J 37:729-738.
35. Bustamante, C., Z. Bryant, and S. B. Smith. 2003. Ten years of tension: single-molecule DNA mechanics. Nature 421:423-427.
36. Bustamante, C., J. F. Marko, E. D. Siggia, and S. Smith. 1994. Entropic elasticity of lambda-phage DNA. Science 265:1599-1600.
37. Chodera, J. D., P. Elms, F. Noé, B. Keller, C. M. Kaiser, A. Ewall-Wice, S. Marqusee, C. Bustamante, and N. S. Hinrichs. 2011. Bayesian hidden Markov model analysis of single-molecule force spectroscopy: Characterizing kinetics under measurement uncertainty. arXiv preprint arXiv:1108.1430.
38. Gao, Y., G. Sirinakis, and Y. Zhang. 2011. Highly anisotropic stability and folding kinetics of a single coiled coil protein under mechanical tension. Journal of the American Chemical Society 133:12749-12757.
39. McKinney, S. A., C. Joo, and T. Ha. 2006. Analysis of single-molecule FRET trajectories using hidden Markov modeling. Biophysical Journal 91:1941-1951.
40. Bell, G. I. 1978. Models for the specific adhesion of cells to cells. Science 200:618-627.
41. Liphardt, J., B. Onoa, S. B. Smith, I. Tinoco, and C. Bustamante. 2001. Reversible unfolding of single RNA molecules by mechanical force. Science 292:733-737.
42. Yap, K. L., J. B. Ames, M. B. Swindells, and M. Ikura. 1999. Diversity of conformational states and changes within the EF-hand protein superfamily. Proteins 37:499-507.
43. Heidarsson, P. O., I. Valpapuram, C. Camilloni, A. Imparato, G. Tiana, F. M. Poulsen, B. B. Kragelund, and C. Cecconi. 2012. A highly compliant protein

native state with a spontaneous-like mechanical unfolding pathway. Journal of the American Chemical Society 134:17068-17075.

44. Shank, E. A., C. Cecconi, J. W. Dill, S. Marqusee, and C. Bustamante. 2010. The folding cooperativity of a protein is controlled by its chain topology. Nature 465:637-640.

45. Sugase, K., H. J. Dyson, and P. E. Wright. 2007. Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 447:1021-U1011.

46. Botelho, H. M., M. Koch, G. Fritz, and C. M. Gomes. 2009. Metal ions modulate the folding and stability of the tumor suppressor protein S100A2. Febs J 276:1776-1786.

47. Lian, L. Y., S. R. Pandalaneni, P. A. Todd, V. M. Martin, R. D. Burgoyne, and L. P. Haynes. 2014. Demonstration of Neuronal Calcium Sensor-1 binding to the Cav2.1 P/Q-type calcium channel. Biochemistry.

48. Lian, L. Y., S. R. Pandalaneni, P. Patel, H. V. McCue, L. P. Haynes, and R. D. Burgoyne. 2011. Characterisation of the interaction of the C-terminus of the dopamine D2 receptor with neuronal calcium sensor-1. Plos One 6:e27779.

49. Woll, M., D. De Cotiis, M. Bewley, D. Tacelosky, R. Levenson, and J. Flanagan. 2011. Interaction between the D2 dopamine receptor and neuronal calcium sensor-1 analyzed by fluorescence anisotropy. Biochemistry 50:8780-8791.

50. Kabbani, N., M. P. Woll, J. C. Nordman, and R. Levenson. 2012. Dopamine receptor interacting proteins: targeting neuronal calcium sensor-1/D2 dopamine receptor interaction for antipsychotic drug development. Current drug targets 13:72-79.

51. Kragelund, B. B., A. Hauenschild, G. Carlstrom, O. Pongs, and B. E. Finn. 2000. 1H, 13C, and 15N assignments of un-myristoylated Ca2+-frequenin, a synaptic efficacy modulator. Journal of biomolecular NMR 16:85-86.

52. Smith, S. B., Y. J. Cui, and C. Bustamante. 2003. Optical-trap force transducer that operates by direct measurement of light momentum. Method Enzymol 361:134-162.

53. Petrie, L. E. B. a. T. 1966. Statistical inference for probabilistic functions of finite state Markov chains. The Annals of Mathematical Statistics:1554-1563.

54. Rabiner, L. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE.

55. Tinoco, I. 2004. Force as a useful variable in reactions: Unfolding RNA. Annu Rev Bioph Biom 33:363-385.

Supporting Material

Single-molecule folding mechanisms of the apo- and Mg^{2+}-bound states of human Neuronal Calcium Sensor-1

Mohsin M. Naqvi, Pétur O. Heidarsson, Mariela R. Otazo, Alessandro Mossa,

Birthe B. Kragelund, and Ciro Cecconi

Hidden Markov Model analysis and landscape reconstruction

Hidden Markov Model (HMM) techniques are a powerful statistical tool to analyze time series. The problem, frequently encountered in data analysis, is the following: assume that the system under investigation can be reasonably well approximated by a Markov chain with *n* states and time-independent transition matrix T_{ij} ($i, j = 1,\ldots,n$). Experiments made on the system can not directly assess which state the system is in (it is *hidden*), but each state emits a characteristic signal (either discretely or continuously distributed) which can be detected. From a sufficiently long time series of observations one can infer both the transition matrix elements and the characteristics of the signal emitted by each state, and then assign to each data point the most likely state of the system at that time. The algorithms to perform these reconstructions were first developed by mathematicians in the '60s (1) and then widely applied to artificial intelligence problems, notably speech recognition (2). Lately they have been adopted by the biophysical community as the tool of choice for analyzing data from single-molecule experiments, at first fluorescence (3) and then optical tweezers (4)

In our case, the states are N, I, U and the signal is the normally distributed end-to-end extension of the molecular construct protein + handles. From our HMM reconstruction, we estimate the elements of the transition matrix T_{ij}, which can be translated into transition rates using the formula $K_{ij} = (\ln M)_{ij}/\Delta t$, where Δt is the time interval between two consecutive measurements (in our experiment, 0.01 s). By collecting rates at different force values we obtain the chevron plots in Figs. 2A, 3F, 4F. Such plots are interpreted according to the Kramers-Bell theory (5) in the simplified version of (6) (see also (7) for a proof of the validity of this approximation):

$$k_{\rightarrow}(f) = k_0 \exp\left[\frac{f x^{\ddagger}_{\rightarrow}}{k_B T}\right]$$

$$k_{\leftarrow}(f) = k_0 \exp\left[\frac{\Delta G - f x^{\ddagger}_{\leftarrow}}{k_B T}\right]$$

$$k_0(f) = k_m \exp\left[\frac{\Delta G^{\ddagger}}{k_B T}\right]$$

From the slopes of the linear fits (in semi-log scale) performed in the chevron plots we can therefore deduce the position of the barriers, while the intercepts give access to the free energy difference between the states. In order to present a free energy landscape like the one in Fig. 2D we need one more data: the pre-exponential factor k_m, which we cannot directly measure with our experimental apparatus. We adopt the value 1.2 $\times 10^{-4}$ Hz, which has been experimentally measured in a setting very similar to ours, albeit with a different protein (8).

Thermodynamic Analysis of a three state system

If we have a three state system with phases *N, I* and *U* then the partition function (equation 45 in reference (9) can be written as:

$$Z = exp\left(-\frac{\Delta G_U(T,x_U) - Fx_U}{k_B T}\right) + exp\left(-\frac{\Delta G_I(T,x_I) - Fx_I}{k_B T}\right) + exp\left(-\frac{\Delta G_N(T,x_N) - Fx_N}{k_B T}\right) \quad (1)$$

Where x_U, x_I, x_N are the average extensions of the three states at temperature *T*, atmospheric pressure and applied force *F*, while

$$\Delta G_U = G_U - G_0$$

(2)

$$\Delta G_I = G_I - G_0$$

(3)

$$\Delta G_N = G_N - G$$

(4)

are the thermodynamic free energies at constant temperature, pressure and average extensions of the three states measured with respect to an arbitrary reference state. From the above relations the probability of a single molecule of being in state *N*, *I* and *U* at temperature *T* and force *F* are:

$$P_U = \frac{1}{Z}\exp\left[-\frac{\Delta G_U(T,x_U) - Fx_U}{k_B T}\right]$$

(5)

$$P_I = \frac{1}{Z}\exp\left[-\frac{\Delta G_I(T,x_I) - Fx_I}{k_B T}\right]$$

(6)

$$P_N = \frac{1}{Z}\exp\left[-\frac{\Delta G_N(T,x_N) - Fx_N}{k_B T}\right]$$

(7)

If we choose the reference state as I, then the partition function in Eq. (1) can be written as:

$$Z = \exp\left(-\frac{\Delta G_U(T,x_U) - Fx_U}{k_B T}\right) + \exp\left(\frac{Fx_I}{k_B T}\right) + \exp\left(-\frac{\Delta G_N(T,x_N) - Fx_N}{k_B T}\right)$$

(8) From which we can rewrite the equilibrium probabilities as:

$$P_U = \left[1 + \exp\left(\frac{\Delta G_{UI} - F\Delta x_{UI}}{k_B T}\right) + \exp\left(\frac{\Delta G_{UN} - F\Delta x_{UN}}{k_B T}\right)\right]^{-1}$$

(9)

$$P_I = \left[1 + \exp\left(-\frac{\Delta G_{UI} - F\Delta x_{UI}}{k_B T}\right) + \exp\left(\frac{\Delta G_{IN} - F\Delta x_{IN}}{k_B T}\right)\right]^{-1}$$

(10)

$$P_N = \left[1 + \exp\left(-\frac{\Delta G_{IN} - F\Delta x_{IN}}{k_B T}\right) + \exp\left(-\frac{\Delta G_{UN} - F\Delta x_{UN}}{k_B T}\right)\right]^{-1}$$

(11)

where we have defined the convenient shorthand notations:

$$\Delta G_{UI} = G_U - G_I = -\Delta G_{IU}\ ; \quad \Delta x_{UI} = x_U - x_I = -\Delta x_{IU}$$

$$\Delta G_{NI} = G_N - G_I = -\Delta G_{IN}\ ; \quad \Delta x_{NI} = x_N - x_I = -\Delta x_{IN}$$

$$\Delta G_{UN} = G_U - G_N = -\Delta G_{NU} ; \Delta x_{UN} = x_U - x_N = -\Delta x_{NU}$$

The fit to the occupation probabilities in Fig. 2C was obtained by fixing Δx_{UN} = 31 nm from the extension distributions in Fig. 1D while treating all the other quantities as adjustable parameters. The best fit was achieved with the values Δx_{UI} = 15 ± 2 nm, Δx_{NI} = 16 ± 1 nm, ΔG_{UI} = 16 ± 3 kcal/mol, ΔG_{NI} = 17 ± 1 kcal/mol, and ΔG_{UN} = 32 ± 3 kcal/mol.

Zero-force free energy for NCS-1 conformers

The comparison between the free energy differences calculated in single molecule experiments and in bulk experiments requires an estimate of the free energy needed for stretching the protein. If we indicate by $F_{1/2}$ the force at which states *U* and *N* have the same free energy, the zero-force free energy difference is (9, 10)

$$\Delta G_{UN}^{0} = F_{\frac{1}{2}}\Delta x_{UN} - \Delta G_{ST} = F_{\frac{1}{2}}\Delta x_{UN} - \int_{0}^{X_U} F(x')dx'$$

(12)

The elastic response of the protein is customarily represented by the worm like chain (WLC) model (11, 12) in which the relation between the force *F* applied to a molecule and its elongation *x* is given by

$$\frac{FP}{k_B T} = \frac{1}{4\left(1-\frac{x}{L}\right)^2} + \frac{x}{L} - \frac{1}{4}$$

(13)

Where P = 0.65 nm is the persistence length and L is the contour length, nominally computed as 0.36 nm per residue. Observe that $x_U = x_N + \Delta x_{UN}$, so for finding the zero-force free energy we need to know the length of the native state (x_N) of the protein. As such information is not available, we have used corresponding values from the Ca^{2+} bound state of NCS-1, whose structure is known (13), confiding in the fact that the resulting systematic error is likely to be small in comparison with other sources of uncertainty. An alternative approach is to directly estimate x_U by solving Eq. (13) at the force $F_{1/2}$: these two ways of estimating x_U agree reasonably well, as we show in the following results. In the case of the Mg^{2+}-bound state of NCS-1 the handles are attached to residues 4 and 188, so we have L = 66.2 nm while x_N = 3.5 nm from X-ray reconstruction of the Ca^{2+}-bound state. From Fig. 2C we get $F_{1/2}$ = 7.3 pN and from Fig. 1D we find Δx_{UN} = 31 nm. With these numbers, Eq. (12) gives ΔG_{UN}^{0} = 16 ± 3 kcal/mol, to be compared with the bulk value of 11 ± 2 kcal/mol. For the apo state only the C-domain (residues 95-188) is folded in the native state, so we set L = 33.5 nm and x_N = 1.32 nm. Using the values $F_{1/2}$ = 6.5 pN and Δx_{UN} = 14 nm from Figs. 3E and 3C, respectively, Eq. (12) yields ΔG_{UN}^{0} = 7.0 ± 1 kcal/mol, in good agreement with the bulk result 6 ± 1 kcal/mol. The alternative estimate of x_U from inversion of Eq. (13) at force $F_{1/2}$ yields x_U = 32.3 nm for the Mg^{2+}-bound state (compare with $x_N + \Delta x_{UN}$ = 34.5 nm) and x_U = 15.1 nm for the apo state (while $x_N + \Delta x_{UN}$ = 15.3 nm).

Equilibrium unfolding experiments

We monitored the denaturation of NCS-1 by incubating in guanidine hydrochloride (GuHCl) followed by measuring intrinsic tryptophan fluorescence emission spectroscopy using a Perkin-Elmer LS50B spectro fluorimeter. Samples of 1-5 µM protein were prepared in 10 mM Tris, 250 mM NaCl, 10mM $\mathrm{MgCl_2}$, 1 mM dithiothreitol, pH 7.0 with GuHCl concentrations in the range 0 - 8 M. The exact denaturant concentration of each sample was determined with a digital refractometer. All experiments were performed at ambient temperatures with excitation at 295 nm monitoring emission at 325 nm. The fraction of folded protein was calculated from the maximum and minimum intensities, and plotted against the concentration of denaturant. The data were fitted to three-state transitions as described (14).

$$f(x) = \frac{(a0+a1x)+(a2+a3x)*\exp\left(m1*\frac{(x-Cm1)}{RT}\right)}{1+\exp\left(m1*\frac{x-Cm1}{RT}\right)} + \frac{(a2+a3x)+(a4+a5x)*\exp\left(m2*\frac{(x-Cm2)}{RT}\right)}{1+\exp\left(m2*\frac{(x-Cm2)}{RT}\right.} \qquad (14)$$

where a0, a1, a2, a3, a4 and a5 are the linear influence of denaturant on the different states, N, [$D_N N_C$], D, respectively. At least 60 data points were used in each experiment.

Supplemental Figures

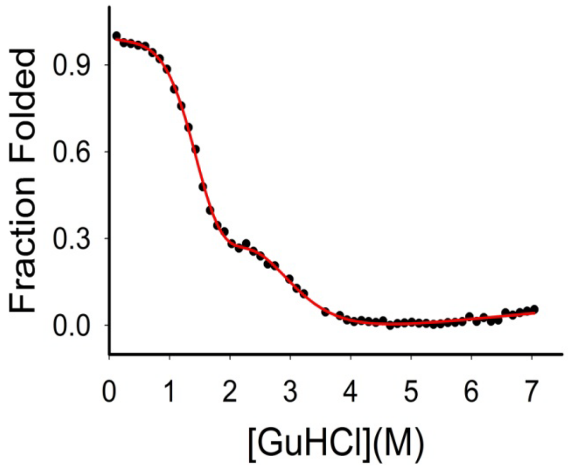


Figure S1.Equilibrium unfolding profile of wild type NCS-1 in 10mM $MgCl_2$.The solid line represents the best fit of the data to Eq. 14. The fitting of the data gave a global free energy change of unfolding of 11.0 ± 2.0 kcal/mol.

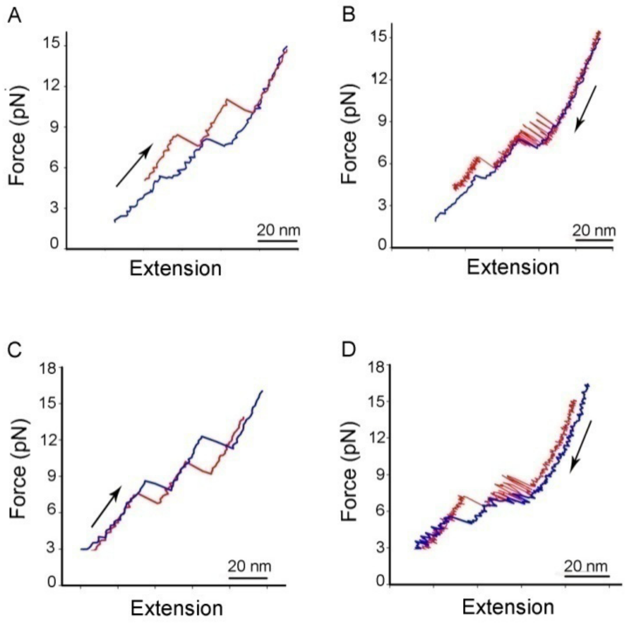


Figure S2. Asignment of unfolding/refolding transitions to individual domains. Stretching and relaxation traces for NCS-1$^{4-188}$ and two mutants designed to alter the size of the N- and C-domain individually. A) Superimposed stretching traces (100 nm/sec) acquired by pulling on NCS-1$^{4-188}$ (red) and NCS-1 $^{38-188}$ (blue). B) Superimposed relaxation traces (20 nm/sec) acquired by manipulating NCS-1$^{4-188}$ (red) and NCS-1 $^{38-188}$. Both in A) and B) the low force transition is smaller for NCS-1 $^{38-188}$. C) Superimposed stretching traces (100 nm/sec) acquired by pulling on NCS-1$^{4-188}$ (red) and NCS-1$^{10\text{ Gly}}$ (blue). D) Superimposed relaxation traces acquired by manipulating NCS-1$^{4-188}$ (red) and NCS-1$^{10\text{ Gly}}$. Both in C) and D) the high force transition is larger for NCS-1$^{10\text{ Gly}}$.

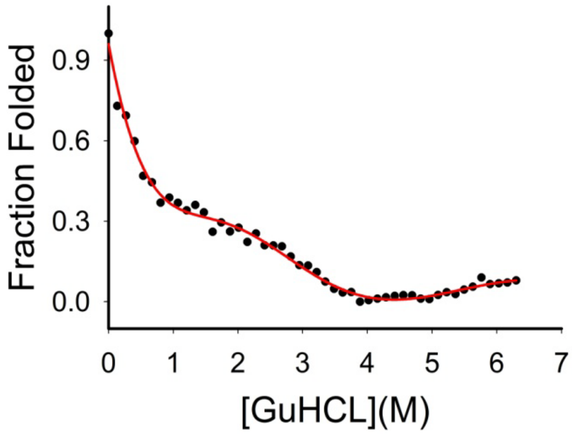


Figure S3.Equilibrium unfolding profile of wild type NCS-1 in absence of divalent ions. The solid line represents the best fit of the data to Eq. 14. The fitting of the data gave a global unfolding free energy change of 6.0 ± 1.0 kcal /mol.

SUPPORTING REFERENCES

1. Petrie LEBaT (1966) Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics* (37):1554-1563.
2. Rabiner L (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*.
3. McKinney SA, Joo C, & Ha T (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophysical Journal* 91(5):1941-1951.
4. Gao Y, Sirinakis G, & Zhang Y (2011) Highly anisotropic stability and folding kinetics of a single coiled coil protein under mechanical tension. *Journal of the American Chemical Society* 133(32):12749-12757.
5. Bell GI (1978) Models for the specific adhesion of cells to cells. *Science* 200(4342):618-627.
6. Tinoco I (2004) Force as a useful variable in reactions: Unfolding RNA. *Annu Rev Bioph Biom* 33:363-385.
7. Mossa A, Manosas M, Forns N, Huguet JM, & Ritort F (2009) Dynamic force spectroscopy of DNA hairpins: I. Force kinetics and free energy landscapes. *Journal of Statistical Mechanics: Theory and Experiment* 2009.
8. Gebhardt JC, Bornschlogl T, & Rief M (2010) Full distance-resolved folding energy landscape of one single protein molecule. *P Natl Acad Sci USA* 107(5):2013-2018.
9. Tinoco I & Bustamante C (2002) The effect of force on thermodynamics and kinetics of single molecule reactions. *Biophysical chemistry* 101-102:513-533.
10. Cecconi C, Shank EA, Bustamante C, & Marqusee S (2005) Direct observation of the three-state folding of a single protein molecule. *Science* 309(5743):2057-2060.
11. Bustamante C, Bryant Z, & Smith SB (2003) Ten years of tension: single-molecule DNA mechanics. *Nature* 421(6921):423-427.
12. Bustamante C, Marko JF, Siggia ED, & Smith S (1994) Entropic elasticity of lambda-phage DNA. *Science* 265(5178):1599-1600.
13. Heidarsson PO*, et al.* (2012) The C-Terminal Tail of Human Neuronal Calcium Sensor 1 Regulates the Conformational Stability of the Ca2+-Activated State. *Journal of Molecular Biology* 417(1-2):51-64.
14. Maxwell SL, Ho HY, Kuehner E, Zhao S, & Li M (2005) Pitx3 regulates tyrosine hydroxylase expression in the substantia nigra and identifies a subgroup of mesencephalic dopaminergic progenitor neurons during mouse development. *Developmental biology* 282(2):467-479.

3. LIST OF PUBLICATIONS

- *Direct single-molecule observation of calcium-dependent misfolding in human neuronal calcium sensor-1*. Pétur O. Heidarsson, **Mohsin M. Naqvi**, Mariela R. Otazo, Birthe B. Kragelund, and Ciro Cecconi. **PNAS**. August 08 2014. 10.1073/pnas.1401065111.

- *Structure, Folding Dynamics, and Amyloidogenesis of D76N β2-Microglobulin. Roles of Shear Flow, Hydrophobic Surfaces and – α Crystallin.*
  P. Patrizia Mangione, Gennaro Esposito, Annalisa Relini, Sara Raimondi, Riccardo Porcari, Sofia Giorgetti, Alessandra Corazza,Federico Fogolari, Amanda Penco, Yuji Goto, Young-Ho Lee, Hisashi Yagi,Ciro Cecconi, **Mohsin M. Naqvi**, Julian D. Gillmore, Philip N. Hawkins, Fabrizio Chiti, Ranieri Rolandi,Graham W. Taylor, Mark B. Pepys, Monica Stoppini, and Vittorio Bellotti**. *The Journal of Biological Chemistry*** .VOL. 288, NO. 43, pp. 30917–30930, October 25, 2013

- *Conformational dynamics of single protein molecules studied by direct mechanical manipulation*.
  Pétur O. Heidarsson**, Mohsin M. Naqvi**, Punam Sonar, Immanuel Valpapuram, Ciro Cecconi Advances in protein chemistry and structural biology 01/2013; 92:93-133. DOI:10.1016/B978-0-12-411636-8.00003-1.

- *Single-molecule folding mechanisms of the apo- and* Mg^{2+}*-bound forms of human Neuronal Calcium Sensor-1*. **Mohsin M. Naqvi,** Pétur O. Heidarsson, Mariela R. Otazo, Alessandro Mossa, Birthe B. Kragelund, and Ciro Cecconi. (submitted).

- *Proteolytic cleavage of transthyretin and disassembly of the tetramer: elucidation of an unexplored amyloidogenic pathway and effect of stabilizers* Julien Marcoux, Patrizia Mangione, Riccardo Porcari, Matteo Degiacomi, Graham W. Taylor, Sofia Giorgetti, Guglielmo Verona, Justin L. P. Benesch,Ciro Cecconi, **Mohsin M. Naqvi**, Julian D. Gillmore, Philip N. Hawkins, Monica Stoppini, Carol V. Robinson, Mark B. Pepys and Vittorio Bellotti. (in preparation).

CONFERENCES AND WORKSHOPS ATTENDED

- **Gordon Research Seminar**: **"Single Molecule Approaches to Biology"** July 12-13 2014 Lucca, Italy (**poster)**.

- **XX International Summer School "Nicolás Cabrera" Biomolecules Techniques And Single Molecule"**. July 21-26 2013. Madrid Spain. (**poster and oral**).

- **Workshop: "Tecniche di Microscopia Innovative e a Risoluzione Atomica"** sponsored by Department of Physics and Material Science of CNR, Institute SPIN Genova Italy.

4. CONCLUSIONS

This thesis is based on the single molecule folding and misfolding mechanisms of NCS-1 protein and the amyloidogenesis of the protein β-2 Microglubulin. The conclusions drawn from the experimental results of these projects (see: paper 1, paper 2 and paper 3) are as follows:

Direct observation of calcium-dependent misfolding in single Neuronal calcium sensor-1 molecules.

In this project (see paper: 1) the misfolding trajectories of NCS-1 were studied at single molecule level using dual beam optical tweezer set up. Our results have shown that NCS-1 enters off pathway or misfolded states from its on pathway intermediate state I2, which acts as a checkpoint between its native state and the misfolding trajectories. By doing force ramp experiments at various Ca^{2+} concentrations (0.5 µM to 10mM) it is shown that the probability of NCS-1 to enter the misfolding trajectories increases with increase in Ca^{2+} concentrations and pulling speeds. Using constant force method and hidden Markov model analysis we analyzed the kinetic and thermodynamic of the equilibrium fluctuations of NCS-1 between the I2 state and the misfolded states at 10mM and 10 µM concentrations. At 10mM concentration, we have characterized two misfolded states of NCS-1 - M1 and M2 while at 10 µM concentration only the M2 state was observed. The energy landscapes governing the misfolding trajectories were reconstructed in this study. Moreover, using mutants experiments: EF- 2 knockout, in which the Ca^{2+} binding site in EF-2 was removed by site directed mutagenesis and similarly for EF - 4 knockout, we have proposed that misfolded states (M1 and M2) are populated due to improper Ca^{2+} binding at both EF-2 and EF-4 binding sites of NCS-1. These results show that in vivo Ca^{2+} binding is responsible for both NCS-1 functions and disorders. This study will prove significant in understanding the mechanism of NCS-1 related neuronal disorders and in the development of new drugs.

Structure, folding dynamics, and amyloidogenesis of D76N β2-microglobulin. Roles of shear flow, hydrophobic surfaces and α crystalline.

In this study the factors responsible for the amyloidogenisis of the protein β-2 Microglubulin were investigated. The results have shown (paper II) that the rate of fibrillization of the structural variant D76N of β -2 Microglubulin, increases as the protein interacts with hydrophobic surfaces in presence of shear flow. However no significant change in the fibrillization of WT β-2 Microglubulin was observed in the above conditions. Since the WT β2-Microglobulin has a more stable native state than its variant D76N, this study shows that the thermodynamic stability of the molecule plays a crucial role in the fibrillization process and the initial step of the fibril formation is the partial denaturation of native state of the protein. The theoretical estimation of forces involved in the fibrillization of the protein has shown that hydrophobic effect plays a dominant role in the initial denaturation of the protein whereas shear forces help in the diffusion of the molecules from the bulk to the hydrophobic surfaces.

Single-molecule folding mechanisms of the apo- and Mg^{2+}-bound forms of human Neuronal calcium sensor-1.

In this project (see paper: 3) the folding (un) trajectories of NCS-1 protein were investigated in presence of Mg^{2+} and apo form (no divalent ions). NCS-1 Mg^{2+} bound form displayed a three state folding (un) behavior involving an intermediate state, half way between its native and fully unfolded states. As previously observed in presence of Ca^{2+} [94] the folding (un) of NCS-1 in Mg^{2+} takes place sequentially in which the N – domain unfolds first followed by the complete unfolding of the C domain. At a given force, using hidden Markov model analysis, the energy landscape of NCS-1 in Mg^{2+} was reconstructed. Our results have shown that NCS-1 in Mg^{2+} has lower thermodynamic stability than its Ca^{2+} bound form and the mechanical strength of the two conformers are same. Moreover, NCS-1 acquires a more compact structure in presence of Mg^{2+} then previously reported. In presence of no divalent ions (apo form), NCS-1 folds (un) reversibly in a two state manner. On pulling the C - domain of NCS -1 in apo form (NCS-1 $^{95\text{-}188}$ mutant) our results show that NCS-1 apo form has a compact C domain and an unstructured N domain, with least thermodynamic stability among the three conformers of NCS-1. This study will prove significant in deciphering the full conformational dynamics of NCS-1 as a function of changes in its ionic conditions.

5. BIBLIOGRAPHY

1 Dobson, C. M. (2003) Protein folding and misfolding. Nature. **426**, 884-890

2 Levinthal, C. (1968) ARE THERE PATHWAYS FOR PROTEIN FOLDING ? Journal de Chimie Physique, p. 44-45

3 Dill, K. A. and Chan, H. S. (1997) From Levinthal to pathways to funnels. Nature structural biology. **4**, 10-19

4 Dill, K. A., Ozkan, S. B., Shell, M. S. and Weikl, T. R. (2008) The protein folding problem. Annual Review of Biophysics. **37**, 289-316

5 Oliveberg, M. and Wolynes, P. G. (2005) The experimental survey of protein-folding energy landscapes. Q Rev Biophys. **38**, 245-288

6 Anfinsen, C. B. (1973) Principles that Govern the Folding of Protein Chains. Science, 223–230

7 Thirumalai, D., Klimov, D. K. and Woodson, S. A. (1997) Kinetic partitioning mechanism as a unifying theme in the folding of biomolecules. Theor Chem Acc. **96**, 14-22

8 Wensley, B. G., Batey, S., Bone, F. A. C., Chan, Z. M., Tumelty, N. R., Steward, A., Kwa, L. G., Borgia, A. and Clarke, J. (2010) Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. Nature. **463**, 685-U122

9 Woodside, M. T. and Block, S. M. (2014) Reconstructing folding energy landscapes by single-molecule force spectroscopy. Annu Rev Biophys. **43**, 19-39

10 Agard, B. N. a. D. A. (2008) How General Is The Nucleation–Condensation Mechanism? Proteins, 754–764.

11 Fersht, A. R. (1995) Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. P Natl Acad Sci USA. **92**, 10869-10873

12 Chaplin, M. (2006) Opinion - Do we underestimate the importance of water in cell biology? Nat Rev Mol Cell Bio. **7**, 861-866

13 Dyson, H., Wright, P. and Scheraga, H. (2006) The role of hydrophobic interactions in initiation and propagation of protein folding. Proceedings of the National Academy of Sciences of the United States of America. **103**, 13057-13061

14 Dill, K. A. and MacCallum, J. L. (2012) The Protein-Folding Problem, 50 Years On. Science. **338**, 1042-1046

15 Capaldi, A. P., Kleanthous, C. and Radford, S. E. (2002) Im7 folding mechanism: misfolding on a path to the native state. Nature structural biology. **9**, 209-216

16 Borgia, M. B., Borgia, A., Best, R. B., Steward, A., Nettels, D., Wunderlich, B., Schuler, B. and Clarke, J. (2011) Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. Nature. **474**, 662-665

17 Stigler, J., Ziegler, F., Gieseke, A., Gebhardt, J. C. and Rief, M. (2011) The complex folding network of single calmodulin molecules. Science. **334**, 512-516

18 Yu, H., Liu, X., Neupane, K., Gupta, A., Brigley, A., Solanki, A., Sosova, I. and Woodside, M. (2012) Direct observation of multiple misfolding pathways in a

single prion protein molecule. Proceedings of the National Academy of Sciences of the United States of America. **109**, 5283-5288

19 Gianni, S., Ivarsson, Y., De Simone, A., Travaglini-Allocatelli, C., Brunori, M. and Vendruscolo, M. (2010) Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. Nature structural & molecular biology. **17**, 1431-1437

20 Zheng, W. H., Schafer, N. P. and Wolynes, P. G. (2013) Frustration in the energy landscapes of multidomain protein misfolding. P Natl Acad Sci USA. **110**, 1680-1685

21 Chiti, F. and Dobson, C. (2006) Protein misfolding, functional amyloid, and human disease. Annual review of biochemistry. **75**, 333-366

22 Chiti, F. and Dobson, C. M. (2009) Amyloid formation by globular proteins under native conditions. Nat Chem Biol. **5**, 15-22

23 Dobson, C. (2003) Protein folding and misfolding. Nature. **426**, 884-890

24 Serpell, L. C., Sunde, M., Benson, M. D., Tennent, G. A., Pepys, M. B. and Fraser, P. E. (2000) The protofilament substructure of amyloid fibrils. J Mol Biol. **300**, 1033-1039

25 Saiki, M., Honda, S., Kawasaki, K., Zhou, D. S., Kaito, A., Konakahara, T. and Morii, H. (2005) Higher-order molecular packing in amyloid-like fibrils constructed with linear arrangements of hydrophobic and hydrogen-bonding side-chains. J Mol Biol. **348**, 983-998

26 Sunde, M. and Blake, C. (1997) The structure of amyloid fibrils by electron microscopy and X-ray diffraction. Advances in protein chemistry. **50**, 123-159

27 Buell, A., Galvagnion, C., Gaspar, R., Sparr, E., Vendruscolo, M., Knowles, T., Linse, S. and Dobson, C. (2014) Solution conditions determine the relative importance of nucleation and growth processes in α-synuclein aggregation. Proceedings of the National Academy of Sciences of the United States of America. **111**, 7671-7676

28 Chandler, D. (2005) Interfaces and the driving force of hydrophobic assembly. Nature. **437**, 640-647

29 Jean, L., Lee, C. F., Lee, C., Shaw, M. and Vaux, D. J. (2010) Competing discrete interfacial effects are critical for amyloidogenesis. Faseb J. **24**, 309-317

30 Jean, L., Lee, C. F. and Vaux, D. J. (2012) Enrichment of Amyloidogenesis at an Air-Water Interface. Biophys J. **102**, 1154-1162

31 Lee, C. F., Bird, S., Shaw, M., Jean, L. and Vaux, D. J. (2012) Combined Effects of Agitation, Macromolecular Crowding, and Interfaces on Amyloidogenesis. Journal of Biological Chemistry. **287**

32 Mangione, P. P., Esposito, G., Relini, A., Raimondi, S., Porcari, R., Giorgetti, S., Corazza, A., Fogolari, F., Penco, A., Goto, Y., Lee, Y. H., Yagi, H., Cecconi, C., Naqvi, M. M., Gillmore, J. D., Hawkins, P. N., Chiti, F., Rolandi, R., Taylor, G. W., Pepys, M. B., Stoppini, M. and Bellotti, V. (2013) Structure, folding dynamics, and amyloidogenesis of D76N beta2-microglobulin: roles of shear flow, hydrophobic surfaces, and alpha-crystallin. The Journal of biological chemistry. **288**, 30917-30930

33 Merlini, G. and Bellotti, V. (2003) Molecular mechanisms of amyloidosis - Reply. New Engl J Med. **349**, 1873-1873

34 Greenfield, N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure. Nat Protoc. **1**, 2876-2890

35 Kern, D., Eisenmesser, E. Z. and Wolf-Watz, M. (2005) Enzyme dynamics during catalysis measured by NMR spectroscopy. Nuclear Magnetic Resonance of Biological Macromolecules, Part C. **394**, 507-524

36 Maity, H., Maity, M., Krishna, M. M. G., Mayne, L. and Englander, S. W. (2005) Protein folding: The stepwise assembly of foldon units. Proceedings of the National Academy of Sciences of the United States of America. **102**, 4741-4746

37 Cecconi, C., Shank, E. A., Bustamante, C. and Marqusee, S. (2005) Direct observation of the three-state folding of a single protein molecule. Science. **309**, 2057-2060

38 Ferreon, A. C. M. and Deniz, A. A. (2011) Protein folding at single-molecule resolution. Bba-Proteins Proteom. **1814**, 1021-1029

39 Flannigan, D., Barwick, B. and Zewail, A. (2010) Biological imaging with 4D ultrafast electron microscopy. Proceedings of the National Academy of Sciences of the United States of America. **107**, 9933-9937

40 Othon, C., Kwon, O.-H., Lin, M. and Zewail, A. (2009) Solvation in protein (un)folding of melittin tetramer-monomer transition. Proceedings of the National Academy of Sciences of the United States of America. **106**, 12593-12598

41 Borgia, A., Williams, P. M. and Clarke, J. (2008) Single-molecule studies of protein folding. Annual review of biochemistry. **77**, 101-125

42 Bustamante, C. (2008) In singulo biochemistry: When less is more. Annual review of biochemistry. **77**, 45-50

43 Deniz, A. A., Mukhopadhyay, S. and Lemke, E. A. (2008) Single-molecule biophysics: at the interface of biology, physics and chemistry. J R Soc Interface. **5**, 15-45

44 Huang, S., Ratliff, K. S. and Matouschek, A. (2002) Protein unfolding by the mitochondrial membrane potential. Nature structural biology. **9**, 301-307

45 Cheolju Lee, M. P. S., Sumit Prakash, Masahiro Iwakura and Andreas Matouschek. (2001) ATP-Dependent Proteases Degrade Their Substrates by Processively Unraveling Them from the Degradation Signal. Molecular Cell. **7**, 627–637

46 Kenniston, J., Baker, T., Fernandez, J. and Sauer, R. (2003) Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. Cell. **114**, 511-520

47 Shtilerman, M., Lorimer, G. H. and Englander, S. W. (1999) Chaperonin function: folding by forced unfolding. Science. **284**, 822-825

48 Oberhauser, A. F., Marszalek, P. E., Erickson, H. P. and Fernandez, J. M. (1998) The molecular elasticity of the extracellular matrix protein tenascin. Nature. **393**, 181-185

49 Gautel, M. and Goulding, D. (1996) A molecular map of titin/connectin elasticity reveals two different mechanisms acting in series. FEBS letters. **385**, 11-14

50 Carrion-Vazquez, M., Oberhauser, A. F., Fowler, S. B., Marszalek, P. E., Broedel, S. E., Clarke, J. and Fernandez, J. M. (1999) Mechanical and chemical unfolding of a single protein: a comparison. P Natl Acad Sci USA. **96**, 3694-3699

51 Carrion-Vazquez, M., Li, H., Lu, H., Marszalek, P. E., Oberhauser, A. F. and Fernandez, J. M. (2003) The mechanical stability of ubiquitin is linkage dependent. Nature structural biology. **10**, 738-743

52 Rief, M., Pascual, J., Saraste, M. and Gaub, H. E. (1999) Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. J Mol Biol. **286**, 553-561

53 Best, R., Li, B., Steward, A., Daggett, V. and Clarke, J. (2001) Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation. Biophysical journal. **81**, 2344-2356

54 Yang, G., Cecconi, C., Baase, W. A., Vetter, I. R., Breyer, W. A., Haack, J. A., Matthews, B. W., Dahlquist, F. W. and Bustamante, C. (2000) Solid-state synthesis and mechanical unfolding of polymers of T4 lysozyme. P Natl Acad Sci USA. **97**, 139-144

55 Ashkin, A. (1970) Acceleration and Trapping of Particles by Radiation Pressure. Physical review letters. **24**, 156-&

56 Ashkin, A., Bjorkholm, J. E. and Chu, S. (1986) Caught in a Trap. Nature. **323**, 585-585

57 Ashkin, A. and Dziedzic, J. M. (1987) Optical Trapping and Manipulation of Viruses and Bacteria. Science. **235**, 1517-1520

58 Smith, S. B., Cui, Y. J. and Bustamante, C. (2003) Optical-trap force transducer that operates by direct measurement of light momentum. Method Enzymol. **361**, 134-162

59 Gordon, J. P. (1973) Radiation Forces and Momenta in Dielectric Media. Phys Rev A. **8**, 14-21

60 Cecconi, C., Shank, E., Dahlquist, F., Marqusee, S. and Bustamante, C. (2008) Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers. Eur Biophys J. **37**, 729-738

61 Cecconi, C., Shank, E., Marqusee, S. and Bustamante, C. (2011) DNA molecular handles for single-molecule protein-folding studies by optical tweezers. Methods in molecular biology (Clifton, N.J.). **749**, 255-271

62 Elms, P., Chodera, J., Bustamante, C. and Marqusee, S. (2012) The molten globule state is unusually deformable under mechanical force. Proceedings of the National Academy of Sciences of the United States of America. **109**, 3796-3801

63 Gao, Y., Sirinakis, G. and Zhang, Y. (2011) Highly anisotropic stability and folding kinetics of a single coiled coil protein under mechanical tension. Journal of the American Chemical Society. **133**, 12749-12757

64 Stigler, J. and Rief, M. (2012) Calcium-dependent folding of single calmodulin molecules. Proceedings of the National Academy of Sciences of the United States of America. **109**, 17814-17819

65 Heidarsson, P. O., Valpapuram, I., Camilloni, C., Imparato, A., Tiana, G., Poulsen, F. M., Kragelund, B. B. and Cecconi, C. (2012) A highly compliant protein native state with a spontaneous-like mechanical unfolding pathway. Journal of the American Chemical Society. **134**, 17068-17075

66 Schlierf, M., Li, H. and Fernandez, J. (2004) The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. Proceedings of the National Academy of Sciences of the United States of America. **101**, 7299-7304

67 Alemany, A., Mossa, A., Junier, I. and Ritort, F. (2012) Experimental free-energy measurements of kinetic molecular states using fluctuation theorems. Nat Phys. **8**, 688-694

68 Shank, E. A., Cecconi, C., Dill, J. W., Marqusee, S. and Bustamante, C. (2010) The folding cooperativity of a protein is controlled by its chain topology. Nature. **465**, 637-640

69 Li, P. T. X., Collin, D., Smith, S. B., Bustamante, C. and Tinoco, I. (2006) Probing the mechanical folding kinetics of TAR RNA by hopping, force-jump, and force-ramp methods. Biophysical Journal. **90**, 250-260

70 Stigler, J. and Rief, M. (2012) Hidden Markov Analysis of Trajectories in Single-Molecule Experiments and the Effects of Missed Events. Chemphyschem. **13**, 1079-1086

71 Elms, P., Chodera, J., Bustamante, C. and Marqusee, S. (2012) Limitations of constant-force-feedback experiments. Biophysical journal. **103**, 1490-1499

72 Ribezzi-Crivellari, M. and Ritort, F. (2012) Force spectroscopy with dual-trap optical tweezers: molecular stiffness measurements and coupled fluctuations analysis. Biophysical journal. **103**, 1919-1928

73 Neuman, K. C. and Nagy, A. (2008) Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. Nat Methods. **5**, 491-505

74 De Vlaminck, I. and Dekker, C. (2012) Recent Advances in Magnetic Tweezers. Annu Rev Biophys. **41**, 453-472

75 De Vlaminck, I., Vidic, I., van Loenhout, M., Kanaar, R., Lebbink, J. and Dekker, C. (2010) Torsional Regulation of hRPA-induced Unwinding of Double Stranded DNA. Biophysical Journal. **98**, 438a-438a

76 Koster, D. A., Crut, A., Shuman, S., Bjornsti, M. A. and Dekker, N. H. (2010) Cellular Strategies for Regulating DNA Supercoiling: A Single-Molecule Perspective. Cell. **142**, 519-530

77 Noom, M. C., van den Broek, B., van Mameren, J. and Wuite, G. J. L. (2007) Visualizing single DNA-bound proteins using DNA as a scanning probe. Nat Methods. **4**, 1031-1036

78 De Vlaminck, I., Henighan, T., van Loenhout, M. T. J., Burnham, D. R. and Dekker, C. (2012) Magnetic Forces and DNA Mechanics in Multiplexed Magnetic Tweezers. Plos One. **7**

79 Sirinakis, G., Ren, Y. X., Gao, Y., Xi, Z. Q. and Zhang, Y. L. (2012) Combined versatile high-resolution optical tweezers and single-molecule fluorescence microscopy. Rev Sci Instrum. **83**

80 Dame, R. T., Noom, M. C. and Wuite, G. J. L. (2006) Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. Nature. **444**, 387-390

81 Bustamante, C., Chemla, Y. R., Forde, N. R. and Izhaky, D. (2004) Mechanical processes in biochemistry. Annual review of biochemistry. **73**, 705-748

82 Bell, G. I. (1978) Models for the specific adhesion of cells to cells. Science. **200**, 618-627

83 Liphardt, J., Onoa, B., Smith, S. B., Tinoco, I. and Bustamante, C. (2001) Reversible unfolding of single RNA molecules by mechanical force. Science. **292**, 733-737

84 Bustamante, C., Bryant, Z. and Smith, S. B. (2003) Ten years of tension: single-molecule DNA mechanics. Nature. **421**, 423-427

85 Bustamante, C., Marko, J. F., Siggia, E. D. and Smith, S. (1994) Entropic elasticity of lambda-phage DNA. Science. **265**, 1599-1600

86 Kramers, H. A. (1940) Brownian motion in a field of force and the diffusion model of chemical reactions. Physica. **7**

87 Gebhardt, J. C., Bornschlogl, T. and Rief, M. (2010) Full distance-resolved folding energy landscape of one single protein molecule. P Natl Acad Sci USA. **107**, 2013-2018

88 Schlierf, M., Berkemeier, F. and Rief, M. (2007) Direct observation of active protein folding using lock-in force spectroscopy. Biophysical journal. **93**, 3989-3998

89 Dudko, O. K., Graham, T. G. and Best, R. B. (2011) Locating the barrier for folding of single molecules under an external force. Physical review letters. **107**, 208301

90 Dudko, O. K., Hummer, G. and Szabo, A. (2008) Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. P Natl Acad Sci USA. **105**, 15755-15760

91 Manosas, M., Collin, D. and Ritort, F. (2006) Force-dependent fragility in RNA hairpins. Physical review letters. **96**, 218301

92 Woodside, M. T., Anthony, P. C., Behnke-Parks, W. M., Larizadeh, K., Herschlag, D. and Block, S. M. (2006) Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. Science. **314**, 1001-1004

93 Hinczewski, M., Gebhardt, J. C., Rief, M. and Thirumalai, D. (2013) From mechanical folding trajectories to intrinsic energy landscapes of biopolymers. P Natl Acad Sci USA. **110**, 4500-4505

94 Heidarsson, P. O., Otazo, M. R., Bellucci, L., Mossa, A., Imparato, A., Paci, E., Corni, S., Di Felice, R., Kragelund, B. B. and Cecconi, C. (2013) Single-Molecule Folding Mechanism of an EF-Hand Neuronal Calcium Sensor. Structure. **21**, 1812-1821

95 Rabiner, L. (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE

96 Prinz, J. H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., Chodera, J. D., Schutte, C. and Noe, F. (2011) Markov models of molecular kinetics: Generation and validation. J Chem Phys. **134**

97 Evans, E. and Ritchie, K. (1997) Dynamic strength of molecular adhesion bonds. Biophysical journal. **72**, 1541-1555

98 Tinoco, I. and Bustamante, C. (2002) The effect of force on thermodynamics and kinetics of single molecule reactions. Biophysical chemistry. **101-102**, 513-533

99 Olga, D., Gerhard, H. and Attila, S. (2006) Intrinsic Rates and Activation Free Energies from Single-Molecule Pulling Experiments. Physical review letters. **96**

100 Yu, H., Gupta, A. N., Liu, X., Neupane, K., Brigley, A. M., Sosova, I. and Woodside, M. T. (2012) Energy landscape analysis of native folding of the prion protein yields the diffusion constant, transition path time, and rates. Proceedings of the National Academy of Sciences of the United States of America. **109**, 14452-14457

101 Alemany, A., Ribezzi, M. and Ritortt, F. (2011) Recent progress in fluctuation theorems and free energy recovery. Aip Conf Proc. **1332**, 96-110

102 Jarzynski, C. (2011) Equalities and Inequalities: Irreversibility and the Second Law of Thermodynamics at the Nanoscale. Annu Rev Conden Ma P. **2**, 329-351

103 Jarzynski, C. (1997) Nonequilibrium equality for free energy differences. Physical review letters. **78**, 2690-2693

104 Liphardt, J., Dumont, S., Smith, S. B., Tinoco, I. and Bustamante, C. (2002) Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. Science. **296**, 1832-1835

105 Gore, J., Ritort, F. and Bustamante, C. (2003) Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. P Natl Acad Sci USA. **100**, 12564-12569

106 Crooks, G. E. (1999) Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. Phys Rev E. **60**, 2721-2726

107 Collin, D., Ritort, F., Jarzynski, C., Smith, S. B., Tinoco, I., Jr. and Bustamante, C. (2005) Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. Nature. **437**, 231-234

108 Bennett, C. H. (1976) Efficient Estimation of Free-Energy Differences from Monte-Carlo Data. J Comput Phys. **22**, 245-268

109 Mossa, A., Huguet, J. M. and Ritort, F. (2010) Investigating the thermodynamics of small biosystems with optical tweezers. Physica E. **42**, 666-671

110 Junier, I., Mossa, A., Manosas, M. and Ritort, F. (2009) Recovery of Free Energy Branches in Single Molecule Experiments. Physical review letters. **102**

111 Burgoyne, R. D. (2007) Neuronal calcium sensor proteins: generating diversity in neuronal Ca2+ signalling. Nat Rev Neurosci. **8**, 182-193

112 Heidarsson, P. O., Bjerrum-Bohr, I. J., Jensen, G. A., Pongs, O., Finn, B. E., Poulsen, F. M. and Kragelund, B. B. (2012) The C-Terminal Tail of Human Neuronal Calcium Sensor 1 Regulates the Conformational Stability of the Ca2+-Activated State. Journal of Molecular Biology. **417**, 51-64

113 Aravind, P., Chandra, K., Reddy, P. P., Jeromin, A., Chary, K. V. R. and Sharma, Y. (2008) Regulatory and structural EF-hand motifs of neuronal calcium sensor-1: Mg2+ modulates Ca2+ binding, Ca2+-induced conformational changes, and equilibrium unfolding transitions. J Mol Biol. **376**, 1100-1115

114 Ames, J. B., Dizhoor, A. M., Ikura, M., Palczewski, K. and Stryer, L. (1999) Three-dimensional structure of guanylyl cyclase activating protein-2, a calcium-sensitive modulator of photoreceptor guanylyl cyclases. The Journal of biological chemistry. **274**, 19329-19337

115 Bourne, Y., Dannenberg, J., Pollmann, V., Marchot, P. and Pongs, O. (2001) Immunocytochemical localization and crystal structure of human frequenin (neuronal calcium sensor 1). Journal of Biological Chemistry. **276**, 11949-11955

116 Strahl, T., Huttner, I. G., Lusin, J. D., Osawa, M., King, D., Thorner, J. and Ames, J. B. (2007) Structural insights into activation of phosphatidylinositol 4-kinase (Pik1) by yeast frequenin (Frq1). Journal of Biological Chemistry. **282**, 30949-30959

117 Cox, J. A., Durussel, I., Comte, M., Nef, S., Nef, P., Lenz, S. E. and Gundelfinger, E. D. (1994) Cation binding and conformational changes in VILIP and NCS-1, two neuron-specific calcium-binding proteins. The Journal of biological chemistry. **269**, 32807-32813

118 Kuboniwa, H., Tjandra, N., Grzesiek, S., Ren, H., Klee, C. B. and Bax, A. (1995) Solution structure of calcium-free calmodulin. Nature structural biology. **2**, 768-776

119 Burgoyne, R. D. and Haynes, L. P. (2012) Understanding the physiological roles of the neuronal calcium sensor proteins. Mol Brain. **5**

120 Reyes-Bermudez, A., Miller, D. J. and Sprungala, S. (2012) The Neuronal Calcium Sensor Protein Acrocalcin: A Potential Target of Calmodulin Regulation during Development in the Coral Acropora millepora. Plos One. **7**

121 Weiss, J. L., Hui, H. and Burgoyne, R. D. (2010) Neuronal Calcium Sensor-1 Regulation of Calcium Channels, Secretion, and Neuronal Outgrowth. Cell Mol Neurobiol. **30**, 1283-1292

122 Ames, J. B. and Lim, S. (2012) Molecular structure and target recognition of neuronal calcium sensor proteins. Bba-Gen Subjects. **1820**, 1205-1213

123 Kabbani, N., Negyessy, L., Lin, R., Goldman-Rakic, P. and Levenson, R. (2002) Interaction with neuronal calcium sensor NCS-1 mediates desensitization of the D2 dopamine receptor. The Journal of neuroscience : the official journal of the Society for Neuroscience. **22**, 8476-8486

124 Burgoyne, R. D., O'Callaghan, D. W., Hasdemir, B., Haynes, L. P. and Tepikin, A. V. (2004) Neuronal Ca2+-sensor proteins: multitalented regulators of neuronal function. Trends in neurosciences. **27**, 203-209

125 Seaton, G., Hogg, E. L., Jo, J., Whitcomb, D. J. and Cho, K. (2011) Sensing change: The emerging role of calcium sensors in neuronal disease. Semin Cell Dev Biol. **22**, 530-535

126 Kumar, A., Bodhinathan, K. and Foster, T. (2009) Susceptibility to Calcium Dysregulation during Brain Aging. Frontiers in aging neuroscience. **1**, 2

127 Saper, M. A., Bjorkman, P. J. and Wiley, D. C. (1991) Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 A resolution. J Mol Biol. **219**, 277-319

128 Berman, H. M. (2000) Nucleic Acid Database: Present and future. Abstr Pap Am Chem S. **220**, U234-U234

129 Smith, D. P. and Radford, S. E. (2001) Role of the single disulphide bond of beta(2)-microglobulin in amyloidosis in vitro. Protein Science. **10**, 1775-1784

130 Trinh, C. H., Smith, D. P., Kalverda, A. P., Phillips, S. E. V. and Radford, S. E. (2002) Crystal structure of monomeric human beta-2-microglobulin reveals clues to its amyloidogenic properties. Proceedings of the National Academy of Sciences of the United States of America. **99**, 9771-9776

131 Giorgetti, S., Rossi, A., Mangione, P., Raimondi, S., Marini, S., Stoppini, M., Corazza, A., Viglino, P., Esposito, G., Cetta, G., Merlini, G. and Bellotti, V. (2005) beta 2-microglobulin isoforms display an heterogeneous affinity for type I collagen. Protein Science. **14**, 696-702

132 Richardson, J. S. and Richardson, D. C. (2002) Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. P Natl Acad Sci USA. **99**, 2754-2759

133 Floege, J. and Ehlerding, G. (1996) Beta-2-microglobulin-associated amyloidosis - Discussion. Nephron. **72**, 9-26

134 Bellotti, V. and Chiti, F. (2008) Amyloidogenesis in its biological environment: challenging a fundamental issue in protein misfolding diseases. Curr Opin Struc Biol. **18**, 771-779

135 Xi, Z. Q., Gao, Y., Sirinakis, G., Guo, H. L. and Zhang, Y. L. (2012) Single-molecule observation of helix staggering, sliding, and coiled coil misfolding. Proceedings of the National Academy of Sciences of the United States of America. **109**, 5711-5716

136 Wang, Q. M., Shah, N., Zhao, J., Wang, C. S., Zhao, C., Liu, L. Y., Li, L. Y., Zhou, F. M. and Zheng, J. (2011) Structural, morphological, and kinetic studies of beta-amyloid peptide aggregation on self-assembled monolayers. Phys Chem Chem Phys. **13**, 15200-15210

137 Valleix, S., Gillmore, J. D., Bridoux, F., Mangione, P. P., Dogan, A., Nedelec, B., Boimard, M., Touchard, G., Goujon, J. M., Lacombe, C., Lozeron, P., Adams, D., Lacroix, C., Maisonobe, T., Plante-Bordeneuve, V., Vrana, J. A., Theis, J. D., Giorgetti, S., Porcari, R., Ricagno, S., Bolognesi, M., Stoppini, M., Delpech, M., Pepys, M. B., Hawkins, P. N. and Bellotti, V. (2012) Hereditary Systemic Amyloidosis Due to Asp76Asn Variant beta(2)-Microglobulin. New Engl J Med. **366**, 2276-2283

138 Swartz, M. A. and Fleury, M. E. (2007) Interstitial flow and its effects in soft tissues. Annu Rev Biomed Eng. **9**, 229-256

139 Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Hawkins, P. N., Dobson, C. M., Radford, S. E., Blake, C. C. F. and Pepys, M. B. (1997) Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. Nature. **385**, 787-793

140 Gifford, J. L., Walsh, M. P. and Vogel, H. J. (2007) Structures and metal-ion-binding properties of the Ca2+-binding helix-loop-helix EF-hand motifs. The Biochemical journal. **405**, 199-221

141 Barclay, J. W., Morgan, A. and Burgoyne, R. D. (2005) Calcium-dependent regulation of exocytosis. Cell calcium. **38**, 343-353

142 Burgoyne, R. D. and Morgan, A. (2003) Secretory granule exocytosis. Physiological reviews. **83**, 581-632

143 Yamniuk, A. P., Silver, D. M., Anderson, K. L., Martin, S. R. and Vogel, H. J. (2007) Domain stability and metal-induced folding of calcium- and integrin-binding protein 1. Biochemistry. **46**, 7088-7098

144 Capozzi, F., Casadei, F. and Luchinat, C. (2006) EF-hand protein dynamics and evolution of calcium signal transduction: an NMR view. J Biol Inorg Chem. **11**, 949-962

145 Grabarek, Z. (2006) Structural basis for diversity of the EF-hand calcium-binding proteins. Journal of Molecular Biology. **359**, 509-525

146 Grabarek, Z. (2011) Insights into modulation of calcium signaling by magnesium in calmodulin, troponin C and related EF-hand proteins. Biochimica et biophysica acta. **1813**, 913-921

147 Zot, H. G. and Potter, J. D. (1982) A structural role for the Ca2+-Mg2+ sites on troponin C in the regulation of muscle contraction. Preparation and properties of troponin C depleted myofibrils. The Journal of biological chemistry. **257**, 7678-7683

148 Peshenko, I. V. and Dizhoor, A. M. (2006) Ca2+ and Mg2+ binding properties of GCAP-1. Evidence that Mg2+-bound form is the physiological activator of photoreceptor guanylyl cyclase. The Journal of biological chemistry. **281**, 23830-23841

149 Wingard, J. N., Chan, J., Bosanac, I., Haeseleer, F., Palczewski, K., Ikura, M. and Ames, J. B. (2005) Structural analysis of Mg2+ and Ca2+ binding to CaBP1, a neuron-specific regulator of calcium channels. The Journal of biological chemistry. **280**, 37461-37470

150 Brunet, S., Scheuer, T., Klevit, R. and Catterall, W. A. (2005) Modulation of CaV1.2 channels by Mg2+ acting at an EF-hand motif in the COOH-terminal domain. The Journal of general physiology. **126**, 311-323

151 Osawa, M., Dace, A., Tong, K. I., Valiveti, A., Ikura, M. and Ames, J. B. (2005) Mg2+ and Ca2+ differentially regulate DNA binding and dimerization of DREAM. The Journal of biological chemistry. **280**, 18008-18014

152 Peshenko, I. V., Olshevskaya, E. V., Yao, S., Ezzeldin, H. H., Pittler, S. J. and Dizhoor, A. M. (2010) Activation of retinal guanylyl cyclase RetGC1 by GCAP1: stoichiometry of binding and effect of new LCA-related mutations. Biochemistry. **49**, 709-717

153 Navarro, G., Hradsky, J., Lluis, C., Casado, V., McCormick, P. J., Kreutz, M. R. and Mikhaylova, M. (2012) NCS-1 associates with adenosine A(2A) receptors and modulates receptor function. Frontiers in molecular neuroscience. **5**, 53

154 Haynes, L. P., Thomas, G. M. and Burgoyne, R. D. (2005) Interaction of neuronal calcium sensor-1 and ADP-ribosylation factor 1 allows bidirectional control of phosphatidylinositol 4-kinase beta and trans-Golgi network-plasma membrane traffic. The Journal of biological chemistry. **280**, 6047-6054

155 Heidarsson, P. O., Naqvi, M. M., Otazo, M. R., Mossa, A., Kragelund, B. B. and Cecconi, C. (2014) Direct single-molecule observation of calcium-dependent misfolding in human neuronal calcium sensor-1. Proceedings of the National Academy of Sciences of the United States of America
111, 13069-13074

156 Ferreon, A. C. and Deniz, A. A. (2011) Protein folding at single-molecule resolution. Biochimica et biophysica acta. **1814**, 1021-1029

157 Heidarsson, P. O., Naqvi, M. M., Sonar, P., Valpapuram, I. and Cecconi, C. (2013) Conformational dynamics of single protein molecules studied by direct mechanical manipulation. Advances in protein chemistry and structural biology. **92**, 93-133

158 Moffitt, J. R., Chemla, Y. R., Smith, S. B. and Bustamante, C. (2008) Recent advances in optical tweezers. Annual review of biochemistry. **77**, 205-228

159 Zoldák, G. and Rief, M. (2013) Force as a single molecule probe of multidimensional protein energy landscapes. Curr Opin Struc Biol. **23**, 48-57

160 Chodera, J. D., Elms, P., Noé, F., Keller, B., Kaiser, C. M., Ewall-Wice, A., Marqusee, S., Bustamante, C. and Hinrichs, N. S. (2011) Bayesian hidden Markov model analysis of single-molecule force spectroscopy: Characterizing kinetics under measurement uncertainty. arXiv preprint arXiv:1108.1430

161 McKinney, S. A., Joo, C. and Ha, T. (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. Biophysical Journal. **91**, 1941-1951

162 Yap, K. L., Ames, J. B., Swindells, M. B. and Ikura, M. (1999) Diversity of conformational states and changes within the EF-hand protein superfamily. Proteins. **37**, 499-507

163 Lian, L. Y., Pandalaneni, S. R., Todd, P. A., Martin, V. M., Burgoyne, R. D. and Haynes, L. P. (2014) Demonstration of Neuronal Calcium Sensor-1 binding to the Cav2.1 P/Q-type calcium channel. Biochemistry

164 Lian, L. Y., Pandalaneni, S. R., Patel, P., McCue, H. V., Haynes, L. P. and Burgoyne, R. D. (2011) Characterisation of the interaction of the C-terminus of the dopamine D2 receptor with neuronal calcium sensor-1. Plos One. **6**, e27779

165 Woll, M., De Cotiis, D., Bewley, M., Tacelosky, D., Levenson, R. and Flanagan, J. (2011) Interaction between the D2 dopamine receptor and neuronal calcium sensor-1 analyzed by fluorescence anisotropy. Biochemistry. **50**, 8780-8791

166 Kabbani, N., Woll, M. P., Nordman, J. C. and Levenson, R. (2012) Dopamine receptor interacting proteins: targeting neuronal calcium sensor-1/D2 dopamine receptor interaction for antipsychotic drug development. Current drug targets. **13**, 72-79

167 Kragelund, B. B., Hauenschild, A., Carlstrom, G., Pongs, O. and Finn, B. E. (2000) 1H, 13C, and 15N assignments of un-myristoylated Ca2+-frequenin, a synaptic efficacy modulator. Journal of biomolecular NMR. **16**, 85-86

168 Petrie, L. E. B. a. T. (1966) Statistical inference for probabilistic functions of finite state Markov chains. The Annals of Mathematical Statistics, 1554-1563

169 Tinoco, I. (2004) Force as a useful variable in reactions: Unfolding RNA. Annu Rev Bioph Biom. **33**, 363-385

170 Mossa, A., Manosas, M., Forns, N., Huguet, J. M. and Ritort, F. (2009) Dynamic force spectroscopy of DNA hairpins: I. Force kinetics and free energy landscapes. Journal of Statistical Mechanics: Theory and Experiment. **2009**

171 Maxwell, S. L., Ho, H. Y., Kuehner, E., Zhao, S. and Li, M. (2005) Pitx3 regulates tyrosine hydroxylase expression in the substantia nigra and identifies a subgroup of mesencephalic dopaminergic progenitor neurons during mouse development. Developmental biology. **282**, 467-479