Cytosolic Interactome Protects Against Protein Unfolding in a Single Molecule Experiment

Scalvini, Barbara; Heling, Laurens W.H.J.; Sheikhhassani, Vahid; Sunderlikova, Vanda; Tans, Sander J.; Mashaghi, Alireza

DOI
10.1002/adbi.202300105

Publication date
2023

Document Version
Final published version

Published in
Advanced Biology

ADVANCED BIOLOGY
www.advanced-bio.com

Cytosolic Interactome Protects Against Protein Unfolding in a Single Molecule Experiment

*Barbara Scalvini, Laurens W.H.J. Heling, Vahid Sheikhhassani, Vanda Sunderlikova, Sander J. Tans, and Alireza Mashaghi**

Single molecule techniques are particularly well suited for investigating the processes of protein folding and chaperone assistance. However, current assays provide only a limited perspective on the various ways in which the cellular environment can influence the folding pathway of a protein. In this study, a single molecule mechanical interrogation assay is developed and used to monitor protein unfolding and refolding within a cytosolic solution. This allows to test the cumulative topological effect of the cytoplasmic interactome on the folding process. The results reveal a stabilization against forced unfolding for partial folds, which are attributed to the protective effect of the cytoplasmic environment against unfolding and aggregation. This research opens the possibility of conducting single molecule molecular folding experiments in quasi-biological environments.

1. Introduction

In recent years, single-molecule techniques have shown their potential in the characterization of folding pathways and molecular interactions,[1–6] in view of their ability to track transient, heterogeneous phenomena. Protein folding transitions have often been studied for proteins in isolation,[7–9] although in physiological conditions the protein often interacts with many other molecules present in the crowded cellular environment. In this respect, chaperones play a crucial role in assisting protein folding,[10,11] by blocking harmful aggregation,[12] rescuing misfolded proteins,[13] or even affecting the folding pathway.[14,15] However fundamental, the exact mechanisms of chaperone – protein interactions remain mostly obscure, due in part to the promiscuity and heterogeneity of chaperone functions.[16] While the interaction between specific client and chaperone/co-chaperone systems by optical tweezers assays has been the object of growing interest in recent years,[12,14,15] we are still lacking the full picture when it comes to protein folding in the cellular environment. This is partly because current single molecule assays are done using buffer solutions including one or a few chaperones, ignoring the complexity of the cytoplasmic content. Besides several major chaperone systems,[17] the cytosol contains a wide spectrum of proteins that may exhibit yet unknown transient interactions affecting protein states and folding. Certain single-molecule investigations have been designed to probe the crowding effect, but these studies lack the specificity of cytoplasmic interactions.[18] Overall, there is a need for assays that probe the folding dynamics of proteins when exposed to the full complexity of the cellular interactome.

Here, we aim at probing the cumulative structural modifications imposed by the cytosolic molecular machinery on a substrate, in this case, maltose-binding protein (MBP), beyond the passive crowding effect. In this proof-of-concept study, we perform single-molecule protein (un)folding experiments in diluted *E. coli* cell extract, and observe notable variations in the force profile of intermediate states. We see a stabilization against forced unfolding for partial folds, indicating "cytoplasmic chaperone" action aimed at protecting these structures from stress-induced unfolding and aggregation. We also present data on single double-stranded DNA (dsDNA) molecules in a cell extract, as DNA handles are often an integral part of protein folding studies by optical tweezers. To our knowledge, this is the first time the mechanical response to the cytosolic interactome of dsDNA molecules is characterized. We observe a lower elastic modulus and a strong hysteretic response after the 65 pN plateau, where

B. Scalvini, L. W. Heling, V. Sheikhhassani, A. Mashaghi
Medical Systems Biophysics and Bioengineering
Leiden Academic Centre for Drug Research
Faculty of Science
Leiden University
Einsteinweg 55, Leiden 2333CC, The Netherlands
E-mail: a.mashaghi.tabari@lacdr.leidenuniv.nl

B. Scalvini, L. W. Heling, V. Sheikhhassani, A. Mashaghi
Centre for Interdisciplinary Genome Research
Faculty of Science
Leiden University
Einsteinweg 55, Leiden 2333CC, The Netherlands

V. Sunderlikova, S. J. Tans
AMOLF
Science Park 104, Amsterdam 1098 XG, The Netherlands

S. J. Tans
Department of Bionanoscience
Kavli Institute of Nanoscience
Delft University of Technology
Van der Maasweg 9, Delft 2629HZ, The Netherlands

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adbi.202300105

DOI: 10.1002/adbi.202300105

the effect of cytosolic proteins binding stretches of ssDNA is predominant. These variations should be appropriately accounted for when using DNA as handle in this type of assays. In view of these results, we discuss the challenges and opportunities of pulling experiments in complex environments.

Considering the heterogeneity of processes and interactions taking place between MBP and the cytosolic environment, we also advance a new way to analyze force spectroscopy data, based on the theoretical framework of circuit topology (CT).[19–22] CT formalizes the arrangement of contacts and loops in a folded chain and thus, disruption and formation of loops as probed in the tweezers assay can be directly modeled by this approach. We can therefore represent such loops and associated contour lengths in the formalism of CT, providing a way to visualize the complete ensemble of unfolding pathways and reconstructing from it the topological fingerprint of the substrate.

2. Results

Pulling experiments in cell extract can present various technical challenges, due to the viscous nature of bacterial cytoplasm[23] and the variety of interactions that can arise between cytosolic biomolecules and the construct under study. In order to limit force contributions arising from the crowding effect, we diluted our cell extract four times in buffer after extraction (see Methods). In this way, physiological concentration ratios of the cytosolic proteome are conserved, while also allowing us to perform traditional optical tweezers experiments with passive (thermal) calibration,[24] since we can expect the viscosity of our solution to be homogeneous. We compared the viscosity of cytosol solution (indicated as cytosol* in the figure) and buffer (**Figure 1**A) by Acoustic Force Spectroscopy (AFS).[25] We measured the viscosity by extracting the limit velocity of free-falling 4 μm silica beads in cytosol solution and buffer, from a starting position determined by the acoustic node generated by the AFS. From the boxplot in Figure 1A, we can observe the spread of the viscosity distribution measured in the two media: buffer viscosity has mean = 9.3×10^{-4} and standard deviation = 0.6×10^{-4} Pa s, while the cytosol solution scores a mean of 9.6×10^{-4} and standard deviation of 0.7×10^{-4} Pa s. The two distributions compare with a *p*-value of 0.057, as calculated by Kolmogorov test (two-tailed).

2.1. DNA Pulling in Cell Extract

Many protocols for force spectroscopy assays involve the coupling of the protein(s) of interest to long DNA overhangs.[26–28] As such, force-extension assays on DNA represent the baseline for pulling experiments. Therefore, a characterization of the interaction between the DNA molecule and the environment where the experiment takes place is important. Moreover, the mechanical behavior of DNA in a biologically relevant environment is a critical aspect of many cellular processes; its elastic properties play a role in protein-DNA interaction, bending, twisting, but also genome compaction and various other structural transitions.[5,29] This is particularly relevant to bacterial cells, where the DNA is embedded in the cytoplasm.

In order to characterize the environmental interplay between dsDNA molecules and cytosolic environment, we performed DNA-pulling experiments in cell extract (Figure 1B). After each pull, the molecule was left for 5 or 30 s in the overstretching region. Subsequently, it was relaxed for other 5/30 s in its relaxed state, before repeating the cycle. After the pull (characterized by the worm-like-chain model[30]) the plateau expected for DNA overstretching (OS) is visible at about 65 pN[31] (Figure 1B). However, substantial changes appear striking in the hysteretic behavior of DNA in the retraction phase, after the OS plateau (see Figure S1, Supporting Information, for comparison). DNA overstretching under (close to) physiological conditions has been presented as a highly cooperative transition.[31] Depending on boundary conditions such as salt content and temperature, two competing processes might occur predominantly during DNA overstretching[32,33]: a reversible fast transition to an overstretched form of dsDNA (S DNA),[33] or the force-induced melting of the DNA strands, resulting in the breaking of hydrogen bonds with the production of ssDNA regions propagating from either a nick or the ends of the molecule.[34] Considering our experimental setup (25 °C and ≈100 mM salt concentration), we expect melting to be visible at overstretching force (65 pN). Therefore, we can assume regions of ssDNA are exposed to the cytosolic environment during overstretching of our DNA molecule. The *E. coli* cytosol proteome presents a variety of proteins capable of binding to ssDNA,[35,36] and it is reasonable to expect that some of them should interact with the exposed regions of ssDNA during the OS phase. Binding with such proteins might induce the ssDNA regions to delay their transition back to B DNA, which is now energetically unfavorable in the range between 50 and 60 pN, where this return transition generally happens in the buffer.[33] We call here "return force" (R_F) the force value for which the retraction curve rejoins the prior pulling curve. We can see from the plots in Figure 1C that, while R_F remains segregated in a narrow interval around 50 pN in buffer, the return force for pulls in cell extract has a wide distribution, populating mostly force ranges below 30 pN. The low force return point becomes more evident when we leave the molecule in the OS phase for longer (Figure 1D). Interestingly, sections of the retraction curves fit a freely-joint chain (FJC) curve, a model generally associated with the force/extension curve ssDNA. What we are observing is probably an ssDNA–dsDNA hybrid complex, somewhat similar to what was obtained by Leila Shokri et al.[37] by stretching dsDNA in presence of glyoxal. This effect is either absent or not relevant enough to be detected by visual inspection of force-extension curves for dsDNA in buffer (Figure S1, Supporting Information).

We could assume that, being the recovery of B DNA structure delayed by ssDNA-protein binding, the retraction curve walks along stretches of curves that are most likely linear combinations between WLC and FJC with weights given by the portion of the chain that is in an ssDNA state in that moment of the transition. We can see a shift towards lower values of the WLC parameters, that is, persistence length L_P and elastic modulus K (Figure 1E). This effect might indicate the cytosolic content is interacting with the dsDNA molecule under tension.

These changes in mechanical properties of the dsDNA molecule should be taken into account when measuring in a complex medium such as cell extract, as wider variations in the mechanical stiffness of the DNA handle can potentially affect the signal to noise ratio of the measurement.[38] It is advisable to perform a separate WLC fit for each pull. Types of OT assays other

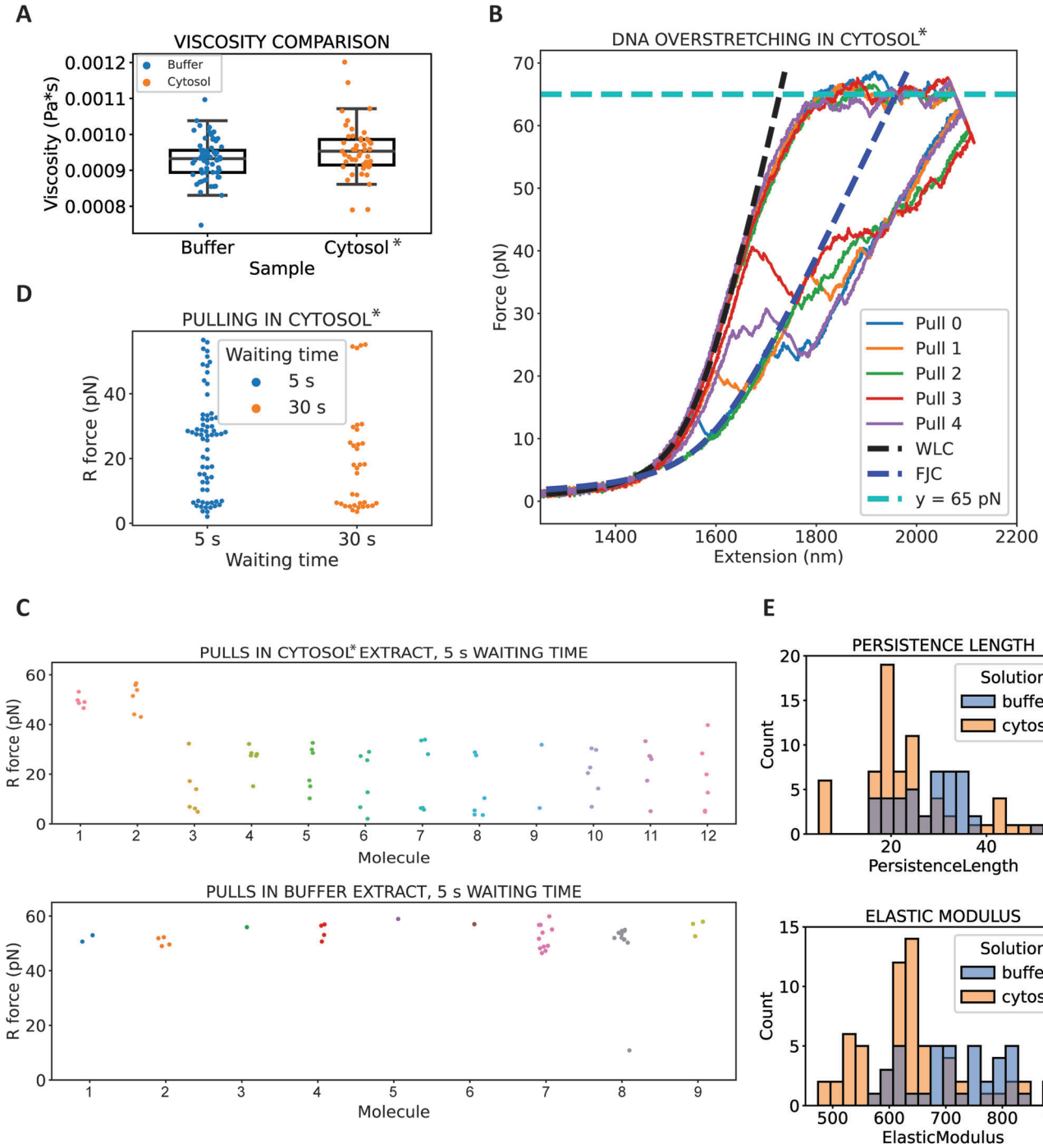

Figure 1. The cytosolic proteome alters the mechanical properties of DNA. A) Boxplot of the viscosity of our cytosol solution compared with buffer viscosity, measured by tracking the limit velocity of freely falling beads in solution in the AFS. For buffer, N = 66 beads were measured, and for the cytosol solution, N = 47. The two distributions compare with a *p*-value of 0.057, as calculated by Kolmogorov test (two-tailed). B) Example of DNA pulls performed in diluted cytosol solution. The two dashed curves represent the worm-like chain (black) and the freely-joint chain (blue) model. Each pull is represented by two curves, corresponding to the stretching (curve following the WLC model) and relaxation of the molecular tether (curve following the FJC model). C) Strip plot of return forces in buffer and in cytosol solution. The data represented were collected from N = 12 DNA molecules (67 pulls) in cytosol solution, and from N = 9 DNA molecules (44 pulls) in buffer. D) Swarm plot of the return force in cytosol, with 5 and 30 s waiting time between pulls. The plot represents data from 67 pulls from N = 12 molecules with a waiting time of 5 s, and data from 43 pulls from N = 14 molecules with a waiting time of 30 s. E) Histogram of the persistence length L_p and elastic modulus K, in buffer and cytosol. The data represented was collected from N = 12 DNA molecules (67 pulls) in cytosol solution, and from N = 9 DNA molecules (44 pulls) in buffer. The label cytosol* in all panels indicates a 4 times dilution of cytosol extract in buffer.

than force-ramp experiments involving long experimental times might need further adjustments to be able to use DNA effectively as a handle in cytosol.

2.2. MBP Pulling in Cell Extract

Individual MBP molecules were tethered between two polystyrene beads (2 μm diameter) by using a 2500 bp DNA handle on each side of the protein, to reduce risk of protein photodamage and bead-bead interaction. Then, we moved the construct into the microfluidics channel containing diluted cell extract. There, we performed cycles of subsequent stretching and relaxation to low forces for a waiting time of 5 s, to provide the protein the opportunity to refold. A complex picture emerged from the unfolding pathway in cytosol solution (**Figure 2**), especially when it comes to the distribution of unfolding forces (Figure 2A). Native MBP presents a widely studied and distinctive unfolding curve in buffer, with one minor unfolding event happening at low force (≈10 pN) and the main unfolding of the MBP core generally measured at ≈25 pN.[12,14,15,39–41] The upper panel in Figure 2A is compatible with this picture, also including all rare intermediate states, generally unfolding in the range between 10 and 20 pN. We found unfolding of the MBP core in our dataset to happen at a force of 28.1 ± 0.6 pN, slightly higher than what is generally reported in Optical Tweezers assays,[12,14,15,39–41] a discrepancy that can be attributed to the pulling speed of the experiments (200 nm/s). However, the histogram of unfolding forces in cytosol solution (Figure 2A, lower panel) presents a more complex process, a wide distribution where it is difficult to identify a preferred pathway. Moreover, the higher force tail of the distribution appears to be more populated in the cytosol solution than in buffer, with 3.1% of unfolding forces having higher values than 40 pN in cytosol solution, versus 1.2% in buffer. The histogram of the contour lengths in Figure 2B provides further insight into this complexity. We detect the frequently visited intermediate *I* state corresponding to the MBP core at 24.5 ± 0.2 nm (see red asterisk in Figure 2B). The same intermediate state in cytosol solution appears to be less prevalent (Figure 2B, lower panel), with a broad peak reaching 40 nm.

In order to make sense of such heterogeneity, we require a smart way to visualize unfolding pathways. We chose to represent transitions on a contour length diagram (Figure 2C). After binning the contour lengths into 5 nm bins, we selected the most frequent unfolding transitions, as defined by initial and final contour lengths. The threshold for frequency is set by the last quartile (0.75) of bin counts, that is, transitions that appear more than three times (in buffer) and two times (in cytosol solution) are displayed. This filtering allows us to improve the signal-to-noise ratio of the representation. The color of the arcs represents the average force at which that transition occurs. The same information, without filtering, is shown in Figure 2D in the form of a heatmap. The initial contour length of the transition is displayed along rows, and the final contour length along columns. A few important observations can be drawn from these representations. One main feature is the abundance of high-force transitions occurring with initial contour lengths higher than 25 nm (therefore after the *I* intermediate state), which show to be much more common when the protein is exposed to the cytosolic proteome. An example of these high-force transitions is displayed in Figure 2E, alongside the most common pathway for MBP unfolding in buffer. These increased unfolding forces show that cytosolic compounds stabilize intermediate states. This effect is consistent with what was previously reported about trigger factor,[15] and Dnak and GroEL in the ADP state.[14,42]

The circuit diagram also shows a tendency of MBP to present a first unfolding event (in its most common pathway) from its native conformation to contour lengths between 20 and 40 nm in the diluted cytosol (Figure 2C, lower panel). The width of this contour length range might suggest small structural variations affecting the *I* intermediate corresponding to the MBP core, which has an associated contour length of ≈30 nm. A similar picture emerges for rare intermediate states detected at contour lengths between 60 and 80 nm, for which the frequency and unfolding force increase in cytosol solution. In order to provide a guideline to interpret these results, we use the concept of mechanical unfoldons (independently unfolding cores), previously identified in MBP by AFM (Atomic Force Microscopy) assays.[43] Four such structural blocks were identified in MBP unfolding: M1, corresponding to an increase in contour length of ≈23 nm, attributed to the unfolding of C-terminal alpha helices (residues 296–366); M2, corresponding to residues 295–244, contributing with a sequential increase of ≈16 nm, and therefore a cumulative contour length M1 + M2 ≈ 39 nm; M3 (N terminal, residues 1–113), often occurring together with M2 as a single unfolding event, yields M1 + M2 + M3 ≈ 81 nm. Finally, M4 unfolds residues 114–243, fully unfolding the protein (**Figure 3**A). These intermediate states (23, 39, 81) can readily be identified in the diagram in Figure 2C; by defining an interval of ±5 nm around the contour length value, we can calculate the average force at which unfolding occurs in that range. Since an intermediate state around 60 nm seems also to be heavily visited in cytosol solution (and was reported in previous studies of MBP alone and with TF[15]), we shall also consider this contour length in the analysis. The results of this analysis present an interesting trend. The unfolding at 23 nm (*I* intermediate state, MBP core) remains basically unaltered in the cytosol solution, in terms of unfolding forces: $F_{C,23}$ = 27.7 ± 0.9 pN as opposed to an unfolding force in buffer of $F_{B,23}$ = 28.1 ± 0.6 pN. However, as expected, all following intermediate states present a noticeable increase in unfolding force, as reported in Table 1. The observed frequency of such intermediates increases as well roughly two- to threefold in cytosol solution. If we rely on the mapping of unfoldons over MBP sequence (Figure 3A), one might conclude that the biggest variation in stability and frequency is observed from the unfolding at 39 nm onward, that is, the M3–M4 independent folding units (1–243 residues). Hence, the data indicate that cytosolic components interact mostly with partially folded MBP, specifically with its N-terminal half, increasing its resistance to unfolding. There also seems to be a mild increase in unfolding force for the initial unfolding corresponding to the native conformation of MBP (0 nm, Table 1), although its observed frequency does not increase significantly. These data could suggest interaction between cytosolic chaperones and native MBP.

Although these data present a clear trend, our intermediate states are somewhat a simplification of the heterogeneous pathways we can observe in cytosol solution (Figure 2C,D). A way to tackle and visualize this complexity could be the formalism

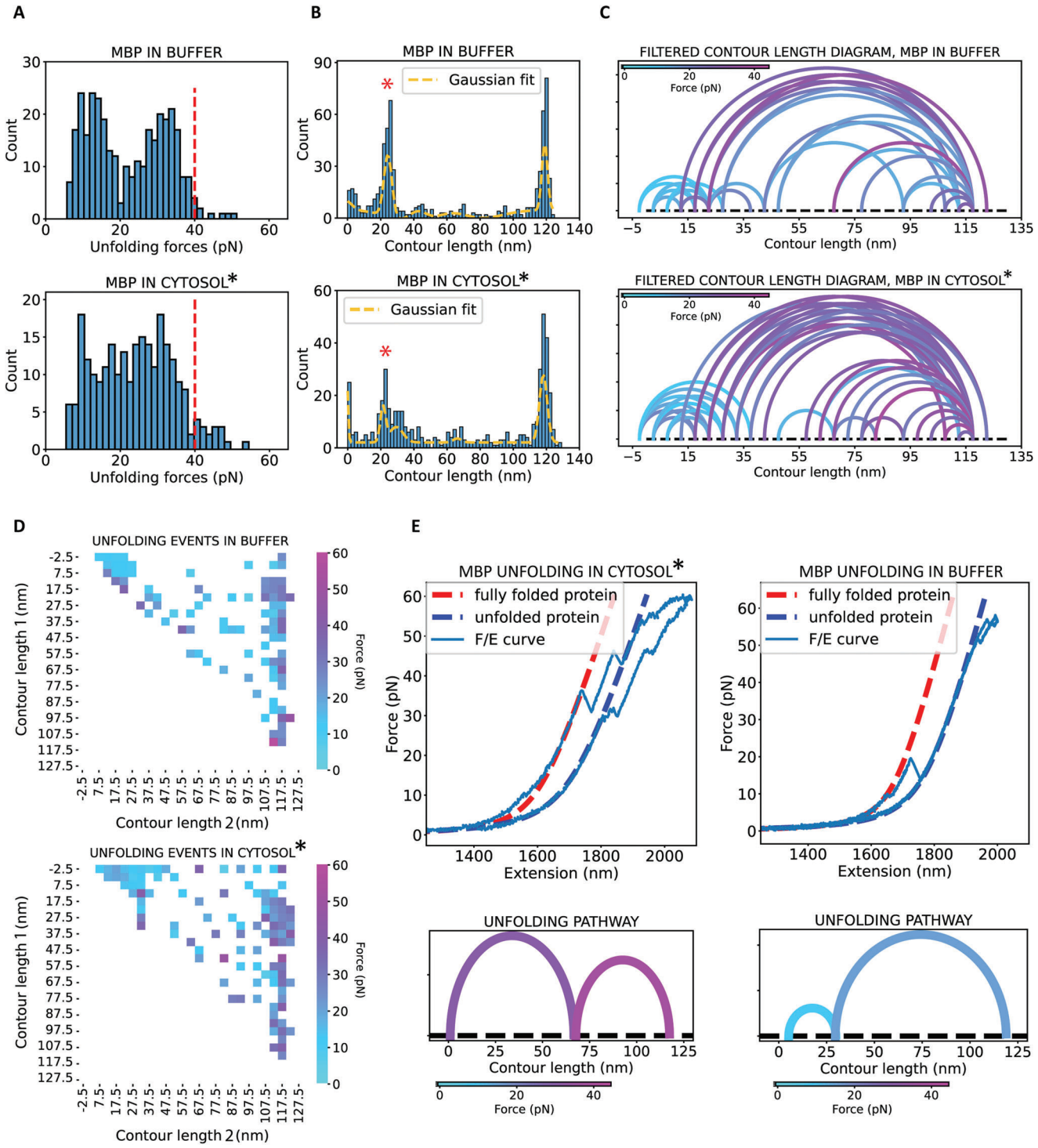

Figure 2. The cytosolic proteome interacts with MBP unfolding intermediates and affects preferred unfolding pathways. Data represented in panels (A), (B), and (D) are collected from N = 42 molecules of MBP (183 pulls) in cytosol, and N = 36 molecules of MBP (221 pulls) in buffer. A) Histogram of MBP unfolding forces in buffer and cytosol. The dashed red line is meant for visualization purposes, to highlight the counts occurring at forces higher than 40 pN. In cytosol, such unfolding forces represent 3.1% of the total, whereas in buffer only 1.2% of unfolding events occur above 40 pN. B) Histogram of MBP contour lengths in buffer and cytosol solution. The multi-gaussian fit (dashed yellow line) is meant for visualization purposes, and was calculated by providing as initial gaussian centers the contour lengths reported in **Table 1**, with the addition of the fully unfolded state (120 nm). C) Representation of unfolding transitions as circuit diagram. Every arc connects initial and final contour lengths of an unfolding event. Only most frequent unfolding events were kept (last quartile). The color-coding indicates the average force for that specific transition. D) Representation of unfolding events as heatmap. The color-coding indicates the average force for that specific transition. E) Example of force/extension curves for unfolding transitions in cytosol solution and buffer. The label cytosol* in all panels indicates a 4 times dilution of cytosol extract in buffer.

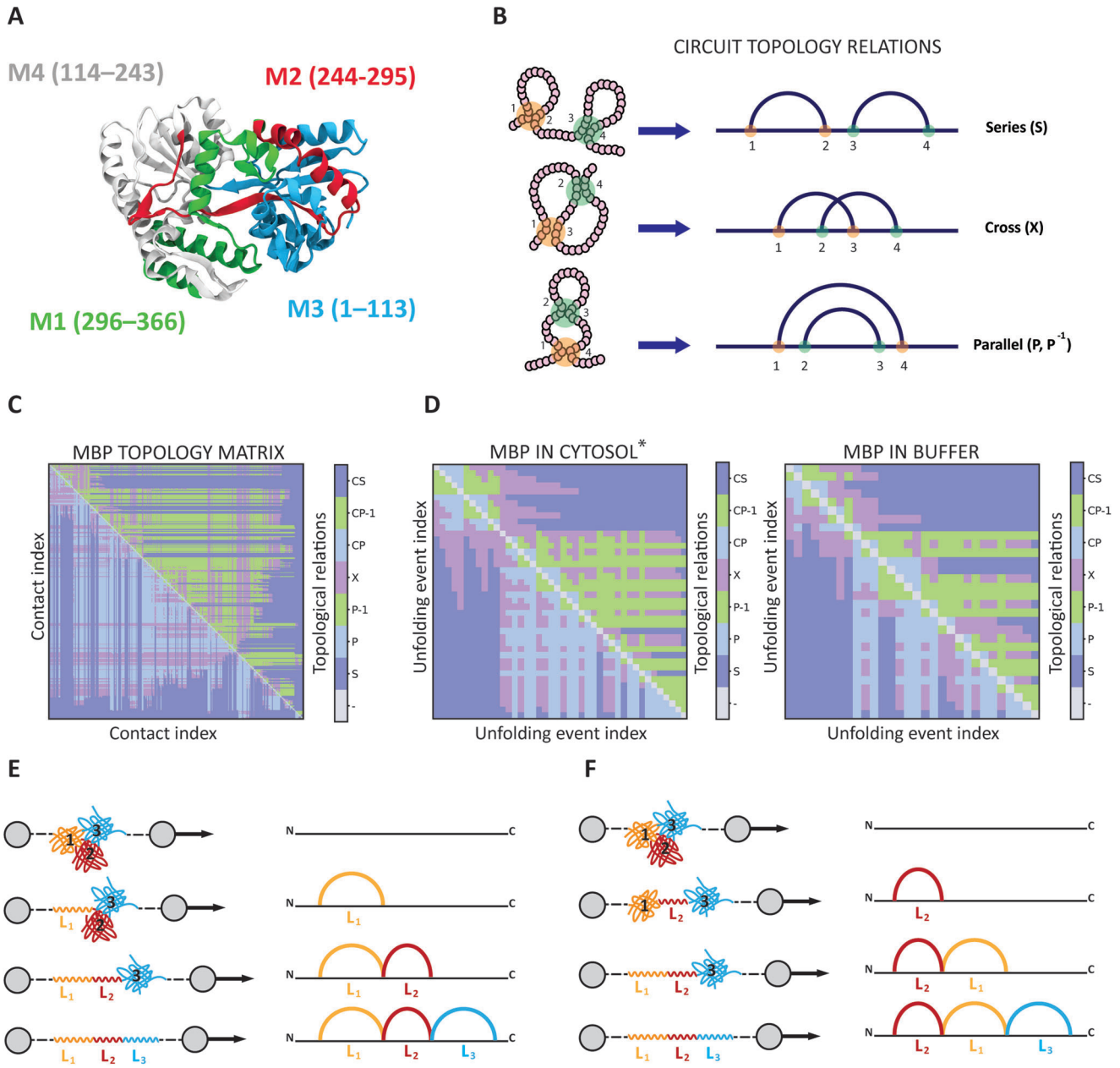

Figure 3. The unfolding pathways can be represented in the topology space. A) Native configuration of MBP. The color coding indicates the subdivision into unfoldons, as identified by Bertz and Rief.[43] B) Summary of the topological relations between loops and their representation as circuit diagram. C) Topology matrix of MBP obtained from its 3D PDB structure. Along rows and columns, specific residue-residue contacts are represented. Each element of the matrix encodes the topological relations between a pair of contacts. In the legend, CP and CS indicate concerted relations: these are a particular class of series and parallel relations, where one contact site is shared among the two contacts. For simplicity, we include these in the count of series and parallel contacts, in the color scheme of the matrix. D) Topology matrix obtained from the unfolding pathway circuit diagram, for transitions in buffer and cytosol solution. The unfolding data used for matrix construction is the same depicted in Figure 2. The label cytosol* indicates a 4 times dilution of cytosol extract in buffer. E) Cartoon of protein pulling where unfolding occurs sequentially, from one end to the other, resulting in a circuit diagram of contour lengths that can be overlapped to protein sequence ($L_1 + L_2 + L_3$). F) Cartoon of protein pulling where unfolding does not occur sequentially, with respective contour length diagram ($L_2 + L_1 + L_3$).

of circuit topology (CT, Figure 3). CT has proven to be successful in the characterization of protein folding in silico.[44–46] The main assumption of CT is that any two pairs of loops created by intra-chain contacts can be in either one of three topological relations, as exemplified by Figure 3B. An overall view of the topology of a biopolymer is then encoded into a topology matrix, for instance for MBP (Figure 3C), where each element represents the topological relation between a specific pair of contact-induced loops. Whenever a contact is disrupted in a force spectroscopy assay, the associated loop is released, resulting in a jump in the

Table 1. All intermediate states after the I transition happen more frequently and at higher forces in cell extract.

	F_B (pN)	F_C (pN)	State % in buffer	State % in cytosol*
0 nm	11.5 ± 0.5	15.7 ± 1.5	12.5% (N = 41)	13.7% (N = 34)
23 nm	28.1 ± 0.6	27.7 ± 0.9	47.6% (N = 156)	28.2% (N = 70)
39 nm	18.4 ± 3.0	25.2 ± 2.1	3.3% (N = 11)	10.0% (N = 25)
60 nm	12.7 ± 1.4	25.6 ± 3. 0	2.7% (N = 9)	6.04% (N = 15)
81 nm	20.0 ± 2.6	27.2 ± 3. 6	2.1% (N = 7)	3.6% (N = 9)

Errors on unfolding forces represent standard errors on the mean. The number N of unfolding events for each intermediate state is reported in brackets. The total number of intermediate unfolding events observed for MBP in cytosol solution are N = 152 (from 183 pulls, 42 molecules), while in buffer N = 224 (from 221 pulls, 36 molecules). Percentages of observed states include also the fully unfolded state (120 nm). According to the classification used in this paper,[43] the intermediate state at 23 nm corresponds to the release of the first unfoldon (M1). Contour lengths at 39 and 81 nm correspond to unfolding of M2 and M3 respectively. The label cytosol* indicates a four times dilution of cytosol extract in buffer.

force/extension diagram of the protein. We can build a topology matrix from any circuit diagram (Figure 2C), by assigning to each arch pair in the diagram a topological relation as described in Figure 3B. The topology matrix of the contour length diagram (Figure 3D) shares the general structure with the one extracted from the native structure of MBP (Figure 3C), specifically the presence of roughly 3/4 lobes made of parallel/cross relations elongating from the diagonal. These lobes correspond to compact, entangled structures. However, the two types of topology matrix do not necessarily always look alike. Structural topology matrices are obtained by drawing a diagram of contacts connecting contact sites, that are then encoded in the topology space (according to the CT formalism) by walking along the sequence from one terminus to the other, generally from N to C. If the unfoldons of a protein were to unfold from one terminus to the other, then one could match the observed unfolding contour lengths back to the structure of the protein. In this case, it would be easy to couple the lobes observed in the unfolding-derived topology matrix with the ones observed in the structural one. (Figure 3E,F). If the unfoldons do not unfold sequentially, one would most likely observe different patterns in the two types of topology matrix. In this sense, a significant mismatch of the overall structure of the topology matrices could help us understand the order in which unfoldons are released in a force spectroscopy experiment.

In the case of MBP, the unfolding is almost completely sequential, from C terminus to N terminus, although the stretch between 1 to 113 (M3) residues unfolds (when this particular intermediate state is visited) before the 114 −243 stretch (M4). The cross/parallel-rich lobe on the upper left part of the matrix indicates the initial, C-terminal unfolding (between 0 and 40 nm in contour length, approximately). The remaining cross-parallel block underneath the lobe corresponds to the unfolding of the MBP core. Here the buffer matrix presents long P/X stripes separated by series relations, indicating that intermediate states are disjointed and easy to identify, while in cytosol solution we do not see this separation, indicating that the compounds in solution are compacting the structure of the core. In the language of CT, when P/X-rich regions are separated by series stretches, we are generally observing tightly knit domains that are disjointed with each other. In this case, unfolding events are disjointed, indicating a clear unfolding hierarchy.

We show here this type of visualization as a proof-of-concept of the potential of topology as a bridge between force spectroscopy data and structural analysis. Further technical development of the framework is needed for the quantitative analysis of these matrices. One limitation is represented by the fact that buffer and cytosolic solution matrices do not have a one-to-one correspondence. Only occurring unfolding events are represented in the matrices. Since events occurring in the two media are not the same in terms of quantity and quality, we do not find correspondence between rows and columns of the matrices, making it impossible to, for example, subtract the two matrices to visualize differences. However qualitatively, we believe CT could be beneficial in the visualization of characteristic patterns, especially in very complex systems. Attributing features in force-extension curves to protein structure is one of the biggest challenges in this type of experiments. MBP is a model protein with a relatively simple unfolding pattern. When faced with bigger, multi-state folding proteins, visualization within the topology space might highlight features not readily visible with traditional methods. As such, matching similarities and differences between the topology of the native structure and that of the unfolding pathway might provide clues to understand protein conformations.

3. Conclusion

Proteins undergo a variety of conformational changes and interactions in the cellular environment. Current single-molecule experiments are performed in buffer, either in isolation[7–9] or in the presence of specific chaperone systems and ligands.[12,14,15,41] However, we are far from being able to portray protein folding as it happens in vivo, in the presence of the full cytosolic proteome. Here, we took a first step in this direction by performing the first-ever single-molecule mechanical interrogation of protein and DNA molecules in a cytosol. We explored how the cytosolic molecular machinery affects the mechanical properties of DNA, which need to be taken into account when using chimeric molecular constructs for protein pulling experiments. One could speculate that the observed lower values of persistence length L_p and elastic modulus K could be caused by the action of helicases, separating locally the two sdDNA strands, resulting overall in a more stretchable and bendable molecular structure. Moreover, we performed the first MBP unfolding assay in the presence of the full cytosolic proteome, where MBP acts as a substrate to observe the cumulative topological action of the molecular machinery. What we observed is an overall stabilization of the rare folding intermediates situated mostly in the N-terminal half of the protein (1–243 residues), indicating most likely the rescuing of partial folds by chaperones to prevent aggregation and to mitigate stress-induced unfolding. Expectedly, this leads to an enhaced refolding efficiency once the unfoldigng stress is resolved.

Our findings concerning the stabilization of intermediate states are compatible with previous results reported for chaperones trigger factor[15,47] and Dnak and GroEL in the ADP state.[14] A particularly striking difference is the emergence of an intermediate state located at 39 nm. Such state was previously detected in AFM studies,[43] but not in optical tweezers assays, to the best of our knowledge. We note that a 39 nm intermediate indicates

chaperone working ranges equal or shorter than 39 nm. According to previous studies, this size is compatible with the working range of both TF and DnaK.[48] However, our study as it is implemented in this paper cannot pinpoint any specific chaperone-client interaction. Regardless, we believe it is meaningful to observe what is the overall effect of the cytosolic molecular machinery over forced unfolding, to test which of the various chaperone mechanisms are dominant. Moreover, there is an abundance of biological information linked to the intricate and multifaceted environment within cells that have been largely overlooked in a single molecule in vitro experiments. We believe our approach can shed light on such complexity.

Experiments performed in environments aimed at mirroring real biological environments could serve as a validation of protein folding parameters obtained in in vitro, while also provide a closer picture to what these processes really look like in the cell. It is worth noting that the findings reported in this paper present a few limitations. First of all, we focused on a single model protein, MBP, which is widely known and used for chaperone–protein interaction studies.[12,14,15,41–43] Moreover, we present here findings obtained with a specific dilution of cell extract to highlight cytoplasmic contributions beyond the widely-known non-specific crowding effects. Here, our aim is to present a proof-of-concept study to explore methods for environments that closely mimic cellular context. However, several next steps can readily be envisioned moving forward in this research line. Pulling experiments in undiluted cytosol could elucidate the added effect of concentration and crowding on protein folding; [49] the use of active calibration[50] would be advisable in this case, to determine the local viscosity perceived by the beads. The CT-based pipeline for visualization and characterization of force spectroscopy-derived folding pathways could prove fundamental for tracking those bacterial processes where chaperone content undergoes critical changes. One example is antibiotic resistance in *E. coli*, which is directly linked to chaperone function in the cytoplasm.[51–53] In this assay, a model protein, MBP, is used as a sensor for the cytosolic interactome, effectively converting topology into a proxy biomarker for chaperone activity. It is still unclear whether the results obtained in this paper could be generalized to other models or more complex proteins. However, further experiments with eukaryotic cells and human proteins are worth to be undertaken, as they would increase the complexity of the systems as well as provide crucial information about disease-triggering phenomena such as protein aggregation and misfolding.

4. Experimental Section

Protein Expression and Purification: Protein expression and purification followed the protocol presented by Avellaneda et al.[28] MBP was modified with cysteine residues using the pET28 vector. Proteins were expressed *E. coli* BL21(DE3) cells with 0.3 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) overnight at 18 °C. Cells were cooled, and collected by centrifugation at 5000 × *g* for 20 min. Cell pellets were resuspended in ice-cold buffer A (50 mM potassium phosphate pH 7.5, 0.15 M NaCl, 3 mM chloramphenicol, 50 mM Glu-Arg, 10 mM Complete Protease Inhibitor Ultra from Roche, 10 mM EDTA) and lysed using a French Press homogenizer. The lysate was centrifuged at 50 000 × *g* for 60 min at 4 °C and incubated with Amylose resin (New England Biolabs) in a gravity column for 20 min at 4 °C. The resin was washed with buffer B (50 mM potassium phosphate pH 7.5, 0.2 M NaCl, 1 mM EDTA, 5 mM DTT) three times and bound proteins were eluted in buffer B supplemented with 20 mM maltose. Purified proteins were analyzed on an SDS gel, aliquoted, flash-frozen in liquid nitrogen, and stored at −80 °C.

Protein-Anchor Coupling Protocol: Protein coupling was performed following the procedure presented by Avellaneda et al.[28] Purified proteins were thawed and passed through a desalting column (PD-10, GE Healthcare) to exchange the buffer to a coupling buffer (Sodium Phosphate 100 mM pH 7.2, NaCl 150 mM, EDTA 10 mM). The proteins were concentrated using 10K Amicon Centrifugal filter at 14 000 × *g* for 10 min. Anchor oligos 5′-modified with maleimide (purchased from biomers.net) were mixed with the protein 4:1 ratio and incubated for 45 min at RT. 5 mM tris(2-carboxyethyl) phosphine (TCEP) was added and incubated for a further 30 min to increase the coupling yield. Uncoupled oligos were removed by affinity chromatography using amylose resin (NEB). The protein-oligo complex was diluted and ligated to 2.5 kbp DNA handles (produced by the Tans lab as per Avellaneda et al.[28]) at a 1:1 ratio with T4 ligase for 1–2 h at room temperature.

DNA Tightrope Generation: DNA tethers were produced by the Tans lab; 5 kb DNA tightropes were generated by PCR using biotin and digoxigenin functionalized primers from the pOSIP-TT plasmid.

Cytosol Extraction: The cytosol solution was obtained from 50 mL DH5α cells transformed with pETM14 grown to an OD660 of 4 to ensure the culture was in the stationary phase[54] and a wider range of homeostatic important proteins were present. The culture was then diluted to an OD600 of 1 to decrease the protein concentration, therefore yielding the empirical 4 dilution factor for the experiments. The cells were centrifuged at 4000 × *g* for 10–20 min at 4 °C. The supernatant was discarded and the cell pellet was diluted in 2 mL HMK buffer (50 mM HEPES, pH 7.5, 5 mM $MgCl_2$, 100 mM KCl), adjusting the final pH to 7.5. The bacterial cells were lysed through five freeze–thaw cycles: for each cycle, it was flash-frozen in liquid nitrogen and subsequently left to thaw on ice. The solution was subsequently filtered and aliquoted for use in OT experiments. Final protein concentration in the cytosol solution was quantified using a Qubit Protein Assay (Invitrogen) to 0.256 mg mL^{-1}.

AFS Viscosity Measurement: In order to measure the viscosity of the cytosol solution, 4.07 µm silica beads (SSD5002, Bangs Laboratories) were added to the cell extract and the solution was injected into the AFS chamber. Beads were then subjected to an acoustic wave until they detached from the surface (voltage: 10–30%). Subsequently, the acoustic wave was turned off, and the beads were left free to settle again on the bottom. The *z*-position of the bead was tracked by observing the diffraction pattern of the bead image and comparing it with a previously calibrated look-up table. The terminal velocity v_t of the bead was extracted by fitting the *z*/time plot in the linear range. The viscosity η' was then obtained by the following equation:

$$\eta' = \frac{gd^2}{18v_t}(\rho_S - \rho) \tag{1}$$

where d and ρ_S are respectively diameter and density of the beads, as reported by the manufacturer, g is the gravitational acceleration, and ρ is the density of the solution. The viscosity values were then corrected by a factor accounting for surface proximity: [55]

$$\eta = \eta'\left(1 - \frac{9r}{8z} + \frac{r^3}{2z^3} - \frac{57r^4}{100z^4} + \frac{r^5}{5z^5} + \frac{7r^{11}}{200z^{11}}\right) \tag{2}$$

where r is the radius of the bead.

Optical Tweezers Assay: MBP-DNA and DNA tethers were incubated for 20 min on a rotary mixer (4 °C) with 15 µL HMK buffer (50 mM HEPES, pH 7.5, 5 mM $MgCl_2$, 100 mM KCl) and 2.12 µm Anti-Dig coated Polystyrene beads (DIGP-20-2, Spherotech). The solution was then rediluted in 120 µL HMK buffer. Connection with 2.08 µm Neutravidin coated Polystyrene beads (NVP-20-5, Spherotech), on the other terminus of the construct, was created during the OT assay. Particles were trapped and brought in close proximity for a few seconds, allowing the tether to bind. Pulling was performed at a constant speed of 200 nm s^{-1}.

Table 2. Persistence length and elastic modulus of DNA overhangs from WLC fitting.

	Mean (buffer)	St. dev (buffer)	Mean (cytosol*)	St.dev (cytosol*)
K' (pN)	774	111	577	95
L'_p (nm)	25	4	24	7

The label cytosol* indicates a four times dilution of cytosol extract in buffer.

Statistical Analysis: Tethers were selected in accordance with the following criteria: DNA overstretching in the expected range of ≈65 pN, and pulling tether size matching the expected construct length. We obtained a total of 44 DNA pulls in buffer, from 9 molecules (5 s waiting time), 67 DNA pulls from 12 molecules in cytosol solution at 5 s waiting time, and 43 pulls from 14 molecules in cytosol solution at 30 s waiting time. For MBP pulling experiments, a total of 221 pulls were measured from 36 molecules in buffer and 183 pulls were measured from 42 molecules in cytosol solution. The WLC fitting were performed with the same procedure presented by Avellaneda et al.,[28] using the approximation of an extensible polymer for DNA,[56] and the Odijk inextensible approximation[57] to account for protein parameters:

$$x = L'_C\left(\frac{4}{3}\left(1-\frac{1}{\sqrt{\beta'+1}}\right)-\left(\frac{10\exp\left(\sqrt[4]{\frac{900}{\beta'}}\right)}{\sqrt{\beta'}\left(\exp\left(\sqrt[4]{\frac{900}{\beta'}}\right)-1\right)^2}+\frac{{\beta'}^{1.62}}{3.55+3.8{\beta'}^{2.2}}\right)+L_c\left(1-\frac{1}{2}\sqrt{\beta}\right)\right) \quad (3)$$

Here, $\beta' = (FL'_p)/(k_BT)$, where T is the temperature and F the force. The constants L'_p, K, and L'_c are persistence length, elastic modulus, and contour length of the DNA molecule (or overhangs). L'_c was set to 1700 in both DNA and protein pulling assays, as the two overhangs are 2.5 kbp each, resulting in a total of 5 kbp. The constants L'_p and K were fit separately for each tether. Often these parameters had to be re-fit for each different pull when performing experiments in cytosol solution, to account for the environmental effects on the DNA tether described in this paper. The average and standard deviation of these values in buffer and cytosol solution are reported in **Table 2**.

For the protein contribution, $\beta = (FL_p)/(k_BT)$, L_p and L_c representing the persistence and the contour length, set respectively to 120 and 0.75 nm. The pulling force was 200 nm s^{-1}. For the histograms in Figure 2, all rupture forces happening at all contour lengths were retained, in order to provide a full picture of all differences between unfolding in buffer and in cytosol solution.

In order to compare viscosity distributions in buffer and cytosol solution (Figure 1A), the Kolmogorov–Smirnov test (two-tailed) was used, in view of the non-normality of the cytosolic viscosity distribution, assessed by Shapiro–Wilk test ($p = 0.011$, two-tailed).

Circuit Topology Analysis: The topology matrix was obtained by custom code, published previously by Moes et al.[58] Cutoffs used for contact identification were: distance cutoff = 4.5 Å, number of atom–atom contacts = 5, neighbor exclusion = 4. The contacts were numbered from N to C terminus (residue–residue).

In order to extract the topology matrix from the unfolding transition diagram (Figure 2C), all transitions happening between the same two 5 nm bins were incorporated into the same loop, so that they would account for one contact only. Bins were numbered, and bin indexes were used to assign topological relations between pairs of transitions (defined by initial and final contour length bin) by the following mathematical relations: [20]

$$C_{i,j}\mathbf{SC}_{r,s} \Leftrightarrow [i,j]\cap[r,s] = \varnothing \quad (4)$$

$$C_{i,j}\mathbf{PC}_{r,s} \Leftrightarrow [i,j]\subset(r,s) \quad (5)$$

$$C_{i,j}\mathbf{XC}_{r,s} \Leftrightarrow [i,j]\cap[r,s]\notin\{[i,j],\ [r,s]\}\cup\mathcal{P}(\{i,j,r,s\}) \quad (6)$$

$$C_{i,j}\mathbf{CSC}_{r,s} \Leftrightarrow ((([i,j]\cap[r,s]=\{i\})\vee([i,j]\cap[r,s]=\{j\})) \quad (7)$$

$$C_{i,j}\mathbf{CPC}_{r,s} \Leftrightarrow (([i,j]\subset[r,s])\wedge(i=r\vee j=s)) \quad (8)$$

where $\mathcal{P}$ denotes the power set, that is, all subsets of a set including the null set ($\varnothing$)). Indexes (i, j, r,s), in this case, correspond to the bin indexes of the transition. The same filtering applied for circuit diagrams (Figure 2C) was retained in the topology matrices in Figure 3.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors acknowledge Nederlandse Organisatie voor Wetenschappelijk Onderzoek (grant numbers: NWA.1228.192.309, OCENW.XS.076, OCENW.XS21.3.103), and Muscular Dystrophy Association (grant number: MDA628071) for financial support. The authors thank Cecilia de Agrela Pinto and Arjen Jakobi lab (Technical University of Delft) for help with cytosol extraction.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

A.M. conceived the project and supervised the research; B.S., L.W.H.J.H., and V.S. performed the experiments; B.S. analyzed the experimental data and performed the circuit topology modeling; V.S. and S.J.T. provided reagents and resources; B.S., L.W.H.J.H., and A.M. wrote the original version of the manuscript; All authors contributed to the revision and approved the final version of the article.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

protein folding, cytosol, single-molecule assay, optical tweezers, chaperone

Received: March 8, 2023
Revised: June 13, 2023
Published online: July 6, 2023

[1] T. Ha, A. G. Kozlov, T. M. Lohman, *Annu. Rev. Biophys.* **2012**, *41*, 295.
[2] A. A. Rebane, L. Ma, Y. Zhang, *Biophys. J.* **2016**, *110*, 441.
[3] C. M. Kaiser, D. H. Goldman, J. D. Chodera, I. Tinoco, Jr. C. Bustamante, *Science* **2011**, *334*, 1723.
[4] B. Jagannathan, S. Marqusee, *Biopolymers* **2013**, *99*, 860.

[5] I. Heller, T. P. Hoekstra, G. A. King, E. J. G. Peterman, G. J. L. Wuite, *Chem. Rev.* **2014**, *114*, 3087.
[6] C. J. Bustamante, Y. R. Chemla, S. Liu, M. D. Wang, *Nat. Rev. Methods Primers* **2021**, *1*, 25.
[7] C. Cecconi, E. A. Shank, F. W. Dahlquist, S. Marqusee, C. Bustamante, *Eur. Biophys. J.* **2008**, *37*, 729.
[8] G. Cecconi, E. A. Shank, C. Bustamante, S. Marqusee, *Science (1979)* **2005**, *309*, 2057.
[9] A. F. Oberhauser, M. Carrión-Vázquez, *J. Biol. Chem.* **2008**, *283*, 6617.
[10] A. Mashaghi, G. Kramer, D. C. Lamb, M. P. Mayer, S. J. Tans, *Chem. Rev.* **2014**, *114*, 660.
[11] H. Saibil, *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 630.
[12] P. Bechtluft, R. G. H. Van Leeuwen, M. Tyreman, D. Tomkiewicz, N. Nouwen, H. L. Tepper, A. J. M. Driessen, S. J. Tans, *Science (1979)* **2007**, *318*, 1458.
[13] R. Imamoglu, D. Balchin, M. Hayer-Hartl, F. U. Hartl, *Nat. Commun.* **2020**, *11*, 365.
[14] A. Mashaghi, S. Bezrukavnikov, D. P. Minde, A. S. Wentink, R. Kityk, B. Zachmann-Brand, M. P. Mayer, G. Kramer, B. Bukau, S. J. Tans, *Nature* **2016**, *539*, 448.
[15] A. Mashaghi, G. Kramer, P. Bechtluft, B. Zachmann-Brand, A. J. M. Driessen, B. Bukau, S. J. Tans, *Nature* **2013**, *500*, 98.
[16] M. J. Avellaneda, E. J. Koers, M. M. Naqvi, S. J. Tans, *Protein Sci.* **2017**, *26*, 1291.
[17] J. C. Young, V. R. Agashe, K. Siegers, F. U. Hartl, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 781.
[18] J. Yuan, C. Chyan, H. Zhou, T. Chung, H. Peng, G. Ping, G. Yang, *Protein Sci.* **2008**, *17*, 2156.
[19] A. Mashaghi, *Not. Am. Math. Soc.* **2021**, *68*, 420.
[20] A. Mashaghi, R. J. van Wijk, S. J. Tans, *Structure* **2014**, *22*, 1227.
[21] D. Moes, E. Banijamali, V. Sheikhhassani, B. Scalvini, J. Woodard, A. Mashaghi, *MethodsX* **2022**, 101861.
[22] B. Scalvini, V. Sheikhhassani, J. Woodard, J. Aupič, R. T. Dame, R. Jerala, A. Mashaghi, *Trends Chem* **2020**, *2*, 609.
[23] I. Golding, E. C. Cox, *Physical Review Letters* **2006**, *098102*, 14.
[24] Y. Jun, S. K. Tripathy, B. R. J. Narayanareddy, M. K. Mattson-Hoss, S. P. Gross, *Biophys. J.* **2014**, *107*, 1474.
[25] T. M. J. Evers, V. Sheikhhassani, M. C. Haks, C. Storm, T. H. M. Ottenhoff, A. Mashaghi, *iScience* **2022**, *25*, 103555.
[26] E. Pfitzner, C. Wachauf, F. Kilchherr, B. Pelz, W. M. Shih, M. Rief, H. Dietz, *Angew. Chem., Int. Ed.* **2013**, *52*, 7766.
[27] Y. Hao, C. Canavan, S. S. Taylor, R. A. Maillard, *Sci. Rep.* **2017**, *7*, 10843.
[28] M. J. Avellaneda, E. J. Koers, D. P. Minde, V. Sunderlikova, S. J. Tans, *Commun Chem* **2020**, *3*, 20.
[29] A. Aggarwal, S. Naskar, A. K. Sahoo, S. Mogurampelly, A. Garai, P. K. Maiti, *Curr. Opin. Struct. Biol.* **2020**, *64*, 42.
[30] C. Bouchiat, M. D. Wang, J. F. Allemand, T. Strick, S. M. Block, V. Croquette, *Biophys. J.* **1999**, *76*, 409.
[31] S. B. Smith, Y. Cui, C. Bustamante, *Science* **1996**, *271*, 795.
[32] G. A. King, P. Gross, U. Bockelmann, M. Modesti, G. J. L. Wuite, E. J. G. Peterman, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 3859.
[33] H. Fu, H. Chen, J. F. Marko, J. Yan, *Nucleic Acids Res.* **2010**, *38*, 5594.
[34] J. Van Mameren, P. Gross, G. Farge, P. Hooijman, M. Modesti, M. Falkenberg, G. J. L. Wuite, E. J. G. Peterman, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 18231.
[35] M. J. McCauley, L. Shokri, J. Sefcikova, Č. Venclovas, P. J. Beuning, M. C. Williams, *ACS Chem. Biol.* **2008**, *3*, 577.
[36] R. R. Meyer, P. S. Laine, *Microbiological Reviews* **1990**, *54*, 342.
[37] L. Shokri, M. J. McCauley, I. Rouzina, M. C. Williams, *Biophys. J.* **2008**, *95*, 1248.
[38] J. R. Moffitt, Y. R. Chemla, D. Izhaky, C. Bustamante, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 9006.
[39] F. Moayed, A. Mashaghi, S. J. Tans, *PLoS One* **2013**, *8*, e54440.
[40] D. B. Ritchie, M. T. Woodside, *Curr. Opin. Struct. Biol.* **2015**, *34*, 43.
[41] V. Aggarwal, S. R. Kulothungan, M. M. Balamurali, S. R. Saranya, R. Varadarajan, S. R. K. Ainavarapu, *J. Biol. Chem.* **2011**, *286*, 28056.
[42] M. M. Naqvi, M. J. Avellaneda, A. Roth, E. J. Koers, A. Roland, V. Sunderlikova, G. Kramer, H. S. Rye, S. J. Tans, *Sci. Adv.* **2022**, *8*, eabl6293.
[43] M. Bertz, M. Rief, *J. Mol. Biol.* **2008**, *378*, 447.
[44] M. Heidari, H. Schiessel, A. Mashaghi, *ACS Cent. Sci.* **2020**, *6*, 839.
[45] B. Scalvini, V. Sheikhhassani, A. Mashaghi, *Phys. Chem. Chem. Phys.* **2021**, *23*, 21316.
[46] A. Mugler, S. J. Tans, A. Mashaghi, *Phys. Chem. Chem. Phys.* **2014**, *16*, 22537.
[47] K. Singhal, J. Vreede, A. Mashaghi, S. J. Tans, P. G. Bolhuis, *PLoS Comput. Biol.* **2015**, *11*, e1004444.
[48] S. A. Teter, W. A. Houry, D. Ang, T. Tradler, D. Rockabrand, G. Fischer, P. Blum, C. Georgopoulos, F. U. Hartl, *Cell* **1999**, *97*, 755.
[49] N. Tokuriki, M. Kinjo, S. Negi, M. Hoshino, Y. Goto, I. Urabe, T. Yomo, *Protein Sci.* **2004**, *13*, 125.
[50] M. Fischer, A. C. Richardson, S. N. S. Reihani, L. B. Oddershede, K. Berg-Sørensen, *Rev. Sci. Instrum.* **2010**, *81*, 015103.
[51] K. I. Wolska, E. Bugajska, D. Jurkiewicz, M. Kuć, A. Jóźwik, *Microbial Drug Resistance* **2000**, *6*, 119.
[52] L. Goltermann, M. v. Sarusie, T. Bentin, *Front. Microbiol.* **2016**, *6*, 1572.
[53] F. Chiappori, M. Fumian, L. Milanesi, I. Merelli, *PLoS One* **2015**, *10*, e0124563.
[54] V. A. Veselovsky, M. S. Dyachkova, D. A. Bespiatykh, R. A. Yunes, E. A. Shitikov, P. S. Polyaeva, V. N. Danilenko, E. I. Olekhnovich, K. M. Klimina, *Microorganisms* **2022**, *10*, 1683.
[55] V. Romanov, G. Silvani, H. Zhu, C. D. Cox, B. Martinac, *Small* **2021**, *17*, 2005759.
[56] R. Petrosyan, *Rheol. Acta* **2017**, *56*, 21.
[57] T. Odijk, *Macromolecules* **1995**, *28*, 7016.
[58] D. Moes, E. Banijamali, V. Sheikhhassani, B. Scalvini, J. Woodard, A. Mashaghi, *MethodsX* **2022**, *9*, 101861.

