Supporting Online Material

Direct observation of chaperone-induced changes in a protein folding pathway.

Philipp Bechtluft[1*], Ruud G. H. van Leeuwen[2,3*], Matthew Tyreman[2*], Danuta Tomkiewicz[1], Nico Nouwen[1,5], Harald L. Tepper[2,4], Arnold J. M. Driessen[1], and Sander J. Tans[2,6]

1. Department of Molecular Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute and the Zernike Institute for Advanced Materials, University of Groningen, Kerklaan 30, 9751 NN Haren, the Netherlands.

2. FOM Institute for Atomic and Molecular Physics (AMOLF), Kruislaan 407, 1098 SJ Amsterdam, the Netherlands.

3. Current address: Philips Research, High Tech Campus 36, 5656 AE Eindhoven, The Netherlands.

4. Current address: McKinsey & Company, Amstel 344, 1017 AS Amsterdam.

5. Current address: Laboratoire de Symbiosis Tropicales et Méditerranéennes TA A-82/J, Campus International de Baillarguet, 34398 Montpellier cedex 5, France.

6. Netherlands Institute of Systems Biology (NISB).

* equally contributing authors

This PDF file includes:

Material and Methods

Figures S1 to S4

References

Materials and Methods

Strains and materials

Escherichia coli strain BL21.1 (BL21 (λDE3*) leu*::Tn10 *secA51*(Ts) (*1*) was used to overproduce the MBP constructs. SecA (*2*) and His-tagged SecB (*3*) were purified as described. *E. coli* strain SF100 (F-, Δ*lacX74, galK, thi, rpsL, strA* Δ*phoA*(pvuII), Δ*ompT*) (*4*) and NN100 (SF100, Δ(*uncB-C*) *zid*::Tn10) (*5*) transformed with pET605 (SecYEG) were grown in the presence of 0.5 mM IPTG, and membrane vesicles were isolated as described (*6*).

Cloning and modification of the *malE* gene

A unique cysteine prior to a GRGS-linker was introduced at the N terminus of MBP and one repeat of the sequence encoding the c-myc-tag (EQKLISEEDL) was introduced at the C terminus of MBP using PCR with plasmid pNN226 as a template yielding pPB01 (*malE*-myc1). In independent restriction reactions by *Afl*III/*Sal*I as well as *Afl*III/*Xho*I followed by ligation, first two then four c-myc sequences were added to the 3' end of the *malE* gene, yielding plasmid pPB02 (*malE*-myc2) and pPB03 (*malE*-myc4). The product of the latter was used for the optical tweezers experiments with single MBP.

To construct plasmids encoding multiple copies of the *malE* gene a *Sal*I restriction site was introduced between the unique cysteine and the linker peptide GRGS at the 5' end of the *malE* gene by PCR using plasmid pNN226 as a template. The fragment was digested with *Xba*I and *EcoR*I (flanking *malE* with the introduced modifications) and ligated into the *Xba*I/*EcoR*I site of pUC18 to yield pPB04. The *malE* gene with myc4 was subcloned from pPB03 into the *Xba*I/*EcoR*I sites of pUC18, yielding pPB05. To generate a plasmid containing two consecutive *malE* genes connected by a GRGS linker, a fragment corresponding to GRGS-MBP was derived from *Afl*III/*Xho*I digested pPB05 which was ligated to a MBP containing fragment derived from *Sal*I/*Afl*III digested pPB04, to yield pPB06 (*malE2*). A quadruple repeat of the *malE* gene was obtained in a similar manner by ligating the *Xho*I/*Afl*III fragment of pPB06 to the *Sal*I/*Afl*III fragment of pPB06, resulting in pPB08 (*malE*4). Finally, the myc4 sequence was introduced at the 3' end of

*malE*4 by ligation of the *Xba*I/*Xho*I fragments of pPB08 and pPB03, yielding pPB09 that encodes MKCGRGS-(4MBP)-myc4.

For the introduction of single mutations into the *malE* gene, the gene was cloned by PCR adding a *Nde*I site at the N-terminus (CCAACAAGGACCATAGCATATGAAAATAAAAACAGGTGC) and *Hind*III site at the C-terminus (CGCATCCGGCATTTCACAAGCTTACTTGGTGATACGAGTC). Newly created restriction sites are underlined. After digestion with *Nde*I and *Hind*III, the *malE* gene was ligated into the *Nde*I/*Hind*III sites of pET3a (pNN226). A unique cysteine residue was introduced at position 51 of the mature domain using the QuikChange site directed mutagenesis kit (Stratagene) resulting in plasmid pNN227. This plasmid was as template for the introduction of the point mutations into the MBP mature domain (Table 1).

Table 1. List of plasmids used in this study

Plasmid	Mutation	Codon
pNN226	-	
pNN227	A51C	GCG/TGC
pEK215	A51C, V8G	GTA/GGC
pEK216	A51C, A276G	GCG/GGG
pEK217	A51C, Y283D	TAT/GAT
pEK218	A51C, T345I	ACC/ATC

Expression and purification of preMBP, MBP and 4MBP proteins

The MBP and 4MBP proteins were expressed at 37º C in *E. coli* strain BL21.1 and the proteins were purified from the crude cell extract using amylose resin affinity chromatography (New England BioLabs). PreMBP and derivatives were overproduced at 30 °C in *E. coli* BL21.1 and purified from inclusion bodies by anion exchange chromatography (HiTap Q Sepharose column, Amersham Pharmacia Biotech) after solubilization in 50 mM Tris pH 9.0, 6 M urea, The unique N-terminal cysteine residue in

the purified MBP/4MBP proteins was labeled with biotin-maleimide (Molecular Probes) (*7*).

Microsphere preparation

Anti-c-myc and anti-digoxigenin antibodies (Roche Diagnostics) were covalently coupled to carboxyl polystyrene beads (1.88 µm, Spherotech) using a carbodiimide crosslinking kit (Polysciences). To prevent unspecific binding of MBP to the polystyrene surface the antibody-coated beads were incubated with 1% (w/v) bovine serum albumin (BSA, Sigma Aldrich) and stored at 4°C until use.

MBP-coated microspheres were made by mixing 1 µl purified biotinylated, c-myc-tagged MBP or 4MBP (2mg/ml) and 2 µl anti-c-myc beads in 20 µl HMK (50 mM Hepes, pH 7.6, 100 mM KCl, 5 mM MgCl2) 0.1%BSA buffer. After 30 minutes incubation on a rotary mixer (4ºC), the beads were dissolved in 400 µl HMK/0.1%BSA buffer for use in optical tweezers experiments.

A dsDNA linker was produced by PCR using primers containing either two biotin or two digoxigenin groups (MWG-Biotech AG) and the plasmid pUC19 (New England BioLabs). The product of this PCR is a 2553bps dsDNA linker with two digoxigenin groups at the 5' end and two biotin groups on the 3' end.

DNA-coated microspheres were made by first incubating 0.4 µg/ml streptavidin (Molecular Probes) with ~250 ng of the digoxigenin and biotinylated dsDNA linker in 10 µl HMK/0.1%BSA buffer for 15 minutes. Next, 2 µl anti-dig beads were diluted in 10 µl HMK/0.1%BSA buffer and mixed with the DNA/streptavidin solution. After 30 minutes on a rotary mixer (4°C), the polystyrene beads were resuspended in 400 µl HMK/0.1%BSA buffer for use in optical tweezers experiments.

Optical tweezers

For trapping, a Nd:YVO_4 laser (Spectra Physics, λ=1064 nm, maximum power 5.4 W) was used. The trap stiffness in the pulling direction was 169±24 pN/µm for a 1.88 µm microsphere. Detection of forces on the trapped microsphere was performed using back focal plane interferometry (*8*). Forces were recorded at 50 Hz after application of an anti-alias filter at 20 Hz. During the experiments, a piezo-nanopositioning stage (Physik

Instrumente) was used to move the sample cell and micropipette at a speed of 50 nm/s. This resulted in a pulling rate on the tethered MBP construct of ~5 pN/s at unfolding.

Translocation ATPase assays

ATPase assays were performed at 37 ºC in 50 mM HEPES/KOH pH 7.5, 5 mM $MgCl_2$, 50 mM KCl, 2 mM DTT, 0.1 mg/ml BSA, 1 mM ATP, with or without 53 μg/ml SecB, 0-140 μg/ml SecA and 500 μg/ml *E. coli* NN100 membrane vesicles containing high levels of SecYEG. Assays were started by the addition of urea-denatured wild-type or mutant preMBP (25 μg/ml), and terminated by chilling on ice. Rates were calculated from the linear increase in released inorganic phosphate measured by the malachite green assay (*9*). All measurements were done in triplicate and values were corrected for background ATPase activity in the absence of SecA.

Molecular dynamics simulations

For the MD simulations, we used the apo-MBP structure 1JW4 from the Protein Database (*10*), with a few side chains flipped and missing atoms added by the program SCWRL 3.0 (*11*) and the PSFGEN module included in VMD1.8.3 (*12*). The system was fully solvated (with a water layer of at least 10 Å in all directions) and neutralized with sodium counterions. After initial energy minimization and a short equilibration run with the protein atoms restrained, the system was further equilibrated by a 2 ns constant pressure and temperature simulation (NPT) at 298 K and 1 bar. All simulations were carried out with NAMD (*13*), with the CHARM22 parameter set (*14*) for the protein and ions and TIP3P parameters (*15*) for the water molecules. Long-range van der Waals interactions were truncated at 10 Å (smooth interpolation from 9 Å) and long-range electrostatics were treated with the Particle Mesh Ewald technique (tolerance 10^{-6}). A Langevin thermostat was applied to non-hydrogen atoms ($\tau = 0.5\ ps^{-1}$), all bonds to hydrogens were fixed, and a timestep of 2 fs was used throughout.

To induce unfolding in the simulated protein, two external spring potentials (force constant 5 kcal/mol/Å^2 = 3.5 x 10^6 pN/μm) were applied to the nitrogen atom of Lys1 and the C_α atom of Lys370. Next, the equilibrium position of one of the two springs was moved along the vector between the initial positions of the two atoms while the other spring was kept fixed. With an applied pulling rate of 1 nm/ns, we found different results when pulling on the C-terminal or N-terminal end of the protein (while keeping the other end fixed in space). Using this fast rate, only the end that was pulled upon unfolded, while the fixed end remained folded, showing that the protein did not have sufficient time to relax. However, when we lowered the pulling speed to 0.1 nm/ns, C- or N-terminal pulling gave identical results. We are therefore confident that the applied perturbation had propagated throughout the protein. The possibility does remain that there are slower rearrangement modes in the protein that do play a role in the unfolding process, but that cannot be tested within current computational limits. Ultimately, comparison with the experiments validates the unfolding pathway. For 10 different simulations with different initial starting conditions, a 5-residue chain at the N-terminal end undocked first from the surface of the protein. A previous example of such force-induced unfolding simulations in comparison with single-molecule experiments can be found in ref. *16*.

Ideally, the abovementioned procedure would be applied to explore the complete unfolding pathway of the protein under study. This would lead to computational problems, however, because the length of the unfolded part of the protein would quickly require an intractable large computational volume. Assuming that already unfolded parts do not dock back unto the protein surface (at least not under the applied force), the problem can be circumvented by cutting of the unfolded part and resuming the procedure on the newly formed 'mutant'. This way, we performed a series of unfolding simulations. In each successive experiment, we started again from the native structure (with the respective residues cut off, new terminal atoms added, and properly solvated and neutralized). This way artificial disruption of the structure by preceding experiments was not allowed to build up in the system.

After the initial unfolding of the N-terminal segment, the structure unfolded by successive peeling off of a series of C-terminal α-helices. From the third experiment onwards, also the fast pulling rate (1 nm/ns) gave identical results independent of the

terminus that was pulled away. From that experiment onwards, we moved to a different routine in which a single spring was attached between the two terminal ends and the equilibrium length of that spring was extended at a rate of 1 nm/ns. After the sequential unfolding of 5 C-terminal α-helices, we arrived at a structure that required significantly higher force to unfold. This structure is proposed to coincide with the metastable intermediate found in the experiments (see main text).

The stability of the core structure versus the native structure is further validated by high-temperature simulations of both constructs. Multi-nanosecond runs at up to 400 K showed a very high stability of the beta-sheet structure. This structure remained largely intact in both cases, while most alpha helices would loose their secondary structure over time. In the native structure, a tendency was observed for the outer helices to partly disconnect from the surface, while a continuation of the sequential unfolding mechanism was not observed for the core construct. The latter is not surprising, since the termini of that construct are almost stripped down to the beta domain.

Supporting figures

Fig.S1

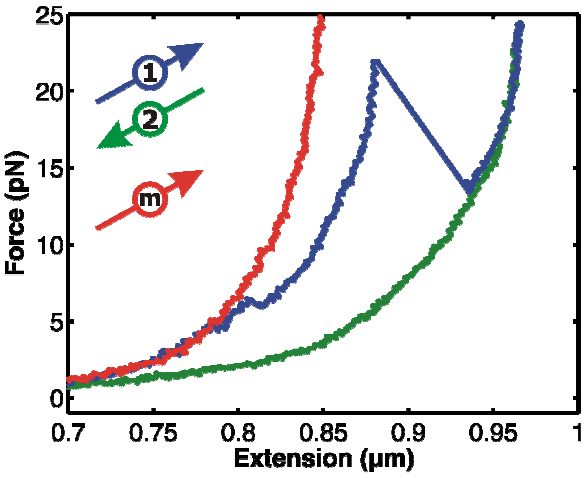

Figure S1: Comparison of force-extension curves of an MBP protein (1&2) **and of a plain biotin c-myc-tag** (m). The latter lacks the sudden extension changes, showing that they can be attributed to unfolding of the protein. The biotin c-myc-epitope was obtained from PerkinElmer.

Fig.S2

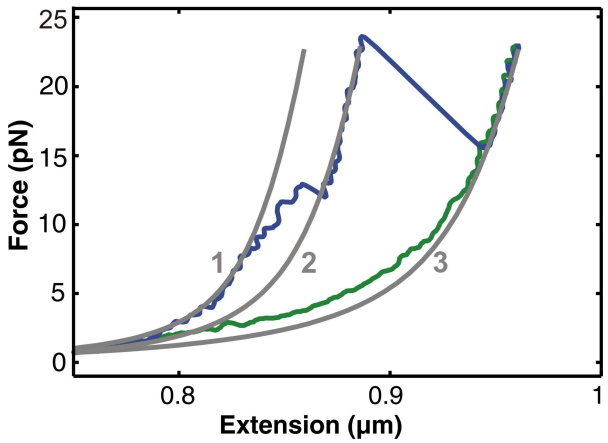

Figure S2: MBP stretching and relaxation curve compared with theoretical curves based on the worm-like chain (WLC) model. The first grey curve (1) is the WLC behaviour for the DNA tether alone, representing native folded MBP. The stretching data (blue curve) follows this curve for the lowest forces until the C-terminal segment starts to gradually detach from the native structure. The second WLC curve (2) represents the DNA tether in series with a polypeptide of 91 residues (persistence length = 2 nm), which is the length of MD predicted unfolded external α-helices. The third WLC (3) reflects the behaviour of the DNA tether in series with a fully unfolded MBP molecule (370 residues). After full unfolding the high force data fits to this theoretical curve for a non-interacting polymer, which can be understood since the polypeptide is stretched and intra-molecular interactions cannot form. During relaxation to lower forces (green curve), the data deviates from the theoretical WLC curve, showing smaller polypeptide extensions. The measured protein extension is seen in the figure as the difference between the grey line (1) and the green line (along the extension axis). The theoretical WLC protein extension is seen in the figure as the difference (along the extension axis) between the first and third grey curves. To indicate the deviation from WLC behaviour, we plot in figure 1D (main text) the average of the measured protein extension for multiple curves within a force window, as a fraction of the WLC protein extension. The energy associated

with forming the intra-molecular interactions can be estimated by the area in between the data (green curve) and the WLC behaviour representing the non-interacting chain (3). We find a non-negligible but small value of ~ 25 k_BT, which corresponds to the hydrolysis of roughly 1 ATP molecule. The rupturing of these compaction interactions requires a similar amount of energy, as evidenced by the overlap between relaxation and stretching curves (taken with short waiting times as to prevent tertiary contacts, Fig. S3).

Fig.S3

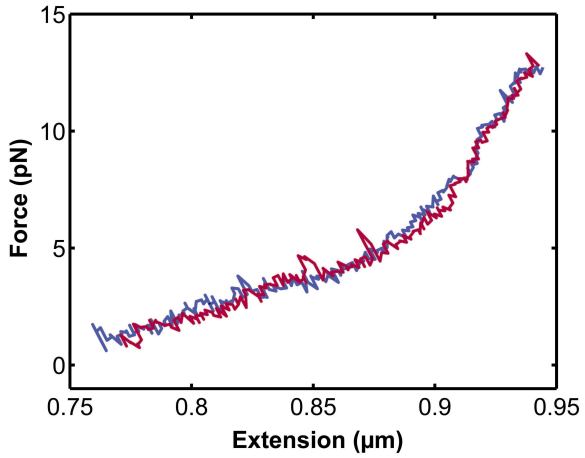

Figure S3: No refolding of MBP after short waiting time at zero force. MBP stretching curve in the absence of SecB (red), performed immediately after full extension and subsequent relaxation to low forces (blue trace).

Fig.S4

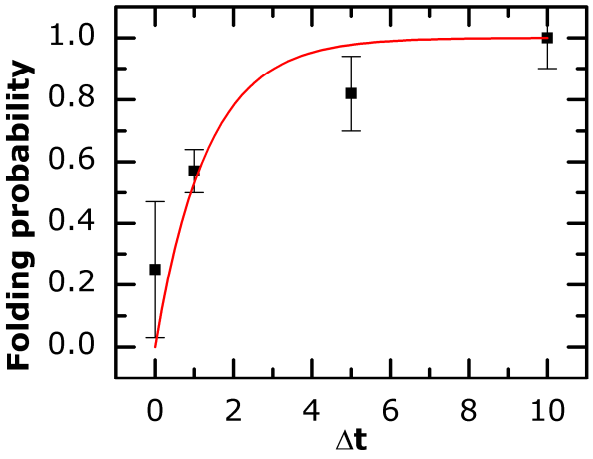

Figure S4: Probability of folding as a function of waiting time. A single MBP molecule was unfolded, relaxed to low force, and held there for a time Δt, after which the folding state was analysed by stretching the molecule. Repeating this experiment multiple times provided an estimate for the folding probability. The solid line shows a fit to the function $P(t)=1-\exp(-\Delta t\, k)$, which yielded a folding rate $k = 0.76\pm0.19\ s^{-1}$.

Supporting references

1. P. McNicholas, T. Rajapandi, D. Oliver, *J Bacteriol.* **177**, 7231 (1995).

2. R.J. Cabelli, L. Chen, P.C. Tai, D.B. Oliver, *Cell* **55**, 683 (1988).

3. P. Fekkes *et al*., *Mol Microbiol* **29**, 1179 (1998).

4. F. Baneyx, G. Georgiou, *J. Bacteriol.* **172**, 491 (1990).

5. N. Nouwen, M. van der Laan, A. J. M. Driessen, *FEBS Lett.* **508**, 103 (2001).

6. A. Kaufmann, E.H. Manting, A.K. Veenendaal, A. J. M. Driessen, C. van der Does, *Biochemistry* **38**, 9115 (1999).

7. J. de Keyzer, C. van der Does, A. J. M. Driessen, *J. Biol. Chem.* **277**, 46059 (2002).

8. F. Gittes, C. F. Schmidt, *Opt. Let.* **23**, 7 (1998).

9. R. Lill *et al*., *EMBO J.* **8**, 961(1989).

10. X. Duan, J.A. Hall, H. Nikaido, F.A. Quiocho, *J. Mol . Biol.* **306**, 1115 (2001).

11. A. A. Canutescu, A .A. Shelenkov, R. L. Dunbrack, *Protein Sci*. **12**, 2001 (2003).

12. W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **14**, 33 (1996).

13. J. C. Phillips *et al*., *J. Comput. Chem*. **26**, 1781, (2005).

14. A. MacKerrell *et al*., *J. Phys. Chem. B.* **102**, 3586 (1998).

15. W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, *J. Chem. Phys.* **79**, 926 (1983).

16. R. B. Best, B. Li, A. Steward, V. Daggett, J. Clarke, *Biophys. J.* **81**, 2344 (2001).