Folding and assembly of the large molecular machine Hsp90 studied in single-molecule experiments

Markus Jahn[a], Johannes Buchner[b,c], Thorsten Hugel[d], and Matthias Rief[a,c,1]

[a]Physik-Department, Technische Universität München, 85748 Garching, Germany; [b]Department Chemie, Technische Universität München, 85748 Garching, Germany; [c]Munich Center for Integrated Protein Science, 81377 München, Germany; and [d]Institute of Physical Chemistry, University of Freiburg, 79104 Freiburg, Germany

Edited by George H. Lorimer, University of Maryland, College Park, MD, and approved December 16, 2015 (received for review September 29, 2015)

Folding of small proteins often occurs in a two-state manner and is well understood both experimentally and theoretically. However, many proteins are much larger and often populate misfolded states, complicating their folding process significantly. Here we study the complete folding and assembly process of the 1,418 amino acid, dimeric chaperone Hsp90 using single-molecule optical tweezers. Although the isolated C-terminal domain shows two-state folding, we find that the isolated N-terminal as well as the middle domain populate ensembles of fast-forming, misfolded states. These intradomain misfolds slow down folding by an order of magnitude. Modeling folding as a competition between productive and misfolding pathways allows us to fully describe the folding kinetics. Beyond intradomain misfolding, folding of the full-length protein is further slowed by the formation of interdomain misfolds, suggesting that with growing chain lengths, such misfolds will dominate folding kinetics. Interestingly, we find that small stretching forces applied to the chain can accelerate folding by preventing the formation of cross-domain misfolding intermediates by leading the protein along productive pathways to the native state. The same effect is achieved by cotranslational folding at the ribosome in vivo.

misfolding | off-pathway | rough energy landscape | optical tweezers

Large protein machines consist of long amino acid chains, often exceeding many hundreds or even over a thousand residues in length. Although the in vitro folding of small and medium-sized proteins is relatively well understood (1–5), very limited information exists about the complete folding process of such large proteins (6). In general, larger proteins often exhibit a multitude of intermediate and aggregation-prone misfolded states (4, 7). Recently, it has been shown that in multidomain proteins with homologous domains, cross-repeat intermediates can greatly slow down productive folding (8) but little is known about how size effects influence the folding of very large (>500 residues) nonhomologous multidomain proteins.

Methods providing dynamic as well as structural information are rare, and many bulk methods often do not provide enough resolution to deal with the multitude of states expected for complex systems such as the aforementioned large protein complexes. Single-molecule force spectroscopy offers kinetic, energetic as well as coarse primary structural information combined with the possibility of actively manipulating systems, making it ideally suited for studying the folding of large proteins (5, 9–12).

In this paper, we study the folding and assembly of the large chaperone machinery heat shock protein 90 from yeast (Hsp90), a protein that needs to fold and self-assemble before it can function as a chaperone in the cell. Hsp90 consists of three domains, the N-terminal domain (N domain, 211 residues), the middle domain (M domain, 266 residues), and the C-terminal domain (C domain, 172 residues). In eukaryotic Hsp90, the N and M domains are connected by a long (62 residues) charged linker that can bind transiently to the N domain (13). In addition, the C domain is partly unstructured (residues 678–709). Hsp90 protomers form biologically active dimers through helix pairs in the C domains (residues 640–672) (14). Early equilibrium bulk studies suggest that the unfolding and refolding of isolated Hsp90 is mostly reversible and an unspecified intermediate is populated (15).

Results

Monomer Unfolding and Refolding. In the first set of experiments, we investigated the overall folding properties of the Hsp90 monomer. To this end, we designed a mutant construct carrying two cysteine-modified ubiquitin domains at each terminus. These serve as attachment points for the DNA handles used to link the construct to the trapped beads (Fig. 1*A* and *Methods*). Fig. 1*B* shows three stretch and relax cycles obtained at a slow pulling velocity of 10 nm/s in which a single Hsp90 monomer was consecutively unfolded and refolded. The unfolding traces (gray) show a characteristic unfolding pattern exhibiting three major peaks that we had previously identified as the unfolding of the C, N, and M domains of Hsp90 (13). At low forces before the domains unfold, fast transitions can be observed that arise from the rapid docking and undocking of the charged linker that connects the N and M domains ("#" in Fig. 1*B*) (13). The completely unfolded Hsp90 monomer starts refolding (purple traces) at high forces (~5 pN) through a complex sequence of near-equilibrium transitions and folding intermediates. Usually, the major refolding transitions fall on top of the unfolding traces, suggesting that the three domains refold sequentially. After successful refolding, the charged linker fluctuations also become visible. All subsequent unfolding traces show the full native Hsp90 unfolding pattern. At a faster pulling velocity of 500 nm/s, we find that just 36% of the traces completely refold, suggesting that complete refolding occurs within less than 2.5 s (see Fig. S1 and *Estimate 1*).

Refolding of the Individual Domains. To better understand the complex behavior of the refolding events observed for the monomer,

Significance

Understanding protein folding is, as yet, an unsolved question in the life sciences that has relevance for many diseases. While the folding of simple and small protein domains is well studied, for large proteins, the abundance of pathways and intermediate states makes them difficult to characterize using standard protein folding experiments. With single-molecule optical tweezers experiments, we can overcome these limitations. We observe in real time the folding of a dimeric, three-domain protein from the fully unfolded chain to the biologically active, quaternary structure. The likelihood of the folding process being hindered by misfolded intermediates increases with chain length. These misfolded states can slow down folding significantly and may lead to aggregation in vivo.

Author contributions: M.J., J.B., T.H., and M.R. designed research; M.J. performed research; M.J. and M.R. analyzed data; and M.J., J.B., T.H., and M.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: mrief@ph.tum.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1518827113/-/DCSupplemental.

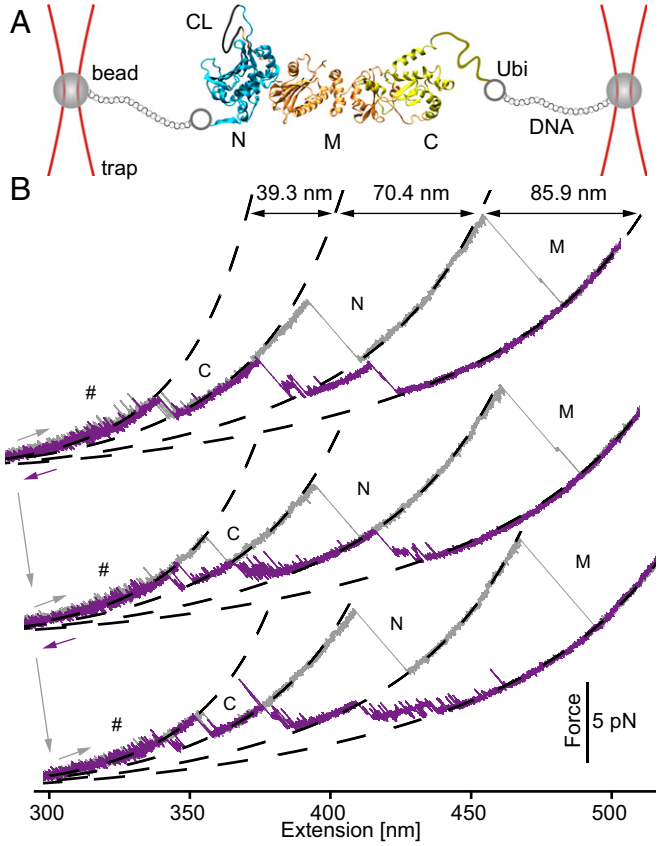

Fig. 1. Hsp90 monomers fold through a complex network of states. (*A*) Optical tweezers assay. An Hsp90 monomer carrying N- and C-terminal ubiquitins (gray circles) is tethered via DNA handles to silica beads (gray spheres), held in optical traps. By moving one of the beads away from the other along the long axis of the protein, we apply force to the Hsp90 monomer (PDB ID 2CG9). Hsp90 consists of an N (blue), an M (orange), and a C (green) domain. The N and M domains are connected by a charged linker (CL, black line). The unstructured part of the C domain is indicated by a green line. The figure is not to scale. (*B*) Three consecutive force-extension traces of an Hsp90 monomer that were acquired by moving the beads apart and together at a slow, constant speed of 10 nm/s. Unfolding traces (gray) show identical, successive unfolding of the three domains. Worm-like chain (WLC) fits to the unfolding events, shown as dashed lines, mark Hsp90's domains (C, N, and M). Average contour length gains (see *SI Structure Sizes* and Table S1) are displayed in the top trace. Refolding traces (purple) from the completely unfolded state show that refolding sets in at ~5 pN. Apart from the major refolding events, many rapid transitions are observed.

we investigated the refolding of the individual domains. We designed separate constructs of the N, M, and C domains (see *SI Methods*). Slow 10-nm/s unfolding traces (Fig. 2 *A*–*C*, gray) show the characteristic unfolding forces and lengths already observed in the monomer (Fig. 1*B*). The refolding traces (Fig. 2 *A*–*C*, colored) exhibit a number of intermediate states for the N (blue) and M domain (orange), whereas the C domain (green) shows two-state behavior.

Several observations suggest that the intermediate observed in the M domain (black arrow in Fig. 2*B*) is an on-pathway folding and unfolding intermediate. First, at the resolution of our experiment, all folding events to the native state have to pass through this intermediate. Second, this intermediate is also populated in unfolding traces with an identical stability and length (see arrows in Fig. S2 *C* and *D*). Third, a C-terminal 10-amino acid residue truncation mutant exhibited slower folding behavior (Fig. S3). This observation suggests that the intermediate state corresponds to the folding/unfolding of the so-called smaller alpha/beta/alpha subdomain of the M domain (14, 16), comprising residues 444–527. The contour length increase of about 28.4 ± 1.1(SD) nm further supports this interpretation (see *SI Structure Sizes* and Table S1).

The other transiently populated intermediates we observe in the N domain as well as in the M domain (red arrows in Fig. 2 *A* and *B*) do not show similarly well-defined length nor kinetics and likely constitute ensembles of intermediate states.

The slow pulling speed traces shown in Fig. 2 *A*–*C* always result in a natively folded domain after relaxation. Faster (500 nm/s) unfolding and refolding cycles give the protein only a short time to refold at low forces and therefore can trap the protein in intermediate states. A series of subsequent fast stretch–relax cycles for the three domains are shown in Fig. S2 *B*, *D*, and *F*. For the N and M domains, we find two possible outcomes: Either the domain has refolded to the native state (green or orange) or it populates one of the intermediate states (red). Scatter plots of mechanical stability vs. contour length gain (Fig. 2 *D* and *E*) show a large spread in both stability and contour length increase of the intermediates, further supporting the notion of an ensemble of states rather than well-defined intermediate structures. In contrast, the native states of all three domains, as well as the on-pathway intermediate of the M domain, show clear overlap characteristic of well-defined states. The C domain shows no intermediate states (Fig. 2*F*), as expected for a two-state folder.

A priori, ensembles of transient intermediate states can comprise on-pathway as well as off-pathway (misfolded) intermediates. For the short-lived states, it is difficult to distinguish between the two from force-extension cycles alone. However, studying the force dependence of the folding kinetics can reveal the nature of the intermediates. In the case of on-pathway intermediates, increasing force should reduce refolding rates, because the population of intermediates will be decreased. In the case of misfolded (off-pathway) states, increasing applied force can increase refolding rates, because a higher load will decrease the population of misfolded states. Inspired by chemical double-jump experiments that are widely used to characterize protein folding pathways (17, 18), we used a mechanical double-jump protocol (19) to study force-dependent refolding. In brief, starting from the unfolded state, we relaxed the polypeptide chain rapidly to a certain low force value, allowing the chain a certain time to refold, then quenched this refolding process by jumping to a force value between 5 pN and 12 pN, from which we started an unfolding force ramp that allowed us to determine the fraction of folded protein in the native state as a function of "waiting force" and "waiting time" (see Fig. S4 and *SI Methods*). In addition to the unfolding ramp, we can observe the behavior of the domains during the waiting time (Fig. S5).

Plots of the force-dependent refolding kinetics are shown in Fig. 2 *G*–*I*. At the lowest waiting force measured (orange markers), the N and M domains fold within seconds, and the C domain folds within tenths of milliseconds. Although the C domain (Fig. 2*I*) exhibits a normal force dependence, where force slows down refolding kinetics, we find a counterintuitive behavior for the N and M domains (Fig. 2 *G* and *H*). Here, folding kinetics first become faster with increasing force, and, at forces exceeding about 3.5 pN, folding slows down again. This effect is even more obvious if folding probabilities are plotted against force (Fig. S6 *D* and *E*), where the folding probabilities peak at around 3.5 pN. This scenario is described above and identifies the ensemble of transiently populated intermediates as misfolded states.

A minimal kinetic model depicted in Fig. 2 *J* and *K* (Eqs. **S8**–**S10**) can quantitatively explain the observed effects. In this model, the unfolded state ensemble is in rapid equilibrium with an ensemble of misfolded, i.e., folding-incompetent, states. The effect of force is twofold: First, it helps the protein to avoid the misfolded states, and, second, it slows down folding from the unfolded to the native state.

The fits of this model to the force- and time-dependent probabilities shown in Fig. 2 *G* and *H* depend only on one set

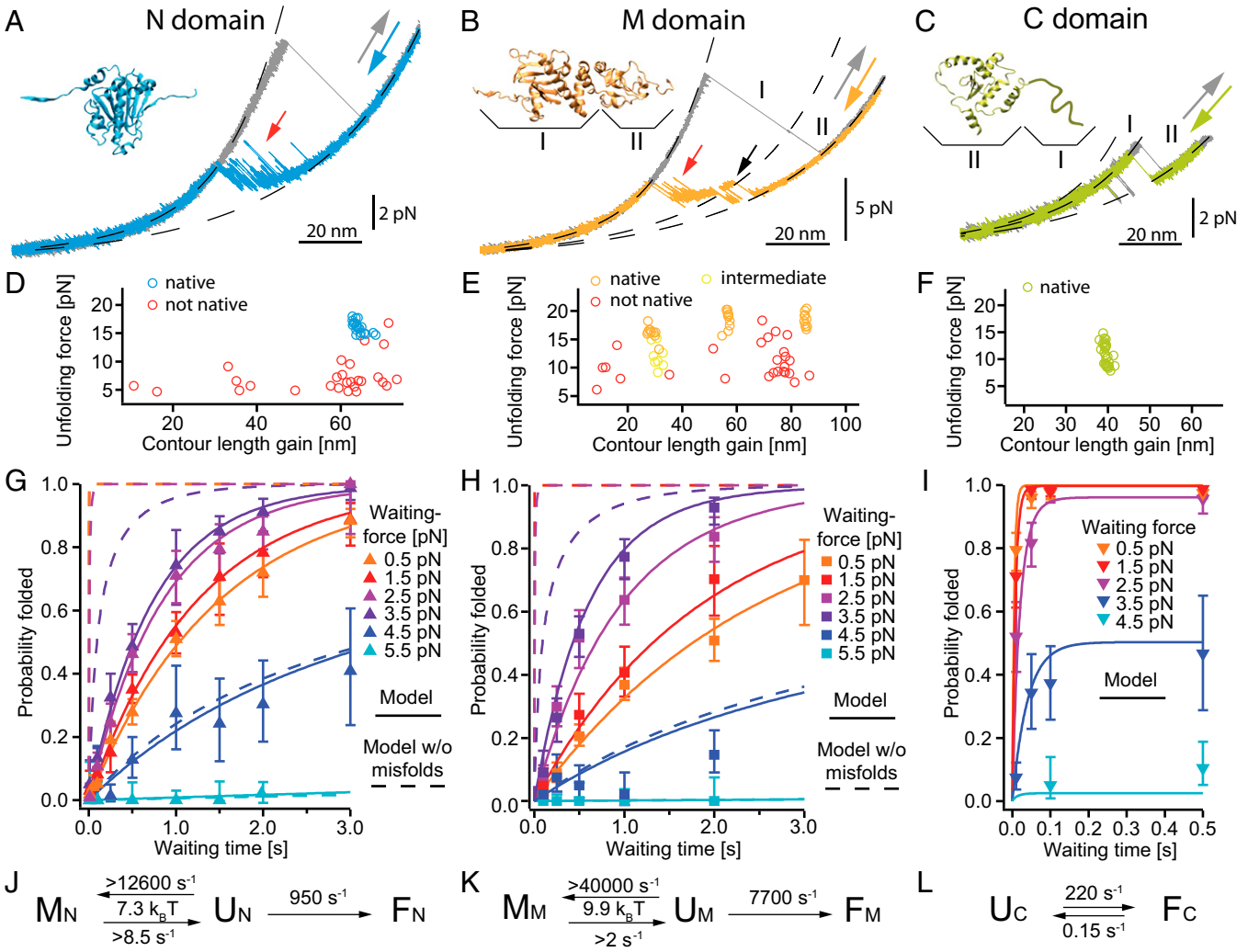

Fig. 2. Detailed analysis of individual domain constructs. (*A*–*C*) Force-extension traces of individual domain constructs (*Insets* from PDB *2cg9*) at a slow pulling speed of 10 nm/s. Unfolding traces (gray) are fitted with WLCs (black dashed). Refolding traces of the N (*A*, blue) and the M (*B*, orange) domains populate transient intermediate states (*A* and *B*, red arrows). The M domain (*B*) populates an on-pathway intermediate (black arrow) that corresponds to region II. The C domain (*C*) shows equilibrium two-state behavior; region I accounts for unstructured regions. More traces are shown in Fig. S2 *A*, *C*, and *E*. (*D*–*F*) Scatter plots display unfolding events (unfolding force vs. contour length gain) for the N (*D*), the M (*E*), and the C (*F*) domains. Each scatter plot is derived from 35 to 75 consecutive traces of one molecule. The native structured M domain can appear as a single event (orange, longest length gain) or as a double event (orange, shortest plus medium length gain) depending on the classification algorithm (see *SI Methods*). If only the on-pathway intermediate is observed (*B*, region II), it is colored in yellow. (*G*–*I*) Averaged refolding probabilities derived from double jump experiments (see Fig. S6 for nonaveraged data) depending on time (*y* axis) and force (color-coded) for the N (*G*), the M (*H*), and the C (*I*) domains. Probabilities are determined from 11,570 traces (45 molecules), 7,927 traces (14 molecules), and 6,157 traces (9 molecules) for the N, M, and C domains, respectively. The N and M domains show increasing refolding probability with increasing force (*G* and *H*, orange to purple markers). Fits are described in *J*–*L*. For uncertainties, see *SI Methods*. (*J*–*L*) The atypical refolding behavior of the N and M domains is fully described by a model (*J* and *K*) assuming an unfolded state ensemble U that is in equilibrium with fast-forming off-pathway intermediates M, that prevent folding to the native state F. Fitting this model to the nonaveraged data (Fig. S6 *A* and *B*) yields fits shown in *G* and *H* as continuous lines. Assuming identical folding parameters and neglecting misfolding yields the refolding probabilities shown in *G* and *H* as dashed lines. The C domain is fitted with a two-state model assuming an unfolded U and a folded state F in equilibrium (*L*, continuous lines in *I* and Fig. S6*C*). The folding rates at zero force, the equilibrium energy between U and M, and lower estimates for the rates from and to the misfolded states (see *Estimate 3*) are displayed.

of four parameters, namely, the folding rate at zero force with the corresponding force dependence and the equilibrium constant of misfolded states with the corresponding force dependence (see *SI Methods* and Fig. S6). Even though the ensemble of misfolded states is quite heterogeneous, average parameters describe our results well. We find that the ensembles of misfolded states have an average free energy of 7.3 (± 0.1) k_BT for the N domain and 9.9 (± 0.1) k_BT for the M domain, consistent with an estimate directly obtained from force-extension traces (*Estimate 2*). We also find fast folding rates from the unfolded ensemble to the native state at zero load of 954 (± 65) s^{-1} for the N and 7651 (± 714) s^{-1} for the M domain. Therefore, the presence of misfolded states reduces the folding rates from thousands per second to about one per second (Fig. 2 *D* and *E*). The expected force-dependent refolding probabilities without misfolded states are shown by the dashed lines in Fig. 2 *G* and *H*.

Since the C domain exhibits equilibrium two-state behavior, a two-state model (Fig. 2*L* and Eqs. **S11** and **S12**) describes its folding probabilities well (fits in Fig. 2*I* and Fig. S6 *C* and *F*). Zero-force folding and unfolding rate constants are 218 (± 16) s^{-1} and 0.154 (± 0.029) s^{-1}, respectively.

Comparing Folding Kinetics of the Hsp90 Monomer and Its Constituting Domains. In the following, we investigated whether the overall

folding kinetics of the complete Hsp90 monomer can be described from the folding kinetics of the individual domains or whether additional complications further affect folding of Hsp90. Using the jump protocol described above, we classified misfolded conformations and folded domains of the whole monomer in a scatter plot of unfolding force vs. contour length gain (Fig. 3*A*). At low waiting forces, we find even more misfolded states (red circles) than for the individual domains with both higher stabilities (unfolding force) and larger contour length gains. We hypothesize that the low forces now allow regions of the protein distant in sequence, i.e., across domains, to interact and misfold into stable intermediates. At higher forces, occurrence of intermediates is reduced and, specifically, those intermediates with long contour length gains from distant regions of the protein are suppressed. It can also be seen that higher loads significantly increase the occurrence of natively folded domains (blue, orange, and green circles). Sample traces are presented in Fig. S7 *A* and *B* for low and high waiting forces, respectively.

A plot that shows the force- and time-dependent probabilities of complete refolding of the monomer at both low and high forces reveals the drastic effect of cross-domain misfolded states (Fig. 3*B*). At low waiting forces (0.3–0.8 pN, red), folding is suppressed dramatically compared with high waiting forces (1.8–2.2 pN, blue). Single exponential fits to the data yield 0.024 ± 0.013 s^{-1} for low waiting forces (Fig. 3*B*, red line) and 0.54 ± 0.07 s^{-1} for high waiting forces (Fig. 3*B*, blue line). A lower-limit comparison with the expected folding kinetics calculated under the assumption of independent folding of the individual domains (dashed lines in Fig. 3*B*; for details, see *Estimate 4*) shows that, under low loads, cross-domain intermediates lead to much slower folding than expected from the domains individually.

Folding and Assembly of the Dimer. The last step of assembly into a functional Hsp90 machine is the dimerization of two monomer chains. To study folding and assembly of Hsp90, we designed a construct where we link two Hsp90 domains through a C-terminal leucine zipper into a single-chain construct. Cysteine residues within the leucine zipper ensure permanent crosslinking of the two chains. We applied force at the N domains of Hsp90 through cysteines at position 61 (for details, see *SI Methods*). Fig. 4 shows a force-extension trace of this construct, where the unfolding of all six Hsp90 domains of the dimer can be seen, proving successful construct design.

An additional unfolding event that is not observed for the monomer reflects the dissociation of the dimer. After dissociation at around 10 pN (red arrow on gray trace), the unstructured portions of the C domains are stretched. The C domains are stabilized by the dimerization, and, after dissociation, the C domains unfold rapidly. Upon relaxation of the unfolded dimer chain, we see folding of all six domains and, as a final event, dimerization of the fully folded chain at about 3 pN (red arrow on purple trace). More dimer traces are shown in Fig. S8.

The contour length change upon dissociation of the dimer is about 40.6 ± 1.1 nm, which matches the elongation of the unstructured parts at the C terminus plus about 12 nm of additional residues, likely from the four-helix bundle (see *SI Structure Sizes* and Table S2). We can exclude significant contributions from the N domains to the dimer stability because, before domain dissociation, we observe the opening fluctuations of the two charged linkers (see "##" in Fig. 4). Therefore, the N and M domains are already spatially separated when the dimer dissociates.

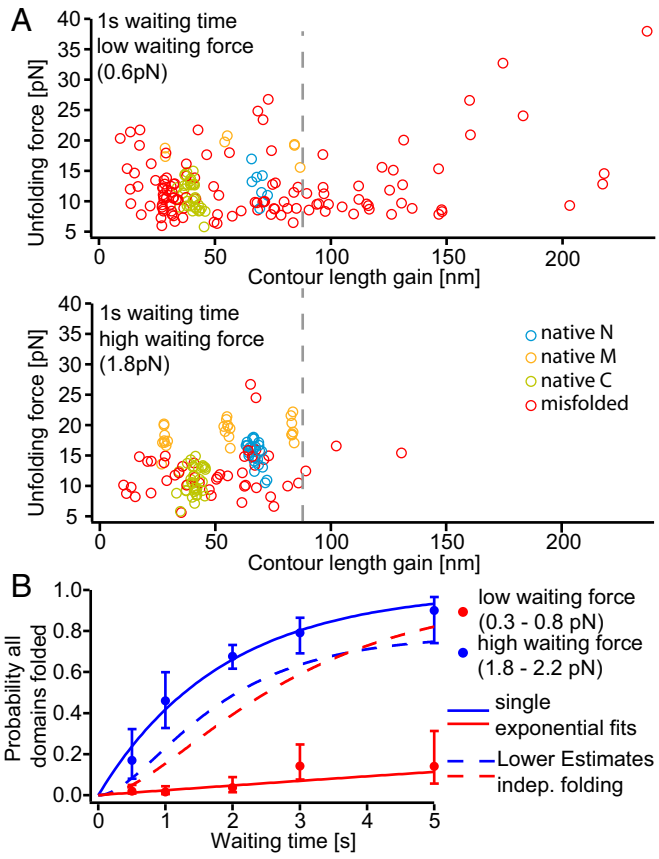

Fig. 3. Cross-domain misfolds strongly decrease refolding rates of the monomer at low forces. (*A*) Example scatter plots of double jump refolding experiments with a low (*Top*) and high (*Bottom*) waiting force, using the same single Hsp90 monomer (60 and 45 unfolding traces, respectively). Blue, orange, and green circles correspond to unfolding events of the N, M, and C domains (see Fig. 2 *D* and *E*). We observe a broad spectrum of nonnative events (red circles). At low forces (*Top*), many of these misfolds show a longer contour length gain than one would expect from the largest native domain (M domain, ~68 nm); hence misfolds to the right of the gray dashed lines must involve two or even all three domains. At slightly higher waiting forces (*Bottom*), misfolds, especially those with longer contour length gain, are greatly reduced. For clarity, the event when the M domain only populates the on-pathway intermediate (~28 nm, especially at low force) is not assigned. More scatter plots for different waiting times are shown in Fig. S7*D*. (*B*) Probability of observing a completely refolded Hsp90 monomer after a double jump experiment, against waiting time. Red and blue dots show averaged probabilities for a low waiting force range (0.3–0.8 pN) and for a high waiting force range (1.8–2.2 pN), respectively. Probabilities were quantified with simple single exponentials (continuous lines), showing that a slightly higher waiting force can greatly improve refolding. We estimated a lower limit from the isolated domain refolding experiments assuming independent folding (see *Estimate 4*). For higher waiting forces, the lower estimate (blue, dashed line) lies below the measured refolding probabilities, indicating little or no influence of cross-domain misfolds. However, for low waiting forces, the estimate is well above the measured refolding probabilities, showing severe disturbance of refolding by cross-domain misfolds. For this graph, 1,067 monomer traces of 10 molecules were analyzed; example traces, refolding probabilities of the domain itself, and nonaveraged probabilities are shown in Fig. S7 *A–C, E, and F*. For uncertainties, see *SI Methods*.

Discussion

Folding of the Individual Domains. The organization of proteins into domains is a common feature that has been recognized as key in facilitating folding and self-assembly (20). We find that all three domains of Hsp90, when held in isolation, fold rapidly and independently. The smallest and weakest C domain (117 residues folded) folds in a two-state-like manner with a folding rate constant $k_{U\to F}$ of 220 s^{-1} without noticeably populating misfolded states. However, the folding kinetics of the larger N and M domains, even though they fold within 1 s, are dominated by misfolded intermediates. As detailed, we find considerable free energies for the misfolded states of 7.3 ± 0.1 k_BT and 9.9 ± 0.1 k_BT as well as lower limits for $k_{U\to M}$ of 12,600 s^{-1} and 40,000 s^{-1} for the

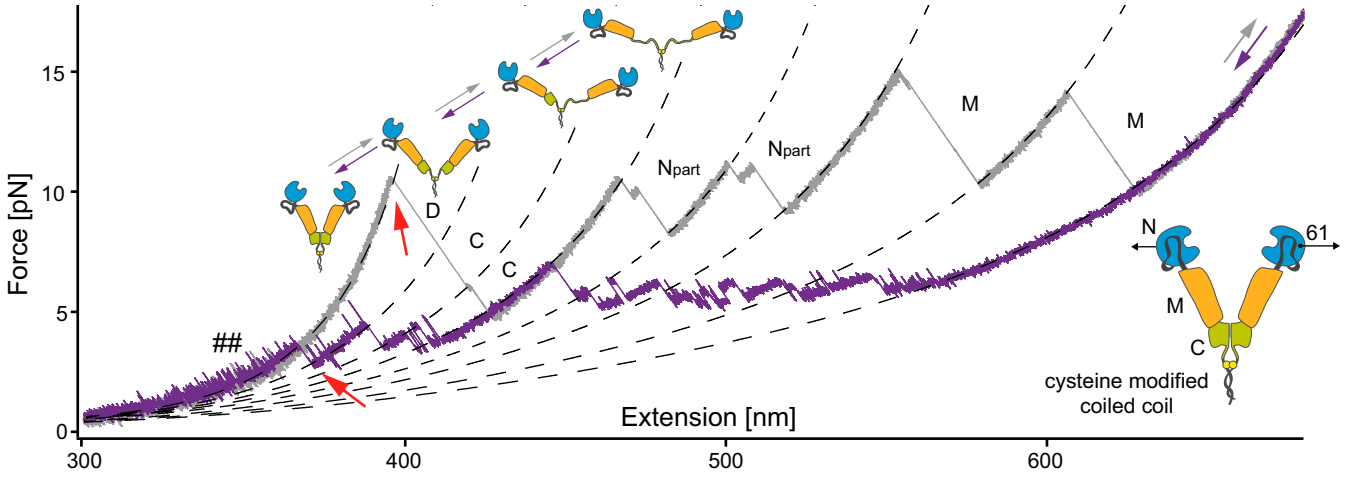

Fig. 4. Dimerization of the Hsp90. Force-extension traces of the Hsp90 dimer, unfolded (gray) and refolded (purple) at a velocity of 10 nm/s. Mechanically stable dimers are engineered using a C-terminal coiled coil motif carrying cysteines (see lower right-hand *Inset*). Force (black arrows) is applied at amino acid positions 61; hence only a part of the N domain is measured, and an additional intermediate is observed (see also legend of Fig. S8). In addition to the duplication of the unfolding events, a high force peak is observed, at the start of the trace in this case (D). This peak (red arrow on gray trace) is due to the dimerization of the C domains. After dissociation, the unstructured regions of both C domains are stretched, and both C domains unfold rapidly. After successful refolding of all domains, the dimerization event is seen (red arrow on purple trace). The unfolding and refolding pathway is shown in a sequence of *Insets*.

N and M domains, respectively, which suggests an extremely fast population of misfolded states. This is supported by the observation that neither the N nor the M domain ever populates the completely unfolded state after fast relaxation of the chain (Fig. S2 *B* and *D*). The productive folding rate constant $k_{U\to F}$ is very fast for both the N (954 ± 65 s^{-1}) and the M (7,651 ± 714 s^{-1}) domains. Given the large size of the M domain, the folding rate of 7,700 s^{-1} may appear very fast. However, the M domain exhibits a simple architecture with a low relative contact order of 0.05 rationalizing fast folding of this domain (2). Note that a similar competition between distinct misfolding intermediates and folded states that can be modulated by force has been reported in single-molecule experiments with the calcium binding proteins NCS-1 (4) and Calmodulin (9). Likely owing to the large sizes of our N and M domains (211 and 266 residues, respectively), the population of misfolded states is more heterogeneous than in those simpler proteins. The large heterogeneity is reflected both in the large spread of lengths and unfolding forces of the scatter plots in Fig. 2 *D* and *E* and in the individual refolding time traces of Fig. S5. Structurally similarly heterogeneous burst-phase intermediates have been inferred from bulk studies of other large proteins like maltose-binding protein (MBP) and Tim barrel protein (6, 21).

Cross-Domain Misfolds. The most striking result of our study is the strong decrease of the overall folding speed when the three Hsp90 domains are linked in one chain forming the 709-amino acid monomer. Clearly, the folding energy landscape is significantly altered. This has previously been observed for protein chains consisting of identical or homologous subunits such as the long immunoglobulin domain chains of the muscle protein titin (8, 22), fibronectin (23), or ubiquitin (24).

We find that misfolded states occur across domains in a long protein even if the domains share no homology. In a protein chain of increasing length and stability, misfolding will be inevitable, even if the subunits have no common fold. In a short single-domain protein, it may be possible to optimize the sequence for a smooth energy landscape without misfolded states of significant free energy as we find for the C domain of Hsp90. However, as the chain length grows, so the number interactions stabilizing misfolded states will increase (25), eventually trapping the protein on its way to the native state. Hence, even if all subdomains can fold rapidly, the mere fact that they are linked in one chain will slow their folding. This is precisely what we find in our Hsp90 monomer. The misfolded states that we find in the large N and M domains already reduce the effective folding time to seconds. However, when integrated into the full chain of the monomer, new cross-domain misfolds slow folding even more dramatically (Fig. 3). Although, for the individual domains, the misfolded states are very dynamic, we often observe states that show high mechanical stability for the full monomer, some of them exceeding our accessible force range of 40 pN (Fig. S7*C*), underscoring the detrimental effects of those cross-domain intermediates on folding. Such cross-domain misfolds are likely a general feature of large proteins and therefore limit refolding rates.

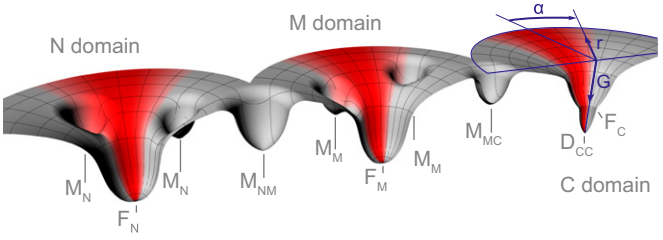

Fig. 5. Simplified energy landscapes for folding and assembly of Hsp90. The folding properties of Hsp90 can be described by three individual energy landscapes, because the individual domains don't directly stabilize each other. Experiments using the individual domains showed that the N and the M domain have a fast productive folding pathway to the native states (F_N, F_M) but that the overall folding rate is greatly slowed down by off-pathway intermediates (M_N, M_M). The C domain exhibits two-state behavior without kinetic traps. Refolding of the full-length monomer showed heterogeneous and stable cross-domain misfolds (M_{NM}, M_{MC}). A cylindrical coordinate system shown for the C domain (blue) applies for all domains. G refers to free energy. The inverse of the radial coordinate r describes the overall number of residues with native conformation. The inverse of the absolute value of the angular coordinate α can be interpreted as the average distance between residues in misfolded conformations. This distance is strongly dependent on force and restricts the conformational search (red shaded areas), speeding up folding by avoiding or depopulating misfolded species. This effect is particularly strong for the cross-domain misfolds in the full-length monomer. After successful formation of the C domain, Hsp90 can dimerize into a functional chaperone (D_{CC}).

Hsp90 Dimer Assembly. After full folding of the monomer has been completed, the two chains have to find each other and form a stable dimer. In our dimer refolding traces, we can observe the formation of the dimer directly in the refolding traces after all of the domains have folded (Fig. 4). In the nucleotide-free state, dimerization of yeast Hsp90 is mainly achieved by association of the C domains (26), as we can rule out significant contributions from N-terminal dimerization.

Chaperoning by Force. An immediate conclusion that can be drawn from our experiments is that small mechanical forces can stretch out the unfolded chain and thus prevent misfolding interactions between distant parts of the protein chain. This leads to the, at first sight paradoxical, effect that full-length Hsp90 as well as its large N and M domains fold faster if small forces are applied (Figs. 2 *G* and *H* and 3*B*). Force speeds up folding in the isolated N and M domains by a factor of 2 or 3, whereas, for the full-length chain, we find a factor of 25 when increasing the force from ∼0.5 pN to ∼2.5 pN. This observation can be visualized in a simplified energy landscape diagram (27). The red shaded areas in Fig. 5 show a restriction of the ensemble of accessible folding pathways by force. This prevents the formation of misfolds involving distant parts of the chain, thus chaperoning the chain toward the folded states of its subunits.

Cotranslational folding has been recognized as an important feature for productive protein folding for a long time (28). The sequential way protein chains are synthesized at the ribosome allows the cell to avoid cross-domain misfolding. We show that a similar goal can also be achieved by force application to the ends of a large protein. In a scenario where Hsp90 folds cotranslationally in vivo, misfolding will therefore be largely avoided. After initial folding of Hsp90 when translation has been accomplished, individual domains or subdomains may transiently unfold and refold under heat shock conditions, but a situation where the whole chain is unfolded is very unlikely to occur again.

Reversing misfolding by actively applying force is also discussed for chaperones, like GroEL/GroES (29). Another mechanism for chaperones to avoid cross-domain misfolding is the stabilization of partially folded, aggregation-prone intermediates. This has been shown for trigger factor on the single-molecule level chaperoning an individual MBP as well as a four-times repeat construct thereof (30).

Methods

Proteins are attached to beads in a multistage reaction. First, small (34 base pairs), maleinimide-modified DNA oligonucleotides are coupled to the free cysteines of the proteins (13). These DNA oligonucleotides are then hybridized with long (545 base pair) DNA handles by a single-stranded overhang on one end that is complementary to the DNA oligonucleotides (9). At the other end, DNA handles are functionalized with biotin or digoxigenin, which in turn can bind to streptavidin-coated or anti-digoxigenin-coated 1-μm silica beads (9, 13). For trapping of beads, we use a custom-built, dual-trap optical tweezers setup with back-focal plane detection and high resolution as described previously (31). The trap stiffnesses of the individual traps are adjusted to around 0.3 pN/nm, and acquisition frequency is 20 kHz or 30 kHz. The temperature at the position of the protein is ∼30 °C. All experiments are performed in a buffer containing 40 mM Hepes, 150 mM KCl, 10 mM $MgCl_2$, pH 7.4. To avoid photo damage, a scavenger system comprising glucose, glucose oxidase, and glucose catalase or trolox is used (13). A detailed description of methods used is given in *SI Methods*.

ACKNOWLEDGMENTS. We thank Marco Grison and Katarzyna Tych for comments on the manuscript, Alena Dudarenka and Matthias Jahn for help with graphics, and the German Research Foundation for financial support (SFB863 A4).

1. Chung HS, Piana-Agostinetti S, Shaw DE, Eaton WA (2015) Structural origin of slow diffusion in protein folding. *Science* 349(6255):1504–1510.
2. Ivankov DN, et al. (2003) Contact order revisited: Influence of protein size on the folding rate. *Protein Sci* 12(9):2057–2062.
3. 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.
4. Heidarsson PO, et al. (2014) Direct single-molecule observation of calcium-dependent misfolding in human neuronal calcium sensor-1. *Proc Natl Acad Sci USA* 111(36): 13069–13074.
5. 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.
6. Walters BT, Mayne L, Hinshaw JR, Sosnick TR, Englander SW (2013) Folding of a large protein at high structural resolution. *Proc Natl Acad Sci USA* 110(47):18898–18903.
7. Brockwell DJ, Radford SE (2007) Intermediates: Ubiquitous species on folding energy landscapes? *Curr Opin Struct Biol* 17(1):30–37.
8. Borgia MB, et al. (2011) Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature* 474(7353):662–665.
9. Stigler J, Ziegler F, Gieseke A, Gebhardt JC, Rief M (2011) The complex folding network of single calmodulin molecules. *Science* 334(6055):512–516.
10. Greenleaf WJ, Frieda KL, Foster DA, Woodside MT, Block SM (2008) Direct observation of hierarchical folding in single riboswitch aptamers. *Science* 319(5863):630–633.
11. Nunes JM, Mayer-Hartl M, Hartl FU, Müller DJ (2015) Action of the Hsp70 chaperone system observed with single proteins. *Nat Commun* 6:6307.
12. Puchner EM, Gaub HE (2010) Exploring the conformation-regulated function of titin kinase by mechanical pump and probe experiments with single molecules. *Angew Chem Int Ed Engl* 49(6):1147–1150.
13. Jahn M, et al. (2014) The charged linker of the molecular chaperone Hsp90 modulates domain contacts and biological function. *Proc Natl Acad Sci USA* 111(50):17881–17886.
14. Ali MM, et al. (2006) Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* 440(7087):1013–1017.
15. Jakob U, et al. (1995) Structural organization of procaryotic and eucaryotic Hsp90. Influence of divalent cations on structure and function. *J Biol Chem* 270(24):14412–14419.
16. Meyer P, et al. (2003) Structural and functional analysis of the middle segment of hsp90: Implications for ATP hydrolysis and client protein and cochaperone interactions. *Mol Cell* 11(3):647–658.
17. Brandts JF, Halvorson HR, Brennan M (1975) Consideration of the possibility that the slow step in protein denaturation reactions is due to cis–trans isomerism of proline residues. *Biochemistry* 14(22):4953–4963.
18. Wildegger G, Kiefhaber T (1997) Three-state model for lysozyme folding: Triangular folding mechanism with an energetically trapped intermediate. *J Mol Biol* 270(2): 294–304.
19. Schwaiger I, Schleicher M, Noegel AA, Rief M (2005) The folding pathway of a fast-folding immunoglobulin domain revealed by single-molecule mechanical experiments. *EMBO Rep* 6(1):46–51.
20. White SH, Jacobs RE (1990) Statistical distribution of hydrophobic residues along the length of protein chains. Implications for protein folding and evolution. *Biophys J* 57(4):911–921.
21. Wu Y, Kondrashkina E, Kayatekin C, Matthews CR, Bilsel O (2008) Microsecond acquisition of heterogeneous structure in the folding of a TIM barrel protein. *Proc Natl Acad Sci USA* 105(36):13367–13372.
22. Oberhauser AF, Marszalek PE, Carrion-Vazquez M, Fernandez JM (1999) Single protein misfolding events captured by atomic force microscopy. *Nat Struct Biol* 6(11):1025–1028.
23. Peng Q, Fang J, Wang M, Li H (2011) Kinetic partitioning mechanism governs the folding of the third FnIII domain of tenascin-C: Evidence at the single-molecule level. *J Mol Biol* 412(4):698–709.
24. Xia F, Thirumalai D, Gräter F (2011) Minimum energy compact structures in force-quench polyubiquitin folding are domain swapped. *Proc Natl Acad Sci USA* 108(17): 6963–6968.
25. Bastolla U, Demetrius L (2005) Stability constraints and protein evolution: The role of chain length, composition and disulfide bonds. *Protein Eng Des Sel* 18(9):405–415.
26. Richter K, Muschler P, Hainzl O, Buchner J (2001) Coordinated ATP hydrolysis by the Hsp90 dimer. *J Biol Chem* 276(36):33689–33696.
27. Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Theory of protein folding: The energy landscape perspective. *Annu Rev Phys Chem* 48:545–600.
28. Fedorov AN, Baldwin TO (1997) Cotranslational protein folding. *J Biol Chem* 272(52): 32715–32718.
29. Shtilerman M, Lorimer GH, Englander SW (1999) Chaperonin function: Folding by forced unfolding. *Science* 284(5415):822–825.
30. Mashaghi A, et al. (2013) Reshaping of the conformational search of a protein by the chaperone trigger factor. *Nature* 500(7460):98–101.
31. von Hansen Y, Mehlich A, Pelz B, Rief M, Netz RR (2012) Auto- and cross-power spectral analysis of dual trap optical tweezer experiments using Bayesian inference. *Rev Sci Instrum* 83(9):095116.
32. Mickler M, Hessling M, Ratzke C, Buchner J, Hugel T (2009) The large conformational changes of Hsp90 are only weakly coupled to ATP hydrolysis. *Nat Struct Mol Biol* 16(3): 281–286.
33. Bornschlögl T, Woehlke G, Rief M (2009) Single molecule mechanics of the kinesin neck. *Proc Natl Acad Sci USA* 106(17):6992–6997.
34. Schlierf M, Berkemeier F, Rief M (2007) Direct observation of active protein folding using lock-in force spectroscopy. *Biophys J* 93(11):3989–3998.