

# Sequence-resolved free energy profiles of stress-bearing vimentin intermediate filaments

Beatrice Ramm<sup>a,1</sup>, Johannes Stigler<sup>a,1,2</sup>, Michael Hinczewski<sup>b</sup>, D. Thirumalai<sup>b</sup>, Harald Herrmann<sup>c</sup>, Günther Woehlke<sup>a</sup>, and Matthias Rief<sup>a,d,3</sup>

<sup>a</sup>Physik Department E22, Technische Universität München, 85748 Garching, Germany; <sup>b</sup>Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742; <sup>c</sup>Functional Architecture of the Cell, German Cancer Research Center, 69120 Heidelberg, Germany; and <sup>d</sup>Nanosystems Initiative Munich, 80799 Munich, Germany

Edited by James A. Spudich, Stanford University School of Medicine, Stanford, CA, and approved June 19, 2014 (received for review February 19, 2014)

**Intermediate filaments (IFs) are key to the mechanical strength of metazoan cells. Their basic building blocks are dimeric coiled coils mediating hierarchical assembly of the full-length filaments. Here we use single-molecule force spectroscopy by optical tweezers to assess the folding and stability of coil 2B of the model IF protein vimentin. The coiled coil was unzipped from its N and C termini. When pulling from the C terminus, we observed that the coiled coil was resistant to force owing to the high stability of the C-terminal region. Pulling from the N terminus revealed that the N-terminal half is considerably less stable. The mechanical pulling assay is a unique tool to study and control seed formation and structure propagation of the coiled coil. We then used rigorous theory-based deconvolution for a model-free extraction of the energy landscape and local stability profiles. The data obtained from the two distinct pulling directions complement each other and reveal a tripartite stability of the coiled coil: a labile N-terminal half, followed by a medium stability section and a highly stable region at the far C-terminal end. The different stability regions provide important insight into the mechanics of IF assembly.**

protein folding | Brownian dynamics simulation | trigger sequence

Intermediate filaments (IFs) serve as “stress buffers” (1) in metazoan cells. They are critical for mechanical resistance of the cell, cell motility, and shape determination (2, 3). IFs are a heterogeneous group of proteins encoded by 70 different genes in humans (4). Three of these genes code for nuclear proteins, termed lamins, the others for various cytoplasmic proteins, such as keratins, found in epithelia, and vimentin, characteristic of mesenchymal cells. All IFs share the same overall structure: a long central  $\alpha$ -helical rod of conserved design that is flanked by non- $\alpha$ -helical head and tail domains of highly varying size. Two individual rod domains wrap around each other to form the basic building block of IF assembly: a dimeric coiled coil (CC) (2).

The structure of the homodimeric rod domain of vimentin may be subdivided into three  $\alpha$ -helical regions termed coils 1A, 1B, and 2 [formerly coil 2A, linker 2, and coil 2B (5)], which are interspaced with non- $\alpha$ -helical linker regions L1 and L12 (Fig. 1A) (6, 7).

The superhelical structure of a left-handed CC reflects the characteristic heptad repeat pattern with positions denoted as *abc-defg*, in which hydrophobic residues occupy positions *a* and *d* (8, 9).

CC motifs are not limited to IFs but occur in an estimated 5–10% of all translated protein sequences (8). Their structural simplicity and diverse functionality make them an important model system for protein folding. Debate still exists regarding the precise events of folding of CCs, ranging from collision of previously unstructured chains to the preformation of short  $\alpha$ -helical segments (10–12). CCs also have become a target of single-molecule force spectroscopy, allowing the study of sequence-resolved folding. Examples are the mapping of the energy landscape of the CC model system GCN4 (13, 14) and the recent characterization of the assembly of the SNARE proteins (15).

Many CCs assemble into higher-order structures, as in the case of IFs. Here, the CC dimers associate laterally into roughly half-staggered antiparallel tetramers that, depending on ionic strength,

further stack upon each other to build so-called unit-length filaments (ULFs) (2, 16, 17). Longitudinal annealing of these ULFs is believed to be mediated by the interaction of both dimer ends from each individual ULF, namely coil 1A and coil 2B, resulting in an “overlap” of successive CC dimers (2, 18). In the last step, the filaments undergo a radial compaction yielding full-length IFs (2) that exhibit unique mechanical properties (19–22). The structural nature of the overlap of coil 1A and coil 2B is still debated (2, 23–25). The CC dimer ends represent hotspots for disease-related mutations in IF proteins (4, 26), further underscoring their functional importance.

Interestingly, coil 2 exhibits high sequence and absolute length conservation across all IFs (1, 27). Moreover, the large C-terminal fragment of coil 2 of vimentin (coil 2B) harbors two structurally interesting features: a stutter (residues 351–354) and the highly conserved region of 25 residues at the C terminus (residues 380–404) (Fig. 1A, *Inset*) (23). A stutter is an insertion of an additional four amino acids at the end of a heptad, thus interrupting the regular *abcdefg* pattern (9). The stutter position in coil 2 of IFs is absolutely conserved (23).

In this study, we present a deconvolution force spectroscopic approach using a combination of experiment and Brownian dynamics simulation, allowing us to measure the full distance-resolved energy landscape of the C-terminal part of vimentin coil 2. Our results yield important insight into the mechanical and thermodynamic stability of IFs, with important consequences for their assembly.

## Significance

**Intermediate filaments (IFs) are an essential part of the cytoskeleton of metazoan cells, responsible for the cell's shape and motility. To exert their mechanical function, IF dimers undergo a complex assembly into long filaments. The dimer ends are thought to mediate assembly. Mutations in these regions are associated with various diseases. We use single-molecule force spectroscopy by optical tweezers to mechanically open the C-terminal part of the dimer of the IF vimentin. Using deconvolution, a specialized analytic method, we can directly assess the stability of the C-terminal dimer end. Relating this stability to the amino acid sequence and structural features of the dimer, we shed light on the early assembly and mechanical properties of IFs.**

Author contributions: B.R., J.S., H.H., G.W., and M.R. designed research; B.R. and J.S. performed research; M.H., D.T., and H.H. contributed new reagents/analytic tools; B.R. and J.S. analyzed data; and B.R., J.S., D.T., H.H., G.W., and M.R. wrote the paper.

The authors declare no conflict of interest.

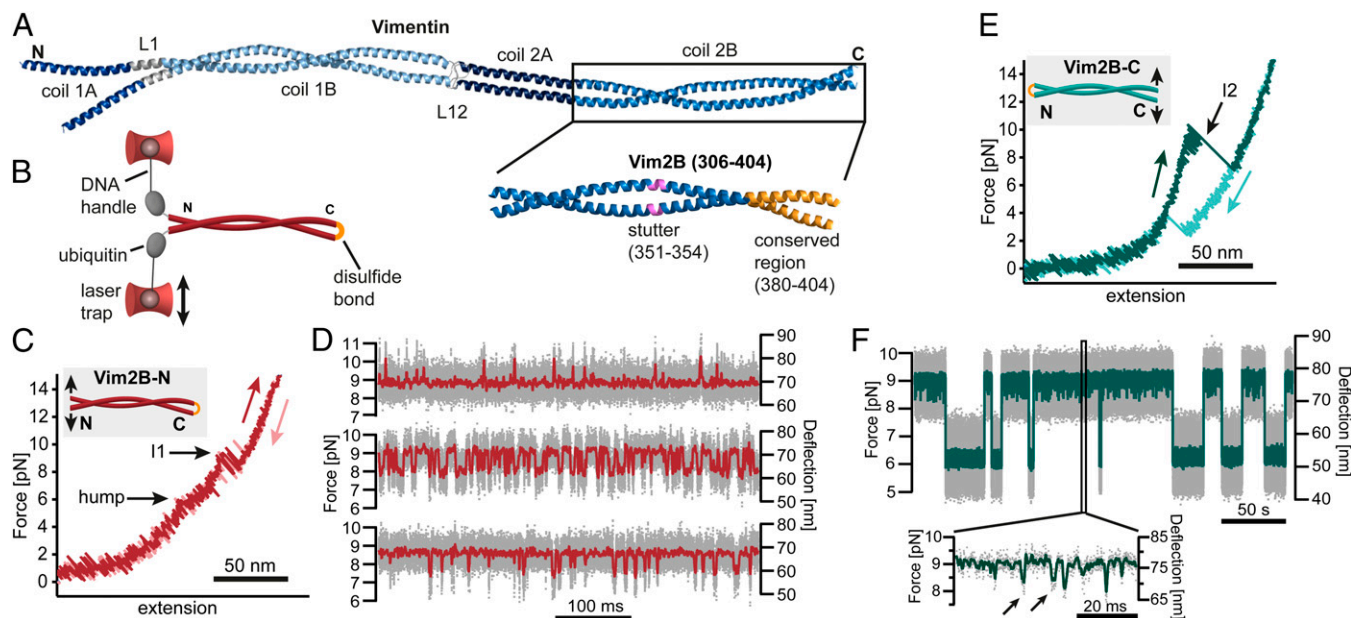
This article is a PNAS Direct Submission.

<sup>1</sup>B.R. and J.S. contributed equally to this work.

<sup>2</sup>Present address: Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032.

<sup>3</sup>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.1403122111/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1403122111/-DCSupplemental).



**Fig. 1.** The C-terminal half of coil 2 (Vim2B) shows different folding and unfolding behaviors when unzipped from two different directions. (A) Three-dimensional structural model of a vimentin dimer, save the head and tail domains, provided by A. Chernyatina and S. Strelkov, Laboratory for Biocrystallography, Department of Pharmaceutical and Pharmacological Sciences, Katholieke Universiteit Leuven, Leuven, Belgium (7). (Inset) Structural characteristics of the vimentin coil 2B, in which the stutter is highlighted in purple and the conserved region is marked in orange. (B) Schematic of the experimental setup using the Vim2B-N construct as an example. (C) Representative stretch-and-relax cycle recorded at 500 nm/s for Vim2B-N. Black arrows indicate the two major transitions. (D) Traces recorded in passive mode show equilibrium fluctuations between an intermediate and the unfolded state of Vim2B-N at varying force bias. The CC fluctuates between states mostly residing in the lower unfolded state at higher force bias (Top) and more in the upper partly folded state for decreasing force bias (Middle and Bottom). Full-bandwidth traces (100 kHz) are gray and low-pass filtered traces are colored in red. (E) A representative force-extension trace for Vim2B-C shows a clear three-state behavior (intermediate marked by an arrow). (F) Fluctuations between all three states of the Vim2B-C coil may be monitored in passive mode at a force bias of about 9.1 pN. (Inset) Transition between the folded and intermediate states (intermediate states are marked by an arrow). Full-bandwidth traces (100 kHz) are gray, whereas low-pass filtered traces are colored in green.

## Results

For single-molecule force measurements with the C-terminal half of vimentin coil 2 (Vim2B), we constructed two different fusion proteins containing the vimentin residues 306–404. A terminal cysteine for covalent intradimer linkage and a ubiquitin domain as a spacer for DNA handle attachment were added. By varying the force attachment points, the unfolding pathway of the constructs can be controlled (28–31). One construct (Vim2B-C) was linked by a disulfide bond at the N terminus where mechanical opening proceeds from the C-terminal end. The other construct (Vim2B-N) is opened from the N terminus while linked at the C terminus (Fig. 1B). The unzipping geometry of the linear CCs allows us to assign a precise unfolding coordinate along the backbone of the  $\alpha$ -helices. Using optical tweezers, we performed nonequilibrium experiments recording force-extension traces by stretching and relaxing the constructs at a constant trap velocity of 500 nm/s. Another set of experiments was conducted in “passive mode,” in which the two traps are held at a constant separation, allowing equilibrium fluctuations of the molecule. As we extracted most free energy values and data discussed from passive mode data, hereafter we use the term “stability” synonymously with equilibrium free energy.

### Vimentin Coil 2B Pulled from the N Terminus Unfolds Gradually.

Stretch and relax curves of Vim2B-N are shown in Fig. 1C. Two major transitions close to equilibrium can be observed. Although transitions very close to equilibrium are difficult to observe in force-extension cycles, they appear clearly in an analysis of the force-dependent SD of the signal, in which transition regions appear as areas with increased noise (Fig. S1). Between 5 and 8 pN, we observe a hump-like transition in which the CC gradually unfolds to a metastable intermediate (I1). In the second transition, at forces between 8.5 and 10 pN, the coil

rapidly fluctuates between I1 and the unfolded state. The average contour length increase of  $35.4 (\pm 2.0)$  nm for the first transition indicates that the N-terminal half of coil 2B is already unfolded at around 8 pN. The contour length increase of  $65.4 (\pm 2.3)$  nm between the completely folded and unfolded states corresponds to the unfolding of  $90 (\pm 3)$  residues, indicating that the first 8 N-terminal residues are likely to be unfolded at our minimally resolvable forces of about 2–3 pN. The close-to-equilibrium nature of the first hump-like transition allows us to analyze the cooperativity of the transition and infer the number of intermediates populated (SI Appendix, section 4.2 and Fig. S2). We find that the hump-like transition involves at least one additional intermediate that exchanges quickly with the folded state and the intermediate I1. Because rapid transitions reflect low transition barriers, we infer that the energy landscape of Vim2B-N is flat at its N terminus. Fig. 1D shows passive mode recordings of force-dependent equilibrium fluctuations between I1 and the unfolded state. Increasing force gradually shifts the probability distribution toward the unfolded state (bottom to top traces). Force-dependent state occupancies allow extraction of equilibrium free energy differences (SI Appendix, section 4.4). The free energy difference between the unfolded state and I1 is  $22.3 k_B T$ , roughly reflecting the stability of the C-terminal half of Vim2B-N. The folding free energy of the whole construct is  $35.2 (\pm 3.5) k_B T$ , as obtained by integrating the force-extension trace across the complete transition (SI Appendix, section 4.4). Hence, two thirds of the total free energy is provided by the C-terminal half.

### Vimentin Coil 2B Unfolds in a Clear Three-State Manner When Pulled from the C Terminus.

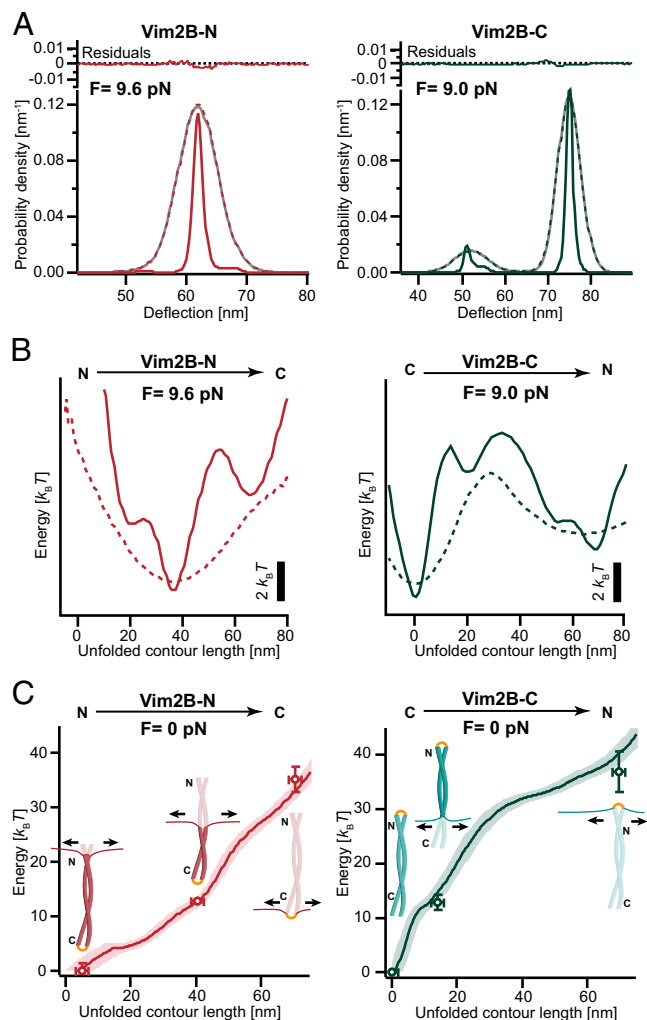
The unfolding/folding pattern of Vim2B-C seemingly is very different from that of Vim2B-N. The stretch-and-relax cycles show a steep rise up to 8 pN, corresponding to the stretching of the DNA handles. Between 8 and 10 pN, the

protein fluctuates quickly between the folded state and an obligatory intermediate (I2) (Fig. 1E). The contour length increase associated with unfolding into I2 is about  $14.2 (\pm 2.0)$  nm ( $19 \pm 3$  amino acids) and corresponds to the unfolding of nearly all residues of the conserved C-terminal region of coil 2B. At about 10 pN, the rest of the CC rips in one step to the fully unfolded state. The overall contour length increase of  $69.5 (\pm 2.0)$  nm corresponds to  $95 (\pm 3)$  folded amino acids and is close to the expected value for a fully folded CC (98 residues). Unfolding and refolding traces display a pronounced hysteresis, with refolding occurring only if the force is lowered below 5 pN at a trap velocity of 500 nm/s. This hysteresis demonstrates that Vim2B-C is much further from equilibrium than Vim2B-N. Nevertheless, we could observe slow folding/unfolding transitions of Vim2B-C between the fully folded state and the unfolded state (Fig. 1F) in passive mode. In addition to the slow folding/unfolding kinetics, rapid population of I2 from the folded state may be observed (Fig. 1F, *Inset*). Analysis of the equilibrium population under force yields a total free energy of  $36.8 (\pm 3.7) k_B T$  for Vim2B-C,  $12.9 (\pm 1.3) k_B T$  of which is contributed by the last 19 C-terminal residues unfolded in the intermediate. Hence, these 19 amino acids contribute about one third of the total free energy while comprising only one fifth of the total number of residues. We find the overall free energy is similar to the one obtained from the Vim2B-N construct. All lengths and energies of Vim2B-N and Vim2B-C are summarized in Table S1.

**Deconvolution Allows a More Elaborate Insight into the Energy Landscape of the Constructs.** To obtain a more detailed view of the energy landscape of coil 2B, we used deconvolution force spectroscopy (13, 32). In brief, in an optical tweezers experiment, the fluctuations of the ends of the CC are blurred by the thermal movements of beads attached to the DNA handles (Fig. 2A, gray line). Hence, bead position is not a faithful reporter of the actual protein extension, because the beads can move by extending the DNA, even if the protein ends do not. Given good knowledge of the thermal fluctuations of beads and handles, we can remove this blur from the passive mode data (Fig. 1D and F) by deconvolution and recover the intrinsic probability distribution of the fluctuations, owing to the opening of the CC (Fig. 2A, solid red and green lines; for details, see *Methods* and *SI Appendix*).

Using the Boltzmann equilibrium (Eq. S8), we can compute the energy landscapes of the CCs directly from the deconvolved probability distributions (Fig. 2B, solid lines). A comparison with the energy landscapes obtained from distributions before deconvolution (Fig. 2B, dotted lines) demonstrates the gain in fine structure upon deconvolution. Back transformation of the energy landscapes measured under force yielded the energy landscapes of Vim2B-N and Vim2B-C when no load is applied (Fig. 2C) (*SI Appendix*, section 5.3). The values obtained by the “conventional” energetic analysis (points in Fig. 2C) fit neatly onto the energy landscapes, validating the deconvolution procedure.

For additional validation of the inferred energy landscapes, we performed simulations of stretch-relax cycles for Vim2B-C and Vim2B-N based on the deconvolved energy landscapes of Fig. 2C. To this end, we modeled the contour length change of the protein as a diffusive process in the corresponding energy landscape. This process was coupled via worm-like chain linkers to beads that were allowed to diffuse in harmonic trap potentials. While gradually increasing/decreasing the trap distance, we simulated the equation of motion of the beads and the trajectory of the end-to-end distance of the protein. The bead trajectories then were converted directly to force-extension traces. For further details of the simulations, see *Methods* and *SI Appendix*. The result of stretch-and-relax cycle simulations performed at 500 nm/s pulling velocity for Vim2B-N and Vim2B-C is shown in *SI Appendix*, Fig. S3. For both constructs, the force-extension profiles of these cycles agree well with measured traces (compare to Fig. 1C and E). The simulations reproduce all basic features, such as intermediate states and the gradual unfolding of the N



**Fig. 2.** Model-free reconstruction of the full energy landscapes of Vim2B-N and Vim2B-C obtained by deconvolution. (A) The protein probability distributions of Vim2B-N (red, *Left*) and Vim2B-C (green, *Right*) may be obtained from the bead deflection probability distributions (gray) by deconvolution. The residuals show the difference between the reconvoled form of the deconvolved probability distributions (colored, dotted lines) and the measured distributions (gray line). (B) The deconvolved energy landscapes of Vim2B-N (continuous red line, *Left*) and Vim2B-C (continuous green line, *Right*) under force provide more details than the original bead position-derived landscapes (dotted lines). (C) Energy landscapes of Vim2B-N (red, *Left*) and Vim2B-C (green, *Right*) in the absence of force. Shaded areas are 68% confidence intervals obtained by bootstrapping. Points are values obtained from conventional data analysis (*SI Appendix*, section 4.4 and Table S1).

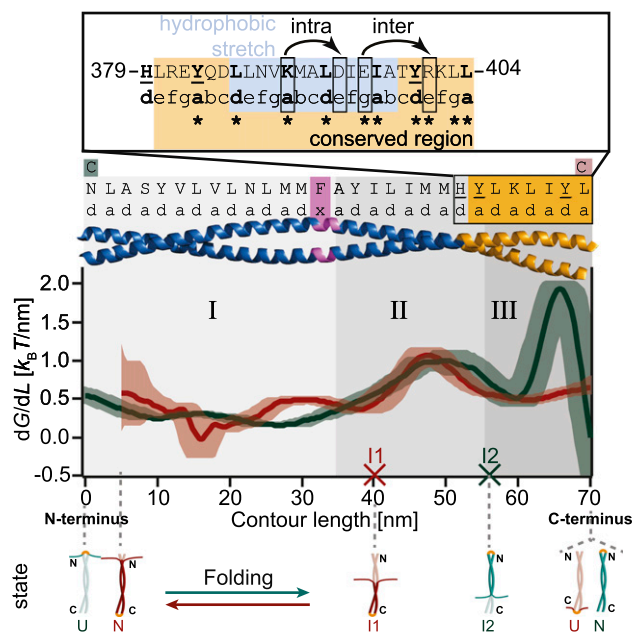
terminus of Vim2B-N, as well as the hysteresis in Vim2B-C, proving the validity of the deconvolved landscapes.

To be able to compare the two energy landscapes obtained by unzipping Vim2B from two different ends, we calculated the differential energy profiles (i.e.,  $dG/dL$ ) for both constructs (*SI Appendix*, section 5.8). These local stability profiles of the CC are shown in Fig. 3, with the length coordinate corresponding to Vim2B-N. Note that in this representation, the profile of Vim2B-C is reversed compared with its energy landscape in Fig. 2C. This local stability can be matched with its primary sequence. To simplify the sequence, only the *a*- and *d*-positions that form the hydrophobic seam are displayed in Fig. 3. Although the two energy profiles agree very well within errors in a large part along the sequence, there are notable differences, which we discuss below.

We find that the CC can be divided roughly into a labile N-terminal half (region I) with stability not exceeding  $0.5 k_B T/nm$  and a stable C-terminal part. In Vim2B-C (green), protein stability at the C terminus is unaffected by contributions from seed formation (*Discussion*). In this construct, the C-terminal region may be subdivided further into two sections (II and III) that are separated by a local stability minimum. The upstream section (region II) displays intermediate stability, whereas the downstream section (region III) is very stable, with values increasing to nearly  $2 k_B T/nm$ .

## Discussion

**Entropic Cost for Seed Formation and Exchanged Sites of Cross-Linking Account for the Differences in the Differential Energy Profiles.** In general, different pulling directions sample different reaction coordinates in a protein's energy landscape (28–31). However, in our experiments, the measured unfolding free energy profiles of Vim2B-N and Vim2B-C reflect the piecewise opening of the CC contacts, starting in a folded conformation and ending in a stretched and fully unfolded conformation. Even though the sequence of opening is reversed for both constructs, one would expect the free energy profiles to proceed along a similar reaction coordinate, albeit with reversed orientation, and hence contain very similar information. However, when looking at the force-extension traces (Fig. 1 *C* and *E*), one notices different intermediates I1 and I2 in the unfolding of Vim2B-N and Vim2B-C.



**Fig. 3.** The energy profiles of Vim2B-N (red) and Vim2B-C (green) reveal a tripartite stability of the CC. Derivative of the energy landscapes of Vim2B-N (red) and Vim2B-C (green) with respect to contour length aligned at the C-terminal end. The derivatives of the energy landscape describe the energy needed to unfold a certain part of the CC and therefore may be interpreted as a local stability profile (*SI Appendix*). The contour lengths of the intermediates I1 and I2 obtained by conventional data analysis are indicated by red and green crosses. Shaded areas are 68% confidence intervals from bootstrapping. The first 5 nm of the stability profile of Vim2B-N are not shown, as errors exceed 100%. The residues occupying *a*- and *d*-positions of the CC are listed on top. Cysteine mutations are colored red/green according to construct. (*Inset*) Features of the conserved region (orange). The position of the stutter is marked in purple. The main hydrophobic stretch of coil 2B is highlighted in blue. Bulky residues causing local stability minima are underlined. Bent arrows indicate the intra- and interhelical salt bridges. Asterisks mark absolutely conserved residues that are unchanged in seven human IF proteins constituting major representatives of all five sequence homology classes of IFs (23).

In addition, the discrepancy between the pronounced hysteresis of Vim2B-C and the close-to-equilibrium fluctuations of Vim2B-N (Fig. 1 *C* and *E*) recorded at identical pulling speeds is striking. Moreover, the energy profiles derived from Vim2B-N and Vim2B-C seem to diverge, especially at the C-terminal end (Fig. 3).

These discrepancies may be explained by two different contributions: First, the need for cross-linking cysteines makes the two protein sequences slightly different. Second, nucleation seed formation that starts folding will occur at different termini in the two constructs. Hence, the free energy profile of Fig. 3 also contains the entropic cost of ordering the residues forming the nucleation seed for further propagation. We argue that the asymmetry caused by the exchanged sites of cross-linking and force application is a major determinant of the observed differences.

Ensemble studies have shown that folding of a CC is initiated by a nucleation mechanism that involves the collision of unstructured or preformed  $\alpha$ -helical chains forming a CC (10, 12, 33). The different cross-linking geometries used in our mechanical unzipping assays allow us to control at which terminus of the protein this nucleation seed forms. According to Sosnick and coworkers (11), introducing a cross-link between the two chains of a CC forces the CC to nucleate at the cross-linked end. Under mechanical load, seed formation at the distal end is inhibited further by the energy costs associated with bringing the untethered ends together. Hence, folding of Vim2B-C demands seed formation at the labile N terminus, whereas the formation of the Vim2B-N coil involves seed formation at the stable C terminus (region III).

The effect of seed formation is clearly visible in region III (Fig. 3) at the C terminus, where the green curve (Vim2B-C) lies above the red curve (Vim2B-N), as the Vim2B-N construct has to form its seed at the C terminus whereas Vim2B-C has already adopted the CC structure at this position. Thus, Vim2B-N consumes the free energy provided by the favorable interactions of the two chains to form the seed, whereas Vim2B-C does not. As the entropic cost paid for seed formation should be similar for both constructs, a comparable discrepancy between the two traces also should occur at the N-terminal end, where Vim2B-C has to pay the price for seed formation instead of Vim2B-N. However, at the N terminus, the red trace seems to lie only marginally above the green trace. At this point, it is important to note that the cross-links introduce a divergence at the sequence level among the two constructs. Vim2B-C is cross-linked at the N terminus where the cysteine replaces an asparagine. In contrast, Vim2B-N is cross-linked at the C terminus, with the cysteine replacing a leucine. Stability studies of CCs have shown that introducing an asparagine at a *d*-position is energetically unfavorable and destabilizes a CC by  $1.0 k_B T$ , whereas leucine at the *a*-position stabilizes by  $5.9 k_B T$  compared with a CC with an alanine in the same position (34–36). Hence, the measured free energy profile of Vim2B-C overestimates the free energy at the N terminus compared with the profile of Vim2B-N. Similarly, the profile of Vim2B-N underestimates the free energy at the C terminus relative to Vim2B-C, because it misses the last favorable leucine contact substituted by a nonopening cysteine cross-link. To illustrate those effects, we show a modified energy profile in Fig. S4, in which we try to restore the energetic equality between the two constructs.

After this correction, the free energies for seed formation (hatched areas in Fig. S4) may be estimated to be about  $4 k_B T$ . Even though direct information about seed length is missing, in this representation, the seed of Vim2B-C is longer, whereas we assume the seed of Vim2B-N is more localized owing to the high stability of region III at the C terminus.

It has been proposed that the highly stable C-terminal end of the CC (region III) harbors a trigger sequence (37) offering a site for seed formation in the initial stage of folding. Trigger sequences have been reported for many CCs; however, whether they are indispensable for folding is debated (38–41). It appears that a main role of trigger sequences is to offer enough free energy (13, 42) to overcome the cost of seed formation, which is

in accord with our findings here. Therefore, the observed folding and seed formation of Vim2B-N is likely to resemble the unconstrained cross-link-less folding pathway more closely than does Vim2B-C. Based on our results here, the two seed lengths and thus the assumed function of the trigger sequence might be tested further by conducting a mutational  $\Phi$ -value analysis (11, 43, 44).

The sequence difference between the constructs (leucine/asparagine) also might account for the slightly higher total free energy we measure for the Vim2B-C coil compared with the Vim2B-N construct. Moreover, the first eight amino acids of Vim2B-N but not of Vim2B-C are likely to be unfolded. The contour length increase of this observed unfolding is too great to result from strain induced by the crosslink. We therefore hypothesize that the shorter measured contour length of Vim2B-N is a result of the destabilizing effect of the asparagine (Table S1).

**The Tripartite Stability of the Vim2B Coiled Coil.** Our energy profiles reveal a tripartite stability of the CC, a labile N-terminal half (region I), a medium stability section (region II), and a highly stable region (region III) at the far C-terminal end. These findings, combined with sequence and structural characteristics, allow us to suggest a plausible role for stability variations along vimentin coil 2B.

Crystallization studies of different vimentin fragments indicate that the N-terminal residues 306–340 alone either do not fold or form very labile CCs at most (6, 23). Indeed, our region I corresponds to this stretch of amino acids and exhibits very low stability, with the first eight N-terminal amino acids presumably being unstructured, as explained above. Within full-length vimentin, region I is flanked by two specific traits: at its N-terminal end, the regular left-handed CC of coil 2B changes into a parallel  $\alpha$ -helical bundle of helices (6) (residues 261–300; formerly coil 2A and linker 2), whereas the C terminus is limited by the presence of a stutter (residues 351–354) (23). The low stability of region I may be essential to allow for the drastic structural changes occurring in the flanking regions.

Concerning the stutter, one would assume that this irregularity has a major influence on CC stability. However, we do not observe a major drop in free energy other than marking the transition to the more stable C terminus. Indeed, Strelkov et al. (23) discovered that the stutter may be accommodated in the structure without destroying the CC. Compensation of the irregularity is achieved by a local unwinding of the CC where the two helices nearly run in parallel, allowing preservation even of the hydrogen bonding pattern and the overall CC geometry.

Intriguingly, the most stable region we find at the far C-terminal end (region III) is largely identical to the highly conserved region (380–404) in coil 2B (Fig. 3, *Inset*) (23). The role of this region is twofold: it contains a trigger sequence for folding, as mentioned above, and simultaneously serves as a clamp with a distinct structural design, thus preventing spontaneous opening of the ends.

Two properties account for its extraordinarily high stability: a long hydrophobic stretch and two salt bridges. The main apolar stretch of coil 2B comprises 13 residues (386–399) at the C-terminal end (18). This region also harbors two salt bridges, an intrahelical one between residues K390 and D394 and an interhelical salt bridge between E396 and R401, which are an integral part of the trigger sequence (23, 37, 45).

Disease-causing mutations in IFs significantly cluster in region III (4, 26). For instance, point mutations affecting the formation of the salt bridges in the related IF protein desmin are responsible for severe cases of desminopathy, with an early disease onset and life-threatening consequences (46). Mutated desmins exhibit compromised assembly, forming mostly very short IFs (46–48).

The energy profile of Vim2B-C reveals that the highly stable region III is bounded by bulky residues in the hydrophobic core defining local energy minima (see underlined residues in Fig. 3, *Inset*). At its N-terminal end, region III is terminated by a tyrosine in the *a*-position 383 and by histidine 379. On the C-terminal

side, the region is terminated by a tyrosine in *d*-position 401. These residues deviate from the residues preferentially found in core positions (Val, Ile, Leu, and Thr) (9). Both tyrosine and histidine are less hydrophobic than the typical core residues and destabilize by decreasing hydrophobic interactions and representing a steric hindrance in the core of the CC, leading to the measured local energy minima (35, 36).

It was proposed that the overlapping helices of coil 1A and coil 2B may participate in a four- $\alpha$ -helical bundle when dimers interact at their adjacent ends during longitudinal annealing of ULFs (Fig. S5, model 1) (2, 18, 23, 24, 49). However, this end-on annealing of two dimers to form a longitudinal type of tetrameric complex (i.e., in the  $A_{CN}$  mode) would demand a major structural reorganization of coil 2B and an unzipping of the coil at its C-terminal part (region III). This event is unlikely in light of the high stability of region III we report here. Instead, our results favor an alternative assembly model (25) in which the elongation occurs via lateral association of the overlap segments of the ends of the two adjacent 1A and 2B CCs (Fig. S5, model 2). Here, coil 2B would react as a stiff fold and the elongation might be stabilized by electrostatic interactions of the conserved acidic residues EGEE downstream of the end of coil 2 with polar residues in coil 1A (25).

Using mutational analysis, future studies might test the importance of the stabilizing effects of the salt bridges in the trigger sequence. We expect these disease-related mutations to decrease the stability of region III, further elucidating the importance of thermodynamic stability of region III for assembly.

In earlier stages of assembly, on the level of tetramers and ULFs, coil 2B protrudes as a whole from the structures and hence does not participate in favorable dimer–dimer interactions (5, 16, 17). Thus, the high stability of region III and the intermediate stability of region II presumably ensure that the CC remains zipped during the first phase of lateral assembly and stabilizes IFs on the level of the dimer in both models of assembly.

## Methods

**Experimental Procedures.** For optical trapping experiments, the two constructs were cross-linked at the respective termini and connected to the beads via DNA–ubiquitin handles at the opposite termini. For a more detailed description, see *SI Appendix*.

**Conventional Data Analysis.** In intermediate force regimes when the protein transitioned between a series of states, we used hidden Markov models on the passive mode data to assign data points to states as described earlier (50). For a more detailed description, see *SI Appendix*.

**Deconvolution Procedure.** We used linear and nonlinear stretching models of the mechanical components in the system to describe the Hamiltonian of bead–protein–bead dumbbells in the optical trap as a sum of the energy landscape of the protein  $\bar{G}_0$  and the mechanical energy involved in stretching the linkers, i.e.,  $\mathcal{H}_d(x, L_p) = \bar{G}_0(L_p) + \mathcal{H}_d^{\text{mech}}(x, L_p)$ , where  $L_p$  is the unfolded contour length,  $d$  is the trap distance, and  $x = x_1 + x_2$  is the sum of the bead deflections. This allowed us to determine the shape of the point spread function  $\Psi_d(x)$  at infinite bandwidth, describing the thermal noise arising from the mechanical components, for varying forces. After calibration for finite bandwidth, we used this set of point spread functions to remove the thermal noise by deconvolution. To this end, we adapted a rigorous deconvolution procedure introduced by Hinczewski et al. (51), which optimizes the estimate for the deconvolved energy landscape by fitting its convolved form to the measured bead–location distribution. By using the full energetic description of the system outlined above, the deconvolved bead distribution then was transformed into a zero-force energy landscape for the unzipping of the protein. From the zero-force energy landscape, we also obtained local stability profiles  $d\bar{G}_0/dL_p$ , which describe the energetic cost of unfolding a specific part of a protein. For a detailed description, see *SI Appendix*.

**Simulation.** To verify the performance of our deconvolution procedure, we used the measured energy landscapes to reproduce stretch-and-relax cycles in simulations. The simulations were designed to capture several important features of the experimental data: the force dependence of the width and

shape of noise, the impact of force on the end-to-end distance of the protein (i.e., the force-dependent folding/unfolding trajectories), and the coupling of the responses of protein and beads. To this end, we concurrently simulated the coupled time evolution of the coordinates  $x$  and  $L_p$  in the potential  $\mathcal{H}_d(x, L_p)$  using Langevin dynamics. These simulations describe the noise of the thermal fluctuations in the trap as well as the coupled dynamics of the end-to-end distance of the protein:

$$\begin{aligned}\Delta x &= \frac{\Delta t}{\gamma_b} \left( -\frac{\partial \mathcal{H}_d}{\partial x}(x, L_p) k_B T + \sqrt{2k_B T \frac{\gamma_b}{\Delta t}} \varepsilon(t) \right), \\ \Delta L_p &= \frac{\Delta t}{\gamma_p} \left( -\frac{\partial \mathcal{H}_d}{\partial L_p}(x, L_p) k_B T + \sqrt{2k_B T \frac{\gamma_p}{\Delta t}} \varepsilon(t) \right),\end{aligned}$$

where  $\varepsilon(t)$  describes white noise with an SD of  $\sigma = 1$ . The friction coefficient  $\gamma_b$  is dominated by hydrodynamic friction on the beads and was calculated

- Herrmann H, Aebi U (2004) Intermediate filaments: Molecular structure, assembly mechanism, and integration into functionally distinct intracellular scaffolds. *Annu Rev Biochem* 73(1):749–789.
- Herrmann H, Strelkov SV, Burkhard P, Aebi U (2009) Intermediate filaments: Primary determinants of cell architecture and plasticity. *J Clin Invest* 119(7):1772–1783.
- Goldman RD, Cleland MM, Murthy SN, Mahammad S, Kuczumarski ER (2012) Inroads into the structure and function of intermediate filament networks. *J Struct Biol* 177(1):14–23.
- Szeverenyi I, et al. (2008) The Human Intermediate Filament Database: Comprehensive information on a gene family involved in many human diseases. *Hum Mutat* 29(3):351–360.
- Steinert PM, Marekov LN, Parry DA (1993) Diversity of intermediate filament structure. Evidence that the alignment of coiled-coil molecules in vimentin is different from that in keratin intermediate filaments. *J Biol Chem* 268(33):24916–24925.
- Chernyatina AA, Strelkov SV (2012) Stabilization of vimentin coil2 fragment via an engineered disulfide. *J Struct Biol* 177(1):46–53.
- Chernyatina AA, Nicolet S, Aebi U, Herrmann H, Strelkov SV (2012) Atomic structure of the vimentin central  $\alpha$ -helical domain and its implications for intermediate filament assembly. *Proc Natl Acad Sci USA* 109(34):13620–13625.
- Moutevelis E, Woolfson DN (2009) A periodic table of coiled-coil protein structures. *J Mol Biol* 385(3):726–732.
- Lupas AN, Gruber M (2005) The structure of  $\alpha$ -helical coiled coils. *Adv Protein Chem* 70:37–78.
- Zitzewitz JA, Ibarra-Molero B, Fishel DR, Terry KL, Matthews CR (2000) Preformed secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system. *J Mol Biol* 296(4):1105–1116.
- Moran LB, Schneider JP, Kentis A, Reddy GA, Sosnick TR (1999) Transition state heterogeneity in GCN4 coiled coil folding studied by using multisite mutations and crosslinking. *Proc Natl Acad Sci USA* 96(19):10699–10704.
- Meisner WK, Sosnick TR (2004) Fast folding of a helical protein initiated by the collision of unstructured chains. *Proc Natl Acad Sci USA* 101(37):13478–13482.
- Gebhardt JC, Bornschlöggl T, Rief M (2010) Full distance-resolved folding energy landscape of one single protein molecule. *Proc Natl Acad Sci USA* 107(5):2013–2018.
- Gao Y, Sirinakis G, Zhang Y (2011) Highly anisotropic stability and folding kinetics of a single coiled coil protein under mechanical tension. *J Am Chem Soc* 133(32):12749–12757.
- Gao Y, et al. (2012) Single reconstituted neuronal SNARE complexes zipper in three distinct stages. *Science* 337(6100):1340–1343.
- Sokolova AV, et al. (2006) Monitoring intermediate filament assembly by small-angle x-ray scattering reveals the molecular architecture of assembly intermediates. *Proc Natl Acad Sci USA* 103(44):16206–16211.
- Mücke N, et al. (2004) Molecular and biophysical characterization of assembly-starter units of human vimentin. *J Mol Biol* 340(1):97–114.
- Smith TA, Strelkov SV, Burkhard P, Aebi U, Parry DAD (2002) Sequence comparisons of intermediate filament chains: Evidence of a unique functional/structural role for coiled-coil segment 1A and linker L1. *J Struct Biol* 137(1–2):128–145.
- Mücke N, et al. (2004) Assessing the flexibility of intermediate filaments by atomic force microscopy. *J Mol Biol* 335(5):1241–1250.
- Staple DB, Loparic M, Kreuzer HJ, Kreplak L (2009) Stretching, unfolding, and deforming protein filaments adsorbed at solid-liquid interfaces using the tip of an atomic-force microscope. *Phys Rev Lett* 102(12):128302.
- Kreplak L, Fudge D (2007) Biomechanical properties of intermediate filaments: From tissues to single filaments and back. *BioEssays* 29(1):26–35.
- Swift J, et al. (2013) Nuclear lamin-A scales with tissue stiffness and enhances matrix-directed differentiation. *Science* 341(6149):1240104.
- Strelkov SV, et al. (2002) Conserved segments 1A and 2B of the intermediate filament dimer: Their atomic structures and role in filament assembly. *EMBO J* 21(6):1255–1266.
- Kapinos LE, et al. (2010) Characterization of the head-to-tail overlap complexes formed by human lamin A, B1 and B2 “half-minilamin” dimers. *J Mol Biol* 396(3):719–731.
- Strelkov SV, Schumacher J, Burkhard P, Aebi U, Herrmann H (2004) Crystal structure of the human lamin A coil 2B dimer: Implications for the head-to-tail association of nuclear lamins. *J Mol Biol* 343(4):1067–1080.

from first principles, whereas the friction coefficient of movement in the protein energy landscape  $\gamma_p$  was matched against measured transition rates. The mechanical parameters were chosen to mimic experimental conditions. Simulations were performed at time steps of  $10^{-7}$  s and sampled at 20 kHz while the trap distance  $d$  was varied continuously. Stretch-relax cycles then were obtained from the bead deflection trajectory  $x(t)$ . *SI Appendix* contains a detailed description.

**ACKNOWLEDGMENTS.** We thank Anastasia A. Chernyatina and Sergei V. Strelkov for providing the three-dimensional structural model of the vimentin dimer. J.S. was supported by the Elitenetzwerk Bayern in the framework of the doctorate program Material Science of Complex Interfaces. D.T. is grateful to the National Institutes of Health for partial support through Grant GM 089685. M.R. acknowledges support through a Sonderforschungsbereich 863 A2 grant from Deutsche Forschungsgemeinschaft.

- Clemen CS, Herrmann H, Strelkov SV, Schröder R (2013) Desminopathies: Pathology and mechanisms. *Acta Neuropathol* 125(1):47–75.
- Parry DAD, Steinert PM (1999) Intermediate filaments: Molecular architecture, assembly, dynamics and polymorphism. *Q Rev Biophys* 32(2):99–187.
- Dietz H, Rief M (2006) Protein structure by mechanical triangulation. *Proc Natl Acad Sci USA* 103(5):1244–1247.
- Brockwell DJ, et al. (2003) Pulling geometry defines the mechanical resistance of a beta-sheet protein. *Nat Struct Biol* 10(9):731–737.
- Shank EA, Ceccoli 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.
- Carrion-Vazquez M, et al. (2003) The mechanical stability of ubiquitin is linkage dependent. *Nat Struct Biol* 10(9):738–743.
- Woodside MT, et al. (2006) Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* 314(5801):1001–1004.
- Bornschlöggl T, Gebhardt JCM, Rief M (2009) Designing the folding mechanics of coiled coils. *ChemPhysChem* 10(16):2800–2804.
- Bornschlöggl T, Rief M (2006) Single molecule unzipping of coiled coils: Sequence resolved stability profiles. *Phys Rev Lett* 96(11):118102.
- Tripet B, Wagschal K, Lavigne P, Mant CT, Hodges RS (2000) Effects of side-chain characteristics on stability and oligomerization state of a de novo-designed model coiled-coil: 20 amino acid substitutions in position “d.” *J Mol Biol* 300(2):377–402.
- Wagschal K, Tripet B, Lavigne P, Mant C, Hodges RS (1999) The role of position a in determining the stability and oligomerization state of  $\alpha$ -helical coiled coils: 20 amino acid stability coefficients in the hydrophobic core of proteins. *Protein Sci* 8(11):2312–2329.
- Wu KC, et al. (2000) Coiled-coil trigger motifs in the 1B and 2B rod domain segments are required for the stability of keratin intermediate filaments. *Mol Biol Cell* 11(10):3539–3558.
- Steinmetz MO, et al. (2007) Molecular basis of coiled-coil formation. *Proc Natl Acad Sci USA* 104(17):7062–7067.
- Kammerer RA, et al. (1998) An autonomous folding unit mediates the assembly of two-stranded coiled coils. *Proc Natl Acad Sci USA* 95(23):13419–13424.
- Burkhard P, Kammerer RA, Steinmetz MO, Bourenkov GP, Aebi U (2000) The coiled-coil trigger site of the rod domain of cortexillin I unveils a distinct network of interhelical and intrahelical salt bridges. *Structure* 8(3):223–230.
- Lee DL, Lavigne P, Hodges RS (2001) Are trigger sequences essential in the folding of two-stranded  $\alpha$ -helical coiled-coils? *J Mol Biol* 306(3):539–553.
- Ibarra-Molero B, Zitzewitz JA, Matthews CR (2004) Salt-bridges can stabilize but do not accelerate the folding of the homodimeric coiled-coil peptide GCN4-p1. *J Mol Biol* 336(5):989–996.
- Matthews CR (1987) Effect of point mutations of the folding of globular proteins. *Methods Enzymol* 154:498–511.
- Fersht AR, Matouschek A, Serrano L (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* 224(3):771–782.
- Herrmann H, et al. (2000) The intermediate filament protein consensus motif of helix 2B: Its atomic structure and contribution to assembly. *J Mol Biol* 298(5):817–832.
- Goudeau B, et al. (2006) Variable pathogenic potentials of mutations located in the desmin  $\alpha$ -helical domain. *Hum Mutat* 27(9):906–913.
- Bär H, et al. (2005) Severe muscle disease-causing desmin mutations interfere with in vitro filament assembly at distinct stages. *Proc Natl Acad Sci USA* 102(42):15099–15104.
- Bär H, et al. (2006) Impact of disease mutations on the desmin filament assembly process. *J Mol Biol* 360(5):1031–1042.
- Meier M, et al. (2009) Vimentin coil 1A-A molecular switch involved in the initiation of filament elongation. *J Mol Biol* 390(2):245–261.
- Stigler J, Rief M (2012) Hidden Markov analysis of trajectories in single-molecule experiments and the effects of missed events. *ChemPhysChem* 13(4):1079–1086.
- Hinczewski M, Gebhardt JCM, Rief M, Thirumalai D (2013) From mechanical folding trajectories to intrinsic energy landscapes of biopolymers. *Proc Natl Acad Sci USA* 110(12):4500–4505.

# Sequence-resolved free energy profiles of stress bearing vimentin intermediate filaments — Supporting Information (SI)

Beatrice Ramm<sup>a,1</sup>, Johannes Stigler<sup>a,1,2</sup>, Michael Hinczewski<sup>b</sup>, D. Thirumalai<sup>b</sup>, Harald Herrmann<sup>c</sup>,  
Günther Woehlke<sup>a</sup> and Matthias Rief<sup>a,d,3</sup>

<sup>a</sup>Physik Department E22, Technische Universität München, James-Frank-Straße, 85748 Garching, Germany

<sup>b</sup>Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, United States

<sup>c</sup>Group Functional Architecture of the Cell, German Cancer Research Center, 69120 Heidelberg, Germany

<sup>d</sup>Nanosystems Initiative Munich, Schellingstraße 4, 80799 München, Germany

<sup>1</sup>BR and JS contributed equally

<sup>2</sup>Present address: Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, United States

<sup>3</sup>Corresponding author

## Contents

<b>1</b>	<b>Sequences</b>	<b>2</b>
<b>2</b>	<b>Experimental procedures</b>	<b>2</b>
<b>3</b>	<b>Elastic linker models</b>	<b>2</b>
<b>4</b>	<b>Conventional data analysis</b>	<b>3</b>
4.1	Length calculations . . . . .	3
4.2	Quasi-equilibrium unfolding description of the first transition in Vim2B-N . . . . .	4
4.3	Hidden Markov identification of states . . . . .	4
4.4	Energy determination . . . . .	4
<b>5</b>	<b>Direct measurement of energy landscapes</b>	<b>5</b>
5.1	Deconvolution procedure . . . . .	5
5.2	Conversion from $G_d(x)$ to $\tilde{G}_d(L_p)$ . . . . .	6
5.3	Transformation of energy landscapes to different trap distances/forces . . . . .	7
5.4	Point spread functions . . . . .	7
5.5	WHAM . . . . .	8
5.6	Combination of energy landscapes . . . . .	9
5.7	Error estimation . . . . .	9
5.8	Local stability plots . . . . .	9
<b>6</b>	<b>Simulation</b>	<b>9</b>

## 1 Sequences

The sequences of Vim2B-N and Vim2B-C are shown below. Vimentin is shown in red, ubiquitin in blue. The cysteine residue for handle attachment is shown in bold, the cysteine for coiled-coil crosslinking is underlined.

**Vim2B-C** MGCNDALRQAKQESTEYRRQVQSLTSEVDALKGTNESLERQMREMEENFAVEAANYQDTIGRLQDEIQN-MKEEMARHLREYQDLLNVKMALDIEIATYRKLGTMQIFVKTTLTGKTITLEVEPSDTIENVKAKIQD KEGIPPDQQR-LIFAGKQLEDGRTLSD YNIQKESTLHLVLRRLRGGKCLE-His<sub>6</sub>

**Vim2B-N** MACKMQIFVKTTLTGKTITLEVEPSDTIENVKAKIQ DKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRRLRGGELNNDALRQAKQESTEYRRQVQSLTSEVDALKGTNESLERQMREMEENFAVEAANYQDTIGRLQDEIQN-MKEEMARHLREYQDLLNVKMALDIEIATYRKLCKLAAALE-His<sub>6</sub>

## 2 Experimental procedures

Vim2B-N and Vim2B-C comprise the residues 306-404 of human vimentin (UniProt accession number: P08670) with the introduced point mutation C328S that prevents unwanted attachment of handle oligonucleotides to the exposed residue 328. Furthermore, both constructs contain a terminal cysteine for chemical intra-homodimeric crosslinking (Vim2B-C: N-terminal; Vim2B-N: C-terminal) and a spacer-ubiquitin with a terminal cysteine (Vim2B-C: C-terminal; Vim2B-N: N-terminal). The crosslinking cysteines replace residues in the hydrophobic core in positions a or d while the ubiquitin cysteines are solvent accessible. The ubiquitin cysteines were used for attachment of about 180 nm long DNA handles with biotin/digoxigenin ends using a protocol derived from work by Cecconi et al. [1], as described in the methods section of Stigler et al. [2]. The DNA-protein construct was assembled in a dumbbell configuration between micron-sized neutravidin/anti-digoxigenin-labeled silica beads for mechanical measurement. Experiments were performed in 10 mM phosphate, 2.7 mM potassium chloride, 137 mM sodium chloride, 26 U/ml glucose oxidase, 17000 U/ml catalase, 0.65 % glucose at pH 7.4 in a custom built dual beam optical tweezers setup with one controllable trap [3]. The trap stiffness and sensitivity were calibrated in a combined method [4]. During measurements we recorded the sum of the bead deflection signals of the two traps,  $x = x_1 + x_2$ . By linearly increasing/decreasing the trap distance, we generated force-distance and force-extension cycles. By keeping the trap distance constant, we also recorded time traces of  $x(t)$  in 'passive mode'.

## 3 Elastic linker models

The proteins in our experiment were tethered to dsDNA molecules, whose elastic properties are best described using an extensible wormlike chain model [5], where the force as a function of extension  $\xi$  is given by

$$F_{\text{eWLC}}(\xi) = \frac{k_B T}{p_D} \left( \frac{1}{4 \left(1 - \frac{\xi}{L_D}\right)} - \frac{1}{4} + \frac{\xi}{L_D} - \frac{F_{\text{eWLC}}}{K} \right). \quad (\text{S1})$$



Here,  $L_D$  is the contour length,  $p_D$  the persistence length and  $K$  is the stretch modulus of the DNA. In our experiments, typical values were  $p_D \approx 20 - 25$  nm,  $L_D \approx 360$  nm,  $K \approx 300$  pN, comparable to values obtained in earlier studies using the same linker system [2, 6].

Unlike globular proteins, which often exhibit an all-or-nothing folding behavior, the folding of coiled coils can arguably be described using a zippering mechanism. We describe the degree of folding using the coordinate of unfolded free contour length of the protein  $L_p$ .

The mechanical properties of the polypeptide part were modeled using a wormlike chain model [7]:

$$F_{\text{WLC}}(\xi) = \frac{k_B T}{p_p} \left( \frac{1}{4 \left(1 - \frac{\xi}{L_p}\right)} - \frac{1}{4} + \frac{\xi}{L_p} \right), \quad (\text{S2})$$

where  $p_p$  is the persistence length of the unfolded polypeptide. In this study we used  $p_p = 0.7$  nm, in accordance with earlier measurements [8, 3].

In the experimental configuration, the two mechanical parts consisting of dsDNA and unfolded polypeptide are in series. Hence, the extension of the full linker consisting of dsDNA and unfolded polypeptide is given by

$$\xi_{\text{construct}}(F) = \xi_{\text{eWLC}}(F) + \xi_{\text{WLC}}(F), \quad (\text{S3})$$

where  $\xi_{\text{eWLC}}$  and  $\xi_{\text{WLC}}$  are given by eq. (S1) and eq. (S2). The inverse of eq. (S3) yields the force on the construct as a function of length of unfolded polypeptide and total extension  $F_{\text{construct}}(L_p, \xi)$ .

## 4 Conventional data analysis

### 4.1 Length calculations

The number of folded amino acids (aa) was determined using a value of 0.365 nm/aa [9]. We chose the construct Vim2B-C, whose unfolding forces are higher, as a reference construct of length. The full measured contour length of Vim2B-C is  $\approx 70$  nm (see Table S1), corresponding to  $95 \pm 3$  folded amino acids in the monomer, close to the expected value of 98 amino acids for a fully folded coiled coil. The measured length for Vim2B-N is  $65.4 \pm 2.0$  nm, shorter than expected for a fully folded coiled coil. The first few N-terminal amino acids in Vim2B-N are hence likely to be unfolded at our lowest resolvable forces ( $\lesssim 2 - 3$  pN). When we compare absolute length coordinates between Vim2B-N and Vim2B-C, we align such that the unfolded state of Vim2B-N,  $U_{\text{Vim2B-N}}$ , corresponds to the folded state of Vim2B-C  $N_{\text{Vim2B-C}}$ .

The discrepancy in measured length between Vim2B-N and Vim2B-C can be explained by the asymmetry of the crosslinking cysteine. In the case of Vim2B-N, the crosslink is at the C-terminus, distant from the putatively unfolded N-terminal part. In the case of Vim2B-C, the crosslink is at the N-terminus, keeping it connected when unzipping from the stable C-terminus.

## 4.2 Quasi-equilibrium unfolding description of the first transition in Vim2B-N

We used the following model to estimate the number of relevant individual unfolding steps that are involved in the first hump-like transitions of Vim2B-N (Fig. 1C). Assuming that this transition occurs in  $N$  individual substeps of equal contour length and the energy of the full transition is  $\Delta G_{\text{tot}} < 0$ , we can write down the Hamiltonian [10]

$$\mathcal{H}_d(x, i) = \left(1 - \frac{i}{N}\right) \cdot \Delta G_{\text{tot}} + \mathcal{H}_d^{\text{mech}}(x, L_p = \frac{i}{N} \cdot L_{p,\text{tot}}) \quad (\text{S4})$$

and use it, assuming quasi-equilibrium conditions, i.e. very slow pulling speeds, to calculate the mean force from the mean bead deflection at a trap distance  $d$ :

$$\langle F \rangle = k_c \cdot \langle x \rangle = k_c \cdot \frac{\int_{-\infty}^{\infty} dx \sum_{i=0}^N x \exp\left(-\frac{\mathcal{H}_d(x, i)}{k_B T}\right)}{\int_{-\infty}^{\infty} dx \sum_{i=0}^N \exp\left(-\frac{\mathcal{H}_d(x, i)}{k_B T}\right)} \quad (\text{S5})$$

The energy describing the purely mechanical properties of the system  $\mathcal{H}_d^{\text{mech}}(x, L_p)$  is defined in eq. (S11).

In the first hump-like transition of Vim2B-N we were unable to discern discrete transitions between states. When fitting this transition region in constant-velocity pulling traces to the above model, it becomes evident that a two state model ( $N = 1$ ) cannot describe the measured data (Fig. S2). A fit to a model with one populated intermediate ( $N = 2$ ) succeeds (Fig. S2). Fits to models with  $N > 2$  also describe the data well but are virtually indistinguishable from the case  $N = 2$ . We thus infer that the transition involves at least one intermediate, possibly more. Since we could not directly resolve the transitions between any of the intermediate states we conclude that the barriers between these states are small, suggesting that the N-terminal half of the energy landscape of Vim2B is generally flat under force.

## 4.3 Hidden Markov identification of states

Whenever transitions between states were identifiable in passive mode traces (transitions between  $N$ ,  $U$  and  $I_2$  in Vim2B-C or transitions between  $I_1$  and  $U$  in Vim2B-N), we used hidden Markov models to identify state positions, occupancies and dwell times as described elsewhere [11, 2].

## 4.4 Energy determination

For Vim2B-C, where the stretch-relax cycles are far from equilibrium, but individual states are discernible, we used the state occupancies from hidden Markov models to calculate the free energy differences between the states under zero-force conditions [3, 2].

For Vim2B-N this approach was not feasible since the first hump-like transition did not show clearly identifiable states. However, stretch-and-relax cycles of this construct were close to equilibrium, which allowed us to determine the difference in free energy between the folded ( $N$ ) and unfolded state ( $U$ ) directly from the mechanical work involved in stretching and unfolding the constructs by integrating the force-distance trace  $F(d)$  and subtracting the mechanical

work required to stretch the linker:

$$\Delta G = \int_0^{d_{\max}} dd F(d) + k_B T \ln \int_{-\infty}^{\infty} dx \exp\left(-\frac{\mathcal{H}_{d_{\max}}^{\text{mech}}(x, L_p = L_{\text{unf}})}{k_B T}\right) \quad (\text{S6})$$

where  $L_{\text{unf}}$  is the contour length of the unfolded state and  $d_{\max}$  is the maximal trap separation (well beyond the unfolding transition to which the construct was stretched). The definition of the mechanical work for stretching the linker and deflecting the beads from the trap centers  $\mathcal{H}_d^{\text{mech}}(x, L_p)$  is given in eq. (S11).

In addition, the difference in energy between the folded state  $N$  and the intermediate  $I_1$  could be determined using eq. (S5) as a fit function to the low-force transition in stretch-relax cycles of Vim2B-N (Fig. S2).

## 5 Direct measurement of energy landscapes

Our experimental observable is the deflection of the beads from the trap centers  $x = x_1 + x_2$ . When the protein construct is held in passive mode at a constant trap separation  $d$ , we record the fluctuations of  $x$  over time. From long mechanical folding trajectories we obtain the equilibrium distribution  $P_d(x)$ .

The fluctuations of the protein are masked by thermal noise coupling to the mechanical components in the system. The distribution  $P_d(x)$  can hence mathematically be described as the result of a convolution of the distribution of the protein fluctuations  $Q_d(x)$  and the distribution of the fluctuations of beads and dsDNA linkers,  $\Psi_a(x)$  (point spread function, PSF):

$$P_d(x) = \Psi_x(x) \otimes Q_d(x) = \int dx' \Psi_x(x - x') Q_d(x') \quad (\text{S7})$$

In our passive mode optical tweezers experiment the trap stiffnesses of the individual traps  $k_1, k_2$  are both greater than zero. Hence, the combined trap stiffness of the traps and linkers changes with applied force [12]. When observing a protein in the folded state, the forces and the total trap stiffness are higher, resulting in a narrow noise distribution. In contrast, when the protein is in the unfolded state, the force drops and the noise is broader. We account for this effect using a force-dependent point spread function  $\Psi_a(x)$  [3].  $\Psi_a(x)$  describes the distribution of bead deflection values  $x$  when the mean deflection is  $a$  and is calculated from the mechanical properties of the system as described in section 5.4.

The free energy profile of the protein  $G_d(x)$  and the probability distribution  $Q_d(x)$  are related via a Boltzmann inversion:

$$G_d(x) = -k_B T \ln Q_d(x). \quad (\text{S8})$$

Fig. S6 shows a flowchart how we obtain the free energy landscape as a function of unfolded protein contour length  $L_p$  at a certain trap distance  $d$ . The individual steps are described in detail below.

### 5.1 Deconvolution procedure

Since iterative deconvolution procedures that were previously used by us and others [3, 13] tended to show poor convergence in our case, we adapted a deconvolution procedure introduced by Hinczewski et al. [14]. In practice, this

method optimizes the estimate for the deconvolved energy landscape by fitting its convolved form to the measured bead-location distribution.

We generated an estimate  $\hat{G}_d(x)$  for  $G_d(x)$  using a cubic spline on typically 20-30 equally spaced base points denoted by  $\tilde{x}_i$ . Deconvolution was performed by numerically minimizing the function

$$\mathcal{O} = \left\langle \left| \ln \left( \exp \left( -\frac{\hat{G}_d(x)}{k_B T} \right) \otimes \Psi_x(x) \right) - \ln P_d(x) \right| \right\rangle_x + \lambda \sum_i \left( \frac{1}{k_B T} \frac{d^2 \hat{G}_d(x)}{dx^2} \Big|_{\tilde{x}_i} \right)^2, \quad (\text{S9})$$

where the second part is a penalty function to ensure solutions are smooth.  $\langle \cdot \rangle_x$  denotes an average over  $x$  and  $\lambda$  is a parameter, which allows us to tune the fidelity of the deconvolution. Higher values of  $\lambda$  smooth the resulting deconvolved energy landscapes. In this study, we chose values of  $\lambda$  that did not significantly smooth the resulting energy landscapes. Figure S7 shows an example for a deconvolution with different values of  $\lambda$  and otherwise unaltered parameters. While for  $\lambda > 10^{-3} \text{ nm}^2$  the blurring of the energy landscapes (top) is clearly visible, for smaller  $\lambda$  this deterioration is negligible. Note that the blurring is also clearly reflected in the residuals of measured bead distributions vs. re-convolved distributions. We typically used  $\lambda$  values of  $10^{-4} - 10^{-5} \text{ nm}^2$  where blurring effects were negligible.

We used a quasi-Newton optimizer with Moré-Hebden steps [15] built into Igor Pro (Wavemetrics) with  $\hat{G}_d(x) = -k_B T \ln P_d(x)$  as starting value to vary the interpolation points in order to get an optimal estimate for  $G_d(x)$ .

It is important to note that we are unable to resolve features in the energy landscape smaller than the spacing of the spline base points, which corresponds to about  $\pm 2 \text{ nm}$  in contour length. A higher number of spline base points and a lower value of  $\lambda$  tend to lead to overfitting of the data and amplification of random noise. To minimize the effects of these possible factors, we used a bootstrapping procedure (see section 5.7) to average out these random fluctuations.

The difference between a re-convolved form of the deconvolved probability distribution and the measured distribution was always close to zero, as exemplified in Fig. 2A. We used these residuals as quality control for the performance of our deconvolution algorithm.

## 5.2 Conversion from $G_d(x)$ to $\tilde{G}_d(L_p)$

To convert an energy landscape  $G_d(x)$  given as a function of bead deflection  $x$  into a landscape as a function of unfolded contour length  $L_p$  we first calculated the corresponding probability distribution  $Q_d(x)$  (eq. (S8)) and used eq. (S10) to change variables.

$$\tilde{Q}_d(L_p) = \left| \frac{1}{\zeta'(\zeta^{-1}(L_p))} \right| Q_d(\zeta^{-1}(L_p)) \quad (\text{S10})$$

where  $\zeta(x) = L_p(x)$ ,  $\zeta'$  is the derivative of  $\zeta$  and  $\zeta^{-1}$  is its inverse.

### 5.3 Transformation of energy landscapes to different trap distances/forces

We define the full energy function of the system when held at a constant trap separation  $d$  as

$$\begin{aligned}\mathcal{H}_d(x, L_p) &= \tilde{G}_o(L_p) + \mathcal{H}_d^{\text{mech}}(x, L_p) \\ &= \tilde{G}_o(L_p) + \frac{1}{2}k_c x^2 + \int_0^{d-x} F_{\text{construct}}(L_p, \xi) d\xi.\end{aligned}\quad (\text{S11})$$

Here,  $\mathcal{H}_d^{\text{mech}}$  describes the mechanical energy involved in stretching the linkers and deflecting the beads from their trap centers,  $x = x_1 + x_2$  is the sum of the two bead deflections,  $k_c = \left(\frac{1}{k_1} + \frac{1}{k_2}\right)^{-1}$  is the combined trap stiffness and  $L_p$  is the unfolded protein contour length.  $\tilde{G}_o(L_p)$  is the energy landscape of the protein as a function of its unfolded protein contour length  $L_p$  under non-denaturing (i.e. zero force or  $d = 0$  conditions).

It should be noted that the description of the two individual traps as one combined trap is an approximation that is valid when the two traps are close to equilibrium, when the energy stored in the traps  $\frac{1}{2}k_1 x_1^2 + \frac{1}{2}k_2 x_2^2$  is equal to  $\frac{1}{2}k_c x^2$ . This assumption is generally true for all equilibrium measurements and analyses based on equilibrium terms, i.e. all long-timescale passive mode measurements and free energy/deconvolution considerations in this study. However, when considering non-equilibrium phenomena such as the simulation of stretch-relax cycles (see section 6), this approximation is not generally valid and is only justified at time scales longer than the equilibration time of the mechanical components of the system. Since the linker components have relatively fast equilibration times (see [12]), the overall time scale for equilibration is dominated by the comparatively slow relaxation time of the beads (on the order of  $10 \mu\text{s}$  [10]). Hence, eq. (S11) is also valid for the non-equilibrium description of phenomena slower than  $\approx 10 \mu\text{s}$ .

For long observation times, the bead dynamics can be integrated out to obtain the effective energy landscape as a function of unfolded protein contour length  $L_p$  at a trap distance  $d$ :

$$\begin{aligned}\tilde{G}_d(L_p) &= -k_B T \ln \int dx \exp\left(-\frac{\mathcal{H}_d(x, L_p)}{k_B T}\right) \\ &= \tilde{G}_o(L_p) - k_B T \ln \int dx \exp\left(-\frac{\mathcal{H}_d^{\text{mech}}(x, L_p)}{k_B T}\right) \\ &= \tilde{G}_o(L_p) - k_B T \ln \int dx \exp\left(-\frac{\frac{1}{2}k_c x^2 + \int_0^{d-x} F_{\text{construct}}(L_p, \xi) d\xi}{k_B T}\right)\end{aligned}\quad (\text{S12})$$

We can use eq. (S12) to transform any energy landscape  $\tilde{G}_{d_1}(L_p)$  at trap distance  $d_1$  to another trap distance  $d_2$  by first transforming to zero force to obtain  $\tilde{G}_o(L_p)$ , and then reapplying eq. (S12) to obtain the wanted  $\tilde{G}_{d_2}(L_p)$ .

### 5.4 Point spread functions

When a tethered molecule is held in a passive-mode optical trap, the distribution of the position noise of the beads can be sufficiently described with a skewed gaussian function [2]. Previous work has shown that we can use statistical mechanics to calculate the exact shape of this noise function from first principles [10]. Here we used this technique to

calculate the mean, standard deviation and skewness, i.e. the first three central moments of this distribution in the case of a fully folded protein and a fully unfolded protein. We then linearly interpolated between these values to calculate the point spread function  $\Psi_a(x)$ . The importance of using a non-constant point spread function has been highlighted elsewhere [3].

When the protein is (and stays) in the folded state ( $L_p = 0$ ), the Hamiltonian for the bead position fluctuations when the trap distance is  $d$  is given by (see eq. (S11))

$$\mathcal{H}_d(x) = \mathcal{H}_d^{\text{mech}}(x, L_p = 0) = \frac{1}{2}k_c x^2 + \int_0^{d-x} F_{\text{construct}}(L_p = 0, \xi) d\xi. \quad (\text{S13})$$

For the unfolded state, we used eq. (S13) with  $L_p = L_{p,\text{unf}}$ , which was obtained from stretch-and-relax cycles.

The  $n$ -th moment of the distribution of  $x$  is then given by [10]

$$\mu_n = \langle x^n \rangle = \frac{\int_{-\infty}^{\infty} dx x^n \exp\left(-\frac{\mathcal{H}_d(x)}{k_B T}\right)}{\int_{-\infty}^{\infty} dx \exp\left(-\frac{\mathcal{H}_d(x)}{k_B T}\right)}. \quad (\text{S14})$$

The mean  $\mu = \mu_1$ , standard deviation  $\sigma = \sqrt{\mu_2 - \mu_1^2}$  and skewness  $\gamma = (\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3) / \sigma^3$  can now be calculated from eq. (S14) for both the folded state (N) and unfolded state (U). We then use this information to calculate the point spread function for the folded state and the unfolded state:

$$\begin{aligned} \Psi_{\mu_N}(x) &= \Gamma(x, \mu_N, \sigma_N, \gamma_N) \\ \Psi_{\mu_U}(x) &= \Gamma(x, \mu_U, \sigma_U, \gamma_U) \end{aligned} \quad (\text{S15})$$

and interpolate the values in between  $\mu_N$  and  $\mu_U$  by linearly interpolating  $\mu$ ,  $\sigma$  and  $\gamma$  between the corresponding N and U values to obtain the full description of the point spread function  $\Psi_d(x)$ . The function  $\Gamma(x, \mu, \sigma, \gamma)$  is a skewed gaussian with mean  $\mu$ , standard deviation  $\sigma$  and skewness  $\gamma$ .

It is important to note that the values calculated using the method described above are only valid for measurements at infinite bandwidth. We therefore used calibration measurements of constructs where beads were tethered only with DNA linkers and no protein was present and compared these 100 kHz measurements with the theoretical infinite bandwidth expectation to obtain calibration factors. The dependence of the measured distribution moments on bandwidth is also discussed in [10].

## 5.5 WHAM

To combine energy landscapes recovered from experiments at differing trap distances we used the method of weighted histograms (WHAM) [16]. The method combines energy landscapes by weighting them with the amount of data that was fed into the deconvolution algorithm. We first transformed individual energy landscapes  $\tilde{G}_{d_i}^i(L_p)$  at trap distances  $d_i$  to a common distance  $d$  (eq. (S12)) and then calculated the combined energy landscape  $\tilde{G}_d^*(L_p)$  as described in ref. [14].

## 5.6 Combination of energy landscapes

The energy landscapes for individual molecules scattered around a common center due to uncertainties in estimating the exact length of the dsDNA linkers. We therefore chose to align the energy landscapes with respect to a state that is populated under all measured conditions (i.e. the intermediate I at  $L_p \approx 40$  nm for Vim2B-N and the folded state N for Vim2B-C).

Fig. S8 shows the results of the deconvolution procedure for a number of molecules of Vim2B-N and Vim2B-C after alignment. The molecules of a kind all show common features.

## 5.7 Error estimation

Errors for the energy landscapes of individual molecules were estimated by bootstrapping from each individual trace. To this end, each trace was split into typically 15 pieces of same length, and a new trace was generated by drawing from the pool of pieces with replacement and the analysis repeated with this newly generated trace [14]. The standard deviation of the resulting energy landscapes from repeating this procedure  $> 100$  times was propagated as additional relative weight into the weighted histogram procedure (see section 5.5). The errors of the final combined energy landscape of all molecules (see section 5.6) was also estimated by bootstrapping from each individual landscape.

## 5.8 Local stability plots

The energy landscape  $\tilde{G}_o^*(L_p)$  describes the energetic cost that is needed to unfold the coiled coil from the native state up to a contour length  $L_p$ . Stable regions in the coiled coil exhibit a steep slope of  $\tilde{G}_o^*(L_p)$  while for unstable regions the slope is flat. To be able to assess the specific stability in certain regions of the coiled coil and to be able to directly compare the energy profiles, we therefore determined local stability graphs by taking the derivative  $\frac{\partial \tilde{G}_o^*(L_p)}{\partial L_p}$  of the energy landscapes (Fig. 3).

# 6 Simulation

To verify that the energy landscapes we obtained from deconvolution are able to reproduce measured results, we conducted simulations of stretch-and relax experiments.

In these simulations we assumed that the dsDNA linkers as well as the unfolded polypeptide are in equilibrium with the beads at all times. In contrast, due to their size, the beads experience a much higher drag and hence are the slowest mechanical components in the system [12]. We therefore based our simulations on a reduced system where we only followed the coordinates of bead deflection and the diffusion in the protein potential.

We based the simulations on the full Hamiltonian of the system  $\mathcal{H}_d(x, L_p)$  (eq. (S11)). At each time step  $\Delta t$ , we updated  $x$  and  $L_p$  based on their stochastic equation of motion ( $x$  is the sum of the two bead deflections  $x_1 + x_2$ ,  $L_p$  is the unfolded

protein contour length):

$$\begin{aligned}\Delta x &= \frac{\Delta t}{\gamma_b} \left( -\frac{\partial \mathcal{H}_d}{\partial x}(x, L_p) k_B T + \sqrt{2k_B T \frac{\gamma_b}{\Delta t}} \varepsilon(t) \right) \\ \Delta L_p &= \frac{\Delta t}{\gamma_p} \left( -\frac{\partial \mathcal{H}_d}{\partial L_p}(x, L_p) k_B T + \sqrt{2k_B T \frac{\gamma_p}{\Delta t}} \varepsilon(t) \right)\end{aligned}\quad (\text{S16})$$

The trap distance  $d$  was hereby increased or decreased linearly, following the pulling protocol.  $\varepsilon(t)$  is a random number from a generator that produces an uncorrelated series with standard deviation  $\sigma = 1$ . The friction coefficient in the protein potential  $\gamma_p$  was chosen such that the transition rates generated from the simulation matched those that we measured experimentally ( $\gamma_p \approx 3 \times 10^{-5}$  pN s nm<sup>-1</sup>). We assumed that the friction of the coordinate  $x$  is dominated by Stokes friction of the largest components of the system, the beads. Its value was therefore chosen to be  $\gamma_b = 6\pi\eta R$ , where  $R = 500$  nm is the radius of a bead and  $\eta = 1$  cP =  $10^{-9}$  pN s nm<sup>-2</sup>.

Force-distance curves were then created by simulating eq. (S16) with a typical time step of  $10^{-7}$  s and sampling at 20 kHz while linearly increasing or decreasing the trap distance  $d$ . Note that this sampling rate equals to a data point spacing of 50  $\mu$ s, a time scale at which all mechanical components of the system are equilibrated (see section 5.3), justifying the use of the approximative eq. (S11).

Examples for stretch and relax cycles based on the measured energy landscapes of Vim2B-N and Vim2B-C (Fig. 2C) are shown in Figure S3.

## References

- [1] Ciro Cecconi, Elizabeth A Shank, Frederick W Dahlquist, Susan Marqusee, and Carlos Bustamante. Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers. *Eur Biophys J*, 37(6):729–738, July 2008.
- [2] Johannes Stigler, Fabian Ziegler, Anja Gieseke, J Christof M Gebhardt, and Matthias Rief. The complex folding network of single calmodulin molecules. *Science*, 334(6055):512–516, October 2011.
- [3] J Christof M Gebhardt, Thomas Bornschlöggl, and Matthias Rief. Full distance-resolved folding energy landscape of one single protein molecule. *Proceedings of the National Academy of Sciences of the United States of America*, 107(5):2013–2018, February 2010.
- [4] SF Tolić-Nørrelykke, Erik Schäffer, J Howard, FS Pavone, F Jülicher, and Henrik Flyvbjerg. Calibration of optical tweezers with positional detection in the back focal plane. *Review of Scientific Instruments*, 77:103101, 2006.
- [5] M D Wang, H Yin, R Landick, J Gelles, and Steven M Block. Stretching DNA with optical tweezers. *Biophysical Journal*, 72(3):1335–1346, March 1997.
- [6] Lorenz Rognoni, Johannes Stigler, Benjamin Pelz, Jari Yläanne, and Matthias Rief. Dynamic force sensing of filamin revealed in single-molecule experiments. *Proceedings of the National Academy of Sciences of the United States of America*, 109(48):19679–19684, November 2012.



- [7] Carlos Bustamante, J F Marko, E D Siggia, and Steven B Smith. Entropic elasticity of lambda-phage DNA. *Science*, 265(5178):1599–1600, September 1994.
- [8] Thomas Bornschlöggl and Matthias Rief. Single molecule unzipping of coiled coils: sequence resolved stability profiles. *Physical Review Letters*, 96(11):118102, 2006.
- [9] Hendrik Dietz and Matthias Rief. Protein structure by mechanical triangulation. *Proceedings of the National Academy of Sciences of the United States of America*, 103(5):1244–1247, January 2006.
- [10] Gabriel Zoldák, Johannes Stigler, Benjamin Pelz, Hongbin Li, and Matthias Rief. Ultrafast folding kinetics and cooperativity of villin headpiece in single-molecule force spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 110(45):18156–18161, November 2013.
- [11] Johannes Stigler and Matthias Rief. Hidden markov analysis of trajectories in single-molecule experiments and the effects of missed events. *Chemphyschem : a European journal of chemical physics and physical chemistry*, 13(4):1079–1086, March 2012.
- [12] M Manosas, Jin-Der Wen, Pan T X Li, Steven B Smith, Carlos Bustamante, Ignacio Tinoco, and Felix Ritort. Force unfolding kinetics of RNA using optical tweezers. II. Modeling experiments. *Biophysical Journal*, 92(9):3010–3021, May 2007.
- [13] Michael T Woodside, PC Anthony, William M Behnke-Parks, K Larizadeh, Daniel Herschlag, and Steven M Block. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science*, 314(5801):1001, 2006.
- [14] Michael Hinczewski, J Christof M Gebhardt, Matthias Rief, and D Thirumalai. From mechanical folding trajectories to intrinsic energy landscapes of biopolymers. *Proceedings of the National Academy of Sciences of the United States of America*, 110(12):4500–4505, March 2013.
- [15] Jr J E Dennis and Robert B Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. SIAM, December 1996.
- [16] AM Ferrenberg and RH Swendsen. Optimized Monte Carlo data analysis. *Physical Review Letters*, 63(12):1195–1198, September 1989.
- [17] B Tripet, K Wagschal, P Lavigne, C T Mant, and R S Hodges. Effects of side-chain characteristics on stability and oligomerization state of a de novo-designed model coiled-coil: 20 amino acid substitutions in position "d". *Journal of molecular biology*, 300(2):377–402, July 2000.
- [18] K Wagschal, B Tripet, P Lavigne, C Mant, and R S Hodges. The role of position a in determining the stability and oligomerization state of alpha-helical coiled coils: 20 amino acid stability coefficients in the hydrophobic core of proteins. *Protein science : a publication of the Protein Society*, 8(11):2312–2329, November 1999.
- [19] Thomasin A Smith, Sergei V Strelkov, Peter Burkhard, Ueli Aebi, and David A D Parry. Sequence comparisons of intermediate filament chains: evidence of a unique functional/structural role for coiled-coil segment 1A and linker L1. *J Struct Biol*, 137(1-2):128–145, January 2002.
- [20] Harald Herrmann, Sergei V Strelkov, Peter Burkhard, and Ueli Aebi. Intermediate filaments: primary determinants of cell architecture and plasticity. *The Journal of clinical investigation*, 119(7):1772–1783, July 2009.

- [21] Sergei V Strelkov, Harald Herrmann, Norbert Geisler, Tatjana Wedig, Ralf Zimbelmann, Ueli Aebi, and Peter Burkhard. Conserved segments 1A and 2B of the intermediate filament dimer: their atomic structures and role in filament assembly. *EMBO J*, 21(6):1255–1266, March 2002.
- [22] Sergei V Strelkov, Harald Herrmann, and Ueli Aebi. Molecular architecture of intermediate filaments. *Bioessays*, 25(3):243–251, 2003.
- [23] Sergei V Strelkov, Jens Schumacher, Peter Burkhard, Ueli Aebi, and Harald Herrmann. Crystal structure of the human lamin A coil 2B dimer: implications for the head-to-tail association of nuclear lamins. *Journal of molecular biology*, 343(4):1067–1080, October 2004.
- [24] Michael Schlierf, Felix Berkemeier, and Matthias Rief. Direct observation of active protein folding using lock-in force spectroscopy. *Biophysical Journal*, 93(11):3989–3998, 2007.
- [25] Leoni Oberbarnscheidt, Richard Janissen, and Filipp Oesterhelt. Direct and Model Free Calculation of Force-Dependent Dissociation Rates from Force Spectroscopic Data. *Biophysical Journal*, 97(9):L19–L21, 2009.

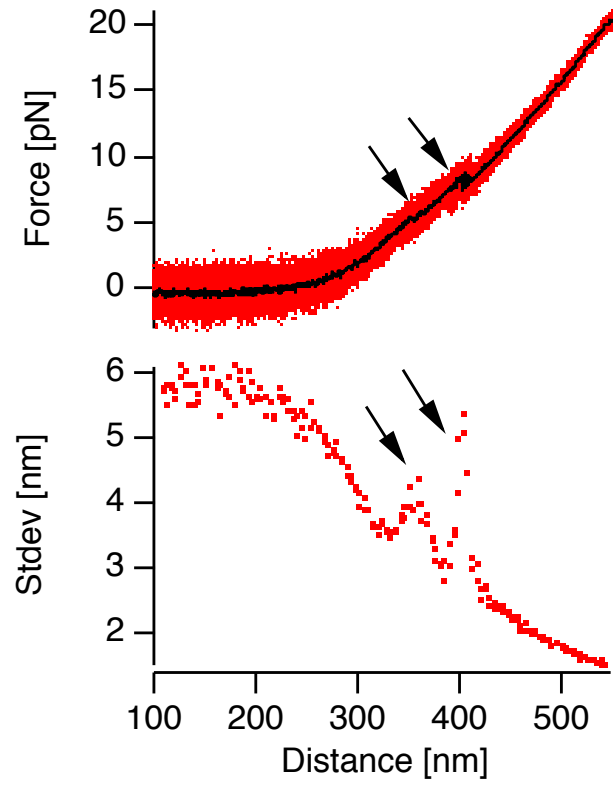


Figure S1: Top: Stretch curve for the Vim2B-N at 50 nm/s pulling speed (about 4–5 pN/nm loading rate) at 20 kHz sampling rate (red points) and low-pass filtered (black line). The two transitions (arrows in top graph) cause an increase in noise and are clearly seen in the corresponding trap distance dependent standard deviations (arrows in lower graph).

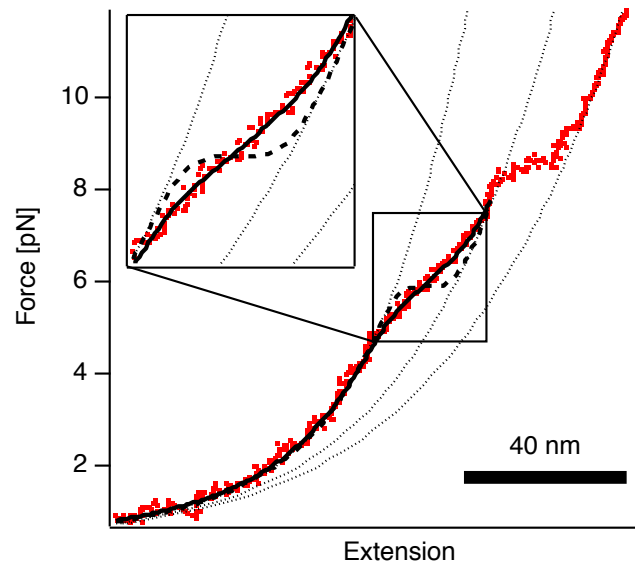


Figure S2: Averaged stretch-and-relax cycles of Vim2B-N with fits to the hump-like first transition according to eq. (S5). Dashed line: two-state transition ( $N = 1$ ), continuous line: three-state transition ( $N = 2$ ). Models for  $N > 2$  are indistinguishable from the case  $N = 2$ .

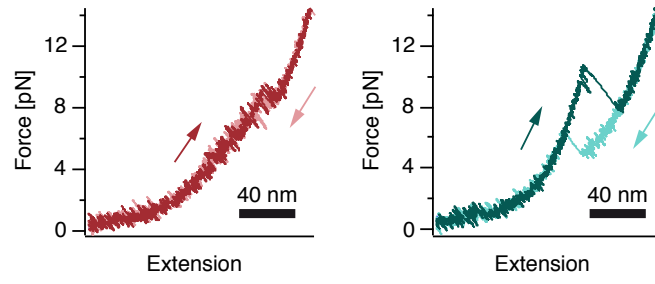


Figure S3: Brownian dynamics simulation of stretch and relax cycles based on the measured energy landscapes for Vim2B-N (left) and Vim2B-C (right).

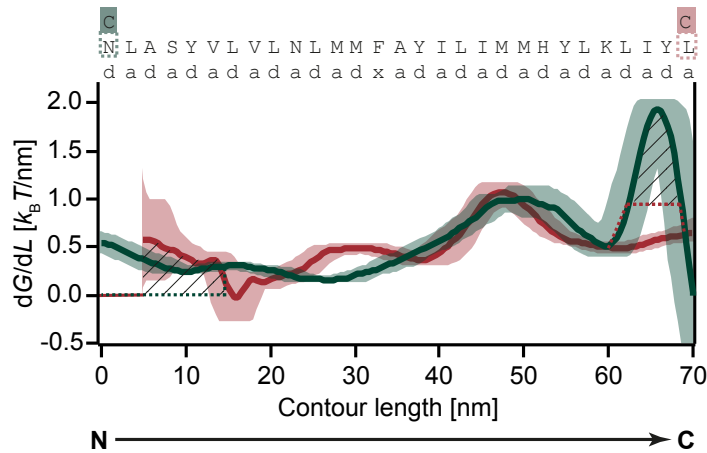


Figure S4: Modified stability profiles for Vim2B-C (green, dotted) and Vim2B-N (red, dotted) to correct for over- and underestimation of the terminal seed energies, respectively. As the first 5 nm of the stability profile of Vim2B-N are not shown as errors exceed 100 %. The stability of Vim2B-N was set to  $0 k_B T$  for this range. Hatched areas denote nucleation seed energy for Vim2B-C (left) and Vim2B-N (right) after correction:

The two different sites of cross-linking introduce an asymmetry into the two constructs. The cross-linking cysteine of Vim2B-C replaces the N-terminal asparagine whereas the cross-link of Vim2B-N replaces the C-terminal leucine. Mutational studies have shown that introducing an asparagine instead of an alanine at a *d*-position destabilizes the coiled-coil by  $1.0 k_B T$  and introducing a leucine for an alanine at an *a*-position stabilizes the coil by  $5.9 k_B T$  [17, 18, 8]. The Vim2B-N and Vim2B-C stability profiles were adjusted such that the free energy difference of the two substitutions of  $6.9 k_B T$  was compensated and the nucleation seed energy of the two constructs are similar. Note that in this representation the nucleation seed of Vim2B-C is longer than that of Vim2B-N.

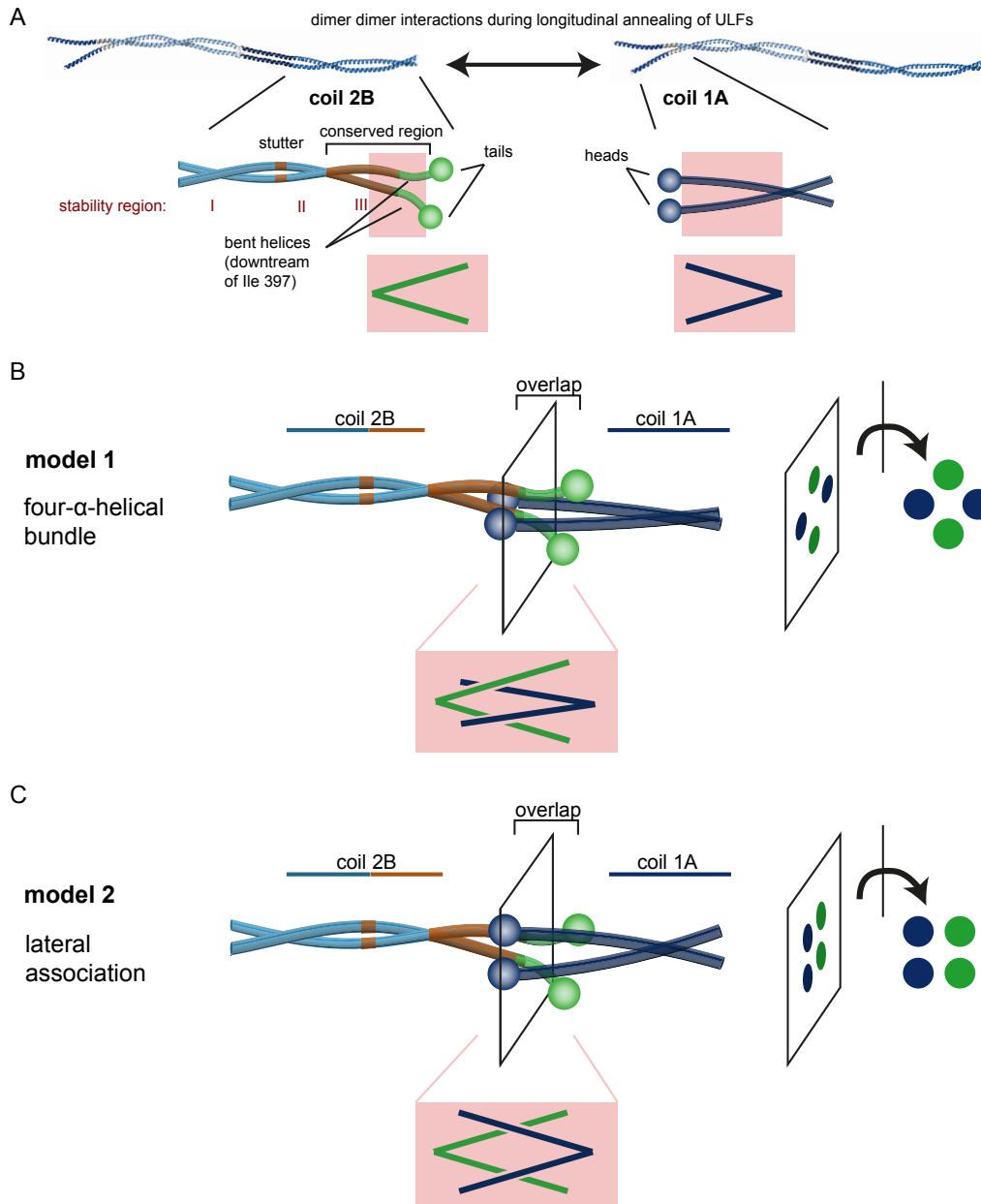


Figure S5: Two different models for the association of adjacent dimers during longitudinal annealing of ULFs. (A) Schematic representation of the annealing ends in coil 2B and coil 1A. The non- $\alpha$ -helical tail domains are connected to the end of coil 2B where the two helices bend away from the coiled coil axis. Coil 1A terminates in the non- $\alpha$ -helical head domains. The parts highlighted in pink are schematically depicted as triangles. (B) Model 1 assumes that the bent helices of coil 2B interlock with the two helices of coil 1A and build a four- $\alpha$ -helical bundle [19, 20, 21, 22]. (C) In model 2 the dimers associate laterally and are stabilized by electrostatic interactions between polar residues [23].

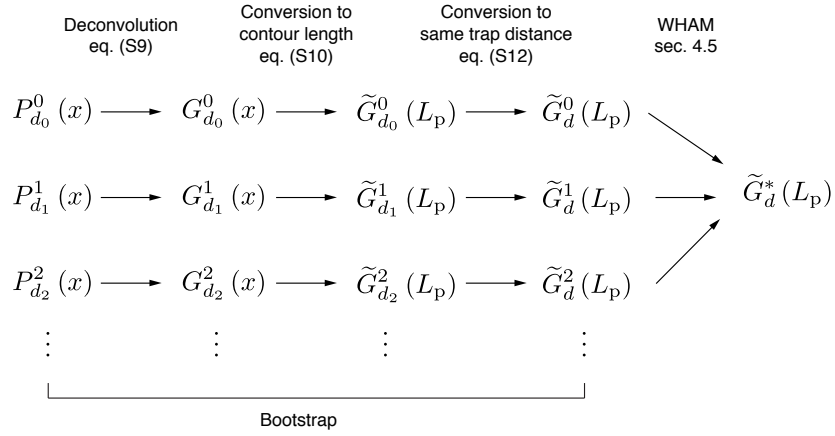


Figure S6: Flowchart for the deconvolution procedure used in this work. For each measured trap distance  $d_n$ , the bead position distribution  $P_{d_n}^n$  was determined, deconvolved, and transformed to contour length space. The resulting energy landscapes were transformed to a common trap distance  $d$  and combined using WHAM.



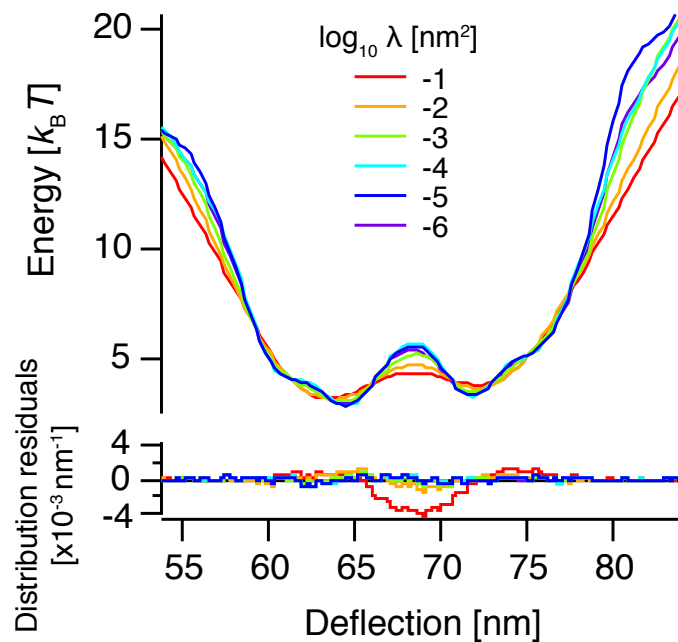


Figure S7: Result of deconvolution for different values of the parameter  $\lambda$  in eq. (S9). Top: resulting energy landscapes from deconvolution. Bottom: Residuals of measured bead distributions compared to re-convolved energy landscapes. While high values of  $\lambda$  deteriorate the performance of the algorithm, values  $\leq 10^{-4} \text{ nm}^2$  do not affect the results and the resulting deconvolved energy landscapes are virtually indistinguishable.

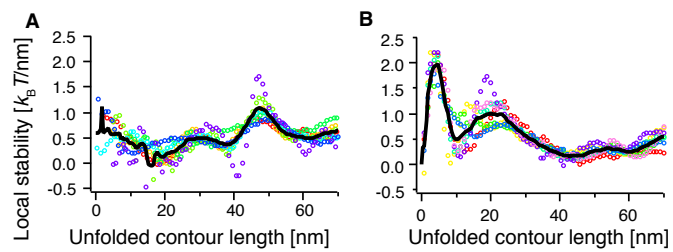


Figure S8: Result of the deconvolution procedure for different molecules of Vim2B-N (A) and Vim2B-C (B) in color. Shown are local stability plots, i.e. derivatives of the energy landscapes  $\frac{\partial \tilde{G}_o^*(L_p)}{\partial L_p}$  (see SI section 5.8). While there are visible variations, the main features of the differential energy landscape are present in all molecules. The black line is an average over all molecules.

	<b>Vim2B-N</b>	<b>Vim2B-C</b>
$\Delta L_{N-I_1}$ (nm)	$35.4 \pm 2.0^*$	—
$\Delta L_{N-I_2}$ (nm)	—	$14.2 \pm 2.0$
$\Delta L_{N-U}$ (nm)	$65.4 \pm 2.0^*$	$69.5 \pm 2.0$
$-\Delta \tilde{G}_{0,N-I_1}$ ( $k_B T$ )	$12.9 \pm 1.3$	—
$-\Delta \tilde{G}_{0,N-I_2}$ ( $k_B T$ )	—	$12.9 \pm 1.3$
$-\Delta \tilde{G}_{0,N-U}$ ( $k_B T$ )	$35.2 \pm 3.5$	$36.8 \pm 3.7$
$-\Delta \tilde{G}_{0,N-U}$ ( $k_B T$ ) <sup>†</sup>	$32.7 \pm 3.5^\dagger$	$40.3 \pm 3.7^\dagger$

Table S1: Measured contour length differences as determined from fits to stretch-relax cycles and free energy differences determined by conventional data analysis for Vim2B-N and Vim2B-C.

\* The first N-terminal 5 nm of Vim2B-N are presumably unfolded at very low pulling forces and hence unresolved in our assay. For comparison between Vim2B-N and Vim2B-C in Fig. 2C and Fig. 3 we therefore aligned the lengths such that the state U lies at about 70 nm (see section 4.1).

<sup>†</sup> Energy values obtained by deconvolution, at a measured (not-aligned) contour length of 65.4 nm for Vim2B-N and 69.5 nm for Vim2B-C; difference from the values at 0 nm.

<b>Molecule</b>	<b>Transition</b>	<b>Transition state distance from originating state [nm]</b>
Vim2B-N	N $\rightarrow$ I <sub>1</sub>	n/a
	I <sub>1</sub> $\rightarrow$ N	n/a
	U $\rightarrow$ I <sub>1</sub>	11.4 $\pm$ 1.0
	I <sub>1</sub> $\rightarrow$ U	15.4 $\pm$ 0.6
Vim2B-C	N $\rightarrow$ I <sub>2</sub>	3.1 $\pm$ 0.7
	I <sub>2</sub> $\rightarrow$ N	13.3 $\pm$ 1.4
	U $\rightarrow$ I <sub>2</sub> *	24.0 $\pm$ 1.3
	I <sub>2</sub> $\rightarrow$ U *	16.0 $\pm$ 3.6

Table S2: Distances to the transition state in units of unfolded contour length. The distances were obtained by fitting the force-dependent transition rates to an energetic model that takes into account the free energy involved in stretching the mechanical components of the system from the originating state up to the transition state [24], as described in [2]. For transitions without asterisks, the force-dependent rates were determined from Hidden Markov analysis of passive-mode traces (see text).

\* Since the kinetics of the U  $\leftrightarrow$  I<sub>2</sub> transition were relatively slow in passive mode, we extracted the force-dependent rate constants for these transitions from force-extension traces following a method by Oberbarnscheidt et al. [25].