Supplemental Data

Full Reconstruction of a Vectorial Protein Folding Pathway by Atomic Force Microscopy and Molecular Dynamics Simulations

Whasil Lee, Xiancheng Zeng, Huan-Xiang Zhou, Vann Bennett, Weitao Yang & Piotr E. Marszalek

Supplemental Note

MATERIALS & METHODS

Protein design, cloning, and expression

Our chimeric polyprotein NI6C-I27 construct consisted of NI6C, a consensus Ankyrin repeat (AR) protein, and three I27 domains each flanking the N- and C-termini [i.e., $(I27)_3$-NI6C-$(I27)_3$]. The sequences of the capping and internal repeats of NI6C were adapted from(1), and are listed in Supplemental Fig. S1*D*. The DNA sequence of NI6C was synthesized by Genescript (Piscataway, NJ). The NI6C gene was inserted into the Poly-I27 pRSETa vector [a kind gift from Jane Clark(2)] using KpnI and NheI restriction sites, and the STOP codon was added before the MluI restriction site. The engineered plasmids were transformed into E. coli C41 (DE3) (Lucigen, #89027), and the NI6C-I27 protein was expressed for 10 hours using IPTG induction. The harvested cells were lysed and purified by using a nickel affinity column (GE Healthcare, #17-5268-02) followed by size exclusion HPLC. Purified protein was checked by SDS-PAGE gel electrophoresis and CD spectroscopy.

AFM-based single molecule force spectroscopy

Purified NI6C-I27 protein was dialyzed in a buffer with 150 mM NaCl, 2 mM TCEP (Thermo Scientific, #77720), and 20 mM Tris (pH 8.0). 50 µl of a diluted protein solution (1-5 µg/ml) was incubated on a substrate for 20 minutes, rinsed once by adding and removing 50 µl of the buffer to remove protein molecules that did not adhere to the substrate surface. Both nickel-NTA functionalized glass substrates(3-5) and clean glass substrates were used for AFM experiments. Force-extension curves of NI6C-I27 measured on both kinds of substrates were similar. All stretching measurements were carried out on custom-built AFM instruments(3,6) equipped with an AFM detector head (Veeco Metrology Group), and high-resolution piezoelectric stages with vertical resolution of 0.1 nm (Physik Instrumente). The spring constant, k_c of each cantilever was calibrated in solution using the energy equipartition theorem as described by Florin *et al.*(7). Protein molecules were picked up by gently touching the substrate with the AFM cantilever tip. All force-extension measurements were performed using Biolever cantilevers (OBL from Veeco, $k_c \approx 6$ pN/nm) at pulling speeds between 5 and 100 nm/s, at room temperature. The force peaks in the force-extension curves were fitted to the Worm-like-chain (WLC) model(8).

SMD simulations using a coarse-grained model

The consensus AR protein NIC6 contains an N-terminal, a C-terminal, and 6 internal repeats, and is composed of 253 amino acids. In the AFM experiments, the AR protein was flanked by I27 modules. To reduce the system size and focus the modeling on NIC6, we used only 6 additional glycine residues as the flanking sequence on each side. Although attempts have been made to study similar systems with all-atom details(9), the sampling convergence in all-atom simulations remains a major concern when it is used to describe large-amplitude protein motions. Therefore, various approaches have been developed to model proteins at a coarse-grained (CG) level in order to efficiently deal with the large-amplitude motions in protein folding(10-14). We adapted a CG model developed by Clementi *et al.*(12) to simulate the AFM-controlled protein unfolding and refolding. In this CG model each amino acid was represented by the C_α atoms only, and the interactions between non-neighboring residues were based on native and non-native contacts. The energy function of the CG model is:

$$E(\Gamma,\Gamma_0)=\sum_{\text{bonds}}K_r(r-r_0)^2+\sum_{\text{angles}}K_\theta(\theta-\theta_0)^2$$
$$+\sum_{\text{dihedral}}K_\phi^{(n)}[1+\cos(n\times(\phi-\phi_0))]$$
$$+\sum_{i<j-3}^{N}\left\{\varepsilon(i,j)\left[5\left(\frac{r_{0,ij}}{r_{ij}}\right)^{12}-6\left(\frac{r_{0,ij}}{r_{ij}}\right)^{10}\right]+\varepsilon_2(i,j)\left(\frac{r_{0,ij}}{r_{ij}}\right)^{12}\right\},$$

The first three summations are over bonds, bond angles, and dihedral angles; the last summation includes the non-local native interactions and the repulsive non-native contacts. The subscript "0" denotes the properties from the native crystal structure; Γ represents the current structure; K_r is the harmonic constant for the bonds; r is the bond length between two C_α atoms; K_θ is the harmonic constant for the bond angles; θ is the angle formed by three consecutive C_α atoms; $K_\phi^{(n)}$ is the harmonic constant for dihedral angles with multiplicity of n; ϕ is the dihedral angle formed by four consecutive C_α atoms; $\varepsilon(i,j)$ and $\varepsilon_2(i,j)$ are the energetic parameters for native and non-native contacts, respectively; r_{ij} is the distance between residue i and j. Based on this force field, we performed SMD(15) simulation using the GROMACS package(16).

Building the structure of NI6C from NI3C (PDB: 2QYJ).
The largest consensus AR protein with determined crystal structure was NI3C. To carry out SMD simulations of the NI6C protein used in the AFM measurements, we needed to extrapolate the structure of NI6C from NI3C. Since our CG model employed a structure-based energy function, building an accurate initial structure was critical. Based on the X-ray crystallography of NI3C (PDB code: 2QYJ)(17), we used a root mean square displacement (RMSD) fitting procedure to build the structure of NI6C. In the fitting procedure, we divided the protein NI3C into 5 domains as listed in Supplemental Table S1 and performed the fitting to minimize the RMSD of the backbone atoms (C_α, C, O, N) among repeats 1, 2, and 3. A step-by-step description of the procedure is shown in Supplemental Fig. S4*B* and the final structure built through the RMSD fitting is shown in Supplemental Fig. S4*A* with overlaps in different colors.

Expected length of a fully stretched Ankyrin repeat.
The expected length of a fully stretched internal repeat (33 amino acids) is 12 nm (33 × 0.365 nm per amino acid)(18). The expected contour length of NI6C (253 amino acids) is 92 nm (253 × 0.365 nm per amino acid), while the length of folded NI6C is ~8 nm. Therefore, the increment or contraction of the contour length due to unfolding or refolding of one repeat in AFM force traces is expected to be around 11 nm (12 nm – 1 nm) when the force peaks are fitted to the WLC model. We should point out that the contour length in the CG model of NI6C is shorter than that in all-atom models or in the AFM measurements. A fully stretched polypeptide chain in the all-atom representation has the dihedral angles N-C-C_α-N near 180° to accommodate the tension in the backbone; however, the equilibrium distance between two neighboring C_α atoms is defined by the crystal structure, and subjected to the harmonic bond restraint $K_r(r-r_0)^2$ when stretched.

REFERENCES

1. Wetzel, S. K., Settanni, G., Kenig, M., Binz, H. K., and Plückthun, A. (2008) *J. Mol. Biol.* **376**, 241-257
2. Steward, A., Toca-Herrera, J. L., and Clarke, J. (2002) *Protein Sci.* **11**, 2179-2183
3. Lee, G., Abdi, K., Jiang, Y., Michaely, P., Bennett, V., and Marszalek, P. E. (2006) *Nature* **440**, 246-249
4. Schmid, E. L., Keller, T. A., Dienes, Z., and Vogel, H. (1997) *Anal. Chem.* **69**, 1979-1985

5. Schmitt, L., Ludwig, M., Gaub, H. E., and Tampe, R. (2000) *Biophys. J.* **78**, 3275-3285
6. Oberhauser, A. F., Marszalek, P. E., Erickson, H. P., and Fernandez, J. M. (1998) *Nature* **393**, 181-185
7. Florin, E. L., Rief, M., Lehmann, H., Ludwig, M., Dornmair, C., Moy, V. T., and Gaub, H. E. (1995) *Biosens. Bioelectron.* **10**, 895-901
8. Bustamante, C., Marko, J. F., Siggia, E. D., and Smith, S. (1994) *Science* **265**, 1599-1600
9. Sotomayor, M., Corey, D. P., and Schulten, K. (2005) *Structure* **13**, 669-682
10. Go, N. (1983) *Annu. Rev. Biophys. Bioeng.* **12**, 183-210
11. Liwo, A., Oldziej, S., Pincus, M. R., Wawak, R. J., Rackovsky, S., and Scheraga, H. A. (1997) *J. Comput. Chem.* **18**, 849-873
12. Clementi, C., Nymeyer, H., and Onuchic, J. N. (2000) *J. Mol. Biol.* **298**, 937-953
13. Izvekov, S., and Voth, G. A. (2005) *J. Phys. Chem. B* **109**, 2469-2473
14. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., and de Vries, A. H. (2007) *J. Phys. Chem. B* **111**, 7812-7824
15. Isralewitz, B., Baudry, J., Gullingsrud, J., Kosztin, D., and Schulten, K. (2001) *J. Mol. Graphics. Modell.* **19**, 13-25
16. Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) *J. Chem. Theory Comput.* **4**, 435-447
17. Merz, T., Wetzel, S. K., Firbank, S., Plückthun, A., Grütter, M. G., and Mittl, P. R. E. (2008) *J. Mol. Biol.* **376**, 232-240
18. Dietz, H., and Rief, M. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16192-16197

Supplemental Table

Table S1. Domains in the X-ray structure of NI3C

Domain	Sequence*	# of residues
N-term	D13-A42	29
Repeat 1	K43-K76	33
Repeat 2	K76-K109	33
Repeat 3	K109-Q142	33
C-Term	Q142-Q166	24

*Residue numbering follows Merz *et al.*(17).

Supplemental Figures

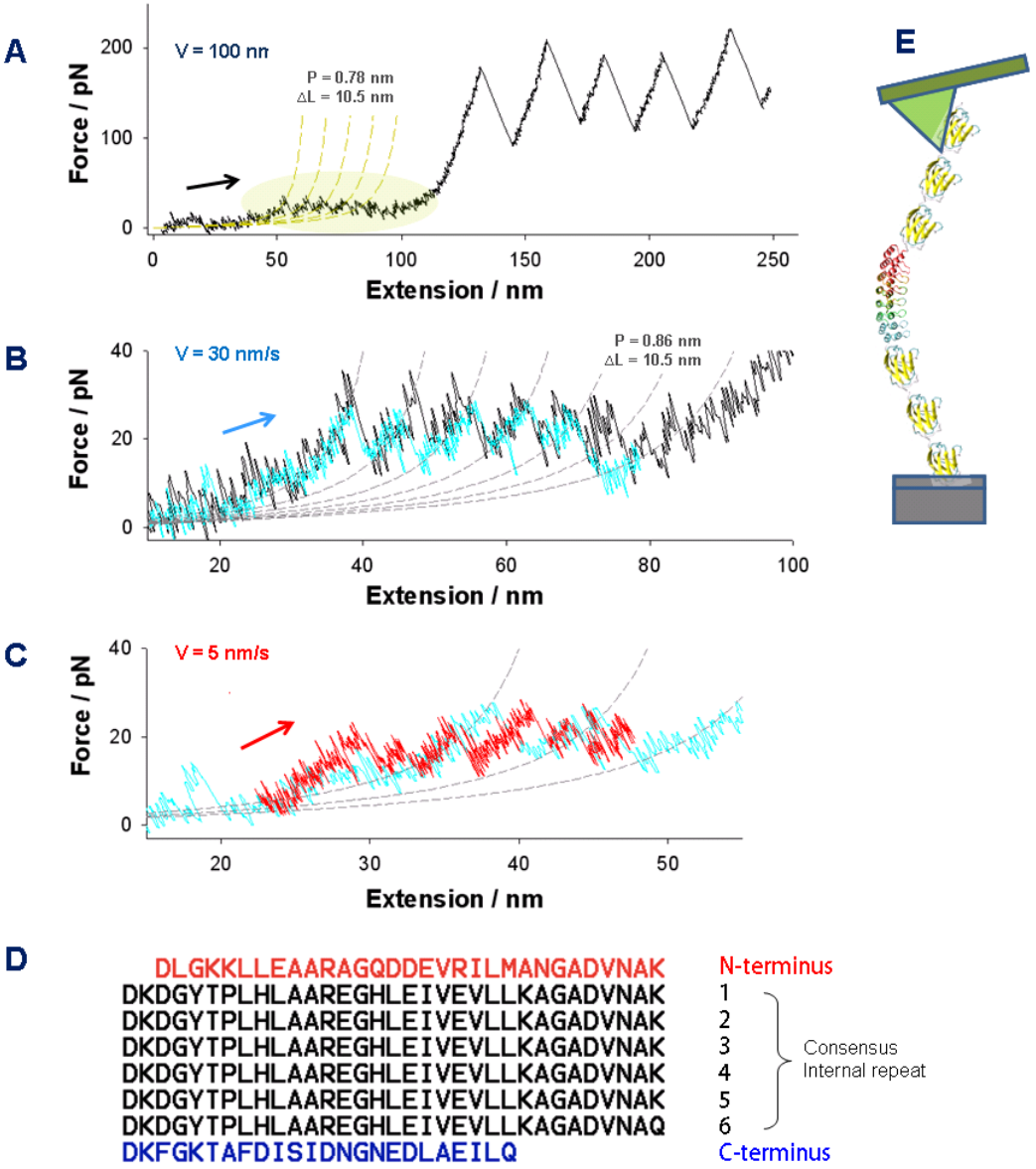


Figure S1. Comparison of the unfolding force extension traces of the NI6C-I27 construct at three different pulling speeds.

A, The force-extension curve obtained at pulling speed of 100 nm/s (black, the same curve shown in Fig. 1*B*, fitted to family of WLC curves (grey dash lines) with ΔL = 10.5 nm and p = 0.78 nm.
B, The force-extension curve obtained at 30 nm/s (cyan, the same curve shown in Fig. 1*C*), fitted to family of WLC curves (grey dash lines) with ΔL = 10.5 nm and p = 0.86 nm, superimposed with the curve at 100 nm/s (grey).
C, The force-extension curve obtained at 5 nm/s (red, the same curve shown in Fig. 3*A*) is superimposed with the curve at 30 nm/s (cyan). The force noise level in *C* is lower than in *A* and *B* thanks to the lower bandwidth of this recording (the photodiode signal low-pass filter was set to 200 Hz, compared to 500 Hz for the recordings shown in *A* and *B*.)
D, The AA sequence of NI6C is shown.

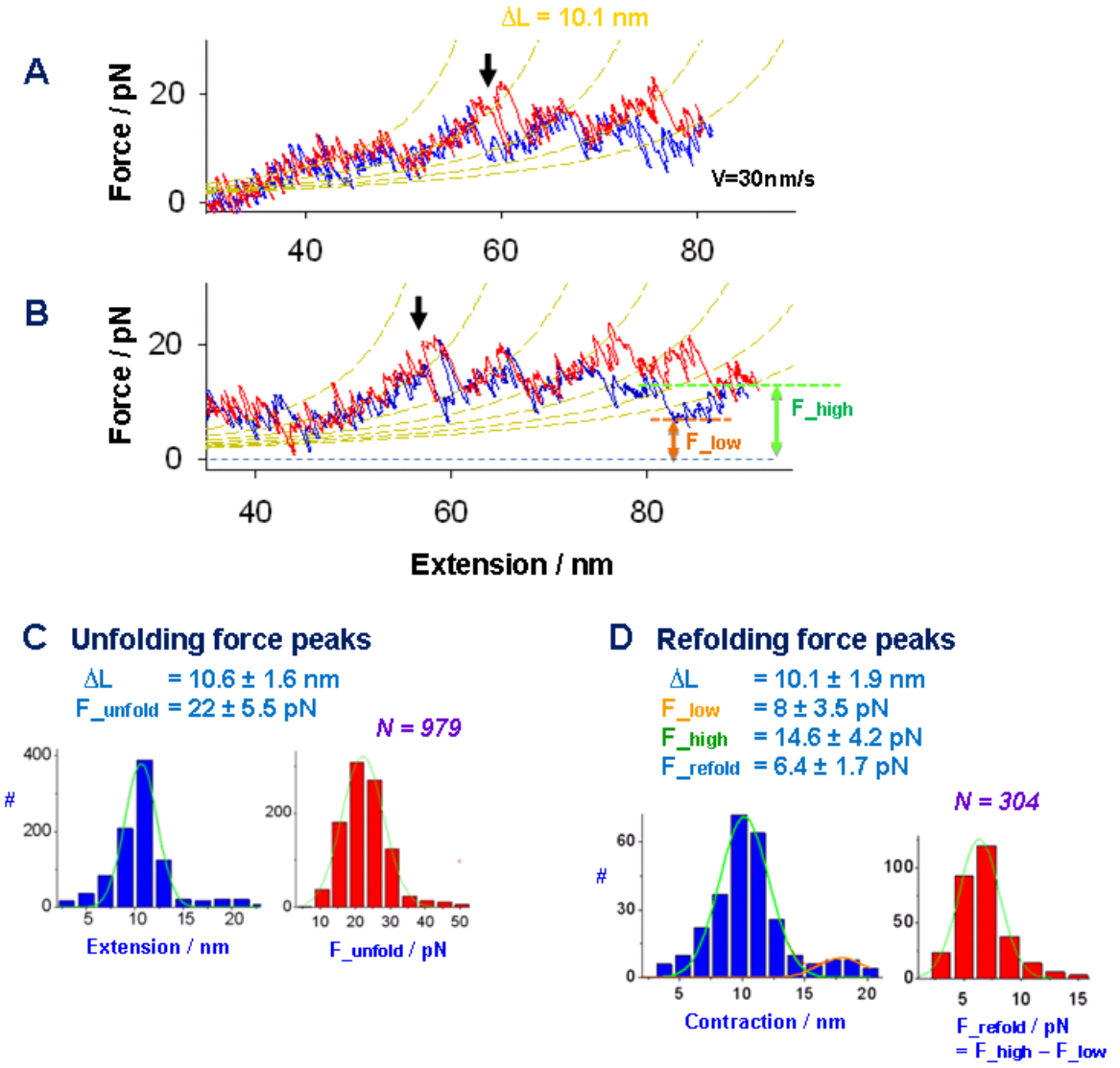


Figure S2. Examples of transient events frequently recorded when NI6C is stretched and relaxed at 30 nm/s.

A, A transient event that captured the transition: folded state → unfolded state → folded state is shown (black arrow).
B, A transient event that captured (a different molecule) the transition: unfolded state → folded state → unfolded state black is shown (black arrow).
C, The contour length increment (ΔL) and the unfolding force histograms determined from the recordings obtained at the pulling speed of 30 nm/s are shown.
D, The contour length contraction (ΔL) and the refolding force histograms determined from the recordings obtained at the pulling speed of 30 nm/s are shown. The refolding active force is measured by subtracting F_{low} from F_{high} as shown in *B*.

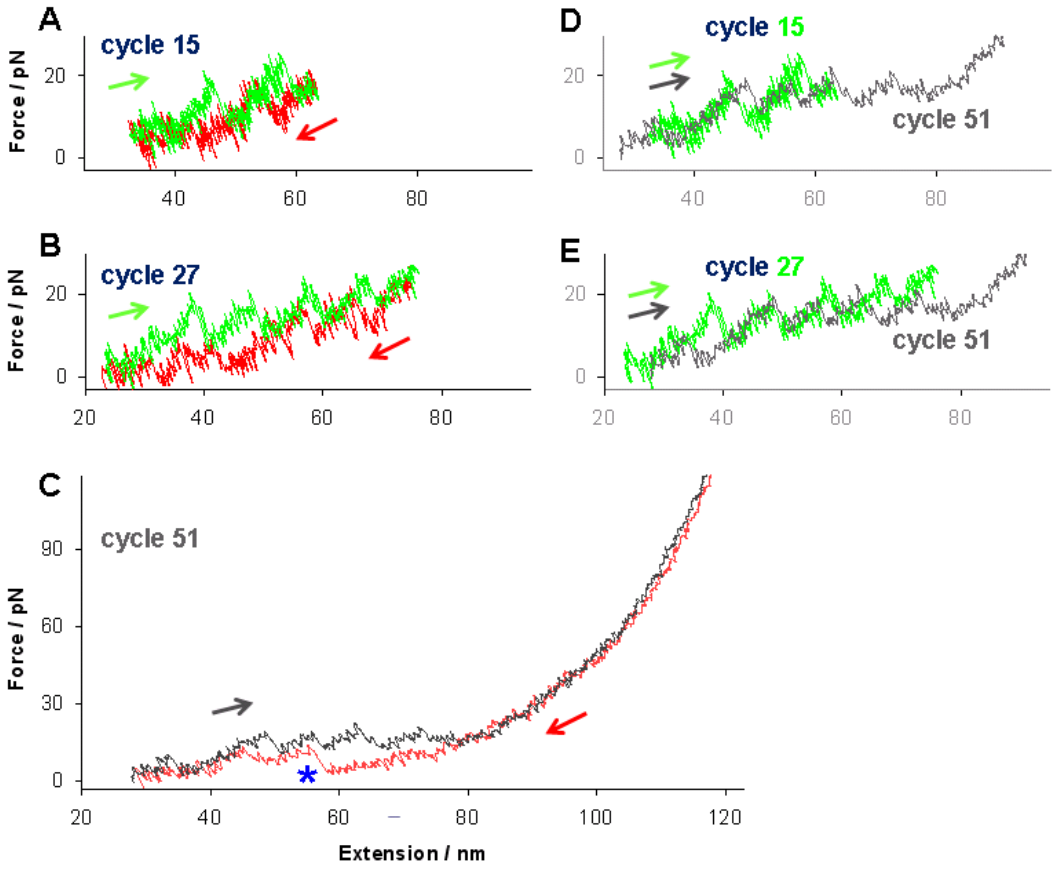


Figure S3. A transition from partial to complete mechanical unfolding and refolding of the same NI6C-I27 molecule by AFM.

A, Cycle 15, captured partial unfolding and refolding of NI6C. Note the refolding force peaks were generated in the relaxing force curve as soon as the molecule started to be relaxed.
B, Cycle 27, captured partial unfolding and refolding of NI6C.
C, Cycle 51, captured complete unfolding and refolding of NI6C. When the molecule was fully stretched (extension over 90 nm) the first refolding force peak was captured only after the molecule has been significantly relaxed (star).
D, The unfolding force extension traces of cycle 15 and cycle 51 are superimposed.
E, The unfolding force extension traces of cycle 27 and cycle 51 are superimposed.

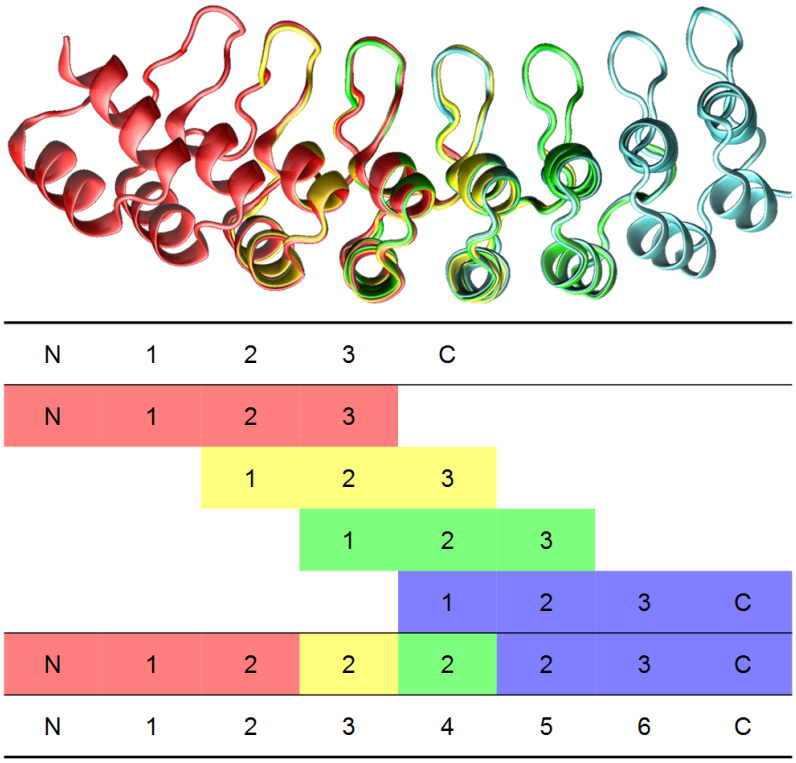


Figure S4. RMSD fitting procedure

1. Duplicate the original PDB structure with 4 copies: red, yellow, green, and blue.
2. Align the internal repeat-1, 2 in the yellow copy with the repeat-2, 3 in the red copy.
3. Align the repeat-1, 2 in the green copy with the repeat-2, 3 in the yellow copy.
4. Align the repeat-1, 2 in the blue copy with the repeat-2, 3 in the green copy.
5. Record the structure from: i) N-terminal, repeat-1, 2, 3 in the red copy, ii) repeat-2 in the yellow copy, iii) repeat-2 in the green copy, iv) repeat-2, 3 and C-terminal repeat in the blue copy.
6. Adding coils composed by six glycine residues as the handles on both ends of the protein.
7. Re-order the residue numbers from 1 to 265 (from N to C) to finish the structure of NI6C.

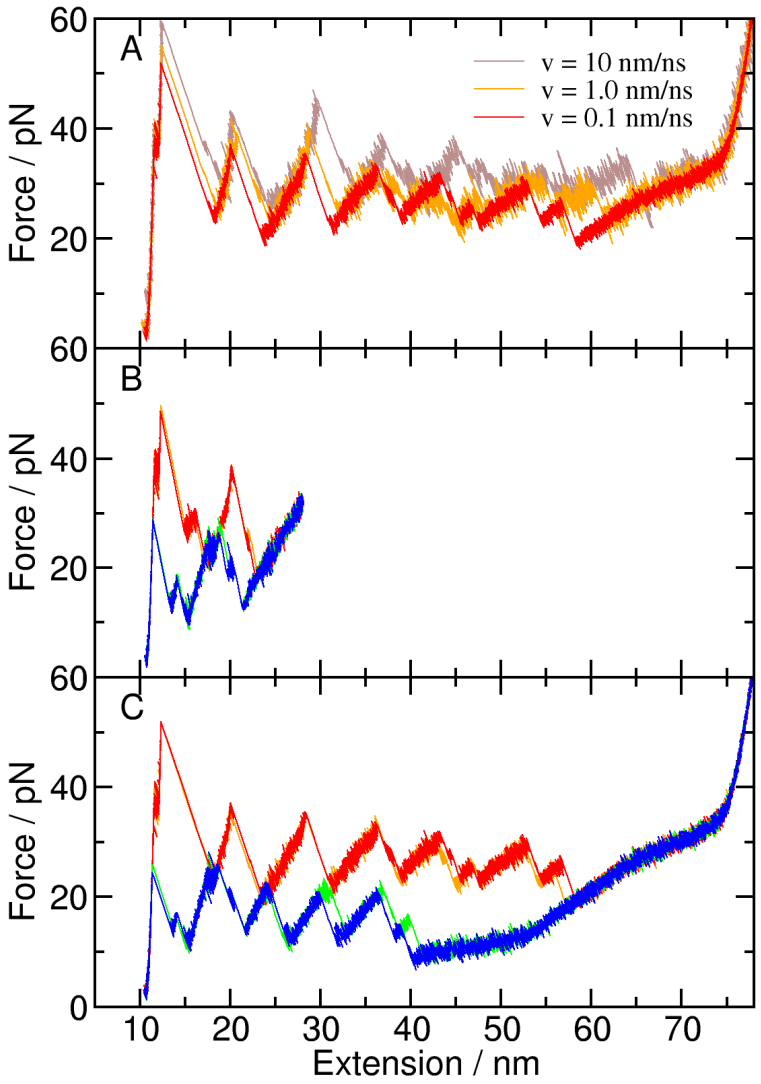


Figure S5. Reproducibility of the force-extension curves in SMD.

A, Force-extension curves obtained at different pulling speed, with the range from 10 nm/ns to 0.1 nm/ns. The curves at v = 1.0 nm/ns and v = 0.1 nm/ns showed similar peak patterns.
B, Multiple force-extension curves for partial folding/refolding from multiple SMD simulation with different initial velocities at v = 0.1 nm/ns. The red and orange traces are for unfolding; the green and blue traces are for the following refolding.
C, Multiple force-extension curves for complete folding/refolding with different initial velocities at v = 0.1 nm/ns. The overlaps demonstrated the reproducibility of the unfolding and refolding processes.

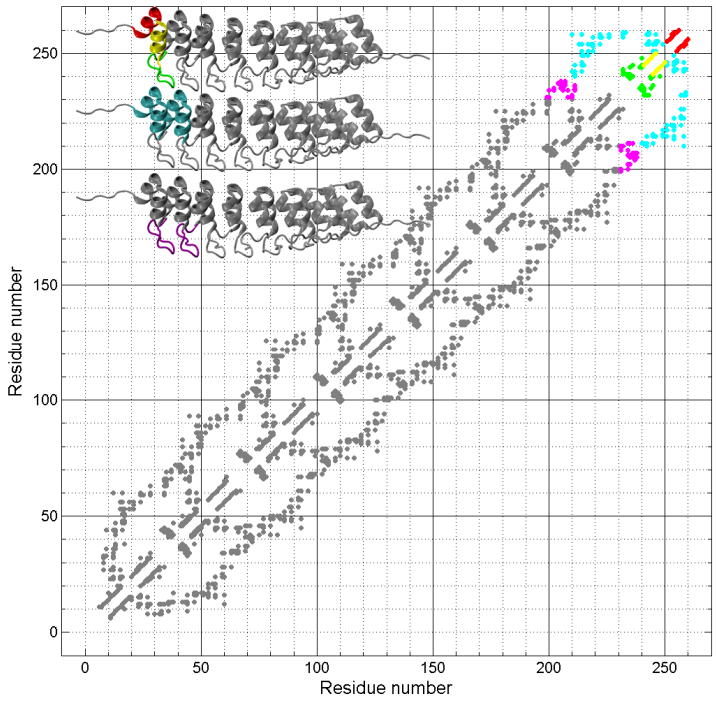
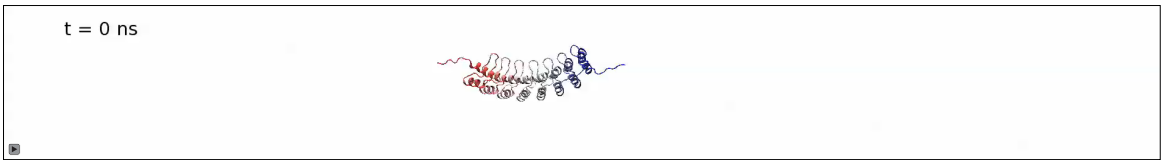


Figure S6. The contact map of the consensus AR NI6C protein and the corresponding structures.

The red, yellow and green dots on the map represent the internal native contacts within the last repeat at the C-terminus as shown in the inset. The dots in cyan and magenta are the inter-repeat contacts between the last two repeats at the C-terminus (shown in the inset).

Supplemental Movies

Movie S1. The complete unfolding process of the consensus AR protein NI6C.



Movie S2. The refolding process of the consensus AR protein NI6C from completely unraveled conformation.