Mutation of Conserved Histidines Alters Tertiary Structure and Nanomechanics of Consensus Ankyrin Repeats

Whasil Lee‡, Johan Strümpfer§, Vann Bennett¶1, Klaus Schulten§2, and Piotr E. Marszalek‡3

From the ‡Center for Biologically Inspired Materials and Material Systems and Department of Mechanical Engineering and Materials Science and the §Howard Hughes Medical Institute and Department of Cell Biology, Duke University, Durham, North Carolina 27708 and the ¶Center for Biophysics and Computational Biology and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

Background: The His → Arg mutation in a native ankyrin-R protein is responsible for converting RBCs to spherocytes, causing hereditary spherocytosis (HS).

Results: The mutant unfolds and refolds at lower forces compared with the wild type.

Conclusion: The His → Arg mutation weakens the mechanical stability of ankyrin repeats.

Significance: The His → Arg mutation in ankyrin-R may cause HS by decreasing the mechanical stability and affecting its structure recovery ability.

The conserved TPLH tetrapeptide motif of ankyrin repeats (ARs) plays an important role in stabilizing AR proteins, and histidine (TPLH)-to-arginine (TPLR) mutations in this motif have been associated with a hereditary human anemia, spherocytosis. Here, we used a combination of atomic force microscopy-based single-molecule force spectroscopy and molecular dynamics simulations to examine the mechanical effects of His → Arg substitutions in TPLH motifs in a model AR protein, NI6C. Our molecular dynamics results show that the mutant protein is less mechanically stable than the WT protein. Our atomic force microscopy results indicate that the mechanical energy input necessary to fully unfold the mutant protein is only half of that necessary to unfold the WT protein (53 versus 106 kcal/mol). In addition, the ability of the mutant to generate refolding forces is also reduced. Moreover, the mutant protein subjected to cyclic stretch-relax measurements displays mechanical fatigue, which is absent in the WT protein. Taken together, these results indicate that the His → Arg substitutions in TPLH motifs compromise mechanical properties of ARs and suggest that the origin of hereditary spherocytosis may be related to mechanical failure of ARs.

Ankyrin repeats (ARs)4 are one of the most common motifs of repeat proteins with a high degree of amino acid sequence homology. ARs fold into nearly identical helix-1-helix-2-loop structures and stack to form elongated superhelical domains that frequently mediate protein-protein interactions (1–3). Previously, single-molecule force spectroscopy experiments of a few AR proteins revealed that mechanically unfolded ARs refold rapidly and generate very robust refolding forces (4–9). Recently, we examined in detail the mechanical properties of a model synthetic AR protein, NI6C (7). NI6C is composed of six identical internal ARs based on a consensus sequence and two capping repeats (10). Our atomic force microscopy (AFM) manipulations on NI6C presented very interesting mechanical properties. We observed that NI6C unfolds stepwise, repeat by repeat, and individual unfolding events produce regular, well-resolved force peaks of ~22 piconewtons (pN). In addition, relaxation traces of NI6C captured robust refolding force peaks and revealed no elastic hysteresis or fatigue, suggesting that this protein possesses superior elastic properties (7).

It has been suggested that the conserved TPLH tetrapeptide motif of ARs plays an important role in stabilizing AR proteins (2, 11–13). Some mutations in the TPLH motif, including Thr → Ala, His → A, and/or His → Gln substitution, were found to significantly affect the thermal stability of gankyrin (11). However, the role of the TPLH motif in the mechanical stability of ARs has not been studied.

Hereditary spherocytosis (HS) is a life-threatening human anemia in which erythrocytes (RBCs) lose their characteristic shape and become spherical and very fragile (14, 15). HS is caused by defects in proteins that mediate membrane-cytoskeleton connections in erythrocytes such as ankyrin-R, spectrins, and band 3 (1, 14, 16). There are two missense mutations identified so far in the AR segment of ankyrin-R that are related to HS: H277R and V463I (14, 16). Importantly, the His → Arg substitution involves a highly conserved histidine of the TPLH motif.

Because RBCs undergo repeated cycles of stress and deformation while flowing through narrow capillaries and slits of the spleen (16, 17), the elastic response of the membrane and associated cytoskeletal proteins to the applied forces is essential for
Mutation Effects on Ankyrin Repeat Mechanics

maintenance of RBC integrity and to enable recovery. Force-induced unfolding of spectrin was already captured in live RBCs under physiological stress (18). It is therefore possible that ankyrin-R, depending on the (presently unknown) direction and magnitude of applied forces, also undergoes stretch-relax cycles in vivo, which may involve all or some of its ARs.

The goal of this work was to test the hypothesis that His → Arg substitutions in the TPLH motif disrupt stabilizing interactions and compromise mechanical properties of ARs. Because the membrane-binding domain of ankyrin-R is very large and its crystal structure is not yet available, we decided to carry out our initial studies of the mechanical effect of His → Arg substitution in the TPLH motif in NlI6C, which proved to be an excellent model AR system amenable to single-molecule force spectroscopy by AFM (7). To amplify the possible effect of the His → Arg mutation, we substituted TPLH histidines in four internal ARs within NlI6C (see Fig. 1B). We used a combination of AFM-based single-molecule force spectroscopy (19–27) and molecular dynamics (MD) simulations (9, 28–34) to examine how mutations in this motif affect the mechanical and structural properties of NlI6C when the proteins are pulled by their N and C termini. We found that for the particular pulling geometry used in this study (stretching by the N and C termini), the mechanical unfolding forces of the mutant were significantly reduced, and as a result, the total consumed mechanical energy during the unfolding process decreased by half (53 versus 106 kcal/mol) compared with WT NlI6C, suggesting that His → Arg substitutions indeed alter the mechanical properties of ARs. Furthermore, we registered that the mechanical refolding events of the mutant were significantly reduced, and observed the mechanical fatigue of the mutant over cyclic stretch-relax experiments. Taken together, these observations suggest that the H277R mutation possibly reduces the mechanical robustness of ankyrin-R and may ultimately cause its mechanical failure, leading to HS.

MATERIALS AND METHODS

Cloning, Expression, and Purification of Repeat Proteins—
The gene of the mutant protein Nl(IH7R)4IC was synthesized by GenScript (Piscataway, NJ). In the mutant protein, four internal consensus ARs (repeats 2–5) were modified by substituting the histidines at position 7 with arginines (see Fig. 1B). Please note that in Fig. 1B, the TPLH histidines are in position 9. However, this position corresponds to position 7 within the AR sequence. Mutation Effects on Ankyrin Repeat Mechanics—The WT NI6C protein contains eight ARs: six identical consensus repeats flanked by slightly different N- and C-terminal capping repeats (see Fig. 2) (7). WT ARs unfold sequentially, generating five to six almost uniform unfolding force peaks of ~23 pN, spaced by ~11 nm. Supported by MD simulations, these sharp and regular five to six unfolding force peaks are interpreted to report the unfolding of six internal consensus repeats. The unfolding of the capping repeats is typically unresolved. Because ARs unfold sequentially and we mutated only the internal four repeats, we expect that the middle peaks in force-extension curves of the mutant...
(see Fig. 2C) are generated when the mutated internal repeats are unruled. Often the unfolding force peaks of the mutant were not sharp but smeared and merged, producing a force curve with a shape of a force plateau. For these reasons, it was rather difficult to determine the individual unfolding force peaks and their amplitudes. To quantify mechanically the effect of the mutation, the consumed energy during the unfolding process was calculated for each curve (WT and mutant). As an example of this energy input, the area enclosed by the unfolding force-extension curve of ARs (corresponding to the extension between 18 and 81 nm) and the worm-like chain curve drawn through the first I27 unfolding peak was calculated (yellow area in supplemental Fig. S2, C and D). The average “mechanical unfolding energy” defined this way was 53.2 ± 10.9 kcal/mol (n = 50) for the mutant and 106.5 ± 12.7 kcal/mol (n = 12) for the WT protein. Thus, the four internal AR His → Arg mutations cause the mechanical unfolding work input to decrease by ~50% compared with the WT protein.

**MD Simulations**—The initial geometry of N16C (253 amino acids) (7) was built on Protein Data Bank structure 2QYJ, corresponding to the NI3C consensus AR protein (10). In N16C, the two capping and all internal repeat amino acid sequences are the same as in NI3C, and only the number of internal repeats is increased in N16C from three to six. Structure building details have been described (7). The histidines of four internal repeats of N16C (see Fig. 1B) were mutated to arginines using the program VMD (39). Equilibrium MD simulations of both N16C and the mutant NI(IH7R)4IC structures in periodic water boxes (110 × 100 × 100 Å^3) with 150 mM NaCl, using the CHARMM22 force field (43, 49), were performed with NAMD (40). The force response of each structure was subsequently probed by steered MD (SMD) simulations (41), where the C-terminal Cα of the protein is pulled with a constant velocity (v; by an attached spring force) while the N-terminal Cα is held fixed. Each structure was extended by 50 nm using a pulling velocity of v = 5 Å/ns and a spring constant of 3 kT/Å^2. Simulations were carried out at 310 K and 1 atm. **supplemental Table S1** presents a summary of all simulations performed.

**RESULTS AND DISCUSSION**

Fig. 1A shows a ribbon diagram of the TPLH region in the WT N16C protein (7, 42). The TPLH motifs are located between the hydrophobic helical bundles and the solvent-exposed loop regions (2). The histidines in the TPLH motifs are located at the beginning of helix 1. As marked by dotted lines in Fig. 1A, histidines at position 7 and tyrosines at position 4 in TPLH motifs form an H-bond network. Each TPLH histidine forms three H-bonds with the TPLH tyrosine within the same repeat and one H-bond with the TPLH tyrosine of the next AR. The intra-repeat H-bonds are Thr-4 H^N–His-7 N^α1, Thr-4 O^Y–His-7 H^N, and Thr-4 H^Y–His-7 N^α1. The inter-repeat H-bond is between His-7 H^C2 and Thr-4 O of the next repeat. Because of these extensive interactions, the conserved TPLH motif is vital to the protein structure and its stability (11). For these reasons, we anticipate that His → Arg substitutions in the TPLH motif will likely affect both the structure and stability of AR polypeptides.

**Equilibrium Structures of N16C and NI(IH7R)4IC**—First we employed MD simulations to test whether the structure of N16C changes from the His → Arg mutations. Four internal consensus ARs (repeats 2–5) (Fig. 1B) of WT N16C (7) were modified in the NI(IH7R)4IC mutant by mutating histidines at position 7 to arginines. Equilibrium all-atom MD simulations of both N16C and NI(IH7R)4IC were performed, totaling 200 ns (see **supplemental Table S1** for details). The WT N16C protein showed little deviation from the initial structures in the equilibrium simulations, with average backbone root mean square (r.m.s.) deviations (C^α r.m.s. deviations) of 1.14, 1.14, and 1.33 Å in each N16C simulation (Fig. 1C). The NI(IH7R)4IC mutant showed significantly more deviation from the initial structure, with average C^α r.m.s. deviations of 1.8, 2.1, and 2.2 Å in each simulation. Much of the deviation occurred in the loop regions between the repeats, with an extension of the overall structure from an average length of 73.5 Å for N16C to 76.2 Å for NI(IH7R)4IC. Significant rearrangement of H-bonding in the TPLH motifs occurs due to the His → Arg mutation: the bulkier arginine displaces the inter-AR loop and forms a single H-bond with the aspartic acid in position 32 (in the loop region between ARs), thus altering the average loop orientation with respect to the AR α-helices (Fig. 1D). The reduction from three H-bonds, made by the histidine in N16C, to only one by the arginine in NI(IH7R)4IC results in a destabilization of the mutant structure, especially the loop regions, as shown by the r.m.s. fluctuations calculated from the equilibrium MD trajectories (Fig. 1E).

**In Silico Force Unfolding of N16C and NI(IH7R)4IC**—Following equilibrium simulation, we performed two SMD simulations for each N16C and NI(IH7R)4IC, pulling on the C-terminal C^α using a pulling velocity of v = 5 Å/ns. The resulting force-extension curves from the simulations are shown in **supplemental Fig. S1A**. The force-extension curves show as much difference between the WT and mutant structures as between repeated SMD simulations of the same structure. All simulations showed major force-rupture peaks spaced every 10 nm, which corresponded to the rupture of each AR. Unfolding of each AR occurred for both the WT and mutant proteins from the C terminus to the N terminus, as reported previously (7). We note, however, that the unfolding pathways as determined through SMD simulations may differ from the pathways explored by the proteins in AFM unfolding measurements, and these issues have to be further examined in future studies. Visual analysis of the SMD simulations showed that the major force-rupture peaks are due to the dissociation of the intra-repeat helices (supplemental Fig. S1B). Given the amplitude of fluctuations seen in equilibrium simulations (Fig. 1E), slower pulling velocities than those achievable with current computational resources are likely required to capture the mechanical differences between the force unfolding of N16C and NI(IH7R)4IC.

**Mechanical Unfolding of NI(IH7R)4IC**—To examine the mechanical unfolding behavior of an intact NI(IH7R)4IC mutant, we performed single-hit stretching measurements on NI(IH7R)4IC-I27 constructs, which were designed to limit possible mechanical damage to the protein resulting from multiple attempts to pick it up by an AFM tip (see “Materials and Methods”). To ease single-molecule force spectroscopy
measurements and their interpretations, the mutant protein Ni(I_{H7R})_{4}IC, similar to the Ni6C protein (7), was flanked on each side by three I_{27} domains of titin, serving as pulling handles and as a force spectroscopy reference for identifying single-molecule recordings (8, 24, 35, 44, 45).

Fig. 2A (inset) shows schematic ribbon diagrams of the chimeric proteins (I_{27})_{3}-Ni6C-(I_{27})_{3} and (I_{27})_{3}-Ni(I_{H7R})_{4}IC-(I_{27})_{3}, which were examined by AFM. Since we studied the mechanical property of WT Ni6C previously (7), we compared the AFM results of Ni(I_{H7R})_{4}IC with the force spectra of Ni6C. In Fig. 2A, we show the representative unfolding force-extension traces that were recorded in single-molecule AFM measurements of Ni6C-I_{27} and Ni(I_{H7R})_{4}IC-I_{27} constructs. Both curves were obtained at the same pulling speed of 0.1 nm/ms. The five I_{27} unfolding force peaks at extensions above 100 nm provide direct evidence that the whole measurement was obtained on a single molecule containing all eight ARs. Small unfolding force peaks at protein extensions below 100 nm in the red and green curves strongly suggest that they correspond to the sequential unfolding of Ni6C and Ni(I_{H7R})_{4}IC, respectively. We used these recordings as the reference force spectra to identify unfolding events of ARs in other measurements on the WT and mutant proteins.

In supplemental Fig. S2 (A and B), we superimposed AFM recordings of different molecules of the mutant and WT proteins, respectively. In Fig. 2C, we compare the AFM recordings of the mutant with a typical unfolding force-extension curve of WT Ni6C (7). It is clear that the force-extension curves of the mutant overlap reasonably well with the force-extension curve of WT Ni6C at the beginning and at the end. However, at the intermediate extensions (indicated by the green bar in Fig. 2C),
the unfolding forces of the mutant are significantly lower compared with the unfolding forces of the WT protein. These observations are consistent with the design of the NI6C mutant, in which only the four internal repeats were mutated. We determined the mechanical energy consumed during the AR unfolding process by integrating the area between each force-extension curve and the worm-like chain curves (yellow area in supplemental Fig. S2, C and D). The average energy input necessary to mechanically unfold the mutant was 53.2 ± 10.9 kcal/mol (n = 50), and that for the WT protein was 106.5 ± 12.7 kcal/mol (n = 12). Thus, the four internal AR His→Arg mutations resulted in the ~53-kcal/mol decrease in the work necessary to unravel the mutated ARs compared with the WT ARs. In summary, the mechanical unfolding forces of the mutated ARs are significantly lower than those of the WT ARs. These observations confirm our expectations that the His→Arg mutations in TPLH motifs may reduce the mechanical stability of ARs.

**Mechanical Refolding of NI(IH7R)4IC in Cyclic Measurements**—To examine the mechanical refolding behavior of NI(IH7R)4IC, we performed cyclic stretch-relax measurements on the NI(IH7R)4IC-I27 construct. During these measurements, the stretching distance was limited to unfold most of the ARs of the NI(IH7R)4IC insert without unfolding of any of the I27 domains. Because the I27 domains remained folded, they did not interfere with the folding of ARs during the relaxation step. In Fig. 3A, a typical stretching force-extension curve of the mutant NI(IH7R)4IC protein obtained under such partial unfolding conditions is superimposed with its refolding force-extension trace. In contrast to the large and sharp refolding force peaks of the WT NI6C protein (black trace in Fig. 3B), the refolding force peaks of the mutant are smaller and somewhat smeared. The refolding traces of the WT and mutant proteins are superimposed in Fig. 3C. During repetitive stretch-relax measurements on the mutant, the refolding force peaks diminished, further indicating some refolding “fatigue” (Fig. 4 and supplemental Fig. S3, A–C) that was not observed for the WT protein (supplemental Fig. S3, D–F; see supplemental data in Ref. 7). In summary, His→Arg substitution in the TPLH motif seems to significantly affect the mechanical unfolding and refolding behavior of consensus ARs.

Our MD and AFM results revealed that His→Arg substitutions in TPLH motifs 1) eliminate inter-repeat H-bonds, greatly increasing the flexibility in the loop regions, which lessens the mechanical stability of the mutated ARs; 2) decrease the observed unfolding and refolding forces compared with those of the WT protein when both proteins are stretched by their termini (mechanical unfolding energy of 53 versus 106 kcal/mol); and 3) cause mechanical fatigue, which is absent in the WT protein. These results indicate that the mutated ARs in NI(IH7R)4IC unfold easily and refold in a less springy, less robust manner compared with the WT protein. We hypothesize that the His→Arg mutation in ankyrin-R may cause HS by altering its tertiary structure and decreasing its mechanical robustness.
Mutation Effects on Ankyrin Repeat Mechanics

Repetitive deformations of ankyrin-R during circulation of RBCs may induce the HS mutant to progressively lose its structural and elastic properties, leading to the mechanical failure of ankyrin-R. This could in turn perturb the protein network mediated by ankyrin-R and disrupt important membrane-cyto-skeleton connections.

Acknowledgments—We thank Jane Clarke for providing the plasmids with poly-I27 domains. Supercomputer time was provided by the Texas Advance Computing Center at The University of Texas at Austin as part of XSEDE (National Science Foundation Grant OCI-1053575), and the parallel computing resource was provided by the Computational Science and Engineering Program at the University of Illinois as part of the Taurus cluster.

REFERENCES
1. Bennett, V., and Baines, A. J. (2001) Spectrin- and ankyrin-based pathways: metazoan inventions for integrating cells into tissues. Physiol. Rev. 81, 1353–1392
2. Michaely, P., Tomchick, D. R., Machius, M., and Anderson, R. G. (2002) Crystal structure of a 12-ANK repeat stack from human ankyrin-R. EMBO J. 21, 6387–6396
3. Ferreiro, D. U., Cho, S. S., Komives, E. A., and Wolynes, P. G. (2005) The energy landscape of modular repeat proteins: topology determines folding mechanisms in the ankyrin family. J. Mol. Biol. 354, 679–692
4. Lee, W., Zeng, X., Zhou, H. X., Bennett, V., Yang, W., and Marszalek, P. E. (2003) Nanospring behavior of ankyrin repeats. Nature 440, 246–249
5. Kim, M., Abdí, K., Jiang, Y., Michaely, P., Bennett, V., and Marszalek, P. E. (2006) Nanospring behavior of ankyrin repeats. Nature 440, 246–249
Unfolding of titin immunoglobulin domains by steered molecular dynamics stimulation. Biophys. J. 75, 662–671
42. Merz, T., Wetzel, S. K., Firbank, S., Plückthun, A., Grütter, M. G., and Mittl, P. R. (2008) Stabilizing ionic interactions in a full consensus ankyrin repeat protein. J. Mol. Biol. 376, 232–240
43. MacKerell, A. D., Jr., Bashford, D., Bellot, M., Dunbrack, R. L., Jr., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., III, Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B 102, 3586–3616
44. Perez-Jimenez, R., Garcia-Manyes, S., Ainavarapu, S. R., and Fernandez, J. M. (2006) Mechanical unfolding pathways of the enhanced yellow fluorescent protein revealed by single-molecule force spectroscopy. J. Biol. Chem. 281, 40010–40014
45. Dietz, H., and Rief, M. (2004) Exploring the energy landscape of GFP by single-molecule mechanical experiments. Proc. Natl. Acad. Sci. U.S.A. 101, 16192–16197
46. Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., Richardson, J. S., and Richardson, D. C. (1999) Visualizing and quantifying molecular goodness of fit: small probe contact dots with explicit hydrogen atoms. J. Mol. Biol. 285, 1711–1733
47. Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side chain amide orientation. J. Mol. Biol. 285, 1735–1747
48. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 35, W375–W383
49. Brooks, C. L., 3rd (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics stimulations. J. Comput. Chem. 25, 1400–1415