Supporting Text

The section contains additional information regarding the simulations and protein structures reported in the main article.

The first two sections provide information about the protein structure considered in the article. In the next sections, we explain mathematical symbols used in the article and discuss background information on the Bell theory and related statistical approaches.

References to specific online material are also included in the main article.

α-Helical Secondary Protein Structure: Background Information

An α-helix (AH) is generated when a single polypeptide chain twists around itself, stabilized by hydrogen bonds (HBs) formed between every fourth residue, linking the O atom of peptide *i* to the N atom of peptide $i+4$ in the residue chain. Consequently, at each convolution, 3.6 HBs are found in parallel arrangement that stabilize the helical configuration (1, 2). The AH was the first protein motif whose atomic structure was known. It was first found in keratin filaments of wool >50 years ago.

As one of the secondary structures, AHs are found in most proteins. Consequently, their mechanical role and its link to associated chemical reactions are of vital importance in answering many questions related to biophysics, biochemistry, or biomedicine. Additional applications are in the *de novo* design and manufacturing of nano-featured protein materials (3-5), where AH motifs are used as a self-assembling building block for bottom-up designed biomaterials.

A variety of AH structures as well as AH coiled-coils have been studied in experiment as well as in simulation (6-23).

The α-helical structure used in the studies reported in this article is taken from the 2B segment of the vimentin coiled-coil dimer. Vimentin belongs to the group of intermediate filaments (IFs), which are in addition to microfilaments (MFs) and microtubules (MTs) one of the three major components of the cytoskeleton in eukaryotic cells. The cytoskeleton plays a critical role in determining the shape and the mechanical properties of the cell, and is vital for many additional functions including protein synthesis, cell motility, as well as cell division or wound healing (24-26). IFs control the large deformation behavior of cells as well as those of cellular tissue, and play a crucial role in mechanical resistance of cells at large deformation. Further, IFs are part of mechanotransduction chains connecting the cellular environment to the DNA (27, 28).

β-Sheet (BS) Amyloid Protein Structure

A BS structure is created when the protein backbone aligns in an extended configuration, in parallel or antiparallel orientation by forming extensive HB networks between the protein strands.

Here we model the mechanics of the structure of Alzheimer's amyloid β-fibrils (29). The coordinates of this structure are obtained from the Protein Data Bank (30) (PDB ID code 2beg).

This protein structure consists of five peptide chains that are stacked along the fibril axis and are stabilized by interchain HBs. Each chain forms a β-strand-turn-β-strand motif that contains two intermolecular, parallel, in-register β-sheets that are formed by residues 18-24 and 35-41. Using the SMD scheme, we pull on the middle chain of the assembly (third chain from top or bottom) at the midpoint of the turn that connects the two β-strands. We fix all the C_α atoms on top and bottom chains during pulling. These boundary conditions are in line with molecular deformation mechanisms hypothesized in recent atomic force microscopy (AFM) experiments on amyloid fibrils (31).

Rupture of HBs between parallel β-strands is the key mechanism that leads to forced unfolding of the protein and thus controls its fracture properties.

Mathematical Symbols Used in This Article

SI Table 2 provides an overview of all mathematical symbols used in the article, including their units and a brief description.

We note that the binomial coefficient is defined as

$$\binom{b}{d} = {}_bC_d = \frac{b!}{(b-d)!\,d!}. \qquad \textbf{[10]}$$

The molecular engineering strain is defined as $\varepsilon = (x - x_0)/x_0$, where x_0 is the initially equilibrated, undeformed length, and *x* is the current, deformed length of the protein.

Strength Models of Protein Materials

Large deformation can, for instance, be induced at macroscopic crack-like defects in tissue, where the stresses and thus molecular forces show a singularity (32, 33). Each protein is exposed to large forces, and the resistance to macroscopic crack growth depends on the resistance of each protein molecule (34).

Protein unfolding in structural protein materials represents a fundamental crack-tip mechanism, in analogy to dislocations in metals or crack tip mechanisms in brittle materials. An understanding of the unfolding behavior of proteins is thus critical to develop fracture theories of biological tissues.

Research Design and Background Information on Bell's Theory

Systematic studies of protein unfolding at different pulling rates provide valuable insight into the protein's mechanical and thermodynamical stability and behavior. Such studies increase the understanding of the protein's behavior on different time scales, because the pulling speed is the key characteristic in defining the time scale.

Here we perform a series of atomistic studies of tensile deformation of a single AH to arrive at a detailed understanding of the mechanical behavior of this molecule over 4 orders of magnitudes of pulling velocities. Development of a theoretical model, based on the phenomenological approach, postulated by Bell, enables us to derive a rigorous understanding of the pulling rate dependence.

Our strategy to determine the dependence of the unfolding force F on pulling speed, associated mechanisms and energy barriers is to carry out MD simulations at different pulling rates, measure the dependence of unfolding force on the pulling rate, and fit the parameters a and b to these data. This information can then be used to calculate the parameters x_b and E_b and thereby reconstruct the free energy landscape.

In addition, to allow for an analytical, continuum-type expression to describe the unfolding dynamics, the extended Bell theory enables us to bridge time scales, because we describe the force as a function of the pulling speed, which is a controlled parameter in simulation and experiment. Understanding the effect of the pulling velocity on the unfolding behavior is of great significance because protein structures display a significant dependence of the unfolding mechanics on the pulling speed. This may control some of their biological functions, for instance, in mechanotransduction or in light of the intermediate filament's security belt function. Further, this insight will eventually enable us to develop continuum-type descriptions of the viscoelastic properties of proteins. As pointed out in ref. 35, understanding the rate-dependent properties is also vital to predicting dissipative properties of proteins, protein networks, and cells under cyclic loading.

Eq. **2** in the main article provides an immediate link between the pulling rate and the pulling force that is necessary to lower the energy barrier in such a way that the bond can be broken with the applied velocity. Increasing the pulling rate means increasing the off-rate and thus the probability of bond rupture. This is only possible through lowering the energy barrier at the transition point, resulting in a higher bond-breaking force f. SI Fig. 6 illustrates the concept behind the Bell theory.

Extension of the Hierarchical Bell Model to a Three-Level Hierarchical System

It was shown in Eqs. **5** and **6** how a system's unfolding rate (off-rate) depends on the number of parallel bonds, where the system consisted of two hierarchies. This can be extended to a third or higher-level hierarchy, which enables us to cover the unfolding rate of a tertiary structure consisting of 2, 3, … *n* AHs, of which *k* unfold simultaneously. This is, for example, the case in a two, three, … *n*-stranded coiled-coil or a filament.

Consequently, *n* over *k* possibilities exist that the unfolding appears in *k* of the *n* helices. Additionally, because the unfolding can begin in each of *n* helices, the probability on the next smaller hierarchy is decreased by the exponential *n*, resulting in $\binom{b}{d}^{-n}$. Further E_b^0 increases *k* times, because now $k \cdot d$ bonds break simultaneously. Finally, we arrive at the following expression:

$$\chi_H = \omega_0 \cdot \binom{n}{k}^{-1} \cdot \binom{b}{d}^{-n} \cdot \exp\left(-\frac{\left(k \cdot d \cdot E_B^0 - f \cdot x_b \cdot \cos(\theta)\right)}{k_B \cdot T} \right). \qquad \textbf{[11]}$$

We have proved this relation by direct MD simulation of different AH configurations (unpublished data).

Negative Force Values in the Bell Theory

We note that the points where the forces are calculated to be zero and below determine the natural bond-breaking speed v_0. At pulling speeds smaller than v_0, the bond dissociates and rebuilds spontaneously because of thermal fluctuations. In other words, as shown in ref. 36, for speeds smaller than v_0 the time scale of thermally activated unfolding is longer than the mechanically activated unfolding and thus leads to the creation of new energy barriers. This is the reason why three to four parallel HBs guarantee the thermodynamic stability of AHs at body temperature but are, at the same time, easy to build and to destroy.

Comments regarding the Fast Deformation Mechanism (FDM)

The smaller x_b in the FDM than in the SDM (see Table 1) is due to a strain localization in the molecule; that is, the local strain is several times larger than the laterally applied strain. This amplification of local strain leads to a smaller value of x_b. Further, before the first HB ruptures at the AP, all HBs in the molecule are stretched by ≈0.2 Å, which equals 7% HB strain (equilibrium HB length is 3 Å). At the AP, one of the HBs rapidly extends to x_b = 1.2 Å and ruptures. In contrast, in the SDM, the protein is not homogeneously stretched in the first regime. Rather, one random convolution is stretched so that three to four HBs extend homogeneously. Subsequently, three HBs rupture simultaneously, as shown in Fig. 4 (main article).

The mechanism is shown schematically in SI Fig. 7, which plots the microscopic strain distribution before and after the angular point (AP) is reached, comparing the FDM (*a*) and the SDM regime (*b*).

Previous Theoretical Work on Rupture Forces and Free Energy Landscapes

Tremendous contributions in this field have been made by Evans and coworkers (37-44) by developing a theory that describes the binding behavior of proteins in dependence of the force rate *r* (increase in force over time, here referred as the “loading-rate theory”).

The experimental and simulation data published by Evans led to the suggestion that the energy landscape is governed by multiple transition states, represented by a spline curve consisting of straight lines in the *f*--ln(*r*) graph. However, in the published work no link was made between the calculated parameters E_b and x_b, which describe the free energy landscape and the according mechanisms.

The loading rate r_f is defined macroscopically as the increase in force over time, and can be seen microscopically as the rupturing force divided by the time for bond breaking (reciprocal of the off-rate) (37):

$$r_f = \frac{\Delta f}{\Delta t} = \frac{k_B \cdot T}{x_B} \cdot \omega_0 \cdot \exp\left(- \frac{\left(E_b - f \cdot x_b \cdot \cos(\theta)\right)}{k_B \cdot T} \right). \qquad \textbf{[12]}$$

The relation between the loading rate and the pulling speed is as follows:

$$r_f = K_m \cdot v\,. \qquad \textbf{[13]}$$

Here, K_m is the molecular spring constant. We receive this relation by the following derivation: For the case of single AHs and also most other proteins, the first regime (until the AP is reached) in the force displacement curve is a linear increase in force (see Fig. 2). K_m is a measure for the slope in the first regime at that particular pulling speed and can be calculated as follows: $K_\mathrm{m} = F_{AP} / x_{AP}$. The pulling speed is the ratio of the displacement until the AP is reached and the time until this happens: $v = x_{AP} / t_{AP}$ (Eq. **2**). By multiplying the velocity *v* with the molecular constant K_m we receive, as indicated in Eq. **7**, the loading rate r_f , which is consequently the force until the AP is reached (beginning of bond breaking) divided by the necessary time to generate this force.

Interestingly, because of viscous friction, K_m itself is a function of the pulling velocity and the geometrical conformation of the protein. Most important for our theory is that in

SMD simulation as well as most single-protein experiments, the pulling speed is the controlled parameter.

In contrast to the phenomenological model, Szabo, Hummer, Dudko, and coworkers (45-49) follow a slightly different approach in their "microscopic theory." They build their theory on the assumption of only one transition state. This transition state not only is lowered with increasing external force f applied to the molecule (similar to the phenomenological theory), but also is moved along the reaction coordinate toward the equilibrium and eventually vanishes, when the barrier disappears (46). This results in a curved instead of a straight line in the f--log(v) space.

To derive the right information about the free energy landscape of the equilibrated system, the AFM pulling experiment or the SMD simulation (both nonequilibrium processes) need to be repeated several times, which might be extremely "expensive," especially for simulations . The data need to be averaged out afterward, by applying the Jarzynski identity, which postulates that the thermodynamic free energy difference of two states equals the work along the nonequilibrium trajectory (50, 51).

The microscopic theory postulates that the simple, straightforward phenomenological regime is a very good approximation, but only over a certain magnitude of pulling velocities. It leads to overestimations of the off-rate at small pulling speeds. The more accurate microscopic approach is in contrast to the more complex phenomenological model, as an additional parameter (the free energy of activation) is necessary.

Some of the advantages and disadvantages of both approaches were recently summarized in refs. 52-54. Other theories related to this topic were published recently (see refs. 55-58).

Definition of Robustness in Biology

In calculating robustness, we use the definition of robustness as parameter insensitivity, postulated by Kitano (59, 60). Details regarding this definition are provided in *Materials and Methods*.

We note that other robustness perspectives in biological materials are adaptation to environmental changes as well as graceful degradation (59, 60).

Background Information on Pareto's Efficiency

Pareto's efficiency rule found broad acceptance and application in many social, economic, political, and natural phenomena. For example, 80% of the wealth is concentrated in 20% of the entire population (61). Similarly, 80% of the revenues, software usage, or published papers typically stem from 20% of all customers, programmed code, or authors (61). The 80/20 rule is an empirical law that is found in many natural phenomena.

The theoretical foundation of this principle was reported by Chen *et al.* (62, 63). They state that this particular distribution is the result of the probability of a new entry, which quantifies the height of the entry barrier. Most importantly, it was shown that the probability of a new entry has an inverse relation with the level of usage concentration. This leads to a concurrence between existing entities and the addition of new entities, leading to the characteristic 80/20 distribution.

This concept can be applied to explain the particular molecular features of the AH structure: The more robust the structure becomes with each additional HB (see Fig. 5), the higher is the barrier to implement an additional HB, because each HB introduces an additional "cost" due to increased material use, that is, the additionally generated weight and additionally required volume. However, the increase of robustness decreases rapidly, so that an optimal balance is found at the characteristic 80% mark. Our analysis therefore suggests that the Pareto rule may also be relevant in explaining the structure of AH proteins in light of their robustness.

1. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002) *Molecular Biology of the Cell* (Taylor & Francis, Oxford, United Kingdom), 4th Ed.

2. Gruber M, Lupas AN (2003) *Trends Biochem Sci* 28:679-685.

3. Bryson JW, Betz SF, Lu HS, Suich DJ, Zhou HXX, Oneil KT, Degrado WF (1995) *Science* 270:935-941.

4. Kirshenbaum K, Zuckermann RN, Dill KA (1999) *Curr Opin Struct Biol* 9:530-535.

5. Ball P (2005) *Nanotechnology* 16:R1-R8.

6. Schwaiger I, Sattler C, Hostetter DR, Rief M (2002) *Nat Mater* 1:232-235.

7. Akkermans RLC, Warren PB (2004) *Philos Trans R Soc London Ser A* 362:1783-1793.

8. Root DD, Yadavalli VK, Forbes JG, Wang K (2006) *Biophys J* 90:2852-2866.

9. Cieplak M, Hoang TX, Robbins MO (2002) *Proteins Struct Funct Genet* 49:104-113.

10. Rohs R, Etchebest C, Lavery R (1999) *Biophys J* 76:2760-2768.

11. Bornschlogl T, Rief M (2006) *Phys Rev Lett* 96:118102.

12. Mitsui K, Nakajima K, Arakawa H, Hara M, Ikai A (2000) *Biochem Biophys Res Commun* 272:55-63.

13. Hanke F, Kreuzer HJ (2006) *Phys Rev E* 74:031909.

14. Wolgemuth CW, Sun SX (2006) *Phys Rev Lett* 97:248101.

15. Forman JR, Clarke J (2007) *Curr Opin Struct Biol* 17:58-66.

16. Brockwell DJ (2007) *Curr Nanosci* 3:3-15.

17. Randles LG, Rounsevell RWS, Clarke J (2007) *Biophys J* 92:571-577.

18. Paramore S, Voth GA (2006) *Biophys J* 91:3436-3445.

19. Ortiz V, Nielsen SO, Klein ML, Discher DE (2005) *J Mol Biol* 349:638-647.

20. Finke JM, Jennings PA, Lee JC, Onuchic JN, Winkler JR (2007) *Biopolymers* 86:193-211.

21. Day R, Daggett V (2007) *J Mol Biol* 366:677-686.

22. Ananikian N, Ananikyan L, Artuso R (2007) *Phys Lett A* 360:615-618.

23. Lucas A, Huang L, Joshi A, Dill KA (2007) *J Am Chem Soc* 129:4272-4281.

24. Wang N, Stamenovic D (2002) *J Muscle Res Cell Motil* 23:535-540.

25. Mucke N, Kreplak L, Kirmse R, Wedig T, Herrmann H, Aebi U, Langowski J (2004) *J Mol Biol* 335:1241-1250.

26. Helfand BT, Chang L, Goldman RD (2004) *J Cell Sci* 117:133-141.

27. Moir RD, Spann TP (2001) *Cell Mol Life Sci* 58:1748-1757.

28. Wilson KL, Zastrow MS, Lee KK (2001) *Cell* 104:647-650.

29. Luhrs T, Ritter C, Adrian M, Riek-Loher D, Bohrmann B, Doeli H, Schubert D, Riek R (2005) *Proc Natl Acad Sci USA* 102:17342-17347.

30. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) *J Mol Biol* 112:535-542.

31. Mostaert AS, Higgins MJ, Fukuma T, Rindi F, Jarvis SP (2006) *J Biol Phys* 32:393-401.

32. Broberg KB (1990) *Cracks and Fracture* (Academic, New York).

33. Freund LB (1990) *Dynamic Fracture Mechanics* (Cambridge Univ. Press, Cambridge, UK).

34. Buehler MJ (2006) *J Comput Theor Nanosci* 3:603-623.

35. Ackbarow T, Buehler MJ (2007) *J Mater Sci* 42:8771-8787.

36. Lacks DJ (2005) *Biophys J* 88:3494-3501.

37. Evans EA, Calderwood DA (2007) *Science* 316:1148-1153.

38. Bayas MV, Leung A, Evans E, Leckband D (2006) *Biophys J* 90:1385-1395.

39. Evans E (2004) *Abstracts of Papers of the American Chemical Society* 227:U469-U470.

40. Evans E (2001) *Annu Rev Biophys Biomol Struct* 30:105-128.

41. Evans E, Leung A, Hammer D, Simon S (2001) *Proc Natl Acad Sci USA* 98:3784-3789.

42. Evans EB (1999) *Biophys Chem* 82:83-97.

43. Merkel R, Nassoy P, Leung A, Ritchie K, Evans E (1999) *Nature* 379:50-53.

44. Evans E, Ritchie K (1997) *Biophys J* 72:1541-1555.

45. Dudko OK, Hummer G, Szabo A (2006) *Phys Rev Lett* 96:108101.

46. Dudko OK, Mathe J, Szabo A, Meller A, Hummer G (2007) *Biophys J* 92:4188-4195.

47. Hummer G, Szabo A (2003) *Biophys J* 85:5-15.

48. Hummer G, Szabo A (2005) *Acc Chem Res* 38:504-513.

49. Hummer G, Szabo A (2001) *Proc Natl Acad Sci USA* 98:3658-3661.

50. Jarzynski C (1997) *Phys Rev E* 56:5018-5035.

51. Jarzynski C (1997) *Phys Rev Lett* 78:2690-2693.

52. Makarov DE (2007) *Biophys J* 92:4135-4136.

53. Li PC, Makarov DE (2003) *J Chem Phys* 119:9260-9268.

54. Schlierf M, Rief M (2006) *Biophys J* 90:L33-L35.

55. Gilli P, Bertolasi V, Pretto L, Ferretti V, Gilli G (2004) *J Am Chem Soc* 126:3845-3855.

56. Wiita AP, Ainavarapu SRK, Huang HH, Fernandez JM (2006) *Proc Natl Acad Sci USA* 103:7222-7227.

57. Maloney CE, Lacks DJ (2006) *Phys Rev E* 73:061106.

58. West DK, Olmsted PD, Paci E (2006) *J Chem Phys* 124:154909.

59. Kitano H (2002) *Nature* 420:206-210.

60. Kitano H (2002) *Science* 295:1662-1664.

61. Chen JCH, Chong PP, Chen YS (2001) *J Org Comput Electronic Commerce* 11:1-14.

62. Chen YS, Chong PP, Tong MY (1994) *Math Comput Model* 19:61-80.

63. Chen YS, Chong PP, Tong YG (1993) *Scientometrics* 28:183-204.