Frequency Factors in a Landscape Model of Filamentous Protein Aggregation

Alexander K. Buell,[1] Jamie R. Blundell,[2] Christopher M. Dobson,[3] Mark E. Welland,[1]
Eugene M. Terentjev,[2,*] and Tuomas P. J. Knowles[1,†]

[1]*Nanoscience Centre, University of Cambridge, J J Thomson Avenue, Cambridge CB3 0FF, United Kingdom*
[2]*Cavendish Laboratory, University of Cambridge, J J Thomson Avenue, Cambridge CB3 0HE, United Kingdom*
[3]*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom*

(Received 21 April 2009; revised manuscript received 28 December 2009; published 1 June 2010)

Using quantitative measurements of protein aggregation rates, we develop a kinetic picture of protein conversion from a soluble to a fibrillar state which shows that a single free energy barrier to aggregation controls the addition of protein molecules into amyloid fibrils, while the characteristic sublinear concentration dependence emerges as a natural consequence of finite diffusion times. These findings suggest that this reaction does not follow a simple chemical mechanism, but rather operates in a way analogous to the landscape models of protein folding defined by stochastic dynamics on a characteristic energy surface.

DOI: 10.1103/PhysRevLett.104.228101 PACS numbers: 87.14.em, 82.20.Db, 82.20.Kh

The phenomenon of amyloid formation from peptides and proteins has intricate connections with a variety of normal and aberrant biological processes [1], ranging from epigenetic information transfer [2] to protein misfolding disorders in humans [1,3]. Protein self-assembly is consequently at the focus of studies, ranging from detailed kinetic experiments [4–6] to coarse grained and atomic level simulations [7], to shed light on the mechanisms underlying the assembly process. In this Letter, we demonstrate that an analytical model considering Langevin dynamics of polymers on a characteristic energy surface captures the key experimental features of the growth of fibrillar protein aggregates and therefore establishes close connections between the phenomenon of amyloid growth and the landscape models of protein folding [8–10].

Amyloid growth occurs through the attachment of soluble precursor proteins to the ends of fibrillar nanostructures [5]. We describe protein molecules as Gaussian polymers [11] with N segments evolving in a potential G_1 which results from the interactions with the fibril end. The physical origin of this potential is in the direct electrostatic interactions between the charged protein and the fibril end, as well as in contributions from van der Waals and hydrophobic interactions; in addition the breakage and formation of noncovalent contacts within and between polypeptide chains as well as with solvent molecules will contribute to this potential. Further implicit contributions to G stem from the projection of the reaction trajectory onto polymer degrees of freedom in the absence of an explicit description of water degrees of freedom [12]. We describe the dynamics of such a system by the overdamped Langevin equation: $\zeta \partial_t x_i = -\partial_{x_i} G(x_1, \ldots, x_{3N}) + \xi_i(t)$, where the $3N$ coordinates of chain segments undergo stochastic motion on the energy surface $G(x_1, \ldots, x_{3N}) = \sum_{i=1}^{3N} G_1(x_i) + \kappa/2 \sum_{i=1}^{3N-3} (x_i - x_{i+3})^2$ given by the connectivity-enforcing harmonic potentials and the external potential G_1. The thermal noise ξ_i satisfies the fluctuation-dissipation theorem $\langle \xi_i \xi_j \rangle = 2 k_B T \zeta \delta_{ij}$. The reaction rate is then given by the total probability flux from one basin of the energy surface, corresponding to a fibril of length j and a separate monomer, $F_j + M$, to another basin describing the energetically more favorable fibril of length $j+1$: F_{j+1} (Fig. 1). This flux can be computed by considering the corresponding Fokker-Planck equation

$$\partial_t \psi = -\mathrm{div} J, \quad \text{with} \quad J_i = -D[\partial_{x_i} \psi + \beta \psi \partial_{x_i} G], \quad (1)$$

where $\beta = (k_B T)^{-1}$ and $\psi(x_1, \ldots, x_{3N}; t)$ is the probability density of the polypeptide chain in a configuration with the segment coordinates $[x_i]$.

Equation (1) is a continuity equation, the right-hand side (rhs) of which represents the n-dimensional divergence of a diffusion current; following [13,14] we rewrite the current components as $J_i = -D e^{-\beta G} \partial_{x_i} \psi e^{\beta G}$ with the diffusion coefficient of the segments $D = k_B T/\zeta \sim 10^{-9}\,\mathrm{m^2\,s^{-1}}$

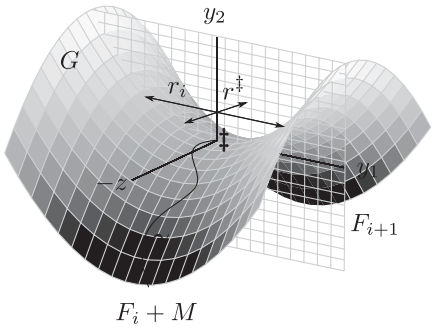

FIG. 1. Amyloid growth through diffusive sampling of an energy surface G. The reaction rate is determined through the flux from the reactants ($A := F_j + M$) to the product ($B := F_{j+1}$). An arbitrary path γ from A to B in the proximity of the saddle point ‡ is shown in black. The current at the saddle point is in the z direction as discussed in the text and the plane for the perpendicular degrees of freedom y_i is shown in gray.

[15]. We then consider the line integral between two points A, B connected by the path γ:

$$\int_\gamma e^{\beta G} J_i dx_i = -D \int_\gamma \partial_{x_i} \psi e^{\beta G} dx_i. \quad (2)$$

As the integrand on the rhs is a total derivative, the expression is simply equal to

$$\psi(B) e^{\beta G(B)} - \psi(A) e^{\beta G(A)} \approx -\psi(A) e^{\beta G(A)}, \quad (3)$$

where we assume that the fibril end B acts as a deep sink and so removes freely diffusing polymers from solution: $\psi(B) = 0$. Because of its exponential form, the integral on the left-hand side of (2) has the main contribution from the portion of the path near the saddle point ‡ between A and B. An appropriate linear transformation [16] represents the energy as

$$G \approx G^\ddagger - \frac{\alpha z^2}{2} + \frac{1}{2} \sum_{i=1}^{3N-1} \alpha_i y_i^2 + k_B T \ln(1 + \tau_R/\tau_D), \quad (4)$$

where the current is parallel to the z direction, Fig. 1. The additional entropic term $S_1 = k_B \ln(\Omega/\Omega_0) = -k_B \ln(1 + \tau_R/\tau_D)$ arises from the constraint imposed by the confined geometry of the fibril end which implies that the barrier crossing through diffusion in conformational space can only be attempted by one molecule at a time; therefore of all the possible microstates Ω_0 we need to consider only a subset $\Omega = \Omega_0 \tau_D/(\tau_R + \tau_D)$ selecting the attempts occurring during the fraction of time $\eta = \tau_D/(\tau_R + \tau_D)$ when the site is free, where τ_R is an average effective residency time and τ_D is a diffusive arrival time which shall be evaluated below. As in the steady state $\partial_{y_i} J_i + \partial_z J_z = 0$, the current density J_z does not vary around the saddle point in the reaction direction, we can compute the total flux Φ over the saddle point as the surface integral $\int dy_1 \ldots dy_{3N-1} J_z$ of J_z over the perpendicular degrees of freedom y_i at the saddle point:

$$\Phi = D \psi(A) e^{\beta G(A)} \frac{e^{-\beta[G^\ddagger - TS_1]}}{r^\ddagger} \prod_{i=1}^{3N-1} r_i, \quad (5)$$

where the length scales $r^\ddagger = \sqrt{2\pi k_B T/\alpha}$ and $r_i = \sqrt{2\pi k_B T/\alpha_i}$ result from the Gaussian integration.

We now examine the term $\psi(A) e^{\beta G(A)}$ from the rhs (3). Point A represents the reactants, i.e., the protein far away from the fibril and the free energy $G(A) \approx \kappa/2 \sum_{i=1}^{3N-3} (x_i - x_{i+3})^2$. Deviations from thermal equilibrium and the Boltzmann distribution are important only in the vicinity of the saddle point (in the region we call the reaction volume) and therefore the probability to find a free polymer chain in equilibrium around A is given by $\psi(A) = N_p e^{-\beta G(A)} Z^{-1} = c e^{-\beta G(A)} r_0^{-(3N-3)}$, where $c = N_p/V$ is the protein concentration in solution and the normalizing partition function is given by $Z = \int dx_1 \ldots dx_{3N} e^{-\beta G(A)} = V r_0^{3N-3}$, with the intrinsic length scale $r_0 = \sqrt{2\pi k_B T/\kappa} = \sqrt{2\pi/3} b_0$ proportional to the bond length b_0 of the free Gaussian polymer.

The total flux in Eq. (5) can now be written as

$$\Phi = D c r_0^2 \left(\prod_{i=1}^{3N-1} \frac{r_i}{r_0} \right) \frac{e^{-\beta G^\ddagger}}{r^\ddagger (1 + \tau_R/\tau_D)}. \quad (6)$$

Note that a broader barrier (greater $r^\ddagger$) results in a lower flux Φ and a flatter saddle point (greater width of the pass r_i) results in a larger flux Φ across the saddle. Similarly to earlier results derived through dimensionality arguments [4], the term $D c r_0^{-(3N-3)} (r^\ddagger)^{-1} \prod_{i=1}^{3N-1} r_i$ represents a flux into an effective reaction volume of linear dimension $r_{\text{eff}} := r_0^{-(3N-3)} (r^\ddagger)^{-1} \prod_{i=1}^{3N-1} r_i$. We therefore identify the diffusion time across this reaction volume as $\tau_D = (D c r_{\text{eff}})^{-1}$, and finally the flux takes the form

$$\Phi = \frac{D r_{\text{eff}} c}{1 + D r_{\text{eff}} c \tau_R} e^{-\beta G^\ddagger}. \quad (7)$$

This expression yields the limit $\Phi = D c r_{\text{eff}} e^{-\beta G^\ddagger}$ at low concentration and saturates at $\Phi = \tau_R^{-1} e^{-\beta G^\ddagger}$ at high concentration. The crossover between these two regimes occurs at $\tilde{c} := (D r_{\text{eff}} \tau_R)^{-1}$.

In order to gain quantitative insight into the magnitude of the frequency prefactor in Eq. (7) we require an estimate for r_{eff}. To achieve this let us present a self-consistency argument about the translational freedom of the polymer center of mass. Consider the time required for the complete rearrangement of internal degrees of freedom of a polymer that is no longer in translational motion as a whole, that is, effectively immobilized near the fibril end. In this case $\psi(A)$ no longer depends on concentration and is simply $1/r_0^{3N-3}$. The $(3N-1)$ terms of the product of the saddle-point integrals will now be reduced by a factor of effective volume for the chain center of mass translations, r_{CM}^3. In this case, the characteristic rate of the free polymer rearrangement would be given by $\Phi_0 = D r_{\text{eff}}/r_{\text{CM}}^3 = \tau^{-1}$, where $\tau \simeq k_B T N^2/\pi D \kappa = N^2 b_0^2/3\pi D$ is the Rouse time [11]. Equating $\Phi_0 = \tau^{-1}$ gives the estimate $r_{\text{CM}} \approx R_g/\pi$, where $R_g = \sqrt{N/6} b_0$ the radius of gyration of the polymer. We can further note that the diffusion rate of the polymer, $\tau_D^{-1} = D c r_{\text{eff}}$, is also expressed through the chain center of mass motion: $\tau_D^{-1} = D_p c r_{\text{CM}}$, where $D_p \approx D/N \simeq 10^{-10}$ m^2 s^{-1} is the Rouse diffusion constant of the center of mass [11]. This gives the relation $r_{\text{CM}} = N r_{\text{eff}}$. The frequency factor τ_D^{-1} can therefore be written as $\tau_D^{-1} \approx D_p c R_g/\pi$, i.e., a time of the order required for the coil center of mass to diffuse into a volume given by the coil size. The Kuhn length $b_0 \approx 1$ nm in polypeptide chains spans approximately 3 residues and, for example, for a protein with 50 residues we obtain the value of $r_{\text{eff}} = 0.3$ Å and $r_{\text{CM}} \approx 5.3$ Å.

This framework for diffusive attachment of proteins on to amyloid fibrils enables insight to be gained into the energy surface which directs the growth of such fibrils. In general, the elongation step of amyloid growth is not

directly accessible using standard assays, as protein aggregation in solution involves nucleation and fragmentation of fibrils simultaneously to their elongation [17]. Recently, however, methods have emerged which allow the elongation step to be monitored independently of other processes using biosensors [18–20]. In the present work we attached preformed fragments of amyloid fibrils to the surface of a quartz transducer [Fig. 2(a)(i)]; when the surface of the sensor is exposed to a monomer solution [Fig. 2(a)(ii)], the elongation of the fibrils [Fig. 2(a)(iii)] can be monitored through the changes in the resonant frequency of the quartz oscillator resulting from the nanogram mass changes consequent upon new molecules adding on to the aggregates [18,21]. The measurements highlight two different concentration regimes: an initial linear concentration dependence, and subsequently a saturation in the growth rate with increasing monomer concentration. In our measurements the growth rate v_{exp} is given by the mass change with time per unit area, which is directly proportional to the flux in Eq. (7): $v_{\mathrm{exp}} = \rho_{2D} M_I \Phi$, where the surface density of fibrils $\rho_{2D} \sim 10^{10}$ cm^{-2} in the case of insulin and $\rho_{2D} \sim 3 \times 10^9$ cm^{-2} for α lactalbumin has been estimated from AFM measurements [insets Figs. 2(b) and 2(c)]. A two-parameter fit to Eq. (7) using the parameters $\tilde{c} = (D r_{\mathrm{eff}} \tau_R)^{-1}$ and $v_{\mathrm{max}} = \tau_R^{-1} \rho_{2D} M_I e^{-G^\ddagger}$ is shown in Figs. 2(b) and 2(c). From these fitted values we can determine the values of the effective residency times for the two proteins; in the case of insulin (51 residues), $\tau_R = (D\tilde{c} r_{\mathrm{eff}})^{-1} \approx 43$ μs and for the longer sequence α lactalbumin (121 residues), $\tau_R = (D\tilde{c} r_{\mathrm{eff}})^{-1} \approx 678$ μs.

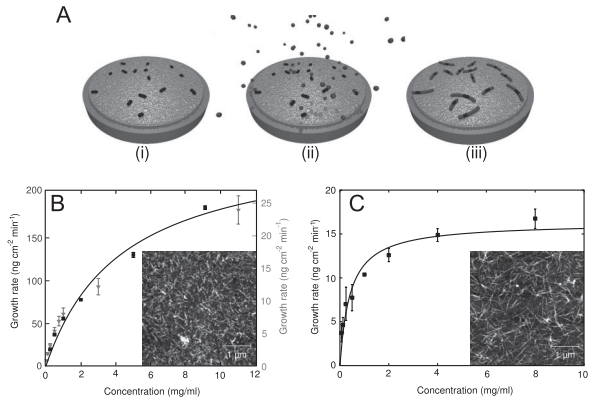

FIG. 2. (a) Schema illustrating the determination of amyloid growth rates through measurements of the mass increase of a fibril population elongating on a quartz crystal transducer, as described in the text. (b) Elongation rate of insulin fibrils as a function of the concentration of soluble insulin; rates under different conditions [with 100 mM NaCl and 20% CH_3COOH (squares) and data from [18] without salt (triangles)] can be scaled to collapse on to a master curve. A two-parameter fit to Eq. (7) gives $\tilde{c} = 4.5 \times 10^{23}$ m^{-3} (equivalent to 4.3 mg/ml for insulin with molar mass $M_I = 5800$ Da), and $v_{\mathrm{max}} = 264$ ng cm^{-3} min^{-1}. (c) analogous experiment carried out with bovine α lactalbumin. Here the fit to Eq. (7) yields $\tilde{c} = 1.8 \times 10^{22}$ m^{-3} (equivalent to 0.42 mg/ml for α lactalbumin with molar mass $M_I = 14200$ Da), and $v_{\mathrm{max}} = 16.3$ ng cm^{-2} min^{-1}. The insets in (b) and (c) show AFM micrographs of a sensor surface before growth from which the seed density can be measured.

The implications of the landscape model are illustrated in this Letter for the *in vitro* fibril assembly of two representative proteins which are widely used in biophysical studies of this phenomenon; insulin, in particular, being the first protein for which this self-assembly process was observed *in vitro* [22]. Characteristic features of the landscape model such as the existence of a linear growth rate at low protein concentrations and a concentration-independent growth rate at higher concentration, however, are observed for a wider class of proteins and growth conditions, including fibril assembly at physiological pH [5].

We can test some of the further predictions of the diffusive mechanism. As the reaction rate Φ is exponentially dependent on the single barrier $G^\ddagger$ and has only weaker dependencies on the other system parameters, we expect perturbations to the system state to result in growth rates with varying monomer concentration which can to a good approximation be scaled to collapse on to a master curve. To illustrate this idea, we compare in Fig. 2(b) insulin fibril growth under different conditions: in 10 mM hydrochloric acid and in 20% acetic acid with 100 mM NaCl. The growth rates for a given concentration of monomer differ by approximately 1 order of magnitude, but the curves can be scaled with a constant factor to collapse on to a master curve (Fig. 2) as predicted by Eq. (7).

This scaling does not emerge from simple chemical kinetics, and it is interesting to discuss briefly how the landscape picture differs from simple chemical kinetics characterized by instantaneous equilibration between reactants and the activated state. In order to account for the less than linear concentration dependence of the growth rates, the prevailing analysis of these types of reactions postulates a two-step mechanism, where a monomer first combines with a fibril end to form an intermediate in a bimolecular reaction with a rate linearly proportional to the monomer concentration; this intermediate then subsequently reorganizes in a concentration-independent step to yield the product. As has been noted by some authors [23], this type of process is formally analogous to the Michealis-Menten model of enzymatic reactions [10], the monomer taking the place of the substrate and the fibril end that of an enzyme. If we write the rates for the two steps as $k_1 c$ and k_2, the overall reaction rate will be given as the reorganization rate k_2 times the steady-state concentration of the intermediate: $v = k_2 c[c + (k_2 + k_{-1})/k_1]^{-1}$, where k_{-1} is a detachment rate. For low monomer concentrations, $c \ll (k_2 + k_{-1})/k_1$, the rate is linear in concentration, $v \approx c k_1$, whereas for high concentrations the rate is constant $v \approx k_2$, leading to an overall sublinear concentration dependence. It is clear, however, that for a true two-step process analogous to a Michaelis-Menten mechanism in general the parameters k_1 and k_2 depend on two different

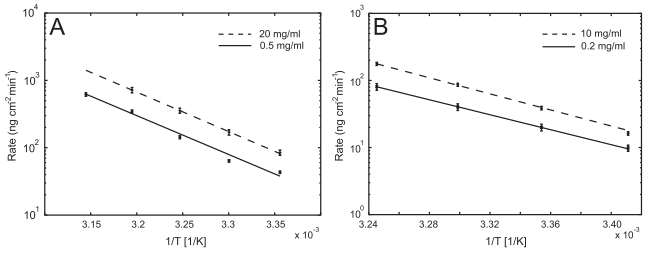

FIG. 3. Measurements of the enthalpic contribution $H^{\ddagger}$ to the activation barrier $G^{\ddagger} = H^{\ddagger} - TS^{\ddagger}$ for insulin and α lactalbumin in the linear low concentration regime (solid black line) and saturated high concentration regime (dashed gray line) through the relation $\log(\Phi) = H^{\ddagger} k_B^{-1} T^{-1} + \text{const}$. Measurements are performed in 20% acetic acid with 100 mM NaCl (insulin) and HCl pH = 1.2 with 100 mM NaCl (α lactalbumin).

barriers $G_1^{\ddagger}$ and $G_2^{\ddagger}$ and therefore growth rates would not be expected to follow a common master curve, at variance to what is observed experimentally (Fig. 2).

The predictions from simple chemical kinetics about the existence of two distinct energy barriers in the high and low concentration regimes as opposed to a single one in the diffusional process can also be directly probed. Biosensor measurements, where the frequency signal is sensitive to the elongation reaction, allow the enthalpic barrier to be measured under both the low concentration and high concentration regime. Interestingly, these activation enthalpies in both regimes are identical within experimental error [Fig. 3(a) averages over 4 independent measurements per concentration: 0.5 mg/ml: $\Delta H^{\ddagger} = 129 \pm 16$ kJ/mol, 20 mg/ml: $\Delta H^{\ddagger} = 115.6 \pm 11.6$ kJ/mol, Fig. 3(b) averages over 2 independent measurements per concentration: 0.2 mg/ml: $\Delta H^{\ddagger} = 102.0 \pm 4.6$ kJ/mol, 10 mg/ml: $\Delta H^{\ddagger} = 112.3 \pm 9.9$ kJ/mol). This finding suggests that in both cases the aggregation reaction is indeed limited by the same barrier. Interestingly, in cases where a true two-state Michaelis-Menten type process occurs, such as in the context of enzymatic processes [10,24], the activation enthalpies between the two steps can vary by a significant fraction of their value [25,26].

The knowledge of the diffusional prefactor further allows us to provide an estimate of the entropic contributions to the free energy barrier. Using the numerical values for the diffusion coefficient, the protein concentrations and $\tau_D^{-1} = D_p R_g c/\pi \approx 1600\ \text{s}^{-1}$ (insulin) and $\tau_D^{-1} = D_p R_g c/\pi \approx 260\ \text{s}^{-1}$ (α lactalbumin) with $R_g \approx 1$ nm $D_p \approx 10^{-10}\ \text{m}^2\,\text{s}^{-1}$, $c \approx 5.2 \times 10^{22}\ \text{m}^{-3}$ (insulin) and $c \approx 8.6 \times 10^{21}\ \text{m}^{-3}$ (α lactalbumin), we obtain entropies of activation $T\Delta S = 115 \pm 25$ kJ/mol (insulin) and $T\Delta S = 87 \pm 20$ kJ/mol (α lactalbumin), i.e., of a similar order of magnitude to the enthalpic term in the free energy—we therefore recover the entropy-enthalpy compensation phenomena characteristic of protein folding [10].

In conclusion, we have computed frequency factors for a landscape model of fibrillar protein self-assembly, and tested experimentally the predictions of this model for the *in vitro* assembly of two representative polypeptide sequences. The availability of estimates for the frequency factors governing amyloid growth kinetics allows the key characteristics of the energy landscape underlying this process to be defined from kinetic measurements.

This work was supported by the U.K. EPSRC, IRC in Nanotechnology, Nokia Research, St. John's College and Magdalene College, Cambridge. We thank David Chandler, Michele Vendruscolo, Benjamin Schlein, and Chris Waudby for helpful discussions.

*emt1000@cam.ac.uk
†tpjk2@cam.ac.uk

[1] C. M. Dobson, Nature (London) **426**, 884 (2003).
[2] R. B. Wickner, Science **264**, 566 (1994).
[3] D. J. Selkoe, Nature (London) **426**, 900 (2003).
[4] Y. Kusumoto, A. Lomakin, D. B. Teplow, and G. B. Benedek, Proc. Natl. Acad. Sci. U.S.A. **95**, 12 277 (1998).
[5] S. R. Collins, A. Douglass, R. D. Vale, and J. S. Weissman, PLoS Biol. **2**, e321 (2004).
[6] W.-F. Xue, S. W. Homans, and S. E. Radford, Proc. Natl. Acad. Sci. U.S.A. **105**, 8926 (2008).
[7] J. Gsponer and M. Vendruscolo, Protein Peptide Lett. **13**, 287 (2006).
[8] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).
[9] T. Lazaridis and M. Karplus, Science **278**, 1928 (1997).
[10] A. Fersht, *Structure and Mechanism in Protein Science* (W. H. Freeman, New York, 1999).
[11] *The Theory of Polymer Dynamics* (Oxford University Press, Oxford, 1986).
[12] Pieter Rein ten Wolde and David Chandler, Proc. Natl. Acad. Sci. U.S.A. **99**, 6539 (2002).
[13] S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943).
[14] J. S. Langer, Ann. Phys. (N.Y.) **54**, 258 (1969).
[15] A. Polson, Biochem. J. **31**, 1903 (1937).
[16] R. Landauer and J. A. Swanson, Phys. Rev. **121**, 1668 (1961).
[17] T. P. J. Knowles, C. A. Waudby, G. L. Devlin, S. I. A. Cohen, A. Aguzzi, M. Vendruscolo, E. M. Terentjev, M. E. Welland, and C. M. Dobson, Science **326**, 1533 (2009).
[18] T. P. J. Knowles, W. Shu, G. L. Devlin, S. Meehan, S. Auer, C. M. Dobson, and M. E. Welland, Proc. Natl. Acad. Sci. U.S.A. **104**, 10 016 (2007).
[19] K. Hasegawa, K. Ono, M. Yamada, and H. Naiki, Biochemistry **41**, 13 489 (2002).
[20] J. A. Kotarek, K. C. Johnson, and M. A. Moss, Anal. Biochem. **378**, 15 (2008).
[21] M. B. Hovgaard, M. Dong, D. E. Otzen, and F. Besenbacher, Biophys. J. **93**, 2162 (2007).
[22] D. F. Waugh, J. Am. Chem. Soc. **68**, 247 (1946).
[23] T. Scheibel, J. Bloom, and S. L. Lindquist, Proc. Natl. Acad. Sci. U.S.A. **101**, 2287 (2004).
[24] J. R. Knowles, Nature (London) **350**, 121 (1991).
[25] D. A. Brant, L. B. Barnett, and R. A. Alberty, J. Am. Chem. Soc. **85**, 2204 (1963).
[26] J. R. Silvius, B. D. Read, and R. N. McElhaney, Science **199**, 902 (1978).