Non-exponential kinetics of unfolding under a constant force

Samuel Bell and Eugene M. Terentjev[1]
Cavendish Laboratory, University of Cambridge, J.J. Thomson Avenue, Cambridge, CB3 0HE, U.K.

We examine the population dynamics of naturally folded globular polymers, with a super-hydrophobic 'core' inserted at a prescribed point in the polymer chain, unfolding under an application of external force, as in AFM force-clamp spectroscopy. This acts as a crude model for a large class of folded biomolecules with hydrophobic or hydrogen-bonded cores. We find the introduction of super-hydrophobic units leads to a stochastic variation in the unfolding rate, even when the positions of the added monomers are fixed. This leads to the average non-exponential population dynamics, which is consistent with a variety of experimental data and does not require any intrinsic quenched disorder that was traditionally thought to be at the origin of non-exponential relaxation laws.

I. INTRODUCTION

Self-assembly and controlled unfolding of biological macromolecules is a fundamental process in cell biology, and is crucial to life itself. It has proven to be a rich area of research, and while there has been much progress and understanding achieved, there is still much left to discover. Part of the reason for this is the unique set of interactions for each protein sequence, and the resulting complexity of the phase space, as well as the many mechanisms for denaturation, including temperature, pH, force and enzymatic action.

The response of biomolecules to mechanical forces has been a popular area of study within biophysics[1]. The sensitivity of experimental tools like optical tweezers and atomic force microscopy (AFM), and their ability to work in a 'wet' environment, have made them ideal for probing biology with mechanical forces at a molecular level. They have been used extensively to characterise the unfolding kinetics of a range of biomolecules[2,3]. As well as working with DNA, many experiments have focused on compact globular structures, such as the Ig domain, an important subdomain of several proteins, including titin[4–6].

AFM experiments can be performed in the position-clamp mode, where the force is measured by the cantilever[5], or in the force-clamp mode, where a constant force is applied and the resulting extension measured. In force-clamp experiments, biomolecules typically show all-or-nothing transitions between folded and unfolded states[7,8], meaning that denaturation occurs abruptly and completely once a critical force is reached in the case of force-ramp, or a characteristic time is reached if a constant force is applied. The transition from stable to metastable configurations with increasing applied force has been treated theoretically, and indeed found to be first order[9–11].

Initially, the kinetics of transitions were studied using a two-state system, with a potential barrier modified by the introduction of force[12]. When the final state is much lower in energy than the initial state, such a reaction is essentially irreversible, and the survival probability of the initial state decays exponentially with time. This is the regime of the original Kramers problem of escape over the barrier[13].

Recent analysis of ubiquitin unfolding data from single-molecule pulling experiments has found this simple model fails at describing the experimental data[14,15]. Strongly non-exponential kinetics have been also found in other biological systems, such as ligand binding in myoglobin, and are usually attributed to random variations in molecule conformations[14]. These ensemble variations may have an additional time-dependence. As such, some early work thought of the free energy landscape as a collection of native globular states (of similar but not identical energies) and extended unfolded states, separated by a single energy barrier, in a globally connected energy landscape[16,17]. This idea of heterogeneity can also be modeled using disorder theory[18]. The variation of unfolding rate in the ensemble of molecules can be described by introducing a stochastic internal parameter, which could follow a chosen pattern of static or dynamic disorder.

When the rate of change of internal parameter is much slower than the unfolding rate, we can regard it as fixed for each molecule, and consider the regime of static, or quenched, disorder. In this limit, Kuo et al.[19] introduced a Gaussian variation in the barrier height to be surmounted by individual molecules. However, Lannon et al.[15] subsequently found this model to be a poorer fit to the data than a stretched exponential distribution.

When the internal parameter can no longer be regarded as fixed on the time-scale of unfolding, one enters the regime of dynamical disorder[18]. Recent work by Hyeon et al.[20] have looked at the role of dynamical disorder in the unfolding kinetics of macromolecules, and other work has had some success at fitting ubiquitin unfolding data using a generalised Langevin equation with fractional Gaussian noise[21]. These approaches are still relatively new, and have yet to be fleshed into a physical model where parameters are obtained from physical quantities.

In this paper we take inspiration from earlier work by Geissler and Shakhnovich[22]. They studied the mechanical response of random copolymers in equilibrium (a problem in static disorder). In attempting to adapt their analysis to look at the unfolding kinetics, we found that a much simpler model is already showing non-exponential kinetic features. Even the introduction of only a single

inhomogeneous residue into an otherwise homogeneous polymer chain can lead to a significant change in the unfolding kinetics. Although the position of this residue could be fully prescribed by the sequence, we found that its exposure to solvent has to be treated as a random variable. This variable affects the unfolding rate constant, making the ensemble-average phenomena highly non-exponential. This could provide a crude but effective model for hydrophobic cores of biomolecules.

To compare our results and predictions to experimental data, a set of realistic model parameters will need to be chosen. For these, we will follow the directly related experiments by Fernandez and Brujic[7,14,15,19], where ubiquitin or Ig-like I27 domains of titin were unfolded at constant force. Hence we will take a polymer chain with $N = 100$ residues, as a value close to the above proteins. Since the minimal force these authors were using was $\sim 90\,\mathrm{pN}$, and we have earlier obtained the theoretical value for the critical unfolding force f^* in this regime[11], quite a reasonable value of the globule 'hydrophobic strength' $u = 5k_BT$ has to be chosen (assuming the characteristic size of an amino acid residue is $b \sim 0.3\,\mathrm{nm}$); at this strength of globule the critical force for a homopolymer would be $f^* \approx 4.48(k_BT/b) = 59.7$ pN, acceptably slightly below what was used in experiment.

II. POLYMER MODEL

We consider a model polymer consisting of a single type of hydrophobic monomer, with just one additional 'core' monomer, which we take to be much harder to remove from the bulk of a polymer globule than all the others. This could be a toy model for the hydrophobic core of a protein, or a particular sequence of residues that binds to its matching counterpart stronger than others. We shall assume that the position M of this particular monomer in the polymer sequence $(1 \ldots N)$ is known, and remains fixed. The reason why we only consider one such 'locked' monomer, and not several (as would be the case in a real folded protein), is simply because the qualitative features emerging due to the presence of such an internal 'lock' are clear already in the singlet case – while the calculations are kept simpler and transparent.

Given poor solvent conditions, we expect that the unstretched polymer is folded in equilibrium, and will form a compact globule. Following earlier works, we model the free energy of the globule as the sum of favourable monomer-monomer pair interactions in the bulk, and unfavourable monomer-solvent interactions at the surface. To minimise the free energy, the core monomer will always be buried deep within the bulk of the globule, and provide no contribution to the surface energy. The free energy of the globule can be written as a function of the number of monomers in the globule N_g, the force-induced extension of the globule x, the bulk energy of the monomers, $-u$, and the additional binding energy of

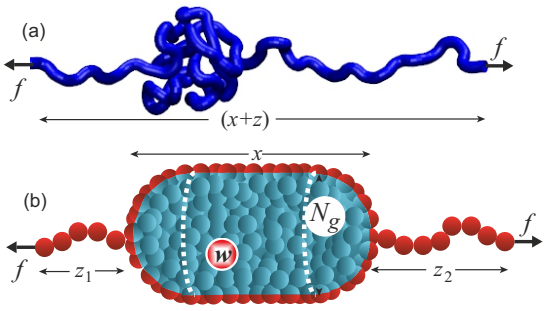

FIG. 1. Spherocylinder model of the globular polymer in poor solvent, being stretched by an application of force f to its two ends. (a) A tube representation of a chain in Brownian dynamics simulation; (b) a scheme illustrating the force-induced asymmetry of the remaining globule with N_g monomers, one of which has a stronger binding energy in the globule, $-(u+w)$. Chain can be removed from both ends when the polymer is unravelling.

the 'lock' monomer w, such that the total energy of this residue in the bulk of the globule is $-(u+w)$. This gives the free energy of the globule:

$$\begin{aligned} F_g(N_g, x) &= -N_g u - w + \frac{A(x)}{b^2}\frac{u}{2} \\ &= -N_g u - w + \left(\frac{2N_g b}{3x} + \sqrt{\frac{\pi N_g x}{b}}\right) u, \end{aligned} \tag{1}$$

where $A(x)$ is the surface area of the deformed globule, and b is the monomer size. Following an earlier work[11], for efficient analytical treatment we take the shape of the globule to be a spherocylinder (see Fig. 1), with a constant volume $N_g b^3$. In Eq. (1) we assume that any monomers on the surface are half-solvated, with the binding energy $-u/2$, and that a completely solvated monomer in the exposed chain segments has the potential energy level of zero. The term $-w$ may or may not be present in F_g, depending on whether the w-monomer is still inside, or has been removed from the globule into the stretched-out segments.

The surface energy $A(x)$ has an inflection point, which indicates an instability in the globule past certain extensions (see[11] for a detailed discussion). In the force-clamp regime, this manifests in the globule's inability to supply restoring forces beyond a certain critical value of its extension, x_{crit}. For a given applied force f, the minimum size of the globule to provide an equal and opposite restoring force is given by:

$$N_g(\mathrm{stab}) = N - N_s = \frac{2}{3\pi^2}\left(\frac{16fb}{3u}\right)^3, \tag{2}$$

Globules below this size cannot sustain the applied force in equilibrium. For compactness of later expressions, we define the amount of chain that must be removed to reach this threshold of stability, N_s. We posit that once a globule has been reduced below its smallest stable size, it will

rapidly unfold and extend to a chain state, with little regard for features of the globule. This issue will become important later, when we start calculating the rate at which a globule transforms into an extended chain.

Since we have chosen to measure the potential energy of fully solvated monomers as zero (in the poor solvent, this leads to the negative potential energy of monomers inside the globule, Eq. (1), the remaining free energy of the expanded chain with contour length $L = b(N - N_g)$ is entirely dependent on the chain's properties. A general form for this free energy in terms of the chain's end-to-end extension, z, was derived by Blundell[23], and here we reduce the full expression to the flexible chain limit, where the persistence length l_p is small – of the order of the monomer size, $l_p \approx b \ll L$:

$$F_{ch}(z) = \frac{2k_BTL}{\pi b\,(1-(z/L)^2)} - \frac{2k_BTL}{\pi b}. \qquad (3)$$

The constant term is added to fix the energy of the chain at zero extension: $F_{ch}(0) = 0$. This expression is valid across the different regimes of stretching as the chain is being pulled. For small deformations, Eq. (3) reduces to the entropic spring expression, $F_{ch}(z) \approx (2k_BT/\pi Lb)z^2$. In the limit of large stretching, $z \to L$, the expression shows the well-discussed divergence for an inextensible chain, known as the Marko-Siggia or Fixman-Kovac limit[24,25].

In the force-clamp mode of a typical AFM stretching experiment, we need to work with the Gibbs free energy,

$$G(f) = F[x_{eq}(f)] - f x_{eq}(f), \qquad (4)$$

where $x_{eq}(f)$ comes from the condition of mechanical equilibrium. To calculate the Gibbs free energy of the whole system, we must add the contributions from any globular part with N_g monomers, $G_1(f)$, any chain parts, $G_2(f)$ with a total of $N_c = N - N_g$ monomers, and whether or not the w-monomer is solvated (i.e. extracted from the globule):

$$G(f, N, N_g) = G_1(f, N_g) + G_2(f, N - N_g) - (w), \qquad (5)$$

To find an expression for the Gibbs free energy of the remaining globule, G_1, we need to use Eq. (1) in (4). In order to obtain a simple and compact form of the equilibrium extension $x_{eq}(f)$, we need to make the assumption that the globule only suffers a small deformation when a force is applied; this will be the case when the solvent is sufficiently poor. Then the deformation is linear with applied force,

$$x_{eq}(f) = \sqrt[3]{\frac{6}{\pi}N_g b^3} + \frac{16b^2}{9\pi u}f, \qquad (6)$$

where the first term is the diameter of spherical globule.

For the pulled-out chain segment, we need to impose the condition of mechanical equilibrium as well, finding the tensile force

$$f = -\frac{\partial F_{ch}(z)}{\partial z} = -\frac{4k_BT}{\pi l_p}\left(\frac{z}{L}\right)\frac{1}{\left(1-(z/L)^2\right)^2}, \qquad (7)$$

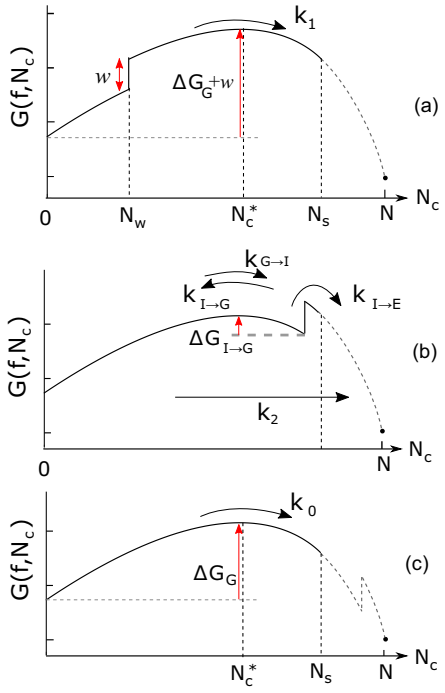

FIG. 2. Gibbs free energy curves $G_1 + G_2$, plotted against the number of monomers in the extended chain (N_c) for a fixed value of pulling force f. We can split N_c into two parts, so exposure of w-monomer can be at different values of N_c. Exposure of the w-monomer at $N_c = N_w$ leads to a jump w in the free energy. In (a), this exposure happens before the top of the free energy barrier, where $N_c = N_c^*$. In (b), exposure occurs after the free energy peak, forming a metastable intermediate state. In (c), the exposure happens once the globule is already unstable, $N_w > N_s$, where exposure is not affecting the unfolding rate.

and inverting it to obtain the equilibrium extension $z_{eq}(f)$ to be then used in Eqs. (3) and (4). Since all expressions depend strictly on the ratio (z/L), we can infer that the total extension in response to an applied force is proportional to the contour length L of the exposed chain. As a consequence, the (negative) Gibbs free energy of the chain, $G_2(f)$, is proportional to the contour length of the chain, $L = N_c b$, where the number of solvated monomers is $N_c = N - N_g$. As a result, we can separate the effective chemical potential factor: $G_2(f) = N_c\mu(f)$. In the limit of large extension, when $(L - z)/L \ll 1$, this chemical potential takes the form[11]:

$$\mu(f) \approx \frac{2k_BT}{\pi}\left(\frac{1}{1-\xi^2} - 1\right) - fb\xi\ . \qquad (8)$$

Here the dimensionless parameter ξ is a measure of the fractional extension of the chain relative to its contour length:

$$\xi = \frac{z_{eq}}{L} = 1 - \frac{1}{\sqrt{1 + \pi fb/k_BT}}\ . \qquad (9)$$

This form of the free energy of the extended portion of the polymer has an important consequence. We can

in principle divide any chain of N_c monomers into two separate segments linked in series (see Fig. 1), and the Gibbs free energy will be the same, regardless of how we distribute the N_c monomers between the two parts. This means that monomers can be pulled out of the globule from either end of the chain, forming two tails of length z_1 and $z_2 = z - z_1$, without any additional free energy penalty. This means that the 'lock', positioned at $M < N/2$ from one end of the chain, could be exposed to the solvent for any N_c in the range $M \leq N_c \leq N$. Let us denote which monomer it is to leave the globule as N_w (meaning that the exposed chain segments have the total of $N_c = N_w$ monomers when the 'lock' is pulled out). The exposure of the w-monomer will manifest itself as a discontinuous jump in the free energy profile $G(N_c, N_w)$ at $N_c = N_w$. There are three qualitatively different situations, depending on the value of N_w, illustrated in Fig. 2 which sketches the dependence of $G(N_c)$ for a fixed value of applied force f (the dashed region on these plots represents the unstable globule, with $N_g \leq N - N_s$, see Eq. 2). The three regimes are:

(a) The lock could be exposed before the free energy barrier ($N_w \leq N_c^*$). This results in a simple two-state kinetics with an enhanced barrier.

(b) The lock could be exposed past the barrier, in the stable region of the free energy curve ($N_c^* < N_w < N_s$). In this case, we have a meta-stable intermediate state, and a resulting three-state kinetics.

(c) The lock is exposed in the unstable region ($N_w > N_s$). This means that the system effectively does not see the free energy jump and the globule unfolds as if there was no w-monomer.

III. DISTRIBUTION OF BARRIER SHAPES

We now show that the value of N_w (the total length of the pulled-out chain for which the w-monomer gets exposed) is a random variable. The randomness arises purely because there is a choice in how to distribute the pulled-out segments z_1 and z_2 (see Fig. 1) before the 'locked' monomer is at the surface. We derive the probability to have the w-monomer pulled out, given its sequential position $M < N/2$ along a chain of N monomers: $P(N_w|N, M)$. Let us assume that each time a monomer is pulled from the globule, we have an equal probability $p = 1/2$ of it being pulled from the left or the right. $P(N_w|N, M)$ is therefore similar to the binomial distribution, with the difference in that we could remove the w-monomer by approaching from the left or the right, equivalently. To remove the w-monomer at N_w from the shorter end, we must first remove $M - 1$ ordinary monomers in a total of $N_w - 1$ exposure events. Let us call the probability of that P_S; this is given by the binomial expression $P_S = 2^{-(N_w-1)}\binom{N_w-1}{M-1}$. In the same way, to remove the w-monomer from the longer end, we must remove $N - M$ ordinary monomers first, with probability $P_L = 2^{-(N_w-1)}\binom{N_w-1}{N-M}$. The total probability is then given by the sum $\frac{1}{2}(P_S + P_L)$, which takes the form:

$$P(N_w|N, M) = \frac{1}{2^{N_w}}\left[\binom{N_w - 1}{M - 1} + \binom{N_w - 1}{N - M}\right]. \quad (10)$$

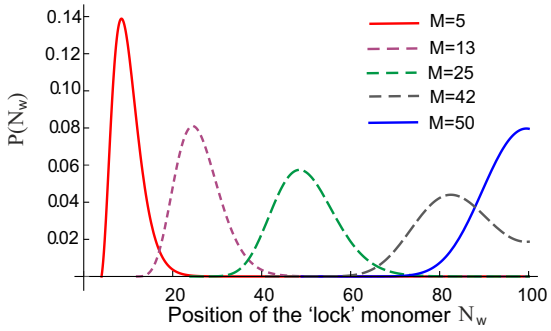

FIG. 3. The likelihood of the jump occurring at N_w is given by $P(N_w)$. The distribution $P(N_w)$ is plotted here for $N = 100$ and several values of the w-monomer's position, M, in the sequence along the chain.

Figure 3 gives the shape of this distribution for various positions of the w-monomer in the chain sequence. $P(N_w|N, M) = 0$ for $N_w < M$: one cannot remove the w-monomer without first removing a certain number of ordinary monomers. It is normalised, as expected for a probability distribution: $\sum_{N_w} P(N_w|N, M) = 1$.

We will shortly determine how the position of the jump N_w determines the rate constant of the transition to the unfolded state. Let us split the population according to each unfolding trajectory, defined by the value of N_w. These sub-populations will decay according to their specific rate constant $k(N_w)$, and the number of folded polymers $n(N_w)$ will vary according to the simple exponential rate law:

$$\dot{n}(N_w, t) = -k(N_w)n(N_w, t)\ . \quad (11)$$

However, in experiment, we cannot distinguish between different sub-populations. Instead, we track how the total population evolves over time. To calculate this, we have to weight each sub-population by its fraction, i.e. average $\langle n(N_w)\rangle$. The fraction of the population in the sub-population $n(N_w)$ is given by $P(N_w)$. Accordingly, the rate equation for the entire population takes the form:

$$\langle n(t, M)\rangle = n_0 \sum_{N_w=1}^{N} P(N_w|N, M)e^{-k(N_w)t}. \quad (12)$$

In order to efficiently evaluate the ensemble-averaged population kinetics for the chains unfolding by the pulling force, we need to make an approximation for the probability distribution of the random variable N_w. This discrete distribution was given by the Eq. (10) and plotted (for $N = 100$) in Fig. 3. It is clear that, although the

binomial expressions involved are skewed, even at these moderate chain lengths the Central Limit Theorem holds, and we can approximate the distribution as a Gaussian with good accuracy. The conversion from a strict binomial distribution of a random variable X with $N_w - 1$ 'attempts' corresponds to replacing it with a continuous Gaussian[26] with the mean $y = (N_w - 1)/2$ and variance $\sigma^2 = (N_w - 1)/4$. For the binomial probability P_S contributing to the first term in Eq. (10), the variable is $X = M - 1$; for the distribution P_L in the second term of Eq. (10), the variable $X = N - M$. Put together, these two expressions (Gaussian in the variable X) produce the final continuous probability density of our actual random variable N_w:

$$P(N_w) = \sqrt{\frac{1}{\pi y}} \left(e^{-\frac{(M-1-y)^2}{y}} + e^{-\frac{(N-M-y)^2}{y}} \right), \quad (13)$$

where the shorthand $y = (N_w - 1)/2$ is employed. This approximate expression turns out to be indistinguishable from the exact curves in Fig. 3, plotted for $N = 100$. The compact analytical expression (13) can now be used to calculate the observed population kinetics, which replaces Eq. (12):

$$\langle n(t) \rangle = \int_0^{\frac{N-1}{2}} dy P(y) e^{-k(y)t}. \quad (14)$$

IV. RATE CONSTANTS

The equilibrium rate constants $k(N_w)$ could be found using a Kramers-like method, first explored by Brinkman for the case of two-well potential[27]. The expression for the rate constant is given by a steady-state limit of the Ornstein-Uhlenbeck theory for the mean first passage time[28], when the ensemble distribution in the initial potential well had enough time to equilibrate before the average transition occurs:

$$k = \frac{k_B T}{\gamma} \frac{1}{\int_{\text{well}} e^{-\beta G(N_c)} dN_c \int_{\text{barrier}} e^{\beta G(N_c)} dN_c}, \quad (15)$$

where γ is the frictional coefficient for the effective energy landscape characterised by the reaction coordinate N_c; as usual $\beta = 1/k_B T$. We use the same approach here to calculate the rate constants for the three separate cases identified above.

A. Jump in the unstable region

We start with this region ($N_w > N_s$), in spite of it appearing the last on the list and in Fig. 2, because the transition rate obtained in this regime is unaffected by the 'lock' and forms the reference for all other cases. When the jump in the free energy happens in this unstable region, we are reduced to the homopolymer problem[11], since the extraction of the w-monomer has no effect on the process (the remaining globule loses its stability before the 'lock' is forced out). The forward rate constant is given by

$$k_0 = \frac{\alpha}{\gamma} \sqrt{\frac{\omega}{2\pi k_B T}} \kappa_\alpha e^{-\beta \Delta G_G}, \quad (16)$$

where ΔG_G is the energy barrier at N_c^*, see Fig. 2(c) for illustration. ω is the curvature of the barrier, and $\alpha = (\partial G/\partial N_c)|_{N_c=0}$ is the slope of the native potential well (which is treated as approximately triangular). The pre-factor κ_α is determined by the geometry of the well, and we define it for the convenience of later expressions:

$$\kappa_\alpha(N_c^*) = \left[1 - \exp\left(-\frac{\alpha N_c^*}{2 k_B T} \right) \right]^{-1}. \quad (17)$$

Note that the rate k_0 is not a function of our random variable N_w.

B. Jump before the barrier

In most cases, when $N_w < N_c^*$, the fixed jump in the free energy has only a minor effect of slightly distorting the pre-exponential factor in the basic homopolymer expression (16) due to the distortion of the native well, see Fig. 2(a). However, the effective height of the barrier is increased by the magnitude of this jump. To estimate the resulting rate constant k_1 we can split the integral over the barrier (having approximated the potential as harmonic) into two pieces,

$$\int_{-\infty}^{\infty} e^{\beta G} dN_c = \int_{-\infty}^{N_w} e^{\beta(G+w)} dN_c + \int_{N_w}^{\infty} e^{\beta G} dN_c.$$

This leads to a modified form of (16):

$$k_1(N_w) = f_w(N_w) \cdot k_0 e^{-\beta w}, \quad \text{with} \quad (18)$$

$$f_w = \frac{2}{(1 - e^{-\beta w}) \operatorname{erf}\left(\sqrt{\frac{\omega}{2 k_B T}} (N_c^* - N_w) \right) + (1 + e^{-\beta w})}.$$

Naturally, the rate constant k_1 reduces to k_0 when $w = 0$, that is, when there is no 'locked' w-monomer present in the polymer globule. In the special case when $N_w = N_c^*$, the exposure happens at the top of the barrier and the rate constant simplifies to:

$$k_1(N_c^*) = \frac{2 k_0}{1 + e^{\beta w}}. \quad (19)$$

C. Intermediate state kinetics

When the jump occurs in the region $N_c^* < N_w < N_s$, we have three (meta-)stable states and two free energy barriers between them, see Fig. 2(b). With three-state

kinetics, one has to make approximations to find an analytic expression for the rate constant. We choose to follow the steady-state approximation[29], where the intermediate state is assumed to be in equilibrium with the native state, and transitions to the extended state are assumed to be permanent (no refolding). In this approximation, the rate constant is expressed as

$$k_2 = \frac{k_{G\to I}k_{I\to E}}{k_{I\to G}+k_{I\to E}} \tag{20}$$

where G, I, and E refer to the globular (native), the intermediate, and the extended states, respectively.

$$k_{G\to I} = k_0 \ , \tag{21}$$

$$k_{I\to G} = \frac{\alpha'}{\gamma}\sqrt{\frac{\omega}{2\pi k_B T}}\kappa_{\alpha'}e^{-\beta\Delta G_{I\to G}} \tag{22}$$

where α' is the reverse gradient of the intermediate well, and $\Delta G_{I\to G} = G(f,N_c^*) - G(f,N_w)$ is free energy barrier for the transition from the intermediate state back to the native globular state, see Fig. 2(b) for illustration. The pre-factor $\kappa_{\alpha'}$ is given by the same form as Eq. (17), but with the arguments α' and $(N_w - N_c^*)$. Finally, the rate constant of escape into the fully extended state is:

$$k_{I\to E} = \frac{\alpha'^2}{k_B T\gamma}\kappa_{\alpha'}^{(1)}\kappa_{\alpha'}^{(2)}e^{-\beta w} \ . \tag{23}$$

α' appears twice in the expression for $k_{I\to E}$, and so is the pre-factor $\kappa_{\alpha'}$: once with the argument $(N_w - N_c^*)$ from the Kramers integral over the triangular well of the intermediate state, and again with the argument $(N_s - N_w)$ from the integral over the $I \to E$ barrier, which is also triangular with the same slope α'.

The full expression for the rate k_2 in the three-state regime can be simplified into a form:

$$k_2(N_w) = \frac{k_0}{\sqrt{\frac{\omega}{8\pi k_B T}}(N_s - N_w)e^{\beta w}e^{-\frac{1}{2}\beta\omega(N_w - N_c^*)^2} + 1} \tag{24}$$

We can plot the rate constants derived above for a range of w and f, as illustrated in Fig. 4. For values of N_w close to N_c^*, the expression for k_2 deviates from its expected value where it needs to join the end-point of k_1: here the approximations made by Brinkman (that of a high barrier) break down. However, we expect that the value of k_1 itself is actually valid at any value of N_w within the region $0 \le N_w \le N_c^*$. Therefore, we get a good idea of the true profile of $k(N_w)$ in spite of approximations.

For small $N_w \ll N_c^*$, the rate constant k_1 differs from the homopolymer rate of escape k_0 only by the Arrhenius factor for the jump, $\exp(-\beta w)$. For $N_w > N_s$, the rate is just a constant k_0. Because of these constant values of rates, the average $\langle n(t,M)\rangle$, given by Eq. (12), will retain a simple exponential time dependence in certain ranges of 'lock' monomer positions M.

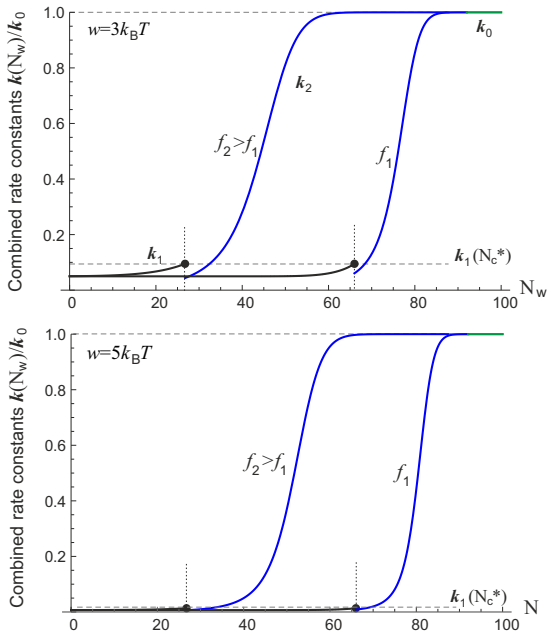

FIG. 4. Plots of the rate constants $k(N_w)/k_0$ for $N = 100$, $u = 5k_BT$ for two values of pulling force: $f_1 = 4.8k_BT/b$ (only slightly above the critical force f^*), and $f_2 = 5.3k_BT/b$. Two values of the 'lock' energy w are shown in the plots. The segments k_1 (at low N_w) terminate at the points when $N_w = N_c^*(f)$, Eq. (19), labelled by •. The segments in the three-state regime, $k_2(N_w)$, reach the value k_0 (1 in the scaled plots) when $N_w = N_s(f)$.

V. ENSEMBLE AVERAGE

We are now in a position to evaluate the integral in Eq. (14), and plot the resulting average population dynamics against time. In a typical AFM experiment[8,14,15] a constant force is applied to a folded protein, and a time of a sharp unfolding transition is recorded. After many repeats, a distribution of rupture times is obtained with great accuracy. The cumulant of this distribution represents the relative population of chains unfolded up to time t. With a single rate of unfolding, one expects this population to grow as $(1 - e^{-k_0 t})$. To represent the same population in our analysis, we plot $1 - \langle n(t)\rangle$ in Fig. 5. In the plot we use constant force slightly above the critical value f^*, and the fixed added strength of the 'lock' that is equal to the hydrophobic strength of the main monomers. In this representation it is difficult to distinguish exponential from non-exponential kinetics: all curves show qualitatively the same cumulative effect of an increasing fraction of unfolded chains as time passes under a constant force. It is, however, remarkable that for all the same parameters of the chain, the mere position of the 'locked' monomer along the sequence has such a strong effect on the apparent rate of unfolding.

Note that the time axis in Fig. 5 and subsequent plots is scaled by a dimensional constant $\tau = \gamma/k_BT$, evident in the original definition of all rates, Eq. (15). This time

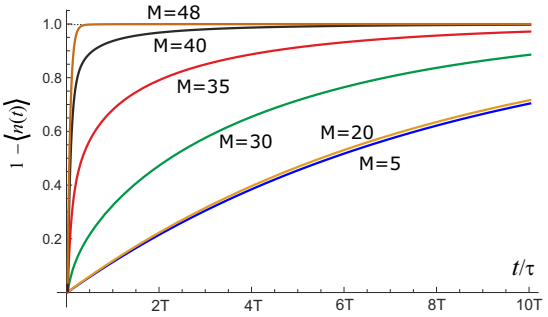

FIG. 5. We plot the cumulant 1-$\langle n(t)\rangle$ as an illustration of a typical experimental trace. We have the same force for each of the curves, and strength of w-monomer ($fb = 4.8k_BT$, $u = 5k_BT$, $w = 5k_BT$ for $N = 100$), but vary the position of the w-monomer, M. The linear non-dimensional time axis is measured in units of 10^{12}, which we shorthand as $T =$'tera'.

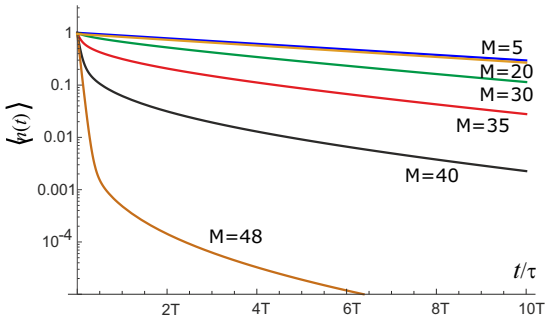

FIG. 6. The plot $\langle n(t)\rangle$ on a logarithmic scale, where $\langle n(t)\rangle$ is the same as in Fig 5, for $fb = 4.8k_BT$, $u = 5k_BT$, $w = 5k_BT$ and $N = 100$. For the core located at $M \approx 20$ or less, we have an approximately single exponential decay. For the intermediate core positions, there is a transition from faster decay to the single exponential at longer times (leading to a long tail). Note that as the core gets close to the middle of the chain the slower (sub-exponential) relaxation becomes increasingly apparent.

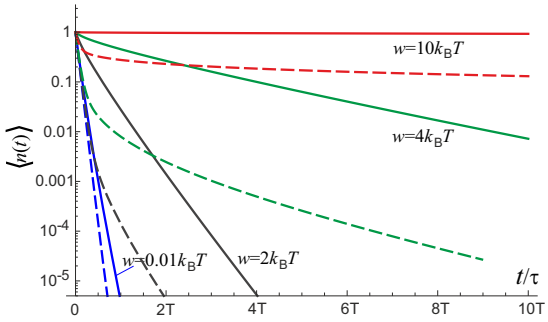

FIG. 7. The plot $\langle n(t)\rangle$ on a logarithmic scale, for $N = 100$, $fb = 4.8k_BT, u = 5k_BT$ and varying core strengths, w, listed in the plot. Solid lines are for a mid-range $M = 30$, the matching dashed lines are for $M = 42$, close to the middle of the chain. A stronger core (larger w) leads to a very long tail in the distribution, even for the same position M.

scale is an inverse of a diffusion constant of fluctuations in the reaction coordinate N_c. To relate it to a more familiar diffusion constant D of a single residue (monomer), we have to use the length scale of our monomer: $\tau = b^2/D$. Taking $D \sim 6 \cdot 10^{-10}\mathrm{m}^2/\mathrm{s}$ for an average amino acid in water (which is almost certainly an overestimate in this case), we obtain $\tau \sim 1.5 \cdot 10^{-10}$s. This means that the real time scale in Fig. 5 and subsequent plots is measured in minutes. It is a bit longer than in experiments we quoted[8,14,15], but our aim was not to reproduce the experimental results quantitatively: we built a minimal model with a single 'lock' to illustrate the point, with both u and w magnitudes chosen somewhat arbitrarily. It is very easy to change these parameters slightly and achieve a much better agreement with measured time scales, but we believe this is not necessary or particularly beneficial.

To what extent does the 'lock' affect the kinetics? To distinguish the exponential relaxation, we first plot $\langle n(t)\rangle$ for different lock positions, M, on a logarithmic scale in Fig. 6. As in Fig. 5, for a range of w-monomer positions near the end of the chain, there is a very little change in the slow single-exponential relaxation. The deviation from single exponential kinetics (showing as a straight line in this logarithmic plot) is increasingly evident for values of M closer to the middle of the chain. Initially, the decay is fast, and follows a simple exponential law, as those globules that transition with the fast rate k_0 unfold first. It is difficult to discern this regime in the log-linear plot in Fig. 6, because we concentrate on the long-time effects. At much longer times, there is a gradual crossover to a different simple-exponential decay with the rate constant $k_0 e^{-\beta w}$, much smaller for significant w. As M gets larger, the time this crossover occurs increases – whereas for small M we mostly see the slow exponential rate.

For small M, the deviation from single exponential kinetics will not be observed in experiment. This is easy to understand: the probability of exposure, $P(N_w)$, is relatively sharply peaked, so for small M, almost all the probability mass will lie in the two opposite regions where the rate constant is independent of N_w. Thus, the ensemble average $\langle n(t)\rangle$ remains a single exponential, with only a small extra contribution from other rate constants. Non-exponential effects are most prominent when M lies close to the middle of the chain where the rate constant undergoes the rapid change, cf. Fig. 4.

We examine the effect of varying the 'lock' strength w in Fig. 7, comparing a mid-range position $M = 30$ and near the middle position $M = 42$. As the 'locked' monomer becomes increasingly hydrophobic, the rate of slow decay $k_0 e^{-\beta w}$ will obviously decrease, and so the folded molecule is more stable for larger w. It is apparent that, at short times of folded population decay, all curves follow the fast simple-exponential relaxation k_0 that is independent of the 'lock' strength w. The position of the crossover to the slow simple-exponential relaxation with the rate constant $k_0 e^{-\beta w}$ occurs at different times depending on the 'lock' position M.

VI. DISCUSSION

In reality, biomolecules have a whole range of varied local and non-local interactions that help stabilise their specific structure. Our model does not have the scope to capture these particular effects, most familiar in protein folding: it would be irresponsible to suggest any quantitative agreement with an experiment. However, if the interactions of the core residues are much stronger than all other interactions in the molecule, we feel justified in taking these other interactions as approximately equal in strength, as in our model.

Not all cores will influence the unfolding kinetics, and we have demonstrated that much depends on the position of such a 'lock' along the chain. When the core residues are close to the terminus of a protein, that protein will unfold on average after a much longer time than those where the core residues are in the middle of the protein's sequence. However, the globules with core residue close to the middle of their sequence exhibit a pronounced non-exponential dynamics (on average) upon the application of constant force. Unsurprisingly, stronger locks stabilise the molecule more than weak locks.

What if there are multiple regions in a protein that are strongly bound together? The probability distribution now has multiple degrees of freedom: the exposures of the different cores. The probability will be sharply peaked around the point $\mathbf{M}$, which is now a vector containing the sequence positions of the different cores. There may be small breakage events into smaller globules (but each larger than the critical size), each protecting a strongly bound region, before the chain fully extends. These might manifest as short-lived intermediate plateaus on the experimental traces.

Experimental work that shows distinctly non-exponential kinetics of unfolding has been mostly done on ubiquitin protein[14,15,19]. The structure of ubiquitin has been resolved in 1987[30], where the authors clearly identify the residues that form the hydrophobic core, which is created by bonding the α-helical segment (residues 23-34) with the β-sheet segment. By examining the sequnce and the folded structure, Fig. 8, we conclude that the main hydrophobic bond is formed between Ala28 and Leu43, out of the total sequence of 76 residues. Since force is applied at the N- and C-termini of the protein, we say that the intermediate chain segment (28-43) between these two residues is only exposed to the force upon the breaking of this bond. Therefore, we introduce an effective chain length of $N = 61$, with a lock at $M = 28$ and $N - M = 33$. This is close to the middle of the effective chain, which agrees with the non-exponential kinetics prediction.

Finally, what about other non-local interactions? We had nothing to say about these in our analysis, but offer some thoughts here. One could consider a collection of springs between two plates, with a constant pulling force applied to the plates. These springs could have different breaking energies, and different spring constants. Force applied will now be spread across these springs, which will break, according to the same principles as discussed here. As each springs break, the force will re-distribute across the remaining springs, and each will have a force increased. This linkage of the bonds will lead to non-exponential kinetics as well, as many other effects such a random variation of sequence.

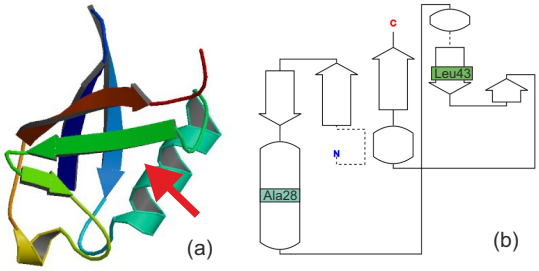

FIG. 8. The structure of ubiquitin: (a) the PDB 1ubq rendering, the arrow pointing at the strongest hydrophobic bond between Ala28 and Leu43; (b) the corresponding sequence annotation with the labeled hydrophobic residues.

In this paper we looked at the simplest possible scenario, and found that the introduction of even a single specifically-placed inhomogeneity to an otherwise homopolymer chain already produces non-exponential kinetics when chains are unfolded under constant force. The inhomogeneity leads to a statistical randomness in the way the chain unfolds and results in stretching of the decay curve, with a long tail decay after an initial fast exponential decay (corresponding to the situations when a significant portion of the chain can be pulled out of the globule before the 'lock' is affected). While this is very much a toy model, such behaviour mimics quite well the behaviour of biological macromolecules in AFM pulling experiments, and we hope this paper offers some insight into the mechanisms for non-exponential kinetics in these experiments.

Acknowledgments

This work is supported by the EPSRC through a studentship award and the Critical Mass Grant for Cambridge Theoretical Condensed Matter EP/J017639.

[1] S. Kumar and M. S. Li, Phys. Rep. **486**, 1 (2010).
[2] A. Engel, H. Gaub, and D. Mller, Curr. Biol. **9**, R133 (1999).
[3] T. E. Fisher, A. F. Oberhauser, M. Carrion-Vazquez, P. E. Marszalek, and J. M. Fernandez, Trends Biochem. Sci. **24**, 379 (1999).
[4] C. Bustamante, Z. Bryant, and S. B. Smith, Nature **421**, 423 (2003).
[5] M. Rief, Science **276**, 1109 (1997).
[6] S. B. Fowler, R. B. Best, J. L. Toca Herrera, T. J. Rutherford, A. Steward, E. Paci, M. Karplus, and J. Clarke, J. Mol. Biol. **322**, 841 (2002).
[7] S. Garcia-Manyes, J. Brujić, C. L. Badilla, and J. M. Fernández, Biophys. J. **93**, 2436 (2007).
[8] A. F. Oberhauser, P. K. Hansma, M. Carrion-Vazquez, and J. M. Fernandez, Proc. Natl. Acad. Sci. USA **98**, 468 (2001).

[9]A. A. Polotsky, E. E. Smolyakova, and T. M. Birshtein, Macromolecules **44**, 8270 (2011).
[10]A. A. Polotsky, E. E. Smolyakova, O. V. Borisov, and T. M. Birshtein, Macromol. Symp. **296**, 639 (2010).
[11]S. Bell and E. M. Terentjev, J. Chem. Phys. **143**, 184902 (2015).
[12]G. Bell, Science **200**, 618 (1978).
[13]H. A. Kramers, Physica **7**, 284 (1940).
[14]J. Brujic, Z. R. I. Hermans, K. A. Walther, and J. M. Fernandez, Nat. Phys. **2**, 282 (2006).
[15]H. Lannon, E. Vanden-Eijnden, and J. Brujic, Biophys. J. **103**, 2215 (2012).
[16]J. G. Saven, J. Wang, and P. G. Wolynes, J. Chem. Phys. **101**, 11037 (1994).
[17]S. J. Hagen and W. A. Eaton, J. Chem. Phys. **104**, 3395 (1996).
[18]R. Zwanzig, Acc. Chem. Res. **23**, 148 (1990).
[19]T.-L. Kuo, S. Garcia-Manyes, J. Li, I. Barel, H. Lu, B. J. Berne, M. Urbakh, J. Klafter, and J. M. Fernández, Proc. Natl. Acad. Sci. USA **107**, 11336 (2010).
[20]C. Hyeon, M. Hinczewski, and D. Thirumalai, Phys. Rev. Lett. **112**, 138101 (2014).
[21]Y. Zheng, Y. Bian, N. Zhao, and Z. Hou, J. Chem. Phys. **140**, 125102 (2014).
[22]P. L. Geissler and E. I. Shakhnovich, Macromolecules **35**, 4429 (2002).
[23]J. R. Blundell and E. M. Terentjev, Macromolecules **42**, 5388 (2009).
[24]M. Fixman and J. Kovac, J. Chem. Phys. **58**, 1564 (1973).
[25]J. Marko and E. Siggia, Macromolecules **28**, 8759 (1995).
[26]G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers*, 2nd ed. (Dover Publications, Mineola, NY, 2013).
[27]H. C. Brinkman, Physica **XXII**, 149 (1956).
[28]C. W. Gardiner, *Handbook of Stochastic Methods*, 2nd ed. (Springer Verlag, Berlin, 1985).
[29]R. T. Raines and D. E. Hansen, J. Chem. Educ. **65**, 757 (1988).
[30]S. Vijay-Kumar, C. E. Bugg, and W. J. Cook, J. Mol. Biol. **194**, 531 (1987).