Localization of denaturation bubbles in random DNA sequences

Terence Hwa[†‡], Enzo Marinari[§], Kim Sneppen[¶], and Lei-han Tang[∥]

[†]Department of Physics and Center for Theoretical Biological Physics, University of California at San Diego, La Jolla, CA 92093-0319; [‡]Statistical Mechanics and Complexity–Istituto Nazionale per la Fisica della Materia, and Dipartimento di Fisica, Università di Roma "La Sapienza", P.A. Moro 2, 00185 Rome, Italy; [¶]Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen, Denmark; and [∥]Department of Physics, Hong Kong Baptist University, Kowloon Tong, Hong Kong

We study the thermodynamic and dynamic behaviors of twist-induced denaturation bubbles in a long, stretched random sequence of DNA. The small bubbles associated with weak twist are delocalized. Above a threshold torque, the bubbles of several tens of bases or larger become preferentially localized to AT-rich segments. In the localized regime, the bubbles exhibit "aging" and move around subdiffusively with continuously varying dynamic exponents. These properties are derived by using results of large-deviation theory together with scaling arguments and are verified by Monte Carlo simulations.

Localized opening of double-stranded DNA is essential in a number of cellular processes such as the initiations of gene transcription and DNA replication (1). Although thermal denaturation is highly unlikely under physiological conditions, *in vitro* experiments show that local denaturation can be readily induced by underwinding the DNA double-helix by an amount that is physiologically reasonable (2–4). The basic physical effect is simple: An underwound double-helix suffers a reduction in binding free energy (5–7). Local openings of the double-helix (referred to as "denaturation bubbles") relieve the twist experienced by the remainder of the double-helix and are thus energetically favorable. The denaturation bubbles may be recruited to a specific location of the genome by a designed (e.g., AT-rich) sequence, since AT pairs bind more weakly than GC pairs (8). On the other hand, entropic effect that favors bubble delocalization is non-negligible for long sequences. Also significant is the kinetic trapping of the bubbles due to statistical agglomeration of AT-rich segments in long heterogenous sequences.

To gain some quantitative understanding on the competing effects of entropy and sequence heterogeneity, we characterize in this study the thermodynamic and dynamic properties of denaturation bubbles in a long, stretched random DNA sequence with no special sequence design. Previously, there have been a number of experimental and theoretical studies (9–12) on the effect of sequence heterogeneity on DNA melting and unzipping transitions. Our study is along this general direction. The specific behaviors exhibited by the denaturation bubbles are rather complex and are typical of those observed in systems dominated by quenched disorders (13): The bubbles are localized upon increase of the applied torque beyond a certain threshold. In the localized regime, their dynamics exhibits "aging" (14, 15) and is subdiffusive with continuously varying exponents.

Interestingly, twist-induced denaturation presents a rare physical example of the celebrated random-energy model (REM) of a disordered system (16). Consequently, detailed analysis of both the thermodynamics and dynamic properties can be made by applying the well-developed theory of disordered systems (13), together with exact results from large-deviation theory familiar in the related sequence alignment problem (17, 18). We will draw on detailed experimental knowledge of thermal denaturation (19–21) throughout the analysis and make our results quantitative whenever possible.

Thermodynamics

Let us consider the application of a torque that underwinds a long, stretched** piece of double-stranded DNA. We are interested in the regime where the applied torque T is below the threshold T_d for bulk denaturation, but sufficiently strong so that denaturation bubbles appear in the system. Due to the highly cooperative nature of the denaturation process, the typical distance $N_\times$ between the large bubbles is large, in which case treating the bubbles as a dilute gas of particles is appropriate. Our strategy will be to first characterize analytically the thermodynamic behavior of a single bubble, and then use this knowledge to determine the length scale $N_\times$ and the many-bubble states for $N \gg N_\times$. We will find that $N_\times > O(10^3)$ bp as long as we are not very close to the threshold T_d, so that the dilute gas approximation is reasonable for a large range of parameters.

The Single-Bubble Model. Consider a denaturation bubble confined in a DNA double-helix between two complementary DNA strands of N bases each. The double-strand is denoted by the base sequence $b_1 b_2 \ldots b_N$ [with $b_k \in \{A, C, G, T\}$] of one of the strands, ordered from the 5′ to 3′ end.

To simplify the notation, we assume that the two ends of the helix are sealed, so that the bubble is always contained in the segment $b_1 \ldots b_N$. Let the index of the first and last open pairs of the open bubble be m and n, with $1 \leq m \leq n \leq N$. We denote the total free energy of the bubble (defined with respect to the helical state) by $\Delta G_L(m)$, where $L \equiv n - m + 1$ is the number of open bases and referred to as the bubble length. Then the partition function of the single-bubble system is given by

$$Z(N) = \sum_{L=1}^{N} \sum_{m=1}^{N-L+1} e^{-\beta \Delta G_L(m)}, \qquad [1]$$

where $\beta^{-1} \equiv k_B T \simeq 0.62$ kcal/mol at 37°C.

In the absence of the external torque, the bubble energy $\Delta G_L(m)$ has two components. First, there is the loss of stacking energy $\delta G_{b,b'}$ between two successive bases b and b'. These stacking energies are in the range of 0.5 to 2.5 $k_B T$s at 37°C, with the AT stacks weaker than the GC stacks. Their values have been measured carefully (19–21). Second, assuming that there is no secondary pairing between bases in the bubble so that the open configuration can be regarded as a polymer loop, there is a well-known polymeric loop entropy cost

$$\gamma_L = \gamma_1 + \alpha \cdot k_B T \ln L \qquad [2]$$

PHYSICS

BIOPHYSICS

for a bubble of length L, with $\alpha \approx 1.8$ (22) for a linearly extended[††] DNA chain. The bubble initiation cost γ_1 depends on the base composition at opening and closing ends, ionic strength, etc., and generally lies[‡‡] in the range of 3 to 5 k_BT. For relevant bubble sizes of few tens of bases in length (see below), the total entropic cost is $\gamma_L = 8 \sim 12\, k_BT$. This large cost justifies the single bubble approximation (at least to the length scale $N_\times \sim e^{\beta\gamma_L} \gtrsim 5 \times 10^3$ bp) and contributes significantly to the sharpness of the observed thermal denaturation transition (21).

An applied negative torque T reduces the thermodynamic stability of the helical state relative to the denatured one by an amount equal to the work done to unwind the helix. This effect is simply modeled here by a linear decrease in the stacking energy in the relevant parameter range (6), i.e., $\delta G_{b,b'} \rightarrow \delta G_{b,b'} - \theta_0 T$, where $\theta_0 = 2\pi/10.35$ is the twist angle per base of the double-helix. Putting the above together, we have

$$\Delta G_L(m) = \gamma_L + \sum_{k=m}^{m+L} \delta G_{b_{k-1}, b_k} - \theta_0 T \cdot L \qquad [3]$$

as the single-bubble energy, which can be computed once the DNA sequence $b_1 \ldots b_N$ is given. Note that although Eq. 3 is formulated specifically for twist-induced denaturation, the general form can be used to describe a number of destabilizing effects, e.g., due to changes in temperature, ionic concentration, etc.

Sequence Heterogeneity. As the torque T increases from zero toward the denaturation point T_d, denaturation bubbles appear in the double-strand and grow in size. We want to know whether the bubbles are free to diffuse along the double-strand, or are they localized in the high AT regions of the DNA where binding is the weakest. For simplicity, we will characterize the typical behavior of an ensemble of random (i.e., independent and identically distributed) sequences described by the single-nucleotide frequencies p_b, although our approach and qualitative findings are also applicable to sequences with short-range correlations.

For a given sequence of bases, the partition function Z can, of course, be efficiently evaluated numerically (including all the multiple-bubble states) by using available programs such as MELTSIM (21). All thermodynamic quantities can subsequently be evaluated from the free energy $F = -k_BT \ln Z$. To obtain the typical behavior of the ensemble, we ideally want to compute the ensemble average of the free energy, $\bar{F} \equiv -k_BT\, \overline{\ln Z}$. [We use the overline to denote average over the ensemble of random sequences, i.e., $\bar{X} \equiv \sum_{b_1,\ldots,b_N} X_{b_1,\ldots,b_N} \Pi_{k=1}^N p_{b_k}$; this is also known as the "disorder average."] Computing $\bar{F}$ numerically, however, will require explicit generation of a large number of random sequences and can be very time consuming for large Ns. Fortunately, we can apply a large body of knowledge accumulated from the statistical mechanics of random systems (13) and provide a detailed characterization of the typical behavior of our system without the need of exhaustive simulation. To introduce notation and concepts in this approach, we examine first the simplified problem of a single bubble with a fixed length.

Bubble with Fixed Length. Let us consider a bubble with a fixed length L (with $1 \ll L \ll N$) embedded in a long, random sequence $b_1 \ldots b_N$. The partition function reads

$$\mathcal{Z}_L = \sum_{m=1}^{N-L+1} \exp[-\beta \Delta G_L(m)], \qquad [4]$$

where the scripted variables refer to properties of the fixed-length bubble. For a random sequence, the energies of the different states labeled by m are uncorrelated with each other beyond the distance L. Such systems belong[§§] to the class of REM and was solved exactly in the 1980s by Derrida (16) for a Gaussian distribution of ΔGs. Discrete distribution of ΔGs was studied in the closely related system involving protein–DNA interaction (26). Below, we will briefly review the salient properties of REM by using the present example.

The REM has a "high-temperature" phase where many (of order N) bubble configurations contribute significantly to the partition sum, and a "low-temperature" phase dominated by only one or a few lowest energy states. It follows that, in the former case, the bubble is delocalized and can freely diffuse along the sequence, whereas in the latter case, the bubble is localized to the lowest energy position. Transition between the two phases is driven by competition between the energetic (variation in ΔG) and entropic ($\ln N$) effects. In the present problem, the magnitude of terms in the partition sum **4** can be tuned not only by varying the temperature, but also by varying the bubble size L. Hence, at a fixed β, whether a bubble is free or localized depends both on the bubble size L and the sequence length N.

An interesting property of the REM is that, in the high-temperature phase, $\mathcal{Z}_L/N$ tends to a finite limit given by the annealed average $\bar{\mathcal{Z}}_L/N$ as $N \rightarrow \infty$, independent of the particular realization of the random sequence. This allows us to replace the average free energy $\bar{F} \equiv -k_BT\, \overline{\ln \mathcal{Z}}$ by its annealed approximation $\tilde{F} \equiv -k_BT \ln \bar{\mathcal{Z}}$, which is much easier to calculate. [We will use the tilde to denote all quantities computed in the annealed approximation.] Introducing a 4×4 matrix $\mathbf{M}(\beta)$ with components $\mathbf{M}_{b,b'} = \sqrt{p_b p_{b'}} \exp[-\beta \delta G_{b,b'}]$, and let the largest eigenvalue of $\mathbf{M}(\beta)$ be $\Lambda(\beta)$, then the disorder average of terms in **4** can be written as

$$\overline{e^{-\beta \sum_{k=m}^{m+L-1} \delta G_{b_{k-1}, b_k}}} = \mathrm{Tr}\, \mathbf{M}^L(\beta) \overset{L \gg 1}{=} \Lambda^L(\beta). \qquad [5]$$

It is convenient to introduce the quantity

$$f(\beta) \equiv -\beta^{-1} \ln \Lambda(\beta), \qquad [6]$$

with which we have (for $N \gg L$)

$$\bar{\mathcal{Z}}_L = \frac{N e^{-\beta \gamma_1}}{L^\alpha} [e^{-\beta[f(\beta) - \theta_0 T]}]^L. \qquad [7]$$

Hence, in the delocalized phase,

$$\bar{F} \simeq \tilde{F} = -k_BT \ln N + L[f(\beta) - \theta_0 T] + \gamma_L.$$

The annealed entropy can be calculated from $\tilde{F}$, with[¶¶]

$$\tilde{S} = -k_B^{-1} \frac{\partial \tilde{F}}{\partial T} = \ln N - \beta[f(\beta) - \varepsilon(\beta)]L, \qquad [8]$$

where $\varepsilon(\beta) \equiv -(\partial/\partial\beta) \ln \Lambda$. It will also be useful to introduce the relative entropy per base for the fixed-length bubble,

$$\mathcal{H}(\beta) \equiv [\ln N - \tilde{S}]/L = \beta[f(\beta) - \varepsilon(\beta)]. \qquad [9]$$

Note that being the difference between f and ε, the quantity $\mathcal{H}$ is a measure of the intrinsic variation in the binding energies δGs for a random sequence with nucleotide frequency p_b and is independent of the average binding energy $\overline{\delta G}$, which external environments such as the temperature or solvent conditions most directly affect.

Derrida's solution of REM (16) shows that the annealed entropy $\tilde{S}$ vanishes at the transition to the low-temperature phase, beyond which the annealed approximation is no longer applicable. Using Eqs. **8** and **9**, we can write the condition for phase transition as

$$L_{\text{loc}} = \ln N/\mathcal{H}(\beta), \qquad [10]$$

which gives the minimal bubble size for localization at a given N. With the values of δGs obtained from ref. 19 at $[\text{Na}^+] = 1$ M and 37°C, and assuming an equal nucleotide distribution (i.e., $p_b = 1/4$ for all bases), we find $f \approx 1.83 \, k_BT$, $\varepsilon \approx 1.50 \, k_BT$, so that $\mathcal{H} \approx 0.33$ and $L_{\text{loc}} \approx 20$ bp for $N \sim 10^3$ bp. From Eq. **10**, it is clear that as $N \to \infty$ any fixed-length bubble remains delocalized.

Bubble Without Length Constraint. The full partition function Z is obtained simply by summing Z_L for different Ls. We will again approach the problem by first applying the annealed approximation and then determining where it breaks down.

Annealed approximation. The annealed partition function $\bar{Z}(N) \equiv \sum_{L=1}^{N} \bar{Z}_L$ has a transition at $T_a = f(\beta)/\theta_0$, where the exponential factor in **7** reaches one: The sum over L is finite and $\bar{Z} \propto N$ only if $T < T_a$. In this regime, the annealed free energy is simply $\bar{F} \approx -k_BT \ln N + \gamma_1$. The annealed energy $\tilde{E} \equiv -(\partial/\partial\beta) \ln \bar{Z}(\beta)$ is also readily computed; it can be expressed as $\tilde{E} = \gamma_1 + [\varepsilon(\beta) - \theta_0 T] \cdot \tilde{L}$ where $\tilde{L}(T) \equiv \sum_{L=1}^{N\to\infty} L \bar{Z}_L/\bar{Z}$ is the average bubble length in the annealed approximation. As T approaches T_a, $\tilde{L}(T)$ diverges, and the annealed entropy

$$\tilde{S} = \ln N - \beta[\theta_0 T - \varepsilon(\beta)] \cdot \tilde{L} \qquad [11]$$

becomes negative.

In the limit $N \to \infty$, the annealed free energy $\bar{F}$ is actually identical to $\bar{F}$ for all $T \leq T_a$. This can be seen from the inequalities $\overline{\ln Z_{L=1}} \leq \overline{\ln Z} \leq \ln \bar{Z}$, and $Z_{L=1} > N \min\{\exp[-\beta\Delta G_1], 1\}$. Since both the lower and upper bounds grow as $\ln N$,

$$\bar{F} \equiv -k_BT \overline{\ln Z} \simeq -k_BT \ln N \qquad [12]$$

for all $T \leq T_a$.

Ground-state properties. To find the ground state of the unconstrained bubble in a long random sequence, we need to study the statistics of stretches of exceptionally high AT content. If we neglect the polymeric contribution γ_L to the bubble energy (to be justified shortly), then the ground-state energy E^* expected in a sequence of length N can be computed exactly from large-deviation theory (17, 27), with

$$\overline{E^*}(N) \approx -\lambda^{-1} \ln N. \qquad [13]$$

The constant λ in Eq. **13** can be expressed as the unique positive root of the equation

$$f(\lambda) = \theta_0 T, \qquad [14]$$

where f is defined by the δGs through Eqs. **5** and **6**. Note that, at $T = T_a$, Eq. **14** is satisfied with $\lambda = \beta$. In this case, **13** coincides with **12**.

The length of the minimal energy bubble is also known from large-deviation theory (18, 27), with

$$L^*(T) = \ln N/H^*(T), \qquad [15]$$

where the relative entropy H^* is given exactly by

$$H^*(T) = \lambda(T) \cdot [\theta_0 T - \varepsilon(\lambda)]. \qquad [16]$$

From the logarithmic dependence of the bubble length L^* on N, it is clear that the corresponding polymeric contribution $\gamma_{L^*} \sim \ln(\ln N)$ can indeed be treated as a constant shift of bubble energy.

Phase transitions. Based on the above discussion, a phase transition can be formally established in the limit $N \to \infty$. This is seen by comparing the expressions **12** and **13**. For $T > T_a$, solution to **14** satisfies $\lambda < \beta$. Consequently, **12** must break down there, yielding a phase transition at $T = T_a$. Since $\bar{F} \leq \overline{E^*}$ in general (e.g., for all $T > T_a$), and at the phase transition point $T = T_a$ the equality $\bar{F} = \overline{E^*}$ already holds, i.e., the ground state already dominates, then we must have the ground state dominating throughout the localized phase. This is exactly the behavior of the REM (16).

A physical understanding of the transition can be obtained by examining the importance of the ground-state contribution $\exp(-\beta E^*) \sim N^{\beta/\lambda}$ to the partition sum Z as the applied twist T is varied. For $T < T_a$, the ratio $\beta/\lambda(T)$ is <1. In this case, the energy gain $\overline{E^*}(N)$ of placing the bubble at the site of the lowest energy is insufficient to overcome the entropy $\ln N$ of placing the bubble in different positions, hence the bubble is typically small and delocalized. When T exceeds T_a, the ground-state contribution grows faster than N, signaling dominance of one or a few low-energy states where the bubble typically resides. The transition is thus identified as the localization transition of the bubble at $T_{\text{loc}} = T_a$.

The onset of the zero entropy point can be obtained from Eq. **11** and written as

$$\ln N = H(\beta, T) \cdot \tilde{L}(\beta, T), \qquad [17]$$

where

$$H(\beta, T) = \beta \cdot [\theta_0 T - \varepsilon(\beta)] \qquad [18]$$

is the relative entropy of the unconstrained bubble. These equations are analogous to the expressions **15** and **16** for the ground-state bubble. In fact, both $L^*(T)$ and $H^*(T)$ are reproduced through the substitution $\beta \to \lambda(T)$, e.g., $L^*(T) = \tilde{L}(\lambda(T), T)$. This turns out to be true also for other thermodynamic variables. Thus the localized phase at different Ts can be viewed as the phase-transition points of systems with different effective temperatures $\lambda^{-1}(T)$; this will be clearly manifested in the bubble dynamics discussed below.

Next, we observe that since $H^* \propto \lambda$ (see Eq. **16**), the bubble length diverges (or approaches N) as $\lambda \to 0$. This defines the point of bulk denaturation[‖] T_d, i.e.,

$$\theta_0 T_d \equiv \lim_{\lambda \to 0} f(\lambda) = \overline{\delta G}, \qquad [19]$$

where the second equality is obtained from manipulating Eqs. **5** and **6**. Using $\overline{\delta G} \approx 1.40 \, k_BT$ (derived from the δGs in ref. 19), we find $T_d \approx 10$ pN·nm. The dependence of λ on T close to T_d can be obtained from the expansion

$$f(\lambda) = \overline{\delta G} - \frac{\lambda}{2} \text{var}(\delta G) + O(\lambda^2). \qquad [20]$$

[‖]Note, however, that the helical segments separating adjacent bubbles can be stable even beyond T_d, so that complete separation of the two strands takes place at $T > T_d$.

Inverting the above for λ by using **14** and **19**, we find

$$\lambda(T) = \frac{2\theta_0}{\beta \text{var}(\delta G)}(T_d - T) + O(T_d - T)^2. \qquad [21]$$

It turns out that the term linear in $T_d - T$ in **21** already gives a very good approximation (to within 1%) of λ throughout the localized phase where $\lambda/\beta < 1$. The localization transition point T_{loc} can be thus obtained by solving Eq. **21** with $\lambda(T_{\text{loc}}) = \beta$. Using $\beta^2 \text{var}(\delta G) \approx 0.565$ (derived from ref. 19), we find $T_d - T_{\text{loc}} \approx 2$ pN·nm. Unlike the value of T_d that is derived from the average stacking energy $\overline{\delta G}$ and hence is sensitive to temperature, ionic strength, etc., the difference $T_d - T_{\text{loc}}$ is set by the variance of $\delta G_{b,b'}$ and should be much less sensitive to experimental conditions. The same is expected for the relative entropy, which has the form

$$H^*(T) \approx 2\theta_0^2 (T_d - T)^2 / \text{var}(\delta G) \qquad [22]$$

throughout the localized phase.

Multiple Bubbles. The localization transition discussed above occurs only as $N \to \infty$. However, for large N, the single-bubble approximation will break down regardless of the large (but finite) bubble cost γ_L. When multiple bubbles are localized, each bubble is effectively in a finite-length system, thereby blurring the localization transition.

We first analyze the delocalized phase for which the annealed approximation is valid. Once multiple bubbles are allowed in the system, we expect a broad range of bubble lengths, as described by the distribution **7**. Qualitatively, we expect only the largest bubbles, of size $\bar{L}(T)$ to be localized as $T \to T_{\text{loc}}$, while the smaller ones remain delocalized. We shall thus focus on these large bubbles. It is the average separation distance $N_\times$ between these large bubbles that sets the effective system size of the single-bubble localization problem.

The Boltzmann weight of one such large bubble in a sequence of length $N \gg \bar{L}$ is $\bar{W}(N) = e^{-\beta\gamma_1} N / \bar{L}^\alpha$ in the vicinity of the localization transition. Setting $\bar{W}(N) = 1$ yields the typical spacing between the large bubbles on the delocalized side,

$$\bar{N}_\times \approx e^{\beta\gamma_1} \bar{L}^\alpha(T). \qquad [23]$$

Note that for bubbles of size 10 bp, the crossover length is already of the order of 10^3 bp. A similar estimate can be made on the localized phase by using the exact expression (28) for the lowest energy for multiple bubbles. We find

$$N_\times^* \approx [e^{\beta\gamma_1}(L^*)^\alpha]^{\lambda(T)/\beta}. \qquad [24]$$

as the average distance between large bubbles of size L^*.

For $N \gg N_\times$, the system consists of $N/N_\times$ effective number of single-bubble subsystems, each of length $N_\times$. At the localization "transition" of an infinite system then we have $\ln \bar{N}_\times = H(\beta, T_{\text{loc}})\bar{L}(T_{\text{loc}})$ (see Eq. **17**). Together with Eq. **23** (or **24** with $\lambda = \beta$), we find $\bar{L}(T_{\text{loc}}) \approx 25$ bp at the onset of localization (using $\gamma_1 \approx 3 k_B T$ and $H \approx 0.33$), with the crossover length $\bar{N}_\times \approx 6,500$ bp. Thus we expect there to be typically one bubble of ~ 25 bp in a random DNA double-strand of length $\sim 6,500$ bp at the localization transition.

Bubble Dynamics

The localization of bubbles is reflected ultimately in their slow dynamics. We expect bubbles to diffuse freely along the DNA double-helix in the delocalized phase, but become trapped in low-energy positions in the localized phase. Details of the bubble movement in the latter case, however, can be rather complicated with nontrivial memory (or aging) effects typical of glassy states (14, 15) as will be described below.

Model. For simplicity, we will restrict ourselves to the description of the movement of a single bubble over its lifetime, which can be rather long in the localized phase. For reasons discussed above, interaction with other bubbles can be neglected when the bubble displacement is within a distance of order $N_\times \sim 10^3$ bp. We will also neglect the polymeric loop entropy γ_L, which provides essentially a constant shift to the bubble energy as shown in the single-bubble section.

In addition to the drift and breathing motion, a bubble may shrink to zero size and disappear from the system. To our knowledge, the time scale involved for the spontaneous collapse of a bubble, particularly under an applied twist, has not been documented so far. Zipping the bubble requires not only pairing of the bases in the open segment, but also rewinding of the helix against the applied undertwist, both of which contribute to the energy barrier to the no-bubble state. This suggests a long lifetime for a bubble, which can be enforced by setting a lower bound (e.g., 10 bp) in the allowed bubble length. However, as we will see, the longtime behavior of bubble dynamics is determined crucially by the occurrence of the large bubble states, and insensitive to the value of the lower bound on L, as long as the $L = 0$ state is excluded. Once accurate estimates of bubble lifetime become available, one may supplement the discussion below with such a cutoff.

Scaling Theory. Eq. **13** gives the lowest energy of an unconstrained bubble in a sequence of length N, while a bubble with its position (but not size) fixed typically has an energy of the order λ^{-1}. For small λ, the energy variation $\Delta E(N) \simeq \lambda^{-1} \ln N$ is large, hence the bubble dynamics is dominated by the thermal escape from the deepest trap. The escape time is thus $t_e(N) \simeq e^{\beta\Delta E(N)} \sim N^{\beta/\lambda}$, i.e., the dynamics is subdiffusive deep in the localized phase (where $\beta \gg \lambda$).

To investigate the dynamical behavior in more detail, especially close to the localization transition where $\lambda \approx \beta$, we need to include also the random motion of the bubble along the double-strand. Toward this end, it is useful to describe the bubble dynamics as a single point moving in the 2D space spanned by the bubble's only two degrees of freedom, its instantaneous length L and the position of one of its ends, say m. The statistics of the 2D energy landscape $\Delta G_L(m)$ is well characterized by the large-deviation theory (18). It consists of a number of valleys, whose depths (denoted by $\Delta\hat{G}$s) are given by the Poisson distribution

$$\mathscr{D}(|\Delta\hat{G}|) \propto e^{-\lambda|\Delta\hat{G}|}, \qquad [25]$$

where λ is the constant defined through **14**. The typical valley length is $\hat{L} \sim 1/H^*$, where H^* is given by **16**. The valleys are spread out along the corridor at $L \lesssim \hat{L}$, separated by a typical distance M, which is also calculable from the large-deviation theory. For much larger Ls, the bubble energy becomes prohibitively high.

Clearly, the dynamics consists of two parts: At short times, it is dominated by the escape of the bubble out of an individual valley and is analogous to the (biased) Sinai problem (29). At longer time scales, the bubble "hops" from one valley to another along the corridor of valleys. This dynamics, which is essentially that of a particle traversing a series of exponentially distributed energy valleys (see Eq. **25**), has been extensively investigated previously in the context of the one-dimensional trap model (30, 31). Here we review some key results and refer the reader to ref. 32 for details.

The basic dynamic quantity is the time $\tau(\Delta\hat{G}) \propto e^{\beta|\Delta\hat{G}|}$ to escape each valley of depth $\Delta\hat{G}$. The average time to traverse K valleys over a length scale $N = K \cdot M$ by random walk is then given by
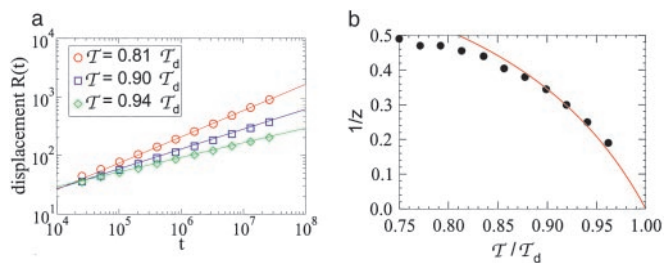
Fig. 1. (*a*) Average bubble position vs. time for various values of Ts in the glass phase: the solid lines are power law fits. (*b*) The extracted exponents vs. T: the solid line is the prediction of the scaling theory, Eq. **27**.

$$t_e = K^2 \cdot \langle \tau \rangle \sim N^2 \int_{\lambda^{-1}}^{|\overline{E^*}|} dx \; \tau(x) \mathscr{D}(x), \quad \quad [26]$$

where $\langle \tau \rangle$ is the average of the trap time $\tau(\Delta \hat{G})$, and the limits of integration in **26** are from the magnitude of the typical valley depth λ^{-1} to that of the deepest valley **13** expected for a segment of length N. The total time according to Eq. **26** can be written as $t_e(N) \propto N^z$, with the dynamic exponent z given by

$$z = \begin{cases} 2 & \text{for } \lambda > \beta \; (\text{or } T < T_{\text{loc}}) \\ 1 + \beta/\lambda & \text{for } \lambda < \beta \; (\text{or } T_{\text{loc}} < T < T_d). \end{cases} \quad [27]$$

The anomalous exponent $z > 2$ in the glass phase shows explicitly that the dynamics is slow, i.e., subdiffusive.

Glassy Dynamics. We next report the result of a Monte Carlo simulation of the bubble dynamics on predefined random nucleotide sequences. We impose local dynamics in which the bubble can only change its length L or shift its end position m by a single base, as long as $L \geq 1$. To remove edge effects and probe the asymptotic dynamics, we use a very large sequence length ($>10^4$ bp) so that the bubble never reaches the boundary of the sequence given the duration of our numerical study. All disorder-averaged quantities reported are performed over 10^4 random sequences.

Anomalous diffusion. To characterize the slow dynamics quantitatively, we show in Fig. 1*a* the time evolution of the average displacement $R(t) = |m(t) - m(0)|$ of the bubble position for a few selected values of Ts in the glass phase. Clearly, the displacement can be described by a power law of the form $R(t) \propto t^\nu$, where we expect $\nu = 1/z$. In Fig. 1*b*, we plot the extracted exponents for different values of Ts in the range $T_{\text{loc}} \lesssim T < T_d$. The expected values $1/z$ according to Eq. **27** (using the linear expression in ref. 21 for λ) is shown as the solid line for comparison. We note that the observed exponents follow the general trend predicted, changing continuously from $1/z = 0.5$, close to the expected location of the glass transition ($T_{\text{loc}} \approx 0.8$ T_d), toward zero as $T \to T_d$. For T close to T_d, the dynamics becomes exceedingly slow, making it difficult to access the asymptotic region. For $T \approx T_{\text{loc}}$, we also observed some finite-size effect. The overall agreement between the scaling theory and numerical results is within $5 \sim 10\%$ over the range tested.

In Fig. 2*a*, we show the dependence of the average bubble length on time for different Ts. The data depict the slow, logarithmic growth of the bubble length. Logarithmic growth is one of the signatures of glassy dynamics. Its occurrence in this particular system can be understood quantitatively as follows: The optimal bubble size $L^*(N)$ in a segment of length N depends logarithmically on N; see Eq. **15**. On the other hand, for a bubble placed at an arbitrary position in a long sequence, the effective sequence length is the distance the bubble can explore within a
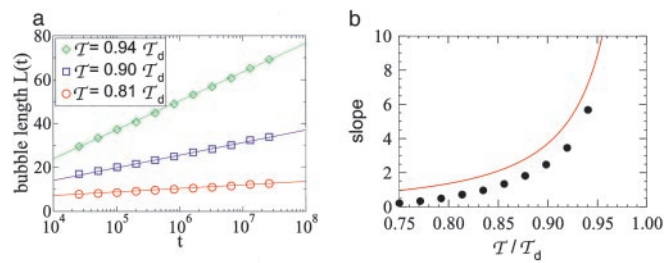
time t, i.e., $N \sim t^{1/z}$ for the subdiffusive dynamics expected in the glassy regime. Hence,

$$L^*(t) \approx \frac{1}{z \cdot H^*} \ln t + \text{const.} \quad [28]$$

is the expected length of the optimal bubble within a time t. Generally, we expect $L^*(t)$ to be the upper bound of the observed bubble length $L(t)$, with $L(t) \approx L^*(t)$ for large t deep in the glass phase. However, outside the glass phase, $L(t)$ must be finite even for $t \to \infty$.

In Fig. 2*b*, we show the coefficients of the observed logarithmic time dependence of $L(t)$ for Ts throughout the range $T_{\text{loc}} < T < T_d$. Also shown is the upper bound $1/(z \cdot H^*)$ (solid line) according to **28**, using the expression **22** for H^*. We note that the difference between the data and the upper bound is nearly constant (≈ 1) for the range studied.

Aging. Perhaps the most characteristic feature of glassy dynamics is that the system ages, e.g., the temporal fluctuation of the system depends on how long the system has evolved from some (arbitrary) initial condition (14, 15): the longer it has evolved, the slower it fluctuates. This is easy to understand in the context of a rough energy landscape with deep valleys and high barriers, since the longer the system evolves, the deeper the energy valley it finds, and hence the higher the barrier it will have to overcome to travel farther. This feature is in marked contrast to subdiffusive hydrodynamic systems that are time-translationally invariant.

Quantitatively, we can define the aging phenomenon via the time-dependent correlation function $C(t_w, \Delta t)$, which measures how much the system changes in time Δt, after first evolving for a waiting period t_w from the initial condition. Let us define a binary variable $\eta_i(t) \in \{0, 1\}$, for each base i of the nucleotide sequence. $\eta_i(t)$ takes on the value 1 if base i is open and belongs to the bubble at time t, and the value 0 if base i is paired. The correlation function, defined as $C(t_w, \Delta t) \equiv \Sigma_i \; \eta_i(t_w) \eta_i(t_w + \Delta t)$ after averaging $>10,000$ random sequences, is a measure of the
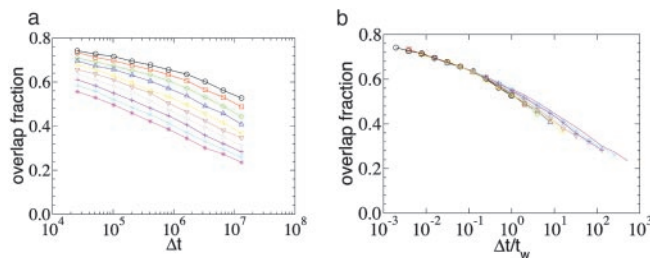

Fig. 2. (*a*) Average bubble length vs. time for several Ts; the solid lines are fits to the form $L(t) = a + b \ln t$. (*b*) Slope b of the logarithmic time dependence of $L(t)$ for various Ts. The solid line is the corresponding quantity for the upper bound of the bubble length $L^*(t)$; see text.

Fig. 3. Aging plot. (*a*) Average overlap fraction $C(t_w, \Delta t)/L(t_w)$ for different t_ws (from 1× to 512× 25,600 Monte Carlo steps) deep in the glass phase, with $T = 0.9 \; T_d > T_{\text{loc}}$. (*b*) Scaling plot of *a* with Δt normalized by t_w.

average self-overlap of the bubble at time t_w and $t_w + \Delta t$. A more convenient quantity to characterize is the fraction of overlap, $C(t_w, \Delta t)/L(t_w)$, where $L(t) = \Sigma_i \eta_i(t)$ is the instantaneous bubble length.

In Fig. 3a, we show the overlap fraction, parameterized by the different waiting time t_ws for the system biased deep in the glass phase with $T = 0.9 \, T_d$. The overlap fraction clearly depends on the waiting time, illustrating the glassy nature of the dynamics. In contrast, the same quantity computed for $T < T_d$ (data not shown) gives no statistically significant dependence on t_w. To characterize more quantitatively the behavior, we replot in Fig. 3b the curves in Fig. 3a with Δt normalized by t_w. We find these curves to collapse reasonably onto a single master curve that exhibits a weak kink at $\Delta t/t_w \sim 1$. A naive explanation of this behavior is that for $\Delta t \ll t_w$, the bubble stays approximately within the energy valley found at time t_w, whereas for $\Delta t \gg t_w$, the bubble makes an excursion faraway from the valley. For the one-dimensional trap model, it was shown rigorously (33) that $C(t_w, \Delta t)$ indeed scales as a function of $\Delta t/t_w$, even though the largest trap time actually scales sublinearly with t_w. This behavior can be understood in terms of the particle making multiple returns to the original valley after escaping it (32), as manifested by the slow decay shown in Fig. 3b for $\Delta t \gg t_w$.

Discussion

In this study we investigated the thermodynamic and dynamic behaviors of twist-induced denaturation bubbles in a long, random sequence of DNA. The small bubbles associated with weak twist are delocalized, e.g., they flicker in and out of existence according to the Boltzmann distribution and are independent of the DNA sequence. The bubbles increase in lengths upon increase in the applied torque. When the largest bubbles reach a critical size L_{loc} which is of the order of a few tens of bases, the bubbles become localized to AT-rich segments which occur statistically in a long random sequence. According to the parameters (19) taken at 37°C with $[Na^+] = 1$ M, the localization "transition" occurs at $T_{loc} \approx 8$ pN·nm, which is

~80% of the torque needed for bulk denaturation T_d. In the localized regime, the bubbles exhibit aging and move along the double-helix subdiffusively, with continuously varying dynamic exponents.

All of the results are obtained under the single-bubble approximation. Thermodynamically, we expect this approximation to be valid for DNA sequences of several thousand bases or less. This is due to the strongly cooperative nature of bubble formation, as manifested in the large initiation energy γ_1. The single-bubble description of dynamics is further restricted by the finite life time of the bubble: Even at length scales where the single-bubble approximation is appropriate thermodynamically, the bubble may annihilate and reappear elsewhere in the sequence, effectively performing long-distance hops. Experimental knowledge of the bubble lifetime in the presence of an applied twist is needed to estimate the crossover time to the long-distance hopping regime. Qualitatively, we expect these bubbles to have much longer lifetimes than the thermally denatured bubbles, since the applied twist plays the role of an energy barrier preventing bubble annihilation.

Finally, we note that bubble localization characterized in this study is a reflection of the statistical background present in long random nucleotide sequences. This background traps the bubble kinetically if the bubble size becomes sufficiently large. Thus, to localize denaturation bubbles at appropriate locations specified by designed sequences (e.g., promoters or replication origins) for biological functions, it is necessary to operate away from the localized regime, i.e., below the onset of localization.

1. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. (2002) *Molecular Biology of the Cell* (Garland, New York).
2. Kowalski, D., Natale, D. A. & Eddy, M. J. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 9464–9468.
3. Kanaar, R. & Cozzarelli, N. R. (1992) *Curr. Opin. Struct. Biol.* **2,** 369–379.
4. Strick, T., Allemand, J. F., Bensimon, D., Bensimon, A. & Croquette, V. (1996) *Science* **271,** 1835–1837.
5. Marko, J. F. & Siggia, E. D. (1995) *Phys. Rev. E* **52,** 2912–2938.
6. Cocco, S. & Monasson, R. (1999) *Phys. Rev. Lett.* **83,** 5178–5181.
7. Strick, T. R., Allemand, J. F., Bensimon, D. & Croquette, V. (2000) *Annu. Rev. Biophys. Biomol. Strut.* **29,** 523–543.
8. Fye, R. M. & Benham, C. J. (1999) *Phys. Rev. E* **59,** 3408–3426.
9. Cule, D. & Hwa, T. (1997) *Phys. Rev. Lett.* **79,** 2375–2378.
10. Tang, L.-H. & Chaté, H. (2001) *Phys. Rev. Lett.* **86,** 830–833.
11. Bockelmann, U., Thomen, P., Essevaz-Roulet, B., Viasnoff, V. & Heslot, F. (2002) *Biophys. J.* **82,** 1537–1553.
12. Lubensky, D. K. & Nelson, D. R. (2002) *Phys. Rev. E* **65,** 031917.
13. Binder, K. & Young, A. P. (1986) *Rev. Mod. Phys.* **58,** 801–976.
14. Struick, L. C. E. (1978) *Physical Ageing in Amorphous Polymers and Other Materials* (Elsevier, Houston).
15. Bouchaud, J.-P., Cugliandolo, L. F., Kurchan, J. & Mézard, M. (1998) in *Spin Glasses and Random Fields*, ed. Young, A. P. (World Scientific, Singapore), pp. 2613–2626.
16. Derrida, B. (1981) *Phys. Rev. B* **24,** 2613–2626.
17. Karlin, S. & Altschul, S. F. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 2264–2268.
18. Yu, Y.-k. & Hwa, T. (2001) *J. Comp. Biol.* **8,** 249–282.
19. SantaLucia, J., Jr., Hatim, T. A. & Senevirante, P. A. (1996) *Biochemistry* **35,** 3555–3562.
20. SantaLucia, J., Jr. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 1460–1465.
21. Blake, R. D., Bizzaro, J. W., Blake, J. D., Day, G. R., Delcourt, S. G., Knowles, J., Marx, K. A. & SantaLucia, J., Jr. (1999) *Bioinformatics* **15,** 370–375.
22. Fisher, M. E. (1966) *J. Chem. Phys.* **45,** 1469–1473.
23. Kafri, Y., Mukamel, D. & Peliti, L. (2000) *Phys. Rev. Lett.* **85,** 4988–4991.
24. Garel, T., Monthus, C. & Orland, H. (2001) *Europhys. Lett.* **55,** 132–138.
25. Rouzina, I. & Bloomfield, V. A. (2001) *Biophys. J.* **80,** 882–893.
26. Gerland, U., Moroz, J. D. & Hwa, T. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 12015–12020.
27. Karlin, S. & Dembo, A. (1992) *Adv. Appl. Probab.* **24,** 113–140.
28. Karlin, S. & Altschul, S. F. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 5873–5677.
29. Marinari, E. & Parisi, G. (1993) *J. Phys. A* **26,** L1149–L1156.
30. Alexander, S. (1981) *Phys. Rev. B* **23,** 2951–2960.
31. Machta, J. (1985) *J. Phys. A* **18,** L531–534.
32. Bertin, E. M. & Bouchaud, J.-P. (2003) *Phys. Rev. E* **67,** 026128.
33. Fontes, L. R. G., Isopi, M. & Newman, C. M. (2002) *Ann. Probab.* **30,** 579–604.