Compaction and Tensile Forces Determine the Accuracy of Folding Landscape Parameters from Single Molecule Pulling Experiments

Greg Morrison,1,2 Changbong Hyeon,3 Michael Hinczewski,2 and D. Thirumalai2

1School of Engineering and Applied Science, Harvard University, Cambridge, Massachusetts 02138, USA
2Biophysics Program, Institute For Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA
3School of Computational Sciences, Korea Institute for Advanced Study, Seoul 130-722, Republic of Korea

(Received 23 August 2010; published 29 March 2011)

We establish a framework for assessing whether the transition state location of a biopolymer, which can be inferred from single molecule pulling experiments, corresponds to the ensemble of structures that have equal probability of reaching either the folded or unfolded states \( P_{\text{fold}} = 0.5 \). Using results for the forced unfolding of a RNA hairpin, an exactly soluble model, and an analytic theory, we show that \( P_{\text{fold}} \) is solely determined by \( s \), an experimentally measurable molecular tensegrity parameter, which is a ratio of the tensile force and a compaction force that stabilizes the folded state. Applications to folding landscapes of DNA hairpins and a leucine zipper with two barriers provide a structural interpretation of single molecule experimental data. Our theory can be used to assess whether molecular extension is a good reaction coordinate using measured free energy profiles.

DOI: 10.1103/PhysRevLett.106.138102 PACS numbers: 87.10.−c, 87.15.Cc, 87.64.Dz, 87.80.Nj

The response of biopolymers to mechanical force \( f \), at the single molecule level, has produced direct estimates of many features of their folding landscapes, which in turn has given a deeper understanding of how proteins and RNA fold. In particular, single molecule pulling experiments directly measure the distribution of forces needed to rupture biomolecules, roughness, and shapes of folding landscapes [1–4]. Such measurements have made it possible to directly measure the distribution of forces needed to rupture biomolecules, roughness, and shapes of folding landscapes [1–4]. This implies that a single degree of freedom accurately describes the behavior of the multiple degrees of freedom explored by the biomolecule. The structural meaning of \( \Delta x^\dagger \), a parameter that is unique to single molecule force spectroscopy, has never been made clear. Despite many subtleties in determining \( \Delta x^\dagger \) from measurements [5–7], \( x_{\text{TS}} \) is most easily identified as a local maximum of the free energy profile at the transition midforce \( f_m \), \( F(R|f_m) \), which can be constructed by measuring the statistics of end-to-end distance \( P(R) \) at \( f = f_m \) [1–4,8,9]. This method has been experimentally used to obtain sequence-dependent folding landscapes of DNA hairpins [2] and, more recently, proteins [3]. In order to render physical meaning to \( \Delta x^\dagger \) we address two questions here: (i) Does \( \Delta x^\dagger \) represent the structures in the transition state ensemble (TSE)? The TSE is a subset of structures that have equal probability of reaching the NBA or unfolded basin of attraction (UBA). (ii) Can a molecular tensegrity (short for tensional integrity) parameter [10] \( s = f_c/f_m \), expressing balance between the internal compaction force \( f_m = \Delta F_{\text{UF}}/\Delta x_{\text{UF}} \) and the tensile force \( f_c = \Delta F_{\text{T}}(f_m)/\Delta x_{\text{T}}(f_m) \), describe the adequacy of \( \Delta x^\dagger(f_m) \) in describing the TSE structures?

We use simulations of a RNA hairpin and an exactly soluble model, both of which are apparent two-state folders as indicated by \( F(R|f) \), to answer the two questions posed above. The TS is a surface in the multidimensional folding landscape (stochastic separatrix [11]) across which the flux to the NBA and the UBA is identical. This implies that the fraction of folding trajectories corresponding to \( \Delta x^\dagger \) that start from the TS should have equal probability \( P_{\text{fold}} = 0.5 \) [12]) of reaching the NBA and the UBA [12,13]. At \( f = f_m \), the mean dwell times in the NBA and the basin of attraction corresponding to unfolded conformations (UBA) are identical, so that \( \int_{0}^{\infty} dR P(R) = \int_{0}^{\infty} dR P(R) = 0.5 \).

However, it is unclear whether or not the barrier top position is consistent with the requirement \( P_{\text{fold}} \approx 0.5 \) in force experiments.

To directly ascertain whether the barrier top position denotes the TSE (requiring \( P_{\text{fold}} = 0.5 \) in force experiments), we study folding of PSGA, a RNA hairpin, for which the NBA and UBA are equally populated at
of the interior bond interaction, respectively, and for this set of parameters, the transition midforce [where $P_{\text{fold}} = 0.65$ [Fig. 1(c)]], which is similar to the value obtained by analyzing the folding trajectory. An examination of the individual trajectories reveals that many molecules, initially with a gradient toward the UBA, recross the transition barrier to reach the NBA [blue trajectories in Fig. 1(b)]. Conversely, most of the molecules directly reach $R = R_N$ if they initially fall into the NBA, showing few recrossing events. Although the precise percentage of molecules reaching the UBA or NBA depends on the particular value of the boundary ($R_N$ and $R_U$), our simulations emphasize the importance of the recrossing dynamics, which is known to cause significant deviations from the transition state theory.

Deviation from $P_{\text{fold}} = 0.50$ suggests that the global coordinate $R$ alone is not sufficient to rigorously describe the hopping kinetics of P5GA. At least one other auxiliary coordinate is needed, and for structural reasons we take it to be the number $N_{bp}$ of base pairs [2]. The broad asymmetric distribution of $N_{bp}$ within the narrow TS region $[R = (4.1-4.2)]$ implied by $F(R|f_m)$ [Fig. 1(d)] shows that hopping kinetics at $f_m$ should be described by a multidimensional (at least two) folding landscape even though the $f$-dependent rates of hopping between the NBA and the UBA can be reliably predicted by using $F(R|f_m)$ [9].

To further illustrate if the $R$ coordinate alone is sufficient to determine the TSE structures, we consider an analytically solvable generalized Rouse model (GRM) [9] that has a single bond in the interior of the chain whose presence corresponds to the NBA. The GRM Hamiltonian [15]

$$\beta H = \frac{3}{2a^2} \int_0^N ds \left[ \frac{1}{2} \left| \frac{\partial s}{\partial t} \right|^2 - \beta F \int_0^N ds \right] + \beta V_c \left[ r(s_1) - r(s_2) \right],$$

(1)

where $V_c [x] = k(x^2 - c^2)/2$ for $x \leq c$ and $V_c [x] = 0$ for $x > c$, describes a simple Gaussian chain under tension with an additional cutoff harmonic interaction at the interior points $s_1$ and $s_2$. The distribution of $R$ for the GRM, with $\Delta s = (s_2 - s_1)$, is

To obtain $P_{\text{fold}}$ we set $N = 22$ for the number of bonds, $a = 0.545 \text{ nm}$ for their spacing, $\beta = 1/k_BT$, $\beta k = 1.65 \text{ nm}^2$, $c = 4 \text{ nm}$ for the strength and cutoff distance of the interior bond interaction, respectively, and $\Delta s = 18$. For this set of parameters, the transition midforce [where $P(x < c) = P(x > c)$] is $f_m = 16.8 \text{ pN}$. The two minima and the position of the barrier top of the free energy $\beta F(R|f_m) = -\log[P(R)]$ are easily determined as $R_N = 3.91 \text{ nm}$, $R_U = 9.39 \text{ nm}$, and $R_{TS} = 6.00 \text{ nm}$, respectively [Fig. 2(a)].

Using Brownian dynamics simulations [14] with the multi-dimensional energy function for the hairpin [8]. We find that $P_{\text{fold}} = 0.65$ [Fig. 1(c)], which is similar to the value obtained by analyzing the folding trajectory. An examination of the individual trajectories reveals that many molecules, initially with a gradient toward the UBA, recross the transition barrier to reach the NBA [blue trajectories in Fig. 1(b)]. Conversely, most of the molecules directly reach $R = R_N$ if they initially fall into the NBA, showing few recrossing events. Although the precise percentage of molecules reaching the UBA or NBA depends on the particular value of the boundary ($R_N$ and $R_U$), our simulations emphasize the importance of the recrossing dynamics, which is known to cause significant deviations from the transition state theory.

Deviation from $P_{\text{fold}} = 0.50$ suggests that the global coordinate $R$ alone is not sufficient to rigorously describe the hopping kinetics of P5GA. At least one other auxiliary coordinate is needed, and for structural reasons we take it to be the number $N_{bp}$ of base pairs [2]. The broad asymmetric distribution of $N_{bp}$ within the narrow TS region $[R = (4.1-4.2)]$ implied by $F(R|f_m)$ [Fig. 1(d)] shows that hopping kinetics at $f_m$ should be described by a multidimensional (at least two) folding landscape even though the $f$-dependent rates of hopping between the NBA and the UBA can be reliably predicted by using $F(R|f_m)$ [9].

To further illustrate if the $R$ coordinate alone is sufficient to determine the TSE structures, we consider an analytically solvable generalized Rouse model (GRM) [9] that has a single bond in the interior of the chain whose presence corresponds to the NBA. The GRM Hamiltonian [15]
projected onto the $R$ coordinate using both sets of parameters exhibit a number of recrossing events.

To understand the relation between $x$ (the structural coordinate which specifies the NBA and UBA in the GRM) and chain extension $R$ (pulling coordinate that is conjugate to $f$), we approximate the sharp interaction in Eq. (1) by a smoothed potential:

$$\beta V_s(x) = \beta V_{KS}(x) = -\log(e^{-\beta(x^2-c^2)/2} + 1),$$

by taking advantage of the clear separation between the UBA and the NBA [Fig. 2(a)]. Defining $\mathcal{N}_{KS} = \langle e^{-\beta(x^2-c^2)/2} \delta(|R| - R) \rangle_0$, with $\langle \ldots \rangle_0$ denoting an average over the Gaussian backbone, we can compute approximately $P_R(x < c) = \int_0^1 dx \delta(|R| - R) = \mathcal{N}_{KS} / (\mathcal{N}_{KS} + \mathcal{N}_F)$, with $\langle \ldots \rangle$ representing an average over the GRM potential in Eq. (3). The probability of bond formation satisfies

$$P_R(x < c) = \left[ 1 + \lambda^{3/2} \exp \left( \frac{3R^2\kappa\sigma^2}{2N^2\alpha^2\lambda} - \frac{\beta k c^2}{2} \right) \right]^{-1},$$

where $\kappa = Na^2\beta k/3$, $\sigma = \Delta s/N$, and $\lambda = 1 + \sigma(1 - \sigma)\kappa$. At $R = 6$ nm, Eq. (4) gives $P_R(x < c) \approx 0.402 \neq 0.5$. Thus, when the dynamics is initiated from the top of the apparent free energy barrier using the $R$ variable, their internal coordinate ($x$) is populated primarily with unfolded conformations [Fig. 2(b)]. Despite the fact that the UBA is primarily populated at $R = R_{TS}$, we find that $P_{\text{fold}} = 0.6$, so that the NBA will be predominantly populated as the trajectories progress. The midpoint of transition $[P_R(x < c) = 0.5]$ occurs at $R_{\text{mid}} = [(N^2a^2\lambda/3k\sigma^2)(\beta k c^2 - 3\log\lambda)]^{1/2} = 5.91$ nm, which deviates slightly from $R_{\text{TS}} = 6.00$ nm. The results from $V_s(x)$, which are in excellent agreement with the simulations as well as the numerical results using $V_s(x)$ (Fig. 2), also suggest that hopping kinetics in the GRM involves coupling between $x$ and $R$. Thus, even in this simple system, accurate location of the TSE should consider two-dimensional free energy profiles (see [6,16]).

To answer the second question, we introduce a molecular tensegrity parameter, which is a ratio of tensile force and a force that determines the stability of the biopolymer. The limit of mechanical stability of the NBA is determined by the critical unbinding force $f_c = \Delta F(t)/\Delta s(t)$. At the midpoint force $f = f_m$, the tensegrity parameter $s = f_c/f_m$ determines whether the applied external tension is sufficient to overcome the stability of the NBA. Models, which approximate the free energy profiles using a cubic or cusp potential [7], could alter $f_c$ or $f_m$ from the form suggested above. However, because $s$ involves the ratio of the two forces, the precise numerical factors are not relevant. Barrier crossings between the NBA over the TS are governed by the competition between $f_c$ and the applied force $f = f_m$. For $f_c \gg f_m$, barrier recrossing from the NBA to the TS will be extremely rare, while if $f_c \ll f_m$, barrier recrossings from within the NBA to the TS will decrease, and the probability of reaching $R = R_N$ will increase. Thus, $P_{\text{fold}}$, the structural link to $\Delta s$, should be determined by the experimentally measurable tensegrity parameter $s$.

To confirm this expectation, we ran 20 different sets of GRM parameters (1000 runs each), with $f_m$ ranging between 5 and 25 pN, $\beta k$ ranging from 0.1 to 600 nm$^{-2}$, $c$ between 0.26 and 6 nm, and $\Delta s$ between $N/2$ and $N$. Figure 3 shows that $P_{\text{unfold}}$ on $s$ (Fig. 3) can be derived by considering diffusion in a 1D version of the GRM potential in Eq. (1): $U(x) = -fx + \frac{1}{2}k_s x^2 + V_s(x)$, with $k_s$ being the curvature in the bottom of the UBA. Assuming constant diffusivity, $P_{\text{unfold}} = \int_{x_N}^{x_U} dx e^{\beta U(x)} / \int_{x_N}^{x_U} dx e^{\beta U(x)}$ [17], where the positions of the transition barrier, native, and unfolded wells are given by $c$, $x_N$, and $x_U$, respectively. For $s \gg 1$, we obtain $P_{\text{unfold}} = (1 + s)^{-1}/2$, with the entire dependence on the energy landscape encoded in $s$. The analytical result, plotted as a solid curve in Fig. 3, captures the overall trend of the GRM and PSGA simulations.

For $s \ll 1$, $P_{\text{unfold}}$ can be approximated as

$$P_{\text{unfold}} = \frac{1}{2} \left[ \frac{\Phi}{1 + \Phi} + \frac{32f_m^2c^2}{\pi k_U(\Phi + 1)^3 s} \right]^{-1},$$

where $\Phi = 4f_m/\sqrt{2\pi k_U T_{\text{K}}}$ and $k_U = \partial^2 U/\partial x^2$ at $x = x_U$. This relation can be used to predict the degree
FIG. 3 (color online). $P_{\text{unfold}}$ as a function of $s = \Delta F^{1}/f_{m}\Delta x^{1}$ for a variety of GRM parameters (colored circles). Red corresponds to small $\beta k \sim (0.1–1) \text{ nm}^{-2}$, purple to intermediate $\beta k \sim (1–50) \text{ nm}^{-2}$, and blue to high $\beta k \sim (50–500) \text{ nm}^{-2}$. The point sizes indicate the magnitude of the midpoint force, with the smallest points at $f_{m} = 5 \text{ pN}$ and the largest at $f_{m} = 25 \text{ pN}$. The large square indicates the results of the P5GA simulations with $P_{\text{unfold}} = 0.35$. The gray line is the theoretical prediction $P_{\text{unfold}}(s) = (1 + s)^{-1}/2$ for $s \gg 1$. The solid line in the inset shows the predicted $P_{\text{unfold}}(s)$ for $s \ll 1$ [Eq. (5)] using the DNA hairpin free energy profiles (not deconvolved) in Ref. [2]. The dashed line is for the leucine zipper, which has two barriers, using data from Ref. [3]. As explained in Ref. [18], numbers 1 and 2 are for NBA $\rightarrow$ I and I $\rightarrow$ UBA transitions, respectively. Because of different $k_{U}$ values [Eq. (5)], dashed and solid lines do not coincide.

Our results can be confirmed by analyzing long time folding trajectories $R(t)$ as long as multiple hopping events occur. Because $\Delta x^{1}$ is $f$-dependent, it follows that tensegrity can be altered by changing $f$. Thus, the extension may be a good reaction coordinate over a certain range of $f$ but may not remain so under all loading conditions.

This work was supported in part by grants from National Research Foundation of Korea (No. 2010-0000602) (C. H.) and National Institutes of Health (No. GM089685) (D. T.).


