# Chain persistency in single-stranded DNA

Anirban Sain,[1] Bae-Yeun Ha,[1] Heng-Kwong Tsao,[2] and Jeff Z. Y. Chen[1,*]
[1]*Department of Physics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1*
[2]*Department of Chemical Engineering, National Central University, Chung-Li, Taiwan 320, Republic of China*
(Received 15 August 2002; revised manuscript received 10 March 2004; published 15 June 2004)

We develop a theoretical approach to hairpin-loop formation of single-stranded (ss) DNA by treating the strand as a two-state system in which bases are either "stacked" or "unstacked." The looping kinetics of ssDNA is shown to be intrinsically different from that of a wormlike chain; it is mainly controlled by stacking-breakage probability, not by the mean curvature of loops, and highly sensitive to the composition of the loop as seen in recent experiments. Our estimate of a stacking energy for poly($dA$), −3.9 kcal/mol, is consistent with known results.

DOI: 10.1103/PhysRevE.69.061913 PACS number(s): 87.15.He, 36.20.Ey, 87.15.Aa

Beyond the genetic code that it carries in the sequence, DNA also displays various conformational properties, which are crucial to its biological functions. For example, the binding affinity of some DNA-binding ligands often relies on the flexibility of their binding site on the DNA [1]. Also, the looping kinetics of DNA is sensitive to the chain stiffness [2–5]. DNA strands are often modeled as a wormlike chain (WLC) [cf. Fig. 1(a)] [6]. However, recent experimental studies unambiguously demonstrated the limitation of the WLC model: it fails to describe adequately conformational fluctuations of homogeneous single-stranded (ss) DNA [e.g., poly($dA$) and poly($dT$)] with attractive stems at both ends of the chain or more simply DNA beacons [2–4,7]. These experiments suggest that the rigidity of these molecules is mainly determined by the so-called "stacking" interactions between two consecutive bases along the strand. This leads to stacking-sensitive (thus sequence-dependent) looping kinetics, intrinsically different from that of a WLC [7].

Despite significant effort in the past [8–10], a consistent theoretical model of ssDNA has, so far, been lacking. The main difficulty lies in that the origin of base stacking, which gives rise to stacking-sensitive chain persistency, has not been well understood despite recent all-atom molecular modeling [11]. The purpose of this paper is twofold. We first introduce a simple but physically motivated model to describe the stacking-induced rigidity. Then, we apply this model to explain some of recent experimental observations on DNA beacons by Goddard *et al.* [7]. Our approach can be complimentary to, but will provide a more systematic interpretation than, a thermodynamic consideration by Aalberts *et al.* [12].

Here, we do not attempt to further clarify the origin of base stacking. Instead we adopt a two-state model of ssDNA, hereafter called a *stacking chain model* [cf. Fig. 1(b)], in which bases are classified as either "stacked" or "unstacked" [13]. This model captures base stacking, as illustrated in Fig. 1(c), at a coarse-grained level. This simplification is invoked by the fact that the hydrophobicity of bases, which is believed to be mainly responsible for base stacking, is short ranged [11]. For simplicity, bases are assumed to have the same segment length $\ell$. If and only if two neighboring bases are aligned in parallel or stacked, they gain an energy of $-\epsilon$ ($\epsilon>0$, by convention). Note that two stacked bases are not allowed to rotate around their axis, losing an entropy associated with rotational degrees of freedom. A typical stacking energy for $AA$ pairs ranges from the earlier estimate [9] of $-\epsilon=-6kT$ (or −4 kcal/mol at room temperature) to more recent values of the same order of magnitudes [11], depending on the solvent environment, where $k$ is the Boltzmann constant and $T$ is the temperature. If unstacked, the bases are free to bend and rotate without any energy cost. The conformation of such a stacking chain is fully characterized by the distribution of $\theta_i$ [cf. Fig. 1(b)], an angle between segments $i$ and $i+1$ [14],

$$P(\{\cos\ \theta_i\}) = \prod_{i=1}^{N-1} [1 + w\delta(\cos\ \theta_i - 1)], \tag{1}$$

where

$$w \equiv \Omega[\exp\ (\beta\epsilon) - 1] \tag{2}$$

and $\beta \equiv 1/kT$. The entropy associated with the degrees of freedom for rotating the base plates is approximately incor-

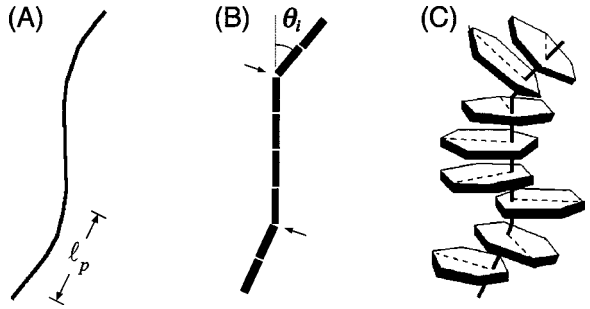


FIG. 1. Schematic representation of (a) a wormlike chain (WLC) that tends to bend smoothly (it remains rodlike within $\ell_p$, the persistence length), and (b) a stacking chain that models ssDNA as schematically shown in (c). Unlike the WLC, the stacking chain allows abrupt bending or stacking breakage, as marked by an arrow in (b).

*Corresponding author.

 ©2004 The American Physical Society

porated into the model through the parameter $\Omega$. Physical quantities, such as loop-formation probability and "equilibrium" closing times, can be derived from the restricted partition function, i.e., a partition function summed over all realizations of chain conformations (or $\theta_i$) with the end-to-end vector $\vec{R}$ fixed in space.

The underlying assumption in our stacking model [cf. Eq. (1)] is that base stacking is a short-ranged effect *and* that there is no other force that influences the chain stiffness [15]. Then it can readily be seen that the energy cost for bending entirely depends on the number of bonds broken, not the radius of curvature as in a WLC. The resulting persistence length should vary inversely with the probability of breaking base stacking: $\ell_p \propto \exp(\epsilon/kT)$. This is in contrast to $\ell_p(\propto 1/kT)$ in a WLC [6]. Also the stacking rigidity is distinct from that of double-stranded (ds) DNA, which is complicated by base pairing between two complimentary base pairs and the geometrical constraint imposed by the doubleness of the chain, for example. To a certain extent, dsDNA can be well characterized as a geometrical object, as implied by the WLC model.

In our model, the end-to-end distance ditribution function, for a given total bond number $N$, assumes the following hierarchical form in the Fourier space:

$$G_{N+1}(k) = \sum_{m=0}^{N-1} G_{N-m}(k)\left(\frac{w}{2}\right)^m S_{m+1}(k), \tag{3}$$

where $G_1(k) = \sin(k\ell)/k\ell$, $G_0(k) = 1$, and $S_m(k) = \sin(mk\ell)/mk\ell$. The expression can be obtained from a diagrammatic expansion of $G_{N+1}(k)$, where each term in the series represents the probability of a particular chain conformation [16]. While the low-temperature, small-$N$ behavior of the model shows unique stacking properties and is relevant for the study of DNA structural formation, the high-temperature, large-$N$ behavior displays typical persistent chain behavior. For example, the mean square end-to-end distance for $N \gg 1$ in this model has the Kratky-Porod form

$$\langle R^2 \rangle = 2\ell_p L\{1 - \ell_p/L[1 - \exp(-L/\ell_p)]\}, \tag{4}$$

where

$$\ell_p = (1 + w)\ell/2 \tag{5}$$

is the effective persistence length and $L \equiv N\ell$.

Recent experiments on ssDNA probed the very nature of chain persistency of poly($dA$) and poly($dT$), which were shown to have a strong and weak stacking tendency, respectively [2,4,7]. In a fluorescence experiment, Goddard *et al.* attached a single strand of poly($dA$) and poly($dT$) of $N$ monomers to stems consisting of five-base sequence *TTGGG* at one end and its complementary *AACCC* at the other end. The closing time of each strand can be obtained by examining the fluorescent intensity of the fluorophore and the quencher attached to the tips of the chain, hence providing an indirect probe of the persistency that controls the closing kinetics of the strand (see Fig. 2).

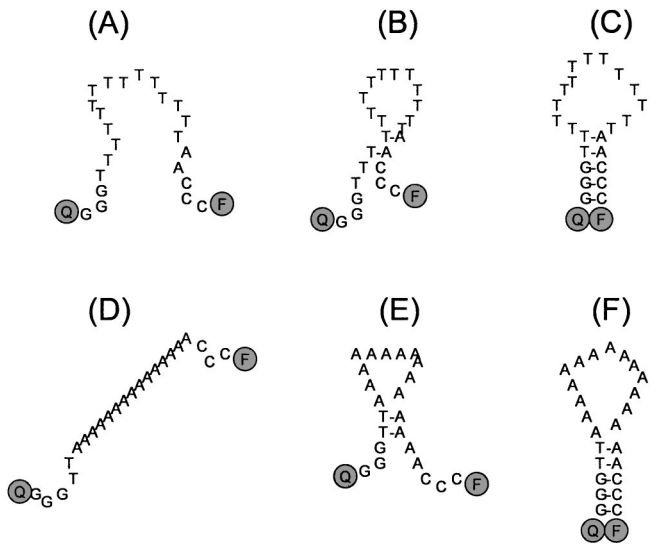


FIG. 2. Conformations considered in text: (a) open poly($dT$) conformation, (b) misfolded poly($dT$) conformation, (c) closed poly ($dT$) conformation, (d) open poly($dA$) conformation, (e) misfolded poly($dA$) conformation, and (f) closed poly-$A$ conformation. Using an analysis of the probability densities and binding energy scales for each of these plots, it can be shown that (b) and (c) should be included in the interpretation of poly($dT$) data, while (d) and (f) are crucial for the interpretation of poly($dT$) data in Ref. [7].

The equilibrium densities of the open and closed structures are related to each other through $\tau_o\rho_o = \tau_c\rho_c$, where $\tau_o$ and $\tau_c$ are, respectively, the characteristic opening and closing time, and $\rho_o$ and $\rho_c$ are, respectively, the chain densities of the open and closed conformations. If the closure of the stem causes an energy gain $\upsilon$ and an entropy loss $s$, we have

$$\tau_c = \tau_o e^{-\beta\upsilon+s}\rho_o^0/\rho_c^0, \tag{6}$$

where $\rho_o^0$ and $\rho_c^0$ are chain densities of the corresponding system *without* the attractive stem part. Since $\tau_0$ is expected to be proportional to $e^{\beta\upsilon-s}$ with no explicit $N$ dependence, $\tau_c$ directly scales as $\rho_c^0/\rho_c^0$. Three essential features emerge from this consideration.

(1) In the Gaussian-chain model, the probability of closed conformations scales as $b/l^3N^{3/2}$, where $b$ is a constant proportional to the volume of the attractive stem. On the other hand, the probability of the open conformations is approximately 1 with a possible correction of order $b/N^{3/2}$. Hence $\tau_c$ varies as $N^{3/2}$ and should be a temperature-*independent* constant according to the Gaussian model. This description is, however, inconsistent with the poly($dT$) data of Goddard *et al.* [7] that show weak temperature dependence of the closing time.

(2) The weak temperature dependence of the scaled $\tau_c/N^{3/2}$ in poly($dT$) can be attributed to misfolded conformations as illustrated in Fig. 2, where a smaller loop is formed due to the *AT* bonding of the *AA* segment in the stem with any *TT* segment in the poly($dT$) loop. In addition to other open conformations, this conformation contributes to $\rho_o^0$ since the flurophore and quencher ends are in an open posi-

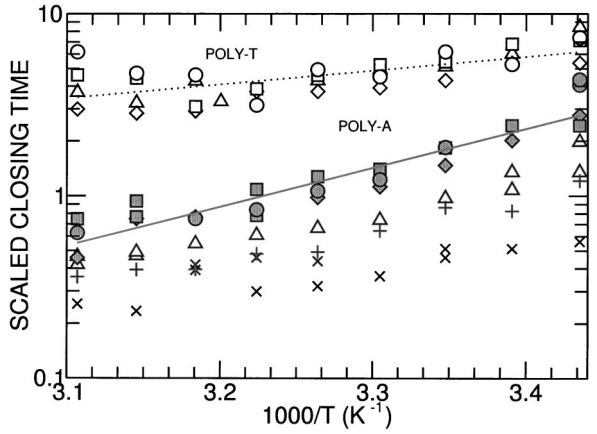


FIG. 3. The closing time ($\tau_c$) of a hairpin loop (in $\mu$s) as a function of temperature $T$: (a) poly($dT$) and (b) poly($dA$). Our results for the poly($dT$) and poly($dA$) are represented by the dotted and solid curves, respectively. Our scaling analysis predicts a straight line in this semilogrithmic plot. Our estimate of staking energy, deduced from the slope of the solid curve, is favorably comparable to previous estimates [see the relevant discussion below Eq. (10)]. Experimental data adopted from Ref. [7] are represented by various symbols: filled circles ($N=30$), filled squares ($N=21$), filled diamonds ($N=16$), triangles ($N=12$), crosses ($N=10$), and plus signs ($N=8$). For poly($dT$), the closing time has been multiplied by $10/[N^{1.5}B(N)]$ [Eq. (8)], where the factor 10 was introduced for clarity of the plot. For poly($dA$), the closing time has been multiplied by $1/N^{1.5}$ [Eq. (10)]. The large-$N$ data, represented by the filled symbols, approach the asymptotic straight line.

tion. Up to a multiplicative constant, the misfolded-chain density can be written as

$$\rho_o^0[\text{Misfolded}] = 1 + B(N)\exp(\beta u - \sigma) + \cdots, \tag{7}$$

where $u$ and $\sigma$ represent the energy gain and entropy loss, respectively, for forming the two $AT$ base bonds. The constant $B$ accounts for the entropy associated with misfolded loop formation and can be shown to have weak $N$ dependence $B(N)=\Sigma_{n=1}^{N-1} n^{-3/2}$. The steric repulsion between "wrong" base pairs (i.e., the $CA$ pairs near the two matched $AT$ pairs), however, contributes to $u$ and hence reduces the magnitude of the total binding energy $u$. The density of closed conformations, $\rho_c^0$, is unchanged in this case. Assuming $\beta u - \sigma \gg 1$, we can estimate the closing time as a function of temperature,

$$\tau_c[\text{poly}(dT)]/N^{3/2} \propto B(N)\exp(\beta u). \tag{8}$$

When misfolding occurs, the closing time $\tau_c$ becomes temperature dependent as implied by Eq. (8). This is in contrast to the previous case of the Gaussian chain model for which the "equilibrium" closing time is solely determined by the chain entropy. Hence, the weak temperature dependence, as seen in experiments [7], arises from the competition between the chain entropy and the excess energy due to misfolding. Figure 3 shows a semilogarithmic plot of the scaled $\tau_c/[B(N)N^{3/2}]$ versus $\beta$ for four poly($dT$) chains, adopted from Ref. [7]. The data collapse into an $N$-independent curve after the scaling. The slope can be used to determine $u$; $\beta u = 5.8$ at $T=300$ K. This misfolding picture in poly($dT$) agrees with the previous assertion that there is no pronounced stacking in poly($dT$) [17].

(3) To examine poly($dA$), we consider the large-$N$ limit $L \gg \ell_p$ in Eq. (3). The free energy change involved in stacking reaction was estimated to be $\Delta G = -1.5$ kcal/mol, implying $\ell_p \cong w\ell/2 \cong 6\ell$ [18]. Hence long chains (or equivalently $N \gg 6$) can be approximated by a Gaussian chain with $2\ell_p$ as an effective segment length. If $n$ is the total number of broken stacking bonds along the chain, and $C_o(N,n)$ and $C_c(N,n)$ are the distribution of $n$ for open and closed sates, respectively, then we have $\rho_o = \Sigma_n C_o(N,n)(4\pi)^n$ and $\rho_c = \Sigma_n C_c(N,n) n^{-3/2}$. While explicit calculations of $\rho_c/\rho_o$ based on Eq. (3) will be reported elsewhere [16], we here provide a meanfield approximation, which amounts to keeping a term that dominates the sum and becomes accurate as $N \to \infty$. We find that the closing probability at a mean-field level becomes $P(0) = \rho_c/\rho_o \sim b(\ell_p L)^{-3/2}$. This consideration, however, does not completely describe hairpin-loop formation. Due to the geometrical constraint on a loop conformation as illustrated in Fig. 2, loop formation should be accompanied by one stacking breakage in the stem with an energy cost $\epsilon$—note that the stem size is smaller than $\ell_p$, preferably assuming a rodlike structure when it is in an open conformation. Hence we have

$$\rho_c^0 \propto w^{-1}P(0) \propto (\omega^{-1}\ell_p^{-3/2})(Nl)^{-3/2}. \tag{9}$$

From Eq. (6) the closing time $\tau_c$ is

$$\tau_c \propto N^{3/2}w^{5/2} \propto N^{3/2}\exp(2.5\epsilon/kT), \tag{10}$$

where a large $w$ approximation has been used in view of Eq. (5). Note here that the origin of the exponential dependence of $\tau_c$ in Eq. (10) is different from that in Eq. (7) in that the former arises from the stacking interaction between two consecutive bases while the latter comes from the hydrogen bonding between a pair of $A$-$T$ separated by a rather long distance along the contour. The exponential dependence in both cases accounts for the energetics involved in chain misfolding [Eq. (7)] or chain closing [Eq. (10)].

In Fig. 3, we plot $\tau_c/N^{3/2}$ for poly($dA$) in a semilogarithmic scale as a function of $1/T$ for various $N$ ranging from $N=8$ to 30. This plot shows how the experimental data of Ref. [7] approach an $N$-independent straight line for large $N$, i.e., $N=30$, 21, 16. This asymptotic behavior is indeed consistent with Eq. (10). A fit to the three higher-$N$ sets of the poly($dA$) data to Eq. (10) gives $\beta\epsilon = 16.3/2.5$ at $T=300$ K, or $\epsilon = 6.5\ kT = 3.9$ kcal/mol per base stacking. This value can be compared with $\epsilon \approx 3.6$ kcal/mol obtained from a recent numerical simulation [11]. It should be noted that the experiments of Goddard *et al.* were conducted in the presence of 0.25 $M$ NaCl, while the solvent effect was not explicitly considered in the modeling of Luo *et al.* [11]. Earlier estimates of the stacking energy also fall within this range. In particular, Dewey and Turner estimated $\epsilon$ to be 4 kcal/mol for poly ($dA$) and 3.2 kcal/mol for poly($dA$) in 0.05$M$ ionic solution [9], based on the thermodynamic analysis of a temperature-jump study; Breslauer and Sturtevant [19] obtained $\epsilon \approx 3.4$ kcal/mol, based on a differential scanning calorimetry

measurement of a chain involving 7 adenine ($A$) bases. Finally, our Eq. (10) implies that, as $T\to\infty$, $\tau_c$ will get saturated at the corresponding value for a Gaussian chain: $\tau_c \propto N^{3/2}$.

The stacking-chain model has very unique implications for loop formation. For WLCs, the energy cost for forming a loop depends on the radius of curvature of the chain and thus the chain length. As a result, longer chains can form a loop more easily if $L \lesssim \ell_p$ [5]. In sharp contrast to this case, the energy cost in the stacking chain model does not depend on the curvature. It rather depends on the number of additional broken bonds caused by chain looping, which were previously stacked next to each other. In this case, the energy cost does not depend on the chain length, since the chain can fold up into a loop as long as a few stacked positions in the middle are broken.

This physical picture presented here, however, deviates from that of Goddard *et al.* [7] that the enthalpic barrier $\Delta H_c$ to looping varies linearly with $N$: $\Delta H_c \sim N$. The $N$ dependence of $\Delta H_c$ in their Fig. 3 was read off by analyzing the slope of the linear fit to their data for $\ln \tau_c$ for various $N$. This estimation is, however, not very conclusive, considering the deviation of their data from the fitting line (see their Fig. 3). On the other hand, we have used a scaling form derived from a microscopic model, which becomes very accurate in the asymptotic limit $N \gg 6$ or $L \gg \ell_p$. As clearly indicated in Fig. 3, the poly($dA$) data for $N=16,21,30$ collapse into a single straight line in this plot, implying an $N$-independent enthalpic barrier. Analysis of short chains (i.e., $N=8,10,12$), however, requires a more complete theory, which will be reported elsewhere [16]. The looping of a short chain would necessitate additional breakage of stacking but otherwise our general picture of stacking rigidity remains valid.

Ansari *et al.* noted that the misfolded conformations, as displayed in their Fig. 2, would trap the poly($dA$) chain energetically and cause a possible nonexponential relaxation of the system [3,20]. A much more involved model was then developed to include the misfolded conformations [3]. Goddard *et al.*, on the other hand, left out the possibility of misfolded intermediate states, and attempted to interpret their experimental results based on an enthalpic consideration alone [21]. This discrepancy can be resolved in our simple energetic analysis presented above; the misfolded states have a typical binding energy of $6kT$ estimated from the poly($dT$) measurement at room temperature. Hence the misfolded states are important in interpreting the poly($dT$) experiment as these states would be treated by the florescence experiments as open states. On the other hand, the misfolded states accompany the breakage of a stacked poly($dA$) at the expense of a typical energy of $2.5\epsilon = 16.3kT$ at room temperature, and are thus disfavored energetically in poly($dA$).

In summary, we have used a simple model of ssDNA, i.e., the stacking chain model, to account for stacking-sensitive looping of ssDNA as seen in recent experiments. It is worth noting that the type of bases can be sensitively reflected in *macroscopic* observables such as $\tau_c$. As a result, the closing dynamics of the stacking chain is intrinsically different from that of a wormlike chain.

Financial support of this work was provided by the Natural Science and Engineering Research Council of Canada (B.Y.H. and J.Z.Y.C.).

[1] Y. Choo and A. Klug, Curr. Opin. Struct. Biol. **7**, 117 (1997), and references therein.
[2] G. Bonnet, O. Krichevsky, and A. Libchaber, Proc. Natl. Acad. Sci. U.S.A. **95**, 8602 (1998).
[3] A. Ansari, S. V. Kuznetsov, and Y. Shen, Proc. Natl. Acad. Sci. U.S.A. **98**, 7771 (2001).
[4] M. I. Wallace, L. Ying, S. Balasubramanian, and D. Klenerman, Proc. Natl. Acad. Sci. U.S.A. **98**, 5584 (2001).
[5] S. Jun, J. Bechhoefer, and B.-Y. Ha, Europhys. Lett. **64**, 420 (2003).
[6] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon Press, Oxford, 1986).
[7] N. L. Goddard, G. Bonnet, O. Krichevsky, and A. Libchaber, Phys. Rev. Lett. **85**, 2400 (2000).
[8] D. N. Holcomband and I. Jr. Tinoco, Biopolymers **3**, 121 (1965).
[9] T. G. Deweyand and D. H. Turner, Biopolymers **18**, 5757 (1979).
[10] D. Pörschke, Biopolymers **17**, 315 (1978).
[11] R. Luo, H. S. R. Gilson, M. J. Potter, and M. K. Gilson, Biophys. J. **80**, 140 (2001).
[12] D. P. Aalberts, J. M. Parman, and N. L. Goddard, Biophys. J. **84**, 3212 (2003).
[13] J. B. Mills, E. Vacano, and P. J. Hagerman, J. Mol. Biol. **285**, 245 (1999).
[14] In this model, pairing between complimentary pairs at stems is not included. See Eq. (6) for the consequence of base pairing on hairpin loop formation.
[15] In reality, the electrostatic repulsion between charged groups on the chain further stiffens the chain. In a low-salt limit, the electrostatic contribution $\ell_{\text{elec}}$ can be equally important. In a high-salt limit ($\sim 0.25M$) as in experiments in Ref. [7], we have estimated $\ell_{\text{elec}}$ to be $<1$ Å and it can be ignored.
[16] A. Sain, J. Z. Y. Chen, and B.-Y. Ha (unpublished)
[17] M. Riley and B. Maling, J. Mol. Biol. **20**, 359 (1966).
[18] T. N. Solie and J. A. Schellman, J. Mol. Biol. **33**, 61 (1968).
[19] K. J. Breslauer and J. M. Sturtevant, Biophys. Chem. **7**, 205 (1977).
[20] A. Ansari, Y. Shen, and S. V. Kuznetsov, Phys. Rev. Lett. **88**, 069801 (2002).
[21] N. Goddard, G. Bonnet, O. Krichevsky, and A. Libchaber, Phys. Rev. Lett. **88**, 069802 (2002).