Stretch-Induced Hairpin-Coil Transitions in Designed Polynucleotide Chains

Haijun Zhou,* Yang Zhang, and Zhong-Can Ou-Yang
Institute of Theoretical Physics, The Chinese Academy of Sciences, P.O. Box 2735, Beijing 100080, China
(Received 17 April 2000)

The structural property of a poly(*dG-dC*) or poly(*dA-dT*) nucleotide is investigated. At low force and room temperatures, the polymer takes on compact hairpin structures. An abrupt transition from hairpin to random coil occurs at certain critical forces, its high cooperativity is related to the unfavorable formation of hairpin and other kinds of looped structures. It is hence necessary to consider the enthalpic effects of single-stranded loops in realistic models of RNA folding. A possible new way to obtain the statistical weights of elementary nucleotide arrangements is by single-macromolecular mechanical measurements on specifically designed polynucleotides.

DOI: 10.1103/PhysRevLett.86.356 PACS numbers: 87.15.By, 05.50.+q, 64.60.Cn, 87.10.+e

The issues of protein and RNA folding [1–4] and DNA configurational transitions [5–9] are of current theoretical interest, and hairpin structures consisting of two paired strands with a single-stranded loop in between are crucial for the understanding of protein and RNA secondary structure formation. Recent models, which focused mainly on a qualitative discussion of the possibility of thermal phase transitions in protein [10] and RNA secondary structures [2–4], have emphasized the "base-pairing" interactions in protein or RNA but have more or less avoided considering the possible effects of single-stranded loops. Continuous or glassy phase transitions were predicted based on these models, indicating that in the vicinity of the transition temperature the structure of the polymer is not dominated by a single native state [3,4].

Recently, experimental data have been accumulated through single-macromolecule manipulation methods, and it appears that the stretch-induced structural transitions in biopolymers are radically different from those caused by heating. For example, a recent work performed by Rief and co-workers [11] showed that *designed* single-stranded DNA (ssDNA) chain made of poly(*dG-dC*) or poly(*dA-dT*) sequence experiences an abrupt transition from compact configurations to random coil upon the action of an external force, just like unzipping a single double-stranded DNA (dsDNA) by pulling its two strands apart [12]. It seems to be difficult for one to explain this experimental observation based on the models of Refs. [2–4,10]: Since any one segment in the designed ssDNA is complementary to any other segment in it, one would have expected that a great many looped structures would be formed in a long ssDNA to maximize conformational entropy; and these loops would then disappear gradually upon stretching, resulting in a smooth structural transition process.

To understand such stretch-induced transitions, we noticed that the above-mentioned experiment was performed at 25 °C, much less than the DNA melting temperature T_m, which is beyond 70 °C [13]. References [2–4] and [10] discussed the behavior of designed biopolymers as the temperature approaches T_m, where the energy penalty for the formation of hairpin and other kinds of loops (enthalpic effects) is only of minor importance and the property of the polymer system is determined by conformational entropy. However, DNA/RNA structural transitions occur at physiological conditions in living organisms; during such and the stretch-induced "melting" processes at low temperatures, it may be possible that enthalpic effects turn to be dominant. In other words, heat- and stretch-induced structural transitions may be governed by different aspects of the intrinsic thermodynamic properties of polynucleotide chains [14]. We have studied the elasticity of designed ssDNA based on this insight, with the enthalpic effects of DNA hairpin, bulge, and interior loops all being quantitatively taken into account [15,16]. As was consistent with the experiment of Reif *et al.* [11], the analytical calculations naturally demonstrated the occurrence of a highly cooperative transition between hairpin and stretched random coil configuration. The estimated free energy for unzipping one *G-C* or *A-T* base pair based on parameters determined by bulk methods [15,16] were in reasonable agreement with those estimated by mechanical measurement of Bockelmann *et al.* [17].

In a ssDNA chain made of tandem repeats of (*dG-dC*) or (*dA-dT*) dinucleotide, because hydrogen bonds can be formed between the complementary bases *G* and *C* (and *A* and *T*), each nucleotide (monomer) can be either unpaired or paired by another monomer to form a base pair. Thus, each monomer belongs either to the single-stranded coiled segment or to the base-paired hairpin segment (Fig. 1a). A coiled segment is just a linear sequence of monomers which are all unpaired, while each hairpin segment can be further divided into a loop region and a stem region (Figs. 1a and 1b). At low temperatures, the formation of multiloops and, hence, of branched structures is extremely difficult according to previous investigations [18], and we neglect such a possibility by assuming that each hairpin region contains only one hairpin loop [19]. The stem region is stabilized by hydrogen bonds between complementary bases and vertical stacking interactions between adjacent base pairs. It is possible for the appearance of bulge and interior loops in the stem (see Fig. 1b), since there may be a length mismatch in the two strands of the stem. To speak

 0031-9007/01/86(2)/356(4)$15.00

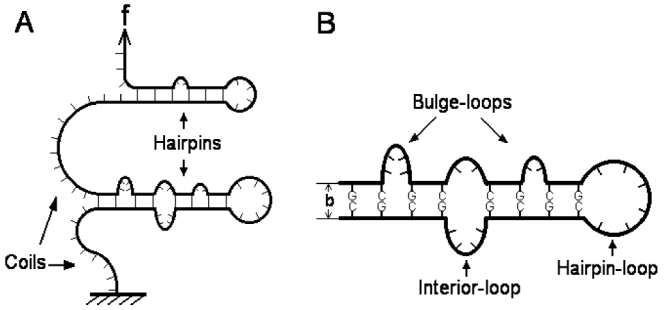

FIG. 1. (A) Schematic of a single-stranded poly(dG-dC) composed of coiled and hairpin regions, with the action of a constant force **F**. (B) A schematic microscopic structure of a hairpin segment.

briefly, we model a ssDNA self-complementary polymer as a linear sequence with coiled and hairpin segments.

To obtain the statistical property of such a model system under the action of a constant stretching force f requires the calculation of the free energy expression. We first obtain the statistical weight for a hairpin segment of $(2l + 2m)$ monomers, with $2l$ $(l = 2, 3, \ldots)$ monomers forming a hairpin loop and $2m$ $(m = 1, 2, \ldots)$ monomers in the stem region. The statistical weight for a hairpin loop of $2l$ monomers is denoted as $\gamma(2l)$ and that for a bulge or interior loop with r and s unpaired bases in the two strands is denoted as $\lambda(r, s)$. The cases of $r > 0$ and $s > 0$ correspond to interior loops, and those of r (or s) $= 0$ to bulge loops; $\lambda(0, 0) = 1$. Based on Fig. 1b, the total statistical weight is calculated to be

$$Z_{\rm hp}(2l, 2m) = \mu(f)\gamma(2l) \sum_{n=1} \sum_{\{r_i\}} \sum_{\{s_i\}} \times \delta\left(2n + \sum_{i=1}^{n-1}(r_i + s_i) - 2m\right) \times (t\tau)^{m-1-k} \prod_{i=1}^{n-1} \lambda(r_i, s_i), \quad (1)$$

where $k = \sum_{i=1}^{n-1}[1 - \delta(r_i)][1 - \delta(s_i)]$. In Eq. (1), t and τ mean, respectively, the statistical weight for the formation of a base pair and that for the stacking interaction between two base pairs. Because of the stretching, the hairpin segment as a whole might align along the force direction. Assuming the two ends of a hairpin are separated by a fixed distance b (see Fig. 1b), the statistical weight for this alignment effect, $\mu(f)$ in Eq. (1), is $\mu(f) = \sinh(\beta f b)/\beta f b$, with $\beta = 1/k_B T$ and T (the temperature) being set to 25 °C. We set $b = 1.0$ nm (about the distance between two $C_{1'}$ atoms in a DNA base pair [20]).

The discrete Laplace transform of $Z_{\rm hp}$ is

$$U_{\rm hp}(x) = \sum_{l=2}^{\infty} \sum_{m=1}^{\infty} Z_{\rm hp}(2l, 2m) x^{-(2l+2m)} = \frac{\mu(f) U_{\rm hploop}(x)}{x^2 - t\tau[1 + 2U_{\rm bulge}(x) + U_{\rm inter}/t\tau]}, \quad (2)$$

where $U_{\rm hploop} = \sum_{l=2}^{\infty} \gamma(2l) x^{-2l}$, $U_{\rm bulge} = \sum_{l=1}^{\infty} \lambda_{\rm bulge}(2l) x^{-2l}$, and $U_{\rm inter} = \sum_{l=1}^{\infty} (2l - 1)\lambda_{\rm inter}(2l) x^{-21}$ are defined, respectively, as the Laplace transforms of the statistical weight for hairpin, bulge, and interior loops. Here we have assumed that the statistical weight for the interior loop is related only with the sum of the number of unpaired bases in the two strands, i.e., $\lambda_{\rm inter}(r, s) \equiv \lambda_{\rm inter}(r + s)$.

The partition function for a coiled segment of n monomers can be obtained based on the observation that the ssDNA segment can be modeled as a wormlike chain of bending persistence length $\ell_p = 0.75$ nm and total contour length $L = na$ [21], with a being the monomer length. If the nucleotide sugar puckering mode in ssDNA is in $C_{3'}$-*endo* (the same as that in dsDNA), then $a = 0.6$ nm, and if in $C_{2'}$-*endo*, then $a = 0.7$ nm [13]; actually the sugar puckering mode may rapidly changing between $C_{3'}$-*endo* and $C_{2'}$-*endo* [13]. The force-related free energy density $g_{\rm coil}$ for a wormlike chain is evaluated variationally by the following equation [22]: $g_{\rm coil}(f) = (k_B T \eta/4\ell_p - f)\mathcal{L}(\eta)$, where $\mathcal{L}(x) = \coth x - 1/x$ is the Langevin function, and η is determined by

$$4\beta f \ell_p = \eta + \mathcal{L}(\eta)/B(\eta), \quad (3)$$

with $B(x) = \partial_x \mathcal{L}(x) = 1/x^2 - 1/\sinh^2 x$. In ssDNA sequences, there is still weak vertical stacking interactions of the order of thermal energy $k_B T$ between adjacent bases [13,23], and the total free energy of each nucleotide monomer in the coiled state should be $g_0 + a g_{\rm coil}$, with g_0 taking account of such stacking interactions. The Laplace transform for the partition function $Z_{\rm coil}(i)$ of the coil state is $U_{\rm coil} = \sum_{i=1}^{\infty} Z_{\rm coil}(i) x^{-i} = \nu/(x - \nu)$, with $\nu = \exp[-\beta(g_0 + a g_{\rm coil})]$.

Denote the total statistical weight for a poly(dG-dC) or poly(dA-dT) chain of N monomers as $Z(N)$; then based on Fig. 1a its Laplace transform is obtained through a lengthy but straightforward calculation to be

$$U(x) = \sum_{i=1}^{\infty} Z(i) x^{-i} = \frac{U_{\rm hp}(x)[1 + U_{\rm coil}(x)] + U_{\rm coil}(x)[1 + U_{\rm hp}(x)]}{1 - U_{\rm hp} U_{\rm coil}}. \quad (4)$$

For a linear polymer composed of a large number of monomers $(N \gg 1)$, the total free energy should be an extensive quantity proportional to N, i.e., $G(f) = -N k_B T \ln x_1$, with x_1 independent of N. Thus, we should expect that the total partition function be expressed asymptotically as $Z(N) = \exp(-\beta G) \simeq x_1^N$. Consequently, Eq. (4) indicates that x_1 is the largest solution of the equation $U_{\rm coil}(x) U_{\rm hp}(x) = 1$ [24], i.e.,

$$(x_1 - \nu)\{x_1^2 - t\tau[1 + 2U_{\rm bulge}(x_1) + U_{\rm inter}(x_1)/t\tau]\} = \mu\nu U_{\rm hploop}(x_1). \quad (5)$$

Once x_1 has been obtained by Eq. (5), the thermodynamic property of ssDNA can then be studied based on the corresponding free energy expression.

Since ν (>1) is a monotonically increasing function of external force f, Eq. (5) indicates that when f is small, DNA monomers stay mainly in the hairpin state, with $x_1 \sim \sqrt{t\tau(1 + 2U_{\text{bulge}} + U_{\text{inter}}/t\tau)}$; while when f becomes very large, the monomers reside mainly in the stretched coil state with $x_1 \sim \nu$. The equation

$$\nu^2(f_c) = t\tau(1 + 2U_{\text{bulge}} + U_{\text{inter}}/t\tau) \qquad (6)$$

determines the characteristic force f_c where transition from hairpin to coil takes place. Nevertheless, the degree of cooperation of this transition is determined by the right-hand side of Eq. (5), $\mu\nu U_{\text{hploop}}$. If $\mu\nu U_{\text{hploop}}$ is always very small in the relevant range of $x \geq \sqrt{t\tau}$, then the transition from hairpin to coil will occur over an extremely narrow force range around f_c and the transition can be regarded as abrupt; in the reverse case, however, this hairpin-coil transition will be gradual and smooth. U_{hploop} is related to hairpin-loop formation in ssDNA. If it is small, formation of a hairpin loop is then quite unfavorable; consequently, each hairpin state has a very long stem region and the ssDNA is composed of only a few hairpins. Therefore, the transition from hairpin to coil state is essentially an all-or-none process, like separating the two strands of a dsDNA apart [12]. But if U_{hploop} is relatively large, then to gain as much entropy as possible, ssDNA will be composed of many hairpins each of which has a short stem region; and when external force is increased from zero, the formed various hairpins disappear one by one and hence a continuous transition process occurs.

Figure 2 demonstrates the numerically calculated values of $\mu\nu U_{\text{hploop}}$ as a function of external force f for

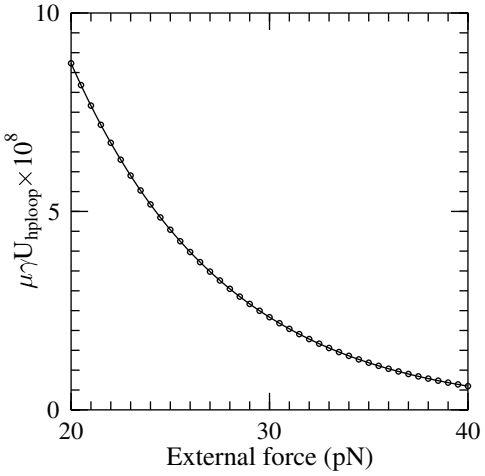

FIG. 2. The value of $\mu\nu U_{\text{hploop}}$ as a function of external force f for the poly(dG-dC) polymer at 25 °C (see text for more details).

poly(dG-dC) [the case of poly(dA-dT) is similar]. We find that it is in the order of 10^{-8}, indeed very small. Therefore the hairpin-coil transition is anticipated to be highly cooperative. We find that U_{bulge} and $U_{\text{inter}}/t\tau$ are also very small (of the order of 10^{-5} and 10^{-8}, respectively), indicating that bulge and interior loop rarely appear in hairpin stem. In the calculation, the experimental data obtained by Tinoco, Jr. *et al.* [15,16] for the statistical weight of a hairpin, a bulge, and an interior loop of $2l$ bases as well as that of the addition of a base pair to a double-stranded region are used without any adjustment. After averaging over two possible stacking manners, $k_BT\ln(t\tau)$ is set to $6.93k_BT$ for poly(dG-dC) and to $3.04k_BT$ for poly(dA-dT), and the values for $\lambda_{\text{bulge}}(2l)$, $\lambda_{\text{inter}}(2l)$, and $\gamma(2l)$ are listed in Table I.

Based on these observations, we can know from Eq. (5) that $x_1 = \sqrt{t\tau}$ for $f < f_c$ and $x_1 = \nu$ for $f \geq f_c$ and, hence, the total free energy $G(f)$ equals $-(N/2)k_BT \times \ln(t\tau)$ for $f < f_c$ and $Nk_BT(g_0 + ag_{\text{coil}})$ for $f \geq f_c$. Consequently, the relative extension z of a ssDNA (compared with its total contour length) is $z = 0$ when $f < f_c$ and then abruptly changes to a nonzero value at the critical force f_c and increases with f according to $z = \mathcal{L}(\eta)$, where η is determined by Eq. (3) as mentioned previously. Thus, the theoretical calculation predicts the same cooperative behavior as that observed by Reif *et al.* (see Figs. 3b and 3c of Ref. [11]).

According to the measurements of Ref. [11], $f_c =$ 20 piconewtons (pN) for poly(dG-dC) and 9 pN for poly(dA-dT). Based on these values of f_c, we estimate the stacking free energy for two adjacent bases to be $g_0 = -(1.61\text{–}1.87)k_BT$ in coiled poly(dG-dC) and $-(0.94\text{–}1.02)k_BT$ in coiled poly(dA-dT). These values are consistent with previous experimental measurements that vertical base stacking interactions are typically in the order of the thermal energy [23] and that the stacking intensity between G and C is stronger than that between A and T [13]. The net free energy $k_BT\ln(t\tau) - 2|g_0|$ required for unzipping a G-C pair is thus estimated to be $(3.18\text{–}3.71)k_BT$, while that for unzipping an A-T pair is $(1.00\text{–}1.16)k_BT$. We notice that such values are close to the phenomenological parameters $E_{G\text{-}C} \simeq 3.0k_BT$ and $E_{A\text{-}T} \simeq 1.3k_BT$ chosen by Bockelmann *et al.* [17] to fit their mechanical measurements.

Such agreements indicate that the present model can quantitatively explain the elastic behavior of designed ssDNA under the action of an external force. At room temperatures, the energetic penalty for the formation of looped structures is so high that complicated multilooped configurations in ssDNA are prohibited and the polymer forms a (single) large hairpin. In biological systems, although many RNA sequences are much more random compared with the sequences studied here, they also have unique native structures. We think such folding processes are governed by the same principle: the polymer chain searches for the configuration which minimizes loop

TABLE I. The statistical weights for the formation of hairpin and bulge loops at 25 °C. Data are based on Ref. [15].

$2l$	2	4	6	8	10–20	20–30	≥30
G-C: $\ln\gamma(2l)$		−8.45	−6.76	−8.45	−10.14	−10.14	$-5.91 - 3.38\ln(2l)$
A-T: $\ln\gamma(2l)$		−11.83	−10.14	−11.83	−13.52	−13.52	$-9.30 - 3.38\ln(2l)$
$\ln\lambda_{\text{bulge}}(2l)$	−6.76	−8.45	−8.45	−10.14	−10.14	$-6.76 - 3.38\ln(2l)$	
$\ln\lambda_{\text{inter}}(2l)$	−3.38	−3.38	−3.38	−5.07	−5.07	$-1.69 - 3.38\ln(2l)$	

formation and maximizes base pairing and base-pair stacking. Provided all the thermodynamic parameters for pairing and stacking between different nucleotides as well as that for looping are known, we may then be able to predict the secondary structure of a given RNA sequence computationally and analyze its stability quantitatively.

Furthermore, the present work indicates a possible way to obtain good estimations for the thermodynamic quantities of nucleotide sequences based on single-macromolecular mechanic experiments. Although we focused on two specific cases, it should be possible to work on other designed chains. There are altogether only two possible base-pairing patterns and ten base-stacking patterns, so the statistical weights for them may all be precisely determined by a combination of measurements and calculations on a handful of carefully designed polynucleotide sequences. Such a method is anticipated to be superior to conventional bulk solution thermal ways.

Note added.—After this paper was submitted, we noticed the work of Montanari and Mézard [25], which predicts a continuous hairpin-coil transition in ssDNA. Such a difference with our work can be accounted for if the energetic contribution of base-pair stacking is included in their model [26].

*Corresponding author.
Present address: Max-Planck-Institut for Colloids and Interfaces, 14424 Potsdam, Germany.

[1] V. S. Pande, A. Yu. Grosberg, and T. Tanaka, Rev. Mod. Phys. **72**, 259 (2000), and references therein.
[2] P. G. Higgs, Phys. Rev. Lett. **76**, 704 (1996).
[3] R. Bundschuh and T. Hwa, Phys. Rev. Lett. **83**, 1479 (1999).
[4] A. Pagnani, G. Parisi, and F. Ricci-Tersenghi, Phys. Rev. Lett. **84**, 2026 (2000).
[5] M. Peyrard and A. R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989).
[6] D. Cule and T. Hwa, Phys. Rev. Lett. **79**, 2375 (1997).
[7] H. Zhou, Y. Zhang, and Z.-C. Ou-Yang, Phys. Rev. Lett. **82**, 4560 (1999); Phys. Rev. E **62**, 1045 (2000).
[8] H. Zhou and Z.-C. Ou-Yang, Mod. Phys. Lett. B **13**, 999 (1999).
[9] S. M. Bhattacharjee, cond-mat/9912297.
[10] C. Guo, H. Levine, and D. A. Kessler, Phys. Rev. Lett. **84**, 3490 (2000).
[11] M. Rief, H. Clausen-Schaumann, and H. E. Gaub, Nat. Struct. Biol. **6**, 346 (1999).
[12] B. Essevaz-Roulet, U. Bockelmann, and F. Heslot, Proc. Natl. Acad. Sci. U.S.A. **94**, 11 935 (1997).
[13] W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).
[14] Under physiological conditions, indeed most naturally occurring RNA or protein molecules have well-defined native structures and structural fluctuations are slight, indicating that energetic considerations dominate over entropic effects.
[15] I. Tinoco, Jr., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla, Nature (London) New Biol. **246**, 40 (1973).
[16] V. A. Bloomfield, D. M. Crothers, and I. Tinoco, Jr., *Physical Chemistry of Nucleic Acids* (Harper & Row, New York, 1974).
[17] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, Phys. Rev. Lett. **79**, 4489 (1997); Phys. Rev. E **58**, 2386 (1998).
[18] P.-G. de Gennes, Biopolymers **6**, 715 (1968).
[19] Each branched structure should at least contain several hairpin loops. As we will show, at the temperature region in which we are interested, hairpin loops appear only rarely; hence branched structures have only higher-order perturbation effects and contribute very slightly to the configurational entropy and free energy.
[20] We have checked that the conclusions of this work were independent of the choice for the value of b.
[21] S. B. Smith, Y. Cui, and C. Bustamante, Science **271**, 795 (1996).
[22] J. F. Marko and E. D. Siggia, Macromolecules **28**, 8759 (1995).
[23] P. O. Ts'o, in *Basic Principles in Nucleic Acid Chemistry,* edited by P. O. Ts'o (Academic Press, New York, 1974), Vol. I, pp. 453–584.
[24] S. Lifson, J. Chem. Phys. **40**, 3705 (1964).
[25] A. Montanari and M. Mézard, cond-mat/0006166.
[26] H. Zhou and Y. Zhang (to be published).