VOLUME 91, NUMBER 14 PHYSICAL REVIEW LETTERS week ending 3 OCTOBER 2003

Length and Statistical Weight of Bubbles in DNA Melting

Yan Zeng, Awrasa Montrichok, and Giovanni Zocchi
Department of Physics and Astronomy, University of California Los Angeles, Los Angeles, California 90095-1547, USA

(Received 26 February 2003; published 30 September 2003)

We present measurements of the length of the bubble versus temperature in the melting transition of DNA oligomers. The sequences are clamped at the ends of GC pairs (strong binding), and have AT rich (weaker binding) middle regions of variable length B relative to the molecule size. We use a quenching technique to trap intermediate states. We find that the average relative bubble size $\langle \ell \rangle$ grows from zero for increasing temperature, but reaches a plateau at a length of order B, producing a discontinuous transition with finite bubble size. The statistical weight of the bubble states decreases as B is reduced, the transition becoming more cooperative.

DOI: 10.1103/PhysRevLett.91.148101 PACS numbers: 87.15.–v

The temperature driven melting of the DNA double helix into separate strands is a relatively simple conformational transition, yet a quantitative understanding, in particular, of the properties of intermediate states, remains challenging. The basic physics of DNA melting is the competition between the energy of the bound states and the entropy of the dissociated states. A typical partially melted conformation will consist of single-stranded regions (“bubbles”) bounded by double-stranded (ds) segments [1–5]. Complications arise because, for instance, a bubble introduces a flexible joint in the stiff duplex (ds) structure, allowing neighboring ds segments to interact with the bubble and with each other. These “excluded volume” interactions may in fact control the order of the transition for large systems [4,6–8].

Sequence specificity is an additional interesting question. In the DNA double helix, the two strands are held together by hydrogen bonds between complementary bases: two bonds for the AT pair and three bonds for the GC pair, which is therefore stronger. Locally, the sequence affects the stability and the structure of the double helix [9]. Globally, sequence heterogeneity may affect the order of the transition in the thermodynamic limit [10].

Experimentally, the transition has been characterized by spectroscopic methods such as UV absorption, circular dichroism spectroscopy, and fluorescence [11–14], by calorimetry [15,16] and electrophoretic mobility assays [17,18]. There is a large accumulated knowledge of the physical chemistry aspects, such as base pairing and stacking enthalpies [11,15,16]. In the case of oligomers, the above methods also revealed sequence-dependent structural transitions below the melting temperature (“premelting”) [14,16]; for several thousand base pairs long DNA, it was found that the derivative of the melting curve exhibits a series of peaks, corresponding to different regions of the molecule melting at different temperatures [19,20]. However, melting curves for oligomers always appear smooth because of finite size effects. Sequence-dependent steps and plateaus in the mechanically induced unzipping of single, kb long DNA molecules have been observed experimentally [21,22] and addressed theoretically [23].

Theoretical investigations of the transition are based on Ising-type models to describe paired and unpaired bases [1–5] or models which introduce a potential energy function of the distance between the bases [6,10,24,25]. Depending on how the entropy of single-stranded loops is estimated, these models predict a continuous [3–5] or discontinuous [6,8,10] transition in the thermodynamic limit.

Our aim here is to characterize the nature and statistical significance of intermediate (partially melted) states, by focusing on the behavior of a single bubble. There are two kinds of bubbles: bubbles opening in the middle, bounded on both sides by ds regions, and bubbles opening at the ends of the molecule. We focus on the former, by choosing sequences clamped at the ends with GC pairs, and with an AT rich middle region of variable relative length B ($B = L_{\mathrm{AT}}/L$, where L_{AT} is the length of the AT rich region and L the length of the molecule).

All the ensemble measurements mentioned above suffer from the following limitation. At the midpoint of the transition, as judged, e.g., by UV absorption, one cannot tell whether half of the molecules are halfway open. One cannot pinpoint intermediate states. We developed a new approach, based on quenched states, to measure the average size of the denaturation bubble and quantify the presence of intermediate states in the melting of DNA oligomers. Our results show that (i) for increasing temperature the average relative bubble length $\langle \ell \rangle$ grows from zero but reaches a plateau at the relative size of the AT rich region B; (ii) opening of the bubble is not continuous, at least for the smaller B; this suggests a nucleation threshold or minimum length for a stable bubble (the average bubble length $\langle \ell \rangle$ is however continuous), and (iii) the statistical weight of the bubble states decreases with decreasing B, the transition becoming more cooperative.

The quenching technique has been described previously [26]. Briefly, we use sequences which are partially self-complementary, so that isolated single strands can form hairpins. The initial state is prepared in the duplex form (which is the ground state) by hybridizing with the exact reverse complement. Aliquots are heated to different temperatures T_i within the transition range, then quenched to ~0 °C (Fig. 1). Molecules which were completely open at temperature T_i form hairpins after quenching, while molecules which were partially open close again as duplexes. The relative amount of hairpins, representing the relative amount of completely open molecules at temperature T_i, is determined by gel electrophoresis. This measurement yields the fraction of completely open molecules versus temperature, which we call p. From UV absorption spectroscopy we determine the fraction of open base pairs versus T, which we call f. Combining these two measurements we can pinpoint the presence of intermediate states. For a two-states transition, $p = f$. If there are intermediate states, $p < f$. Since the fraction of open base pairs can be written

$$f(T) = [1 - p(T)]\langle \ell \rangle + p(T), \tag{1}$$

where $\langle \ell \rangle$ is the average fractional length of the bubble (averaged over the subset of the partially open molecules), the average fractional bubble length is

$$\langle \ell \rangle = (f - p)/(1 - p). \tag{2}$$

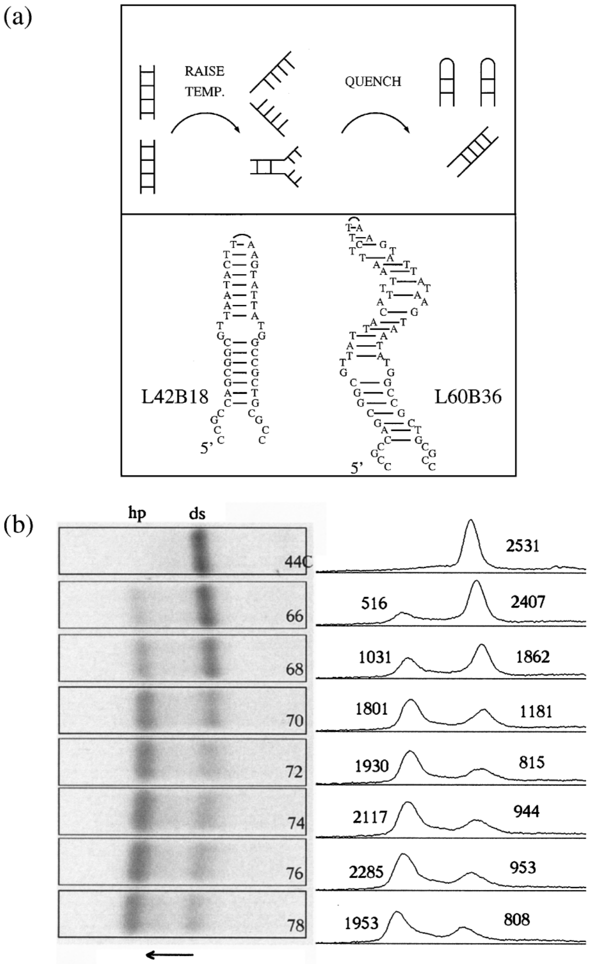

FIG. 1. (a) Sketch of the quenching method used to tap intermediate states. The lower part of the figure shows the two sequences used in this study, in the hairpin conformation. (b) Example of a gel, running right to left. The slow (fast) band corresponds to duplexes (hairpins). The temperatures to which the aliquots were heated before quenching are indicated on the lanes. The plot on the right shows the intensity profiles; the numbers are proportional to the areas under the peaks and are used to calculate the fraction of open molecules p.

The two sequences, L60B36 (length = 60 bases, bubble forming region = 36 bases) and L42B18 have identical GC rich regions at the ends, and an AT rich middle portion of length 36 (18) bases [Fig. 1(a)]. We designed some mismatches into the hairpins in order to be able to prepare an initial state predominantly in the duplex form and also to reduce the importance of alternative structures (crosses) in the duplex; however there are no mismatches built into the duplexes. The synthetic DNA oligonucleotides were purchased from Operon Technologies. For all experiments DNA concentration was 1 μM, in phosphate buffer saline at an ionic strength of 50 nM. Total volumes were 800 μL in standard quartz cuvettes for the UV measurements, and 30 μL in polymerase chain reaction tubes for the quenching measurements.

Figure 1(b) shows one of the gels (3% agarose, stained with ethidium bromide). Gel electrophoresis sorts DNA fragments by length, the longer fragments having slower drift velocities through the gel under an applied electric field. In Fig. 1(b) the drift direction is from right to left; the "slow" band, marked "ds," corresponds to duplexes; the "fast" band, marked "hp," to hairpins. The slight curvature in the position of the bands is a common occurrence caused by inhomogeneities in the gel such as lateral temperature gradients. The graph on the right shows the intensity integrated across the lane. The numbers represent the areas under the peaks (in arbitrary units) and are proportional to the total intensity of the band. These numbers were used to obtain the fraction of open molecules p at each temperature. The straightforward normalization for p is $p(T) = hp(T)/[hp(T) + ds(T)]$ where $hp(T)$ is the intensity of the hairpin band of the aliquot which was brought to temperature T, and $ds(T)$ the intensity of the duplex band. However there is some $hp + hp \rightarrow ds$ recombination after the quench; this is visible in Fig. 1(b), where the ds band does not completely disappear at high T. The importance of this effect varies with sequence, but it can be taken into account through a correction to the normalization of p. In the simplest model, the recombination rate is proportional to the square of the hairpin concentration; one then finds that the normalization of p should include a correction factor:

$$p = \frac{hp_q}{hp_q + ds_q} = \frac{hp_{\text{meas}}}{hp_{\text{meas}} + ds_{\text{meas}}} \frac{1}{1 - \gamma hp_{\text{meas}}}, \quad (3)$$

where the subscript q means "quenched" (i.e., the value of the quantity right after the quench), and meas means "measured" (a time t after the quench, the same t for all aliquots in one run). The factor γ is found from the band intensities, by enforcing $hp_q/(hp_q + ds_q) = 1$ at high T [e.g., 74, 76, 78 °C in Fig. 1(b)].

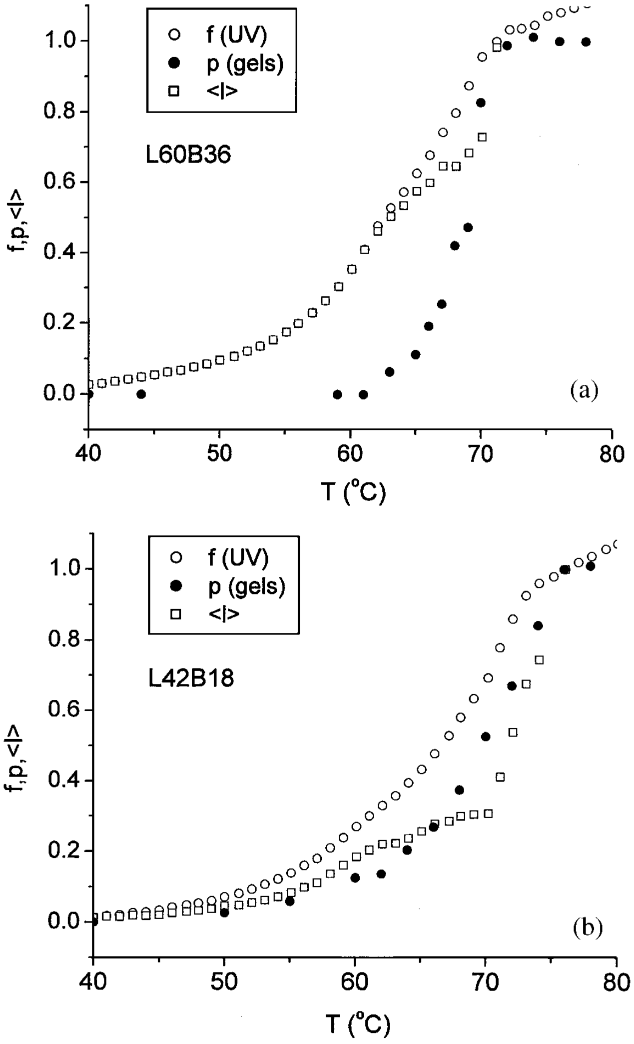

FIG. 2. Melting curves for the two sequences; the open circles represent the fraction of open base pairs f (from the UV absorption measurements); the filled circles represent the fraction of open molecules p (from the gels), and the squares represent the average relative length of the bubble $\langle \ell \rangle$, calculated from Eq. (2). (a) $\langle \ell \rangle$ grows smoothly from zero and reaches a plateau for $\langle \ell \rangle \approx 0.6 = B$, the relative size of the AT rich middle region for this sequence. (b) With a shorter AT rich middle region ($B = 0.42$) the average bubble length reaches a plateau at a correspondingly smaller value $\langle \ell \rangle \approx 0.3$, and the transition is not continuous ($\langle \ell \rangle \neq f$ even before the plateau).

Figure 2 shows the melting curves and the average fractional length of the bubble $\langle \ell \rangle$, for the two sequences. The UV absorption data (open circles), which represent the fraction of open bases f, were normalized so that $f = 1$ coincides with $p = 1$ (the end point of the transition). The further increase in absorption at higher T ($f > 1$) is well known and corresponds to base unstacking in the single strands. Our normalization of f ($f = 1$ when $p = 1$) is consistent with the usual approach of setting $f = 1$ at the shoulder of the S-shaped melting curve. $\langle \ell \rangle$ (open squares) was calculated from Eq. (2), using an interpolation of the p points to obtain a better definition. In Fig. 3 we show the quantity $\sigma = f - p$, which represents the fraction of bases in intermediate states.

Both sequences show plateaus in $\langle \ell \rangle$ at values corresponding to the relative length of the AT rich region, B. The significance is that as the temperature is increased within the plateau region, the increase of f comes entirely from an increase in the fraction of completely dissociated molecules p, as can be seen by setting $\langle \ell \rangle = \text{const}$ in (2). Thus in the plateau region bubble states turn to dissociated states, through the abrupt melting of the GC rich end regions. These observations suggest a finite bubble size at the transition (for a continuous transition with diverging bubble size we expect $\langle \ell \rangle = f$ and no plateaus).

For L60B36 [Fig. 2(a)] we find that the average bubble length $\langle \ell \rangle$ reaches a plateau for $\langle \ell \rangle \approx 0.6$, corresponding to the relative length of the AT rich region $B = 36/60 = 0.6$ for this sequence. The fraction of bases in intermediate states, σ (Fig. 3), peaks at the beginning of the plateau region, with a max value $\sigma_{\max} \approx 0.5$. Since these states

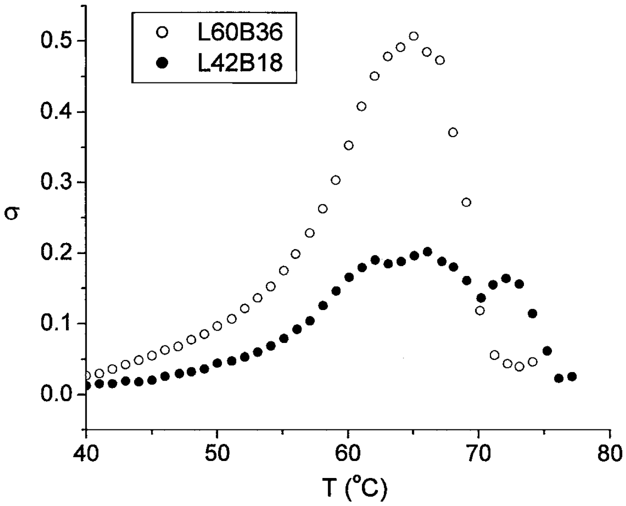

FIG. 3. The quantity $\sigma = f - p$, representing the fraction of bases in a bubble state. The statistical weight of intermediate states is smaller for the shorter sequence.

correspond to a bubble of fractional length B, the fraction of molecules in intermediate states at this point, i.e., the statistical weight of these intermediates, can be estimated as $\sigma_{\max}/B = 0.5/0.6 = 0.83$. Thus more than 80% of the molecules are in a bubble state at this point.

For L42B18 [Fig. 2(b)] we find that the bubble length reaches a plateau value $\langle \ell \rangle \approx 0.3$, somewhat shorter than $B = 18/42 = 0.43$ for this sequence. However there is a significant difference with respect to L60B36, namely, now $\langle \ell \rangle \neq f$ from the beginning (at low T); this shows that the opening of the bubble is not continuous. The peak value of σ (Fig. 3) is $\sigma_{\max} \approx 0.2$, much smaller than in the case of L60B36. The fraction of molecules in intermediate states at the peak is also smaller: $\sigma_{\max}/0.3 = 0.67$.

The extension of the plateau is smaller for L60B36 ($B = 0.6$) than for L42B18 ($B = 0.42$); indeed, we expect the plateau's extension to shrink with growing B, since for $B = 1$ there is, presumably, no plateau. Thus the $\langle \ell \rangle$ vs T curves for different values of B are reminiscent of the isotherms in the P-V plane for a liquid-gas transition (where also the phase coexistence plateau shrinks as one approaches the critical temperature).

In summary, for both sequences we see plateaus in the average bubble size $\langle \ell \rangle$, which indicates a discontinuous transition with bubble size less than molecule size. The finite size of the molecules allows us in this case to characterize the transition, through the quantities p and $\langle \ell \rangle$. L42B18 shows that there is a minimum bubble size, or nucleation threshold, because as soon as the bubble starts to open, there are also some completely dissociated molecules ($\langle \ell \rangle \neq f$). This is not just a finite size effect: in a forthcoming study we show that shorter sequences (down to 19 bp), which unzip from one end, can open continuously ($\langle \ell \rangle = f$), with large populations of intermediate states. Thus the present data suggest a discontinuous transition with coexistence of bubbles and paired tracts; this scenario is supported by earlier studies on long sequences [19,20].

The novelty of the experiment lies in measuring the order parameter $\langle \ell \rangle$ in the case of oligomers; indeed, none of these results would be apparent from the spectroscopy curves alone, which in all cases have the same smooth shape; thus the method yields new data for comparing models to experiments. Our future work will include studying the role of end effects and sequence design in shaping the nature of the melting transition for oligomers. Presumably the knowledge thus gained on the finite size system will also impact the understanding of the infinite system.

This research was partly supported by the U.S.-Israel Binational Science Foundation under Grant No. 2000298.

[1] T. L. Hill, J. Chem. Phys. **30**, 383 (1959).
[2] B. H. Zimm, J. Chem. Phys. **33**, 1349 (1960).
[3] D. Poland and H. A. Scheraga, J. Chem. Phys. **45**, 1456 (1966); **45**, 1464 (1966).
[4] M. E. Fisher, J. Chem. Phys. **45**, 1469 (1966).
[5] M. Ya. Azbel, Phys. Rev. A **20**, 1671 (1979).
[6] M. S. Causo, B. Coluzzi, and P. Grassberger, Phys. Rev. E **62**, 3958 (2000).
[7] Y. Kafri, D. Mukamel, and L. Peliti, Phys. Rev. Lett. **85**, 4988 (2000).
[8] E. Carlon, E. Orlandini, and A. L. Stella, Phys. Rev. Lett. **88**, 198101 (2002).
[9] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry* (Freeman, New York, 1980).
[10] D. Cule and T. Hwa, Phys. Rev. Lett. **79**, 2375 (1997).
[11] G. Vesnaver and K. J. Breslauer, Proc. Natl. Acad. Sci. U.S.A. **88**, 3569 (1991).
[12] C. A. Gelfand *et al.*, Proc. Natl. Acad. Sci. U.S.A. **96**, 6113 (1999).
[13] G. Bonnet *et al.*, Proc. Natl. Acad. Sci. U.S.A. **96**, 6171 (1999).
[14] S. S. Chan *et al.*, Biochemistry **29**, 6161 (1990).
[15] M. M. Senior, R. A. Jones, and K. J. Breslauer, Proc. Natl. Acad. Sci. U.S.A. **85**, 6242 (1988).
[16] S. S. Chan *et al.*, Biochemistry **32**, 11 776 (1993).
[17] R. M. Myers *et al.*, Nucleic Acids Res. **13**, 3111 (1985).
[18] J. Zhu and R. M. Wartell, Biochemistry **36**, 15 326 (1997).
[19] W. Beers, A. Cerami, and E. Reich, Proc. Natl. Acad. Sci. U.S.A. **58**, 1624 (1967).
[20] R. D. Blake and S. G. Delcourt, Nucleic Acids Res. **26**, 3323 (1998).
[21] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, Phys. Rev. Lett. **79**, 4489 (1997).
[22] C. Danilowicz *et al.*, Proc. Natl. Acad. Sci. U.S.A. **100**, 1694 (2003).
[23] D. K. Lubensky and D. R. Nelson, Phys. Rev. E **65**, 031917 (2002).
[24] T. Dauxois and M. Peyrard, Phys. Rev. E **51**, 4027 (1995).
[25] N. Theodorakopoulos, T. Dauxois, and M. Peyrard, Phys. Rev. Lett. **85**, 6 (2000).
[26] A. Montrichok, G. Gruner, and G. Zocchi, Europhys. Lett. **62**, 452 (2003).