

Intermediate phase in DNA melting

Richard A. Neher* and Ulrich Gerland

Arnold-Sommerfeld-Center for Theoretical Physics (ASC) and Center for NanoScience (CeNS), LMU München, Theresienstrasse 37, 80333 München, Germany

(Received 14 September 2005; published 23 March 2006)

We predict a temperature-driven phase transition of DNA below the melting transition. The additional, intermediate phase exists for repetitive sequences, when the two strands have different lengths. In this phase, the excess bases of the longer strand are completely absorbed as bulge loops inside the helical region. When the temperature is lowered, the excess bases desorb into overhanging ends, resulting in a contour length change. This continuous transition is in many aspects analogous to Bose–Einstein condensation. When the sequence is weakly disordered, the contour length changes discontinuously with temperature.

DOI: [10.1103/PhysRevE.73.030902](https://doi.org/10.1103/PhysRevE.73.030902)

PACS number(s): 87.14.Gg, 05.70.Fh, 87.15.Aa, 87.15.He

The base-pairing interaction between the two strands of DNA is not only pivotal to its biological function [1], but also leads to intriguing applications in nanotechnology [2]. One approach to probe this interaction is to monitor the DNA conformation as a function of temperature. Experimentally, one can observe the number of base pairs formed (using UV absorption [3,4]), as well as changes of intramolecular distances on the nanometer scale (using modern single-molecule techniques [5]). On the theoretical side, the temperature dependence of DNA conformations has been studied for almost 50 years, using models of various degrees of complexity [6–14]. Particular attention has been paid to the characteristics of the melting transition, where the two strands separate completely. Whereas early models yielded only a crossover [6], the Poland-Scheraga (PS) model [8] was the first to display a phase transition, albeit a continuous one, which appeared to be at variance with the experimentally observed sharp jump in the fraction of bound base pairs [3]. Only recently have mechanisms been proposed [11,12] which yield an abrupt, first-order transition. So far, however, most analyses of DNA melting have incorporated only native interactions, i.e., base pairs that occur in the ground state of the molecule (see [7,13,14] for notable exceptions). It is our aim here to show that such non-native interactions can introduce an intermediate phase in the melting behavior of DNA, associated with an additional conformational transition before strand separation.

Non-native interactions are particularly relevant for repetitive DNA sequences, which are common in genomes [15]. Periodic DNA, with, e.g., a single base repeat such as TTT..., or a higher-order repeat such as CAGCAG..., can take on base-pairing patterns with asymmetric loops and the two complementary strands can be shifted relative to each other. Here, we consider the general situation where the two strands can have arbitrary lengths N , M . We find that for $N \neq M$, the bound phase splits into two separate phases. The low-temperature phase is characterized by an extensive length of the unbound end on the longer strand, whereas in the new intermediate phase these overhanging bases are ab-

sorbed into the helical region. Mathematically, and also conceptually, many aspects of this transition are analogous to Bose–Einstein condensation (BEC), as “particles” (bases) condense into a single “state” (the overhanging end), which thereby acquires macroscopic “occupation” (length). Obviously, the analogy extends only to the behavior of the partition function, as there is no quantum coherence in the DNA problem. Effectively, the transition amounts to a temperature-sensitive change in the contour length of the DNA molecule, which should be observable with optical or single-molecule methods. While the transition is continuous for perfectly periodic sequences, we find that the contour length shortens discontinuously once (weak) sequence disorder is introduced.

DNA model. We consider two DNA strands with lengths N and $M \geq N$, respectively, and describe their interaction with the “imperfect matching” generalization of the PS model [8,13,14,16]. Specifically, a base $i \leq N$ of the lower strand can form a base pair (i, j) with every complementary base $j \leq M$ of the upper strand, whereas the formation of base pairs within a strand can be neglected (since we are interested only in sequences with a high degree of complementarity and a low degree of self-complementarity). Due to geometrical constraints, we may neglect the “crossing” of base pairs, e.g., two base pairs (i_1, j_1) and (i_2, j_2) with $i_1 < i_2$ but $j_1 > j_2$. The base-pairing pattern \mathcal{S} , i.e., the set of all formed base pairs, then creates a DNA conformation consisting of bound segments alternating with (possibly asymmetric) loops, see Fig. 1. To simplify the discussion, we enforce the base pair (N, M) at the right end, so that we need to consider

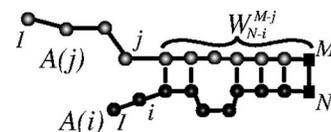


FIG. 1. A possible configuration of two complementary DNA strands with a repetitive sequence (a bead represents one repeat unit). Note that repetitive sequences can form base-pairing patterns with asymmetric loops. In general, we allow for different strand lengths N , M . The last repeat units (squares) are permanently bound.

*Electronic address: richard.neher@physik.lmu.de

only one overhanging end. However, our main findings remain equally valid without this constraint. Experimentally, this constraint could be realized, e.g., by a few particularly strong base pairs at one end.

Each base-pairing pattern \mathcal{S} , receives a statistical weight $Q(\mathcal{S})$, which takes the form of a product with factors of four different types: (i) a Boltzmann factor $q=e^{\varepsilon_b/k_B T}$ for every base pair with binding energy $-\varepsilon_b < 0$, (ii) a Boltzmann factor $g^2=e^{-\varepsilon_\ell/k_B T}$ for every loop with loop initiation cost $\varepsilon_\ell > 0$, (iii) an entropic factor $B_\ell(m)=s^m m^{-c}$ for each loop, which is the increase in the number of polymer configurations when m bases form a (floppy) loop instead of being in a (rigid) double helical conformation, and (iv) a similar entropic factor $A(n)=s^n n^{-\bar{c}}$ for a single-stranded end of n bases. Here, the exponents c, \bar{c} in the entropic factors are universal in that they are independent of the detailed polymer properties, but are sensitive to excluded volume interactions. For interacting self-avoiding loops one has $c \approx 2.15$, while $\bar{c} \approx 0.1$ [17]. Whereas the value of c determines the critical behavior at the melting transition, the nonuniversal constant s has no qualitative effect on the melting behavior (we use $s=10$ in all numerical examples).

In the following, we first apply the DNA model to perfectly periodic sequences, where each repeat unit can be treated as an effective base with renormalized parameters (we use $\varepsilon_b=6$ and $\varepsilon_\ell=3$ in temperature units, $k_B=1$). We emphasize that our simplistic model for the involved energies and entropies is meant to illustrate the physical phenomena in a transparent way, but leads to an unrealistic temperature scale. With a more detailed description [4], we find that all of the interesting behavior happens at accessible temperatures [18].

Free energy of periodic DNA. To obtain the equilibrium properties of the DNA model, we calculate the partition sum over all base-pairing patterns, $Z_N^M = \sum_{\mathcal{S}} Q(\mathcal{S})$. By separating the single-stranded ends from the double-stranded part, see Fig. 1, we write Z_N^M as

$$Z_N^M = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} A(i)A(j)W_{N-i}^{M-j}. \quad (1)$$

Here, W_r^t is the partition function of two complementary and periodic strands of length r and t with the first and last base pair formed. W_r^t obeys this recursion relation

$$W_{r+1}^t = qW_r^t + g^2q \sum_{\substack{k < r, m < t \\ k+m > 0}} B_\ell(k+m)W_{r-k}^{t-m}, \quad (2)$$

with the initial conditions $W_1^1 = q$ and $W_1^i = W_i^1 = 0$ for $i > 1$. For sequences with weak disorder considered further below, q has to be replaced by q_{r+1}^t , where $q_r^t = e^{\varepsilon_b/k_B T}$ for original bases at r and s , $q_r^s = e^{\varepsilon_b/k_B T}$ for mutated bases at $r=s$ and $q_r^t = 0$ otherwise. We use the recursion relation to calculate Z_N^M and expectation values for finite lengths N, M [13, 19]. To extract the thermodynamic behavior in the limit of long strands, we take the z transform $\hat{Z}(x, y) = \sum_{N, M=0}^{\infty} Z_N^M x^N y^M$ and solve for the transformed partition sum. This procedure is equivalent to the method of sequence generating function employed in Ref. [14]. One obtains

$$\hat{Z}(x, y) = \frac{\hat{A}(x)\hat{A}(y)qxy}{1 - qxy + \frac{qg^2xy}{x-y}[y\hat{B}(y) - x\hat{B}(x)]}, \quad (3)$$

where the transforms of the entropic factors are given by $\hat{A}(z) = \phi_{\bar{c}}(sz) + 1$ and $\hat{B}(z) = \phi_c(sz)$, with the polylogarithm $\phi_c(z) = \sum_{n=1}^{\infty} z^n n^{-c}$.

The Z -transform carried out above amounts to a change from the canonical to the grand canonical ensemble. The transformation variables x, y play the role of fugacities for bases in the lower and upper strands, respectively. However, for the ensuing discussion, it is advantageous to keep the length N of the shorter strand fixed as a reference. Hence, we perform the inverse transformation for the lower strand by contour integration in x , to obtain the partition sum $Z_N(y_0)$ for N bases on the lower strand and the upper strand coupled to a ‘‘nucleotide reservoir’’ with fixed fugacity y_0 . For large N and temperatures below the melting temperature, $Z_N(y_0)$ is given by $\hat{A}(y_0)x^*(y_0)^{-N}$, where $x^*(y_0)$ is the smallest real zero of the denominator of Eq. (3) for given $y=y_0$ [14]. Hence, the free energy of the bound phase is given by $Nf_b(y_0) - T \ln \hat{A}(y_0)$, where the first term is the contribution of the helical region with a free energy per length $f_b(y_0) = T \ln x^*(y_0)$, and the second term is the contribution from the unbound end of the longer strand. The free energy for given N and M is then obtained by saddle point integration,

$$\frac{F(T, N, M)}{T} = -\ln \hat{A}(y_0) + N \frac{f_b(y_0)}{T} + M \ln(y_0), \quad (4)$$

where the fugacity y_0 is determined by

$$M = \langle M \rangle_{y_0} = y_0 \frac{\partial \ln \hat{A}(y_0)}{\partial y_0} - N \frac{y_0}{T} \frac{\partial f_b(y_0)}{\partial y_0}. \quad (5)$$

Phase diagram. To extract the physical behavior of the DNA model from Eqs. (4) and (5), we focus on two observables, the total number of base pairs, $N\theta$, and the length of the single-stranded overhang. The fraction θ of bound base pairs is calculated from the free energy per length of the helical region as

$$\theta = -\frac{q}{T} \frac{\partial f_b(y_0)}{\partial q}. \quad (6)$$

To obtain the overhang length, we note that the right-hand side of (5) decomposes the total length M of the upper strand into two contributions, where the first term is the expected overhang length and the second term corresponds to the number of bases in the helical region. The dashed line in Fig. 2 (top) shows the overhang length as a function of temperature, for $N=1000$ and $M=1150$. At low temperatures, the two DNA strands are completely aligned, so that all $M-N$ excess bases of the longer strand form an overhanging end. However, we observe that the overhang length decreases with increasing temperature, dropping almost to zero before it rises again sharply at even higher temperature. We see in Fig. 2 (bottom) that this drop occurs in a temperature range where almost all possible base pairs are formed, and the rise

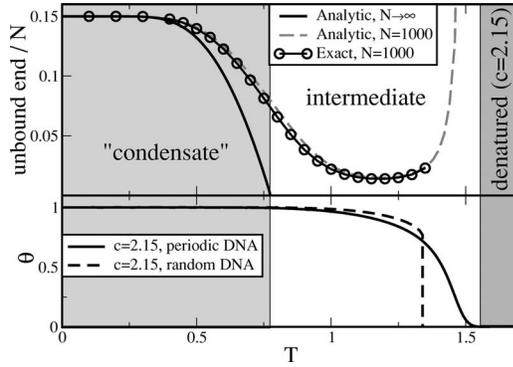


FIG. 2. Top: The length of the unbound end, normalized by the number of excess bases $N-M$ on the longer strand. For finite systems ($N=1000$, dashed line), the unbound end shrinks to a minimal value and increases again, as the melting temperature is approached. In the $N \rightarrow \infty$ limit, the overhang length diverges below $T_c=0.7752$ and is of order 1 for $T > T_c$. Expectation values calculated numerically using Eq. (2) agree well with the analytic result. Bottom: The fraction of bound base pairs θ as a function of temperature. For periodic sequences with $c=2.15$, θ vanishes with zero slope at $T_m=1.424$, whereas a random sequence shows a first-order phase transition.

occurs when the two strands separate. These observations suggest that a temperature-driven conformational transition occurs before the melting transition.

This transition is in fact completely analogous to BEC, as Eq. (5) parallels the behavior of the equation of state for an ideal Bose gas: If we divide Eq. (5) by our system size N and introduce the “particle density” $\alpha=M/N$, we obtain

$$\alpha = \frac{1}{N} \frac{\phi_{c-1}(sy_0)}{\phi_c(sy_0) + 1} + \bar{\alpha}(y_0), \quad (7)$$

where $\bar{\alpha}(y_0) = -(y_0/T)[\partial f_b(y_0)/\partial y_0] \geq 1$ is the density inside the helical region. In Eq. (7), the first term on the right-hand side corresponds to the occupation of the ground state of an ideal Bose gas, whereas $\bar{\alpha}(y_0)$ is analogous to the occupation of the excited states. The fugacities of a Bose gas and our DNA are bounded: for the former, by the energy of the ground state, and for the DNA by the weight of an unbound monomer, i.e., $y_0 \leq s^{-1}$. The population of the excited states increases monotonically with the fugacity, and attains a finite maximal value, in our case $\bar{\alpha}_{max} = \bar{\alpha}(s^{-1})$, provided the loop exponent $c > 2$ [22]. When the temperature is lowered, $\bar{\alpha}_{max}$ decreases (see bottom panel of Fig. 3), and when it falls below α , the length of the unbound end must become extensive to accommodate the remaining bases. In an analogous way, the ground state of a Bose gas is macroscopically populated at low temperatures. In this “condensate” phase, the fugacity is locked to the value s^{-1} in the thermodynamic limit ($N, M \rightarrow \infty$, $\alpha = \text{const.}$). The deviation for finite systems scales as $s^{-1} - y_0 \sim 1/N$, see Fig. 3 (top). In the opposite case, where $\alpha < \bar{\alpha}_{max}$, there is a solution to Eq. (7) with $y_0 < s^{-1}$ and the unbound end remains finite for all system sizes.

It is easily shown that $\bar{\alpha}_{max}$ approaches 1 at low temperatures, and consequently all excess bases of the longer strand are condensed in the overhang, as illustrated in Fig. 3 (bot-

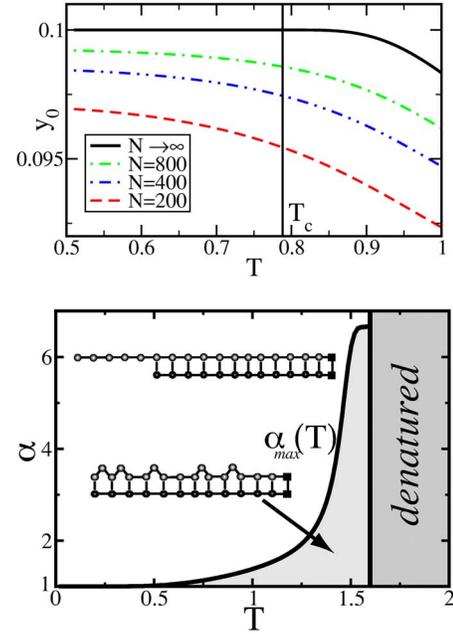


FIG. 3. (Color online) Top: The fugacity y_0 vs T for different system sizes N . In the thermodynamic limit, $y_0 = s^{-1}$ for $T < T_c$. As for BEC, y_0 approaches its limiting value as $s^{-1} - y_0 \sim 1/N$. Bottom: Phase diagram of periodic DNA. At low temperatures, both strands are completely aligned and excess bases of the longer strand form an unbound end. In the intermediate phase, all excess bases are absorbed into the helical region.

tom). As T increases, more and more bases are absorbed in the helical region ($\bar{\alpha}_{max}$ increases), and the system enters the intermediate phase at $T=T_c$, where $\bar{\alpha}_{max} = \alpha$. At T_c the condensate fraction vanishes, as the solid line shows in Fig. 2 (top). Note that the intermediate phase exists only when α is not too large.

The melting temperature T_m , where the strands separate and θ vanishes (denatured phase), is independent of α . For periodic sequences, the loop size distribution at $T=T_m$ decays as $\sim n^{-(c-1)}$ instead of $\sim n^{-c}$, since n bases of a loop can be distributed in $n+1$ ways among both strands [8]. Hence, periodic DNA displays a continuous melting transition for $2 < c \leq 3$ and a first-order transition only if $c > 3$. For $2 < c \leq 3$, we obtain $\theta \sim |T - T_m|^{(3-c)/(c-2)}$, using the same method as [20] for the standard PS model. To illustrate this, we plot θ for periodic sequences and for the standard PS model in Fig. 2 (bottom). Whereas for the latter θ drops discontinuously to zero, θ of periodic DNA vanishes with zero slope.

Weak sequence disorder. Is the intermediate phase identified above robust against sequence disorder? To address this question, we replace a small fraction of base pairs by bases that can pair with each other ($\bar{\epsilon}_b=2$), but not with other bases in the sequence. Figure 4 shows the average overhang length calculated using the generalized Eq. (2) for sequences with evenly spaced mutations every $d=25, 50$, and 100 bases. The unbound end keeps its ground-state length up to a certain temperature, and then shortens rapidly. The temperature at which the drop occurs increases with the density of mutations. The width of the transition region scales inversely with the system size N , see Fig. 4(b). These numerical observa-

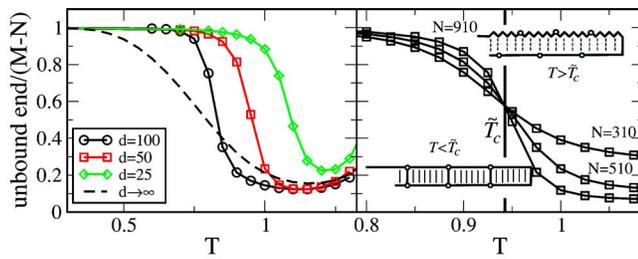


FIG. 4. (Color online) Sequence disorder renders the transition to the intermediate phase discontinuous. Both panels show the average length of the unbound end, normalized by $M-N$. The left panel shows data for different mutation densities (no mutations, mutations every $d=100$, 50, and 25 bases) for $N=510$ and $\alpha=1.15$. The right panel shows data with $d=50$ for different system sizes. The transition sharpens with increasing N and the slope of the curves in the transition region is compatible with $\sim N$.

tions suggest that in the limit of infinite system size, the transition becomes an abrupt first order transition.

Indeed, the first-order behavior can also be understood theoretically by comparing the energy barriers for forming bulge loops with and without mutations: The formation of a bulge loop on the longer strand of a perfectly periodic molecule requires only the initiation energy ε_ℓ . In the presence of mutations, however, shifting both strands breaks mutated base pairs. Hence, to form a bulge loop, all mutations to the left of the loop have to be broken and the energy barrier becomes extensive. For a sufficiently low density of muta-

tions, there is a temperature \tilde{T}_c at which the entropy gained by distributing excess bases in loops along the molecule outweighs the energetic costs to break all mutations [21]. Below \tilde{T}_c all mutations are bound, if $T > \tilde{T}_c$ as many mutations open, as are necessary to absorb all excess bases.

Discussion and outlook. We have identified a BEC-like conformational transition in periodic DNA, which occurs below the melting transition. The hallmark of this transition is the shortening of the unbound end, which could be directly observed experimentally by resonant energy transfer between fluorescent dyes located at the ends of the two strands. The transition is also associated with a change in the contour length of the DNA molecule, roughly proportional to $M-N$. The increased density of bulge loops in the helical region may additionally yield an effect on the persistence length. We found that the existence of the intermediate phase is robust against weak sequence disorder and expect that it is also independent of the details of our model. The temperature range of the intermediate phase rapidly narrows as the mutation density is increased. We conjecture that the intermediate phase disappears completely at a finite mutation density. The exploration of the critical behavior in the complete temperature—mutation density phase diagram is left as an interesting theoretical challenge for the future.

We are grateful for important comments by E. Frey and H. Wagner. We acknowledge financial support by the Deutsche Forschungsgemeinschaft through the Emmy Noether Program.

- [1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell* (Garland Science, New York, 2002).
- [2] N. C. Seeman, *Nature (London)* **421**, 427 (2003).
- [3] R. M. Wartell and A. S. Benight, *Phys. Rep.* **126**, 67 (1985).
- [4] J. SantaLucia, Jr., *Annu. Rev. Biophys. Biomol. Struct.* **33**, 415 (2004).
- [5] X. Zhuang and M. Rief, *Curr. Opin. Struct. Biol.* **13**, 88 (2003).
- [6] B. H. Zimm and J. K. Bragg, *J. Chem. Phys.* **31**, 526 (1959).
- [7] T. L. Hill, *J. Chem. Phys.* **30**, 383 (1959).
- [8] D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers* (Academic, New York, 1970).
- [9] M. Peyrard and A. R. Bishop, *Phys. Rev. Lett.* **62**, 2755 (1989).
- [10] D. Cule and T. Hwa, *Phys. Rev. Lett.* **79**, 2375 (1997).
- [11] Y. Kafri, D. Mukamel, and L. Peliti, *Phys. Rev. Lett.* **85**, 4988 (2000).
- [12] N. Theodorakopoulos, T. Dauxois, and M. Peyrard, *Phys. Rev. Lett.* **85**, 6 (2000).
- [13] T. Garel and H. Orland, *Biopolymers* **75**, 453 (2004).
- [14] A. Litan and S. Lifson, *J. Chem. Phys.* **42**, 2528 (1965).
- [15] S. T. Lovett, *Mol. Microbiol.* **52**, 1243 (2004).
- [16] R. A. Neher and U. Gerland, *Phys. Rev. Lett.* **93**, 198102 (2004).
- [17] Y. Kafri, D. Mukamel, and L. Peliti, *Eur. Phys. J. B* **27**, 135 (2002).
- [18] R. A. Neher and U. Gerland (unpublished).
- [19] R. Bundschuh and T. Hwa, *Phys. Rev. E* **65**, 031903 (2002); V. Guttal and R. Bundschuh, *Phys. Rev. Lett.* **96**, 018105 (2006).
- [20] M. E. Fisher, *J. Chem. Phys.* **45**, 1469 (1966).
- [21] R. A. Neher and U. Gerland, *Biophys. J.* **89**, 3846 (2005).
- [22] For $y_0=s^{-1}$ and $x^*(y_0)<s^{-1}$, the loop size distribution decays as $\sim n^{-c}$. Since the mean loop size diverges for $c<2$, there is never a condensed phase in this case.